PRC - 2

QUANTITATIVE METHODS

The Institute of Chartered Accountants of Pakistan

CA PAKISTAN

First edition published by
The Institute of Chartered Accountants of Pakistan
Chartered Accountants Avenue
Clifton
Karachi – 75600 Pakistan
Email: studypacks@icap.org.pk
www.icap.org.pk

TABLE OF CONTENTS

CHAPTER 1

MATHEMATICAL EQUATIONS AND COORDINATE SYSTEM

IN THIS CHAPTER

AT A GLANCE

A Coordinate System is a method of representing points in a space of given dimensions. Points are positioned in a three or two dimensional plane (x and y axis).

Graphs are plots of points on a two dimensional coordinate system.

Any point representing a value of y given a value or x will fall on the line. The equations of straight lines are represented using x and y variables.

1 EXPRESSIONS AND EQUATIONS

An expression is a number, a variable, or a combination of numbers and variables and their related operation symbol(s). When this is the case, expression is separated by operations into components known as **terms**.

- *For example:*

  A simple mathematical expression can be 5+7 or 8÷3.

  Where '+' or '÷' are operations.

A polynomial is a mathematical expression consisting of a sum of terms, each term including a variable or variables raised to a power and multiplied by a coefficient. A variable is an unknown term that stands for any number in an expression and can be represented by letters say x, y, a, b or z.

- *For example:*

  $2x$ is an expression made up of a single term – a monomial.

  $2x + 3y$ is an expression made up of two terms – a binomial.

  $2x + 3y + z$ is an expression made up of three terms – a trinomial.

Algebraic expressions are used to clarify or simplify situations or mathematical calculations. For instance, if a hotel charges Rs. x per person for a room how much would it cost for 25 such persons. This problem can simply be expressed as $25x$.

When two mathematical expressions are considered equal (indicated by the sign =), they constitute an **equation.**

- *For example:*

  $2x = 8$

  $2x + 3y = 15$

  $2x + 3y + z = 54$

Equations consist of coefficients, variables as well as constants. A **variable** is a value that may change within the scope of a given problem or set of operations - often represented by letters like x, y and z. A **constant** is a value that remains unchanged (though it might be of unknown value). Unknown constants are often represented by letters like a, b and c. Finally, **coefficients** are multiples within an equation.

- *Illustration:*

  $y = 2x + c$

  Where:

  x and y are variables

  c is constant

  2 is coefficient

An equation can have one or more variables. In this equation, y is described as the dependent variable and x as the independent variable. The dependent variable changes as a result of changes in the independent variable.

It is not always easy to identify the dependent and independent variables in equations but by convention, if an equation is arranged as above, the dependent variable is the one on its own to the left hand side of the equal sign (y).

Degree of terms and equations

The degree of terms or equations, in a polynomial is the sum of all the exponents of the variables used in expression or equations (where degree means the highest power).

- *For example:*

  For $8xy^7+5x^2y^3+8x-5$ degree of polynomial is 8

  Similarly, for the equation $8x^3+6x^2+3x=0$ degree of polynomial equation is 3.

First Degree or linear equations:

A linear equation is one in which each term is either a constant or the product of a constant and a single variable. Linear equations do not contain terms that are raised to a power or reduced to a root. A linear equation is one in which the degree of the equation is one.

The equation of a straight line is a linear equation with two variables.

- *Formula: standard form*

  $y = a + mx$

  Where:

  y = dependent variable.

  x = independent variable.

  a = the value of the intercept on the y axis.

  m = the slope (gradient) of the line.

- *For example:*

  A student gets 2 points every time he answers a correct objective question. Express total score in terms of mathematical equation when the student got a bonus of 6 points for completing the paper in time.

  In order to express the above situation in mathematical equation, let's assume

  Number of Objective questions to be = x

  Total score to be = y

  Objective score + Bonus score = total score

  $(2)(x) + 6 = y$

  $2x + 6 = y$

AT A GLANCE

SPOTLIGHT

STICKY NOTES

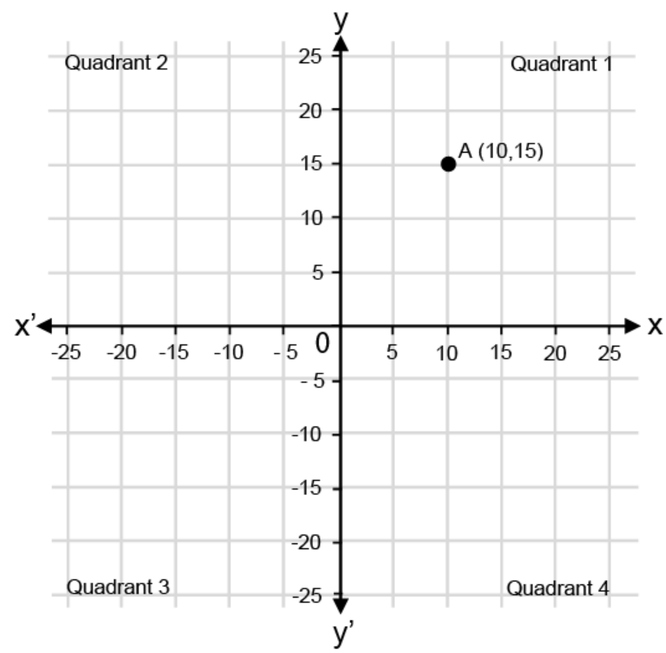
2 THE COORDINATE SYSTEM

A coordinate system is a method of representing points in a space of given dimensions by coordinates.

A coordinate is a group of numbers used to indicate the position of a point or a line.

The coordinate system is a method of specifying a location with a series of numbers which refer to other fixed points. Coordinates can specify position in a three dimensional space or on a two dimensional plane. Most graphs that you have seen are plots on a coordinate plane.

► *Illustration*



Features:

The plane has two scales called the x axis (horizontal) and y axis (vertical) which are at right angles to each other. The point where the axes intersect is called the ***point of origin*** and is usually denoted ***0***.

In the above diagram there are negative values for both x and y. This divides the plane into 4 sections called quadrants.

- Quadrant 1 where values of x and y are both positive;
- Quadrant 2 where values of x are negative but values of y are positive;
- Quadrant 3 where values of x and y are both negative; and
- Quadrant 4 where values of x are positive but values of y are negative.

Coordinates

Any point on the plane can be identified by a value of x and a value of y.

On the above diagram point A has been identified as being located at (10, 15). By convention the location in relation to the x coordinate is always given first.

The coordinates of A mean that it is located on a line at right angle from the x axis at 10 on the x axis scale and on a line at right angle from the y axis at 15 on the y axis scale. Identifying the exact location is described as plotting the point.

All graphs in this syllabus are plots of points on a two dimensional coordinate system

3 LINEAR EQUATIONS

A linear equation is one in which each term is either a constant or the product of a constant and a single variable. Linear equations do not contain terms that are raised to a power or reduced to a root. A linear equation is one in which the degree of the equation is one (where degree means the highest power). In other words a linear equation does not include terms like x^2.

The equation of a straight line is a linear equation with two variables

▶ *For example:*

Given the equation y = 3 + 5x

For various values of x, different values of y can be arrived at

Value of x	Solution	Value of y
0	= 3+5(0) = 3+0	3
1	= 3+5(1) = 3+5	8
2	= 3+5(2) = 3+10	13
3	= 3+5(3) = 3+15	18
4	= 3+5(4) = 3+20	23

Slope-Intercept form of the equation of a straight line

As it can be observed from the above table, for every increasing value of x, y increases. This can be plotted on a graph using a coordinate system.

A straight line on a graph can be a line crossing various points on a coordinate system in an increasing or decreasing manner.

Equation of a straight line (slope intercept version) would be as follows:

y = b + mx (or y = b – mx)

Where:

y = the dependent variable.

x = the independent variable.

b = the value of the intercept on the y axis.

This is the value at which the line crosses the y axis (where x=0). It could have a negative value

m = the slope (gradient) of the line.

The gradient may be positive (upward sloping from left to right), negative (downward sloping from left to right), zero (parallel to x-axis) or undefined (parallel to y-axis).

▶ *For example:*

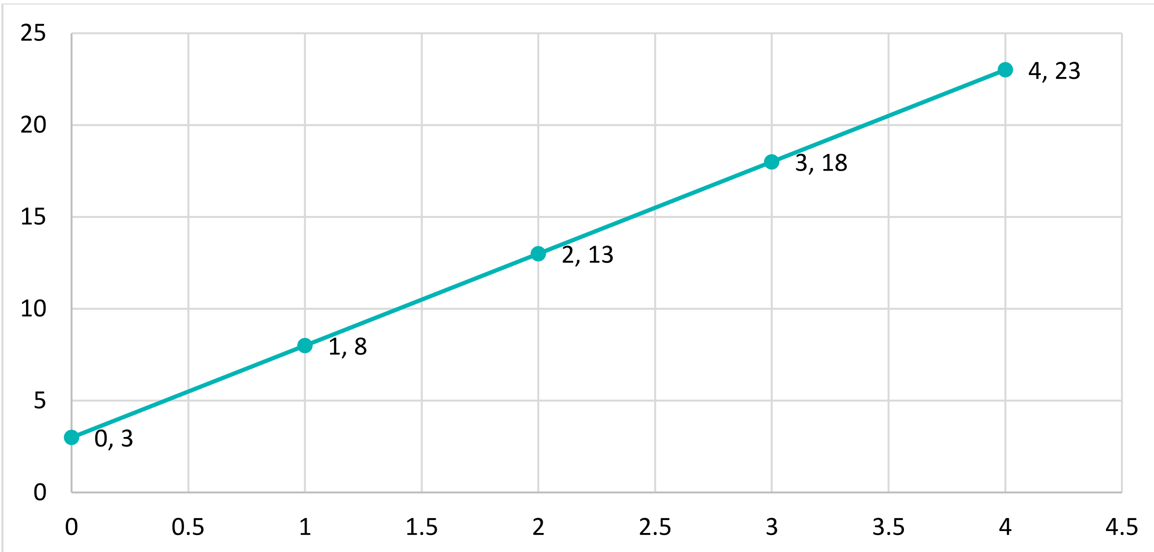
Drawing a straight line graph requires information about two coordinates or information about one coordinate and the slope of the graph.

Information about two coordinates might be given or can be estimated from the equation for the line by solving it to find two different values of y for two different values of x.

Using example above, x and y coordinates can be plotted using a two-dimensional graph as follows:

x	0	1	2	3	4
y	3	8	13	18	23

AT A GLANCE
SPOTLIGHT
STICKY NOTES



Slope (gradient)

The slope of a line is a number which describes how steep it is. It is a measure of the change in the value of the dependent variable (y) which results from a change in the independent variable.

▶ *Formula: Slope of a straight line*

$$\text{Slope} = \frac{y_2 - y_1}{x_2 - x_1}$$

Where:

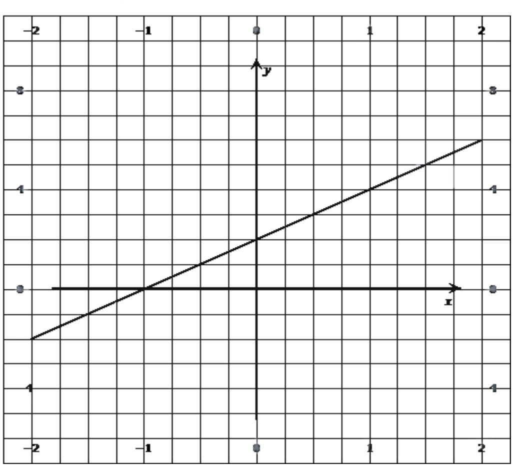
(x_1, y_1) and (x_2, y_2) are pairs of coordinates on the line

The gradient may be positive, negative, infinite (a vertical line) or zero (a flat line).

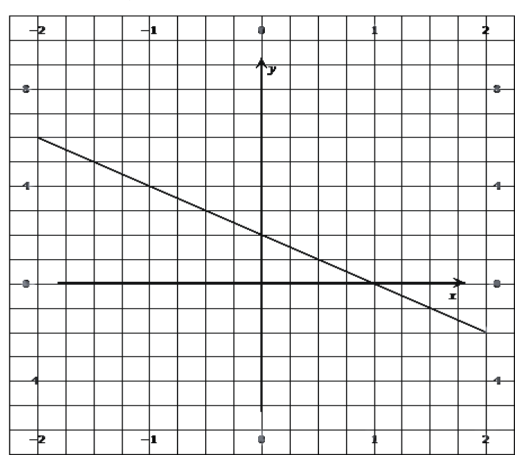
- Positive: upward sloping from left to right – increase in x causes an increase in y;
- Negative: downward sloping from left to right – increase in x causes a decrease in y

▶ *Illustration: Positive and negative slopes*

y = 2 + 2x (slope = +2)



y = 2 – 2x (slope = –2)



AT A GLANCE
SPOTLIGHT
STICKY NOTES

Point slope form

Equation of a straight line in a Point slope form

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

Where:

(x_1, y_1) = A known set of coordinates on the line

x = A chosen value

m = The slope (gradient) of the line

A straight line can be drawn with information of a single set of coordinates and the slope of the line. In order to do any other value of x is substituted into the equation.

- *For example:*

Calculate a second pair of coordinates of the line which passes through (6, 32) and has a slope of 4.

Known values in the example are

$x_1 = 6$

$y_1 = 32$

$m = 4$

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

Substitute in the known values

$y = 4(x - 6) + 32$

Pick any other value of x (say 3).

$y = 4(3 - 6) + 32$

$y = -12 + 32$

$y = 20$

The coordinates (3, 20) and (6, 32) could be used to draw the line.

The point slope form could be used to calculate the value of the intercept. To do this, set the chosen value of x to zero.

- *For example:*

Calculate the value of the y intercept of the line which passes through (6, 32) and has a slope of 4.

Known values in the example are

$x_1 = 6$

$y_1 = 32$

$m = 4$

$x = 0$

Substitute in the known values in the equation $y = m(x - x_1) + y_1$

$y = 4(x - 6) + 32$

Set the other value of x to zero

$y = 4(0 - 6) + 32$

$y = 4(-6) + 32$

$y = -24 + 32$

$y = 8$

AT A GLANCE
SPOTLIGHT
STICKY NOTES

Intercept form

This equation does not apply unless the line intercepts both the x and the y axis. Most lines do, as the intercept could be in a negative quadrant. The equation does not apply to any line that is parallel to either axis or to a line that goes through the point of origin (where the axes meet).

The equation can be used to derive the slope–intercept form.

- *Formula:*

Intercept form of equation of a straight line

Intercept form

$$\frac{x}{a} + \frac{y}{b} = 1$$

Where:

(x, y)	=	A pair of coordinates
a	=	the intercept value on the x axis
b	=	the intercept value on the y axis

- *For example:*

A line has an x intercept of 6 and a y intercept of 3. Derive the slope–intercept form of the equation of a straight line.

$$\frac{x}{a} + \frac{y}{b} = 1$$

Substitute in the known values

$$\frac{x}{6} + \frac{y}{3} = 1$$

Multiply by 18 (6 × 3) to eliminate the fractions

$$18(\frac{x}{6})+18(\frac{y}{3})=18(1)$$

3x+6y = 18

Rearrange:

6y = 18 - 3x

Divide by 6 on both sides

$$\frac{6y}{6} = \frac{18}{6} - \frac{3x}{6}$$

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

- *Practice example:*

A line with a negative slope passes through (3, 2). It has an x intercept value which is twice the y intercept value. Derive the slope–intercept form of the equation of a straight line.

$$\frac{x}{a} + \frac{y}{b} = 1$$

But a = 2b so the equation would be

$$\frac{x}{2b} + \frac{y}{b} = 1$$

Substitute in the known values

$$\frac{3}{2b}+\frac{2}{b}=1$$

Multiply by 2b

$$2b(\frac{3}{2b})+2b(\frac{2}{b})=2b(1)$$

$3+4=2\text{b}$

$7=2\text{b}$

$\text{b}=\frac{7}{2}$ or 3.5

Substituting the value of b to arrive at a

a = 2(3.5)

a = 7

Insert these values into the intercept form equation

$$\frac{x}{7}+\frac{y}{3.5}=1$$

Multiply by 7

$$y=3.5-\tfrac{1}{2}x$$

Rewrite the equation y – 4x – 5 = 0 in a slope intercept form, find slope m and y intercept b.

y - 4x - 5= 0

y - 4x = 5

y = 5 + 4x

As per the formula

y = b + mx

so b = 5 and m = 4

However, y-intercept can be calculated by keeping x = 0

y = 5 + 4(0)

y = 5

In identifying the slope, lets calculate various values of y for corresponding values of x

x	Solution	Value of y
0	= 5 + 4(0) = 5 =	5
1	= 5 + 4(1) = 5+4 =	9
2	= 5 + 4(2) = 5+8 =	13
3	= 5 + 4(3) = 5+12 =	17
4	= 5 + 4(4) = 5+16 =	21

Using the formula for slope

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

m = $\frac{9-5}{1-0}$

m = 4

It can be useful here to calculate x intercept where y=0

y = 5 + 4x

(0) = 5 + 4x

- 5 = 4x ; $-\frac{5}{4} = x$

4 SOLVING LINEAR EQUATIONS

Two or more equations are equivalent when they are said to have same solution

▶ *For example:*

Value of 'x' in the equation $2x=22$ would only be true when $x=11$. Which is also true for $2x+11=33$.

Sometimes solutions are not known for two equations. Same solution for different equations can be reached by solving equations simultaneously by way of any of the following

- Adding or subtracting a value from each side of the equation
- Multiplying or dividing a value from each side of the equation
- Simplify equations step by step using above mentioned properties
- Interchanging values or terms from both sides

In solving the equation, each successive step leads to a simpler equation which is equivalent to the earlier.

▶ *For example:*

Solve for x when 34 - 6x = 5x - 21

Adding 6x on both sides

$34 - 6x + 6x = 5x + 6x - 21$

$34 = 11x - 21$

Adding 21 on both sides

$34 + 21 = 11x – 21 + 21$

$55 = 11x$

Dividing both sides with 11

$x=\frac{55}{11}$

$x=5$

Solve for 4(z + 7) = z - 2

Expanding the brackets, the given equation can be written as

4z + 28 = z - 2

Adding 2 from both sides

4z + 28 + 2 = z - 2+2

4z + 30 = z

Subtracting z from both sides

4z - z + 30 = z - z

3z + 30 = 0

The equation can be written as

3z = - 30

Dividing both sides with 3

z = - 10

Simultaneous Linear Equations – two variables

For linear equations with two variables, the equations can be solved simultaneously by the elimination of one variable.

The aim is to manipulate one of the equations so that it contains the same coefficient of x or y as in the other equation. The two equations can then be added or subtracted from one another to leave a single unknown variable. This is then solved and the value substituted back into one of the original equations to find the other variable.

- *For example:*

Solve for $3x + 4y = 36$ and $x + 2y = 16$

Multiply second equation by 2 (to make the y coefficients the same)

$2(x + 2y) = 2(16)$

$2x + 4y = 32$

Subtract this equation from first equation

$+ 3x + 4y = + 36$

$+ 2x + 4y = + 32$

$\underline{(-) \quad (-) \quad\quad (-)}$

$\underline{\quad x \quad\quad = + 4}$

Substitute x = 4 in the first equation

$3x + 4y = 36$

$3(4) + 4y=36$

$12 + 4y=36$

Subtract 12 from both sides

$4y = 24$

Divide 4 from both sides

$y = 6$

Solve for $2x + y = 7$ and $3x + 5y = 21$

Multiply first equation with 3 and second equation with 2 (to make the x coefficients the same)

$3(2x + y) = 3(7)$ and $2(3x + 5y) = 2(21)$

$6x + 3y = 21$ and $6x + 10y = 42$

Subtracting later equation from first equation

$+ 6x + 10y = + 42$

$+ 6x + \ \ 3y = + 21$

$\underline{(-) \quad (-) \quad\quad (-) \quad}$

$\underline{\quad\quad 7y = \ \ 21}$

$y = \frac{21}{7}$

$y = 3$

substituting value of y in first equation

$2x + (3) = 7$

$2x + 3 = 7$

Subtracting 3 from both sides

$2x = 4$

Dividing 2 from both sides

$x = \frac{4}{2}; \quad x=2$

Plotting two equations

Two simultaneous equations can be plotted in a single graph.

Where two lines cross there is a single value for x and a single value for y that satisfies both equations.

Solving simultaneous equations using graphs means finding the point where two lines intersect.

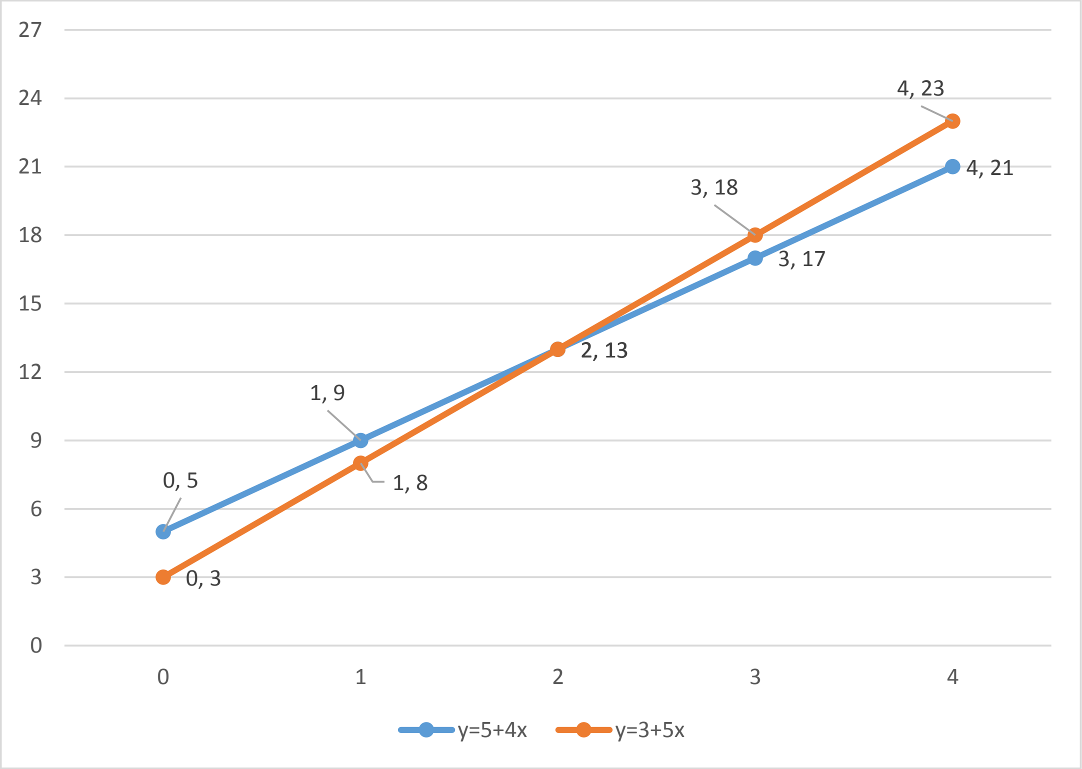
It involves finding the values of x and y which are true to both equations

- *Practice example:*

Solve for $y = 5 + 4x$ and $y = 3 + 5x$ (as above) using graphs

Following table shows values of y using multiple values of x:

$y = 5 + 4x$		$y = 3 + 5x$	
x	y	x	y
0	5	0	3
1	9	1	8
2	13	2	13
3	17	3	18
4	21	4	23



Solve for equations $3x + 4y = 36$ and $x + 2y = 16$ using graphs.

AT A GLANCE

SPOTLIGHT

STICKY NOTES

$3x + 4y = 36$	$x + 2y = 16$
$4y = 36 - 3x$	$2y = 16 - x$
$y = \frac{(36 - 3x)}{4}$	$y = \frac{(16 - x)}{2}$

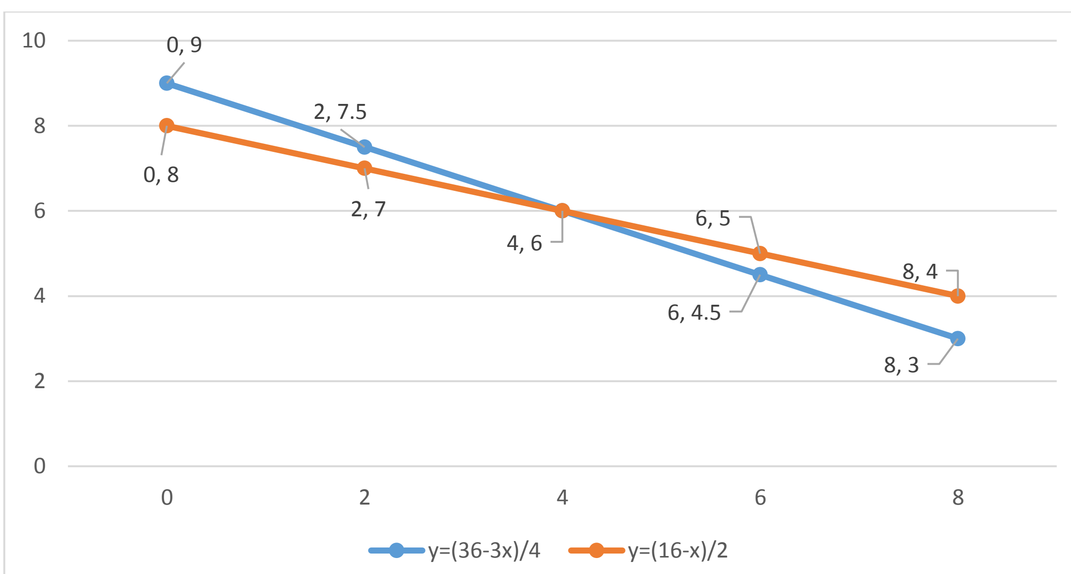
Following table shows values of y using multiple values of x:

$y = \frac{(36 - 3x)}{4}$		$y = \frac{(16 - x)}{2}$	
x	**y**	**x**	**y**
0	9	0	8
2	7.5	2	7
4	6	4	6
6	4.5	6	5
8	3	8	4

In plotting the points as above, value of x and y are

$x = 4$

y = 6



Application for Linear Equations:

Equations of straight lines are very important in accounting and other business operations. This knowledge can be applied to evaluate

- Cost behavior
- Simple Progression or prediction
- Depreciation analysis
- Interest rates and amounts
- Resource allocation
- Product pricing

Problems relating to each of the above applications can be solved using linear equations as follows:

- Thoroughly read the question with is requirements and given information
- List the unknown variables and assign appropriate letter (say x)
- Express relationship between variables as per the question or common formula.
- Solve the equation for the unknown variable (say x).
- Recheck the solution one more time.

▶ *Practice example:*

A factory pays workers at the rate of Rs. 20 per hour per person. If there are 500 workers working for 8 hours a day and 20 days a month, what is the total cost of the factory when it also pays the monthly rent Rs. 85,000/-? What would the cost be if factory reduces their employees to 240?

Rate = Rs. 20 per hour

Number of hours: 8 × 20 = 160 hours per month

Number of workers = 500

Rent = 85,000/-

Variable cost = 160 × 20 = Rs. 3,200

Equation would be

Total = Variable cost × units + Fixed cost

For total cost at 500 units

T_1 = (3,200×500) + 85,000

T_1 = 1,600,000 + 85,000

T_1 = Rs. 1,685,000

For total cost at 240 units.

T_2 = (3,200 + 240) + 85,000

T_2 = (3,200 + 240) + 85,000

T_2 = 768,000 + 85,000

T_2 = 853,000

Mr. G works as a Freelance for a News agency. He earns Rs. 2,500 per hour. One day he earned a total of Rs. 8,000, including Rs. 5,000 from tips. How many hours did he work that day?

Using the linear equation,

2,500×h + 5,000 = 8,000

Subtracting from both sides

2,500h = 8,000 - 5,000

2,500h = 7,500

h = $\frac{7,500}{2500}$

h = 3

He worked three hours that day.

A car manufacturer is building a garage. The length of the garage must be 6 times the width. If the perimeter of the garage is 1,400m, then find the length of the garage.

Using the linear equations:

Perimeter of a rectangle (P) = 2(L+B)

P = 2(6w + w)

1,400 = 12w + 2w

1,400 = 14w

Dividing both sides with 14

$\frac{1,400}{14} = w$

100=w.

Length must be 6(100) = 600m.

A raw material used for production of a chemical compound is on sale for 35% less than its usual price and if the current price is Rs. 78,000, how much would it cost for the buyer?

Using Linear Equations:

P_1 = 78,000

discount 35% or 0.35 less than the current price

P_2 = 78,000 (1 - 0.35)

P_2 = 78,000(0.65)

P_2 = 50,500

Two Swimmers are 4 km apart and moving directly towards each other. One Swimmer is moving at a speed of 1.5 kmph and the other is moving at 0.5 kmph. Assuming that the Swimmer start moving at the same time how long does it take for the two to meet??

Using Linear Equations:

Distance between two = D1 + D2 = 4

Distance = Speed × time

1.5t + 0.5t = 4

2t = 4

t = 2 hours

they will meet in 2 hours.

A company makes and sells a single product. The company is required to calculate its variable production cost of per unit. The product has a variable selling cost of Rs.1.5 per unit and a selling price of the product is Rs.6. Number of units sold are 50 and a contribution margin of 50.

Known:

Units (x) = 50

Variable selling cost (v_1) = 1.5

Selling price (s) = 6

Contribution Margin (CM) = 50

Equation would be formed as follows:

CM = R - VC

CM = (s)(x) - ($v_1 + v_2$)(x)

75 = 6 × 50 - (1.5 + v_2)(50)

75 = 300 - 75 - 50v_2

75 = 225 - 50v_2

Subtracting 225 from both sides

75 - 225 = - 50v_2

- 150= - 50v_2

Dividing both sides with (- 50)

$\frac{-150}{-50} = v_2$

3 = V_2

A company makes a single product that has a selling price of Rs.16,000 per unit and variable costs of Rs. 9,000 per unit. Fixed costs are expected to be Rs.50,000,000. What volume of sales is required to break even?

Known:

Fixed Cost = 50,000,000

Variable selling cost (v) = 9,000 per unit

Selling price (s) = 16,000 per unit

Break even sales

First let's calculate contribution margin per unit

CM = s – v

CM = 16,000 - 9,000

CM = 7,000

Breakeven point: $\frac{FC}{CM}$ per unit

BE = $\frac{50,000,000}{7,000}$

BE sales = 7,142.857 units. (7,143 units)

In sales revenue:

It is 7143×16,000= Rs.114,288,000.

AT A GLANCE

SPOTLIGHT

STICKY NOTES

A family member is 15 years younger than a second member. Three times of the age of the first member added to twice the age of second member combines to 80 years. Find age of the two family members.

Using Linear Equations:

Age of the member 1 = x and Age of the member 2 = y

As per the first statement

$x = y - 15$

then as per the second statement

$3x + 2y = 80$

Substituting first equation in the second.

$3(y - 15) + 2y = 80$

$3y - 45 + 2y = 80$

$5y - 45 = 80$

Adding 45 on both sides

$5y = 125$

Dividing both sides by 5

$y = 25$

so x would be

$x = 25 - 15$ years ; $x = 10$ years

STICKY NOTES

A coordinate system is a method of representing points in a space of given dimensions and coordinates

Expression is a number, a variable, or a combination of numbers and variables and their related operation symbol(s). When two mathematical expressions are considered equal (indicated by the sign =), they constitute an equation.

A linear equation is one in which each term is either a constant or the product of a constant and a single variable. Slope intercept form of equation of a straight line is represented by :

$$y = b + m x \text{ (or } y = b - m x\text{)}$$

Two or more equations are equivalent when they are said to have same solution. In solving the equation, each successive step leads to a simpler equation which is equivalent to the earlier. For linear equations with two variables, the equations can be solved simultaneously by the elimination of one variable.

Two simultaneous equations can be plotted in a single graph.

Where two lines cross there is a single value for x and a single value for y that satisfies both equations. Solving simultaneous equations using graphs means finding the point where two lines intersect.

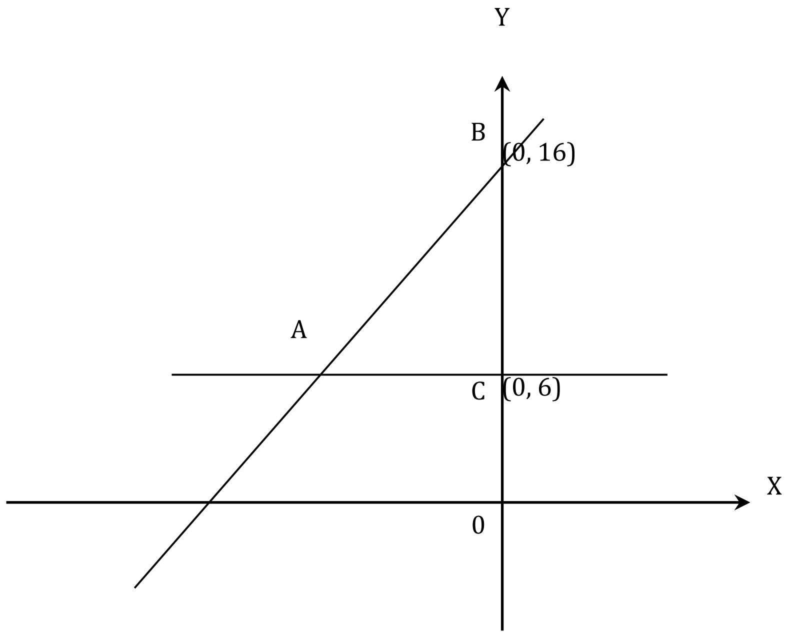
AT A GLANCE

SPOTLIGHT

STICKY NOTES

SELF-TEST

1.1. In the diagram, B is the point (0, 16) and C is the point (0, 6). The sloping line through B and the horizontal line through C meet at the point A. Then the equation of the line AC is:



(a)	$y = 6$	(b)	$x = 6$
(c)	$y = -6$	(d)	$y = 0$

1.2. A line passes through the point (0, 5) and has gradient –2. The equation of the line is:

(a)	$y = 5 + 2x$	(b)	$y = -5 - 2x$
(c)	$y = 5 - 2x$	(d)	$y = -5 + 2x$

1.3. The slope of the line perpendicular to the line $3x - 4y + 5 = 0$ is:

(a)	3/4	(b)	–3/4
(c)	4/3	(d)	–4/3

1.4. Whether the pair of lines $3x = y + 7$ and $x + 3y = 7$ are parallel, perpendicular or neither:

(a)	Perpendicular	(b)	Parallel
(c)	Neither	(d)	Not possible

1.5. y-intercept and slope of the equation $3y = 9 - 12x$ are:

(a)	y-intercept 4, slope –3	(b)	y-intercept 3, slope 4
(c)	y-intercept 3, slope –4	(d)	y-intercept –3, slope –4

1.6. y = c is the equation of straight line parallel to:

(a)	x-axis	(b)	y-axis
(c)	x, y axis	(d)	None of these

1.7. A firm's fixed costs are Rs.50,000 per week and the variable cost is Rs.10 per unit. The total cost function for the firm is:

(a)	$C(x) = 10x + 50{,}000$	(b)	$C(x) = 10x + 5{,}000$
(c)	$C(x) = 10x - 50{,}000$	(d)	$C(x) = 10x - 5{,}000$

1.8. The total cost curve of the number of copies of a particular photograph is linear. The total cost of 5 and 8 copies of a photograph are Rs.80 and Rs.116 respectively. The total cost for 10 copies of the photograph will be:

(a)	Rs.100	(b)	Rs.120
(c)	Rs.130	(d)	Rs.140

1.9. A manufacturer produces 80 T.V. sets at a cost Rs.220000 and 125 T.V. sets at a cost of Rs.287500. Assuming the cost curve to be linear. The cost of 95 sets with the help of equation of the line is:

(a)	Rs.242,600	(b)	Rs.242,500
(c)	Rs.245,500	(d)	Rs.242,400

1.10. The cost of production of a product in rupees is: $C = 15x + 9{,}750$ where x is the number of items produced. If selling price of each item is Rs.30, the sales quantity at which there would be no profit or loss is:

(a)	560 units	(b)	600 units
(c)	650 units	(d)	500 units

1.11. A manufacturer sells a product at Rs.8 per unit. Fixed cost is Rs.5,000 and the variable cost is Rs.22/9 per unit. The total output at the break-even point is:

(a)	600 units	(b)	900 units
(c)	700 units	(d)	800 units

1.12. The slope of the straight line $y = 2 - 3x$ is:

(a)	2	(b)	-3
(c)	3	(d)	-2

1.13. Pairs of coordinates (x, y) for which x and y are positive integers, such that $4x + 3y = 29$ are:

(a)	(2, 7) and (3, 5)	(b)	(7, 2) and (5, 3)
(c)	(2, 7) and (5, –3)	(d)	(2, 7) and (5, 3)

1.14. Line has equation $y = 3x + 7$ and passes through the point $(h, h + 15)$. Then the value of h is:

(a)	± 4	(b)	4
(c)	3	(d)	-4

1.15. The point (p, 2p) lies on the straight line $x + 4y = 36$. The value of p is:

(a)	4	(b)	-2
(c)	-4	(d)	2

1.16. For the profit function $P = -Q^2 + 17Q - 42$, break even points are:

(a)	13 and 4	(b)	12 and 5
(c)	14 and 6	(d)	14 and 3

1.17. The equation of line joining the point (3, 5) to the point of intersection of the lines $4x + y - 1 = 0$ and $7x - 3y - 35 = 0$ is:

(a) $2x - y = 1$ (b) $3x + 2y = 19$
(c) $12x - y - 31 = 0$ (d) None of these

1.18. A firm is introducing a new washing detergent. The firm plans to sell the family size box for Rs.24. Production estimates have shown that the variable cost of producing one unit of the product is Rs.21.60. Fixed cost of production is Rs.36,000. It is assumed that both the total revenue and total cost functions are linear over the relevant sale quantity range. Then the break-even volume of sales is:

(a) 150,000 boxes (b) 1,500 boxes
(c) 150 boxes (d) 15,000 boxes

1.19. The number of colour T.V sets sold by a firm was three times the combined sale of C.D players and radios. If the sales included 72 T.V. sets and 8 radios, then the numbers of C.D. players sold are:

(a) 15 (b) 14
(c) 16 (d) 13

1.20. $4y - x = 10$ and $3x = 2y$ then xy is:

(a) 2 (b) 3
(c) 6 (d) 12

1.21. If $7x - 5y = 13$; $2x - 7y = 26$; then $5x + 2y$ is:

(a) 11 (b) 13
(c) –11 (d) –13

1.22. Zain is x years old and his sister Ifrah is $(5x - 12)$ years old. Given that Ifrah is twice as old as Zain, then the age of Ifrah is:

(a) 8 years (b) 10 years
(c) 7 years (d) 9 years

1.23. Solution set for the simultaneous equations $2x - 3y = 19$ and $3x + 2y = -4$ is:

(a) (2, 5) (b) (–2, –5)
(c) (–2, 5) (d) (2, –5)

1.24. Ten years ago the age of a father was four times of his son. Ten years hence the age of the father will be twice that of his son. The present ages of the father and the son are:

(a) (50, 20) (b) (60, 20)
(c) (55, 25) (d) None of these

1.25. Two numbers are such that twice the greater number exceeds twice the smaller one by 18 and 1/3 of the smaller and 1/5 of the greater numbers are together 21. The numbers are:

(a) (66, 75) (b) (45, 36)
(c) (50, 41) (d) (55, 46)

1.26. A piece of iron rod costs Rs.60. If the rod was 2 metre shorter and each metre costs Re.1.00 more, the cost would remain unchanged. The length of the rod is:

(a)	13 metres	(b)	12 metres
(c)	14 metres	(d)	15 metres

1.27. Three times the square of a number when added to seven times the number results in 26. The numbers are:

(a)	–2 or –13/3	(b)	2 or –13/3
(c)	–2 or 13/3	(d)	2 or 13/3

1.28. Find out the point of intersection of following lines:

i. $3x + y = 10$

ii. $x + y = 6$

(a)	(4,2)	(b)	(2,4)
(c)	(3,3)	(d)	(1,5)

1.29. A table manufacturer has to pay Rs. 30,000 as a fixed cost and Rs. 10,000 as variable cost per table manufactured. If the Selling price of each table is Rs. 15,000 compute break-even point.

(a)	6 units	(b)	4 units
(c)	8 units	(d)	2 units

1.30. A company sells two products A and B. Total profit earned by the company in one particular period is Rs. 150,000 by selling 4 units of A and 10 units of B. The ratio of profits between A is to B is 1.5:1. Compute profit per unit from product B.

(a)	Rs. 9,375	(b)	Rs. 14,062.5
(c)	Rs. 13,062.5	(d)	Rs. 10,375

1.31. Sana distributed Rs. 1,800 amongst her three brothers in such a way that each one of them received Rs. 150 at least. What can be the maximum difference between amounts received by each brother?

(a)	Rs. 1,500	(b)	Rs. 1,350
(c)	Rs. 1,650	(d)	Not possible to determine

1.32. A manufacturer has three production workers working eight hours per day and producing twelve units in total. How many hours will this manufacturer need to operate per day if one worker leaves the company and company has to produce sixteen units in total.

(a)	8 hours	(b)	16 hours
(c)	24 hours	(d)	10 hours

1.33. A company has total budget of Rs. 105,000 to purchase tables and chairs. The cost of each table is twice that of a chair and company wants to have 3 table and 15 chairs. If the company uses its entire budget, how much more will be spent on buying chairs as compared to tables

(a)	Rs. 30,000	(b)	Rs. 75,000
(c)	Rs. 45,000	(d)	Rs. 50,000

1.34. y = 3x + c is the equation of a straight line. The line passes through the point (2,11). Determine value of y-intercept

(a)	5	(b)	6
(c)	8	(d)	10

1.35. Find the equation of a line which passes through the points (2,1) and (4,9)

(a)	y = 4x - 7	(b)	y = 2x + 7
(c)	y = 6x - 7	(d)	y = 4x + 7

1.36. The product of two positive integers is 28 and when added together the result is 11. Find the integers

(a)	2 and 14	(b)	4 and 7
(c)	28 and 1	(d)	Not possible to determine

1.37. A restaurant sells fries for Rs. 120 each, burgers for Rs. 200 each and sandwiches for Rs. 250 each. If the profit on fries is 1/3rd of its selling price, profit on burgers is Rs. 60 more than the profit on fries and profit on sandwiches is 125% of profit on burgers, what will be the total profit for the sale of 5 fries, 10 burgers and 15 sandwiches

(a)	Rs. 3,075	(b)	Rs. 2,075
(c)	Rs. 5,075	(d)	Rs. 2,175

1.38. Determine x-intercept of a line passing through the points (1,7) and (2,10)

(a)	3/4	(b)	– ¾
(c)	– 4/3	(d)	4/3

1.39. A person buys 20units of a product for Rs. 950 in total, this includes discount of 5% which is applicable if more than 10 units are purchased. Compute cost per unit if 5 units are purchased

(a)	Rs. 250	(b)	Rs. 45
(c)	Rs. 50	(d)	Rs. 200

1.40. A father gives his son certain amount of pocket money every week. Son saves 30% of amount and spends the rest, with spending on Saturday and Sunday being twice than what is spent on other days of the week. If total savings in the week are Rs. 420. How much is spent each Saturday?

(a)	Rs. 217.78	(b)	Rs. 317.78
(c)	Rs. 117.78	(d)	Rs. 17.78

1.41. The slope of the line parallel to x-axis is:

(a)	0	(b)	cannot be determined
(c)	+1	(d)	-1

1.42. The slope of the line which passes through origin and such that every coordinate has equal x and y values is:

(a)	0	(b)	1
(c)	-1	(d)	infinity

AT A GLANCE SPOTLIGHT STICKY NOTES

ANSWERS TO SELF-TEST QUESTIONS

1.1	**1.2**	**1.3**	**1.4**	**1.5**	**1.6**
(a)	(c)	(d)	(a)	(c)	(a)
1.7	**1.8**	**1.9**	**1.10**	**1.11**	**1.12**
(a)	(d)	(b)	(c)	(b)	(b)
1.13	**1.14**	**1.15**	**1.16**	**1.17**	**1.18**
(d)	(b)	(a)	(d)	(c)	(d)
1.19	**1.20**	**1.21**	**1.22**	**1.23**	**1.24**
(c)	(c)	(d)	(a)	(d)	(a)
1.25	**1.26**	**1.27**	**1.28**	**1.29**	**1.30**
(b)	(b)	(b)	(b)	(a)	(a)
1.31	**1.32**	**1.33**	**1.34**	**1.35**	**1.36**
(b)	(b)	(c)	(a)	(a)	(b)
1.37	**1.38**	**1.39**	**1.40**	**1.41**	**1.42**
(a)	(c)	(c)	(a)	(a)	(b)

CHAPTER 2

QUADRATIC EQUATIONS

IN THIS CHAPTER

AT A GLANCE

A quadratic expression is one in which the degree of variable is two (the highest power is a square). Plots of a quadratic equation are always U shaped or have an inverted U shape. Solving quadratic equations of the form $ax^2 + bx + c = 0$ involves finding the two values of x where the line cuts the x axis (where the line $y = 0$).

The possible approaches also involve factorization, completion of the square and using a formula.

Quadratic equations have various applications for its everyday uses including calculation and plotting of area and distance, forecasting and progression, interest rates and amounts and other areas.

1 SECOND DEGREE OR QUADRATIC EQUATIONS

A quadratic expression is one in which the degree of variable is two (the highest power is a square).

The most general form of a quadratic equation is written as

- *Formula: standard form*

  $y = ax^2 + bx + c$

  Where:

  a, b and c are constant (provided that a does not equal 0)

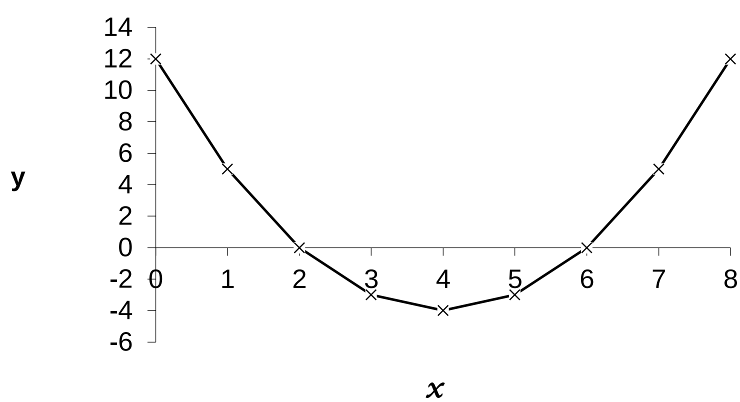
The value of a quadratic equation ($y = ax^2 + bx + c$) can be computed for different values of x and these values plotted on a graph.

- *Illustration:*

  For the equation $x^2 - 8x + 12 = y$, values can be computed as

x	x^2	$-8x$	**12**	**=**	**y**
0	0	0	**12**	**=**	12
1	1	–8	**12**	**=**	5
2	4	–16	**12**	**=**	0
3	9	–24	**12**	**=**	–3
4	16	–32	**12**	**=**	–4
5	25	–40	**12**	**=**	–3
6	36	–48	**12**	**=**	0
7	49	–56	**12**	**=**	5
8	64	–64	**12**	**=**	12

These points can be plotted to give the following graph



Plots of a quadratic equation are always U shaped or have an inverted U shape. In the above illustration the plot cuts the x axis at two places so there are two values of x that result in ***$y = x^2 - 8x + 12 = 0$***. These are ***$x = 2$*** and ***$x = 6$***.

Not all quadratic expressions cut the x axis. In this case the quadratic equation is insoluble.

Solving quadratic equations of the form $ax^2 + bx + c = 0$ involves finding the two values of x where the line cuts the x axis (where the line **y = 0**). The possible approaches involve using:

- factors of the quadratic equation (found by factorisation);
- factors of the quadratic equation (found by completion of the square); and
- a formula

▶ *For example:*

A worker painted a rectangular room. The room's length was 6m less than 3 times its width. If the area of the room was 426m2. Express dimensions in mathematical equation.

In order to express the above situation in mathematical equation, let's assume

Width of the rectangle = w

Length of the rectangle (l) = 3(w) – 6

Area of a rectangle = Length (l) × Width (w)

Expressing the same in equation

A = (3w - 6)(w)

$426 = 3w^2 - 6w$

AT A GLANCE

SPOTLIGHT

STICKY NOTES

2 SOLVING QUADRATIC EQUATIONS

Solving quadratic equations of the form $ax^2 + bx + c = 0$ involves finding the two values of x. The possible approaches involve using:

- factors of the quadratic equation (found by factorization);
- factors of the quadratic equation (found by completion of the square); and
- a formula

Solving quadratic equations by factorization

A factor is a term that multiplies something. Factorisation is about finding the terms that multiply together to give the quadratic equation. In other words, factorization involve finding the factors of the equation, and equating each factor to zero.

Factorising is the reverse of expanding brackets.

- Expanding brackets involves multiplying terms together to result in a new expression.
- Factorization starts with an expression and finds what terms could be multiplied to result in that expression.

Factorisation is like working backwards through a multiplication.

The overall approach to finding the factors of the equation differs slightly depending on whether a (i.e. the coefficient of x^2)= 1 or not.

Note that it is not always possible to factorise a quadratic equation. In such cases another approach must be used.

▶ *Illustration:*

Consider the following quadratic expression in its standard form.

$$0 = x^2 + 9x + 20$$

The coefficient of x in the quadratic expression (9) is the sum of the numbers (4 + 5) and the constant in the quadratic expression (20) is the product of the numbers in the original expressions (4 × 5).

Factorizing a quadratic expression, would then involves finding two numbers to complete the following:

$$x^2 + 9x + 20 = (x \pm ?)(x \pm ?)$$

The two numbers needed must sum to the coefficient of x and multiply to the constant. (This is only true if the coefficient of $x^2 = 1$)

The two numbers that do this are +4 and +5 so the factors become:

$$x^2 + 9x + 20 = (x + 4)(x + 5)$$

▶ *For example:*

Solve $x^2 + 5x + 6 = 0$

What two numbers add to 5 and multiply to 6? 2 and 3

Therefore the factors are: $(x + 2)(x + 3)$

This means that:

if $x^2 + 5x + 6 = (x + 2)(x + 3) = 0$

Therefore:

either $(x + 2) = 0$ or $(x + 3) = 0$

if $x + 2 = 0$

$x = -2$

if $x + 3 = 0$

$x = -3$

Therefore, $x = -2$ or -3

Solve for $x^2 - 5x - 14 = 0$

The two numbers that add to -5 and multiply to 14 are - 7 and +2.

Therefore the factors are $(x - 7)(x + 2)$

This means that:

if $x^2 - 5x - 14 = (x - 7)(x + 2) = 0$

Therefore, either $(x - 7) = 0$ or $(x + 2) = 0$

if $x - 7 = 0$ then $x = +7$

if $x + 2 = 0$ then $x = -2$

Therefore, $x = +7$ and -2

Solve for $x^2 - 9x + 20 = 0$

The two numbers that add to - 9 and multiply to 20 are - 4 and - 5.

Therefore the factors are $(x - 4)(x - 5)$

This means that:

if $x^2 - 9x + 20 = (x - 4)(x - 5) = 0$

Therefore, either $(x - 4) = 0$ or $(x - 5) = 0$

if $x - 4 = 0$ then $x = +4$

if $x - 5 = 0$ then $x = +5$

Therefore, $x = +4$ and $+5$

Factorizing quadratic expression where the coefficient of x^2 is more than 1

A quadratic equation is one which takes the form $y = ax^2 + bx + c$

There are a series of steps which must be followed.

- **Step 1:** Multiply the coefficient of x^2 by the constant (a × c).
- **Step 2:** Find two numbers that sum to the coefficient of x and multiply to the number (ac).
- **Step 3:** Rewrite the original equation by replacing the x term with two new x terms based on the two numbers.
- **Step 4:** Bracket the first two terms and the second two terms. Remember that if there is a minus sign before a set of brackets the sign inside the bracket must change.
- **Step 5:** Factor the first two terms and the last two terms separately. The aim is to produce a factor in common.
- **Step 6:** Complete the factorisation.

- *Illustration:*

Factorise: $2x^2 - 7x + 3$

Step 1: Multiply the coefficient of x^2 by the constant	$2 \times 3 = 6$
Step 2: Find two numbers that sum to –7 and multiply to 6.	The two numbers are -6 and -1
Step 3: Rewrite the original equation by replacing the x term with two new x terms based on the two numbers.	$2x^2 - 6x - 1x + 3$
Step 4: Bracket the first two terms and the second two terms (remembering to change the sign in the brackets if needed).	$(2x^2 - 6x) - (x - 3)$

Note the sign change due to bracketing the second term.

Step 5: Factor the first two terms and the last two terms separately. The aim is to produce a factor in common.	$2x(x - 3) - 1(x - 3)$

The end result is that the expression now has two x – 3 terms.

Step 6: Complete the factorisation.	$(2x - 1)\ (x - 3)$

Solving the equation:

$(2x - 1)\ (x - 3) = 0$ (therefore either $(2x - 1) = 0$ or $(x - 3) = 0$.

if $2x - 1 = 0$	if $x - 3 = 0$.
$x = 0.5$	$x = 3$

Therefore $x = 0.5$ or 3

- *For example:*

Solve $12x^2 - 20x + 3 = 0$

Step 1: Multiply the coefficient of x^2 by the constant	$12 \times 3 = 36$
Step 2: Find two numbers that sum to the coefficient of -20 and multiply to 36.	The two numbers are – 2 and –18
Step 3: Rewrite the original equation by replacing the x term with two new x terms based on the two numbers.	$12x^2 - 18x - 2x + 3 = 0$
Step 4: Bracket the first two terms and the second two terms (remembering to change the sign in the brackets if needed).	$(12x^2 - 18x) - (2x - 3) = 0$
Step 5 Factor the first two terms and the last two terms separately. The aim is to produce a factor in common.	$6x(2x - 3) - 1(2x - 3) = 0$

Note the sign change due to multiplying the second $2x - 3$ term by -2. The end result is that the equation now has a $2x - 3$ term in each half.

Step 5: Complete the factorisation.	$(6x - 1)(2x - 3) = 0$

Solving the equation:

$(6x - 1)(2x - 3) = 0$ therefore either $(6x - 1) = 0$ or $(2x - 3) = 0$.

if $6x - 1 = 0$	if $2x - 3 = 0$.
$x = {}^1/_6$	$x = 2.5$

Therefore $x = {}^1/_6$ or 2.5

Solve for $24x^2 + 10x - 4 = 0$

Step 1: Multiply the coefficient of x^2 by the constant	$24 \times (-4) = -96$
Step 2: Find two numbers that sum to the coefficient of +10 and multiply to –96.	The two numbers are +16 and – 6
Step 3: Rewrite the original equation by replacing the x term with two new x terms based on the two numbers.	$24x^2 + 16x - 6x - 4 = 0$
Step 4: Bracket the first two terms and the second two terms (remembering to change the sign in the brackets if needed).	$(24x^2 + 16x) - (6x + 4) = 0$
Step 4: Factor the first two terms and the last two terms separately. The aim is to produce a factor in common.	$4x(6x + 4) - 1(6x + 4) = 0$
Step 5: Complete the factorisation.	$(4x - 1)(6x + 4)$

Solving the equation:

$(4x - 1)(6x - 4) = 0$ therefore either $(4x - 1) = 0$ or $(6x - 4) = 0$.

if $4x - 1 = 0$	if $6x + 4 = 0$.
$x = {}^1/_4$	$x = -{}^2/_3$

Therefore $x = {}^1/_4$ or ${}^2/_3$

Solve for $5x^2 + 9x - 2 = 0$

Step 1: Multiply the coefficient of x^2 by the constant	$5 \times -2 = -10$
Step 2: Find two numbers that sum to the coefficient of +9 and multiply to –10.	The two numbers are +10 and – 1
Step 3: Rewrite the original equation by replacing the x term with two new x terms based on the two numbers.	$5x^2 + 10x - x - 2 = 0$
Step 4: Bracket the first two terms and the second two terms (remembering to change the sign in the brackets if needed).	$(5x^2 + 10x) - (x + 2) = 0$
Step 4: Factor the first two terms and the last two terms separately. The aim is to produce a factor in common.	$5x(x + 2) - 1(x + 2) = 0$
Step 5: Complete the factorisation.	$(5x - 1)(x + 2)$

Solving the equation:

$(5x - 1)(x + 2) = 0$ therefore either $(5x - 1) = 0$ or $(x + 2) = 0$.

if $5x - 1 = 0$	if $x + 2 = 0$.
$x = {}^{1}/_{5} = 0.2$	$x = -2$

Therefore $x = 0.2$ or -2

Solving quadratic equations by completing the square

Not all quadratic equations can be factorised. If this is the case another method must be used to solve the equation. One such method is known as completing the square.

In solving quadratic equation using completing the square method, the idea is to solve for a perfect square in the equation. Reference formula in this respect are

$(a + b)^2 = (a + b)(a + b) = a^2 + 2ab + b^2$

$(a - b)^2 = (a - b)(a - b) = a^2 - 2ab + b^2$

The approach is as follows:

- **Step 1:** Rearrange the equation from the format $ax^2 + bx + c = 0$ into $ax^2 + bx = -c$ (signs as appropriate).
- **Step 1b:** Divide the equation by the coefficient of x^2 if this is not one. (The method only works if the coefficient of x^2 equals unity).
- **Step 2:** Take half of the coefficient of x (including its sign).
- **Step 3:** Square this number (note that as a short cut be aware that this appears in the final equation).
- **Step 4:** Add this number to both sides of the equation.
- **Step 5:** Rearrange into the required format (by factorising the left hand side as a perfect square).

► *Illustration:*

The equation $x^2 - 6x - 21=0$ cannot be factorised but if it could be turned back into $(x - 3)^2 - 30$ then that expression could be used to solve for x.

This is what the technique of completing the squares does.

$(x - 3)^2 - 30 = 0$

Adding 30 on both sides

$(x - 3)^2 = 30$

Taking a square root on both sides

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

$x - 3 = +/- 5.477$

adding 3 on both sides

$x = +/- 5.477 + 3$

$x = -2.477$ or $+ 8.577$.

► *For example:*

Complete the square of $2x^2 - 5x - 12 = 0$

Step 1: Rearrange the equation	$2x^2 - 5x = 12$
Step 1b: Divide by the coefficient of x^2 (2)	$x^2 - 2.5x = 6$

Step 2: Take half of the coefficient of x (including its sign).	- 2.25
Step 3: Square this number.	2.5625
Step 4: Add this number to both sides of the equation.	$x^2 - 2.5x + 2.5625 = 6 + 2.5625$
Step 5: Rearrange into the required format.	$(x - 2.25)(x - 2.25) = 7.5625$ $(x - 2.25)^2 = 7.5625$
Solve the equation	$x - 2.25 = \sqrt{7.5625}$ $x = 2.25 \pm \sqrt{7.5625}$ $x = 2.25 \pm 2.75 = 4$ *or* -2.5

Solving quadratic equations by using a formula

Some quadratic equations can be difficult or impossible to factorise. The completion of squares method will always work for those quadratic equations that can be solved but there is another method which involves using a formula (which in fact completes the squares automatically).

- *Formula:*

For Quadratic equation

$$ax^2 + bx + c = 0$$

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

- *For example:*

Solve $x^2 - 5x - 14 = 0$ using the formula

a=1; b = - 5, c= - 14

Substituting the values within the formula, we get

$$x = \frac{-(-5) \pm \sqrt{(-5)^2 - 4(1)(-14)}}{2(1)}$$

$$x = \frac{5 \pm \sqrt{25 - (-56)}}{2}$$

$$x = \frac{5 \pm \sqrt{25 + 56}}{2}$$

$$x = \frac{5 \pm \sqrt{81}}{2}$$

$$x = \frac{5 \pm 9}{2}$$

$$x = \frac{5 + 9}{2} \; or \; x = \frac{5 - 9}{2}$$

$$x = \frac{14}{2} \; or \; x = \frac{-4}{2}$$

$$x = 7 \; or \; x = -2$$

Application for Quadratic Equations:

Quadratic equations when plotted on a graph are always U shaped or have an inverted U shape. This gives various applications for its everyday uses. This knowledge can be applied to evaluate

- Area and distance
- Forecasting and progression
- Interest rates and amounts
- Modeling a relationship between prices and products
- Identifying minimum and maximum – Optimization

Problems relating to quadratic functions can be solved using the same steps as of solving linear equation, however, nature of variable would have 2 as their highest degree.

- *Practice examples:*

The profit function of a company selling product J is given as P(x) = – 0.025x² + 9.85x – 60. What will be the profit of the company if none of the product J is sold on the market? What would the breakeven units of product?

In order to calculate the distance, x is the units of product J.

(i) So for no product sold in the market $x = 0$

$P(0) = -0.025(0)^2 + 9.85(0) - 60$

$P(0) = 0+0 - 60$

$P(0) = - 60$

This means the company would bear a loss of Rs. 60.

(ii) for breakeven units, the profit must be zero. That is $P(x) = 0$

Using the standard form of quadratic equation and formula for quadratic equation

$0= - 0.025x^2 + 9.85x - 60$

a= – 0.025; b =9.85, c= – 60

Substituting the values within the formula, we get

$$x = \frac{-(9.85) \pm \sqrt{(9.85)^2 - 4(-0.025)(-60)}}{2(-0.025)}$$

$$x = \frac{-9.85 \pm \sqrt{97.0225 - (6)}}{-0.05}$$

$$x = \frac{-9.85 \pm \sqrt{97.0225 - 6}}{-0.05}$$

$$x = \frac{-9.85 \pm \sqrt{92.0225}}{-0.05}$$

$$x = \frac{-9.85 \pm 9.5406}{-0.05}$$

$x = -6.189 \ or \ x = 387.812$

Approximately, 388 units of J would yield break even.

STICKY NOTES

A quadratic expression is one in which the degree of variable is two (the highest power is a square).
Formula:

standard form: $y = ax^2 + bx + c$

The value of a quadratic equation can be computed for different values of x and these values plotted on a graph.

Factorisation is like working backwards through a multiplication. There are a series of steps which must be followed.

- Step 1: Multiply the coefficient of x^2 by the constant ($a \times c$).
- Step 2: Find two numbers that sum to the coefficient of x and multiply to the number (ac).
- Step 3: Rewrite the original equation by replacing the x term with two new x terms based on the two numbers.
- Step 4: Bracket the first two terms and the second two terms. Remember that if there is a minus sign before a set of brackets the sign inside the bracket must change.
- Step 5: Factor the first two terms and the last two terms separately. The aim is to produce a factor in common.
- Step 6: Complete the factorisation.

Not all quadratic equations can be factorised. If this is the case another method must be used to solve the equation. One such method is known as completing the square.

- Step 1: Rearrange the equation from the format $ax^2 + bx + c = 0$ into $ax^2 + bx = c$ (signs as appropriate).
- Step 1b: Divide the equation by the coefficient of x^2 if this is not one. (The method only works if the coefficient of x^2 equals unity).
- Step 2: Take half of the coefficient of x (including its sign).
- Step 3: Square this number (note that as a short cut be aware that this appears in the final equation).
- Step 4: Add this number to both sides of the equation.
- Step 5: Rearrange into the required format (by factorising the left hand side as a perfect square).

Some quadratic equations can be difficult or impossible to factorise.
Formula:

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

SELF-TEST

2.1. If the solution of a quadratic equation has two values 0 and 4, the equation is:

(a) $x^2 - 4 = 0$ (b) $x^2 + 16 = 0$

(c) $x^2 - 16 = 0$ (d) $x^2 - 4x = 0$

2.2. The equation $x^2 + kx - 18 = 0$, where k is a constant, is satisfied by $x = 2$. Then the value of k is:

(a) -7 (b) 5

(c) ± 7 (d) 7

2.3. Solution of the equation $(x + 5)^2 = 16$ is:

(a) 1 or 9 (b) - 1 or - 9

(c) - 1 or 9 (d) 1 or - 9

2.4. Given that $8x^2 - 2xy - 3y^2 = 0$; then 'y' in term of 'x' is:

(a) $4x/3$ and $2x$ (b) $4x/3$ and $-2x$

(c) $-4x/3$ and $-2x$ (d) $4/3x$ and $-2x$

2.5. A quadratic function can have:

(a) Only one x-intercept (b) Two x-intercepts

(c) No x-intercepts (d) All options are possible

2.6. The profit function of a company is given as $P(x) = -0.05x^2 + 4x - 30$. What will be the profit of the company if company sells 25 units?

(a) 38.75 (b) 48.75

(c) 58.75 (d) 68.75

2.7. The profit function of a company is given as $P(x) = -0.05x^2 + 4x - 30$. What will be the profit of the company if company is not able to sell any unit?

(a) - 30 (b) 0

(c) 30 (d) Cannot be determined

2.8. Which of the following statements is/are correct?

i. Quadratic equations when plotted on a graph are always U shaped.

ii. Quadratic equations when plotted on a graph are always inverted U shaped.

iii. Quadratic equations when plotted on a graph are can either be U shaped or inverted U shaped.

iv. Quadratic equations when plotted on a graph can never be inverted U shaped.

(a) i (b) i and ii

(c) iii (d) iv

AT A GLANCE
SPOTLIGHT
STICKY NOTES

2.9. A quadratic equation is one which takes the form:

(a) $y = mx + c$ (b) $y = ax^2 + bx + c$

(c) $y = ax^3 + bx^2 + c$ (d) Any of the above

2.10. Solve: $(5x - 1)(x - 4)=0$

(a) - 1/5 and 4 (b) - 1/5 and - 4

(c) 1/5 and 4 (d) 4 only

2.11. Expand the following equation:

$(x+2)(x+3) = 0$

(a) $x+6 = 0$ (b) $x^2+5x+6 = 0$

(c) $x^2+3x+4 = 0$ (d) $x^2+6x = 0$

2.12. A quadratic expression is one in which the degree of variable is ____.

(a) One (b) Two

(c) Three (d) Four

2.13. If the solution of the equation $(kx - 3)(kx+1)=0$ has two roots 3/2 and - 1/2, the value of k is:

(a) 3 (b) -1

(c) 2 (d) -2

2.14. Roots of the equation $(ax+b)(cx+d) = 0$ are:

(a) - b/a and - d/c (b) - a/b and - c/d

(c) a/b and c/d (d) b and d

2.15. For the equation $kx^2+2k^2x+k = 0$. The possible values of k and k^2 are:

(a) 1 and 3 (b) 2 and 3

(c) 2 and 5 (d) 2 and 4

2.16. By comparing $5x^2+3x-6 = 0$ with $ax^2+bx+c = 0$ the values of a, b and c are:

(a) 5, 3 and -6 (b) 3, 5 and 6

(c) 5, 3 and 6 (d) - 5, - 3 and 6

2.17. By comparing $-5x^2-3x-6 = 0$ with $ax^2+bx+c = 0$ the values of a, b and c are:

(a) - 5, - 3 and - 6 (b) - 5, 3 and 6

(c) 5, 3 and 6 (d) - 5, 3 and -6

2.18. Number of real roots for the quadratic equation $3x^2+2x+1=0$ are:

(a) One (b) Two

(c) Three (d) None

2.19. Number of real roots for the quadratic equation - $3x^2+2x+1=0$ are:

(a)	Two	(b)	One
(c)	Three	(d)	None

2.20. Number of real roots for the quadratic equation $x^2+2x+1=0$ are:

(a)	One	(b)	Two
(c)	Three	(d)	Four

2.21. Which of the following statements is / are correct:

i. A quadratic equation will always have two equal roots.

ii. A quadratic equation will always have two roots.

iii. A quadratic equation will always have one root.

iv. A quadratic equation will either have one or two roots.

(a)	i	(b)	i and ii
(c)	All	(d)	None

2.22. Which of the following statements is / are correct:

i. The roots of a quadratic equation can never be negative.

ii. The roots of a quadratic equation can never be positive.

iii. One of the roots of a quadratic equation is always zero.

iv. A quadratic equation can have no real roots.

(a)	i	(b)	ii
(c)	iii	(d)	iv

2.23. Solve $x^2+7x+12 = 0$

(a)	-3 and 4	(b)	3 and -4
(c)	-3 and -4	(d)	3 and 4

2.24. Which of the following statements is/are correct:

i. Graph of a quadratic equation will always intersect x-axis.

ii. Graph of a quadratic equation may intersect x-axis.

iii. Graph of a quadratic equation will always have two x-intercepts.

iv. Graph of a quadratic equation will never have two x-intercepts.

(a)	i	(b)	ii
(c)	iv	(d)	iii and iv

2.25. Which of the following statements is/are correct:

i. Graph of a quadratic equation will always have a negative slope.

ii. Graph of a quadratic equation never intersects y-axis.

iii. Graph of a quadratic equation will always be U shaped.

iv. Graph of a quadratic equation may have two x-intercepts

(a)	iv	(b)	iii
(c)	ii	(d)	i

ANSWERS TO SELF-TEST QUESTIONS

2.1	**2.2**	**2.3**	**2.4**	**2.5**	**2.6**
(d)	(d)	(b)	(b)	(d)	(a)
2.7	**2.8**	**2.9**	**2.10**	**2.11**	**2.12**
(a)	(c)	(b)	(c)	(b)	(b)
2.13	**2.14**	**2.15**	**2.16**	**2.17**	**2.18**
(c)	(a)	(d)	(a)	(a)	(d)
2.19	**2.20**	**2.21**	**2.22**	**2.23**	**2.24**
(a)	(a)	(d)	(d)	(c)	(b)
2.25					
(a)					

CHAPTER 3

MATHEMATICAL PROGRESSION

IN THIS CHAPTER

AT A GLANCE

A numerical sequence is a succession of numbers each of which is formed according to a definite law that is the same throughout the sequence. An arithmetic progression is one where each term in the sequence is linked to the immediately preceding term by adding or subtracting a constant whereas, geometric progression is formed when each term in the sequence is linked by multiplying a constant. In order to identify the progression in the sequence, knowing law that links the numbers is critical and for that various mathematical formulae are used.

1. ARITHMETIC PROGRESSION

Two important sequences are formed when:

- Each term is formed by adding or subtracting a constant to the previous term (known as an arithmetic progression); and
- Each term is formed by multiplying the previous term by a constant (known as a geometric progression).

An arithmetic progression is one where each term in the sequence is linked to the immediately preceding term by adding or subtracting a constant number.

The number added or subtracted to construct the progression is known as the **common difference.** In other words, there is a common difference between each pair of successive numbers in the sequence.

▶ *For example:*

Sequence	Common difference
0,1, 2, 3, 4, 5, 6 etc.	+1
0, 3, 6, 9, 12, 15 etc.	+3
25, 20, 15, 10, 5, 0, – 5 etc.	– 5

Value of any term in an arithmetic progression

▶ *Formula:*

n^{th} term in a series can be identified using:

n^{th} term= a + (n - 1)d

Where:

a = the first term in the series

d = the common difference

▶ *For example:*

What is the 8th term? When a=10 and d=5

8^{th} term = 10 + (8 - 1)5 = 45

What is the 19th term? when a=10 and d= - 6

19^{th} term = 10 + (19 - 1) (- 6) = - 98

▶ *Practice examples:*

Calculate the 6th and 11th term of a series which starts at 7 and has a common difference of 6

a = 7 and d = 6

6^{th} term = 7 + (6 - 1)6 = 37

11^{th} term = 7 + (11 - 1)6 = 67

Calculate the 5th and 12th term of a series which starts at 8 and has a common difference of - 7.

a = 8 and d = - 7

5^{th} term = 8 + (5 - 1) (- 7) = - 20

12^{th} term = 8 + (12 - 1) (- 7) = - 69

The first two numbers in an arithmetic progression are 6 and 4. What is the 81st term?

a = 6 and d = - 2

81^{st} term = 6 + (81 - 1) (- 2) = - 154

Ali intends to start saving in July. He will set aside Rs 5,000 in the first month and increase this by Rs 300 in each of the subsequent months. How much will he set aside in March next year?

a = 5,000 and d = 300

n = 9 (March is 9 months after July)

9^{th} term = 5,000 + (9 – 1)300 = Rs 7,400

Sum of number of terms in an arithmetic progression

The sum of all the terms of an arithmetic sequence is called an arithmetic series.

There are two methods of finding the sum of a number of terms in an arithmetic progression.

- *Formula:*

Sum of terms in an arithmetic progression (Method 1)

$$s = \frac{n}{2}\{2a + (n - 1)d\}$$

Where:

a = the first term in the series

d = the common difference

n = total count of terms

Sum of terms in an arithmetic progression (Method 2)

$$s = \frac{1}{2}n(a + l)$$

Where:

a = the first term in the series

l = the last term in a series

n = total count of terms

- *For example:*

What is the sum of the first 8 terms? When a = 10; d = 5

Method 1:

$$s = \frac{8}{2}\{2 \times 10 + (8 - 1)5\}$$

s = 4 (20 +35)

s = 220

Method 2:

<u>Step 1</u> – find the 8^{th} term

nth term = a + (n - 1)d

8th term = 10 + (8 - 1)5 = 45

<u>Step 2</u> – sum the terms

s = ½ n(a + l)

s = ½ 8(10 + 45) = 220

What is the sum of the first 19 terms? a = 10; d = - 6

Method 1:

$$s = \frac{19}{2}\{2 \times 10 + (19 - 1)6\}$$

s = 9.5 {(20) + (18) × - 6}

s = 9.5 (20 - 108)

s = - 836

Method 2:

Step 1 – find the 19^{th} term

nth term = a + (n - 1)d

19^{th} term = 10 + (19 - 1) - 6 = 10 - 108 = - 98

Step 2 – sum the terms

s = ½ n(a + l)

s = ½ 19(10 + (- 98)) = ½ (- 1672) = - 836

Problem of a more complex nature and practical applications

So far the above equations have been used to solve for values of *s*. A question might require to solve for *a*, *d* or *n* instead. Using the equation to solve for *a* or *d* is relatively straight forward. However, a quadratic equation results when solving for *n*. This must be solved in the usual way.

- *For example:*

Solve for n When the first term is 10 and the series has a common difference of 2. How many terms are needed to produce a sum of 252?

$$s = \frac{n}{2}\{2a + (n - 1)d\}$$

$$252 = \frac{n}{2}\{2 \times 10 + (n - 1)2\}$$

Multiply by 2 on both sides

504 = n{(2 × 10) + (n - 1)2}

504 = n(20 + 2n - 2)

504 = 18n + $2n^2$

Rearrange in the form of quadratic equation

$2n^2$ + 18n - 504 = 0

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

a=2; b=18, c= - 504

$$n = \frac{-18 \pm \sqrt{18^2 - 4 \times 2 \times -504}}{2 \times 2}$$

$$n = \frac{-18 \pm \sqrt{324 - (-4032)}}{4}$$

$$n = \frac{-18 \pm \sqrt{324 + 4032}}{4}$$

$$n = \frac{-18 \pm \sqrt{4356}}{4}$$

$$n = \frac{-18 \pm 66}{4}$$

$$n = \frac{-18 - 66}{4} \quad n = \frac{-18 + 66}{4}$$

$$n = \frac{-84}{4} \quad n = \frac{48}{4}$$

$$n = -21 \quad n = 12$$

The value cannot be -21 in the context of this problem n = 12

The first term is 10 and the series has a common difference of 5. How many terms are needed to produce a sum of 325?

$$s = \frac{n}{2}\{2a + (n - 1)d\}$$

$$325 = \frac{n}{2}\{2 \times 10 + (n - 1)5\}$$

Multiply 2 on both sides

650 = n{(2 × 10) + (n - 1)5}

650 = n(20 + 5n – 5)

650 = 15n + $5n^2$

Rearranging the same as standard quadratic equation

$5n^2$ + 15n – 650 = 0

Solve for n using the quadratic equation

Where a = 5; b=15; c = -650

$$n = \frac{-15 \pm \sqrt{15^2 - 4 \times 5 \times -650}}{2 \times 5}$$

$$n = \frac{-15 \pm \sqrt{225 - (-13000)}}{10}$$

$$n = \frac{-15 \pm \sqrt{225 + 13000}}{10}$$

$$n = \frac{-15 \pm \sqrt{13225}}{10}$$

$$n = \frac{-15 \pm 115}{10}$$

$$n = \frac{-15 - 115}{10} \quad n = \frac{-15 + 115}{10}$$

$$n = \frac{-130}{10} \quad n = \frac{100}{10}$$

$$n = -13 \quad n = 10$$

The value cannot be -13 in the context of this problem n = 10

Mathematical progression has its implications for various accounting and other everyday problems including identifying installments, determining the time sequence and even locating a number in a series. For such situations, formula for determining the n^{th} term be same but now it would be used to solve for a particular number or amount in a given situation.

AT A GLANCE

SPOTLIGHT

STICKY NOTES

- *For example:*

A person sets aside Rs 1,000 in the first month, 1,100, in the second, 1,200 in the third and so on. How many months will it take to save Rs 9,100?

Here in this situation, a= 1000, and d = 100 and s=9,100

$$\boldsymbol{s = \frac{n}{2}\{2a + (n - 1)d\}}$$

9,100 = n/2{(2 × 1,000) + (n - 1)100}

Multiplying both sides with 2

18,200 = n{(2 × 1,000) + (n - 1)100}

18,200 = n(2,000 + 100n - 100)

18,200 = 1,900n + $100n^2$

Rearranging in the standard quadratic equation format

$100n^2$ + 1,900n - 18,200 = 0

Where a = 100; b = 1900; c = -18200

$$n = \frac{-1{,}900 \pm \sqrt{1{,}900^2 - 4 \times 100 \ \times -18{,}200}}{2 \times 100}$$

$$n = \frac{-1{,}900 \pm \sqrt{3610000 - (-7280000)}}{200}$$

$$n = \frac{-1{,}900 \pm \sqrt{3610000 + 7280000}}{2 \times 100}$$

$$n = \frac{-1{,}900 \pm \sqrt{10890000}}{200}$$

$$n = \frac{-1{,}900 \pm 3300}{200}$$

$$n = \frac{-1900 - 3300}{200} \quad n = \frac{-1900 + 3300}{200}$$

$$n = \frac{-5200}{200} \quad n = \frac{1400}{200}$$

$$n = -26 \quad n = 7$$

The value cannot be -26 in the context of this problem n = 7

It will take 7 months to save the desired sum.

A company is funding a major expansion which will take place over a period of 18 months. It will commit Rs. 100,000 to the project in the first month followed by a series of monthly payments increasing by Rs. 10,000 per month.

What is the monthly payment in the last month of the project?

What is the total amount of funds invested by the end of the project?

In order to find the 18^{th} payment:

n = 18; d = 10,000; and a = 100,000.

n^{th} term = a + (n - 1)d

18^{th} term = 100,000 + (18 - 1)10,000

18^{th} term = 100,000+170,000

18^{th} term = Rs.270,000

For total investment in the 18 months

$$s = \frac{n}{2}\{2a + (n - 1)d\}$$

$$s = \frac{18}{2}\{2(100{,}000) + (18 - 1)10{,}000\}$$

$$s = \frac{18}{2}\{200{,}000 + (17)10{,}000\}$$

$$s = \frac{18}{2}\{200{,}000 + 170{,}000\}$$

$$s = \frac{18}{2}\{370{,}000\}$$

s = 9{370,000}

s = Rs. 3,330,000

A company has spare capacity of 80,000 hours per month. (This means that it has a workforce which could work an extra 80,000 hours if needed).

The company has just won a new contract which requires an extra 50,000 hours in the first month. This figure will increase by 5,000 hours per month.

The company currently has enough raw material to service 300,000 hours of production on the new contract. It does not currently use this material on any other contract. January is the first month of the contract

In what month will the company need to hire extra labour?

In what month will the existing inventory of raw material run out?

In order to find the month (n) when 80,000 extra hours would be reached.

Here: d=5000; and a= 50,000.

80,000 = a+(n - 1)d

80,000 = 50,000 + (n - 1)5,000

80,000 = 50,000 +5,000n - 5,000

80000 = 45000 + 5000n

80000 - 45000 = 5000n

35000 = 5000n

35000/5000 = n

7 = n

The company will need to hire extra labour in August.

In order to determine in which month (n) existing inventory of raw material will run out, we use the equation:

$$s = \frac{n}{2}\{2a + (n - 1)d\}$$

$$300{,}000 = \frac{n}{2}\{2(50{,}000) + (n - 1)5{,}000\}$$

$$300{,}000 = \frac{n}{2}\{(100{,}000) + 5000\text{n} - 5{,}000\}$$

$$300{,}000 = \frac{n}{2}\{95{,}000 + 5{,}000\text{n}\}$$

600,000 = n(95,000 + 5,000n)

600,000 = 95,000n + 5,000n^2

Rearranging in the standard quadratic equation format

5,000n^2 + 95,000n - 600,000 = 0

Dividing the equation by 5000 we get

n^2 + 19n - 120 = 0

Solving the equation using formula

Where a = 1; b=19; c= - 120

$$n = \frac{-19 \pm \sqrt{19^2 - 4 \times 1 \ \times -120}}{2 \times 1}$$

$$n = \frac{-19 \pm \sqrt{361 - (-480)}}{2}$$

$$n = \frac{-19 \pm \sqrt{361 + 480}}{2}$$

$$n = \frac{-19 \pm \sqrt{841}}{2}$$

$$n = \frac{-19 \pm 29}{2}$$

$$n = \frac{-48}{2} \ n = \frac{10}{2}$$

$$n = -24 \ n = 5$$

The value cannot be -24 in the context of this problem n = 5

The supply of inventory will be used up in May.

A property owner has to repay his load of Rs. 425,000. He can pay Rs. 20,000 in the first year and then increases payment by Rs. 5,000 in every installment. How many installments will it take for him to clear his loan?

In order to determine in number of installments to reach the sum, we use the equation:

$$s = \frac{n}{2}\{2a + (n\ - 1)d\}$$

Where a = 20,000; d = 5000 and s = 425000

$$425{,}000 = \frac{n}{2}\{2(20{,}000) + (n\ - 1)5{,}000\}$$

$$425000 = \frac{n}{2}\{(40{,}000) + 5{,}000n - 5{,}000\}$$

$$425000 = \frac{n}{2}\{35000 + 5000n\}$$

Multiplying 2 on both sides

850,000 = n(35,000 + 5,000n)

850,000 = 35,000n + 5,000n^2

Rearranging in the standard quadratic equation format

5000n^2 + 35,000n – 850,000 = 0

Dividing the equation by 5,000 we get

n^2 + 7n – 170 = 0

Solving the equation using formula

Where a = 1; b=7; c=-170

$$n = \frac{-7 \pm \sqrt{7^2 - 4 \times 1\ \times -170}}{2 \times 1}$$

$$n = \frac{-7 \pm \sqrt{49 - (-680)}}{2}$$

$$n = \frac{-7 \pm \sqrt{49 + 680}}{2}$$

$$n = \frac{-7 \pm \sqrt{729}}{2}$$

$$n = \frac{-7 \pm 27}{2}$$

$$n = \frac{-7 - 27}{2}\ n = \frac{-7 + 27}{2}$$

$$n = \frac{-34}{2}\ n = \frac{20}{2}$$

$$n = -17\ n = 10$$

It will take 10 annual installments to reach the loan amount.

A firm agrees to hire an employee at the pay of Rs. 65,000 per month. Annual increase in salary would be 78,000. However, it does not agree to increase the salary after it reaches Rs. 1,638,000 per year. How long would it take the employee to exhaust his increment?

In order to determine in number of year to reach the sum, we use the equation:

Where annual salary = Rs 65,000 x 12 = Rs 780,000

Nth term = a + (n - 1)d

1,638,000 = 780,000 + (n - 1)78,000

1,638,000 = 780,000+78,000n - 78,000

1,638,000 = 702,000+78,000n

936,000 = 78,000n

12 = n

Employee will be able to get 12 such annual increments.

Mr. Z has opted for a loan from the bank. In his tenth installment he paid Rs. 25,000 and in the fifteenth he paid Rs. 30,000. How much he paid in the first month? And how many installments it would reach until it does not exceed Rs. 72,000.

In order to determine the first installment, we use the equation:

N^{th} term = a + (n - 1) d and solve it simultaneously

N^{th} $term_1$ = a + (n_1 - 1) d	Nth $term_2$ = a + (n_2 - 1) d
25,000 = a + (10 - 1)d	30,000 = a + (15 - 1)d
25,000 = a + 9d	30,000 = a + 14d

Eliminating a from both the equations

25,000 - 9d = a	30,000 - 14d = a

Solving it simultaneously we get

25,000 - 9d = 30,000 - 14d

14d - 9d = 30,000 - 25,000

5d = 5,000

d = 1,000

Substituting d in any of the above equation to solve for the 'a' we get:

25,000 - 9(1,000) = a

25,000 - 9,000 = a

16,000 = a

First installment was Rs. 16,000

Now if required to calculate the time:

72,000 = 16,000 + (n - 1) 1,000

7,2000 = 16,000 + 1,000n - 1,000

72,000 = 15,000 + 1,000n

72,000 - 15,000 = 1,000n

57,000 = 1,000n

57 = n

It would require 57 installments to reach that limit.

Ms. A's annual installment for the loan taken increases by the same amount every year. At its fourth installment, she has to pay Rs. 3,700. In her thirteenth, she has to pay Rs. 6,400. How much she will have to pay in her last twentieth installment? What is the total amount paid?

In order to determine the first installment, we use the equation:

N^{th} term = a + (n - 1) d and solve it simultaneously

N^{th} term$_1$ = a + (n_1 - 1) d	N^{th} term$_2$ = a + (n_2 - 1) d
3,700 = a + (4 - 1)d	6,400 = a + (13 - 1)d
3,700 = a + 3d	6,400 = a + 12d

Eliminating a from both the equations

3,700 – 3d = a	6,400 – 12d = a

Solving it simultaneously we get

3,700 – 3d = 6,400 – 12d

12d – 3d = 6,400 – 3,700

9d = 2,700

d = 300

Substituting d in any of the above equation to solve for the 'a' we get:

3700 – 3(300) = a

3700 – 900 = a

2,800= a

First installment was Rs. 2,800

Now if we are required to calculate twentieth the time:

20^{th} installment = 2,800 + (20 - 1) 300

20^{th} installment = 2,800 + (19) 300

20^{th} installment = 2,800 + 5,700

20^{th} installment = 2,800 + 5,700

20^{th} installment = 8,500

Her 20^{th} installment would be Rs. 8500.

In solving for the loan amount, we would use the equation

$$s = \frac{n}{2}\{2a + (n - 1)d\}$$

Where a=2,800; d=300 and n= 20

$$s = \frac{20}{2}\{2(2{,}800) + (20 - 1)300\}$$

$$s = \frac{20}{2}\{5{,}600 + (19)300\}$$

$$s = \frac{20}{2}\{5{,}600 + 5{,}700\}$$

$$s = 10(11{,}300)$$

$$s = 113{,}000$$

She in all paid Rs. 113,000 in her installments.

2. GEOMETRIC PROGRESSION

A sequence in which each term is obtained by multiplying the previous term by a constant. (Neither the first term in the series nor the constant can be zero).

A geometric progression is one where the ratio between a term and the one that immediately precedes it is constant throughout the whole series.

- *For example:*

Sequence	Common ratio
1, 2, 4, 8, 16, 32, 64, 128 etc.	2
1, -2, 4, -8, 16, -32, 64, -128 etc.	-2
1, ½, ¼, $^1/_8$, $^1/_{16}$ etc.	½

The ratio that links consecutive numbers in the series is known as the ***common ratio.***

The common ratio can be found by dividing a term from the term preceding it in the sequence. Note that for the three consecutive terms x, y and z in a geometric progression:

$$\frac{y}{x} = \frac{z}{y}$$

- *For example:*

Series:	Common ratio:
6, -12, 24, -48…..	$-^{12}/_6 = -2$ or $-^{48}/_{24} = -2$

Value of any term in a geometric progression and their sum

The following expression can be used to give the value of any term in a geometric progression.

- *Formula: nth term in a geometric progression*

$$n^{th}\ term = ar^{n-1}$$

Where:

a = the first term in the sequence

r = the common ratio

- *Formula: Sum of a number of terms in a geometric series.*

$$s = \frac{a(r^n - 1)}{r - 1} \qquad s = \frac{a(1 - r^n)}{1 - r}$$

Where:

a = the first term in the series

r = the common ratio

n = total number of terms

- *For example:*

What is the 9th term and their sum? When a= 3; r = 2

$n^{th}\ term = ar^{n-1}$

9^{th} term = $3 \times 2^{9-1}$

9^{th} term = 3 × 256 = 768

In order to find the sum we use the formula:

$$s = \frac{a(r^n - 1)}{r - 1}$$

$$s = \frac{3(2^9 - 1)}{2 - 1}$$

$$s = \frac{3(512 - 1)}{1}$$

$$s = 3(511)$$

$$s = 1{,}533$$

What is the 9th term and their sum? When a = 100; r = 0.5

$$n^{th}\ term = ar^{n-1}$$

9^{th} term = $100 \times 0.5^{9-1}$

9^{th} term = = 0.390625

In order to find the sum we use the formula:

$$s = \frac{a(1 - r^n)}{1 - r}$$

$$s = \frac{100(1 - 0.5^9)}{1 - 0.5}$$

$$s = \frac{100(1 - 0.001953125)}{0.5}$$

$$s = \frac{100(0.998047)}{0.5}$$

$$s = \frac{99.8047}{0.5}$$

$$s = 199.6094$$

Find the 10th term of the geometric progression, 5, 10, 20... and their sum

$$n^{th}\ term = ar^{n-1}$$

Here a=5, r=2

10^{th} term = $5 \times 2^{10-1}$

10^{th} term = 5 × 512 = 2,560

In order to find the sum, we use the formula:

$$\boldsymbol{s = \frac{a(r^n - 1)}{r - 1}}$$

$$s = \frac{5(2^{10} - 1)}{2 - 1}$$

$$s = \frac{5(1024 - 1)}{1}$$

$$s = 5(1{,}023)$$

$$s = 5{,}115$$

AT A GLANCE

SPOTLIGHT

STICKY NOTES

Find the 12th term of the geometric progression, 5, –10, 20, –40 ……..

To find the common ratio:

$r = \frac{-10}{5} = 2$

$n^{th}\ term = ar^{n-1}$

Here, a=5, r = -2

12^{th} term = 5 × (-2^{12-1})

12^{th} term = 5 × (-2^{11})

12^{th} term = 5 × (-2048) = -10240

In order to find the sum we use the formula:

$$s = \frac{a(1 - r^n)}{1 - r}$$

$$s = \frac{5(1 - (-2)^{12})}{1 - (-2)}$$

$$s = \frac{5(1 - 4{,}096)}{3}$$

$$s = \frac{5(-4{,}095)}{3}$$

$$s = \frac{-2{,}0475}{3}$$

$$s = -6{,}825$$

Find the 6th term of the geometric progression, 1,000, 800, 640…

To find the common ratio:

$r = \frac{800}{1000} = 0.8$

$n^{th}\ term = ar^{n-1}$

Here, a = 1,000; r =0.8

6th term = 1,000 × 0.8^{6-1}

6th term = 1,000 × 0.32768 = 327.68

In order to find the sum, we use the formula:

$$s = \frac{a(1 - r^n)}{1 - r}$$

$$s = \frac{1{,}000(1 - 0.8^6)}{1 - 0.8}$$

$$s = \frac{1{,}000(1 - 0.262144)}{0.2}$$

$$s = \frac{1{,}000(0.737856)}{0.2}$$

$$s = \frac{737.856}{0.2}$$

$$s = 3{,}689.28$$

Infinite series

A sequence may be infinite means there would not be any last term. With common ratio greater than 1, the sum of infinite sequence be a larger number upto infinity. If the common ratio is less than 1 (|r| <1), the sum of infinite series may be found.

- *Formula:*

As the number of terms approaches infinity n becomes very large and rn becomes very small so that it can be ignored.

In other words, as n approaches infinity then rn approaches zero.

$$s = \frac{a}{1 - r}$$

Where:

a = the first term in the sequence

r = the common ratio

- *For example:*

Find the sum of the given geometric series:

$$\frac{\mathbf{1}}{\mathbf{2}} + \frac{\mathbf{1}}{\mathbf{4}} + \frac{\mathbf{1}}{\mathbf{8}} + \frac{\mathbf{1}}{\mathbf{16}} + \cdots \frac{\mathbf{1}}{\boldsymbol{n}}$$

To find the common ratio r

¼ ÷ ½ = ½

$$s = \frac{a}{1 - r}$$

$$s = \frac{\frac{1}{2}}{1 - \frac{1}{2}}$$

$$s = \frac{0.5}{0.5} = 1$$

Find the sum of the given geometric series:

6, - 1, $^1/_6$, - $^1/_{36}$...

To find the common ratio r

–1 ÷ 6 = - $^1/_6$

$$s = \frac{a}{1 - r}$$

$$s = \frac{6}{1 - (-\frac{1}{6})}$$

$$s = \frac{6}{1 + 1/6}$$

$$s = \frac{6}{1\frac{1}{6}} = 5.143$$

AT A GLANCE

SPOTLIGHT

STICKY NOTES

Problem of a more complex nature and practical applications

For geometrical sequence, questions of more complexity can require solving for the term or sum of in finite terms. In this respect, knowledge and applications for log and exponential functions would be used.

- *For example:*

What is the last term in the sequence 1.2, 1.44, ... , 2.0736. Also find its sum.

$n^{th}\ term = ar^{n-1}$

a = 1.2; r= 1.44/1.2 = 1.2

nth term = $1.2 \times 1.2^{n-1}$

$2.0736 = 1.2 \times 1.2^{n} \times 1.2^{-1}$

$$2.0736 = \frac{1.2 \times 1.2^{n}}{1.2}$$

$2.0736 = 1.2^{n}$

This can either be solved using log or substituting for n

$1.2^{4} = 1.2^{n}$

n = 4

In order to find the sum, we use the formula:

$$s = \frac{a(r^{n}-1)}{r-1}$$

$$s = \frac{1.2(1.2^{4}-1)}{1.2-1}$$

$$s = \frac{1.2(2.0736-1)}{0.2}$$

$$s = \frac{1.2(1.0736)}{0.2}$$

$$s = \frac{1.28832}{0.2}$$

$$s = 6.4416$$

What is the last term in the sequence 240, 144, ... , 4.0310784. Also find its sum.

$n^{th}\ term = ar^{n-1}$

a = 240; r= 144/240 = 0.6

nth term = $240 \times 0.6^{n-1}$

$4.0310784 = 240 \times 0.6^{n} \times 0.6^{-1}$

$$4.0310784 = \frac{240 \times 0.6^{n}}{0.6}$$

$$4.0310784 = 400 \times 0.6^{n}$$

$0.0100777 = 0.6^{n}$

This can either be solved using log or substituting for n

$0.6^{9} = 0.6^{n}$

n = 9

In order to find the sum, we use the formula:

$$s = \frac{a(1 - r^n)}{1 - r}$$

$$s = \frac{240(1 - 0.6^9)}{1 - 0.6}$$

$$s = \frac{240(0.989922)}{0.4}$$

$$s = \frac{237.581}{0.4}$$

$$s = 593.95338$$

Geometrical progression has also its implications to everyday problems. Some of the problems involve solving for the nth term while others would require summing up the potential amount.

- *For example:*

Mr K. is planning for a loan from a bank. The installment increases 25% every year. If the installment starting in 2020 is Rs. 4,000, what would be the installment in the year 2027.

If the installment increases 25% each year then the common ratio is 1.25.

Year 2020 : 4000

Year 2021 : 4000 + 4000(0.25) = 5000

Common Ratio = 5000/4000 = 1.25

$n^{th}\ term = ar^{n-1}$

a = 4000; r= 1.25; n = 8

2027 = 8th term = 4000 × 1.25^{8-1}

2027 = 8th term = 4000 × 1.25^{7}

2027 = 8th term = 4000 × 4.7684

2027 = 8th term = 19,073.6

Installment in the year 2027 would be Rs. 19,073.6

A company earns a profit of Rs. 75000 in its first year. In the second year, it made Rs. 86250. If the profit continues to grow in geometric sequence, what will be the profit in its 4th year. What will be the total earned profit.

If the installment increases 15% each year then the common ratio is 1.15.

Common Ratio = 86,250/75,000 = 1.15

$n^{th}\ term = ar^{n-1}$

a = 75000; r= 1.15; n = 4

4th term = 75,000 × 1.15^{4-1}

4th term = 75,000 × 1.15^{3}

4th term = 75,000 × 1.520875

4th term = 114,065.625

Profit in the fourth year would be Rs. 114,065.625

AT A GLANCE
SPOTLIGHT
STICKY NOTES

$$s = \frac{a(r^n - 1)}{r - 1}$$

$$s = \frac{75{,}000(1.15^4 - 1)}{1.15 - 1}$$

$$s = \frac{75{,}000(1.749 - 1)}{0.15}$$

$$s = \frac{75{,}000(0.749)}{0.15}$$

$$s = \frac{56{,}175}{0.15}$$

$$s = 374{,}500$$

Total profit earned in four years would be Rs. 374,500 approximately.

STICKY NOTES

An arithmetic progression is one where each term in the sequence is linked to the immediately preceding term by adding or subtracting a constant number.

Value of any term in an arithmetic progression can be found by:

nth term= a+(n-1)d

The sum of all the terms of an arithmetic sequence is called an arithmetic series. Sum of terms in an arithmetic progression can be found by:

$s = \frac{n}{2}\{2a + (n - 1)d\}$ **and** $s = \frac{1}{2}n(a + l)$

A geometric progression is one where the ratio between a term and the one that immediately precedes it is constant throughout the whole series.

For nth term in a geometric progression

$n^{th}\ term = ar^{n-1}$

Sum of a number of terms in a geometric series can be found using the formula:

$$s = \frac{a(r^n - 1)}{r - 1}$$

A sequence may be infinite means there would not be any last term. As the number of terms approaches infinity n becomes very large and r^n becomes very small so that it can be ignored. Therefore, in order to find the sum following formula can be used.

$$s = \frac{a}{1 - r}$$

SELF-TEST

3.1. 7^{th} term of an A.P. 8, 5, 2, −1, −4... is:

(a)	9	(b)	10
(c)	−9	(d)	−10

3.2. (a – b), a, (a + b) are in ______ progression.

(a)	Geometric	(b)	Arithmetic
(c)	Harmonic	(d)	None of these

3.3. The sum of the series 1 + 2 + 3 .... + n is:

(a)	(n + 2)/2	(b)	n(n – 1)/2
(c)	n(n + 1)/2	(d)	None of these

3.4. 2 + 4 + 6 + ... + 100 = ______.

(a)	2,500	(b)	2,550
(c)	2,575	(d)	None of these

3.5. Which term of the AP $\frac{3}{\sqrt{7}}, \frac{4}{\sqrt{7}}, \frac{5}{\sqrt{7}}$ ...... is $\frac{17}{\sqrt{7}}$?

(a)	15	(b)	17
(c)	16	(d)	18

3.6. The value of x such that $8x + 4$, $6x - 2$, $2x + 7$ will form an AP is:

(a)	15	(b)	2
(c)	15/2	(d)	None of these

3.7. The sum of a certain number of terms of an AP series −8, −6, −4, ... is 52. The number of terms is:

(a)	12	(b)	13
(c)	11	(d)	None of these

3.8. The 7^{th} term of the series 6, 12, 24...... is:

(a)	384	(b)	834
(c)	438	(d)	None of these

3.9. The last term of the series x^2, x, 1,...... to 31 terms is:

(a)	x^{28}	(b)	$1/x$
(c)	$1/x^{28}$	(d)	None of these

3.10. If the terms $2x$, $(x + 10)$ and $(3x + 2)$ be in A.P., the value of x is:

(a)	7	(b)	10
(c)	6	(d)	None of these

3.11. The sum of all odd numbers between 200 and 300 is:

(a)	11,600	(b)	12,490
(c)	12,500	(d)	24,750

3.12. The sum of the series 1, $\frac{3}{5}, \frac{9}{25}, \frac{27}{125}$ ...... to infinity is:

(a)	2/5	(b)	3/2
(c)	-2/5	(d)	5/2

3.13. Sum of the series $6 + \frac{3}{2} + \frac{3}{8} + \ldots\ldots\ldots$ to infinity is:

(a)	6	(b)	8
(c)	7	(d)	9

3.14. The sum of the infinite Series 1 + 1/2 + 1/4 + ... is:

(a)	1.99999	(b)	2.00001
(c)	2	(d)	1.999

3.15. Sum of infinity of the following geometric progression

$\frac{1}{1.1} + \frac{1}{(1.1)^2} + \frac{1}{(1.1)^3} + \ldots\ldots\ldots$ is:

(a)	9	(b)	–10
(c)	10	(d)	–9

3.16. The first term of an A.P is 14 and the sum of the first five terms and the first ten terms are equal in magnitude but opposite in sign. The 3rd term of the AP is:

(a)	$6\frac{4}{11}$	(b)	6
(c)	$\frac{4}{11}$	(d)	None of these

3.17. The first and the last term of an AP are –4 and 146. The sum of the terms is 7171. The number of terms is:

(a)	101	(b)	100
(c)	99	(d)	None of these

3.18. If you save 1 paisa today, 2 paisas the next day 4 paisas the succeeding day and so on, then your total savings in two weeks will be:

(a)	Rs.163	(b)	Rs.183
(c)	Rs.163.83	(d)	None of these

3.19. A person is employed in a company at Rs.3,000 per month and he would get an increase of Rs.100 per year in his monthly salary. The total amount which he receives in 25 years and the monthly salary in the last year are:

(a)	Rs.5,400 and Rs.1,250,000	(b)	Rs.540 and Rs.1,260,000
(c)	Rs.5,500 and Rs.1,260,000	(d)	Rs.5,400 and Rs.1,260,000

AT A GLANCE SPOTLIGHT STICKY NOTES

3.20. Three numbers are in AP and their sum is 21. If 1, 5, 15 are added to them respectively, they form a G.P. the numbers are:

(a)	5, 7, 9	(b)	9, 5, 7
(c)	7, 5, 9	(d)	None of these

3.21. The sum of three numbers in G.P. is 70. If the two extreme terms are each multiplied by 4 and the geometric mean of these three terms is multiplied by 5, the first term multiplied by 4, the geometric mean and the third term multiplied by 4are in A.P. the numbers are:

(a)	12, 18, 40	(b)	10, 30, 90
(c)	40, 20, 10	(d)	None of these

3.22. The sum of all natural numbers between 500 and 1000 which are divisible by 13 is:

(a)	28,405	(b)	24,805
(c)	28,540	(d)	None of these

3.23. If unity is added to the sum of any number of terms of the A.P. 3, 5, 7, 9,...... the resulting sum is:

(a)	'a' perfect cube	(b)	'a' perfect square
(c)	both (a) and (b)	(d)	none of these

3.24. A person has to pay Rs.975 by monthly installments each less than the former by Rs.5. The first installment is Rs.100. The time by which the entire amount will be paid is:

(a)	10 months	(b)	15 months
(c)	14 months	(d)	None of these

3.25. Divide 12.50 into five parts in A.P. such that the first part and the last part are in the ratio of 2:3.

(a)	2, 2.25, 2.5, 2.75, 3	(b)	–2, –2.25, –2.5, –2.75, –3
(c)	4, 4.5, 5, 5.5, 6	(d)	–4, –4.5, –5, –5.5, –6

3.26. The p^{th} term of an A.P. is 1/q and q^{th} term is 1/p. The sum of the pq^{th} term is:

(a)	1/2(pq + 1)	(b)	1/2(pq – 1)
(c)	pq + 1	(d)	pq – 1

3.27. The least value of n for which the sum of n terms of the series 1 + 3 + 32 + ... is greater than 7,000 is:

(a)	9	(b)	10
(c)	8	(d)	7

3.28. If the sum of infinite terms in a G.P. is 2 and the sum of their squares is 4/3 the series is:

(a)	1, 1/2, 1/4......	(b)	1, –1/2, 1/4......
(c)	–1, –1/2, –1/4......	(d)	None of these

3.29. The infinite G.P. with first term 1/4 and sum 1/3 is:

(a)	1/4, 1/16, 1/64......	(b)	1/4, –1/16, 1/64......
(c)	1/4, 1/8, 1/16......	(d)	None of these

3.30. The numbers x, 8, y are in G.P. and the numbers x, y, –8 are in A.P. The values of x, y are:

(a)	16, 4	(b)	4, 16
(c)	Both (a) and (b)	(d)	None of these

3.31. How many terms are there in the sequence of 1/128, 1/64, 1/32….., 32, 64?

(a)	13	(b)	14
(c)	15	(d)	16

3.32. An auditorium has 20 seats in the front row, 25 seats in the second row, 30 seats in the third row and so on for 13 rows. Numbers of seats in the thirteenth row are:

(a)	70	(b)	80
(c)	82	(d)	90

3.33. Sum of the series 1, 1/3, 1/9, 1/27 ……… to infinity is:

(a)	–3/2	(b)	2/3
(c)	1/3	(d)	3/2

3.34. The sum of the infinite series 2 + $\sqrt{2}$ + 1 + ……… is:

(a)	$2-4\sqrt{2}$	(b)	$-2(1+\sqrt{2})$
(c)	$2+4\sqrt{2}$	(d)	$4+2\sqrt{2}$

3.35. Shaheer bought equipment for Rs. 450,000. The amount is payable in 11 annual installments, where first 10 installments are in arithmetic progression and 11th installment will settle the remaining balance. If the first two installments are Rs. 10,000 and Rs. 12,000 respectively compute the amount of 11th installment.

(a)	240,000	(b)	230,000
(c)	260,000	(d)	280,000

3.36. Saqib will invest Rs. 2 on the first day of January, Rs. 4 on the second day of January, Rs. 6 on the third day of January and so on for the first five days of January. He will then spend Rs 1 per day from sixth to tenth day of January. Compute what amount will he have with him on the eleventh day of January

(a)	25	(b)	30
(c)	40	(d)	45

3.37. A company produces 100 units in first week. It is expected that re-engineering production process will increase output by 10% every week of the amount of units produced in previous week till fifth week, after which the output per week will be constant.
If re-engineering is carried out compute the number of total units produced in ninth and tenth week (to the nearest whole number)

(a)	293	(b)	280
(c)	275	(d)	288

3.38. The first and last terms of an arithmetic progression are 5 and 25 respectively. Find the middle term of the series

(a)	20	(c)	25

(b) 5 (d) 15

3.39. Sami borrowed a loan of Rs 11,000 with the condition that the repayments will be made in seven annual installments. The first and the last installments will be equal to each other and the first six installments will be in arithmetic progression. If common difference between first six installments is Rs 500. Compute fifth installment.

(a)	2,500	(c)	500
(b)	3,000	(d)	1,000

3.40. A company earned a profit of Rs. 50,000 in first year. The profit is expected to increase by Rs 5,000 each year. How many years will it take the company to earn a total profit of Rs. 165,000

(a)	4	(b)	3
(c)	5	(d)	2

3.41. Find the value of Y such that Y+2, 2Y+3, 6Y+1 will form an AP

(a)	2	(b)	3
(c)	1	(d)	5

3.42. If the first and the last terms of an arithmetic progression are 10 and 50 respectively and the total number of terms is 10. Find the sum of all the terms

(a)	200	(b)	400
(c)	300	(d)	100

3.43. A worker produces 1,092 units in three days. If the production is expected to increase by 20% percent each day compute the total number of units produced on first and last day.

(a)	732	(b)	432
(c)	300	(d)	360

3.44. The first and the last term of an arithmetic progression are in the ratio of 3:4. The sum of all terms is 210 and there are 6 terms in total. Compute common difference.

(a)	2	(b)	3
(c)	5	(d)	10

3.45. Rizwan borrowed a loan of Rs 6,620 to be paid in three annual payments in geometric progression. If the first installment is Rs. 2,000 compute the common ratio.

(a)	1.15	(b)	1.1
(c)	1.2	(d)	1.22

3.46. The sum of 5 terms of an AP, whose last term is 17 is 65. The first term and common difference are:

(a)	9 and 2	(b)	13 and 2
(c)	11 and 7	(d)	11 and 2

3.47. The last term of the sequence, 1, -4, 16, …… to 5th term is

(a)	64	(b)	256
(c)	-256	(d)	-64

3.48. Find the value of x such that $x+3$, $2x+11$, $2x+41$ are in G.P.

(a)	3	(b)	4
(c)	2	(d)	1

3.49. If the first term of a G.P is a and the common ratio is 2. What will be the value of third term?

(a)	a	(b)	2a
(c)	3a	(d)	4a

3.50. If the fourth term of a G.P is 8 times its first term compute common ratio

(a)	3	(b)	4
(c)	2	(d)	5

3.51. The first term of an A.P. is 5 and fourth term is 17. Compute the value of common difference

(a)	4	(b)	12
(c)	3	(d)	5

3.52. A particular A.P. has first term as 3 and third term as 7. The common difference of this A.P. is equivalent to the first term of a G.P. with common ratio of 5. Compute third term of the G.P.

(a)	60	(b)	50
(c)	40	(d)	20

3.53. Sum of a G.P. is 175, if the total number of terms is 3 and common ratio is 2 compute the value of first term

(a)	25	(b)	50
(c)	40	(d)	70

3.54. The infinite G.P. with common ratio of ½ and sum 10 is:

(a)	5, 2.5, 1.25…..	(b)	1.25, 2.5, 5….
(c)	2.5, 1.25, 5….	(d)	None of these

ANSWERS TO SELF-TEST QUESTIONS

3.1	3.2	3.3	3.4	3.5	3.6
(d)	**(b)**	**(c)**	**(b)**	**(a)**	**(c)**
3.7	3.8	3.9	3.10	3.11	3.12
(b)	**(a)**	**(c)**	**(c)**	**(c)**	**(d)**
3.13	3.14	3.15	3.16	3.17	3.18
(b)	**(c)**	**(c)**	**(a)**	**(a)**	**(c)**
3.19	3.20	3.21	3.22	3.23	3.24
(d)	**(a)**	**(c)**	**(a)**	**(b)**	**(b)**
3.25	3.26	3.27	3.28	3.59	3.30
(a)	**(a)**	**(a)**	**(a)**	**(a)**	**(a)**
3.31	3.32	3.33	3.34	3.35	3.36
(b)	**(b)**	**(d)**	**(d)**	**(c)**	**(a)**
3.37	3.38	3.39	3.40	3.41	3.42
(a)	**(d)**	**(a)**	**(b)**	**(c)**	**(c)**
3.43	3.44	3.45	3.49	3.47	3.48
(a)	**(a)**	**(b)**	**(a)**	**(b)**	**(c)**
3.49	3.50	3.51	3.52	3.53	3.54
(d)	(c)	(a)	(b)	(a)	(a)

CHAPTER 4

LINEAR PROGRAMMING

IN THIS CHAPTER:

AT A GLANCE

A linear programming problem involves maximizing or minimizing a linear function (the objective function) subject to linear constraints. As per the problem, defined constraints will lead to the maximum or minimum value of the objective function. Plotting a graph helps in identifying the value that satisfies the objective function. The straight line for each constraint is the boundary edge of that constraint – its outer limit in the case of maximum amounts (and inner limit, in the case of minimum value constraints). The optimal combination of values of x and y can be found using corner point theorem; or slope of the objective function.

AT A GLANCE
SPOTLIGHT
STICKY NOTES

1. LINEAR PROGRAMMING

Linear programming is a mathematical method for determining a way to achieve the best outcome (such as maximum profit or lowest cost) subject to a number of limiting factors (constraints). It is the process of taking various linear inequalities relating to some situation, and finding the "best" value obtainable under those conditions.

Both the constraints and the "best outcome" are represented as linear relationships.

What constitutes the best outcome depends on the objective. The equation constructed to represent the best outcome is known as the **objective function.**

▶ *For example:*

Typical examples would be to work out the maximum profit from making two sorts of goods when resources needed to make the goods are limited or the minimum cost for which two different projects are completed.

Linear inequality can be solved in a way similar to that used to solve simultaneous equations. Overall approach for linear programming requires following steps:

- **Step 1:** Define variables.
- **Step 2:** Construct inequalities to represent the constraints.
- **Step 3:** Plot the constraints on a graph
- **Step 4:** Identify the feasible region. This is an area that represents the combinations of x and y that are possible in the light of the constraints.
- **Step 5:** Construct an objective function
- **Step 6:** Identify the values of x and y that lead to the optimum value of the objective function. This might be a maximum or minimum value depending on the objective. There are different methods available of finding this combination of values of x and y.

An inequality solves to give a value of a variable at the boundary of the inequality. A value could be one side of this variable but not the other. Double inequalities are solved to give the range within which a variable lies.

Defining the variables and constructing inequalities:

An inequality defines a relationship where one expression is greater than or less than another. Variable x and y are usually used to demonstrate these relationships.

▶ *For example:*

A company owns a machine which runs for 50 hours a week. The machine is used to make bowls. Each bowl takes 1 hour of machine time

Let x = the number of bowls made

Then Weekly production would be $x \leq 50$

This suggest that x can be any number up to 50.

Note that it does not have to be 50 but it cannot be more than 50

A company owns a machine which runs for 50 hours a week. The machine is used to make bowls and vases. Each bowl takes 1 hour of machine time and each vase takes two hours of machine time

The example suggest that the company could make bowls and vases in any combination but the total machine hours cannot exceed 50

Let x = the number of bowls made

Let y = the number of vases made

Weekly production of bowls and vases would be $x + 2y \leq 50$

The statement that $x + 2y \leq 50$ means that $x + 2y = 50$ is possible

Plotting constraints

The constraints in a linear programming problem can be drawn as straight lines on a graph, provided that there are just two variables in the problem (x and y).

The straight line for each constraint is the boundary edge of the constraint – its outer limit in the case of maximum amounts (and inner limit, in the case of minimum value constraints).

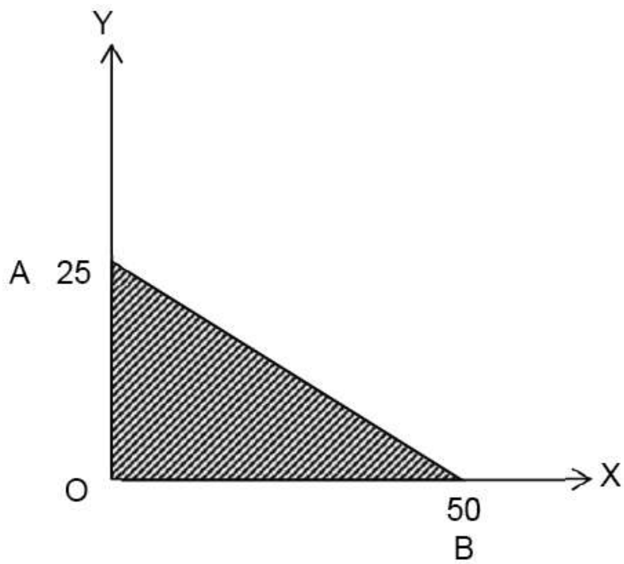
- *For example:*

Information from the example is used to plot the graph as follows:

The company could make 50 bowls if it made no vases and the company could make 25 vases if it made no bowls. The line is plotted by establishing the coordinates of the points where the line cuts the axes. The constraint is drawn as a straight line. The easiest approach to finding the points is to set x to zero and calculate a value for y and then set y to zero and calculate a value for x.

Equation of the line: $x + 2y = 50$

Point A ($x = 0$)	Point B ($y = 0$)
Therefore:	Therefore:
$0 + 2y = 50$	$x + 0 = 50$
$y = 25$	$x = 50$
A: (0, 25)	B: (50, 0)



Feasible region

The feasible area for a solution to the problem is shown as the shaded area AOB in the example above. Combinations of values for x and y within this area can be achieved within the limits of the constraints. Combinations of values of x and y outside this area are not possible, given the constraints that exist.

The line represents a boundary of what is possible (feasible). These combinations could be on the line AB or they could be under the line AB.

- *For example:*

In the example above, shaded area suggest that any combinations of x and y are possible as long as no more than 50 hours are used. Therefore, they must be above the x axis and to the right of the y axis since there cannot be negative vases and bows.

The area bounded by AOB is described as a feasibility region.

Maximising (or minimising) the objective function

The combination of values for x and y that maximises the objective function will be a pair of values that lies somewhere along the outer edge of the feasible area.

The solution is a combination of values for x and y that lies at one of the 'corners' of the outer edge of the feasible area. In the graph above, the solution to the problem will be the values of x and y at A, B or C. The optimal solution cannot be at D as the objective function is for maximisation and values of x and/or y are higher than those at D for each of the other points.

The optimal combination of values of x and y can be found using:

- corner point theorem; or
- slope of the objective function.

Corner point theorem.

The optimum solution lies at a corner of the feasible region. The approach involves calculating the value of x and y at each point and then substituting those values into the objective function to identify the optimum solution.

► *For example:*

In the previous example, the solution has to be at points A, B or C.

Calculate the values of x and y at each of these points, using simultaneous equations. Having established the values of x and y at each of the points, calculate the value of the objective function for each.

The solution is the combination of values for x and y at the point where the total contribution is highest.

Maximise C = $50x + 40y$

Point A is at the intersection of:	$x = 2$	1
	$4x + 8y = 48$	2
Substitute for x in equation 2	$4(2) + 8y = 48$	
	$8y = 48 - 8$	
	$y = 5$	
Point B is at the intersection of:	$4x + 8y = 48$	1
	$6x + 3y = 36$	2
Multiply equation 1 by 1.5	$6x + 12y = 72$	3
Subtract 2 from 3	$9y = 36$	
	$y = 4$	
Substitute for y in equation 1	$4x + 8(4) = 48$	
	$4x = 16$	
	$x = 4$	
Point C is at the intersection of:	$y = 3$	1
	$6x + 3y = 36$	2
Substitute for y in equation 2	$6x + 3(3) = 36$	

$$6x = 27$$
$$x = 4.5$$

Substitute coordinates into objective function

$$C = 50x + 40y$$

Point A ($x = 2$; $y = 5$) $C = 50(2) + 40(5)$

$C = 300$

Point B ($x = 4$; $y = 4$) $C = 50(4) + 40(4)$

$C = 360$

Point C ($x = 4.5$; $y = 3$) $C = 50(4.5) + 40(3)$

$C = 345$

Conclusion: The optimal solution is at point B where $x = 4$ and $y = 4$ giving a value for the objective function of 360.

Slope of the objective function

There are two ways of using the slope of the objective line to find the optimum solution.

Measuring slopes

This approach involves estimating the slope of each constraint and the objective function and ranking them in order. The slope of the objective function will lie between those of two of the constraints. The optimum solution lies at the intersection of these two lines and the values of x and y at this point can be found as above.

- *For example:*

Objective function:	Rearranged	Slope
$C = 50x + 40y$	$y = {}^{C}/_{40} – 1.25x$	– 1.25
Constraints:		
$6x + 3y = 36$	$y = 12 – 2x$	– 2
$4x + 8y = 48$	$y = 6 – 0.5x$	– 0.5
$x = 2$		∞
$y = 3$		0

The slope of the objective function (– 1.25) lies between the slopes of $6x + 3y = 36$ (– 2) and $4x + 8y = 48$ (– 0.5). The optimum solution is at the intersection of these two lines.

Values of x and y at this point would then be calculated as above ($x = 4$ and $y = 4$) and the value of the objective function found.

Plotting the objective function line and moving it from the source

The example above showed that the objective function ($C = 50x + 40y$) can be rearranged to the standard form of the equation of a straight line ($y = C/40 – 1.25x$).

This shows that the slope of the line is not affected by the value of the equation (C). This is useful to know as it means that the equation can be set to any value in order to plot the line of the objective function on the graph.

The first step is to pick a value for the objective function and construct two pairs of coordinates so that the line of the objective function can be plotted. (The total value of the line does not matter as it does not affect the slope.

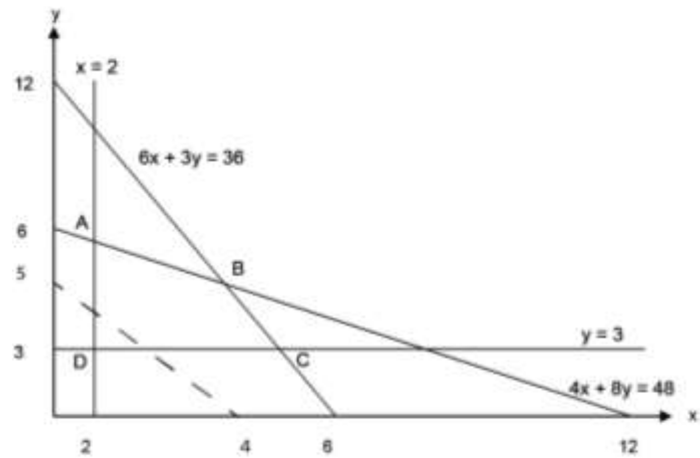
- *For example:*

  Let C = 200

  200 = 50x + 40y

  If x = 0 then y = 5 and if y = 0 then x = 4

  This is plotted as the dotted line below.



Note that the value of C in the above plot of the objective function is constant (at 200). This means that any pair of values of x and y that fall on the line will result in the same value for the objective function (200). If a value higher than 200 had been used to plot the objective function it would have resulted in a line with the same slope but further out from the source (the intersection of the x and y axis).

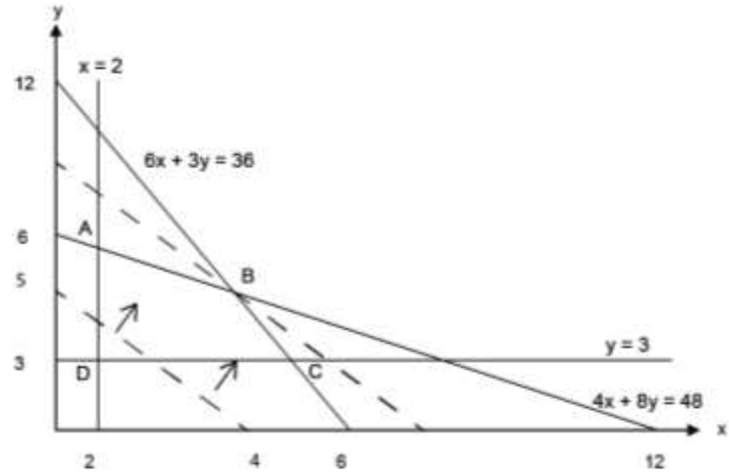
To maximise the value of the objective function, the next step is to identify the point in the feasible area where objective function line can be drawn as far from the origin of the graph as possible.

This can be done by putting a ruler along the objective function line that you have drawn and moving it outwards, parallel to the line drawn, until that point where it just leaves the feasibility area. This will be one of the corners of the feasibility region. This is the combination of values of x and y that provides the solution to the linear programming problem.

- *For example:*

  Move a line parallel to the one that you have drawn until that point where it is just about to leave the feasible region.

AT A GLANCE SPOTLIGHT STICKY NOTES



The optimum solution is at point B (as before). This is at the intersection of $6x + 3y = 36$ and $4x + 8y = 48$. The next step is to solve these equations simultaneously for values of x and y that optimise the solution. These values can then be inserted into the objective function to find the value of C at the optimum. (This has been done earlier).

- *For example:*

Objective function: Maximise C = 5x + 5y, subject to the following constraints:

Direct labour	**2x + 3y**	**≤**	**6,000**
Machine time	**4x + y**	**≤**	**4,000**
Sales demand, y	**y**	**≤**	**1,800**
Non-negativity	**x, y**	**≥**	**0**

In order to identify the feasible region, let's find combinations of values for x and y that represent the 'feasible region' on the graph for a solution to the problem

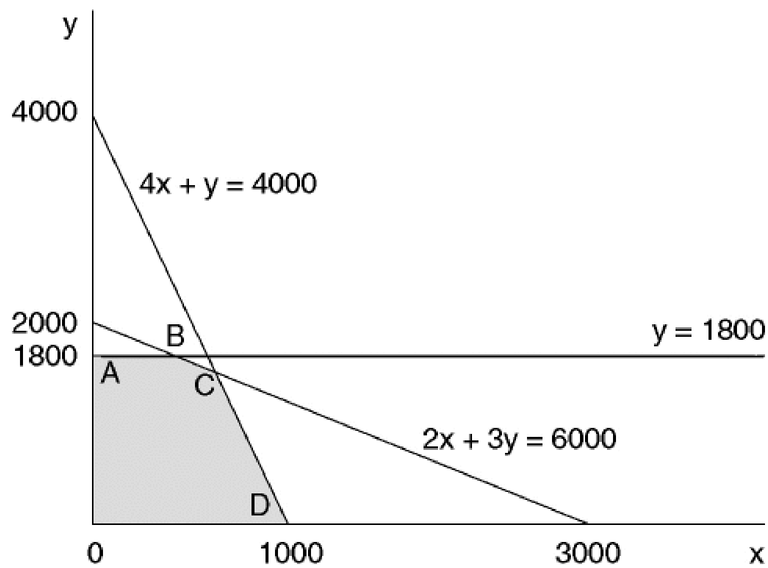
(1) Constraint: $2x + 3y \leq 6{,}000$

When $x = 0$, $y \leq 2{,}000$. When $y = 0$, $x \leq 3{,}000$.

(2) Constraint: $4x + y \leq 4{,}000$

When $x = 0$, $y \leq 4{,}000$. When $y = 0$, $x \leq 1{,}000$.

(3) Constraint: $y = 1{,}800$



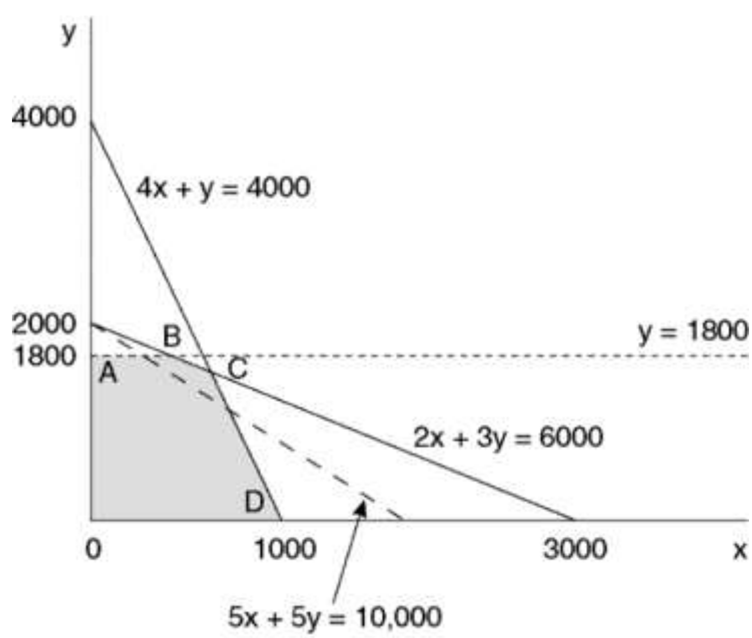
Now, in order to plot the objective function – Let P = 10,000 (The value 10,000 is chosen as a convenient multiple of the values 5 and 5 that can be drawn clearly on the graph.)

$5x + 5y = 10,000$

$x = 0$, $y = 2,000$ and $y = 0$, $x = 2,000$.

The combination of values of x and y that will maximise total contribution lies at point C on the graph.

The combination of x and y at point C is therefore the solution to the linear programming problem.



The optimum solution is at point C. This can be found be examining the objective function line to see where it leaves the feasible region, using corner point theorem or by comparing the slope of the objective function to the slopes of the constraints.

Point C is at the intersection of: $2x + 3y = 6,000$ and $4x + y = 4,000$

Multiply first equation with 2

$2(2x + 3y) = 2(6,000)$

$4x + 6y = 12,000$

Subtract earlier equation from this equation

$+ 4x + 6y = 12,000$

$+ 4x + y \;\; = 4,000$

(-) (-) (-)

$5y = 8,000$

Dividing both sides by 5

$y = 1600$

Substituting the value of y in any of the above equations

$2x + 3(1,600) = 6,000$

$2x + 4,800 = 6,000$

Subtracting 4800 from both sides

$2x = 1,200$

$x = 600$

The objective function is maximised by producing 600 units of x and 1,600 units of **y.**

The objective in this problem is to maximise $5x + 5y$ giving a maximum value of:

$5(600) + 5(1,600) = 11,000.$

However, the same problem can be solved using Optimum solution using corner point theorem as follows:

Maximise $C = 5x + 5y$		
Point A is where	$y = 1,800$	
	$x = 0$	
Point B is at the intersection of:	$y = 1,800$	1
	$2x + 3y = 6,000$	2
	$2x + 3(1,800) = 6,000$	3
	$x = 300$	
Point C is at the intersection of:	$2x + 3y = 6,000$	1
	$4x + y = 4,000$	2
Solved previously at	$x = 600$ and $y = 1,600$	
Point D is at the intersection of:	$y = 0$	1
	$4x + y = 4,000$	2
	$4x + 0 = 4,000$	
	$x = 1,000$	
Substitute coordinates into objective function		
	$C = 5x + 5y$	
Point A ($x = 0$; $y = 1,800$)	$C = 5(0) + 5(1,800)$	
	$C = 9,000$	

Point B (x = 300; y = 1,800)	C = 5(300) + 5(1,800)
	C = 10,500
Point C (x = 600; y = 1,600)	C = 5(600) + 5(1,600)
	C = 11,000
Point D (x = 1,000; y = 0)	C = 5(1,000) + 5(0)
	C = 5,000

Conclusion: The optimal solution is at point C where x = 600 and y = 1,600 giving a value for the objective function of 11,000.

And Optimum solution can be obtained by measuring slopes as follows:

Objective function:	**Rearranged**	**Slope**
C = 5x + 5y	y = $^C/_5$ – x	– 1
Constraints:		
y = 1,800		0
2x + 3y = 6,000	y = 2,000 – $^2/_3x$	– $^2/_3$
4x + y = 4,000	y = 4,000 – 4x	– 4

The slope of the objective function (– 1) lies between the slopes of 2x + 3y = 6,000 (– $^2/_3$) and 4x + y = 4,000 (– 4). The optimum solution is at the intersection of these two lines.

Values of x and y at this point have been calculated above.

Minimising functions

The above illustration is one where the objective is to maximise the objective function. For example, this might relate to the combination of products that maximise profit.

You may also face minimisation problems (for example, where the objective is to minimise costs). In this case the optimal solution is at the point in the feasibility region that is closest to the origin. It is found using similar techniques to those above.

2. LINEAR PROGRAMMING OF BUSINESS PROBLEMS

Decisions about what mix of products should be made and sold in order to maximise profits can be formulated and solved as linear programming problems. Overall approach to solving such problems is similar to solving inequalities using constraints.

Solving for business problems:

The first step in any formulation is to define variables for the outcome that require optimisation. For example, if the question is about which combination of products will maximise profit, a variable must be defined for the number of each type of product in the final solution.

Then, a separate constraint must be identified for each item that might put a limitation on the objective function. Each constraint, like the objective function, is expressed as a formula. Each constraint must also specify the amount of the limit or constraint.

- For a **maximum limit**, the constraint must be expressed as 'must be equal to or less than'.
- For a **minimum limit**, the constraint must be expressed as 'must be equal to or more than'.

▶ *For example:*

Maximum limit:	Maximum sales demand for Product X of 5,000 units
Define variables	Let x be the number of Product X made.
Constraint expressed as:	$x \leq 5{,}000$
Constraints with two variables	A company makes two products, X and Y. It takes 2 hours to make one unit of X and 3 hours to make one unit of Y. Only 18,000 direct labour hours are available.
Define variables	Let x be the number of Product X made and let y be the number of Product Y made.
Constraint expressed as:	$2x + 3y \leq 18{,}000$
Minimum limit:	There is a requirement to supply a customer with at least 2,000 units of Product X
Define variables	Let x be the number of Product X made.
Constraint expressed as:	$x \geq 2{,}000$
Non-negativity constraints	It is not possible to make a negative number of products.
Define variables	Let x be the number of Product X made and let y be the number of Product Y made.
Constraint expressed as:	$x, y \geq 0$ are called positive constraints
	These constraints do not have to be plotted as they are the x and y axes of the graph

The usual assumption is that a business has the objective of maximising profits. This then requires, expressing the same using an objective function.

- *For example:*

  A company makes and sells two products, Product X and Product Y. The contribution per unit is \$8 for Product X and \$12 for Product Y. The company wishes to maximise profit.

  Let x be the number of Product X made and let y be the number of Product Y made.

  To maximize contribution, the formula can be expressed as C = $8x + 12y$

Once a linear programming problem has been formulated, it must then be solved to decide how the objective function is maximised (or minimised) as explained earlier.

- *For example:*

  Construct the constraints and the objective function taking into account the following information:

  A company makes and sells two products, Product X and Product Y. The contribution per unit is \$8 for Product X and \$12 for Product Y. The company wishes to maximise profit.

  The expected sales demand is for 6,000 units of Product X and 4,000 units of Product Y.

	Product X	Product Y
Direct labour hours per unit	3 hours	2 hours
Machine hours per unit	1 hour	2.5 hours

	Total hours
Total direct labour hours available	20,000
Total machine hours available	12,000

For the above problem a linear programming problem can be formulated as follows:

Let the number of units (made and sold) of Product X be x

Let the number of units (made and sold) of Product Y be y

The objective function is to maximise total contribution given as C = $8x + 12y$.

Subject to the following constraints:

Direct labour	$3x + 2y$	≤	20,000
Machine time	$x + 2.5y$	≤	12,000
Sales demand, x	x	≤	6,000
Sales demand, Y	y	≤	4,000
Non-negativity	x, y	≥	0

Islamabad Manufacturing Limited makes and sells two versions of a product, Mark 1 and Mark 2. Identify the quantities of Mark 1 and Mark 2 that should be made and sold during the year in order to maximise profit and contribution.

	Mark 1	Mark 2
Direct materials per unit	2 kg	4 kg
Direct labour hours per unit	3 hours	2 hours
Maximum sales demand	5,000 units	unlimited
Contribution per unit	10 per unit	15 per unit

	Resource available
Direct materials	24,000 kg
Direct labour hours	18,000 hours

In defining the constraints, let the number of units of Mark 1 be x and number of units of Mark 2 be y.

Using the given information of contribution per unit for each product, maximised total contribution can be given by: C = 10x + 15y.

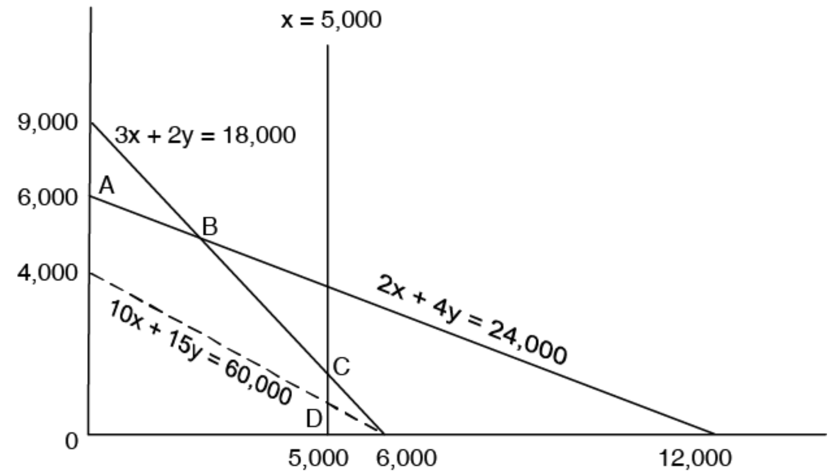
Let C = 60,000 to allow a plot of the line

Constraints are given as:

Direct materials	$2x + 4y$	≤	24,000
Direct labour	$3x + 2y$	≤	18,000
Sales demand, Mark 1	x	≤	5,000
Non-negativity	x, y	≥	0

For plotting the constraints and objective function intercepts can be used

	C = 10x + 15y	**2x + 4y = 24,000**	**3x + 2y = 18,000**
x = 0	y = 4,000	y = 6,000	y = 9,000
y = 0	x = 6,000	x = 12,000	x = 6,000



The feasible solutions are shown by the area 0ABCD in the graph.

The optimum solution is at point B. This can be found by examining the contribution line to see where it leaves the feasible region, using corner point theorem or by comparing the slope of the objective function to the slopes of the constraints (workings given on next page)

Point B is at the intersection of: $2x + 4y = 24{,}000$ and $3x + 2y = 18{,}000$

Solving for x and y using the constraint equations

Multiplying equation for constraint 2 by 2.

$2(3x + 2y) = 2(18{,}000)$

$6x + 4y = 36{,}000$

Subtract equation for constraint 1 from this equation

$+ 6x + 4y = 36{,}000$

$+ 2x + 4y = 24{,}000$

$(-) \quad (-) \quad (-)$

$4x \quad = 12{,}000$

$x = 3{,}000$

Substituting the value of x in any of the above equations

$2(3{,}000) + 4y = 24{,}000$

$6{,}000 + 4y = 24{,}000$

$4y = 18{,}000$

$y = 4{,}500$

The total contribution is maximised by producing 3,000 units of X and 4,500 units of Y.

The objective in this problem is to maximise $10x + 15y$ giving a maximum value of:

$10(3{,}000) + 15(4{,}500) = 97{,}500.$

Using the corner point theorem, optimal solution can be found as follows:

Maximise C = $10x + 15y$		
Point A is where	$y = 6,000$	
	$x = 0$	
Point B is at the intersection of:	$2x + 4y = 24,000$	
	$3x + 2y = 18,000$	
Solved previously at	$x = 3,000$ and $y = 4,500$	
Point C is at the intersection of:	$x = 5,000$	1
	$3x + 2y = 18,000$	2
Substituting	$3(5,000) + 2y = 18,000$	
	$y = 1,500$	
Point D is where	$y = 0$	
	$x = 5,000$	
Substitute coordinates into objective function		
	$C = 10x + 15y$	
Point A ($x = 0$; $y = 6,000$)	$C = 10(0) + 15(6,000)$	
	$C = 90,000$	
Point B ($x = 3,000$; $y = 4,500$)	$C = 10(3,000) + 15(4,500)$	
	$C = 97,500$	
Point C ($x = 5,000$; $y = 1,500$)	$C = 10(5,000) + 15(1,500)$	
	$C = 72,500$	
Point D ($x = 5,000$; $y = 0$)	$C = 10(5,000) + 15(0)$	
	$C = 50,000$	
Conclusion: The optimal solution is at point B where $x = 3,000$ and $y = 4,500$ giving a value for the objective function of 97,500.		

Solution can also be reached using measuring slopes methods as follows:

Objective function:	**Rearranged**	**Slope**
C = $10x + 15y$	$y = {}^{C}/_{15} - {}^{10}/_{15}x$	$-{}^{10}/_{15} = -{}^{2}/_{3}$
Constraints:		
$2x + 4y = 24{,}000$	$y = 6{,}000 - 0.5x$	-0.5
$3x + 2y = 18{,}000$	$y = 9{,}000 - {}^{3}/_{2}x$	$-{}^{3}/_{2}$
$x = 5{,}000$		∞

The slope of the objective function (– 2/3) lies between the slopes of $2x + 4y = 24{,}000$ (– 0.5) and $3x + 2y = 18{,}000$ (– 3/2). The optimum solution is at the intersection of these two lines.

Values of x and y at this point would then be calculated as above ($x = 3{,}000$ and $y = 4{,}500$) and the value of the objective function found.

A company manufactures two models of car engines. Model S requires 1 units of labor and 5 units of raw materials. Model T requires 1 units of labor and 4 units of raw materials. If 110 units of labour and 480 units of raw materials are available and company makes a contribution of Rs. 7 per Model S and Rs. 6 per Model T. How many of each models of car engine to be manufactured to maximize the profit? If the contribution changes to Rs. 3 for Model S and Rs. 4 for Model T, how many each engine type to be manufactured to maximize the profit?

Let's consider number of units of Model S as x and number of units of Model T as **y**

Using the given information of Contribution per unit for each product, maximise total contribution can be given by: $C_1 = 7x + 6y$ and $C_2 = 3x + 4y$.

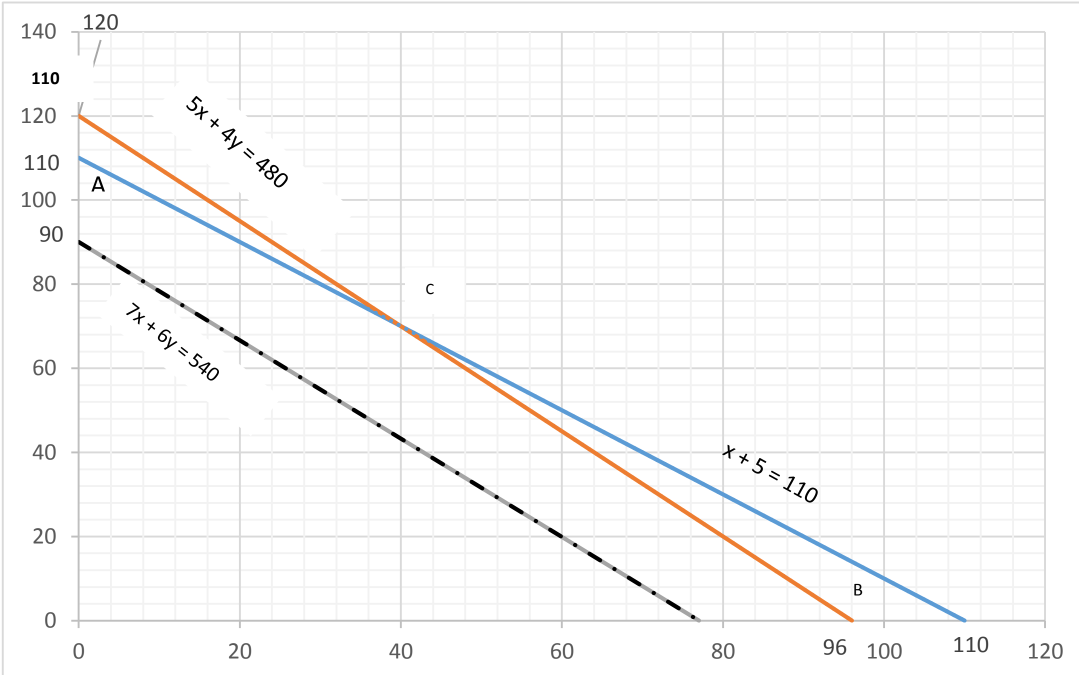
Let C = 540 to allow a plot of the line

Constraints are given as:

labour	$x + y$	≤	110
Raw materials	$5x + 4y$	≤	480
Non-negativity	x, y	≥	0

For plotting the constraints and objective function intercepts can be used

	540 = 7x + 6y	**x + y = 110**	**5x + 4y = 480**
x = 0	y = 90	y = 110	y = 120
y = 0	x =77	x = 110	x = 96



The feasible solutions are shown by the area 0ABC in the graph.

The optimum solution can be found using corner point theorem or by comparing the slope of the objective function to the slopes of the constraints.

Point C is at the intersection of x + y = 110 and 5x + 4y = 480

Solving for x and y using the constraint equations

Multiplying equation for constraint 1 by 4

4(x + y) = 4(110)

4x + 4y = 440

Subtract the above equation from equation constraint 1

+ 5x + 4y = 480

+ 4x + 4y = 440

(-) (-) (-)

x = 40

Substituting the value of x in any of the above equations

40 + y = 110

4 = 110 - 40

y = 70

Using the corner point theorem, optimal solution can be found as follows:

Maximise C = $7x + 6y$	
Point A is where	$y = 110$
	$x = 0$
Point B is at the intersection of:	$x + y = 110$
	$5x + 4y = 480$
Solved previously at	$x = 40$ and $y = 70$
Point C is where	$y = 0$
	$x = 96$
Substitute coordinates into objective function	

Point A (0, 110)	**Point C (40, 70)**	**Point B (96, 0)**
C = $7x + 6y$	C = $7x + 6y$	C = $7x + 6y$
C = 7(0) + 6(110)	C = 7(40) + 6(70)	C = 7(96) + 6(0)
C = 0 + 660	C = 280 + 420	C = 672 + 0
C = 660	C = 700	C = 672

Conclusion: The optimal solution is at point C where $x = 40$ and $y = 70$ giving a value for the objective function of Rs. 700

And if the profit function changes to C = $3x + 4y$ then the profit would be maximized at Point A

Point A (0, 110)	**Point C (40, 70)**	**Point B (96, 0)**
C = $3x + 4y$	C = $3x + 4y$	C = $3x + 4y$
C = 3(0) + 4(110)	C = 3(40) + 4(70)	C = 3(96) + 4(0)
C = 0 + 440	C = 120 + 280	C = 288 + 0
C = 440	C = 400	C = 288

Conclusion: The optimal solution is at point A where $x = 0$ and $y = 110$ giving a value for the objective function of Rs. 440

STICKY NOTES

Linear programming is a mathematical method for determining a way to achieve the best outcome (such as maximum profit or lowest cost) subject to a number of limiting factors (constraints).

Overall approach for linear programming requires following steps:

- **Step 1: Define variables.**
- **Step 2: Construct inequalities to represent the constraints.**
- **Step 3: Plot the constraints on a graph**
- **Step 4: Identify the feasible region. This is an area that represents the combinations of x and y that are possible in the light of the constraints.**
- **Step 5: Construct an objective function**
- **Step 6: Identify the values of x and y that lead to the optimum value of the objective function.**

The constraints in a linear programming problem can be drawn as straight lines on a graph. The line represents a boundary of what is possible (feasible).

The combination of values for x and y that maximises the objective function will be a pair of values that lies somewhere along the outer edge of the feasible area.

The optimal combination of values of x and y can be found using:

- **corner point theorem; or**
- **slope of the objective function.**

SELF-TEST

4.1. (i) An employer recruits experienced (x) and fresh workmen (y) for his firm under the condition that he cannot employ more than 9 people, x and y can be related by the inequality:

(a) $x + y \neq 9$ (b) $x + y \leq 9, x \geq 0, y \geq 0$
(c) $x + y \geq 9, x \geq 0, y \geq 0$ (d) None of these

(ii) On the average, an experienced person does 5 units of work while a fresh recruit does 3 units of work daily and the employer has to maintain an output of at least 30 units of work per day. This situation can be expressed as:

(a) $5x + 3y \leq 30$ (b) $5x + 3y > 30$
(c) $5x + 3y \geq 30, x \geq 0, y \geq 0$ (d) None of these

(iii) The rules and regulations demand that the employer should employ not more than 5 experienced hands to 1 fresh one. This fact can be expressed as:

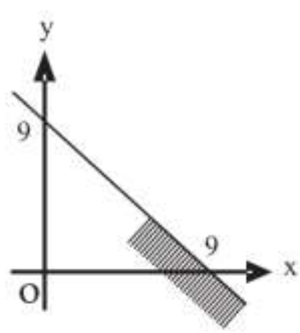
(a) $y \geq x/5$ (b) $5y \leq x$
(c) $x > 5y$ (d) None of these

(iv) The union however forbids him to employ less than 2 experienced person to each fresh person. This situation can be expressed as:

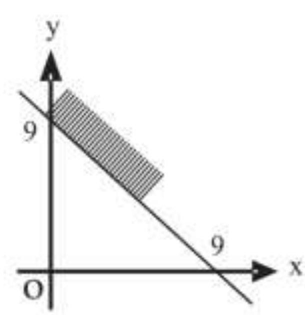
(a) $x \leq y/2$ (b) $y \leq x/2$
(c) $y \geq x/2$ (d) $x > 2y$

4.2. The graph to express the inequality $x + y \leq 9$ is:

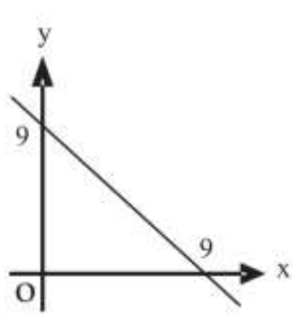
(a)



(b)



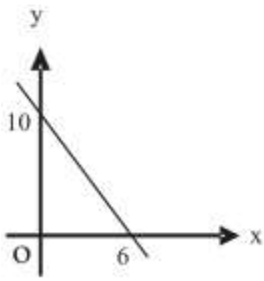
(c)



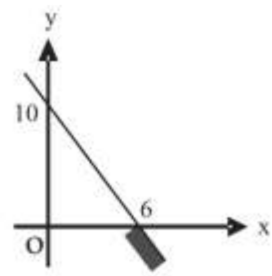
(d) None of these

4.3. The graph to express the inequality $5x + 3y \geq 30$ is:

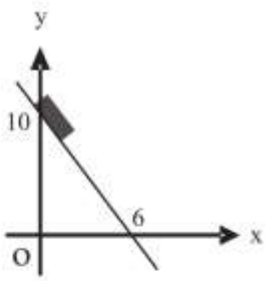
(a)



(b)



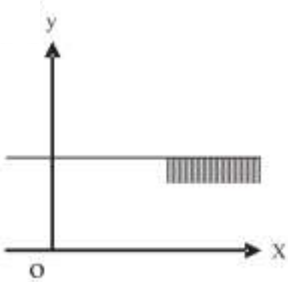
(c)



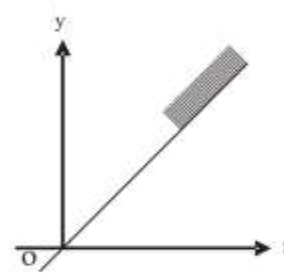
(d) None of these

4.4. The graph to express the inequality $y \leq \left(\frac{1}{2}\right) x$ is indicated by:

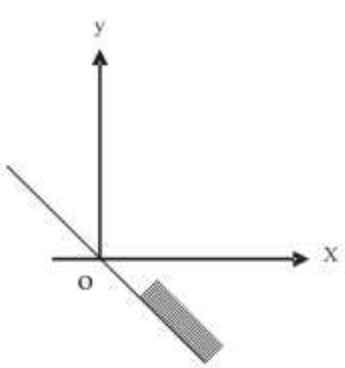
(a)



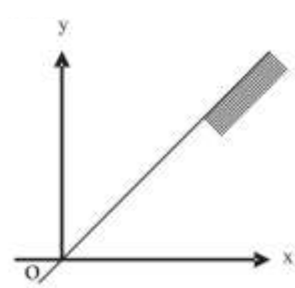
(b)



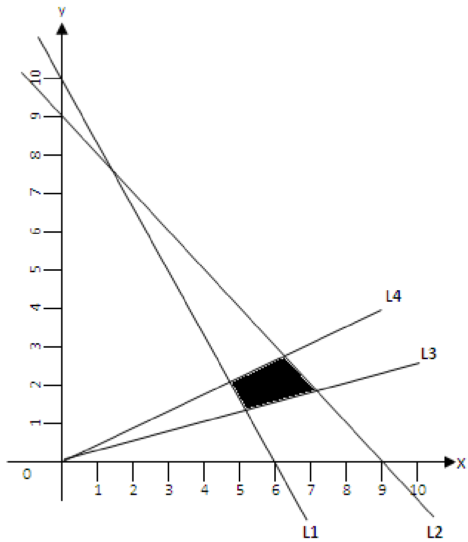
(c)



(d)



4.5.



L1: $5x + 3y = 30$, L2: $x + y = 9$, L3: $y = x/3$, L4: $y = x/2$

The common region (shaded part) shown in the diagram refers to:

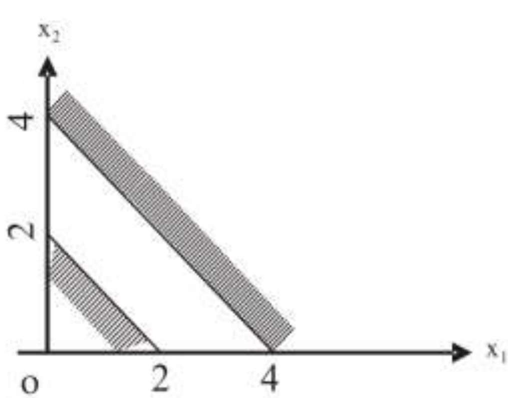
(a) $5x + 3y \leq 30$
$x + y \leq 9$
$y \leq 1/5x$
$y \leq x/2$

(b) $5x + 3y \geq 30$
$x + y \leq 9$
$y \geq x/3$
$y \leq x/2$
$x \geq 0, y \geq 0$

(c) $5x + 3y \geq 30$
$x + y \geq 9$
$y \leq x/3$
$y \geq x/2$
$x \geq 0, y \geq 0$

(d) None of these

4.6. The region indicated by the shading in the graph is expressed by inequalities:



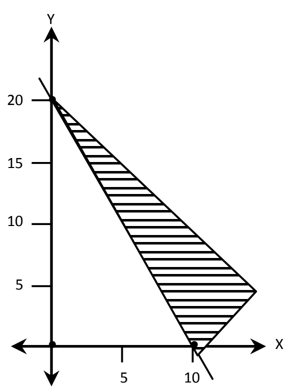
(a)	$x_1 + x_2 \leq 2$ $2x_1 + 2x_2 \geq 8$ $x_1 \geq 0, x_2 \geq 0$	(b)	$x_1 + x_2 \leq 2$ $x_2x_1 + x_2 \leq 4$
(c)	$x_1 + x_2 \geq 2$ $2x_1 + 2x_2 \geq 8$	(d)	$x_1 + x_2 \leq 2$ $2x_1 + 2x_2 > 8$

4.7. If A is the number of batsmen and B is the number of bowlers, the inequality constraint that the number of batsmen must be no more than 50% of the total players is:

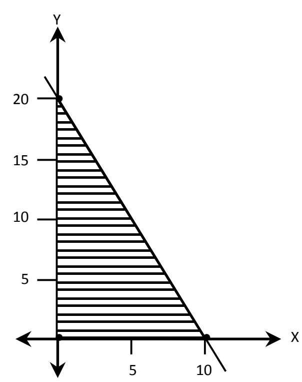
(a) $A \geq B$ (b) $A \leq B$ (c) $B \leq A$ (d) $A < B$

4.8. A firm manufactures two products. The products must be processed through one department. Product A requires 6 hours per unit, and product B requires 3 hours per unit. Total production time available for the coming week is 60 hours. There is a restriction in planning the production schedule, as total hours used in producing the two products cannot exceed 60 hours. This situation can be expressed as:

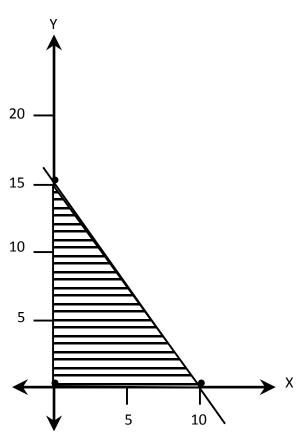
(a)



(b)



(c)



(d) None of these

4.9. A manufacturer produce two products P and Q which must pass through the same processes in departments A and B having weekly production capacities of 240 hours and 100 hours respectively. Product P needs 4 hours in department A and 2 hours in department B. Product Q requires 3 hours and 1 hour respectively, in department A and B. Profit yields for product P is Rs.700 and for Q is Rs.500. The manufacturer wants to maximize the profit with the given set of inequalities. The objective function and all the constraints are:

(a) $Z = 700x + 500y$
$2x + 3y \leq 240$
$2x + y \leq 100$
$x, y \geq 0$

(b) $Z = 700x + 500y$
$4x + 3y \leq 240$
$2x + y \leq 100$
$x, y \geq 0$

(c) $Z = 700x + 500y$
$4x + 3y \leq 240$
$2x + y \leq 100$

(d) None of these

4.10. A dietician wishes to mix together two kinds of food so that the vitamin content of the mixture is at least 9 units of vitamin A, 7 units of vitamin B, 10 units of vitamin C and 12 units of vitamin D. The vitamin content per Kg. of each food is shown below:

	A	B	C	D
Food I:	2	1	1	2
Food II:	1	1	2	3

Assuming x units of food I is to be mixed with y units of food II the situation can be expressed as:

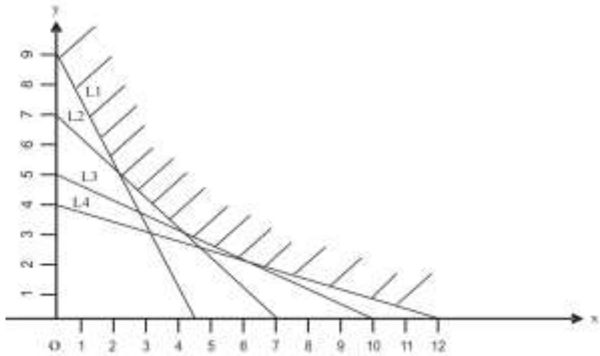
(a) $2x + y \leq 9$
$x + y \leq 7$
$x + 2y \leq 10$
$2x + 3y \leq 12$
$x > 0, y > 0$

(b) $2x + y \geq 30$
$x + y \leq 7$
$x + 2y \geq 10$
$x + 3y \geq 12$

(c) $2x + y \geq 9$
$x + y \geq 7$
$x + y \leq 10$
$x + 3y \geq 12$

(d) None of these

4.11. Graphs of four equations are drawn below:



L1: $2x + y = 9$, L2: $x + y = 7$, L3: $x + 2y = 10$, L4: $x + 3y = 12$

The common region (shaded part) indicated on the diagram is expressed by the set of inequalities.

(a) $2x + y \leq 9$
$x + y \geq 7$
$x + 2y \geq 10$
$x + 3y \geq 12$

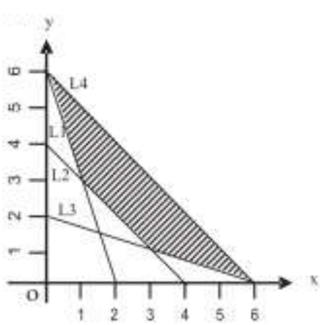
(b) $2x + y \geq 9$
$x + y \leq 7$
$x + 2y \geq 10$
$x + 3y \geq 12$

(c) $2x + y \leq 9$
$x + y \geq 7$
$x + 2y \geq 10$
$x + 3y \geq 12$
$x \geq 0, y \geq 0$

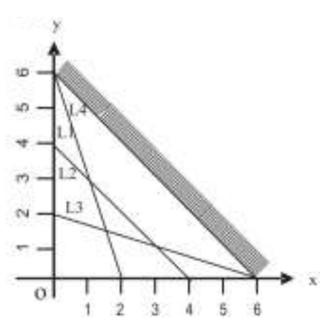
(d) None of these

4.12. The common region satisfied by the inequalities L1: $3x + y \geq 6$, L2: $x + y \geq 4$, L3: $x + 3y \geq 6$ and L4: $x + y \leq 6$ is indicated by:

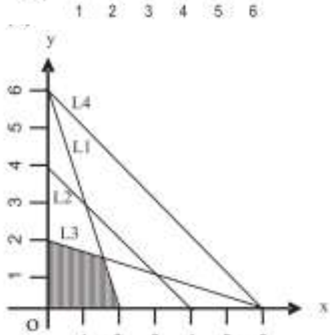
(a)



(b)



(c)



(d) None of these

4.13. A firm makes two types of products: Type A and Type B. The profit on product A is Rs.20 each and that on product B is Rs.30 each. Both types are processed on three machines M1, M2 and M3. The time required in hours by each product and total time available in hours per week on each machine are as follow:

Machine	Product A	Product B	Available Time
M1	3	3	36
M2	5	2	50
M3	2	6	60

The constraints can be formulated taking x_1 = number of units A and x_2 = number of unit of B as:

(a)
$x_1 + x_2 \leq 12$
$5x_1 + 2x_2 \leq 50$
$2x_1 + 6x_2 \leq 60$

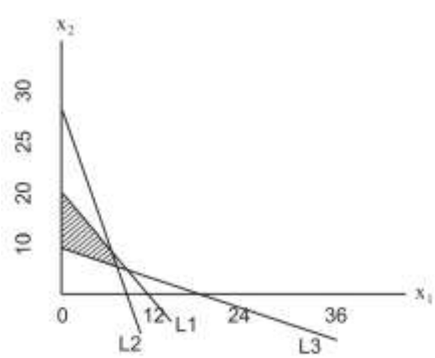
(b)
$3x_1 + 3x_2 \geq 36$
$5x_1 + 2x_2 \leq 50$
$2x_1 + 6x_2 \geq 60$
$x_1 \geq 0, x_2 \geq 0$

(c)
$3x_1 + 3x_2 \leq 36$
$5x_1 + 2x_2 \leq 50$
$2x_1 + 6x_2 \leq 60$
$x_1 \geq 0, x_2 \geq 0$

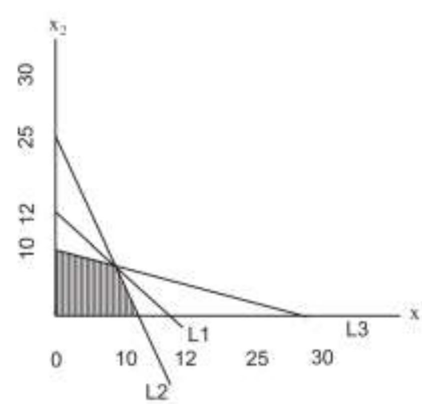
(d) None of these

4.14. The set of inequalities L1: $x_1 + x_2 \leq 12$, L2: $5x_1 + 2x_2 \leq 50$, L3: $x_1 + 3x_2 \leq 30$, $x_1 \geq 0$ and $x_2 \geq 0$ is represented by:

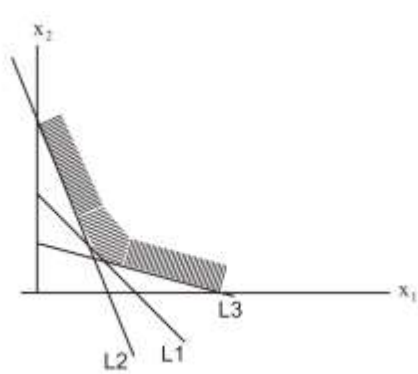
(a)



(b)



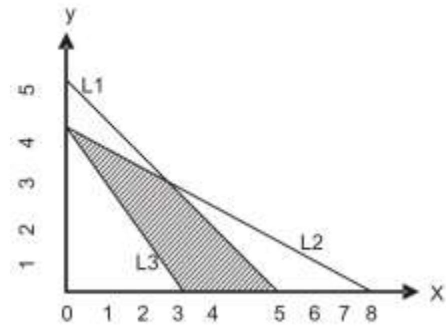
(c)



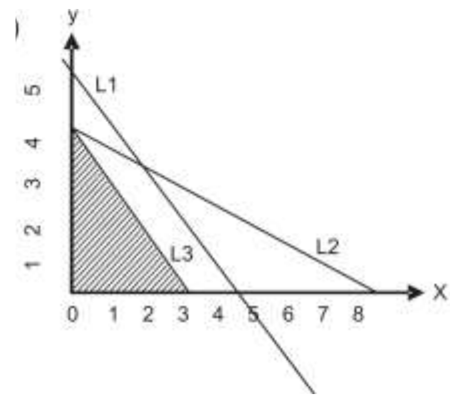
(d) None of these

4.15. The common region satisfying the set of inequalities $x \geq 0$, $y \geq 0$, L1: $x + y \leq 5$, L2: $x + 2y \leq 8$ and L3: $4x + 3y \geq 12$ is indicated by:

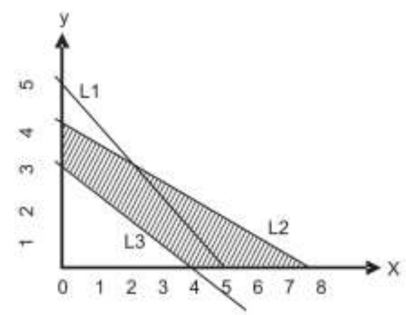
(a)



(b)



(c)



(d) None of these

4.16. A manufacturer produces two products X1 and X2. Resources available for the production of these two items are restricted to 200 support staff hours, 320 machine hours and 280 labour hours. X1 requires for its production 1 support staff hour, 1 machine hour and 2 labour hours. X2 requires 1 support staff hour, 2 machine hours and 0.8 labour hour. X1 yields Rs.300 profit per unit and X2 yields Rs.200 profit per unit. The manufacturer wants to determine the profit maximizing weekly output of each product while operating within the set of resource limitations. Situation of the above data in the form of equations and inequalities is:

(a) $Z = 300x + 200y$
$x + y \leq 200$
$x + 2y \leq 320$
$2x + 0.8y \leq 280$
$x \geq 0, y \geq 0$

(b) $Z = 300x + 200y$
$x + y \leq 200$
$3x + 2y \leq 320$
$2x + 0.8y \leq 280$
$x \geq 0, y \geq 0$

(c) $Z = 300x + 200y$
$x + y \leq 200$
$x + 2y \leq 320$
$2x + 0.8y \leq 280$

(d) None of these

4.17. A factory is planning to buy some machine to produce boxes and has a choice of B-1 or B-9 machines. Rs.9.6 million has been budgeted for the purchase of machines. B-1 machines costing Rs.0.3 million each require 25 hours of maintenance and produce 1,500 units a week. B-9 machines costing Rs.0.6 million each require 10 hour of maintenance and produce 2,000 units a week. Each machine needs 50 square meters of floor area. Floor area of 1,000 square meters and maintenance time of 400 hours are available each week. Since all production can be sold, the factory management wishes to maximize output. Situation of above data in the form of objective function and constraints is:

(a) $Z = 1500x + 2000y$
$0.3x + 0.6y \geq 9.6$
$25x + 10y \leq 400$
$50x + 50y \leq 1000$
$x \geq 0, y \geq 0$

(b) $Z = 1500x + 2000y$
$0.3x + 0.6y \geq 9.6$
$25x + 10y \leq 400$
$50x + 50y \geq 1000$
$x \geq 0, y \geq 0$

(c) $Z = 1500x + 2000y$
$0.3x + 0.6y \leq 9.6$
$25x + 10y \leq 400$
$50x + 50y \leq 1000$
$x \geq 0, y \geq 0$

(d) $Z = 1500x + 2000y$
$0.3x + 0.6y \leq 9.6$
$25x + 10y \geq 400$
$50x + 50y \leq 1000$
$x \geq 0, y \geq 0$

4.18. A company produces two products a and b. Production of A will be at least five times the production of B. This situation can be expressed as:

(a)	$b \geq 5a$	(b)	$b \leq 5a$
(c)	$a \leq 5b$	(d)	$a \geq 5b$

4.19. A manufacturer produces two products A and B. Each unit of A requires 3kg of raw material and 2 labour hours whereas each unit of B requires 2kg or raw material and 2 labour hours. Total raw material available is 200kg whereas total labour hours available are 150 hours. This situation can be expressed as:

(a)	$3a+2b \geq 200$ and $2a+2b \geq 150$	(b)	$3a+2b \leq 200$ and $2a+2b \leq 150$
(a)	$3a+2b \leq 150$ and $2a+2b \leq 200$	(d)	$3a+2b \geq 150$ and $2a+2b \geq 200$

4.20. Inequalities $x \geq 0$ and $y \geq 0$ suggest that feasible region will be in ____ quadrant.

(a)	1st	(b)	2nd
(c)	3rd	(d)	4th

4.21. Which of the following values of x and y does not satisfy the inequality $2x+3y \leq 6$

(a)	(1,1)	(b)	(0,1)
(c)	(3,1)	(d)	(1,0)

4.22. A machine can work for 16 hours per day at most. It takes 4 hours to produce one unit of product A and 8 hours to produce one unit of product B. This information can be expressed as:

(a)	$4a+8b \leq 16$	(b)	$4a+8b \geq 16$
(c)	$4a+8b = 16$	(d)	Not possible to express information

4.23. Which of the following constraint will have a feasible region with coordinate (2, 3) in it?

(a) $10x+5y \leq 40$ (b) $10x+5y \leq 30$
(c) $10x+5y \leq 10$ (d) $10x+5y \leq 20$

4.24. Pharma limited produces two medicines. Each of which requires two types of raw material with following details:

	Product A(kg/unit)	Product B(kg/unit)	Total available (kg)
Raw material X	2	3	120
Raw material Y	4	6	360

Form appropriate constraints as per above information

(a) $2a+3b \leq 120$ and $4a+6b \leq 360$ (b) $2x+4y \leq 120$ and $3x+6y \leq 360$
(c) $4a+6b \leq 120$ and $2a+3b \leq 360$ (d) $2x+3y \leq 120$ and $4a+6b \leq 360$

4.25. The maximum number of constraints while solving a business problem using linear programming is

(a) 1 (b) 2
(c) 3 (d) there is no specific limit

4.26. Feasible region is a ____

(a) Region which satisfies all given constraints
(b) Region which is in first quadrant
(c) Region above the last mentioned constraint
(d) Always below all the constraints in a given scenario

4.27. Corner point theorem states ____

(a) Optimal solution is at one of the corners of feasible region.
(b) Optimal solution is within feasible region.
(c) Profit is maximum at all corners of a feasible region.
(d) Revenue is maximum at all corners of a feasible region.

4.28. Linear programming model can be used for ____

(a) Profit maximization (b) Cost minimization
(c) Both (d) None

4.29. For the following set of inequalities and objective function. Identify the optimal solution using corner point theorem

Constraints

i. $x \geq 0$
ii. $y \geq 0$
iii. $x+y \leq 8$

Objective function

Profit = $10x+2y$

(a) (8, 0) (b) (0, 8)
(c) (4, 4) (d) (0, 0)

4.30. For the following set of inequalities and objective function. Identify the optimal solution using corner point theorem

Constraints

i. $x \geq 0$
ii. $y \geq 0$
iii. $2x+y \leq 8$
iv. $2x+2y \leq 10$

Objective function

Profit = $10x+2y$

(a) (0, 8) (b) (5, 0)
(c) (0, 5) (d) (8, 0)

4.31. For the following set of inequalities and objective function. Identify the optimal solution using corner point theorem

Constraints

i. $x \geq 1$
ii. $y \geq 2$
iii. $x+y \leq 8$

Objective function

Cost = $10x+2y$

(a) (0, 0) (b) (1, 2)
(c) (2, 1) (d) (0, 0)

4.32. Identify the redundant constraint from the following set of constraints:

i. $x+y \leq 6$
ii. $x \geq 0$
iii. $y \geq 0$
iv. $4x+2y \leq 8$

(a) $4x+2y \leq 8$ (b) $x \geq 0$
(c) $y \geq 0$ (d) $x+y \leq 6$

4.33. A company wants to make product A which will contain both chemical x and chemical y. The product must contain at least 1 unit of x and 2 units of y. The company has to be careful that total of x and y units in the product must not exceed 8 units.

If the cost of each unit of x and y is 10 and 2 respectively, formulate constraints and objective function respectively.

(a) Constraints
i. $x \geq 1$
ii. $y \geq 2$
iii. $x+y \leq 8$
Objective function
Cost = $10x+2y$

(b) Constraints
i. $x \geq 1$
ii. $y \geq 2$
iii. $x+y \leq 8$
Objective function
Cost = $2x+10y$

(c) Constraints
i. $x \geq 0$
ii. y ≥0
iii. x+y ≤ 8
Objective function
Cost = 10x+2y

(d) Constraints
i. $x \geq 1$
ii. y ≥2
iii. x+y ≤ 2
Objective function
Cost = 10x+2y

4.34. A company produces two types of cola drinks. Each of which requires two types of raw material with following details:

	Product A(kg/unit)	Product B(kg/unit)	Total available (kg)
Raw material X	3	4	480
Raw material Y	2	6	540

How many units of each type of cola drink may be produced to maximise the profit, if the profit on each unit of A and B is Rs 1,800 and Rs. 6,300 respectively?

(a) (72, 66) (b) (0, 90)
(c) (160, 0) (d) (66, 72)

4.35. For the following constraints:

i. x+y ≥ 40
ii. x+y ≤ 10

(a) There is no feasible region (b) One feasible region
(c) Two feasible regions (d) Three feasible regions

4.36. Minimum number of constraints to draw a feasible region is/are:

(a) 1 (b) 2
(c) 3 (d) 4

4.37. A bounded feasible region will:

(a) Have both maximum and minimum value (b) Have maximum value only
(c) Have minimum value only (d) None of these

4.38. Y ≤ X will have a feasible region: __

(a) Above Y axis only (b) Below Y axis only
(c) Above X axis only (d) Below or on the line Y=X

AT A GLANCE
SPOTLIGHT
STICKY NOTES

ANSWERS TO SELF-TEST QUESTIONS

4.1(i)	**4.1(ii)**	**4.1(iii)**	**4.1(iv)**	**4.2**	**4.3**
(b)	(c)	(a)	(b)	(a)	(c)
4.4	**4.5**	**4.6**	**4.7**	**4.8**	**4.9**
(d)	(b)	(a)	(b)	(b)	(b)
4.10	**4.11**	**4.12**	**4.13**	**4.14**	**4.15**
(d)	(d)	(a)	(c)	(b)	(a)
4.16	**4.17**	**4.18**	**4.19**	**4.20**	**4.21**
(a)	(c)	(d)	(b)	(a)	(c)
4.22	**4.23**	**4.24**	**4.25**	**4.26**	**4.27**
(a)	(a)	(a)	(d)	(a)	(a)
4.28	**4.29**	**4.30**	**4.31**	**4.32**	**4.33**
(c)	(a)	(b)	(b)	(d)	(a)
4.34	**4.35**	**4.36**	**4.37**	**4.38**	
(b)	(a)	(a)	(a)	(d)	

CHAPTER 5

FINANCIAL MATHEMATICS

IN THIS CHAPTER

AT A GLANCE

Interest is the extra amount an investor receives after investing a certain sum, at a certain rate for a certain time. In other words, interest is the additional amount of money paid by the borrower to the lender for the use of money loaned to him. The interest can be simple (fixed percentage on the principal amount) and compound (accumulated interest over period on the principal amount). Based on the interest charged, original amount invested or loaned might be repaid on a periodic basis or accumulated and be repaid at the end of the period.

Discounting estimates the present value of a future cash flow at a given interest rate. The amount expected at a specified time in the future is multiplied by a discount factor to give the present value of that amount at a given rate of interest. Discounting and Compounding help evaluate and compare present and future cash flows for decisions on investment and borrowing matters.

1. SIMPLE INTEREST

Simple interest is where the annual interest is a fixed percentage of the original amount (the principal) borrowed or invested.

- *For example:*

A person borrows Rs 10,000 at 10% with principal and interest to be repaid after 3 years.

	Rs
Amount owed at the start of year 1	10,000
Interest for year 1 (10%)	1,000
Interest for year 2 (10%)	1,000
Interest for year 3 (10%)	1,000
Total interest	3,000
Amount owed at the end of year 3	13,000

The closing balance of 13,000 must be repaid to the lender at the end of the third year.

Calculating interest for a small number of periods is easy but becomes time consuming as the length of a loan grows.

The interest based on the original principal might be repaid on a periodic basis or accumulated and be repaid at the end. For which formula can be used

- *Formula:*

Amount repayable at the end of a loan	$S = P(1 + rn)$
Interest due on a loan	$I = Prn$

Where:

S = amount to be paid or received at the end of period n or Future Value

I = interest charge

P = principal (amount borrowed or invested) or Present value

r = period interest rate

n = number of time periods that the loan is outstanding

- *For example:*

A person borrows Rs 10,000 at 10% simple interest for 5 years. Calculate annual interest, interest for the period, and amount payable at the end of 5 years.

Annual interest charge can be found using

I=Prn where P = 10,000; r= 0.1; n=1

I = 10,000×0.1×1 =1,000

Interest charge over five years

Here the n = 5

I = 10,000×10%×5 years=5,000

The person might pay the interest on an annual basis (1,000 per annum for 5 years) and then the principal at the end of the loan (10,000 after 5 years) or might pay the principal together with the five years interest at the end of the loan (15,000 after 5 years).

This may be calculated as follows:

Amount repayable at the end of 5 years

S=P(1+rn)

$S = 10{,}000 \times (1 + [0.1 \times 5])$

$S = 10{,}000 \times (1 + [0.5])$

$S = 10{,}000 \times 1.5$

$S = 15{,}000$

What amount of interest would a person receive at end of 5 years, if he decides to deposit Rs. 1,600 in the bank at 8% per annum?

I=Prn where P = 1600; r= 0.08; n=5

I = 1600×0.08×5 =640

What amount of loan a business had borrowed if repayment of loan requires Rs. 32,000 to be paid after 10 years (the simple interest was 15% per annum)?

Borrowed amount can be found by

S=P(1+rn)

S = 32,000; r= 0.15; n=10

$32{,}000 = \mathrm{P} \times (1 + [0.15 \times 10])$

$32{,}000 = P \times (1 + [1.5])$

$32{,}000 = P \times 2.5$

$\mathrm{P} = \mathrm{Rs.}\,12{,}800$

A person borrows Rs.100,000 for 4 years at an interest rate of 7%. What must he pay to clear the loan at the end of this period?

Amount repayable at the end of 4 years can be found by substituting the values

P = 100,000; r =0.07; n = 4

$\mathrm{S} = \mathrm{P}(1 + \mathrm{rn})$

$\mathrm{S} = 100{,}000 \times (1 + [0.07 \times 4])$

$\boldsymbol{S} = \mathbf{100{,}000} \times (\mathbf{1} + \mathbf{0.28})$

$\boldsymbol{S} = \mathbf{128{,}000}$

A person invests Rs. 60,000 for 5 years at an interest rate of 6%. What is the total interest received from this investment?

Interest charged for the period can be found by putting the values P = 60,000; r = 0.06; and n = 5

I=Prn

I = 60,000×0.06×5

I = 18,000

AT A GLANCE

SPOTLIGHT

STICKY NOTES

Periods less than a year and other problems

Interest rates given in questions are usually annual interest rates. This means that they are the rate for borrowing or investing for a whole year.

For shorter periods, the annual simple interest rate is pro-rated to find the rate that relates to a shorter period.

- *For example:*

A person borrows Rs 10,000 at 10% simple interest for 8 months. What will be charged interest and the amount paid after the period?

In this case n is a single period of 8 months

Annual interest charge can be found by putting the values P = 10,000, r = $(0.1)(\frac{8}{12})$

I=Prn

I = 10,000×[(10%×8/12)×1]=667

Amount repayable at the end of 8 months

$S = P(1 + rn)$

$S = 10{,}000 \times (1 + [0.1 \times 8/12] \ \times 1)$

S=10,667

A person borrows Rs. 75,000 for 4 months at an interest rate of 8%. (This is an annual rate) What must he pay to clear the loan at the end of this period?

Amount repayable at the end of 4 months can be found by substituting the values

P = 75,000; r = 0.08 × 4/12 ; n = 1

$S = P(1 + rn)$

$S = 75{,}000 \times \left(1 + \left[(0.08 \times \frac{4}{12}) \times 1\right]\right)$

$\boldsymbol{S = 75,000 \times (1 + 0.02667)}$

$\boldsymbol{S = 77,000}$

A person borrows Rs.60,000 at 8%. At the end of the loan he repays the loan in full with a cash transfer of Rs.88,800. What was the duration of the loan?

Time period n for the loan can be found by substituting the values

P = 60,000; r = 0.08 ; S = 88,800

$S = P(1 + rn)$

$88{,}800 = 60{,}000 \times (1 + [0.08 \times n])$

$\boldsymbol{88,800 = 60,000\ (1 + 0.08n)}$

$\boldsymbol{88,800 = 60,000 + 4,800n}$

88,800 – 60,000 = 4,800n

28,800 = 4,800n

n=6

Time duration for the loan was 6 years.

A person invests Rs.90,000 for 6 years. At the end of the loan she receives a cash transfer of Rs.122,400 in full and final settlement of the investment. What was the interest rate on the loan?

Interest rate r for the loan can be found by substituting the values

P = 90,000; n=6 ; S = 122,400

$S = P(1 + rn)$

$122{,}400 = 90{,}000 \times (1 + [r \times 6])$

$\mathbf{122,400 = 90,000\ (1 + 6}\boldsymbol{r}\mathbf{)}$

$\mathbf{12,2400 = 90,000 + 540,000}\boldsymbol{r}$

122,400 – 90,000 = 540,000r

$32{,}400 = 540{,}000\text{r}$

r=0.06

Loan interest rate was 6%.

AT A GLANCE
SPOTLIGHT
STICKY NOTES

2. COMPOUNDING

Compounding is the process of accumulating interest on an investment over time to earn more interest[1]. Annual interest based on the amount borrowed plus interest accrued to date is hence referred to as Compound Interest.

The interest accrued to date increases the amount in the account and interest is then charged on that new amount.

One way to think about this is that it is a series of single period investments. The balance at the end of each period (which is the amount at the start of the period plus interest for the period) is left in the account for the next single period. Interest is then accrued on this amount and so on.

The accumulated interest and the investment is also referred to as the Future Value (FV) of the investment that has grown over time at the given interest rate.

- *For example:*

A person borrows Rs 10,000 at 10% to be repaid after 3 years.

	Rs
Amount owed at the start of year 1	10,000
Interest for year 1 (10%)	1,000
Amount owed at the end of year 1 (start of year 2)	11,000
Interest for year 2 (10%)	1,100
Amount owed at the end of year 2 (start of year 3)	12,100
Interest for year 3 (10%)	1,210
Amount owed at the end of year 3	13,310

The closing balance of 13,310 must be repaid to the lender at the end of the third year.

In the example, it is to see that interest for the year is increasing year after year, unlike simple interest which remains constant. This is because the amount on which interest is applied increases.

This suggests the following formula.

- *Formula:*

$$S_n = \mathrm{So} \times (1 + \mathrm{r})^{\mathrm{n}}$$

Where:

S_n = final cash flow at the end of the loan (the amount paid by a borrower or received by an investor or lender). Or Future Value

S_o = initial investment or Principle amount or Present Value

r = period interest rate

n = number of periods

Note that the $(1 + r)^n$ term is known as a compounding factor

[1] Ross, Westerfield & Bradford (2008). Fundamentals Of Corporate Finance. 8th Edition. Mcgraw-Hill/Irwin. Pg. 122

- *For example:*

A person borrows Rs 10,000 at 10% to be repaid after 3 years. In order to calculate the due amount at the end of the period following values must be substituted in the formula:

$S_o = 10{,}000$; $r = 0.1$; $n = 3$

$S_n = So \times (1 + r)^n$

$S_n = 10{,}000 \times (1 + 0.1)^3$

$S_n = 10{,}000 \times (1.1)^3$

$S_n = 13310$

What will be the compound interest charged if an investor borrowed Rs. 20,000 from bank for 8 years at 10% per annum?

$S_o = 20{,}000$; $r = 0.1$; $n = 8$

$S_n = 20{,}000 \times (1 + 0.1)^8$

$S_n = 20{,}000 \times (1.1)^8$

$S_n = 20{,}000 \times 2.1436$

$S_n = 42{,}872$

Interest charged 42,872 – 20,000 = Rs. 22872

Periods less than a year

Similar to simple interest calculations, it is important to remember that the rate of interest r must be consistent with the length of the period n.

There are two methods for calculating a rate for a shorter period from an annual rate:

- **Compounding by parts**: When an investment pays interest at an annual rate compounded into parts of a year, the rate that relates to the part of the year is simply the annual interest divided by the number of parts of the year. The compounding formula then becomes
  - *Formula:*

$$S_n = S_0 \times (1 + \frac{r}{m})^{n \times m}$$

Where:

r = period interest rate

n = number of periods

m = number of parts of the year

- **Continuous compounding:** When a loan is compounded continuously, it means that interest is being calculated at every smallest instance. Although continuous compounding plays significant role in financial compounding, it is not considered realistic to have an infinite number of periods for interest calculation.
  - *Formula:*

$$S_n = S_o\, e^{rn}$$

Where:

r = period interest rate

n = number of period

- *For example:*

A bank lends Rs 100,000 at an interest rate of 12.55% per annum compounded quarterly. The duration of the loan is 3 years. How much will the bank's client have to pay?

Since compounded in parts, the amount due to the bank after the period would be calculated using following formula and values;

$S_o = 100{,}000; \quad r = 0.1255; \quad m=4; \quad n=3$

$$S_n = S_o \times (1+\frac{r}{m})^{n \times m}$$

$$S_n = 100{,}000 \times (1+\frac{0.1255}{4})^{3 \times 4}$$

$$S_n = 100{,}000 \times (1+0.031375)^{12}$$

$$S_n = 100{,}000 \times (1.031375)^{12}$$

$$S_n = 144{,}876.919$$

Mr. A invested Rs. 100,000 at a 12% interest rate. How much he would earn by the end of 5 years if compounded continuously?

$$S_n = So\ e^{rn}$$

Here So=100000; r = 0.12; n= 5

$$S_n = 100000\ e^{0.12(5)}$$

$$S_n = 100000\ e^{0.6}$$

$$S_n = 182211.88$$

A person borrows Rs. 100,000 for 3 years at an annual interest rate of 8%. What must he pay to clear the loan at the end of the period if compounded a) annually b) semiannually c) quarterly d) monthly e) weekly f) daily and g) hourly?

Annually	*S_o = 100,000; r = 0.08, m=1; n=3*	$S_n = S_o \times (1+\frac{r}{m})^{n \times m}$ $S_n = 100{,}000 \times (1+0.08)^3$ $S_n = 100{,}000 \times (1.08)^3$ $S_n = 100{,}000 \times 1.2597$ S_n = 125,971.2
Semi annually	*S_o = 100,000; r = 0.08; m=2; n=3*	$S_n = S_o \times (1+\frac{r}{m})^{n \times m}$ $S_n = 100{,}000 \times (1+\frac{0.08}{2})^{3 \times 2}$ $S_n = 100{,}000 \times (1+0.04)^6$ $S_n = 100{,}000 \times (1.04)^6$ $S_n = 100{,}000 \times 1.26531$ Sn = 126,531.9

Quarterly	*S₀ = 100,000; r = 0.08; m=4; n=3*	$S_n = S_o \times (1+\frac{r}{m})^{n \times m}$ $S_n = 100,000 \times (1+\frac{0.08}{4})^{3 \times 4}$ $S_n = 100,000 \times (1+0.02)^{12}$ $S_n = 100,000 \times (1.02)^{12}$ $S_n 100,000 \times 1.26824$ $\boldsymbol{S_n}$ = 126,824.17
Monthly	*S₀ = 100,000; r = 0.08; m=12; n=3*	$S_n = S_o \times (1+\frac{r}{m})^{n \times m}$ $S_n = 100,000 \times (1+\frac{0.08}{12})^{3 \times 12}$ $S_n = 100,000 \times (1+0.00666)^{36}$ $S_n = 100,000 \times (1.00066)^{36}$ $S_n = 100,000 \times 1.270237$ $\boldsymbol{S_n}$ = 127,023.7
Weekly	*S₀ = 100,000; r = 0.08; m=52; n=3*	$S_n = So \times (1+\frac{r}{m})^{n \times m}$ $S_n = 100,000 \times (1+\frac{0.08}{52})^{3 \times 52}$ $S_n = 100,000 \times (1+0.001538)^{156}$ $S_n = 100,000 \times (1.001538)^{156}$ $S_n = 100,000 \times 1.271014$ $\boldsymbol{S_n}$ = 127,101.47
Daily	*S₀ = 100,000; r = 0.08; m=365; n=3*	$S_n = S_o \times (1+\frac{r}{m})^{n \times m}$ $S_n = 100,000 \times (1+\frac{0.08}{365})^{3 \times 365}$ $S_n = 100,000 \times (1+0.0002191)^{1095}$ $S_n = 100,000 \times (1.0002191)^{1095}$ $S_n = 100,000 \times 1.271215$ $\boldsymbol{S_n}$ = 127,121.57
Hourly	*S₀ = 100,000; r = 0.08; m=8760; n=3*	$S_n = So \times (1+\frac{r}{m})^{n \times m}$ $S_n = 100,000 \times (1+\frac{0.08}{8760})^{3 \times 8760}$ $S_n = 100,000 \times (1+0.00000913)^{26280}$ $S_n = 100,000 \times (1.00000913)^{26280}$ $S_n = 100,000 \times 1.271247$ $\boldsymbol{S_n}$ = 127,124.7

A person invests Rs. 60,000 for 6 years at an interest rate of 6%. What is the total interest received from this investment if compounded annually? Continuously?

Annually	*S_o = 60,000; r = 0.06, n=6*	$S_n = S_o \times (1 + r)^n$ $S_n = 60{,}000 \times (1 + 0.06)^6$ $S_n = 60{,}000 \times (1.06)^6$ $S_n = 60{,}000 \times 1.41851$ S_n = 85,111.146 Interest received = 85,111.146 – 60,000 = 25,111.146
Continuously	*S_o = 60,000; r = 0.06, n=6*	$S_n = S_o\, e^{rn}$ $S_n = 60{,}000\, e^{0.06(6)}$ $S_n = 60{,}000\, e^{0.36}$ $S_n = 60{,}000 \times 1.4333294$ $S_n = 85{,}999.76$ Interest received = 85999.76 – 60000 = 25,999.76

A person invests his inheritance at an interest rate of 6% compounding annually. He receives Rs. 337,900 at the end of the 9 years. What is inherited amount?

$S_n = S_o \times (1 + r)^n$

S_n=337900; r=0.06; n=9

$337{,}900 = S_o \times (1.06)^9$

$337{,}900 = S_o \times 1.68947$

S_o = 337,900/ 1.68947

S_o = 200,002.49 or ~ Rs. 200,000

Nominal and effective rates

The nominal rate, also referred as stated rate, is the "face value" of a loan. However, this might not be the true annual cost because it does not take into account the way the loan is compounded. Effective rates take into account compounding periods and compare annual interests for different compounding periods.

- *For example:*

  A company wants to borrow Rs.1,000. It has been offered two different loans.

  Loan A charges interest at 10% per annum and loan B at 5% per 6 months.

  Which loan should it take?

  The nominal interest rates are not useful in making the decision because they relate to different periods. They can be made directly comparable by restating them to a common period. This is usually per annum.

	Loan A	Loan B
Amount borrowed	1,000	1,000
Interest added:		
		50
		1,050
	100	52.5
	1,100	1,102.5
Effective rate per annum	10%	10.25%
	$\frac{100}{1,000} \times 100$	$\frac{102.5}{1,000} \times 100$
	$= 10\%$	$= 10.25\%$

Generally, nominal interest rate is less than the effective rate. This is because, compounding occurs more frequently than once a year. The more often compounding in a year, the higher the effective interest rates.

- *Formula:*

  $$\text{Effective annual rate} = (1 + \text{r})^{\text{n}} - 1$$

  Where:

  r = Known interest rate for a known period (usually annual)

  n = number of times the period for which the rate is unknown fits into one single year

In the above example:

The effective annual rate could be calculated using:

	Effective annual rate $= (1 + \text{r})^{\text{n}} - 1$
Loan A	Effective annual rate $= (1.1)^1 - 1 = 0.1$ or 10%
Loan B	Effective annual rate $= (1.05)^2 - 1$ $= 0.1025$ or 10.25%

- *For example:*

  If the 6-monthly and 3 monthly interest rates are 4.9% and 2.41% what will be the annual effective interest rates?

  Given by: Effective annual rate $= (1 + \text{r})^{\text{n}} - 1$

  Effective rate for 6 monthly compounding: $(1 + 0.049)^2 - 1 = 0.100$ or 10%

  Effective rate for 3 monthly compounding: $(1 + 0.0241)^4 - 1 = 0.00999$ or $\sim 10\%$

  If a bank charges 16% per year for a loan compounded monthly. What will be the effective interest rate?

  Effective annual rate $= (1 + \text{r})^{\text{n}} - 1$

  Effective annual rate $= (1 + \frac{0.16}{12})^{12} - 1$

  Effective annual rate $= (1 + 0.01333)^{12} - 1$

  Effective annual rate $= (1.01333)^{12} - 1$

  Effective annual rate $= 1.17227 - 1$

  Effective annual rate $= 0.17227$ or $\sim 17\%$

AT A GLANCE SPOTLIGHT STICKY NOTES

What will be the effective interest rate when nominal interest rate is 12% per annum compounded monthly?

Period rate $= (1 + 0.12/12)^{12(1)} - 1$

Period rate $= (10.1)^{12} - 1$

Period rate $= 1.1268 - 1$

Period rate = 0.1268 or ~12.7 %

3. DISCOUNTING

Discounting estimates the present day equivalent (present value which is usually abbreviated to ***PV***) of an amount at a specified time in the future at a given rate of interest. Discounting is the reverse of compounding.

PV is thus the current value of future cash flows discounted at the appropriate discount rate[2].

- An amount expected at a specified time in the future is multiplied by a discount factor to give the present value of that amount at a given rate of interest.
- The discount factor is the inverse of a compound factor for the same period and interest rate. Therefore, multiplying by a discount factor is the same as dividing by a compounding factor.

▶ *Formula:*

$$S_o = S_n \times \frac{1}{(1+r)^n}$$

Or

$$\text{Present value (PV)} = \text{Future Value (FV)} \times \frac{1}{(1+r)^n}$$

Where:

r = the period interest rate (cost of capital)

n = number of periods

Discount Factor = $\frac{1}{(1+r)^n}$

The present value of a future cash flow is the amount that an investor would need to invest today to receive that amount in the future.

▶ *For example:*

A person expects to receive Rs 13,310 in 3 years. If the person faces an interest rate of 10% what is the present value of this amount?

Here Future cash flow would be 13310, n = 3 and r = 0.1

$$\text{Present value} = 13{,}310 \times \frac{1}{(1.1)^3}$$

$$\text{Present value} = 10{,}000$$

It is important to realize that the present value of a cash flow is the equivalent of its future value in the future point in time. Using the above example to illustrate this, Rs 10,000 today is exactly the same as Rs 13,310 in 3 years at an interest rate of 10%. The person in the example would be indifferent between the two amounts. He would look on them as being identical at different points in time.

Also the present value of a future cash flow is a present day cash equivalent. The person in the example would be indifferent between an offer of Rs 10,000 cash today and Rs 13,310 in 3 years.

▶ *Complex Examples:*

What is the value of Rs. 50,000 to be paid 6 years from now at 6.5% interest rate compounded annually?

Sn = 50,000; r = 0.065; n = 6

$$50{,}000 = \text{So} \times (1 + 0.065)^6$$

$$50{,}000 = \text{So} \times (1.065)^6$$

$$50{,}000 = \text{So} \times 1.459$$

$$\text{So} = 34267$$

[2] Ross, Westerfield & Bradford (2008). Fundamentals Of Corporate Finance. 8th Edition. Mcgraw-Hill/Irwin. Pg. 129

How much would an investor need to invest now in order to have Rs 100,000 after 12 months, if the compound interest on the investment is 0.5% each month?

The investment 'now' must be the present value of Rs 100,000 in 12 months, discounted at 0.5% per month. This makes FCF = 100,000, n=12 and r = 0.005

$$PV = 100{,}000 \times \frac{1}{(1.005)^{12}} = \text{Rs } 94{,}190$$

The price of a car increases from Rs. 20,000 to Rs. 48,000 in just 4 years. What is the rate at which price of the car increases?

For the question above, both compounding and discounting formula would yield same result. We will use the formula for Present Value:

Using the formula and substituting the values FCF = 48,000, n = 4, PV=20,000

$$20{,}000 = 48{,}000 \times \frac{1}{(1+r)^4}$$

$$(1 + r)^4 = \frac{48{,}000}{20{,}000}$$

$$(1 + r)^4 = 2.4$$

Multiplying ¼ on both sides

$$(1 + r) = 2.4^{1/4}$$

$$r = 2.4^{1/4} - 1$$

r = 24.46%

Discount tables

Tables of discount rates which list discount factors by interest rates and duration are also used in order to reach present value.

- *Illustration:*

(Full tables are given as an appendix to this text).

	Discount rates (r)					
(n)	**5%**	**6%**	**7%**	**8%**	**9%**	**10%**
1	0.952	0.943	0.935	0.926	0.917	0.909
2	0.907	0.890	0.873	0.857	0.842	0.826
3	0.864	0.840	0.816	0.794	0.772	0.751
4	0.823	0.792	0.763	0.735	0.708	0.683

Where:

n = number of periods

r = rate of interest

- *For example:*

Calculate the present value of Rs 60,000 received in 4 years assuming a cost of capital of 7%.

Using the formula and substituting the values FCF = 60,000, r = 0.07, n = 4

$$\mathrm{PV} = 60{,}000 \times \frac{1}{(1.07)^4} = 45{,}773$$

From the above table

Column 3 and row 4 is the cell value at 7% rate for four years. Factor = 0.763

The factor multiplied by the FCF gives the Present Value:

$$\mathrm{PV} = 60{,}000 \times 0.763 = 45{,}780$$

What is the present value of an investment of Rs. 6,000 due in 3 years with 8% interest compounded quarterly?

Using the formula and substituting the values FCF = 6,000, r = 0.08/4, n = 3*4

$$\mathrm{PV} = 6{,}000 \times \frac{1}{(1.02)^{12}}$$

$$\mathrm{PV} = \mathrm{Rs.}\ 4{,}730.959$$

From the Present Value table in the Annexure

Table 1 = Column 2 and row 12 is the cell value at 2% rate for 12 periods. Factor = 0.788

The factor multiplied by the FCF gives the Present Value:

$$\mathrm{PV} = 6{,}000 \times 0.788 = 4{,}728$$

A father intends to buy a car for his daughter on her 18th birthday. The car cost Rs. 2,000,000 and is expected to remain same. How much he needs to save today at 12% interest per annum if his daughter is 10 years old as of today.

Using the formula and substituting the values FCF = 2000000, r = 0.12, n = 8

$$\mathrm{PV} = 2{,}000{,}000 \times \frac{1}{(1.12)^8}$$

$$\mathrm{PV} = \mathrm{Rs.}\ 807{,}766.456$$

From the Present Value table in the Annexure

Table 2 Column 2 and row 8 is the cell value at 12% rate for 8 periods. Factor = 0.404

The factor multiplied by the FCF gives the Present Value:

$$\mathrm{PV} = 2{,}000{,}000 \times 0.404 = 808{,}000$$

Moreover, future returns of business ventures can be converted into present values to compare two possible investments decisions.

AT A GLANCE
SPOTLIGHT
STICKY NOTES

Comparing Cash flows:

Discounting cash flows to their present value is a very important technique. It can be used to compare future cash flows expected at different points in time by discounting them back to their present values.

- *For example:*

A borrower is due to repay a loan of Rs 120,000 in 3 years. He has offered to pay an extra Rs 20,000 as long as he can repay after 5 years. The lender faces interest rates of 7%. Is the offer acceptable?

For existing contract:

FCF = 120,000, n=3, r=0.07

$$\mathrm{PV} = 120{,}000 \times \frac{1}{(1.07)^3} = \mathrm{Rs}\ 97{,}955$$

Client's offer need additional 20,000 which makes FCF 140,000 keeping interest rate and period same present value can be found as

$$\text{Present value} = 140{,}000 \times \frac{1}{(1.07)^5} = \mathrm{Rs}\ 99{,}818$$

The client's offer is acceptable as the present value of the new amount is greater than the present value of the receipt under the existing contract.

An investor wants to make a return on his investments of at least 7% per year. He has been offered the chance to invest in a bond that will cost Rs 200,000 and will pay Rs 270,000 at the end of four years.

In order to earn Rs 270,000 after four years at an interest rate of 7% the amount of his investment now would need to be:

$$\mathrm{PV} = 270{,}000 \times \frac{1}{(1.07)^4} = \mathrm{Rs}\ 205{,}982$$

The investor would be willing to invest Rs 205,982 to earn Rs 270,000 after 4 years.

However, he only needs to invest Rs 200,000.

This indicates that the bond provides a return in excess of 7% per year.

A Bank offers to give Rs. 400,000 after 4 years to any individual who will invest Rs. 300,000 today. The interest rate is 10%. Mr. A is interested in getting the amount invested at the bank for likely profit in future. Is it a wise decision or he can explore other options?

In weighing the prospects, let's find out the Present Value of Rs. 400,000. If it would be less than the amount asked for then the option is not viable for Mr. A.

$$\mathrm{PV} = 400{,}000 \times \frac{1}{(1.10)^4} = \mathrm{Rs}\ 273{,}205.38$$

PV for the possible offer is less than the investment required. This means that Mr. A can invest lesser amount in any other bank for the same return in future. Hence, not a wise decision to take that offer.

STICKY NOTES

Simple interest is where the annual interest is a fixed percentage of the original amount borrowed or invested (the principal).

The interest based on the original principal might be repaid on a periodic basis or accumulated and be repaid at the end. Formula for amount repayable at the end of a loan is: S=P(1+rn) and for calculating interest due on a loan I=Prn

For shorter periods, the annual simple interest rate is pro-rated to find the rate that relates to a shorter period

Compounding is the process of accumulating interest on an investment over time to earn more interest. Annual interest based on the amount borrowed plus interest accrued to date is hence referred to as Compound Interest.

$$S_n = So \times (1 + r)^n$$

Discounting estimates the present day equivalent of an amount at a specified time in the future at a given rate of interest. Discounting is the reverse of compounding.

An amount expected at a specified time in the future is multiplied by a discount factor to give the present value of that amount at a given rate of interest.

$$\text{Present value (PV)} = \text{Future Value (FV)} \times \frac{1}{(1 + r)^n}$$

SELF-TEST

5.1. The formula for simple interest is:

(a) $\dfrac{P \times R \times N}{100}$ (b) $\dfrac{P \times R}{100 \times N}$

(c) $\dfrac{100\ P}{R \times N}$ (d) $\dfrac{100 \times R \times N}{P}$

5.2. Future value of Rs.1,355/- invested @ 8% p.a. simple interest for 5 years is:

(a) Rs.1,897 (b) Rs.1,798

(c) Rs.1,987 (d) Rs.1,789

5.3. The present value of Rs.1,400 at 8 percent simple interest for 5 years is:

(a) Rs.2,000 (b) Rs.900

(c) Rs.3,000 (d) Rs.1,000

5.4. A bank charges mark-up @ Rs.0.39 per day per Rs.1,000/-, rate of mark-up as percent per annum is (assume 365 days in a year):

(a) 14.4% (b) 14.004%

(c) 14.24% (d) 14.0004%

5.5. A person borrowed Rs.20,000 from a bank at a simple interest rate of 12 percent per annum. In how many years will he owe interest of Rs.3,600?

(a) 1.5 years (b) 1.55 years

(c) 1.6 years (d) 1.45 years

5.6. How long will it take for a sum of money to double itself at 10% simple interest?

(a) 7 years (b) 9 years

(c) 5 years (d) 10 years

5.7. The sum required to earn a monthly interest of Rs.1200 at 18% per annum SI is:

(a) Rs.50,000 (b) Rs.60,000

(c) Rs.80,000 (d) None of these

5.8. In what time will Rs.1,800 yield simple interest of Rs.390 at the rate of 5% per annum?

(a) 5 years 2 months (b) 4 years 4 months

(c) 4 years 5 months (d) None of these

5.9. An amount of Rs.20,000 is due in three months. The present value if it includes simple interest @8% is:

(a) Rs.19,608.84 (b) Rs.19,607.84

(c) Rs.18,607.84 (d) Rs.19,507.84

5.10. A sum of money would amount to Rs.6,200 in 2 years and Rs.7,400 in 3 years. The principal and rate of simple interest are:

(a)	Rs.3,800, 31.57%	(b)	Rs.3,000, 20%
(c)	Rs.3,500, 15%	(d)	None of these

5.11. A sum of money would double itself in 10 years. The number of years it would be four times is (assume compound interest):

(a)	25 years	(b)	15 years
(c)	20 years	(d)	None of these

5.12. A total of Rs.14,000 is invested for a year at simple interest, part at 5% and the rest at 6%. If Rs.740 is the total interest, amount invested at 5% is:

(a)	Rs.9,000	(b)	Rs.8,000
(c)	Rs.6,000	(d)	Rs.10,000

5.13. If the simple interest on a certain sum for 15 months at 7½% per annum exceeds the simple interest on the same sum for 8 months at 12½% per annum by Rs.32.50, then the sum (in Rs.) is:

(a)	Rs.3,000	(b)	Rs.3,060
(c)	Rs.3,120	(d)	Rs.3,250

5.14. A certain sum is invested for T years. It amounts to Rs.400 at 10% simple interest per annum. But when invested at 4% simple interest per annum, it amounts to Rs.200. Then time (T) is:

(a)	41 years	(b)	39 years
(c)	50 years	(d)	None of these

5.15. A sum of Rs.7700 is to be divided among three brothers Zain, Zaid and Zoaib in such a way that simple interest on each part at 5% per annum after 1 year for Zain, 2 years for Zaid and 3 years for Zoaib, respectively remains equal. The share of Zain is more than that of Zoaib by:

(a)	Rs.2,800	(b)	Rs.2,500
(c)	Rs.3,000	(d)	None of these

5.16. A person borrowed Rs.500 @ 3% per annum S.I. and Rs.600 @ 4.5% per annum on the agreement that the whole sum will be returned only when the total interest becomes Rs.126. The number of years, after which the borrowed sum is to be returned, is:

(a)	2	(b)	3
(c)	4	(d)	5

5.17. A lends Rs.2,500 to B and a certain sum to C at the same time at 7% p.a. simple interest. If after 4 years, A altogether receives Rs.1,120 as interest from B and C, then the sum lent to C is:

(a)	Rs.700	(b)	Rs.1,500
(c)	Rs.4,000	(d)	Rs.6,500

AT A GLANCE SPOTLIGHT STICKY NOTES

5.18. An investor receives a total of Rs.5,700 per annum in interest from 3 stocks yielding 4%, 5% and 8% per annum respectively. The amount at 4% is Rs.20,000 more than the amount invested at 5%, and the interest from the 8% investment is 8 times the interest from the 5% investment. Amount of money invested in each stock is:

(a)	Rs.10,000, 30,000 & 50,000	(b)	Rs.10,000, 30,000 & 20,000
(c)	Rs.20,000, 30,000 & 50,000	(d)	Rs.10,000, 20,000 & 50,000

5.19. An individual has purchased Rs.275,000 worth of Savings Certificate. The Certificate expires in 25 years and a simple interest rate is computed quarterly at a rate of 3 percent per quarter. Interest cheques are mailed to Certificate holders every 3 months. The interest the individuals can expect to earn every three months is:

(a)	Rs.8,450	(b)	Rs.8,250
(c)	Rs.8,150	(d)	Rs.8,350

5.20. If P = Rs.1,000, i = 5% p.a, n = 4; amount and C.I. is:

(a)	Rs.1,215.50, Rs.215.50	(b)	Rs.1,125, Rs.125
(c)	Rs.2,115, Rs.115	(d)	None of these

5.21. The C.I on Rs.16,000 for 1½ years at 10% p.a. payable half-yearly is:

(a)	Rs.2,222	(b)	Rs.2,522
(c)	Rs.2,500	(d)	None of these

5.22. A person will receive Rs.5,000 six years from now. Present value at a compounded discount rate of 8 percent is:

(a)	Rs.3,150.85	(b)	Rs.3,170.99
(c)	Rs.3,160.99	(d)	Rs.3,180.99

5.23. At what rate of interest compounded semi-annually will Rs.6,000 amount to Rs.9,630 in 8 years?

(a)	5%	(b)	6.1%
(c)	6%	(d)	5.1%

5.24. The effective rate of interest corresponding to a nominal rate 3% p.a. payable half yearly is:

(a)	3.2% p.a	(b)	3.25% p.a
(c)	3.0225% p.a	(d)	None of these

5.25. The population of a town increases every year by 2% of the population at the beginning of that year. The number of years by which the total increase of population be 40% is:

(a)	7 years	(b)	10 years
(c)	17 years (approx)	(d)	None of these

5.26. The effective rate of interest corresponding a nominal rate of 7% p.a. convertible quarterly is:

(a)	7%	(b)	7.5%
(c)	5%	(d)	7.18%

5.27. Osama invested Rs.8,000 for 3 years at 5% C.I in a post office. If the interest is compounded once in a year, what sum will he get after 3 years?

(a)	Rs.9,261	(b)	Rs.8,265
(c)	Rs.9,365	(d)	None of these

5.28. The compound interest on Rs.1,000 at 6% compounded semi-annually for 6 years is:

(a)	Rs.425.76	(b)	Rs.450.76
(c)	Rs.475.76	(d)	Rs.325.76

5.29. A sum of money invested at compound interest amounts to Rs.4,624 in 2 years and to Rs.4,913 in 3 years. The sum of money is:

(a)	Rs.4,096	(b)	Rs.4,260
(c)	Rs.4,335	(d)	Rs.4,360

5.30. The population of a country increases at the rate of 3% per annum. How many years will it take to double itself?

(a)	21.45 years	(b)	22.45 years
(c)	23 years	(d)	23.45 years

5.31. The number of fishes in a lake is expected to increase at a rate of 8% per year. How many fishes will be in the lake in 5 years if 10,000 fishes are placed in the lake today?

(a)	14,693 fishes	(b)	14,683 fishes
(c)	15,693 fishes	(d)	14,583 fishes

5.32. Compute effective rate of interest where nominal rate is 8% compounded quarterly?

(a)	7.24%	(b)	8.0%
(c)	8.42%	(d)	8.24%

5.33. An investor can earn 9.1% interest compounded semi-annually or 9% interest compounded monthly. Determine which option he should prefer?

(a)	Option I	(b)	Option II
(c)	Both (a) and (b)	(d)	None of these

5.34. The population of a city was 8 million on January 1, 2010. The population is growing at the exponential rate of 2 percent per year. What will the population be on January 1, 2015?

(a)	8.74 million	(b)	8.84 million
(c)	8.64 million	(d)	7.84 million

5.35. A trust fund for a child's education is being set up by a single payment so that at the end of 10 years there will be Rs.240,000. If the fund earns at the rate of 8% compounded semi-annually, amount of money should be paid into the fund initially is:

(a)	Rs.108,533	(b)	Rs.109,533
(c)	Rs.109,433	(d)	Rs.100,533

5.36. If Rs.110,000 is to grow to Rs.250,000 in ten years period, at what annual interest rate must it be invested, given that interest is compounded semiannually?

(a)	8.05%	(b)	8.25%
(c)	8.15%	(d)	8.38%

5.37. The nominal interest rate on an investment is 12 percent per year. Determine the effective annual interest rate if interest is compounded quarterly.

(a)	11.55%	(b)	12.05%
(c)	12.55%	(d)	11.55%

5.38. The nominal interest rate on an investment is 16 percent per annum. Determine the effective annual interest rate if interest is compounded quarterly.

(a)	16%	(b)	16.98%
(c)	15%	(d)	15.98%

5.39. Find out the effective rate of interest equivalent to the nominal rate of 10 percent compounded semi-annually.

(a)	10.25%	(b)	10.45%
(c)	10.35%	(d)	10.15%

5.40. A car was moving at a speed of 135 km per hour. When brakes were applied, the speed of the car reduced to 43.2 km per hour in five seconds. Find the rate of decline in the speed of the car per second, if the percentage decrease after each second was the same.

(a)	20.37%	(b)	20.74%
(c)	20.73%	(d)	20.47%

5.41. Find the effective rate of interest equivalent to nominal rate of 8% compounded monthly?

(a)	8.29%	(b)	8.20%
(c)	8.19%	(d)	8.39%

5.42. The difference between C.I and S.I on a certain sum of money invested for 3 years at 6% p.a. is Rs.110.16. The sum is:

(a)	Rs.3,000	(b)	Rs.3,700
(c)	Rs.12,000	(d)	Rs.10,000

5.43. What will be the difference in the compound interest on Rs.50,000 at 12% for one year, when the interest is paid yearly and half-yearly?

(a)	Rs.500	(b)	Rs.600
(c)	Rs.180	(d)	Rs.360

5.44. A sum of money lent at compound interest for 2 years at 20% per annum would fetch Rs.482 more, if the interest was payable half-yearly than if it was payable annually. The sum is:

(a)	Rs.10,000	(b)	Rs.20,000
(c)	Rs.40,000	(d)	Rs.50,000

5.45. The compound interest on a certain sum for 2 years at 10% per annum is Rs.525. The simple interest on the same sum for double the time at half the rate percent per annum is:

(a)	Rs.400	(b)	Rs.500
(c)	Rs.600	(d)	Rs.800

5.46. A shopkeeper sold goods worth Rs.3.0 million during 2008. If he is able to increase his sale by 15% annually, in which year he would achieve annual sale of Rs.25 million?

(a)	Year 2023	(b)	Year 2024
(c)	Year 2021	(d)	Year 2000

5.47. How much should an individual deposit now to yield Rs.600,000 at the end of five years at 9% compounded half yearly?

(a)	Rs.379,965	(b)	Rs.389,864
(c)	Rs.386,357	(d)	Rs.387,964

5.48. A person deposited Rs.100,000 in a bank for three years. The bank paid interest at the rate of 8% per annum compounded half yearly during the first year and at the rate of 12% per annum compounded quarterly during the last two years. His balance after three years is:

(a)	Rs.137,013.85	(b)	Rs.136,013.85
(c)	Rs.147,013.85	(d)	Rs.157,013.85

5.49. Mr. Rashid invested Rs.60,000 in a company but found that his investment was losing 6% of its value per annum. After two years, he decided to pull out what was left of the investment and place at 4% interest compounded twice a year. He would recover his original investment in the_____ year after investing at 4%

(a)	4th year	(b)	3rd year
(c)	2nd year	(d)	5th year

5.50. Rashid wants to obtain a bank loan. Bank A offers a nominal rate of 14% compounded monthly; Bank B a nominal rate of 14.5% compounded quarterly and bank C offers an effective rate of 14.75%. Which option he should prefer, if all other terms are same?

(a)	Bank A	(b)	Bank B
(c)	Bank C	(d)	Cannot be determined

5.51. A firm's labour force is growing at the rate of 2 percent per annum. The firm now employs 500 people. How many employees is it expected to hire during the next five years?

(a)	52 employees	(b)	550 employees
(c)	552 employees	(d)	50 employees

5.52. The population of a country is growing exponentially at a constant rate of 2 percent per year. How much time this population will take to double itself?

(a)	31.65 years	(b)	32.65 years
(c)	30.65 years	(d)	34.66 years

5.53. The capital of a business grows @ 12% per annum compounded quarterly. If present capital is Rs.300,000 the capital after 12 years would be:

(a)	Rs.2,159,019	(b)	Rs.7,656,784
(c)	Rs.1,158,792	(d)	Rs.1,239,676

5.54. If annual interest rate falls from 12 to 8 percent per annum, how much more be deposited in an account to have Rs.600,000 in 5 years, if both rates are compounded semi annually?

(a)	Rs.70,000	(b)	Rs.70,302
(c)	Rs.70,600	(d)	Rs.70,900

5.55. Bank A offers 12.25% interest compounded semi-annually, on its saving accounts, while Bank B offers 12% interest compounded monthly. Which Bank offers the higher effective rate?

(a)	Bank B	(b)	Bank A
(c)	Both rates are same effectively	(d)	Cannot be determined

5.56. Mr. X borrowed Rs.5,120 at 12½% p.a. C.I. At the end of 3 years, the money was repaid along with the interest accrued. The amount of interest paid by him is:

(a)	Rs.2,100	(b)	Rs.2,170
(c)	Rs.2,000	(d)	None of these

5.57. Dawood has to repay a loan along with interest, three years from now. The amount payable after three years is Rs. 428,000. The amount of loan presently if interest rate is 8% compounded Semi-annually is:

(a)	Rs. 345,161	(b)	Rs. 339,760
(c)	Rs. 338,254	(d)	Rs. 336,421

5.58. Dawood has to repay a loan along with interest, three years from now. The amount payable after three years is Rs. 428,000. The amount of loan presently if interest rate is 8% compounded Quarterly is:

(a)	Rs. 329,760	(b)	Rs.337,475
(c)	Rs. 341,475	(d)	Rs.335,475

5.59. The banker's interest to the nearest paisa's. Principal: Rs.2500; Rate: 9%; Time 180 days: Interest is:

(a)	Rs.104.32	(b)	Rs.109.75
(c)	Rs.112.50	(d)	Rs.110.96

5.60. The maturity value of a loan of Rs.2,800 after three years. The loan carries a simple interest rate of 7.5% per year is:

(a)	Rs.3,429	(b)	Rs.3,430
(c)	Rs.3,431	(d)	Rs.3,440

5.61. To increase present value, the discount rate should be adjusted:

(a)	Upward	(b)	Downward
(c)	Not relevant	(d)	Will depend on time period

5.62. A borrower has agreed to pay Rs.10,000 in 9 months at 10% simple interest. How much did this borrower receive?

(a)	Rs.9,090	(b)	Rs.9,250
(c)	Rs.9,500	(d)	Rs.9,302

5.63. Bashir owes Rs.50,000 to Arshad due to a court decision. The money must be paid in 10 months with no interest. Suppose Bashir wishes to pay the money now. What amount should Arshad be willing to accept? Assume simple interest of 8% per annum.

(a)	Rs.45,875	(b)	Rs.47,875
(c)	Rs.46,875	(d)	Rs.46,575

5.64. Mr. Junaid received Rs.48,750 in cash as the proceeds of a 90 day loan from a bank which charges 10% simple interest. The amount he will have to pay on the maturity date is (assume 360days in a year):

(a)	Rs.49,969	(b)	Rs.50,000
(c)	Rs.47,548	(d)	Rs.53,625

5.65. Present Value of a future cash flow is the amount that

(a) Investor invests in future to get the amount today
(b) Investor invests today to get the amount in future
(c) Is not equivalent to the value of future cash flow in present day
(d) Is not equivalent to the value of present cash flow same time in the future

5.66. How would the future value of an amount invested today change in 8 years if the amount is invested at the rate of 9% compound interest?

(a)	Double	(b)	Half
(c)	Quadruple	(d)	Will remain same

5.67. Laura had invested Rs. 250,000 in a land ten years ago. Now, she wants to evaluate the worth of the land she bought. What will be the value of land today if it is compounded at 6% per annum

(a)	Rs. 265,000	(b)	Rs. 235,850
(c)	Rs. 447,712	(d)	Rs. 416,667

5.68. A three-year investment yields a return of 25% over the period. Compute the effective annual rate of return.

(a)	8.33%	(b)	7.72%
(c)	6%	(d)	5%

5.69. The annual cost of a company is Rs 500,000 at the moment. It is expected that due to re-engineering the cost will reduce by 20% per annum for next two years. Compute the annual cost for fourth year from now.

(a)	Rs. 320,000	(b)	Rs. 163,840
(c)	Rs. 200,000	(d)	Rs. 260,000

5.70. Value of a particular asset decreases by 10% every year of its cost. If the cost is Rs 700,000 today, compute its value two years from today.

(a)	Rs. 630,000	(b)	Rs. 560,000
(c)	Rs. 567,000	(d)	Rs. 500,000

5.71. A sum of Rs. 100,000 is invested at 10% per annum compounded quarterly. Compute the amount to which it will grow after 3 years.

(a)	134,489	(b)	143,984
(c)	413,489	(d)	489,134

5.72. The present value of Rs. 5,000 at 16% per annum for 4 years is:

(a)	Rs. 1,800	(b)	Rs. 3,761.46
(c)	Rs. 4,600	(d)	Rs. 2,761.46

5.73. The discount factor for 3 years at 10% is:

(a)	0.751	(b)	0.683
(c)	0.826	(d)	0.909

5.74. Taha invested Rs. 200,000 for 5 years at simple interest at 10%. At the end of fifth year he withdrew entire amount and invested at 10% per annum compound interest for 2 years. Compute the amount that he will have with him at the end of 7th year from now.

(a)	Rs. 463,000	(b)	Rs. 363,000
(c)	Rs. 563,000	(d)	Rs. 663,000

5.75. Marfani lends Rs. 500,000 to Taha at 12% per annum compounded annually for 3 years and Rs. 300,000 to Jawwad at 10% per annum compounded quarterly for a period of 3 years. Compute the effective annual rate of interest for amount lent to Jawwad

(a)	10%	(b)	12.38%
(c)	10.38%	(d)	11.38%

5.76. Luqman invests Rs 450,000 at 10% simple interest for 4 years and Rs 900,000 at 10% compound interest for the same time period. Compute the difference between the receipts of two investments at the end of 4th year.

(a)	Rs. 417,690	(b)	Rs. 687,690
(c)	Rs. 450,000	(d)	Rs 867,690

5.77. A sum of money invested at compound interest amounts to Rs. 5,280 in 2 years and to Rs. 5,808 in 3 years. The sum of money is:

(a)	Rs. 4,363.64	(b)	Rs. 4,080
(c)	Rs. 4,750	(d)	Rs. 4,500

5.78. Find the future value of Rs. 7,000 invested at 10% per annum compounded quarterly for three years and further invested at 12% per annum compounded semi-annually for two years.

(a)	Rs. 11,858.24	(b)	Rs. 11,885.24
(c)	Rs. 11,588.24	(d)	Rs. 11,558.24

5.79. Which is more in present value terms if interest rate is 16% per annum compounded annually?

i. Rs 5,000 one year from now.
ii. Rs 10,000 in three years from now.
iii. Rs, 15,000 in two years from now.
iv. Rs 20,000 in four years from now.

(a)	i	(b)	ii
(c)	iii	(d)	iv

5.80. Which is most in future value terms if interest rate is 16% per annum?

i. Rs 5,000 invested for one year.
ii. Rs 10,000 invested for three years.
iii. Rs 15,000 invested for two years.
iv. Rs 20,000 invested for four years.

(a) i (b) ii
(c) iii (d) iv

5.81. Mr. Khalid invested Rs. 200,000 at 12% simple interest and Rs. 189,744 at 14% simple interest. The two amounts were left to grow till such time that the total of interest and principal of both investments become equal.

(a) 4 (b) 3
(c) 2 (d) 5

5.82. Mr. Asim has following investment options:

i. Invest in Bank A at 12% per annum compounded quarterly
ii. Invest in Bank B at 10% per annum compounded semi-annually
iii. Invest in Bank C at 8% per annum compounded monthly
iv. Invest in Bank D at 15% per annum.

Which of the above options must he opt

(a) Bank A (b) Bank B
(c) Bank C (d) Bank D

5.83. Compute the effective annual rate of following options:

i. A: Investment of Rs. 200,000 resulting in a single 25% return on investment after 4 years.
ii. B: Investment of Rs. 300,000 resulting in a 15% per annum compounded semi –annual return.

(a) A: 5.74% B: 15.56% (b) A: 5.54% B: 15.56%
(c) A: 5.64% B: 14.56% (d) A: 5.64% B: 15%

5.84. Bakar has to pay Zeeshan Rs. 900,000 three years from today. If the market interest rate is 10% per annum compounded annually how much will Zeeshan be willing to accept today to settle the debt

(a) Rs. 676,183 (b) Rs. 576,183
(c) Rs 767,183 (d) Rs. 876,183

5.85. Sheeraz bought a bike on cash basis for Rs. 170,000. The seller is willing to offer it on following terms as well:

i. Pay Rs. 200,000 at the end of 1st year.
ii. Pay Rs. 300,000 at the end of 4th year.
iii. Pay Rs. 430,000 at the end of 5th year.

If the market interest rate is 17% per annum compounded annually. Shall Sheeraz accept any of the offers given by the dealer, if yes which offer is most beneficial?

(a) No (b) Yes, Option # i
(c) Yes, Option # ii (d) Yes, Option # iii

5.86. Asim bought a machine for Rs 500,000. The value of the machine decreases by 15% per annum of its net book value. The life of the machine is three years. At the end of its life Asim will sell it and the expected receipts from disposal will be 12% above its net book value. Asim will also need a replacement machine which will cost Rs 600,000 in future value terms. How much more will he need at the end of three years to buy the machine, if the proceeds from initial machine will also be used to pay for the new machine?

(a) Rs. 331,525 (b) Rs. 600,000
(c) Rs. 213,525 (d) Rs. 256,090

5.87. Rs. X are invested for 7 years at Y% per annum compound interest. Compute future value in terms of x and y.

(a)	X(1+Y/100)7	(b)	X(1+100/Y)7
(c)	Y(1+X/100)7	(d)	Not possible to determine

ANSWERS TO SELF-TEST QUESTIONS

5.1	**5.2**	**5.3**	**5.4**	**5.5**	**5.6**
(a)	(a)	(d)	(c)	(a)	(d)
5.7	**5.8**	**5.9**	**5.10**	**5.11**	**5.12**
(c)	(b)	(b)	(a)	(c)	(d)
5.13	**5.14**	**5.15**	**5.16**	**5.17**	**5.18**
(c)	(c)	(a)	(b)	(b)	(a)
5.19	**5.20**	**5.21**	**5.22**	**5.23**	**5.24**
(b)	(a)	(b)	(a)	(c)	(c)
5.25	**5.26**	**5.27**	**5.28**	**5.29**	**5.30**
(c)	(d)	(a)	(a)	(a)	(d)
5.31	**5.32**	**5.33**	**5.34**	**5.35**	**5.36**
(a)	(d)	(b)	(b)	(b)	(d)
5.37	**5.38**	**5.39**	**5.40**	**5.41**	**5.42**
(c)	(b)	(a)	(a)	(a)	(d)
5.43	**5.44**	**5.45**	**5.46**	**5.47**	**5.48**
(c)	(b)	(b)	(b)	(c)	(a)
5.49	**5.50**	**5.51**	**5.52**	**5.53**	**5.54**
(a)	(c)	(a)	(d)	(d)	(b)
5.55	**5.56**	**5.57**	**5.58**	**5.59**	**5.60**
(a)	(b)	(c)	(b)	(d)	(b)
5.61	**5.62**	**5.63**	**5.64**	**5.65**	**5.66**
(b)	(d)	(c)	(a)	(b)	(a)
5.67	**5.68**	**5.69**	**5.70**	**5.71**	**5.72**
(c)	(b)	(a)	(c)	(a)	(d)
5.73	**5.74**	**5.75**	**5.76**	**5.77**	**5.78**
(a)	(b)	(c)	(b)	(a)	(b)
5.79	**5.80**	**5.81**	**5.82**	**5.83**	**5.84**
(c)	(d)	(a)	(d)	(a)	(a)
5.85	**5.86**	**5.87**			
(c)	(d)	(a)			

CHAPTER 6

DISCOUNTED CASH FLOWS

IN THIS CHAPTER

AT A GLANCE

Discounted cash flow is a technique for evaluating proposed investments, to decide whether they are financially worthwhile. There are two methods of DCF:

- Net present value (NPV) method: the cost of capital r is the return required by the investor or company
- Internal rate of return (IRR) method: the cost of capital r is the actual return expected from the investment.

All cash flows are assumed to arise at a distinct point in time (usually the end of a year). Net present value (NPV) is one of the method to analyze the value of discounted cash flow for proposed investments and to decide whether they are financially worthwhile.

An annuity is a series of regular periodic payments of equal amount whereas a perpetuity is a constant annual cash flow 'forever'. Analysis of cashflows helps in investment decisions.

1 FUTURE & PRESENT VALUE WITH MULTIPLE CASH FLOW

Future and present value for multiple cash flow situations occur when deposits or receipt for certain amount is multiple times in a given period. This situation is a little different from the earlier discussed situations where lump sum amount at the end of the period was received or deposited.

- *For example:*

A person would receive Rs. 400 in the year 1, Rs. 600 in the year 2 and Rs. 800 in the year 3 and 4 for four consecutive years. If he can earn 9% interest on these amounts. How much he would need to save today to receive these amounts in future?

Present value of amounts would be

Year 1: $400 \div 1.09^{1} = 366.972$

Year 2: $600 \div 1.09^{2} = 505.008$

Year 3: $800 \div 1.09^{3} = 617.746$

Year 4: $800 \div 1.09^{4} = 566.740$

Sum that up

$366.972 + 505.008 + 617.746 + 566.740 = 2056.466$

Or

In understanding problems for cash flows, timings of the cash flow is important. Cash flows are mostly assumed to be at the end of the period. However, there may be cash flows due or expected at the beginning of the year or a certain point during the identified period. In all such cases, basic principle for present and future value of cash flow remain same, however, need inclusion of periodic changes in mind.

2 NET PRESENT VALUE OF AN INVESTMENT

In determining the investment's worth, Net Present Value is used. The difference between an investment's market value and its cost is called the Net Present Value (NPV) of the investment. Investments with positive net present value means that more value is created or added for the investors.

Calculating the NPV of an investment project

Step 1: List all cash flows expected to arise from the project. This will include the initial investment, future cash inflows and future cash outflows.

Step 2: Discount these cash flows to their present values using the cost that the company has to pay for its capital (cost of capital) as a discount rate. All cash flows are now expressed in terms of 'today's value'.

Step 3: The NPV of a project is difference between the present value of all the costs incurred and the present value of all the cash flow benefits (savings or revenues).

- The project is acceptable if the NPV is positive.
- The project should be rejected if the NPV is negative.

► *For example:*

A company with a cost of capital of 10% is considering investing in a project. Initial payment required is Rs. 10,000 while it will yield Rs. 6,000 per year for the initial two years. Should the project be undertaken?

One way to understand the problem is to analyze cash flows with their discount factors as below:

Year	Cash flow	Discount factor (10%)	Present value
0	(10,000)	1	(10,000)
1	6,000	$\frac{1}{(1.1)}$	5,456
2	6,000	$\frac{1}{(1.1)^2}$	4.959
NPV			415

The project is viable since the NPV is positive.

The same problem can be solved using PV formula:

$$\text{Present value} = \text{X} \times \left[\frac{1-\frac{1}{(1+r)^n}}{\text{r}}\right]$$

$$\text{Present value} = 6{,}000 \times \left[\frac{1-\frac{1}{(1+0.1)^2}}{0.1}\right]$$

$$\text{Present value} = 6{,}000 \times \left[\frac{0.17355}{0.1}\right]$$

$$\text{Present value} = 6{,}000 \times 1.7355$$

$$\text{Present value} = 10{,}413.22$$

PV is more than the initial outlay. This means that project's worth for investment is more than the cost. And hence must be accepted.

A company is considering whether to invest in a new item of equipment costing Rs.53,000 to make a new product. The product would have a four-year life. Calculate the NPV assuming the discount rate of 11% and the estimated cash profits over the four-year period as follows.

Year	Rs.
1	17,000
2	25,000
3	16,000
4	12,000

In solving for the above problem, each year's cash flow would be discounted at the given rate to calculate the PV.

Year	Cash flow	Discount factor (11%)	Present value
0	(53,000)	1	(53,000)
1	17,000	$\frac{1}{(1.11)}$	15,315
2	25,000	$\frac{1}{(1.11)^2}$	20,291
3	16,000	$\frac{1}{(1.11)^3}$	11,699
4	12,000	$\frac{1}{(1.11)^4}$	7,905
NPV			2,210

The NPV is positive so the project should be accepted

A company is considering whether to invest in a new item of equipment costing Rs. 65,000 to make a new product. The product would have a three-year life. Calculate the NPV of the project using a discount rate of 8% if the cash flows are as follows:

Year	Rs.
1	27,000
2	31,000
3	15,000

In solving for the above problem, each year's cash flow would be discounted at the given rate to calculate the PV.

Year	Cash flow	Discount factor (8%)	Present value
0	(65,000)	1	(65,000)
1	27,000	$\frac{1}{(1.08)}$	25,000
2	31,000	$\frac{1}{(1.08)^2}$	26,578
3	15,000	$\frac{1}{(1.08)^3}$	11,907
NPV			(1,515)

The NPV is negative so the project should be rejected.

A company is considering whether to invest in a project which would involve the purchase of machinery with a life of five years. The machine would cost Rs. 556,000 and would have a net disposal value of Rs. 56,000 at the end of Year 5. The project would earn annual cash flows (receipts minus payments) of Rs. 200,000. Calculate the NPV of the project using a discount rate of 15%

One way to solve for the above problem is to discount cash flow each year for the calculation of the PV.

Year	Cash flow	Discount factor (15%)	Present value
0	(556,000)	1	(556,000)
1	200,000	$\frac{1}{(1.15)^1}$	173,913
2	200,000	$\frac{1}{(1.15)^2}$	151,229
3	200,000	$\frac{1}{(1.15)^3}$	131,503
4	200,000	$\frac{1}{(1.15)^4}$	114,351
5	200,000	$\frac{1}{(1.15)^5}$	99,435
5	56,000	$\frac{1}{(1.15)^5}$	27,842
NPV			142,273

The same problem can be solving using PV of an annuity as well as PV of a lump sum amount formula

$$\text{Present value} = X \times \left[\frac{1-\frac{1}{(1+r)^n}}{\text{r}}\right] + X\left[\frac{1}{(1+r)^n}\right]$$

$$\text{Present value} = 200{,}000 \times \left[\frac{1-\frac{1}{(1+0.15)^5}}{0.15}\right] + 56{,}000\left[\frac{1}{(1+0.15)^5}\right]$$

$$\text{Present value} = 200{,}000 \times \left[\frac{0.50282}{0.15}\right] + 56{,}000(0.4972)$$

$$\text{Present value} = 670{,}431 + 27{,}842$$

$$\text{Present value} = 698{,}273$$

PV is more than the initial outlay. This means that project's worth for investment is more than the cost. And hence must be accepted.

AT A GLANCE SPOTLIGHT STICKY NOTES

A company is considering whether to invest in a project which would involve the purchase of machinery with a life of four years. The machine would cost Rs.1,616,000 and would have a net disposal value of Rs.301,000 at the end of Year 4. The project would earn annual cash flows (receipts minus payments) of Rs.500,000. Calculate the NPV of the project using a discount rate of 10%

One way to solve for the above problem is to discount cash flow each year for the calculation of the PV.

Year	Cash flow	Discount factor (15%)	Present value
0	(1,616,000)	1	(1,616,000)
1	500,000	$\frac{1}{(1.10)^1}$	454,546
2	500,000	$\frac{1}{(1.10)^2}$	413,223
3	500,000	$\frac{1}{(1.10)^3}$	375,657
4	500,000	$\frac{1}{(1.10)^4}$	341,507
4	301,000	$\frac{1}{(1.10)^4}$	205,587
NPV			174,520

The same problem can be solving using PV of an annuity as well as PV of a lump sum amount formula

$$\text{Present value} = \text{X} \times \left[\frac{1-\frac{1}{(1+r)^n}}{\text{r}}\right] + X\left[\frac{1}{(1+r)^n}\right]$$

$$\text{Present value} = 500{,}000 \times \left[\frac{1-\frac{1}{(1+0.10)^4}}{0.10}\right] + 301{,}000\left[\frac{1}{(1+0.10)^4}\right]$$

$$\text{Present value} = 500{,}000 \times \left[\frac{0.316987}{0.10}\right] + 301{,}000(0.6830135)$$

$$\text{Present value} = 1{,}584{,}933 + 205{,}587$$

$$\text{Present value} = 1{,}790{,}520$$

PV is more than the initial outlay. This means that project's worth for investment is more than the cost. And hence, must be accepted.

AT A GLANCE

SPOTLIGHT

STICKY NOTES

3 ANNUITIES AND PERPETUITIES

Annuity

An annuity is a series of regular periodic payments of equal amount. It is a constant cash flow for a given number of time periods. A capital project might include estimated annual cash flows that are an annuity.

▶ *For example:*

Rs.30,000 each year for years 1 – 5

Rs.20,000 each year for years 3 – 10

Rs.500 each month for months 1 – 24.

The present value / future value of an annuity can be computed by multiplying / dividing each individual amount by the individual discount factor and then adding each product. This is fine for annuities of just a few periods but would be too time consuming for long periods. An alternative approach is to use the annuity factor.

An annuity factor for a number of periods is the sum of the individual discount factors for those periods.

An annuity factor can be constructed by calculating the individual period factors and adding them up but this would not save any time. In practice a formula or annuity factor tables are used.

There are two types of annuities:

- Ordinary annuity – payments (receipts) are in arrears i.e. at the end of each payment period
- Annuity due – payments (receipts) are in advance i.e. at the beginning of each payment period.

Future value of Annuity:

Amount of an annuity or future value of an annuity is the total of all the installments together with the compound interest of each payment for the period.

▶ *Formula:*

Ordinary annuity $$Sn = X\left[\frac{(1+r)^n - 1}{r}\right]$$

Annuity due $$Sn = X(1+r)\left[\frac{(1+r)^n - 1}{r}\right]$$

Where:

Sn = final cash flow at the end of the loan (the amount paid by a borrower or received by an investor or lender).

X = Annual investment

r = period interest rate

n = number of periods

AT A GLANCE
SPOTLIGHT
STICKY NOTES

- *For example:*

A savings scheme involves investing Rs.100,000 per annum for 4 years (on the last day of the year). If the interest rate is 10% what is the sum to be received at the end of the 4 years?

$$\mathrm{Sn} = \mathrm{X}\left[\frac{(1+\mathrm{r})^{\mathrm{n}}-1}{\mathrm{r}}\right]$$

Here X = 100,000; n=4; i=0.1

$$\mathrm{Sn} = 100{,}000\left[\frac{(1.1)^{4}-1}{0.1}\right]$$

$$\mathrm{Sn} = 100{,}000\left[\frac{1.4641-1}{0.1}\right]$$

$$\mathrm{Sn} = \frac{46{,}410}{0.1}$$

$$\mathrm{Sn} = \mathrm{Rs.}\,464{,}100$$

A savings scheme involves investing Rs.100,000 per annum for 4 years (on the first day of the year). If the interest rate is 10% what is the sum to be received at the end of the 4 years?

$$\mathrm{Sn} = \mathrm{X}\left[\frac{(1+\mathrm{r})^{\mathrm{n}}-1}{\mathrm{r}}\right]\times(1+r)$$

Here X = 100,000; n=4; i=0.1

$$\mathrm{Sn} = 100{,}000\left[\frac{(1.1)^{4}-1}{0.1}\right]\times 1.1$$

$$\mathrm{Sn} = 100{,}000\left[\frac{(1.4641)-1}{0.1}\right]\times 1.1$$

$$\mathrm{Sn} = \frac{46{,}410}{0.1}\times 1.1$$

$$\mathrm{Sn} = \mathrm{Rs.}\,510{,}510$$

A man wants saving to meet the expense of his son going to university. He intends to puts Rs. 50,000 into a savings account at the end of each of the next 10 years. The account pays interest of 7%. What will be the balance on the account at the end of the 10-year period?

$$\mathrm{Sn} = \mathrm{X}\left[\frac{(1+\mathrm{r})^{\mathrm{n}}-1}{\mathrm{r}}\right]$$

Here X = 50,000; n=10; i=0.07

$$Sn = 50{,}000\left[\frac{(1+0.07)^{10}-1}{0.07}\right]$$

$$Sn = 50{,}000\left[\frac{(1.96715)-1}{0.07}\right]$$

$$Sn = \frac{48{,}357.567}{0.07}$$

$$Sn = \text{Rs.}\,690{,}822.4$$

A proud grandparent wishes to save money to help pay for his grand daughter's wedding. He intends to save Rs.5,000 every 6 months for 18 years starting immediately. The account pays interest of 6% compounding semi-annually (every 6 months) What will be the balance on the account at the end of 18 years.

$$Sn = X(1+r)\left[\frac{(1+r)^n-1}{r}\right]$$

Here X = 5,000; n=2× 18; i=0.06/2

$$Sn = 5{,}000(1.03)\left[\frac{(1.03)^{36}-1}{0.03}\right]$$

$$Sn = 5{,}000(1.03)\left[\frac{(2.89827-1)}{0.03}\right]$$

$$Sn = \frac{9{,}491.391}{0.03}\times 1.03$$

$$Sn = \text{Rs.}\,325{,}871.113$$

A man invests Rs. 250 at the end of each 6-month period in a bank at the rate of 5% per annum. What will be the accumulated amount if he has deposited 8 such installments.

$$Sn = X\left[\frac{(1+r)^n-1}{r}\right]$$

Here X=250; n=8; i=0.05/2

$$Sn = 250\left[\frac{(1.025)^8-1}{0.025}\right]$$

$$Sn = 250\left[\frac{(0.2184)}{0.025}\right]$$

$$Sn = \frac{54.6}{0.025}$$

$$Sn = \text{Rs.}\,2{,}184$$

AT A GLANCE SPOTLIGHT STICKY NOTES

Present value of Annuity:

If we are to receive equal installments for the certain years to come at a given rate, present value of the annuity signifies how much it is worth or offered today. Present value of an annuity also reflects how much amount can be borrowed if exact amount for installments is known.

- *Formula:*

$$\text{Ordinary annuity} \qquad \text{Present value} = X \times \left[\frac{1 - \frac{1}{(1+r)^n}}{r}\right]$$

Where:

Present Value (PV_A) = Initial cash flow at the beginning of the loan (the amount to be received by a borrower or deposited by an investor or offered by a lender).

X = Annual investment / periodic investment / installments

r = period interest rate

n = number of periods

The present value of an annuity can be computed by multiplying each individual amount by the individual discount factor and then adding each product. This is fine for annuities of just a few periods but would be too time consuming for long periods. An alternative approach is to use the annuity factor.

An annuity factor for a number of periods is the sum of the individual discount factors for those periods.

- *For example:*

Calculate the present value of Rs. 50,000 per year for years 1 – 3 at a discount rate of 9%.

$$\text{Present value} = X \times \left[\frac{1 - \frac{1}{(1+r)^n}}{r}\right]$$

$$\text{Present value} = 50{,}000 \times \left[\frac{1 - \frac{1}{(1+0.09)^3}}{0.09}\right]$$

$$\text{Present value} = 50{,}000 \times \left[\frac{1 - 0.7722}{0.09}\right]$$

$$\text{Present value} = 50{,}000 \times 2.5313$$

$$\text{Present value} = 126{,}564.73$$

Mr. G wants to buy a car for which he can pay Rs. 7,500 per year for 4 years. A bank offers 12% rate of interest. How much Mr. G can borrow?

$$\text{Present value} = \text{X} \times \left[\frac{1 - \frac{1}{(1+r)^n}}{\text{r}}\right]$$

X = 7500, r = 0.12, n=4

$$\text{Present value} = 7{,}500 \times \left[\frac{1 - \frac{1}{(1+0.12)^4}}{0.12}\right]$$

$$\text{Present value} = 7{,}500 \times \left[\frac{1 - 0.63552}{0.12}\right]$$

$$\text{Present value} = 7{,}500 \times 3.0373$$

$$\text{Present value} = 22{,}780.12$$

A Bank is offering a loan of Rs. 30,000 to be repaid in 15 equal annual instalments. If the compound interest is charged at 8% per annum, what would be the amount of installments?

$$\text{Present value} = \text{X} \times \left[\frac{1 - \frac{1}{(1+r)^n}}{\text{r}}\right]$$

PV = 30,000, r = 0.08, n=15

$$30{,}000 = \text{X} \times \left[\frac{1 - \frac{1}{(1+0.08)^{15}}}{0.08}\right]$$

$$30{,}000 = \text{X} \times \left[\frac{1 - 0.31524}{0.08}\right]$$

$$30{,}000 = \text{X} \times 8.5595$$

$$\text{X} = 3504.88 \sim \text{Rs. } 3{,}505 \text{ per installment.}$$

AT A GLANCE

SPOTLIGHT

STICKY NOTES

A bank has scheduled a payment of Rs. 3000 per year for a deposit of Rs. 18000. If the interest charge is 9%, how long will it take to repay the loan?

$$\text{Present value} = X \times \left[\frac{1 - \frac{1}{(1+r)^n}}{r}\right]$$

PV = 18,000, r = 0.09, X = 3000

$$18{,}000 = 3{,}000 \times \left[\frac{1 - \frac{1}{(1+0.09)^n}}{0.09}\right]$$

$$\frac{18{,}000}{3{,}000} = \left[\frac{1 - \frac{1}{(1.09)^n}}{0.09}\right]$$

$$6 \times 0.09 = \left[1 - \frac{1}{(1.09)^n}\right]$$

$$0.54 = \left[1 - \frac{1}{(1.09)^n}\right]$$

Or

$$1 - 0.54 = \left[\frac{1}{(1.09)^n}\right]$$

$$0.46 = \frac{1}{(1.09)^n}$$

We use the annuity table and find n

Where annuity factor = p/r = 18,000 / 3,000 = 6 and therefore,

i = 9%.

Annuity Tables

An annuity factor can be constructed by calculating the individual period factors and adding them up but this would not save any time. This factor is multiplied by the installment amount to calculate the annuity. An annuity factor for a number of periods is the sum of the individual discount factors for those periods. In practice a formula or annuity factor tables are used

- *Illustration:*

(Full tables are given as an appendix to this text).

Discount rates (r)						
(n)	**5%**	**6%**	**7%**	**8%**	**9%**	**10%**
1	0.952	0.943	0.935	0.926	0.917	0.909
2	1.859	1.833	1.808	1.783	1.759	1.736
3	2.723	2.673	2.624	2.577	2.531	2.487
4	3.546	3.465	3.387	3.312	3.240	3.170
5	4.329	4.212	4.100	3.993	**3.890**	3.791
Where:						
n = number of periods						

- *For example:*

A company is considering an investment of Rs.70,000 in a project. The project life would be five years. What must be the minimum annual cash returns from the project to earn a return of at least 9% per annum?

Investment = Rs.70,000

Annuity factor at 9%, years 1 – 5 = 3.890

Minimum annuity required = Rs.17,995 (= Rs.70,000/3.890)

A company borrows Rs 10,000,000. This to be repaid by 5 equal annual payments at an interest rate of 8%. Calculate the payments.

The approach is to simply divide the amount borrowed by the annuity factor that relates to the payment term and interest rate.

		Rs
Amount borrowed		10,000,000
Divide by the 5 year, 8% annuity factor		3.993
Annual repayment		2,504,383

Sinking funds

A business may wish to set aside a fixed sum of money at regular intervals to achieve a specific sum at some future point in time. This is known as a sinking fund.

To build to a required amount at a given interest rate over a given period of years, fixed annual amount necessary can be calculated. the calculations use the same approach as above but this time solving for ***X*** as ***Sn*** is known.

- *For example:*

A company will have to pay Rs.5,000,000 to replace a machine in 5 years. The company wishes to save up to fund the new machine by making a series of equal payments into an account which pays interest of 8%. The payments are to be made at the end of the year and then at each year end thereafter. What fixed annual amount must be set aside so that the company saves Rs.5,000,000?

$$\mathrm{Sn} = \mathrm{X}\left[\frac{(1+\mathrm{i})^{\mathrm{n}} - 1}{i}\right]$$

Here Sn = 5,000,000; i = 0.08; n= 5

$$5{,}000{,}000 = \mathrm{X}\left[\frac{(1.08)^{5} - 1}{0.08}\right]$$

$$5{,}000{,}000 = \mathrm{X}\left[\frac{1.46932 - 1}{0.08}\right]$$

$$5{,}000{,}000 = \mathrm{X}\left[\frac{0.46932}{0.08}\right]$$

$$X = \frac{5{,}000{,}000 \times 0.08}{0.46932} = \mathrm{Rs.}\,852{,}296.94$$

A business wishes to start a sinking fund to meet a future debt repayment of Rs. 100,000,000 due in 10 years. What fixed amount must be invested every 6 months if the annual interest rate is 10% compounding semi-annually if the first payment is to be made in 6 months?

$$\mathrm{Sn} = \mathrm{X}\left[\frac{(1+\mathrm{i})^{\mathrm{n}} - 1}{\mathrm{i}}\right]$$

Here Sn = 100,000,000; i = 0.1/2; n= 10×2

$$100{,}000{,}000 = \mathrm{X}\left[\frac{(1+0.1/2)^{10\times 2} - 1}{0.05}\right]$$

$$100{,}000{,}000 = \mathrm{X}\left[\frac{(1.05^{20} - 1)}{0.05}\right]$$

$$100{,}000{,}000 = \mathrm{X}\left[\frac{(1.65329)}{0.05}\right]$$

$$X = \frac{100{,}000{,}000 \times 0.05}{1.65329} = \mathrm{Rs.}\,3{,}024{,}272.814$$

A man wishes to invest equal annual amount in a fund so that he accumulates 5,000,000 by the end of 10 years. The annual interest rate available for investment is 6%. What equal annual amounts should he set aside?

$$\mathrm{Sn} = \mathrm{X}\left[\frac{(1+\mathrm{i})^{\mathrm{n}}-1}{\mathrm{i}}\right]$$

Here Sn = 5,000,000; i = 0.06; n= 10

$$5{,}000{,}000 = \mathrm{X}\left[\frac{(1.06)^{10}-1}{0.06}\right]$$

$$5{,}000{,}000 = \mathrm{X}\left[\frac{(1.79085-1)}{0.06}\right]$$

$$5{,}000{,}000 = \mathrm{X}\left[\frac{(0.79085)}{0.06}\right]$$

$$X = \frac{5{,}000{,}000 \times 0.06}{0.79085} = \mathrm{Rs.}\,379{,}338.7$$

The same problem can be solved using the present value:

Step 1: Calculate the present value of the amount required in 10 years.

$$\mathrm{PV} = 5{,}000{,}000 \times \frac{1}{(1.06)^{10}} = 2{,}791{,}974$$

Step 2: Calculate the equivalent annual cash flows that result in this present value

	Rs
Present value	2,791,974
Divide by the 10 year, 6% annuity factor	7.36
Annual repayment	379,344

If the man invests 379,344 for 10 years at 6% it will accumulate to 5,000,000.

Perpetuities

A perpetuity is a constant annual cash flow 'forever', or into the long-term future. It is an annuity in which the cash flows continue forever[1].

In investment appraisal, an annuity might be assumed when a constant annual cash flow is expected for a long time into the future.

- *Formula:*

$$\text{Perpetuity factor} = \frac{1}{\mathrm{r}}$$

Where:

r = the cost of capital

[1] Ross, Westerfield & Bradford (2008). Fundamentals Of Corporate Finance. 8th Edition. Mcgraw-Hill/Irwin. Pg. 162

and

$$\text{Present Value of Perpetuity} = \frac{X}{r}$$

Where:

X = annual cash flow or periodic installments

r = the cost of capital

► *For example:*

An investment offers Rs. 2,000 per year in perpetuity. When cost of capital is 8% what would be the value of investment?

$$\text{Present Value} = \frac{X}{r}$$

$$= \frac{2{,}000}{0.08} = 25{,}000$$

Now suppose the investment ask for 28,000 value for the same investment. At what interest rate this would be the better deal?

$$\text{Present Value} = \frac{X}{r}$$

$$28{,}000 = \frac{2{,}000}{r}$$

$$r = \frac{2{,}000}{28{,}000}$$

$$r = 7.15\%$$

Find out the present value of an investment that promises to pay Rs. 3,000 every alternate year forever. Discount rate is 5.5% compounded daily.

This is the problem of perpetuity but the installment occurs every two years. Therefore, effective rate would be calculated to be used for the present value of the perpetuity.

$$\text{Period rate} = (1 + r)^n - 1$$

r=0.055/365, n = 365 × 2

$$\text{Period rate} = (1 + 0.0001507)^{365\times 2} - 1$$

$$\text{Period rate} = (1.0001507)^{730} - 1$$

$$\text{Period rate} = 0.116 \sim 11.6\%$$

Now this two-year periodic interest rate can be used for the calculation of the present value of perpetuity;

$$\text{Present Value} = \frac{X}{r}$$

$$PV = \frac{3{,}000}{0.116}$$

$$PV = 25{,}802.27$$

4 INTERNAL RATE OF RETURN (IRR)

The internal rate of return method (IRR method) is another method of investment evaluation with discounted cash flow. It is an alternate method to net present value but is closely related to it[2].

The internal rate of return of a project is the discounted rate of return on the investment.

- It is the average annual investment return from the project.
- Discounted at the IRR, the NPV of the project cash flows must come to 0.
- The internal rate of return is therefore the discount rate that will give a net present value = Rs.0.

A company might establish the minimum rate of return that it wants to earn on an investment. If other factors such as non-financial considerations and risk and uncertainty are ignored:

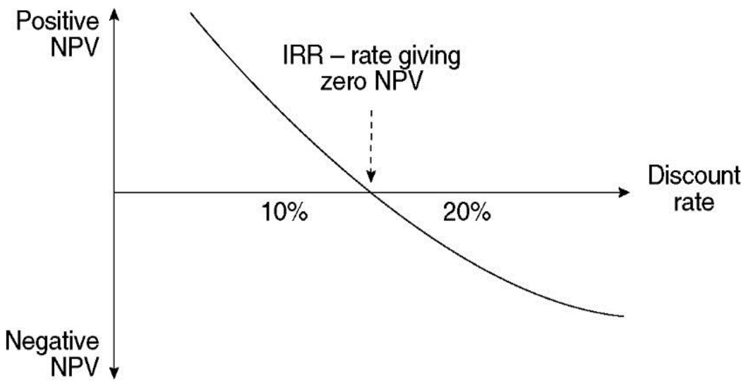
- If a project IRR is equal to or higher than the minimum acceptable rate of return, it should be undertaken.
- If the IRR is lower than the minimum required return, it should be rejected.

In the investment decisions, project is accepted or considered worthy if the required rate of return is less than the Internal Rate of Return or IRR. This is called the IRR Rule. This suggest that an investment is acceptable if the IRR exceeds the required return. It should be rejected otherwise. In other words, this investment is economically a break-even proposition when the NPV is zero because value is neither created nor destroyed[2].

Since NPV and IRR are both methods of DCF analysis, the same investment decision should normally be reached using either method.

The internal rate of return is illustrated in the diagram below:

▶ *Illustration:*



Calculating the IRR of an investment project

An approximate IRR can be calculated using interpolation or by trial and error. To calculate the IRR, you should begin by calculating the NPV of the project at two different discount rates.

- One of the NPVs should be positive, and the other NPV should be negative. (This is not essential. Both NPVs might be positive or both might be negative, but the estimate of the IRR will then be less reliable.)
- Ideally, the NPVs should both be close to zero, for better accuracy in the estimate of the IRR.

When the NPV for one discount rate is positive NPV and the NPV for another discount rate is negative, the IRR must be somewhere between these two discount rates.

Although in reality the graph of NPVs at various discount rates is a curved line, as shown in the diagram above, using the interpolation method we assume that the graph is a straight line between the two NPVs that we have calculated. We can then use linear interpolation to estimate the IRR, to a reasonable level of accuracy.

[2] Ross, Westerfield & Bradford (2008). Fundamentals Of Corporate Finance. 8th Edition. Mcgraw-Hill/Irwin. Pg. 277

- *Formula:*

$$IRR = A\% + \left(\frac{NPV_A}{NPV_A - NPV_B}\right) \times (B - A)\%$$

Ideally, the NPV at A% should be positive and the NPV at B% should be negative.

Where:

NPV_A = NPV at A%

NPV_B = NPV at B%

- *For example:*

A business requires a minimum expected rate of return of 12% on its investments. A proposed capital investment has the following expected cash flows. Estimate IRR:

Year	Cash flow
0	**(80,000)**
1	**20,000**
2	**36,000**
3	**30,000**
4	**17,000**

In calculating IRR, PV at 10% and 15% are calculated below:

Year	Cash flow	Discount factor at 12%	Present value at 12%	Discount factor at 10%	Present value at 10%	Discount factor at 15%	Present value at 15%
0	(80,000)	1.0000	(80,000)	1.000	(80,000)	1,000	(80,000)
1	20,000	0.8929	17,857	0.909	18,180	0.870	17,400
2	36,000	0.7972	28,699	0.826	29,736	0.756	27,216
3	30,000	0.7118	21,353	0.751	22,530	0.658	19,740
4	17,000	0.6355	10,804	0.683	11,611	0.572	9,724
NPV			(1,287)		+ 2,057		(5,920)

Now Using the formula:

$$IRR = A\% + \left(\frac{NPV_A}{NPV_A - NPV_B}\right) \times (B - A)\%$$

$$IRR = 10\% + \left(\frac{2{,}057}{2{,}057 - (-5{,}920)}\right) \times (15 - 10)\%$$

$$IRR = 10\% + \left(\frac{2{,}057}{2{,}057 + 5{,}920}\right) \times 5\%$$

$$IRR = 10\% + \left(\frac{2{,}057}{2{,}057 + 5{,}920}\right) \times 5\%$$

$$IRR = 10\% + \left(\frac{2{,}057}{7{,}977}\right) \times 5\%$$

$$IRR = 10\% + 0.258 \times 5\% = 10\% + 1.3\%$$

$$IRR = 11.3$$

The IRR of the project (11.3%) is less than the target return (12%).

The project should be rejected.

The following information is about a project.

Year	Rs.
0	(53,000)
1	17,000
2	25,000
3	16,000
4	12,000

This project has an NPV of Rs. 2,210 at a discount rate of 11%. Estimate the IRR of the project.

In calculating IRR, PV at two different points, NPV at 11% is given at 2,210. We should estimate a discount rate which yields a negative NPV. Let's calculate at 15%.

Year	Cash flow	Discount factor at 15%	Present value at 15%
0	(53,000)	1,000	(53,000)
1	17,000	0.870	14,790
2	25,000	0.756	18,900
3	16,000	0.658	10,528
4	12,000	0.572	6,864
NPV			(1,918)

Now Using the formula:

$$IRR = A\% + \left(\frac{NPV_A}{NPV_A - NPV_B}\right) \times (B - A)\%$$

$$IRR = 11\% + \left(\frac{2{,}210}{2{,}210 - (-1{,}918)}\right) \times (15 - 10)\%$$

$$\text{IRR} = 11\% + \left(\frac{2{,}210}{2{,}210 + 1{,}918}\right) \times 5\%$$

$$\text{IRR} = 11\% + \left(\frac{2{,}210}{4{,}128}\right) \times 5\%$$

$$\text{IRR} = 11\% + 0.535 \times 5\% = 11\% + 2.7\%$$

$$\text{IRR} = 13.7\%$$

The IRR of the project (13.7%) is greater than the target return (11%).

The following information is about a project.

Year	Rs.
0	(65,000)
1	27,000
2	31,000
3	15,000

This project has an NPV of Rs.(1,515) at a discount rate of 8%. Estimate the IRR of the project.

NPV at 8% is Rs.(1,515). A lower rate is needed to produce a positive NPV. (say 5%)

Year	Cash flow	Discount factor at 5%	Present value at 5%
0	(65,000)	1,000	(65,000)
1	27,000	0.952	25,704
2	31,000	0.907	28,117
3	15,000	0.864	12,960
NPV			1,781

Now Using the formula:

$$\text{IRR} = \text{A}\% + \left(\frac{\text{NPV}_\text{A}}{\text{NPV}_\text{A} - \text{NPV}_\text{B}}\right) \times (\text{B} - \text{A})\%$$

$$\text{IRR} = 5\% + \left(\frac{1{,}781}{1{,}781 - (-1{,}515)}\right) \times (8 - 5)\%$$

$$\text{IRR} = 5\% + \left(\frac{1{,}781}{1{,}781 + 1{,}515}\right) \times 3\%$$

$$\text{IRR} = 5\% + \left(\frac{1{,}781}{3{,}296}\right) \times 3\%$$

$$\text{IRR} = 5\% + 0.540 \times 3\% = 5\% + 1.6\%$$

$$\text{IRR} = 6.6\%$$

Annually a project provides for Rs. 12,000 for 12 years. The initial investment required is Rs. 90,000. Should the project be acceptable at 9% rate of return or 6%. What would be the rate where investor would be indifferent about the offer?

Since the investment returns is in the form of an annuity, NPV would be calculated at both the given rates.

$$\text{Present value} = X \times \left[\frac{1-\frac{1}{(1+r)^n}}{r}\right]$$

$$PV = 12{,}000 \times \left[\frac{1-\frac{1}{(1+0.09)^{12}}}{0.09}\right]$$

$$PV = 12{,}000 \times \left[\frac{0.6445}{0.09}\right]$$

$$PV = 85{,}929$$

$$NPV = -90{,}000 + 85{,}929 = (4{,}071)$$

At 9% the proposal must be rejected

$$\text{Present value} = X \times \left[\frac{1-\frac{1}{(1+r)^n}}{r}\right]$$

$$PV = 12{,}000 \times \left[\frac{1-\frac{1}{(1+0.06)^{12}}}{0.06}\right]$$

$$\text{Present value} = 12{,}000 \times \left[\frac{0.5030}{0.06}\right]$$

$$\text{Present value} = 100{,}606$$

$$NPV = -90{,}000 + 100{,}606 = 10{,}606$$

While at 6% the proposal can be accepted.

The rate at which the investor would be indifferent, can be estimated using the formula:

$$IRR = A\% + \left(\frac{NPV_A}{NPV_A - NPV_B}\right) \times (B - A)\%$$

$$IRR = 6\% + \left(\frac{10{,}606}{10{,}606 - (-4{,}071)}\right) \times (9 - 6)\%$$

$$IRR = 6\% + \left(\frac{10{,}606}{10{,}606 + 4{,}071}\right) \times 3\%$$

$$IRR = 6\% + \left(\frac{10{,}606}{14{,}677}\right) \times 3\%$$

$$IRR = 6\% + 0.723 \times 3\% = 6\% + 2.17\%$$

$$IRR = 8.17\%$$

IRR must be around 8%

STICKY NOTES

Future and present value for multiple cash flow situations occur when deposits or receipt for certain amount is multiple times in a given period.

Calculating the NPV of an investment project

- **Step 1: List all cash flows expected to arise from the project. This will include the initial investment, future cash inflows and future cash outflows.**
- **Step 2: Discount these cash flows to their present values using the cost that the company has to pay for its capital (cost of capital) as a discount rate.**
- **Step 3: The NPV of a project is difference between the present value of all the costs incurred and the present value of all the cash flow benefits (savings or revenues).**

The project is acceptable if the NPV is positive.

The project should be rejected if the NPV is negative.

An annuity is a series of regular periodic payments of equal amount. It is a constant cash flow for a given number of time periods whereas a perpetuity is a constant annual cash flow 'forever'.

There are two types of annuities:

- **Ordinary annuity – payments (receipts) are in arrears i.e. at the end of each payment period**
- **Annuity due – payments (receipts) are in advance i.e. at the beginning of each payment period.**

$$\text{Present value} = X \times \left[\frac{1-\frac{1}{(1+r)^n}}{r}\right]$$

Future Value of Ordinary Annuity =

$$Sn = X\left[\frac{(1+r)^n-1}{r}\right]$$

Future Value of Annuity due

$$Sn = X(1+r)\left[\frac{(1+r)^n-1}{r}\right]$$

SELF-TEST

6.1. Find the future value of an annuity of Rs.500 for 7 years at interest rate of 14% compounded annually.

(a) Rs.5,465.35 (b) Rs.5,565.35
(c) Rs.5,365.35 (d) Rs.5,665.35

6.2. Rs.200 is invested at the end of each month in an account paying interest 6% per year compounded monthly. What is the future value of this annuity after 10th payment?

(a) Rs.2,400 (b) Rs.2,044
(c) Rs.2,404 (d) Rs.2,004

6.3. Z invests Rs.10,000 every year starting from today for next 10 years. Suppose interest rate is 8% per annum compounded annually. Future value of the annuity is:

(a) Rs.156,654.87 (b) Rs.157,454.87
(c) Rs.156,555.87 (d) Rs.156,454.87

6.4. Rs.5,000 is paid every year for ten years to pay off a loan. What is the loan amount if interest rate is 14% per annum compounded annually?

(a) Rs.27,080.55 (b) Rs.25,080.55
(c) Rs.26,080.55 (d) Rs.24,080.55

6.5. The present value of an annuity of Rs.3,000 for 15 years at 4.5% p.a. C.I. is:

(a) Rs.23,809.41 (b) Rs.32,218.63
(c) Rs.32,908.41 (d) None of these

6.6. What is the present value of Rs.15,000 received at the end of the current year & next four years if the applicable rate is 7% per annum?

(a) Rs.61,502.96 (b) Rs.64,502.96
(c) Rs.62,502.96 (d) Rs.63,502.96

6.7. M/s. ABC Limited is expected to pay Rs.18 every year on a share of its stock. What is the present value of a share if money worth is 9% compounded annually?

(a) Rs.300 (b) Rs.200
(c) Rs.400 (d) Rs.100

6.8. 'A' borrows Rs.500,000 to buy a house. If he pays equal installments for 20 years and 10% interest on outstanding balance what will be the equal annual installment?

(a) Rs.58,729.84 (b) Rs.53,729.84
(c) Rs.56,729.84 (d) Rs.54,729.84

6.9. 'Y' bought a TV costing Rs.13,000 by making a down payment of Rs.3,000 and agreeing to make equal annual payment for four years. How much would each payment be if the interest on unpaid amount is 14% compounded annually?

(a) Rs.3,232.05 (b) Rs.3,332.05
(c) Rs.3,132.05 (d) Rs.3,432.05

AT A GLANCE
SPOTLIGHT
STICKY NOTES

6.10. How much amount is required to be invested every year so as to accumulate Rs.300,000 at the end of 10 years if interest is compounded annually at 10%?

(a) Rs.18,823.62 (b) Rs.18,523.62
(c) Rs.18,723.62 (d) Rs.18,623.62

6.11. ABC Ltd. Wants to lease out an asset costing Rs.360,000 for a five years period. It has fixed a rental of Rs.105,000 per annum payable annually starting from the end of first year. This agreement would be favourable to the company if the interest rate which the company earns on its investments is:

(a) 16% (b) 15%
(c) 17% (d) 14%

6.12. A loan of Rs.10,000 is to be paid back in 30 equal annual installments. The amount of each installment to cover the principal and 4% p.a. C.I. is:

(a) Rs.587.87 (b) Rs.597.87
(c) Rs.578.31 (d) None of these

6.13. A person desires to create a fund to be invested at 10% C.I. per annum to provide for a prize of Rs.300 every year. The amount he should invest is:

(a) Rs.2,000 (b) Rs.2,500
(c) Rs.3,000 (d) None of these

6.14. A firm wants to establish a library maintenance fund for a university. The firm would provide Rs.25,000 every 6 months. The fund yields a 10 percent annual rate of interest compounded semi annually. What is the initial deposit required to establish a perpetual stream of payments from the interest every 6 months after making the first payment from the principal?

(a) Rs.525,000 (b) Rs.526,000
(c) Rs.525,500 (d) Rs.525,800

6.15. A person has borrowed Rs.19,000 for a small business. The loan is for five years at an annual interest rate of 8 percent compounded quarterly. What is the amount of quarterly payments to pay back the loan?

(a) Rs.1,361.97 (b) Rs.1,261.97
(c) Rs.1,461.97 (d) Rs.1,161.97

6.16. A person deposits Rs.30,000 every six months into a retirement account. The account pays an annual interest rate of 12 percent compounded semi-annually. The value of account after 15 years would be:

(a) Rs.2,341,746 (b) Rs.2,371,746
(c) Rs.2,331,746 (d) Rs.2,351,746

6.17. A firm has set up a contingency fund yielding 16 percent interest per year compounded quarterly. The firm will be able to deposit Rs.1,000 into the fund at the end of each quarter. The value of the contingency fund at the end of 3 years is:

(a) Rs.15,225.80 (b) Rs.15,025.80
(c) Rs.15,325.80 (d) Rs.15,125.80

6.18. A Rs.680,000 loan calls for payment to be made in 10 annual installments. If the interest rate is 14% compounded annually. Annual payment to be made is:

(a) Rs.125,365.20 (b) Rs.130,365.20
(c) Rs.133,365.20 (d) Rs.135,365.20

6.19. Monthly payment necessary to pay off a loan of Rs.8,000 at 18% per annum compounded monthly in two years is:

(a)	Rs.419.40	(b)	Rs.409.40
(c)	Rs.399.40	(d)	Rs.389.40

6.20. A man agrees to pay Rs.4,500 per month for 30 months to pay off a car loan. If the interest of 18% per annum is charged monthly, the present value of car is:

(a)	Rs.108,271.27	(b)	Rs.108,171.27
(c)	Rs.108,671.27	(d)	Rs.108,071.27

6.21. A company is considering proposal of purchasing a machine either by making full payment of Rs.4,000 or by leasing it for four years requiring annual payment of Rs.1,250 or by paying Rs. 4,800 at the end of 2nd year. Which course of action is preferable if the company can borrow money at 14% compounded annually?

(a)	Leasing	(b)	Full payment
(c)	Rs. 4,800 after 2 years	(d)	Either (a) or (c)

6.22. A machine with useful life of seven years costs Rs.10,000 while another machine with useful life of five years costs Rs.8,000.The first machine saves labour expenses of Rs.1,900 annually and the second one saves labour expenses of Rs.2,200 annually. Determine the preferred course of action. Assume cost of borrowing as 10% compounded per annum.

(a)	First Machine	(b)	Second Machine
(c)	Both are same	(d)	Cannot be determined

6.23. A company borrows Rs.10,000 on condition to repay it with compound interest at 5% p.a. by annual installments of Rs.1,000 each. The number of years by which the debt will be clear is:

(a)	14.2 years	(b)	10 years
(c)	12 years	(d)	11 years

6.24. Mr. Dawood borrows Rs.20,000 on condition to repay it with C.I. at 5% p.a. in annual installments of Rs.2,000 each. The number of years for the debt to be paid off is:

(a)	10 years	(b)	12 years
(c)	11 years	(d)	None of these

6.25. A person invests Rs.500 at the end of each year with a bank which pays interest at 10% p.a. C.I. The amount standing to his credit one year after he has made his yearly investment for the 12th time would be:

(a)	Rs.11,764	(b)	Rs.10,000
(c)	Rs.12,000	(d)	None of these

6.26. Shiraz acquired a new car worth Rs.850,000 through a leasing company. He made a down payment of Rs.200,000 and has agreed to pay the remaining amount in 10 equal semi-annual installments. The leasing company will charge interest @ 19% per annum, over the lease term. Amount of semi-annual installment and total amount of interest is:

(a)	Rs.103,623 and Rs.386,230	(b)	Rs.103,533 and Rs. 386,000
(c)	Rs.103,523 and Rs.385,230	(d)	Rs.103,554 and Rs. 385,830

6.27. Ashraf purchased a new car and made a down payment of Rs.50,000. He is further required to pay Rs.30,000 at the end of each quarter for five years. The cash purchase price of the car, if the quarterly payments include 12% interest compounded quarterly, is:

(a)	Rs.498,324.25	(b)	Rs.496,324.25
(c)	Rs.499,324.25	(d)	Rs.497,324.25

6.28. Shahab has an opportunity to invest in a fund which earns 6% profit compounded annually. How much should he invest now if he wants to receive Rs.6,000 (including principal) from the fund, at the end of each year for the next 10 years? How much interest he would earn over the period of 10 year?

(a)	Rs.44,560.52 and Rs.15,939.48	(b)	Rs.44,260.52 and Rs.15,739.48
(c)	Rs.44,760.52 and Rs.15,239.48	(d)	Rs.44,160.52 and Rs.15,839.48

6.29. A firm wants to deposit enough in an account to provide for insurance payments over the next 5 years. Payment of Rs.27,500 must be made each quarter. The account yields an 8% annual rate compounded quarterly. How much be deposited to pay all the insurance payments?

(a)	Rs.448,664.41	(b)	Rs.447,664.41
(c)	Rs.449,664.41	(d)	Rs.446,664.41

6.30. How much money must be invested in an account at the end of each quarter if the objective is to have Rs.225,000 after 10 years. The account can earn an interest rate of 9 percent per year compounded quarterly. How much interest will be earned over the period?

(a)	Rs.3,547.87 and Rs.83,895.03	(b)	Rs.3,527.87 and Rs.83,885.03
(c)	Rs.3,557.87 and Rs.83,875.03	(d)	Rs.3,537.87 and Rs.83,845.03

6.31. A home buyer made a down payment of Rs.200,000 and will make payments of Rs.75,000 each 6 months, for 15 years. The cost of fund is 10% compounded semi-annually. What would have been equivalent cash price for the house? How much will the buyer actually pay for the house?

(a)	Rs.1,360,933.8 and Rs.2,440,000	(b)	Rs.1,362,933.8 and Rs.2,455,000
(c)	Rs.1,352,933.8 and Rs.2,450,000	(d)	Rs.1,372,933.8 and Rs.2,456,000

6.32. An individual plans to borrow Rs.400,000 to buy a new car. The loan will be for 3 years at a 12 percent annual rate compounded monthly. He can pay Rs.12,500 per month during the first year. What amount would he be required to pay during the next two years in order to repay the loan?

(a)	Rs. 13,755	(b)	Rs. 13,705
(c)	Rs. 13,655	(d)	Rs. 13,605

6.33. A person calculated that by depositing Rs.12,500 each year, starting from the end of the first year, he shall be able to accumulate Rs.150,000 at the time of nth deposit if the rate of interest is 4%. The number of years in which he can accumulate the required amount is:

(a)	9 years	(b)	10 years
(c)	12 years	(d)	11 years

6.34. A food distributor has borrowed Rs.950,000 to buy a warehouse. The loan is for 10 years at an annual interest rate of 12 percent compounded quarterly. The amount of quarterly payments which he must make to pay back the loan and the interest he would pay is:

(a)	Rs.41,199.26 & Rs.693,971.36	(b)	Rs.41,299.26 & Rs.693,972.36
(c)	Rs.41,099.26 & Rs.693,970.36	(d)	Rs.41,399.26 & Rs.693,973.36

6.35. A firm will need Rs.300,000 at the end of 3 years to repay a loan. The firm decided that it would deposit Rs.20,000 at the start of each quarter during these 3 years into an account. The account would yield 12 percent per annum compounded quarterly during the first year. What rate of interest should it earn in the remaining 2 years to accumulate enough amount in this account to pay the loan at the end of 3 years?

(a) 13.4% (b) 13.8%

(c) 14.2% (d) 14.6%

6.36. A machine costs a company Rs.1,000,000 & its effective life is estimated to be 20 years. If the scrap is expected to realize Rs.50,000 only. The sum to be invested every year at 13.25% compounded annually for 20 years to replace the machine which would cost 30% more than its present value is:

(a) Rs.14,797.07 (b) Rs.14,897.07

(c) Rs.14,697.07 (d) Rs.14,997.07

6.37. To clear a debt, a person agrees to pay Rs.1,000 now, another Rs.1,000 a year from now and another Rs.1,000 in two years. If the future payments are discounted at 8% compounded quarterly, what is the present value of these payments?

(a) Rs.2,777.41 (b) Rs.2,760.41

(c) Rs.2,767.41 (d) Rs.2,762.41

6.38. An equipment is bought for Rs.2,000 as down payment & a monthly installment of Rs.400 each for one year. If money worth 12% compounded monthly, what is the cash price of the equipment?

(a) Rs.6,602.03 (b) Rs.6,502.03

(c) Rs.6,402.03 (d) Rs.6,302.03

6.39. A __________ is a constant cash flow for a given number of periods. Fill in the blank.

(a) Perpetuity (b) discounting factor

(c) annuity (d) Net present value

6.40. A sinking fund is

(a) saving of a constant amount for future

(b) saving of a continuously changing amount for benefit in the future

(c) lending of a constant amount for specific benefit in future

(d) continuous cash flow for a long time in future

6.41. A company is deciding for an endowment fund which is likely to generate Rs. 20,000 per year. How much can be invested today when the rate of interest is 9%?

(a) Rs. 18,348 (b) Rs. 222,222

(c) Rs. 180,00 (d) Rs. 21,8000

6.42. In order to purchase an equipment, five annual instalments of Rs. 30,000 are required. Given the interest rate of 10%, what is the price that you will be paying for that equipment today?

(a) Rs. 27,273 (b) Rs. 33,000

(c) Rs. 28,500 (d) Rs. 14,560.3 correct option must be: 113,723

6.43. When calculating IRR, Net Present Value of the project cash flow must be:

(a) zero (b) neglected

(c) one (d) double

6.44. If an investment of Rupees 100,000 is made on a real estate business with a promise of 7% interest earned each year for the next 10 years. The annual interest would be______ .

(a)	7,000	(b)	700
(c)	17,000	(d)	7,700

6.45. A project requires Rs. 800,000 to be invested today. In return it will yield two inflows of certain amount at the end of first and second year respectively in such a way that the first amount is twice that of the second amount. If the cost of capital and net present value is 10% and Rs. 1,200,000 respectively compute the amount to be received at the end of second year.

(a)	1,512,500	(b)	756,250
(c)	2,000,000	(d)	400,000

6.46. A project has an internal rate of return of 15%. There is only one cash outflow of Rs 750,000 today and a single cash inflow of certain amount three years from today. Compute the net present value if cost of capital is 13%.

(a)	Rs. 40,532 positive	(b)	Rs 40,532 negative
(c)	Rs. 790,532 positive	(d)	Rs. 790,532 negative

6.47. An investor invests Rs 500,000 today with the objective of receiving two equal annual receipts at the end of first year and second year (inclusive of the original invested amount). If the interest rate is 16%, compute the amount of interest received in the second receipt.

(a)	Rs. 268,519	(b)	Rs. 42,963
(c)	Rs. 80,000	(d)	Rs. 268,518

6.48. A project requires certain amount to be invested today. The project will yield a single return at the end of four years which is four times the original investment. Compute Internal rate of return of the project

(a)	10 %	(b)	21.42 %
(c)	31.42 %	(d)	41.42 %

6.49. Shariq needs Rs. 400,000 at the end of fifth year. In order to have this amount he will invest certain amount at the end of each year for three years. If the interest rate is 8% compute the amount to be invested at the end of each year.

(a)	Rs.105,636	(b)	Rs. 214,086
(c)	Rs. 14,086	(d)	Rs. 314,086

6.50. Compute present value of a perpetual stream of cashflows of Rs. 200 each starting from the end of fifth year if discount rate is 10%.

(a)	Rs. 1,366	(b)	Rs. 2,000
(c)	Rs. 1,242	(d)	Rs. 2,400

6.51. Sadiq invests Rs. 40,000 at the end of each month for the first year and Rs 50,000 at the end of each month for the second year. Compute the total value of funds with him at the end of second year if interest rate is 12% per annum compounded monthly

(a)	Rs. 1,205,763	(b)	Rs. 1,305,763
(c)	Rs. 1,405,763	(d)	Rs. 2,105,763

6.52. Saqib has following investment options:

i. Invest Rs. 500,000 today and get Rs. 200,000 at the end of each year for four years.

ii. Invest Rs. 500,000 today and get Rs. 900,000 at the end of third year.

iii. Invest Rs. 500,000 today and get Rs. 200,000 each at the end of third and fourth year respectively.

If cost of capital is 15%. Which of the above option(s) must be selected based on net present value technique?

(a) None (b) All

(c) i and ii (d) ii and iii

6.53. Sadiq has following investment options:

i. Invest Rs. 500,000 today and get Rs. 300,000 at the end of each year for two years.

ii. Invest Rs. 500,000 today and get Rs. 900,000 at the end of third year.

Compute internal rate of return for each of the above options

(a) 18% and 21.64% (b) 13.07% and 18.05%

(c) 14.63% and 16.35% (d) 13.07% and 21.64%

6.54. Asif is thirty years old today, he wishes to set aside a certain amount every year starting from today till he turns 40. He intends to have Rs. 5,000,000 at the age of 40 to buy his own house. If interest rate is 11% per annum, compute the annual deposits he must make.

(a) Rs 255,605 (b) Rs. 299,007

(c) Rs. 268,905 (d) Rs. 300,000

6.55. An investment project has a net present value of Rs. 524,564 when its cash flows are discounted at 13% per year. Cash inflows from the project are Rs. 500,000 at the end of each year for three years starting from the end of third year. Compute the investment made today for the project.

(a) 350,000 (b) 375,000

(c) 400,000 (d) 425,000

6.56. Which of the following statement is **incorrect? (drag down mtq)**

(a) Net present value decreases if discounting rate is increased

(b) Net present value increases if discounting rate is decreased

(c) Net present value is zero if internal rate of return is used to discount project cash flows

(d) There is a direct relationship between net present value and discount rate, that is if one increases the other increases as well

6.57. Which of the following statement is incorrect? (drag down mtq)

(a) An annuity is a series of regular periodic payments of equal amount for a certain time interval

(b) A perpetuity is a constant cash flow forever.

(c) A project with a positive net present value must be carried out.

(d) A project with a negative net present value must be carried out.

6.58. Which of the following statement in correct? (drag down mtq)

(a) A project must be undertaken if internal rate of return is more than cost of capital.

(b) A project must be undertaken if internal rate of return is less than cost of capital.

(c) A project with a positive net present value must not be undertaken.

(d) A project with a negative net present value must be undertaken.

6.59. A project requires Rs. 600,000 today and will result in two annual cash inflows of Rs 300,000 at the end of first year and Rs. 500,000 at the end of second year respectively. Compute net present value and internal rate of return of the project if cost of capital is 5%

(a)	Rs. 239,229 and 7%	(b)	Rs. 139,229 and 19.65%
(c)	Rs. 239,229 and 19.65%	(d)	Rs. 139,229 and 7%

6.60. Present value of Rs. 200,000 at the end of each year for five years is Rs. 820,039.5.

What rate of discounting leads to above results?

(a)	6%	(b)	5%
(c)	7%	(d)	8%

6.61. Qasim wants to invest certain equal amount at the end of each year for two years in order to have sufficient funds for following expenses:

i. Rs. 300,000 at the end of third year

ii. Rs. 400,000 at the end of fourth year

If discounting rate is 10% per annum compute the required investment at the end of each of the two years.

(a)	Rs. 287,289	(b)	Rs. 294,289
(c)	Rs. 277,289	(d)	Rs. 267,289

6.62. Fahad will invest Rs. 35,000 at the start of each year for three years starting one year from today. Compute the total amount he will have at the end of third year if interest rate is 10% per annum

(a)	Rs. 115,850	(b)	Rs. 125,850
(c)	Rs. 110,850	(d)	Rs. 112,850

6.63. A project has a net present value of Rs 15,028 when discount rate is 15%. At what discount rate will the net present value of project be Rs. 11,111.

(a)	13%	(b)	14%
(c)	10%	(d)	20%

6.64. A project has a net present value of Rs 15,028 when discount rate is 15%. What will be the net present value if discount rate is increased to 20%

(a)	Rs. 11,111	(b)	Rs. 16,028
(c)	Rs. 17,028	(d)	Rs. 18,028

ANSWERS TO SELF-TEST QUESTIONS

6.1	**6.2**	**6.3**	**6.4**	**6.5**	**6.6**
(c)	(b)	(d)	(c)	(b)	(a)
6.7	**6.8**	**6.9**	**6.10**	**6.11**	**6.12**
(b)	(a)	(d)	(a)	(d)	(c)
6.13	**6.14**	**6.15**	**6.16**	**6.17**	**6.18**
(c)	(a)	(d)	(b)	(b)	(b)
6.19	**6.20**	**6.21**	**6.22**	**6.23**	**6.24**
(c)	(d)	(a)	(b)	(a)	(d)
6.25	**6.26**	**6.27**	**6.28**	**6.29**	**6.30**
(a)	(c)	(b)	(d)	(c)	(b)
6.31	**6.32**	**6.33**	**6.34**	**6.35**	**6.36**
(c)	(a)	(b)	(c)	(b)	(d)
6.37	**6.38**	**6.38**	**6.40**	**6.41**	**6.42**
(a)	(b)	(c)	(b)	(b)	
6.43	**6.44**	**6.45**	**6.46**	**6.47**	**6.48**
(a)	(a)	(b)	(a)	(b)	(d)
6.49	**6.50**	**6.51**	**6.52**	**6.53**	**6.54**
(a)	(a)	(a)	(c)	(d)	(a)
6.55	**6.56**	**6.57**	**6.58**	**6.59**	**6.60**
(c)	(d)	(d)	(a)	(b)	(c)
6.61	**6.62**	**6.63**	**6.64**		
(a)	(a)	(d)	(a)		

CHAPTER 7

DATA - COLLECTION AND REPRESENTATION

IN THIS CHAPTER

AT A GLANCE

Data is the source of information. Figures or values obtained, while collecting data, are usually processed further for its usefulness in business situation. This chapter discusses various data types, data collection methods and various methods used to organize and summarize available data.

For data representation, multiple graphical tools can be used to exhibit relation between variables, concepts and proportions. Some of the methods, that the chapter provides for include bar or pie charts, histograms, ogives and box plots.

1. DATA, TYPES AND COLLECTION METHODS

Data is a term that refers to facts or figures or known terms or values. Information is derived when facts are processed, structured and analysed for decision making and other implications.

In quantitative analysis, the information gathered through experiments, surveys or observations collectively can be referred to as data.[1] Values recorded can be against a particular variable. It can assume different values for different entities (where an entity is an observation made about persons, places ord things).

▶ *For example:*

- Values against two variables can be recorded as below:

Age of salesman (in yrs)	Distance travelled (in km)
35	561
61	550
42	499

- Heights (in cms) of 5 students in a competition are 125, 111, 98, 120 and 103.
- There were 56 participants in a conference, 40 of whom where foreigners.

Types of data:

There are multiple ways to categorize data and variables. This is because of the nature of observations required and values acquired.

Numerical and categorical: For numerical data, arithematic operations can be performed; whereas not for categorical data. Numeric data can be refered to as quantitative (including or of numeric values) where as categorial data can be qualtitative in nature.

▶ *For example:*

- Numeric data: age, salary or receipts, temperature.
- Categorical data: color of the eyes, location, gender, religion.

Discrete and continuous: Within numerical data, discrete data variables are those that can only take on certain and separate values (results from a count[2]). The measurement of such variable increases or jumps in gaps between possible values. Whereas, continuous variables can take any value within a given range.

▶ *For example:*

- Discrete: number of contracts or children
- Continuous: height, weight.

Nominal and ordinal: When data has an order it can be referred to as ordinal data. When no natural order can be determined then the data is nominal. With nominal data, change in order would not change the meaning of data.

▶ *For example:*

- Ordinal: Result scorecard (Am B or C), likert scale (Agree, Neutral, Disagree), sizes (small-medium-large)
- Nominal: Gender (male-female), good-bad, pass-fail.

[1] Agresti,A., Franklin, C., & Klingenberg, B. (2018). Statistics. The Art and Science of Learning from Data. Fourth Edition. Global Edition. With contributions by Michael Posner. (pg. 29)

[2] Albright, S. C., & Winston, W. L. (2013). Business Analytics: Data Analysis and Decision Making. Fifth Edition. Cengage Learning. Pg. 26

Cross-sectional or time series: Data on a cross section of a population at a distint point as the name suggest is cross-sectional data. Data collected can contain multiple observations but would be particular to that point in time. Time series data (usually observations of a single phenomenon) are collected and tracked over a period of time.

- *For example:*
  - Cross-sectional: Sales revenue, production of units for year 1, multiple stock positions for quarter 1.
  - Time-series: Year-wise sales revenue, quartely production of units over a period of time.

Primary and secondary: Primary data is collected first hand specifically for the investigation process. Secondary data, on the other hand, is data previously collected that may be relevant to the investigation under process.

- *For example:*
  - Primary: Consumer survey results, dosage output for a drug experiment.
  - Secondary: Published statistics on consumers' preferences in the market, earlier known experiemental analysis of the drug dosage and reactions.

Population and sample: Data resulting from investigating every member of a population can be referered to as population data. When representative data or the observations are selected within population for analysis it can be referred to as sample data. Sample data is usually extracted or reflective of the population data.

- *For example:*
  - Population: A classroom research involving behavioural response of all students in the class.
  - Sample: Behavioural sample of random group of students within a class.

Structured and unstructured: Data can be referred to as structured when there is a pre-defined model or organization. When data represents mere numbers, observations or qualities without any organized form, it can be referred to as unstructured.

- *For example:*
  - Unstructured: weather statistics, survelliance videos. Email messages.
  - Structured: hourly wheather staistics, customer addresses and occupancies, product size and value.

Quality of Data:

For Data to be useful and be of qualtiy, it should be:

- Complete;
- Relevant and significant
- Timely
- Detailed as required
- Clear and understandable
- Communicated via an appropriate channel to the right person
- Reliable and consistent

Data collection and methods:

Data can be required for various purposes. The objective of the observation or purpose of an intended study determines what type of data is required and how can it be collected.

Some of the areas of concerns include identification of factors affecting the variable. Understanding these factors will help in selecting the most appropriate method for data collection.

Some of the data collection methods are as follows:

1. Secondary data includes examination of existing data that are relevant to the specific study.
2. Surveys are used to gather behavioural responses, feedback or input from certain sample or population. For this pupose, a questionnaire is usually designed and used that collects data in a specific format that can be used for analysis later.
3. Experiments can be used to observe a phenomenon or study effects or causes of certain conditions and factors. Experiemtnal methods generally include control groups and involve factors that are controlled to observe specific reactions or effects towards a stimulus.
4. Interviews and recordings are used to gather qualitative data . Interviews can be conduced from focus groups or one-to-one.
5. Direct observations include counting or measurement of certain factors, phenomenon or behavior.

2. DATA ORGANIZATION AND SUMMARY

Raw data is often collected in large volumes. In order to make it useful, data must be rearranged into a format that is easier to understand and analyze further. For this purpose, both numeric and graphical formats can be used.

▶ *Illustration:*

The following list of figures is the distance travelled by 100 salesmen in one week (kilometres).

561	581	545	487	565	512	495
549	526	492	530	489	499	481
548	517	531	538	534	538	528
502	515	491	482	572	503	500
550	577	529	500	561	523	524
521	553	486	527	564	534	531
584	594	579	541	539	541	557
551	530	517	536	533	574	486
532	554	587	532	497	505	512
529	526	556	515	543	498	539
494	590	509	536	569	535	511
518	576	541	504	573	510	509
502	520	553	588	503	513	547
514	547	511	537	550	558	560
542	577					

The human mind cannot assimilate data in this form. The same can be summarized by rearranging numeric information into meaningful order. There may be following ways to organize data:

1. Array:

The obvious first step is to arrange the data into ascending or descending order. This is called an array.

▶ *Illustration:*

Distances can be arranged from shortest to longest in one week (in kilometers).

481	502	515	529	538	550	569
482	502	515	530	538	550	572
486	503	517	530	539	551	573
486	503	517	531	539	553	574
487	504	518	531	541	553	576
489	505	520	532	541	554	577
491	509	521	532	541	556	577
492	509	523	533	542	557	579
494	510	524	534	543	558	581
495	511	526	534	545	560	584
497	511	526	535	547	561	587
498	512	527	536	547	561	588
499	512	528	536	548	564	590
500	513	529	537	549	565	594
500	514					

AT A GLANCE
SPOTLIGHT
STICKY NOTES

Frequency distribution:

A frequency distribution shows various values a group of items may take, and the frequency with which each value arises within the group.

Many values in the above list occur more than once. The next step might be to simplify the array by writing the number of times each value appears. This is called **frequency** and is denoted with the letter ***f.***

- *Illustration:*

The data in this form is known as an ungrouped frequency distribution

	f		f		f		f		f
481	1	504	1	526	2	542	1	564	1
482	1	505	1	527	1	543	1	565	1
486	2	509	2	528	1	545	1	569	1
487	1	510	1	529	2	547	2	572	1
489	1	511	2	530	2	548	1	573	1
491	1	512	2	531	2	549	1	574	1
492	1	513	1	532	2	550	2	576	1
494	1	514	1	533	1	551	1	577	2
495	1	515	2	534	2	553	2	579	1
497	1	517	2	535	1	554	1	581	1
498	1	518	1	536	2	556	1	584	1
499	1	520	1	537	1	557	1	587	1
500	2	521	1	538	2	558	1	588	1
502	2	523	1	539	2	560	1	590	1
503	2	524	1	541	3	561	2	594	1

The distribution can be simplified further by grouping data. For example, instead of showing the frequency of each individual distance, frequency of weekly distances of 480 and above but less than 500 and 500 and above but less than 520 can be grouped together. Such groupings are known as ***classes***.

- *Illustration:*

Grouped frequency distribution for the distances travelled by the salesmen can be as follows:

Distance in km	Frequency (f)
480 to under 500	13
500 to under 520	22
520 to under 540	27
540 to under 560	19
560 to under 580	13
580 to under 600	6

Tally

In preparing for frequency distribution, data tally is used to organize data. Tally involves writing out the list of values (or classes) and then proceeding down the list of data in an orderly manner and for each data value placing a tally mark against each class which then can be counted for accurate frequency. Generally, counts of five are used to mark a tally that makes the counting of frequency easier.

▶ *Illustration:*

Distance in km	Tally	Frequency (f)
480 to under 500	𝍸 𝍸 |||	13
500 to under 520	𝍸 𝍸 𝍸 𝍸 ||	22
520 to under 540	𝍸 𝍸 𝍸 𝍸 𝍸 ||	27
540 to under 560	𝍸 𝍸 𝍸 ||||	19
560 to under 580	𝍸 𝍸 |||	13
580 to under 600	𝍸 |	6

Points to be noted in the construction of grouped frequency distribution are:

- Identify the minimum and maximum term in the dataset to create classes.
- The number of classes should be kept relatively small (say 6 or 7 or if the data is large 15 to 20).
- In general, classes should be of the same width (of equal class interval) for easier comparison.
- Class interval (class size or width) can be determined by dividing the range (Maximum-Minimum term) with the number of classes required.
- Each class would have a lower and upper class limit. Continuous data is already arranged into classes without any gaps. The upper limit of one class is the lower limit of the next so there are no gaps.
- The end limits of the classes must be unambiguous. For example, '0 to 10' and '10 to 20' leaves it unclear in which 10 would appear. '0 to less than 10' and '10 to less than 20' is clear.
- For discrete data, class boundaries can be created. Class boundary is the midpoint of the upper class limit of one class and the lower class limit of the subsequent class.

▶ *Illustration:*

Number of visits	Frequency	Class boundaries	Class width	Mid-point of class
41 to 50	6	40.5 to 50.5	10	45.5
51 to 60	8	50.5 to 60.5	10	55.5
61 to 70	10	60.5 to 70.5	10	65.5
71 to 80	12	70.5 to 80.5	10	75.5
81 to 90	9	80.5 to 90.5	10	85.5
91 to 100	5	90.5 to 100.5	10	95.5

Here,

- There are 6 classes in the data
- Mid-point is calculated by dividing the upper and lower limit in each class by 2.
- Class boundaries are created by subtracting 0.5 from the lower limit and adding 0.5 to the upper limit.
- Each class has a width of 10.

Principles of data tabulation

The following principles should be applied whenever data is tabulated

- Simplicity: The material must be classified and details kept to a minimum.
- Title: The table must have a clear title that explains its purpose.
- Source: Source of information should be clearly labelled.
- Units: The units of measurement should be clearly stated.
- Headings: Column and row headings should be concise and unambiguous.
- Totals: Totals and subtotals should be shown where meaningful.

3. GRAPHICAL REPRESENTATION OF DATA:

Data representation involves display of collected data into meaningful categories or graphical flowcharts.

One of the simplest ways of data representation is to distribute data into various categories and use graphical formats to present it. Usually measurements or set of measurements are placed into categories that are meant to define them. Frequency distributions, discussed earlier can be one of the ways of data representation too.

Once categorization is achieved, data can be displayed graphically. Some of the methods are relatively straightforward, whereas others are quite sophisticated.

Straightforward methods include:

- Bar charts; and
- Pie charts

More complex methods include

- Histograms
- Frequency polygons
- Ogives
- Graphs (including semi-log graphs)
- Stem and leaf displays
- Box and whisker plots
- Time-series plots

Note that histograms, frequency polygons and ogives are ways of plotting grouped frequency distributions.

All graphs, diagrams and charts should:

- have a heading;
- be neat;
- be large enough to be understood;
- be well labelled;
- show scales on axis;
- be accurate.

Bar charts:

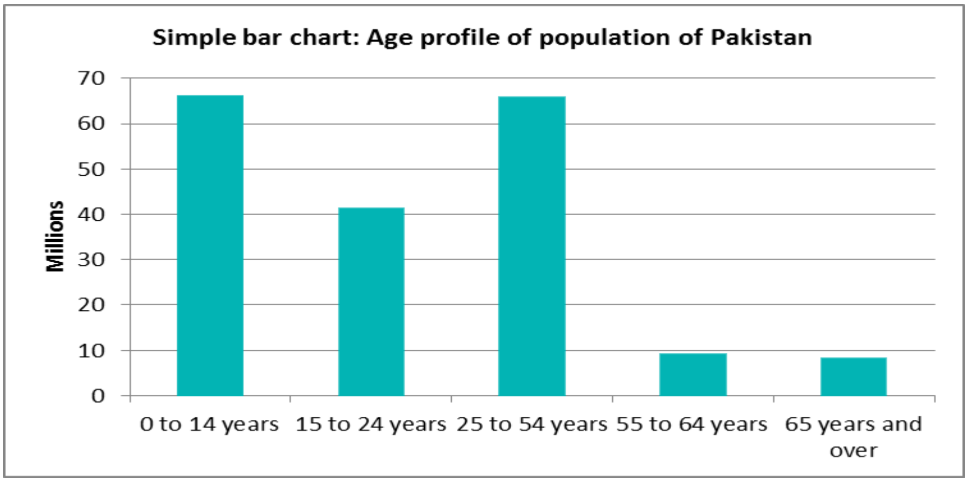
A bar chart or bar graph is a chart with rectangular bars with lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally and can take several forms (For Example, simple or multiple bar charts). Bar charts are usually used for plotting discrete (or 'discontinuous') data.

► *Illustration:*

Extract from Pakistan demographics profile 2013 (all figures in millions) is provided below

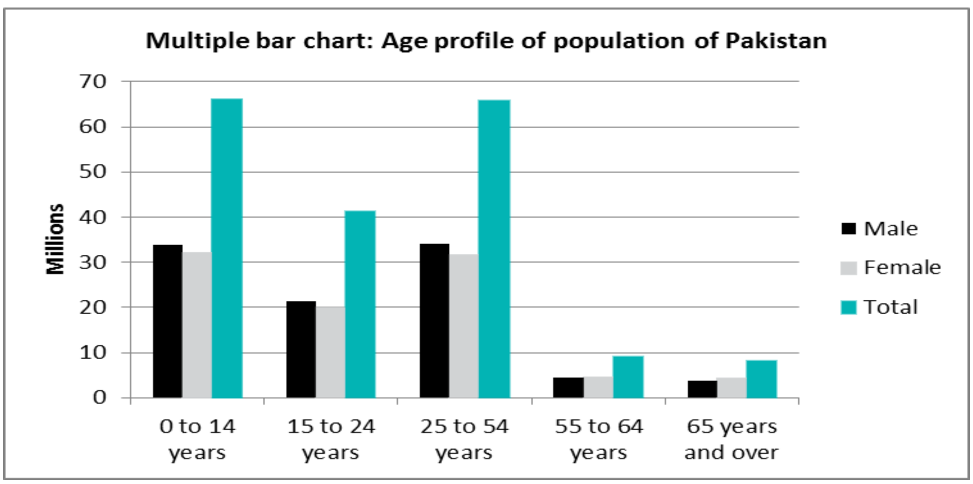
	Male	Female	Total
0 to 14 years	33.9	32.1	66.0
15 to 24 years	21.3	20.0	41.3
25 to 54 years	34.2	31.6	65.8
55 to 64 years	4.5	4.6	9.1
65 years and over	3.8	4.3	8.1

This data can be represented as a simple bar chart:



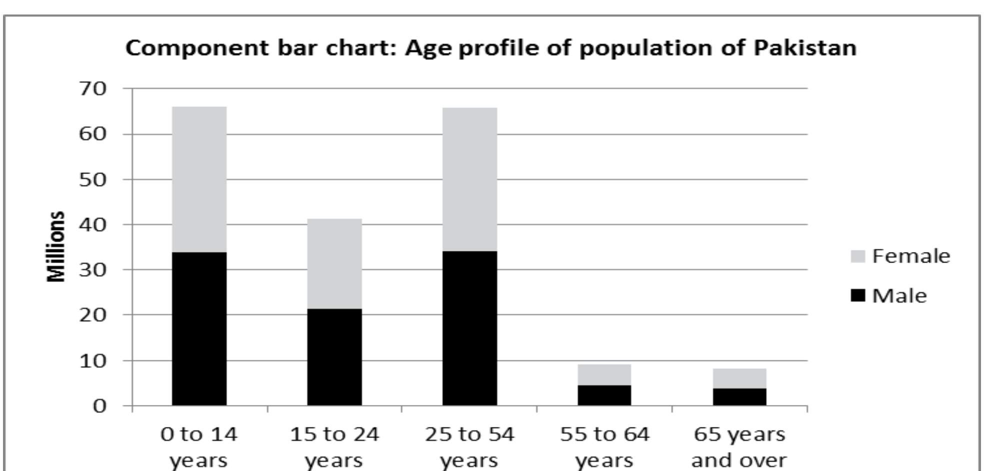
- *Multiple bar chart:*

  Multiple bar chart shows the components of totals as separate bars adjoining each other.



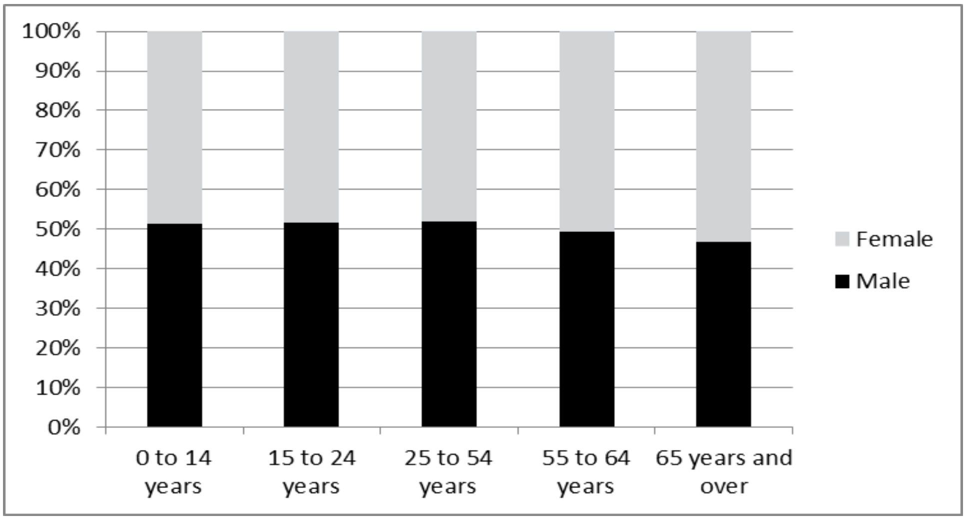
- *Component bar chart:*

  It is similar to a simple bar chart (since each bar indicates the size of the figure represented for a class) but the bars are subdivided into component parts.



- *Percentage component bar chart*

  This shows each class as a bar of the same height. However, each bar is subdivided into separate lengths representing the percentage of each component in a class.

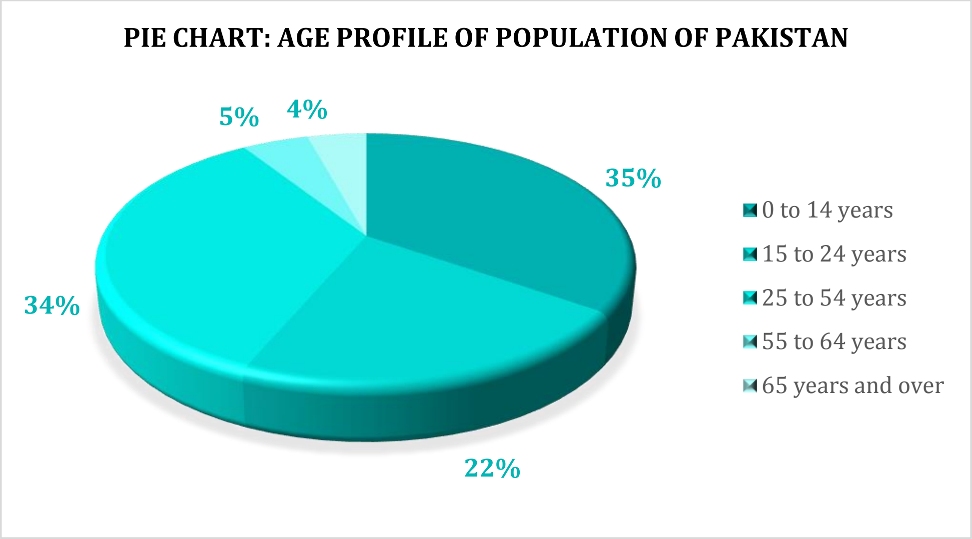


Pie charts

A pie chart (or a circle graph) is a circular chart that displays variables in proportion (or percentage) of the quantity within a circle (slice of a pie).

Both pie charts and bar charts are used for representing categorical data.

- *Pie chart:*



Histograms

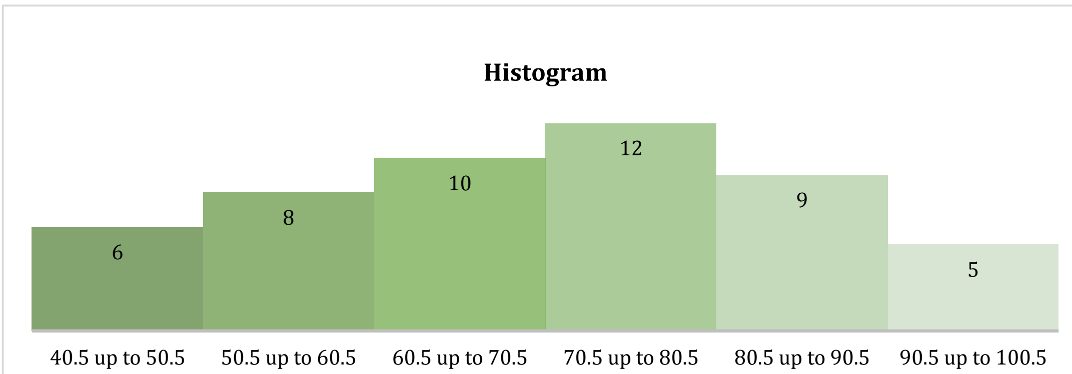
A histogram is a graph of a grouped frequency distribution.

It is constructed a little different from bar charts. A vertical rectangle is drawn to represent the frequencies (or relative frequency) for a quantitative grouped data. There are probably no gaps between the bars or vertical rectangles.

- *Illustration:*

When the class intervals are same, histogram can be constructed from the data in the example before:

Number of visits	Class boundaries	Mid-points	Frequency
41 to 50	40.5 to 50.5	45.5	6
51 to 60	50.5 to 60.5	55.5	8
61 to 70	60.5 to 70.5	65.5	10
71 to 80	70.5 to 80.5	75.5	12
81 to 90	80.5 to 90.5	85.5	9
91 to 100	90.5 to 100.5	95.5	5



When the class intervals are not same, there can be three methods to construct histograms:

- *By calculating adjusted frequency of the class:.*

Adjusted frequency of the class: $\frac{Minimum\ class\ size}{Class\ size} \times Frequency$

Number of visits	Class interval	No. of invoices (f)	Adjustment to frequency (f)	Height of bar
50 to under 100	50	5	50/50 ×5	5
100 to under 200	100	12	50/100 ×12	6
200 to under 300	100	17	50/ 100 × 17	8.5
300 to under 500	200	16	50/200 × 16	4
500 and over	200	8	50/200 × 8	2

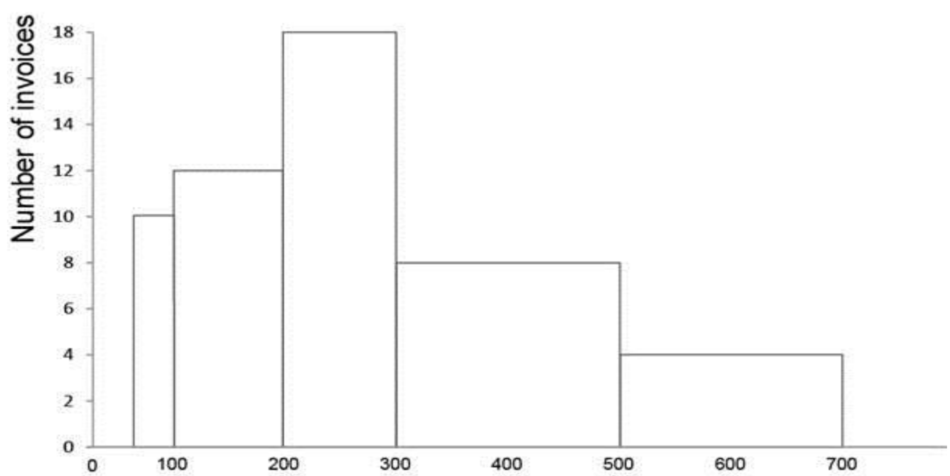
- *By calculating Frequency Density.*

Frequency Density : $\frac{Frequency}{Class\ width}$

Number of visits	Class interval	No. of invoices (f)	Frequency Density
50 to under 100	50	5	5/50 = 0.1
100 to under 200	100	12	12/100 = 0.12
200 to under 300	100	17	17/100 = 0.17
300 to under 500	200	16	16 / 200 = 0.08
500 and over	200	8	8 / 200 = 0.04

- *By adjusting the class intervals to equal:*

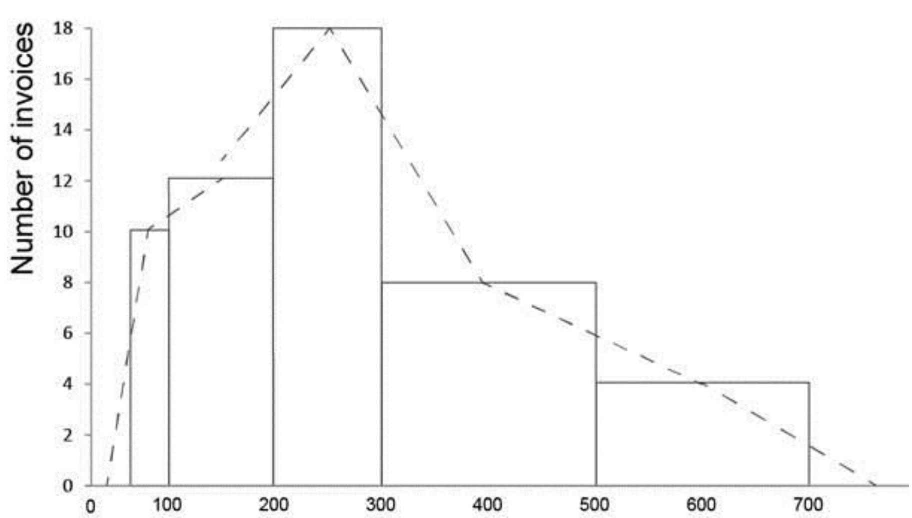
Number of visits	Class interval	No. of invoices (f)	Adjustment to frequency (f)	Height of bar
50 to under 100	50	5	× 2	10
100 to under 200	100	12		12
200 to under 300	100	17		17
300 to under 500	200	16	÷ 2	8
500 and over	200	8	÷ 2	4



Frequency polygons

A histogram (with equal class intervals) can be converted to a curve (or line graph) by joining the midpoint of the top of each rectangle with a straight line.

The curve is extended to the x axis at a point (mid-point) outside the first and last class interval.

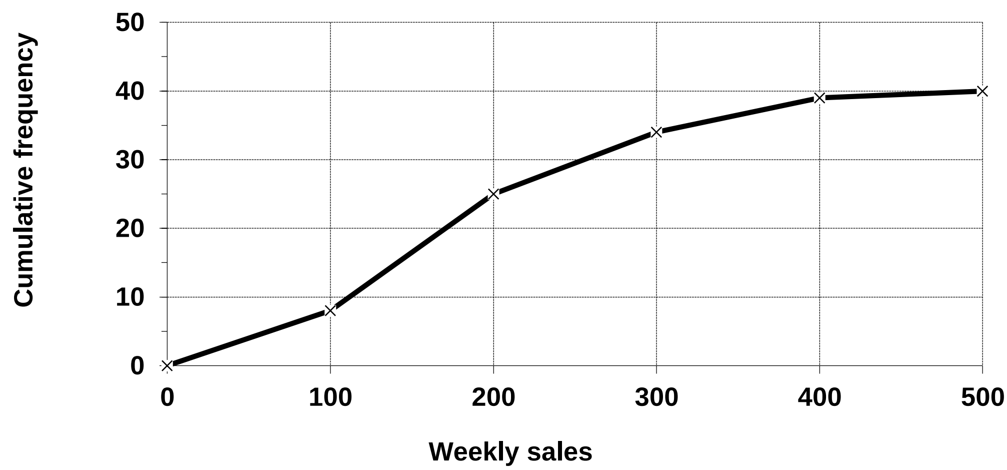


Ogives:

An ogive is the graph of a cumulative frequency distribution. Frequencies are accumulated at each level and are plotted against the upper class limits of each class, and the points are joined with a smooth curve.

► *Illustration:*

Weekly sales (Units)	*Frequency*	*Cumulative frequency*
0 to under 100	8	8
100 to under 200	17	8+17 = 25
200 to under 300	9	25+9 = 34
300 to under 400	5	34+5 = 39
400 to under 500	1	39+ 1 = 40



Interpretation

- Each point on the curve shows how many weeks had sales of less than the corresponding sales value. (It is a "less than" ogive).

Fractiles (or quantiles):

A fractile, also known as quantiles divides dataset into equivalent subgroups. The value within the fractile is the point of an item which is a given fraction of the way through a distribution.

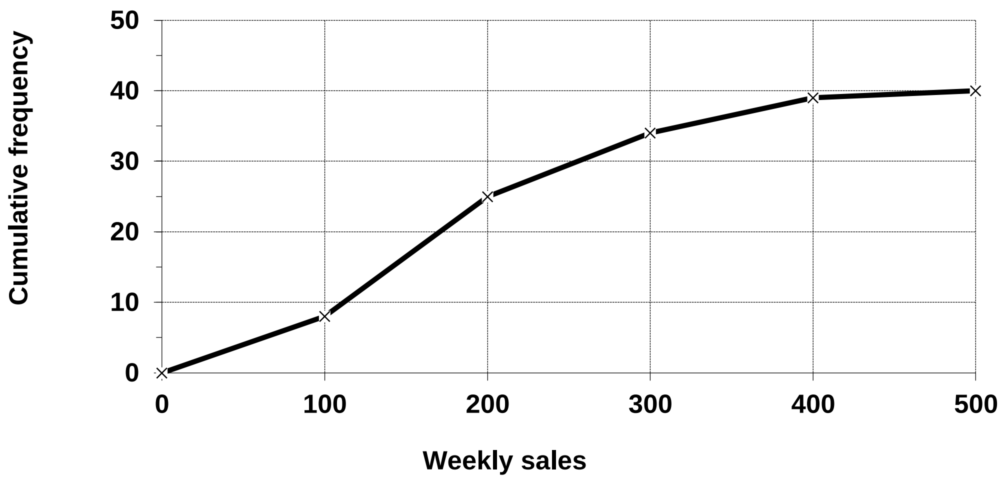
A fractile can be:

- A **percentile:** The value of an item one hundredth of a way through the distribution. The ***pth*** percentile is the value of an item such that ***p*** percent of the items fall below or at that value. 50th percentile is the median.
- A **decile:** The value of an item one tenth of a way through the distribution.
- A **quartile**: The value of an item one quarter of a way through the distribution. 25th , 50th, 75th and 100th percentiles represents the end of Q1, Q2, Q3 and Q4. Q2 is the median.

An ogive can be used to measure a fractile.

- *For example:*

  What is the 1st quartile of the distribution represented on the following ogive?



There are 40 items in the distribution so the 1st quartile is the value of the 10th item (25% of 40).

Reading off 10 on the vertical scale the 1st quartile is about 110.

Stem and leaf plots:

A stem-and-leaf display can be used to show frequency distribution while also retaining the raw numerical data.

A stem and leaf display is drawn as two columns separated by a line. The stem sits on the left hand side of the line and the leaves on the right hand side.

Units must be defined for both the stem and the leaves and data is sorted in ascending order. Usually, leaves are the last digit in a value and stem is made up of all the other digits in a number. For example, a stem of 48 followed by a leaf of 2 is a plot of the number 482.

- *Illustration:*

  Take this grouped data that shows classes of distances in km and their respective frequencies:

Distance in km	Frequency (f)
480 to under 500	13
500 to under 520	22
520 to under 540	27
540 to under 560	19
560 to under 580	13
580 to under 600	6
Total	**100**

For each number of values, the stem-leaf plot can be as follows:

AT A GLANCE
SPOTLIGHT
STICKY NOTES

Stem (units = 10)	Leaves (units = 1)
48	1, 2, 6, 6, 7, 9
49	1, 2, 4, 5, 7, 8, 9
50	0, 0, 2, 2, 3, 3, 4, 5, 9, 9
51	0, 1, 1, 2, 2, 3, 4, 5, 5, 7, 7, 8
52	0, 1, 3, 4, 6, 6, 7, 8, 9, 9
53	0, 0, 1, 1, 2, 2, 3, 4, 4, 5, 6, 6, 7, 8, 8, 9, 9
54	1, 1, 1, 2, 3, 5, 7, 7, 8, 9
55	0, 0, 1, 3, 3, 4, 6, 7, 8
56	0, 1, 1, 4, 5, 9
57	2, 3, 4, 6, 7, 7, 9
58	1, 4, 7, 8,
59	0, 4

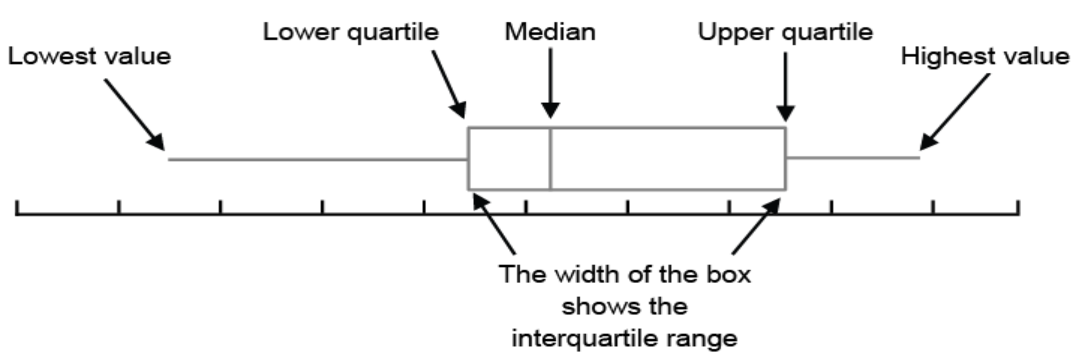
Here, distances between 480 and 499 would include 13 terms:
481, 482, 486, 486, 487, 489, 491, 492, 494, 495, 497, 498, 499

Box and whisker plots

A box plot is a visual representation of data clustered around some central value using a box drawn to incorporate median and quartiles values with whiskers extended to show the range of the data. It is usually drawn alongside a numerical scale.

The box in the boxplot encloses lower Quartile (Q1) to the upper Quartile (Q3) containing central 50% of the distribution[3]. Dividing line within the box is the median. The extending lines from each side are called whiskers.

- *Illustration:*

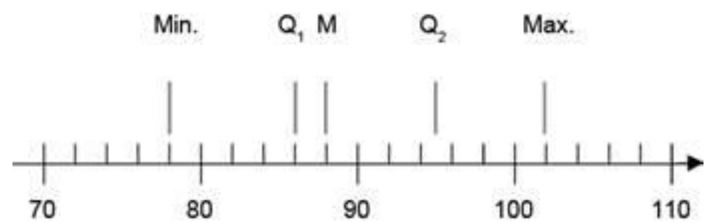
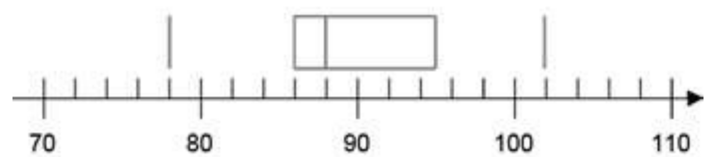
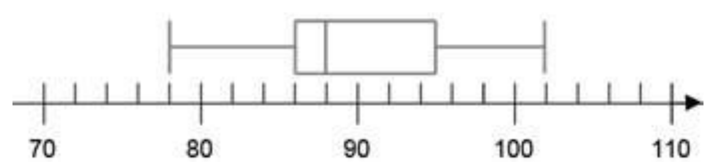


Constructing the plot:

- Arrange the data in ascending order.
- Identify the median that separates the data into dataset below and above it.
- Identify the top points of the first and third quartile (the medians of each half labelled Q1 and Q3). Medians in each data set (below and above the median value) is calculated.
- Construct an appropriate scale and draw in lines at M, Q1 and Q3 and the maximum and minimum values.
- Complete the box by drawing lines to join the tops and bottoms of the Q1 and Q3 lines.
- Draw lines (whiskers) from each end of the box to the maximum and minimum lines.

[3] Agresti,A., Franklin, C., & Klingenberg, B. (2018). Statistics. The Art and Science of Learning from Data. Fourth Edition. Global Edition. With contributions by Michael Posner. (pg. 96)

- *For example:*

Data set	86, 102, 78, 90, 88 98, 100, 94, 82, 92, 88, 86, 96, 88, 84, 90, 88
Ascending order	78, 82, 84, 86, 86, 88, 88, 88, 88, 90, 90, 92, 94, 96, 98, 100, 102
Median (M)	The 9th term = 88.
Finding Q1 and Q3	M 78, 82, 84, **86**, **86**, 88, 88, 88 88 90, 90, 92, **94**, **96**, 98, 100, 102 $Q_1 = \frac{86+86}{2} = 86$ $Q_3 = \frac{94+96}{2} = 95$
Scale	
Boxplot	
Whiskers	

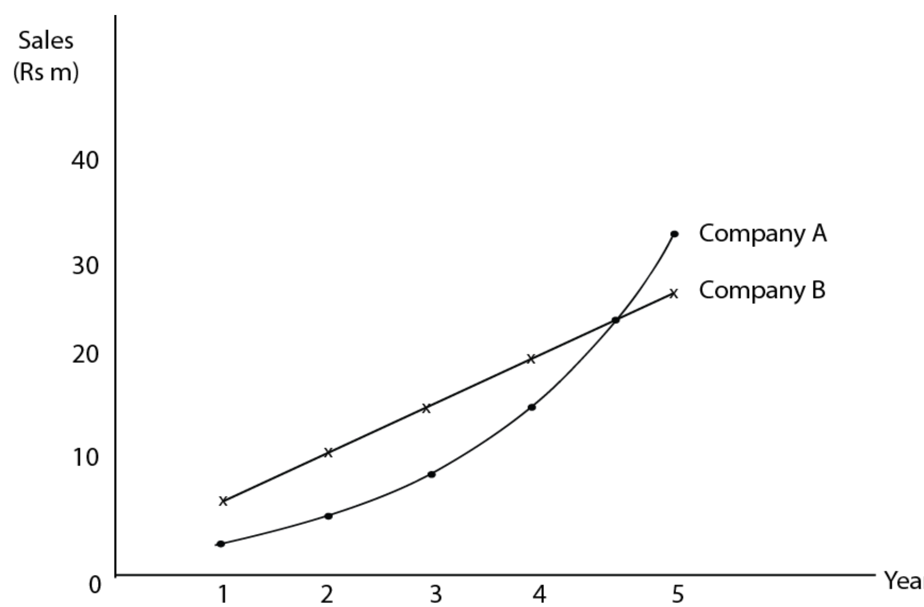
Time series graph

A time series graph is one that has time on the horizontal axis and a variable on the vertical axis.

▶ *Illustration:*

	Sales (Rs m)	
Year	**Company A**	**Company B**
1	2	5
2	4	10
3	8	15
4	16	20
5	32	25

The sales can be represented graphically as follows:



Shape of distribution:

Graphical representation of data can also provide for overall feel of the data and its inclination. Peaks and lows within the distribution and graphs constitutes overall shape of the distribution.

The shape of the distribution can be

- Symmetric: If the peak is in the middle of the histogram and the frequencies either side are similar to each other; OR
- Skewed: If peak lies to one side or the other of a histogram.

Skewness can be described in terms of:

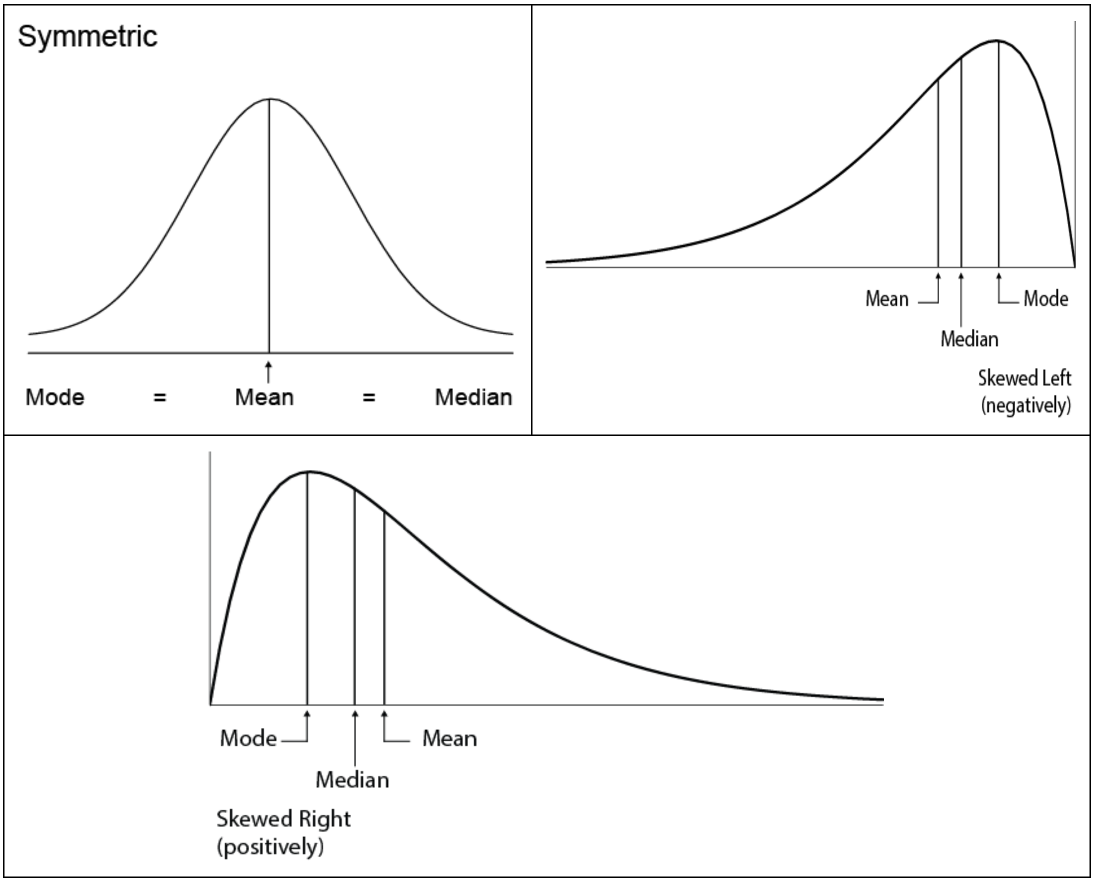
- Direction of skew: Right or positively skewed graphs are where tail of the graph is pulled towards positive numbers; whereas left or negatively skewed graphs are pulled towards negative numbers.

- Degree of skew: When the degree of skewness is 0 the distribution is symmetrical. Positive and negative value corresponds to positive (right) and negative (left) skewed distribution.
    - Pearson's Coefficient of skewness can be determined using the formula:

$$\text{Sk} = \frac{3(\text{Mean} - \text{Median})}{\text{Standard deviaton}}$$

Where: Sk = Coefficient of skewness

► *Illustration:*



STICKY NOTES

Types of Data include Numerical (quantitative) or Categorical (qualitative); Discrete (resulting from a count) or Continuous (any value within a given range); ordinal (where order matters) or nominal (change in order do not change the meaning of data); primary (collected first hand) or Secondary (collected previously).

Data Collection Methods can include examining existing data (secondary sources) or from experiments, direct observations or from interviews (primary sources).

From understanding large amount of data, various formats are used to make it easier to comprehend including use of arrays (rearranging data in ascending or descending order); frequency distribution, tally and data tabulations.

Graphically, data can be represented using bar charts, pie charts, histograms, ogives, box plots or other frequency polygons.

SELF-TEST

7.1. Primary data can be collected:

(a)	From newspapers	(b)	From State Bank
(c)	By carrying out surveys	(d)	All of these

7.2. Any recording of information, whether it be quantitative or qualitative is called:

(a)	Observation	(b)	Data
(c)	Sample space	(d)	None of these

7.3. The totality of the observations with which an statistician is concerned is known as:

(a)	Data	(b)	Sampling distribution
(c)	Population	(d)	Sample

7.4. A discrete variable is that which can assume:

(a)	Only integral values	(b)	Only fractional values
(c)	Whole number as well as fraction	(d)	None of these

7.5. A continuous variable is that which can assume:

(a)	Only integral values	(b)	Only fractional values
(c)	Whole number as well as fraction	(d)	None of these

7.6. Raw data, or unprocessed data or originally collected data are known as:

(a)	Sample data	(b)	Primary data
(c)	Secondary data	(d)	None of these

7.7. Processed or published data is known as:

(a)	Sample data	(b)	Primary data
(c)	Secondary data	(d)	None of these

7.8. An inquiry form comprising of a number of pertinent questions is knows as:

(a)	Inquiry form	(b)	Data collection form
(c)	Questionnaire	(d)	None of these

7.9. The arrangement of data according to magnitude of data is known as:

(a)	Classification	(b)	Tabulation
(c)	Array	(d)	None of these

7.10. The process of arranging data in groups or classes according to their resemblance or affinities is known as:

(a)	Classification	(b)	Frequency distribution
(c)	Tabulation	(d)	None of these

7.11. The purpose of classification and tabulation is to present data in:

(a)	Visual form	(b)	Easy to understand form
(c)	Frequency distribution	(d)	None of these

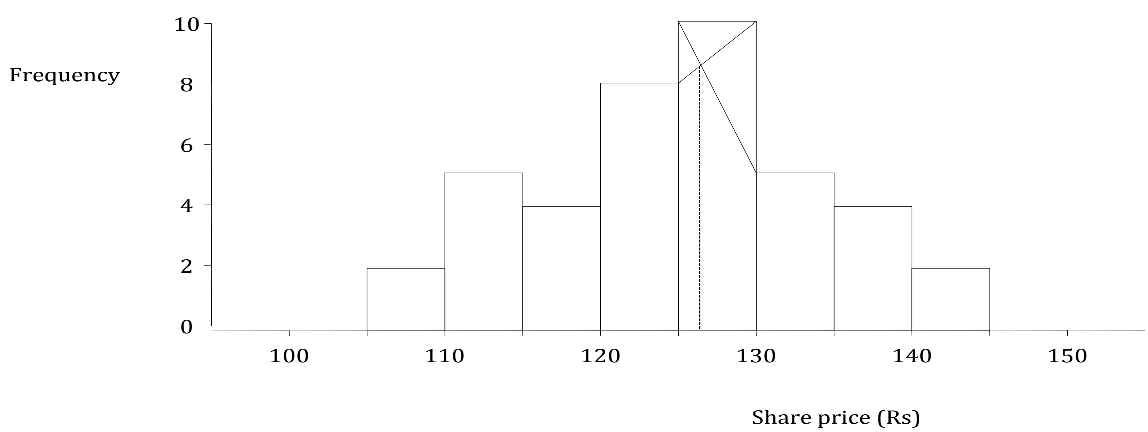
7.12. Any numerical value describing a characteristic of a population is called:

(a)	Parameter	(b)	Sample
(c)	Statistic	(d)	None of these

7.13. ______ data variables are those that results from counts or jumps in the gaps between possible values.

(a)	Continuous	(b)	Nominal
(c)	Discrete	(d)	Ordinal

7.14. Answer the following questions based on the given plot.



(i) How many stocks are between Rs. 110 and 120?

(a)	4	(b)	9
(c)	5	(d)	15

(ii) Total stocks recorded in the histogram are

(a)	10	(b)	40
(c)	145	(d)	125

(iii) What are the maximum and minimum number of stocks within the given intervals?

(a)	8 and 4	(b)	8 and 2
(c)	10 and 4	(d)	10 and 2

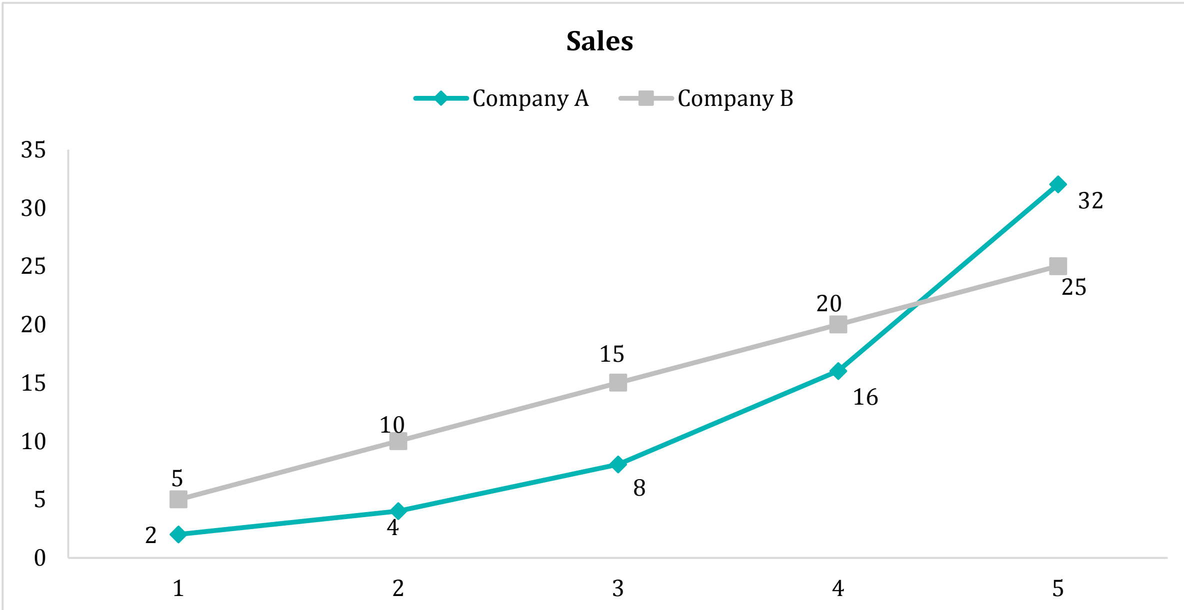
(iv) Determine the class width

(a)	5	(b)	10
(c)	20	(d)	100

(v) How many stocks are in between stock prices of 100 and 125

(a)	11	(b)	29
(c)	19	(d)	None of the above.

7.15. Answer the following questions based on the given plot.



(i) For how many years, company A sales have been less than company B's?

(a)	4 years	(b)	All 5 years
(c)	More than 4 but less than 5 years	(d)	Cannot tell from the graph

(ii) Sales recorded in year 3 for Company B are:

(a)	8	(b)	15
(c)	20	(d)	23

(iii) Total sales for year 5 for both the companies are:

(a)	32	(b)	7
(c)	25	(d)	57

7.16. A distribution that lacks symmetry with respect to a vertical axis is said to be:

(a)	Normal	(b)	Probability distribution
(c)	Skewed	(d)	None of these

7.17. In a given distribution if mean is less than median, the distribution is said to be:

(a)	Symmetrical	(b)	Positively skewed
(c)	Negatively skewed	(d)	None of these

7.18. In a given distribution if mean is greater than median, the distribution is said to be:

(a)	Symmetrical	(b)	Positively skewed
(c)	Negatively skewed	(d)	None of these

AT A GLANCE
SPOTLIGHT
STICKY NOTES

7.19. Which of the following is an example of discrete data?

(a)	Height of trees	(b)	Weight of bags
(c)	Length of tables	(d)	Number of students

7.20. Which of the following is an example of continuous data?

(a)	Number of books	(b)	Number of students
(c)	Number of chairs	(d)	Height of trees

7.21. Which of the following is an example of primary data?

(a)	Published statistics on consumers in the market	(b)	Earlier known experimental analysis of the drug reactions
(c)	Published story of an accident in newspaper	(d)	Consumer survey results

7.22. Select the qualities of good data from the following list:

(a)	Timely	(b)	Clear and understandable
(c)	Reliable	(d)	Consistent

7.23. ____ is called the arrangement of data into ascending or descending order.

(a)	Array	(b)	Display
(c)	Arrangement	(d)	Frequency distribution

7.24. A ____ shows the various values a group of items may take, and the frequency with which each value arises within the group.

(a)	Array	(b)	Histogram
(c)	frequency distribution	(d)	Ogive

7.25. A ____ is a circular graph that represents data.

(a)	Bar chart	(b)	Pie chart
(c)	Histogram	(d)	Stem and leaf

7.26.

Stem	Leaf
4	5, 5, 5, 6, 6, 7, 8
5	0, 0, 1, 1, 2, 2,
6	-
7	0
8	1, 2
9	9

What is the lowest value in above stem and leaf diagram

(a) 45 (b) 4

(c) 40 (d) 48

7.27.

Stem	Leaf
4	5, 5, 5, 6, 6, 7, 8
5	0, 0, 1, 1, 2, 2,
6	-
7	0
8	1, 2
9	9

How many values are there in above stem and leaf diagram?

(a) 18 (b) 17

(c) 16 (d) 20

7.28.

Stem	Leaf
4	5, 5, 5, 6, 6, 7, 8
5	0, 0, 1, 1, 2, 2,
6	-
7	0
8	1, 2
9	9

What is the highest value in above stem and leaf diagram?

(a) 99 (b) 45

(c) 100 (d) 90

7.29.

Stem	Leaf
4	5, 5, 5, 6, 6, 7, 8
5	0, 0, 1, 1, 2, 2,
6	-
7	0
8	1, 2
9	9

What are the number of values in 60s in above stem and leaf diagram

(a) 60 (b) None

(c) One (d) Cannot be determined

AT A GLANCE
SPOTLIGHT
STICKY NOTES

7.30.

Stem	Leaf
4	5, 5, 5, 6, 6, 7, 8
5	0, 0, 1, 1, 2, 2,
6	-
7	0
8	1, 2
9	9

What is the mid-value in above stem and leaf diagram?

(a) 50 (b) 51

(c) 99 (d) 45

7.31.

Stem	Leaf
4	5, 5, 5, 6, 6, 7, 8
5	0, 0, 1, 1, 2, 2,
6	-
7	0
8	1, 2
9	x

What can be the highest possible value of x in above box and whisker diagram.

(a) 90 (b) 99

(c) 95 (d) 98

7.32. Select the correct statements:

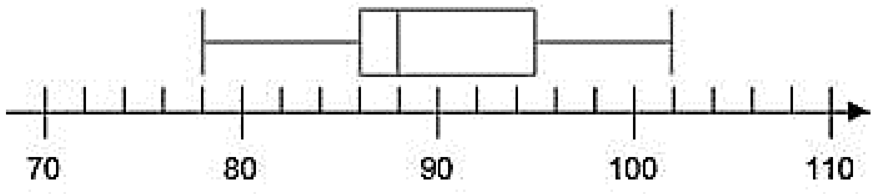
(a) Multiple bar chart shows the components of totals as separate bars adjoining each other.

(b) Component bar chart is similar to a simple bar chart but the bars are subdivided into component parts.

(c) Percentage component bar chart shows each class as a bar of the same height. However, each bar is subdivided into separate lengths representing the percentage of each component in a class.

(d) Histogram is a circular chart that displays variables in proportion of the quantity within a circle

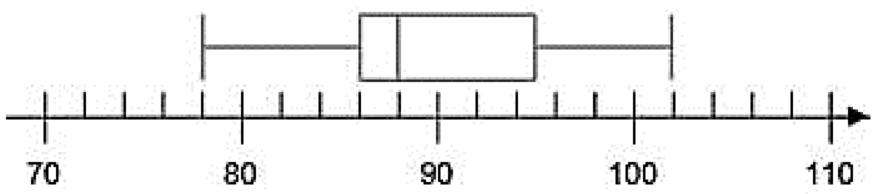
7.33. From the following box and whisker plot identify highest value:



(a) 78 (b) 86

(c) 88 (d) 102

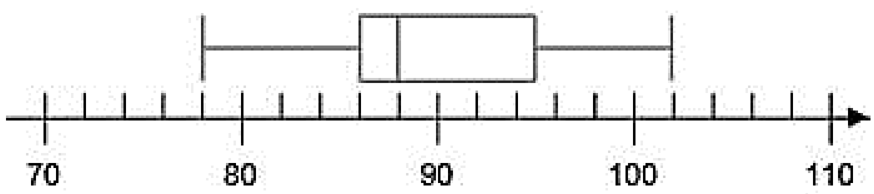
7.34. From the following box and whisker plot identify lowest value:



(a) 78 (b) 86

(c) 88 (d) 102

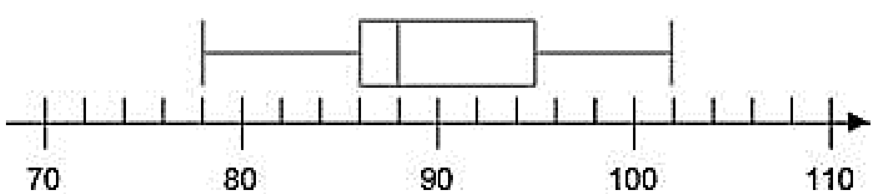
7.35. From the following box and whisker plot identify lower quartile value:



(a) 78 (b) 86

(c) 88 (d) 102

7.36. From the following box and whisker plot identify median value:



(a) 78 (b) 86

(c) 88 (d) 102

7.37. Which diagram is most suitable to represent data if following values are available only:

1. Lower quartile
2. Upper quartile
3. Median
4. Highest value
5. Lowest value

(a) Pie chart (b) Bar chart

(c) Histogram (d) Box and whisker

7.38. Following information is relevant for a pie chart:

Type of cost	Amount in Rupees	Degree of angle
Material	5,000	45^o
Labour	15,000	X
Overheads	20,000	Y

Based on above information identify the angle covered by labour cost on pie chart.

(a) 135^o (b) 180^o

(c) 45^o (d) 360^o

ANSWERS TO SELF-TEST QUESTIONS

7.1	**7.2**	**7.3**	**7.4**	**7.5**	**7.6**
(c)	(a)	(c)	(a)	(c)	(b)
7.7	**7.8**	**7.9**	**7.10**	**7.11**	**7.12**
(c)	(c)	(b)	(a)	(b)	(a)
7.13	**7.14(i)**	**7.14(ii)**	**7.14(iii)**	**7.14(iv)**	**7.14(v)**
(c)	(b)	(b)	(d)	(a)	(c)
7.15(i)	**7.15(ii)**	**7.15(iii)**	**7.16**	**7.17**	**7.18**
(c)	(b)	(d)	(c)	(c)	(b)
7.19	**7.20**	**7.21**	**7.22**	**7.23**	**7.24**
(d)	(d)	(d)	(a),(b),(c),(d)	(a)	(c)
7.25	**7.26**	**7.27**	**7.28**	**7.29**	**7.30**
(b)	(a)	(b)	(a)	(b)	(a)
7.31	**7.32**	**7.33**	**7.34**	**7.35**	**7.36**
(b)	(a),(b),(c)	(d)	(a)	(b)	(c)
7.37	**7.38**				
(d)	(a)				

CHAPTER 8

STATISTICAL MEASURES OF DATA

IN THIS CHAPTER

AT A GLANCE

From descriptive to predictive analysis, statistical measures help in better understanding of available data. For example, in summarizing the datasets, measures of central tendency (mean, median and mode) as well as measures of dispersion (range, standard deviation and variance) are used for primary statistical analysis.

The chapter deals with measures of central tendency and dispersion in detail with examples.

1 MEASURES OF CENTRAL TENDENCY

Data can be described through various means. Part of statistics that describes data or summarizes information from the data is referred to as Descriptive Statistics. Apart from describing data through distribution and graphical representation, measures of central tendency are used to understand given data by the location of its central value.

Median

The median is the middle value when the data is arranged in ascending or descending order. If there is an even number of items, then there is no middle item. For example, if there had been 8 values in the data the position of the median would be mid-value of the two middle terms.

In cases where there is a large number of values the position of the median can be found by using the formula.

There are two ways to find the median of a grouped frequency distribution. The first of these is to construct an ogive. The median is then the 50th percentile, 5th decile, or 2nd quartile which can be read off the ogive.

The second way involves finding which class the median is in and then estimating which position it occupies in this class.

- *Formula*

For ungrouped data:

$$\text{Position of the median} = \frac{n+1}{2}$$

Where:

n = number of items in the distribution or total observations

For Grouped data:

Step1: Identify the position of the median class by using the above formula

Step 2: calculate the median using the formula

$$median = L + \left[\frac{\left(\frac{\sum f}{2} - cf\right)}{f}\right] \times C$$

Where:

L = Lower boundary of the median class.

$\sum f$ = sum of frequencies

cf = cumululative frequency before the median class

f = frequency of the median class

C = size of the median class

- *For example:*

For Ungrouped data

Calculate Median for: 22, 18, 19, 24, 19, 21, 20

Arrange the numbers in ascending order: 18, 19, 19, 20, 21, 22, 24

Finding the middle value by dividing the sequence into halves. <u>18, 19, 19</u>, **20**, <u>21, 22, 24</u>

Median here is 20.

Or:

Position of the median can be found by:

$$\text{Position of the median} = \frac{7+1}{2} = 4^{th} item$$

The 4th item in the array is 20

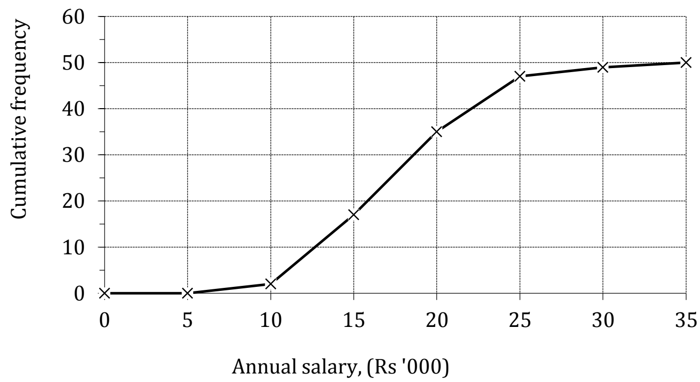
Method 1: For Grouped Data

Monthly Salary (Rs 000)	Number of people (f)	Cumulative frequency distribution	
5 and under 10	2	2	
10 and under 15	15	17	
15 and under 20	18	35 ⇐	Median Class
20 and under 25	12	47	
25 and under 30	2	49	
30 and over	1	50	

A diagram representing above data with annual salary on x-axis and cumulative frequency on y-axis can be drawn. This is known as Ogive.

Ogive focuses on cumulative frequency and is helpful in finding how many data values are below certain level of salary.

Ogive can be constructed as follows:



Median = 17 (perhaps 17.2)

Method 2: For Grouped Data

Position of the median $= \frac{\text{n} + 1}{2} = \frac{50 + 1}{2} = 25.5^{th}$ item

This places into the "15 and under 20" class

Therefore,

L = 15, $\sum$f= 50, cf=17, 7=18 and C=5

Therefore, the median is:

$$median = L + \left[\frac{\left(\frac{\sum \text{f}}{2} - cf\right)}{f}\right] \times C$$

$$median = 15 + \left[\frac{\left(\frac{50}{2} - 17\right)}{18}\right] \times 5$$

$$median = 15 + \left[\frac{(25 - 17)}{18}\right] \times 5$$

$$15 + \left(\frac{8}{18}\right) \times 5 = 17.2$$

Mode

This is the most frequently occurring value (or the highest frequency in grouped data). There may be more than one mode, or there may be no mode at all.

In Grouped data, mode can be estimated using the formula. In addition, a histogram can be constructed to find the mode of a grouped frequency distribution. This is found by drawing two straight lines:

- From the top left hand corner of the modal class to the top left hand corner of the class above it;
- From the top right hand corner of the modal class to the top right hand corner of the class below it.

A vertical line is drawn straight down to the x axis where these two lines cross. The mode is where this vertical line hits the x axis.

► *Formula:*

$$mode = L + [\frac{(f_m - f_1)}{(f_m - f_1) + (f_m - f_2)}] \times C$$

Where:

L = Lower boundary of the median class.

f_m= Frequency of the modal class

f_1 = frequency of the class before the modal class

f_2 = frequency of the class after the modal class

C = size of the median class

► *For example:*

Identify mode for: 22, 18, 19, 24, 19, 21, 20

Mode = 19

For Grouped data: modal class is the one with the highest frequency.

Monthly Salary (Rs 000)	Number of people	
5 and under 10	2	
10 and under 15	15	
15 and under 20	18	⇦ Modal Class
20 and under 25	12	
25 and under 30	2	
30 and over	1	

L = 15, f_m = 18, f_1 = 15, f_2 = 12, C = 5

$$mode = L + [\frac{(f_m - f_1)}{(f_m - f_1) + (f_m - f_2)}] \times C$$

$$mode = 15 + \left[\frac{(18 - 15)}{(18 - 15) + (18 - 12)}\right] \times 5$$

$$mode = 15 + [\frac{(3)}{(3) + (6)}] \times 5$$

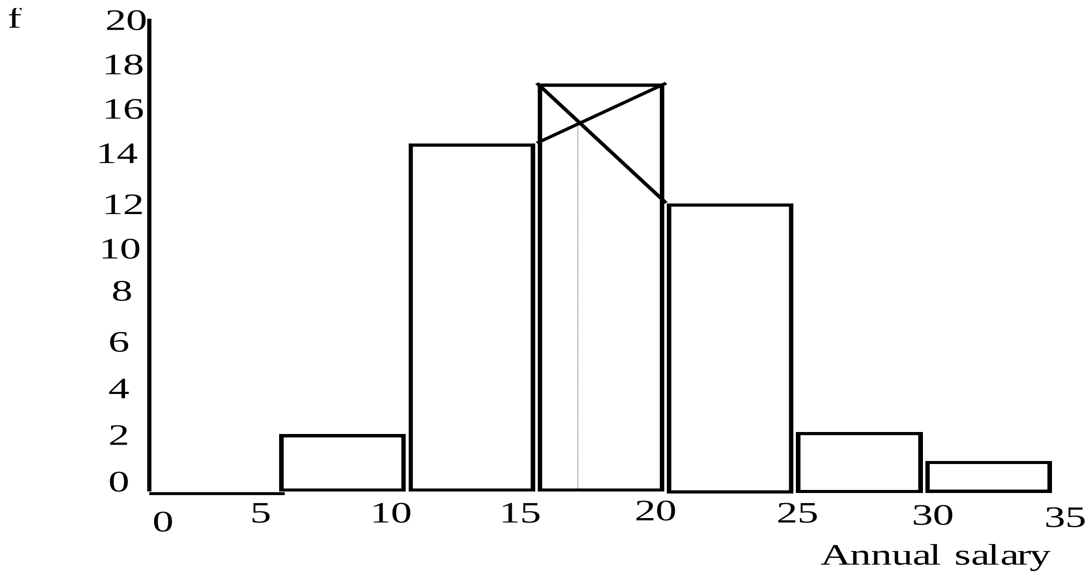
$$mode = 15 + [\frac{3}{9}] \times 5$$

$$mode = 15 + 1.6667$$

$$mode = 16.667$$

- *For example:*

For mode using histogram,



The mode is about 17. This is found by:

- Drawing straight line from the top left hand corner of the modal class to the top left hand corner of the class above it.
- Drawing straight line from the top right hand corner of the modal class to the top right hand corner of the class below it.
- A vertical line is drawn straight down the x-axis where these two lines cross. The mode is where this vertical line hits the x-axis.

Mean or Averages

The arithmetic mean is a mathematical representation of the typical value of a series of numbers. It is computed as the sum of the numbers in the series divided by the count of numbers in that series.

For Grouped data, the mean can be calculated by multiplying the midpoints with the class frequencies and dividing it by sum of all frequencies.

- *Formula:*

For Ungrouped data

$$\bar{x} = \frac{\sum x}{n}$$

Where:

$\bar{x}$ = mean

n = number of items

$\sum x$ = the sum of all values of x

For Grouped data:

$$\bar{x} = \frac{\sum fx}{\sum f}$$

Where:

$\bar{x}$= mid point of the class

f = number of incidences in a class

The mean need not be one of the values found in the series. In fact it might not even be a round number.

The mean is usually taken as the best single figure to represent all the data.

- *For example:*

For Ungrouped data

Calculate mean for: 22, 18, 19, 24, 19, 21, 20

$$\bar{x} = \frac{\sum x}{n}$$

$$\bar{x} = \frac{22 + 18 + 19 + 24 + 19 + 21 + 20}{7}$$

$$= \frac{143}{7}$$

$$= 20.4$$

For Grouped Data:

Monthly Salary (Rs 000)	Midpoint of class (x)	Number of people (f)	fx
	x	f	
5 and under 10	7.5	2	15
10 and under 15	12.5	15	187.5
15 and under 20	17.5	18	315
20 and under 25	22.5	12	270
25 and under 30	27.5	2	55
30 and over (closed at 35)	32.5	1	32.5
		50	875

$$\bar{x} = \frac{\sum \text{fx}}{\sum \text{f}} = \frac{875}{50} = 17.5$$

Combined Arithmetic Mean

For several groups of data, the combined mean is not as straightforward as it might sound. It is not a case of simply taking each mean, adding them up and dividing by the number of means as we would when we usually calculate the arithmetic mean. This is because this approach ignores the different sample sizes that gave rise to each mean. To use an extreme example to illustrate this suppose there were two groups of students. The first group consisted of 100 students all of whom achieved a mark of 50% in an exam. The second "group" was one student who achieved 100% on the same exam. Clearly the mean mark for all students is **not** the mean of the two means (75%).

The formula required to calculate combined means would be as follows:

- *Formula:*

$$= \frac{(\text{Mean}_A \times n_A) + (\text{Mean}_B \times n_B)}{n_A + n_B}$$

Where:

n_A = number of observations in group A

n_B = number of observations in group B

- *For example:*

The following information is about the average marks achieved by two groups of students in the same exam

	Group A	Group B
Number of students	1,200	400
Mean score	60	70

$$\frac{(\text{Mean}_A \times n_A) + (\text{Mean}_B \times n_B)}{n_A + n_B}$$

$$\frac{(60 \times 1{,}200) + (70 \times 400)}{1{,}200 + 400}$$

$$\frac{72{,}000 + 28{,}000}{1{,}600} = \frac{100{,}000}{1{,}600} = 62.5$$

Weighted arithmetic mean

The weighted arithmetic mean is similar to an arithmetic mean, but instead of each of the data points contributing equally to the final average, some data points contribute more than others.

The weighted arithmetic mean is computed by taking into account the relative importance of each item. The weighting is assigned to each number under consideration, to reflect the relative importance of that number according to a required criterion.

- *Formula:*

$$\bar{x} = \frac{\sum \mathrm{w}x}{\sum \mathrm{w}}$$

Where:

x = Value of the item

w = Weight of the item

- *For example:*

A university screens students for entrance by assigning different weights to the students' exam scores.

The following table shows the marks of two students along with the weighting the university assigns to the exam subjects

Subject	Weighting	Student A	Student B
Maths	5	20	70
Science	4	50	80
Geography	2	80	50
Art	0	70	20
Total	11	220	220
Arithmetic mean		55	55

Weighted arithmetic mean:

Student A $$\frac{(5\times 20)+(4\times 50)+(2\times 80)+(0\times 70)}{11}$$

$$\frac{100+200+160+0}{11}=\frac{460}{11}=42$$

Student B $$\frac{(5\times 70)+(4\times 80)+(2\times 50)+(0\times 20)}{11}$$

$$\frac{350+320+100+0}{11}=\frac{770}{11}=70$$

Geometric mean

The geometric mean is useful whenever several quantities are multiplied together to arrive at a product. It can be used, for example, to work out the average growth rate over several years.

If a property increased in value by 10% in year 1, 50% in year 2 and 30% in year 3 the average growth rate cannot be calculated as the arithmetic mean of 30%.

This is because the value of the property in the first year was multiplied by 1.1 and this total was then multiplied by 1.5.

The geometric mean calculates the average growth (compounding rate) over the period.

- *Formula:*

Geometric mean

$$=\sqrt[n]{x_1\times x_2\times x_3\ \ldots\ldots n}$$

Where:

n = number of terms in the series (e.g. years of growth)

x = Numbers used to multiply the starting value (with percentage increases these values are 1 + the percentage growth rate in each year).

Geometric mean

$$=\sqrt[n]{\frac{value\ at\ end\ of\ period\ of\ \ n\ years}{value\ at\ start}}$$

Where:

n = number of terms in the series (e.g. years of growth)

- *For example:*

If a property increased in value by 10% in year 1, 50% in year 2 and 30% in year 3

$$\sqrt[n]{x_1\times x_2\times x_3\ \ldots\ldots n}$$

$$\sqrt[3]{1.1\times 1.5\times 1.3}$$

$$\sqrt[3]{2.145}=1.28966\text{ or }1.29$$

Therefore, the average growth rate over the three-year period = 29%

As for the formula 2: Let the value at the start = 100, then the value at the end would be calculated as:

$$= 100 \times 1.1 \times 1.5 \times 1.3 = 214.5$$

This can be used for the calculation of Geometric mean

$$\sqrt[n]{\frac{\text{value at end of period of } n \text{ years}}{\text{value at start}}}$$

$$\sqrt[3]{\frac{214.5}{100}} = 1.2896 \text{ or } 1.29$$

The second method is useful when you are given actual values at each point in time rather than year by year growth rates.

Harmonic mean

This mean is the number of observations divided by the sum of the reciprocals of the values of each observation.

It is not used much but is useful in working out average speeds.

- *Formula:*

$$\text{Harmonic mean} = \frac{n}{\sum \frac{1}{x}}$$

Where:

n = number of items

x = values of items

- *For example:*

A lorry drove from Karachi to Lahore at an average speed of 60 kilometres per hour and back at an average speed of 40 kilometres per hour.

What was the average speed of the journey?

$$\text{Harmonic mean} = \frac{n}{\sum \frac{1}{x}}$$

$$\text{Harmonic mean} = \frac{2}{\frac{1}{60} + \frac{1}{40}} = \frac{2}{0.04167} = 48$$

The average speed over the whole journey was 48 kilometres per hour

AT A GLANCE
SPOTLIGHT
STICKY NOTES

2 MEASURES OF DISPERSION

Knowledge of the measures of central tendency are helpful to an extent, however they give no indication of the degree of clustering around that value. The variability in the data can provide deeper insights. Following are some of the measures that might be used to provide such information including:

Range

The range is simply the difference between the highest and lowest value in a distribution.

For grouped data range can be calculated by taking the difference between the "upper limit of the last interval and the lower limit of the first interval"[1].

Range is simple to calculate and easy to understand. However, it is distorted by extreme values.

- *Formula:*

  Range = $n_{(max)} - n_{(min)}$

- *For example:*

  For Ungrouped data

  The following are the ages of 7 students:

  22, 18, 19, 24, 19, 21, 20

  Age range = 24 - 18 = 6 years

  For Grouped Data:

Monthly Salary (Rs 000)	Number of people
5 and under 10	2
10 and under 15	15
15 and under 20	18
20 and under 25	12
25 and under 30	2
30 and over	1

  If the last class closed off at 35 (to create a class, the same size as the others).

  Range = 35 - 5 = 30 (i.e. Rs 30,000)

Inter-quartile range:

The Inter-quartile range is the range of the middle 50% of the data i.e. those range in values of items that are in between Q1 and Q3 or between 75th and 25th percentile.

Semi interquartile range is half of inter-quartile range.

When the data is concentrated about the median then the interquartile range can be misleading. This is because it ignores the extremes in dataset[2]. The semi inter-quartile range is useful, if a distribution is skewed (i.e. asymmetrical about the mean). When it is used it is usually used with the median (which is Q2) as a pair of measures (average and dispersion).

[1] Ott, R.L. & Longnecker, M. (2010). An Introduction to Statistical Methods & Data Analysis. Seventh Edition. Cengage Learning. Pg. 91

[2] Ott, R.L. & Longnecker, M. (2010). An Introduction to Statistical Methods & Data Analysis. Seventh Edition. Cengage Learning. Pg. 95

- *Formula:*

  Interquartile Range = $Q_3 – Q_1$ or 75th percentile – 25th percentile

  $$\textit{Semi interquartile range} = \frac{Q_3 - Q_1}{2}$$

  Where:

  Q_3 is the third quartile (upper quartile) and

  Q_1 is the first quartile (lower quartile).

- *For example:*

Monthly Salary (Rs 000)	Number of people (f)	Cumulative frequency distribution
5 and under 10	2	2
10 and under 15	15	17
15 and under 20	18	35
20 and under 25	12	47
25 and under 30	2	49
30 and over	1	50

The ogive is as follows:

There are 50 items so Q_1 is the value of the 12.5th item and Q_3 is the value of the 37.5th item.

These can be read off (very approximately) as Q1 = 13.5 and Q3 = 21

$$\text{Semi interquartile range} = \frac{Q_3 - Q_1}{2} = \frac{21 - 13.5}{2} = 3.75$$

Variance:

Another measure of variability that takes into account for the whole dataset is variance.

The deviation from the mean and the observations are calculated. Deviations can be positive or negative depending upon observations above and below mean respectively. The average of the squared deviations is called **variance**.

- *Formula:*

For Ungrouped Data:

$$\text{Variance } (\sigma 2) \ or \ s^2 = \frac{\sum(x-\bar{x})^2}{\text{n}}$$

Where:

n = number of items in the sample

x = variable or observations

$\bar{x}$= observation mean

For Grouped Data:

$$s^2 = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$$

Where:

x = mid-point of a class

f = frequency of a class

- *For example:*

Find Variance for the given dataset: 41, 68, 40, 28, 22, 41, 30, 50

n = 8

x	$(x-\bar{x})$	$(x-\bar{x})^2$
41	1	1
68	28	784
40	0	0
28	-12	144
22	-18	324
41	1	1
30	-10	100
50	10	100
320	0	1,454

$$\bar{x} = \frac{\sum x}{\text{n}} = \frac{320}{8} = 40$$

$$s^2 = \frac{\sum(x-\bar{x})^2}{\text{n}} = \frac{1{,}454}{8} = 181.75$$

Find Variance for the given distribution:

Weekly sales (Units)	Number of weeks
0 to under 100	8
100 to under 200	17
200 to under 300	9
300 to under 400	5
400 to under 500	1

Weekly sales (Units)	Mid-point (x)	Number of weeks (f)	fx	fx^2
0 to under 100	50	8	400	20,000
100 to under 200	150	17	2,550	382,500
200 to under 300	250	9	2,250	562,500
300 to under 400	350	5	1,750	612,500
400 to under 500	450	1	450	202,500
	1,250	40	7,400	1,780,000

$$\bar{x} = \frac{\sum \mathrm{f}x}{\sum \mathrm{f}} = \frac{7{,}400}{40} = 185$$

$$\mathrm{s}^2 = \frac{\sum \mathrm{f}x^2}{\sum \mathrm{f}} - \left(\frac{\sum \mathrm{f}x}{\sum \mathrm{f}}\right)^2$$

$$\mathrm{s}^2 = \frac{1{,}780{,}000}{40} - \left(\frac{7{,}400}{40}\right)^2$$

$$\mathrm{s}^2 = 44{,}500 - (185)^2$$

$$\mathrm{s}^2 = 44{,}500 - 34{,}225$$

$$\mathrm{s}^2 = 10{,}275$$

Standard Deviation

The square root of variance is called standard deviation.

The measures of variation are around the mean and are used with arithmetic mean as a pair of summary measures to describe distribution of a dataset.

- *Formula:*

For Ungrouped Data:

$$\mathrm{s} = \sqrt{\frac{\sum(x-\bar{x})^2}{\mathrm{n}}}$$

For Grouped Data:

$$s = \sqrt{\frac{\sum f x^2}{\sum f} - \left(\frac{\sum f x}{\sum f}\right)^2}$$

AT A GLANCE

SPOTLIGHT

STICKY NOTES

- *For example:*

Standard deviation of six numbers: 50, 51, 48, 51, 49, 51 can be found by:

x	$(x-\overline{x})$	$(x-\overline{x})^2$
50	0	0
51	1	1
48	-2	4
51	1	1
49	-1	1
51	1	1
300	0	8

n = 6

$$\bar{x} = \frac{\sum x}{\text{n}} = \frac{300}{6} = 50$$

$$\text{s} = \sqrt{\frac{\sum(x-\bar{x})^2}{\text{n}}} = \sqrt{\frac{8}{6}} = 1.154$$

Mean and standard deviation of the following distribution can be found by:

Number of goals scored per hockey match (x)	Number of matches (f)	fx	fx^2
0	8	0	0
1	7	7	7
2	3	6	12
3	1	3	9
4	1	4	16
	20	20	44

$$\bar{x} = \frac{\sum \text{f}x}{\sum \text{f}} = \frac{20}{20} = 1$$

$$\text{s} = \sqrt{\frac{\sum \text{f}x^2}{\sum \text{f}} - \left(\frac{\sum \text{f}x}{\sum \text{f}}\right)^2} = \sqrt{\frac{44}{20} - \left(\frac{20}{20}\right)^2} = 1{,}095 \text{ goals}$$

Coefficient of variation

The coefficient of variation is a measure of relative dispersion that expresses standard deviation in terms of the mean. The aim of this measure is to allow comparison of the variability of two sets of figures.

- *Formula:*

$$\text{Coefficient of variation} = \frac{s}{\bar{x}} \times 100$$

- *For example:*

A class of children sat 2 exams (one was marked out of 20 and the other was marked out of 100). The results were as follows:

	Test 1	Test 2
Maximum possible mark	20	100
Mean ($\bar{x}$)	12	64
Standard deviation (σ)	3	10

Coefficient of variation

$$\text{Test 1:} = \frac{s}{\bar{x}} \times 100 = \frac{3}{12} \times 100 = 25\%$$

$$\text{Test 2:} = \frac{s}{\bar{x}} \times 100 = \frac{10}{64} \times 100 = 15.625\%$$

Test 1 results have the highest relative dispersion.

AT A GLANCE
SPOTLIGHT
STICKY NOTES

3 MEASURES COMPARED

Measure	Advantages	Disadvantages
Mean	Uses every value in the distribution.	Can be distorted by extreme values when samples are small.
	Used extensively in other statistical applications.	The mean of discrete data will usually result in a figure that is not in the data set.
Median	Not influenced by extreme values.	Little use in other applications.
	Is an actual value (as long as there is an odd number of observations).	
Mode	Is an actual value.	Little use in other applications.
Range	Easy to calculate and understand	Distorted by extreme figures
Semi inter-quartile deviation	Easy to calculate and understand	Does not take all figures into account.
		Does not give any insight into the degree of clustering
Standard deviation	Uses all values	Difficult to understand
	Very useful in further applications	Difficult to calculate (but perhaps this is no longer true due to programmes like excel)
Variance	Can be used to find the standard deviation of a distribution	
	Useful in further applications	

Skewness and measures of central tendency:

Skewness with respect to the shape of distribution was discussed in the earlier section. In comparing the measures, generally, if the shape is

- perfectly symmetric, the mean equals the median.
- skewed to the left/negative, the mean is smaller than the median.
- skewed to the right/positive, the mean is larger than the median.

STICKY NOTES

The median is the middle value when the data is arranged in ascending or descending order.

***For ungrouped data*: Position of the median=(n+1)/2**

For Grouped data: $\boldsymbol{median = L + \left[\frac{\left(\frac{\sum f}{2} - cf\right)}{f}\right] \times C}$

Mode is the most frequently occurring value (or the highest frequency in grouped data). There may be more than one mode, or there may be no mode at all.

Formula for grouped data: $\boldsymbol{mode = L + \left[\frac{(f_m - f_1)}{(f_m - f_1) + (f_m - f_2)}\right] \times C}$

The arithmetic mean is a mathematical representation of the typical value of a series of numbers. It is computed as the sum of the numbers in the series divided by the count of numbers in that series.

For Ungrouped data: $\bar{x} = \frac{\sum x}{n}$

For Grouped data: $\bar{x} = \frac{\sum fx}{\sum f}$

The range is simply the difference between the highest and lowest value in a distribution.

The deviation from the mean and the observations are calculated. Deviations can be positive or negative depending upon observations above and below mean respectively. The average of the squared deviations is called variance.

For Ungrouped Data: Variance (σ2) *or* $s^2 = \frac{\sum (x - \bar{x})^2}{n}$

For Grouped Data: $s^2 = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$

The square root of variance is called standard deviation.

For Ungrouped Data: $s = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$

For Grouped Data: $s = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$

SELF-TEST

8.1. Any numerical value describing a characteristic of a sample is called:

(a) Sample mean (b) Sample variance

(c) Statistic (d) None of these

8.2. If each observation is decreased by 2, the mean of new set of observations will:

(a) Increase by 2 (b) Decrease by 2

(c) Remain unchanged (d) None of these

8.3. If each observation is increased by 5, the mean of new set of observations will:

(a) Increase by 5 (b) Decease by 5

(c) Remain unchanged (d) None of these

8.4. If each observation is multiplied by 5, the mean of new set of observations will:

(a) Increase by 5 times (b) Divided by 5 times

(c) Remain unaltered (d) None of these

8.5. If each observation is divided by 10, the mean of new set of observations will:

(a) Decrease by 10 times (b) increase by 10 times

(c) Remain unchanged (d) None of these

8.6. The middle most value of arranged set of observations is known as:

(a) Mode (b) Median

(c) Mean (d) None of these

8.7. The unit of measurement of coefficient of variation of speed of balls bowled by a bowler is:

(a) Feet per second (b) Yards per second

(c) Neither (a) nor (b) (d) Both (a) as well as (b)

8.8. A family travels 500 kms each day for 3 days. The family averages 80 kms per hour the first day, 93 kms per hour the second day, 87 kms per hour the third day. The average speed for the entire trip is:

(a) 87 kms \Hr (b) 85 kms \Hr

(c) 86.7 kms \Hr (d) None of these

8.9. The average preferred, to such data as rates of change in ratios, economic index numbers, population sizes over consecutive time periods, is:

AT A GLANCE
SPOTLIGHT
STICKY NOTES

(a) Arithmetic mean (b) Geometric mean

(c) Harmonic mean (d) None of these

8.10. Over a period of 4 years, an employee's salary has increased in the ratios, 1.072, 1.086, 1.069 and 1.098. The average of these ratios and hence the average percent increase are:

(a) 1.08125 and 8.125% (b) 1.086 and 8.6%

(c) 1.08119 and 8.6% (d) None of these

8.11. Median of 82, 86, 93, 92, 79 is:

(a) 86 (b) 93

(c) $\frac{93+92}{2}$ (d) None of these

8.12. Median of 2.5, 3.6, 3.1, 4.3, 2.9, 2.3 is:

(a) Does no exist (b) 2.9

(c) 3 (d) None of these

8.13. The value which occurs most or the value with the greatest frequency is called:

(a) Mode (b) Q_3

(c) Q_1 (d) None of these

8.14. Mode of 2.5, 3.6, 3.1, 4.3, 2.9, 2.3 is:

(a) 4.3 (b) $\frac{3.1+4.3}{2}$

(c) Does not exist (d) None of these

8.15. The real __________ of the mean is that it may be affected by extreme values.

(a) Advantage (b) Disadvantage

(c) Choice (d) None of these

8.16. The median is ______ by extreme values and gives a truer average.

(a) Not influenced (b) Influenced

(c) Affected (d) None of these

8.17. It is the only average that can be used for qualitative as well as quantitative data:

(a) Mean (b) Mode

(c) Median (d) None of these

8.18. A car averages 20 kilometres per litre on the highway. How many litres are required to complete 300 kilometre trip?

(a)	60	(b)	25
(c)	15	(d)	None of these

8.19. A person invests Rs. 5,000/= at 10.5% interest, Rs. 6,300/= at 10.8% and Rs. 4,500 at 11%. What is the average percentage return to the saving?

(a)	10.762%	(b)	10.8%
(c)	10.766%	(d)	10.666%

8.20. What is the average for a student who received grades of 85, 76 and 82 on 3 mid term tests and a 79 on the final examination, if the final examination counts 3 times as much as each of the three tests?

(a)	81	(b)	80
(c)	80.5	(d)	None of these

8.21. Three sections of a statistics class containing 28, 32 and 35 students averaged 83, 80 and 76 respectively. What is the average of all 3 sections?

(a)	79.41	(b)	79.67
(c)	80	(d)	None of these

8.22. The sum of deviations from a mean is always:

(a)	Zero	(b)	Less than mean
(c)	Greater than mean	(d)	None of these

8.23. If $\sum(x-7)=0$, then $\bar{x}$ = ________ .

(a)	8	(b)	7
(c)	5	(d)	None of these

8.24. The sum of square of deviations from mean is always:

(a)	More than zero	(b)	Zero
(c)	Less than zero	(d)	Less than Mean

8.25. The average daily sale and the related standard deviation of Ali, Atif, Ahmed and Azeem in thousands of Rupees are 41 & 4.3, 36 & 3.2, 26 & 2.9 and 24 & 2.5 respectively, then the most inconsistent among them is:

(a)	Ali	(b)	Atif
(c)	Ahmed	(d)	Azeem

8.26. If for two observations, deviations from mean are -3 and +3 respectively, then the variance is:

(a)	0	(b)	Does not exist
(c)	9	(d)	None of these

8.27. The positive square root of variance is called:

(a)	Mean deviation	(b)	Standard deviation
(c)	Mean deviation from mean	(d)	None of these

8.28. The co-efficient of variation of 3 numbers i.e. x, x+4 and x+11 is y. If x is increased, the value of y would:

(a)	Decrease	(b)	Increase with the same amount
(c)	Increase in the same ratio	(d)	Remain the same

8.29. For the observations 7, 7, 7, mean and variance are respectively:

(a)	0 and 7	(b)	7 and 0
(c)	7 and 49	(d)	None of these

8.30. If the variance of 2, 5, 7, 10 and 14 is 17.04 then the variance of 5, 8, 10, 13 and 17 is:

(a)	17.04	(b)	20.04
(c)	51.12	(d)	None of these

8.31. If the variance of 2, 5, 7, 10 and 14 is 17.04 then the variance of 6, 15, 21, 30 and 42 is:

(a)	17.04	(b)	51.12
(c)	153.36	(d)	None of these

8.32. If some constant is added to each observation of a given set of data then the variance will:

(a)	Increase by that constant	(b)	Decrease by that constant
(c)	Remain unchanged	(d)	None of these

8.33. If some constant is subtracted from each observation of the given set of data, variance will:

(a)	Decrease by the that constant	(b)	Increase by that constant
(c)	Remain unchanged	(d)	None of these

8.34. If each observation of the given set of data is multiplied by a constant x, then the variance will:

(a)	Increase by x times	(b)	Decrease by x times

(c) Increase by x^2 times (d) None of these

8.35. If each observation of the given set of data is divided by a constant x, then the variance will:

(a) Decrease by x times (b) Increase by x times

(c) Decrease by x^2 times (d) None of these

8.36. For the values -2, -5, -7, -10 and -14 the variance is:

(a) -17.04 (b) +17.04

(c) -8.54 (d) None of these

8.37. If the variance of a given set of observations is 81, the standard deviation is:

(a) ± 9 (b) +9

(c) -9 (d) None of these

8.38. The 50th percentile, fifth decile, second quartile and median of a distribution are:

(a) Unequal (b) Equal

(c) Approximately equal (d) None of these

8.39. The measure of variation which measures the length of interval that contains the middle 50% of data is known as:

(a) Quartile-deviation (b) Semi-inter quartile range

(c) Inter-quartile range (d) None of these

8.40. Statistical measures which define the centre of a set of data are called:

(a) Median (b) Measures of data

(c) Measures of central tendency (d) None of these

8.41. Statistical measures which provide a measure of variability among the observations are called:

(a) Quartiles (b) Measures of dispersion

(c) Co-efficient of variation (d) None of these

8.42. Arithmetic mean for a set of data with six observations was calculated as 30, however during data input an observation valuing 31 was considered 13 by mistake. Compute correct value of mean

(a) 30 (b) 33

(c) 36 (d) 39

8.43. Which of the following is/are correct?

(1) The formula of arithmetic mean is: sum of the observations / total number of observations.

(2) The algebraic sum of the deviations of the observations from their mean is always zero.

(3) Harmonic mean is not used in averaging speeds.

8.44. Select one or more of the possible correct answers:

(1) Mode is that value which divides the sets of observations into two equal parts.

(2) Sum of squares of deviations of observations from their mean is minimum.

8.45. Select one or more of the possible correct answers:

(1) Coefficient of variation is independent of units and expressed in percentages

(2) Coefficient of variation is an absolute measure of dispersion.

(3) Coefficient of variation is a relative measure of dispersion.

(4) Coefficient of variation cannot be calculated when standard deviation and mean are available.

8.46. Select one or more of the possible correct answers:

(1) Variance and standard deviation cannot be negative.

(2) Variance and standard deviation cannot be zero.

(3) Variance and standard deviation can be zero.

(4) Variance and standard deviation does not change if each value in data is multiplied by a constant.

8.47. Select one or more of the possible correct answers:

(1) Range is the difference of highest and lowest observation in the data.

(2) Standard deviation is the positive square root of variance.

(3) Arithmetic mean of 7 and 9 is 8.5

8.48. Select one or more of the possible correct answers:

(1) The weighted arithmetic mean is computed by taking into account the relative importance of each item.

(2) The geometric mean is useful to work out the average growth rate over several years.

(3) Coefficient of variation is an absolute measure of dispersion.

8.49. For the following data compute mean

Number of goals	Number of matches (frequency)
0	5
1	3
2	2
3	1

(a) 2.1 (b) 0.81

(c) 1.1 (d) 0.91

8.50. For the following data compute standard deviation

Number of goals	Number of matches (frequency)
0	5
1	3
2	2
3	1

(a) 0.9959 (b) 0.9121
(c) 0.8812 (d) 0.9359

8.51. A person wants to calculate average growth rate in his property value over past few years. Which method of average is suitable

(a) Arithmetic mean (b) Geometric mean
(c) Harmonic mean (d) None of these

8.52. Compute coefficient of variation for following data:
Mean: 80
Standard deviation: 20

(a) 400% (b) 20%
(c) 80% (d) 25%

8.53. Arithmetic mean for a set of data with four observations is 50. Which of the following cannot be a possible value of observation?

(a) 190 (b) 200
(c) 50 (d) 250

8.54. Arithmetic mean for a set of data with four observations is 50. What can be the maximum possible value a single observation?

(a) 200 (b) 50
(c) 250 (d) 400

8.55. Arithmetic mean of three observations is 12. The observations are such that they have a common difference of 2 between them. Compute median value.

(a) 12 (b) 10
(c) 14 (d) 36

8.56. Arithmetic mean of three observations is 15. The observations are such that they have a common difference of 3 between them. If Unity is subtracted from mid value of these three observations find the mean after changes

(a)	15	(b)	14.67
(c)	16	(d)	18

8.57. If the standard deviation of a and b, two observations is 2.5 then the standard deviation of a+3 and b+3 is:

(a)	2.5	(b)	5.5
(c)	0	(d)	Cannot be determined with limited information

8.58. If the standard deviation of a and b, two observations is 2.5 then the standard deviation of 3a and 3b is:

(a)	5.5	(b)	7.5
(c)	0	(d)	Cannot be determined with limited information

8.59. A combined mean is simply a _____, where the weights are the size of each group.

(a)	Weighted	(b)	Geometric
(c)	Harmonic	(d)	None of the options are valid

8.60. A class of 20 students had mean marks of 25. If there are 18 boys in the class with mean marks of 20. What can be the highest possible marks of a single girl student?

(a)	140	(b)	150
(c)	20	(d)	70

8.61. A class of 20 students had mean marks of 25. If there are 18 boys in the class with mean marks of 20. Compute total marks scored by the two girls.

(a)	140	(b)	70
(c)	150	(d)	180

ANSWERS TO SELF-TEST QUESTIONS

8.1	**8.2**	**8.3**	**8.4**	**8.5**	**8.6**
(c)	(b)	(a)	(a)	(a)	(b)
8.7	**8.8**	**8.9**	**8.10**	**8.11**	**8.12**
(c)	(c)	(b)	(a)	(a)	(c)
8.13	**8.14**	**8.15**	**8.16**	**8.17**	**8.18**
(a)	(c)	(b)	(a)	(b)	(c)
8.19	**8.20**	**8.21**	**8.22**	**8.23**	**8.24**
(a)	(b)	(a)	(a)	(b)	(a)
8.25	**8.26**	**8.27**	**8.28**	**8.29**	**8.30**
(c)	(c)	(b)	(a)	(b)	(a)
8.31	**8.32**	**8.33**	**8.34**	**8.35**	**8.36**
(c)	(c)	(c)	(c)	(c)	(b)
8.37	**8.38**	**8.39**	**8.40**	**8.41**	**8.42**
(b)	(b)	(c)	(c)	(b)	(b)
8.43	**8.44**	**8.45**	**8.46**	**8.47**	**8.48**
1,2	2	1,3	1,3	1,2	1,2
8.49	**8.50**	**8.51**	**8.52**	**8.53**	**8.54**
(d)	(a)	(b)	(d)	(d)	(a)
8.55	**8.56**	**8.57**	**8.58**	**8.59**	**8.60**
(a)	(b)	(a)	(b)	(a)	(a)
8.61					
(a)					

CHAPTER 9

INDICES

IN THIS CHAPTER:

AT A GLANCE

Index numbers (indices) are used to measure changes over time, location or some other characteristics. These changes usually relate to price or quantity.

Index numbers might be constructed in a relatively straightforward way by taking a list of values and choosing one as a base against which other values are compared. Alternatively, index numbers might be constructed in a complexed way, combining the price and quantity of many items. The retail price index is constructed in this way.

Index numbers can be used to judge the effects of changes, such as price changes, in the past. They can also be used to predict the future, by forecasting what the index number will be at a future date.

Various indices are explained in this chapter including Laspeyre and Paasche indices.

AT A GLANCE
SPOTLIGHT
STICKY NOTES

1. INDEX NUMBERS

Index numbers are statistical devices that are used to express the relationship between quantitative variables. They can be used as measures of fluctuations within the economy or production.

▶ *For example:*

Consumer price indices (or Retail Price Indices – RPI) measure changes in price levels over time; and

Stock Market Indices measure changes in a group of share prices over time (for example, the S&P 500 index measures changes in the prices of the top 500 shares traded in the USA).

Uses of Index numbers:

- At economic levels index numbers can be used to get feel of the economic tendencies, inflation and deflation. Government might choose the contents of the basket to reflect typical purchases in a society in order for the price index to provide a measure of inflation of relevance to government planning.
- In formulating policies, indices can help vouch for any production or trade fluctuations
- It can also help in exploring the trends prevailing at the market.
- It can also be helpful in determining purchasing powers as well as real incomes of the people.

Base Number and period:

Indices are calculated with reference to a base number which is usually given a value of 100 or 1,000. All other numbers are calculated in relation to the base.

A Base period number can be a relative period against which all other variables are compared. Selecting the base period is critical. Ideal base period should not be too far but should reflect stablity within the given variables.

▶ *For example:*

In Year 1 a base of 100 was set for a price index. This represents price levels at this time.

The price index at Year 3 is 112.5 and Year 4 it is 118.1.

Price increase between:	
Year 1 and Year 3	$\left(\frac{112.5}{100.0}\right) - 1 = 0.125 \text{ or } 12.5\%$
Year 1 and Year 4	$\left(\frac{118.1}{100.0}\right) - 1 = 0.181 \text{ or } 18.1\%$
Year 3 and Year 4	$\left(\frac{118.1}{112.5}\right) - 1 = 0.05 \text{ or } 5.0\%$

Types of Indices:

A price index is an index that measures changes in prices of a group of items over time.

A quantity index measures changes over time in the quantities of items in a ‘basket’. The price levels applied to the baskets at different points in time are kept constant so that the index only reflects movements in the quantity of goods and services purchased.

Constructing index numbers

There are following ways to construct index numbers

Simple index numbers

Where we are interested in information about a single item the calculation of the index numbers is straightforward. One of the period is taken as a base period, and price or quantity of each item is taken as the percentage of the price or quantity of the base period.

At times, index is taken as the average of the percentage.

- *Formula:*

Simple price index (known as the price relative)	**Simple quantity index (known as the quantity relative)**
Price index $= \frac{p_1}{p_0} \times 100$	Quantity index $= \frac{q_1}{q_0} \times 100$
Where:	
p_0 = prices in the base period	q_0 = quantities in the base period
p_1 = prices in the current period	q_1 = quantities in the current period

- *For example:*

The following information relates to the price and quantity of coffee machines sold over a five year period.

	Price (Rs.)	Quantity
2011	6,500	12,178
2012	6,800	13,493
2013	6,900	15,149
2014	7,200	16,287
2015	7,500	17,101

Considering 2013 as the base year:

Simple price (price relative)		$\frac{p_1}{p_0} \times 100$	Index number
2011	6,500	$\frac{6,500}{6,900} \times 100$	94.20
2012	6,800	$\frac{6,800}{6,900} \times 100$	98.55
2013	6,900	$\frac{6,900}{6,900} \times 100$	100
2014	7,200	$\frac{7,200}{6,900} \times 100$	104.35
2015	7,500	$\frac{7,500}{6,900} \times 100$	108.69

Simple quantity (quantity relative)		$\frac{q_1}{q_0} \times 100$	Index
2011	12,178	$\frac{12,178}{15,149} \times 100$	80.38
2012	13,493	$\frac{13,493}{15,149} \times 100$	89.06
2013	15,149	$\frac{15,149}{15,149} \times 100$	100
2014	16,287	$\frac{16,287}{15,149} \times 100$	107.51
2015	17,101	$\frac{17,101}{15,149} \times 100$	112.88

Sometimes an index is adjusted to change its base year. The new index numbers are constructed by dividing the existing index number in each year by the existing index number for new base year and multiplying the result by 100. Alternatively, the price or quantity of the period is divided by the price or quantity of the base year, multiplying the result by 100.

- *For example:*

Consider the base year as 2011

Simple price (price relative)		Index number (2013)	$\frac{p_1}{p_0} \times 100$	Index number (2011)
2011	6,500	94.2	$\frac{6,500}{6,500} \times 100$ or $\frac{94.2}{94.2} \times 100$	100
2012	6,800	98.5	$\frac{6,800}{6,500} \times 100$ or $\frac{98.55}{94.2} \times 100$	104.61
2013	6,900	100	$\frac{6,900}{6,500} \times 100$ or $\frac{100}{94.2} \times 100$	106.15
2014	7,200	104.3	$\frac{7,200}{6,500} \times 100$ or $\frac{104.35}{94.2} \times 100$	110.76
2015	7,500	108.6	$\frac{7,500}{6,500} \times 100$ or $\frac{108.70}{94.2} \times 100$	115.38

Note that this does not change the story told by the index numbers. It simply tells it in a different way.

Index number with more than one item

Consider price index numbers where there is more than one item.

The aim is to construct a figure (index number) to compare costs (or quantities) in a year under consideration to a base year.

One way of doing this would be to calculate the **simple aggregate price index**

- *Formula:*

$$\text{Simple aggregate price index} = \frac{\sum p_1}{\sum p_0} \times 100$$

Where

$\sum p_0$ = sum of prices in the base period

$\sum p_1$ = sum of prices in the current period

- *For example:*

A manufacturing company makes an item which contains four materials.

The prices of each material were as follows in the base period, January Year 1, and the most recent period, December Year 2

	January Year 1 (base period)	December Year 2 (Current period)
Material	**Price per kilo**	**Price per kilo**
	p_0	p_1
A	0.5	0.75
B	2.0	2.1
C	4.0	4.5
D	1.0	1.1
	7.5	8.45

$$\text{Simple aggregate price index} = \frac{\sum p_1}{\sum p_0} \times 100 = \frac{8.45}{7.5} \times 100 = 113$$

The index indicates that prices have increased from a base of 100 to 113, an increase of 13%.

Limitation for this approach is that it gives no indication of the relative importance of each material. The price of material A has increased from 0.5 to 0.75. This is an increase of 50%.

If each unit of production used 1 kg of B, C and D but 100mg of A, the simple aggregate price index would not capture the serious impact that the price increase has had on the cost base.

A more sophisticated method is needed to take the relative importance of items into account.

This is very important for governments who need to collect information about inflation. To do this they must construct a price index to measure changes in the prices of a group of items over time.

The group of items is often described as a basket of goods (and services).

A government must decide what goods and services to include in the group how much of these goods and services to reflect typical purchases in a society in order for the price index to provide a measure of inflation of relevance to government planning.

2. WEIGHTED INDICES

The index where relative weights are assigned as per the relative significance of each item in the 'basket'.

There are different ways of weighting an index. The two most common types of indices are:

- Laspeyre indices; and
- Paasche indices.

For price index numbers, either the original quantities or the current quantities of items in the 'basket' can be used.

Laspeyre price index

The Laspeyre price index measures price changes with reference to the quantities of goods in the basket at the date that the index was first established. That is, base year quantities are taken as weights.

This is the usual form of a weighted price index.

► *Formula:*

$$\frac{\sum(p_1 \times q_0)}{\sum(p_0 \times q_0)} \times 100$$

Where:

p_0 = prices in the base period for the index

p_1 = prices in the current period for which an index value is being calculated

q_0 = the original quantities in the base index.

► *For example:*

A manufacturing company makes an item which contains four materials.

The price of each material, and the quantities required to make one unit of finished product, were as follows in the base period, January Year 1, and the most recent period, December Year 2

	January Year 1 (Base period)		December Year 2 (Current period)			
Material	**Price per kilo**	**Kilos per unit**	**Price per kilo**	**Kilos per unit**		
	p_0	q_0	p_1	q_1	$p_0 q_0$	$p_1 q_0$
A	0.5	10	0.75	11.0	5.0	7.5
B	2.0	3	2.1	2.5	6.0	6.3
C	4.0	2	4.5	3.0	8.0	9.0
D	1.0	2	1.1	5.0	2.0	2.2
					21.0	**25.0**

Laspeyre price index as at December Year 2:

$$\frac{\sum(p_1 \times q_0)}{\sum(p_0 \times q_0)} \times 100$$

$$\frac{25.0}{21.0} \times 100 = 119.04$$

The Laspeyre index shows that prices have risen by 19.04% between January Year 1 and December Year 2.

Paasche price index:

The Paasche price index measures price changes with reference to current quantities of goods in the basket.

- *Formula:*

$$\frac{\sum(p_1 \times q_1)}{\sum(p_0 \times q_1)} \times 100$$

Where:

p_0 = prices in the base period for the index

p_1 = prices in the current period for which an index value is being calculated

q_1 = quantities in the current period for which an index value is being calculated.

- *For example:*

A manufacturing company makes an item which contains four materials.

The price of each material, and the quantities required to make one unit of finished product, were as follows in the base period, January Year 1, and the most recent period, December Year 2

	January Year 1 (Base period)		December Year 2 (Current period)			
Material	**Price per kilo**	**Kilos per unit**	**Price per kilo**	**Kilos per unit**		
	p_0	q_0	p_1	q_1	$p_0\,q_1$	$p_1\,q_1$
A	0.5	10	0.75	11.0	5.50	8.25
B	2.0	3	2.1	2.5	5.00	5.25
C	4.0	2	4.5	3.0	12.00	13.50
D	1.0	2	1.1	5.0	5.00	5.50
					27.50	**32.50**

Paasche price index as at December Year 2:

$$\frac{\sum(p_1 \times q_1)}{\sum(p_0 \times q_1)} \times 100$$

$$\frac{32.5}{27.5} \times 100 = 118.18$$

The Paasche index shows that prices have risen by 18.18% between January Year 1 and December Year 2.

Limitations of Laspeyre and Paache prince indices:

- **The denominator:** In the calculation of the Laspeyre price index (p_0q_0), the denominator does not change from year to year. The only new information that has to be collected each year is the prices of items in the index.

  The denominator in the Paasche price index (p_0q_1) has to be recalculated every year to take account of the most recent quantities consumed. This information might be difficult to collect.

  In summary more information has to be collected to construct the Paasche price index than to construct the Laspeyre price index. This helps to explain why the Laspeyre price index is used more than the Paasche price index in practice.

- **Inflation:** Laspeyre price index tends to overstate inflation whereas the Paasche price index tends to understate it. This is because consumers react to price increases by changing what they buy.

  The Laspeyre index which is based on quantities bought in the base year fails to account for the fact that people will buy less of those items which have risen in price more than others. These are retained in the index with the same weighting even though the volume of consumption has fallen.

  The Paasche index is based on the most recent quantities purchased. This means that it has a focus which is biased to the cheaper items bought by consumers as a result of inflation.

Fisher index

The Fisher index is also known as the ideal price index.

The Fisher index is a price index, computed for a given period by taking the square root of the product of the Paasche index value and the Laspeyre index value. The result is that the Fisher price index measures price change on the basis of the baskets from both the base and the current period.

It is said to be the geometric mean of the Paasche and Laspeyre indices.

- *Formula:*

$$= \sqrt{\frac{\Sigma(p_1 \times q_0)}{\Sigma(p_0 \times q_0)} \times \frac{\Sigma(p_1 \times q_1)}{\Sigma(p_0 \times q_1)}} \times 100$$

Or

$$= \sqrt{P_L \times P_P}$$

- *For example:*

Revisiting the previous example.

Where

P_L = Laspeyre price index = 119.04

P_P = Paasche price index = 118.18

$$= \sqrt{P_L \times P_P} = \sqrt{119.04 \times 118.18} = 118.60$$

Laspeyre quantity index

Laspeyre index uses the price levels at the base period date.

- *Formula:*

$$\frac{\Sigma(p_0 \times q_1)}{\Sigma(p_0 \times q_0)} \times 100$$

Where:

p_0 = prices in the base period for the index

q_0 = the original quantities in the base index.

q_1 = quantities used currently.

- *For example:*

Using the earlier example:

	January Year 1 (Base period)		December Year 2 (Current period)			
Material	**Price per kilo**	**Kilos per unit**	**Price per kilo**	**Kilos per unit**		
	p_0	q_0	p_1	q_1	$p_0 q_0$	$p_0 q_1$
A	0.5	10	0.75	11.0	5.0	5.50
B	2.0	3	2.1	2.5	6.0	5.00
C	4.0	2	4.5	3.0	8.0	12.00
D	1.0	2	1.1	5.0	2.0	5.00
					21.0	**27.50**

Laspeyre quantity index as at December Year 2:

$$\frac{\sum(p_0 \times q_1)}{\sum(p_0 \times q_0)} \times 100$$

$$\frac{27.5}{21.0} \times 100 = 130.95$$

The Laspeyre index shows that there has been an increase of 30.95% in the quantities used between January Year 1 and December Year 2.

Paasche quantity index

A Paasche index uses the current price levels.

- *Formula:*

$$\frac{\sum(p_1 \times q_1)}{\sum(p_1 \times q_0)} \times 100$$

Where:

p_1 = prices in the current period for which an index value is being calculated

q_0 = the original quantities in the base index.

q_1 = quantities used currently or quantities in the current period for which an index value is being calculated.

- *For example:*

Using the earlier example:

	January Year 1 (Base period)		December Year 2 (Current period)			
Material	**Price per kilo**	**Kilos per unit**	**Price per kilo**	**Kilos per unit**		
	p_0	q_0	p_1	q_1	$p_1 q_0$	$p_1 q_1$
A	0.5	10	0.75	11.0	7.5	8.25
B	2.0	3	2.1	2.5	6.3	5.25
C	4.0	2	4.5	3.0	9.0	13.50
D	1.0	2	1.1	5.0	2.2	5.50
					25.0	**32.50**

AT A GLANCE
SPOTLIGHT
STICKY NOTES

Paasche quantity index as at December Year 2:

$$\frac{\sum(p_1 \times q_1)}{\sum(p_1 \times q_0)} \times 100$$

$$\frac{32.5}{25.0} \times 100 = 130.0$$

The Paasche index shows that there has been an increase of 30% in the quantities used between January Year 1 and December Year 2.

Chained index series

Another way of calculating an index is to calculate a chained index series.

A chained index series is an index series where each new value in the index is calculated based on the previous years.

This type of index can be used to monitor trends in the past and use past trends to prepare a forecast for the future.

- *Formula:*

$$\textit{Link Relative} = \frac{\text{Value in the current period}}{\text{Value in the previous period}} \times 100$$

$$\textit{Chain Index} = \frac{\textit{Relative Index of current year} \times \textit{Chain Index of previous year}}{100}$$

- *For example:*

A company has achieved the following sales over the past five years.

Year	Sales	Relative	Index	Chain index
2009	1,760	$\frac{1,760}{1,760} \times 100$	100.0	100
2010	1,883	$\frac{1,883}{1,760} \times 100$	107.0	$\frac{107.0 \times 100}{100} = 107$
2011	2,024	$\frac{2,024}{1,883} \times 100$	107.5	$\frac{107.5 \times 107.0}{100} = 115.02$
2012	2,166	$\frac{2,166}{2,024} \times 100$	107.0	$\frac{107.0 \times 115.02}{100} = 123.07$
2013	2,320	$\frac{2,320}{2,166} \times 100$	107.1	$\frac{107.1 \times 123.07}{100} = 131.80$

The chained index shows that annual sales growth has been about 7% per year, compound.

Consumer Price Index

Consumer price index (CPI), measures the cost of a consumer basket (goods and services) purchased by a "typical" urban family at a point in time as compared to the base year. The basket may include items of food, clothing, housing, fuels, transportation, and medical care[1].

CPI is used to calculate the purchasing power or cost of living of the concerned group of people or class. It can also help identify periods of inflation and deflation.

[1] Lee, F. C, Lee, J. C., & Lee, A. C. (2013). Statistics for Business and Financial Economics. Third Edition. Springer.

- *Formula:*

CPI for a single item can be found by:

$$\text{CPI} = \frac{\text{Cost of market Basket in current year}}{\text{Cost of market Basket in base year}} \times 100$$

- *For example:*

A household has following basket of commodities

	Prices of commodities			
Commodity	**2009**	**2010**	**2011**	**2012**
Eggs	1.90	2.50	2.11	2.00
Milk	1.00	1.50	1.10	1.30
Bread	0.90	0.95	0.90	0.90
Butter	0.55	0.60	0.65	0.65
Meat	1.20	1.12	1.50	1.60
	5.55	**6.67**	**6.26**	**6.45**

Considering base year as 2009

Year	Cost of Basket		CPI
2009	5.55	$\frac{5.55}{5.55} \times 100$	100
2010	6.67	$\frac{6.67}{5.55} \times 100$	120.18
2011	6.26	$\frac{6.26}{5.55} \times 100$	112.79
2012	6.45	$\frac{6.45}{5.55} \times 100$	116.21

Calculating the rate of inflation/deflation:

The percentage change in the CPI can be used to determine rate of inflation/deflation. If the CPI increases there is inflation and if the CPI decreases there is deflation.

- *Formula:*

$$Inflation\ Rate = \frac{CPI_2 - CPI_1}{CPI_1} \times 100$$

- *For example:*

For the above example, rate of inflation /deflation can be calculated as

Year	Change	Inflation/ Deflation
2009 to 2010	$\frac{120.18 - 100}{100} \times 100$	20.18%
2010 to 2011	$\frac{112.79 - 120.18}{120.18} \times 100$	-6.15%
2011 to 2012	$\frac{116.21 - 112.79}{112.79} \times 100$	3.03%

Nominal and Real Variables:

Effects of inflation and deflation may change the real prices or values. Actual value, a consumer or bank would get, would be determined by excluding the effects of inflation or deflation.

- *For example:*

  If the bank charges interest rate of 10% for the current year. If the inflation rate is 3% then the actual (real) interest charged would be 7%.

GDP (gross domestic product) is a measure of output of an economy. It is the sum of gross value added by all resident producers in the economy. Nominal GDP accounts for current prices in the economy and therefore affected by inflation in prices. Real GDP in this respect, provides for the actual effects of the output in the economy.

- *Illustration:*

  For a market, two years basket of goods is presented below.

	2019			2020		
Commodity	**Qty**	**Pr.**	**Output Value**	**Qty.**	**Pr.**	**Output Value**
Eggs	50	1.90	95	64	2.00	128
Milk	60	1.00	60	50	1.50	75
Bread	55	0.90	49.5	60	0.95	57
	Nominal GDP		**204.5**	Nominal GDP		**260**

The increase in GDP from 2019 to 2020 can be considered as dramatic increase, from 260 to 204.5. However, the GDP in 2020 is likely to overstate the quantity of commodities and increase in GDP may be because of price increase of inflation.

In order to determine the actual value of output or real GDP, the output value needs to be adjusted for changes in cost of commodities in the base year.

Real GDP = quantity of output in the current year ×price of output in base year

	2019	2020	Output Value
Commodity	**Pr.**	**Qty.**	
Eggs	1.90	64	121.6
Milk	1.00	50	50
Bread	0.90	60	54
	Real GDP (2020)		**225.6**

Now, in this case the Real GDP has increased but by a lower amount.

Inflation rate can be calculated by determining the GDP deflator. The GDP deflator is nominal GDP divided by real GDP. In this case:

$$\frac{260}{225.6} = 1.1524$$

In terms of index numbers, it can be written as 115 approximately.

Inflation rate is approximately:

$$\frac{115 - 100}{100} \times 100 = 15\%$$

The GDP inflator is the broader measure of inflation than the CPI.

STICKY NOTES

Index numbers are statistical devices that are used to express the relationship between quantitative variables

Indices are calculated with reference to a base number which is usually given a value of 100 or 1,000. All other numbers are calculated in relation to the base.

$$\text{Price index} = \frac{p_1}{p_0} \times 100$$

$$\text{Quantity index} = \frac{q_1}{q_0} \times 100$$

Consider price index numbers where there is more than one item.

$$\text{Simple aggregate price index} = \frac{\Sigma p_1}{\Sigma p_0} \times 100$$

The Laspeyre price index measures price changes with reference to the quantities of goods in the basket at the date that the index was first established. That is, base year quantities are taken as weights.

$$\frac{\sum(p_1 \times q_0)}{\sum(p_0 \times q_0)} \times 100$$

The Paasche price index measures price changes with reference to current quantities of goods in the basket.

$$\frac{\sum(p_1 \times q_1)}{\sum(p_0 \times q_1)} \times 100$$

SELF-TEST

9.1. The period with which other periods are to be compared is called:

(a) Current period (b) Base period

(c) Chain-base period (d) None of these

9.2. The index that uses quantities of base period as weights, so that only prices are allowed to change, in calculating weighted price aggregate is known as:

(a) Laspeyer's Index Number (b) Paasche's Index Number

(c) Fisher's Index Number (d) None of these

9.3. If the quantities of current year are used to calculate weighted price aggregate, the index number so calculated is called:

(a) Laspeyer's Index Number (b) Paasche's Index Number

(c) Fisher's Index Number (d) None of these

9.4. The geometric mean of Laspeyer's and Paasche's Index Number is called:

(a) Mean Index Number (b) Marshall Edge Worth's

(c) Fisher's Ideal Index Number (d) None of these

9.5. The equation 1995 = 100 indicates that the index numbers are calculated using:

(a) Base year 1995 (b) Current year 1995

(c) Previous year 1995 (d) None of these

9.6. The index that measures changes in prices that affect the cost of living of a large fraction of population is called:

(a) Whole sale price Index Number (b) Simple price relatives

(c) Consumer Price Index Number (d) None of these

9.7. Index numbers are used as the barometers of:

(a) Prices (b) Quantities

(c) Inflation (d) None of these

9.8. If the reciprocal of consumer price Index is expressed as the percentage the resulting value is called:

(a) Rate of inflation (b) Rate of deflation

(c) Purchasing power of money (d) None of these

9.9. If the purchasing power of money is multiplied by the current per capita income the resulting value is known as:

(a) Rate of inflation (b) Error in per capital income

(c) Real (or deflated) per capita income (d) None of these

9.10. If the relative changes in the current year prices are expressed on the basis of previous year prices the simple Index so calculated is known as:

(a) Simple Price Relative (b) Fixed base Index

(c) Chain-base Index or Link Relative (d) None of these

9.11. An increase in price index from 120 to 128 means that the prices have increased by____

(a) 8% (b) 8.667%

(c) 6% (d) 6.667%

9.12. The increase in index from 100 to 113 and a price increase from 0.9 to 1.35 in terms of percentage is:

(a) 13% for both (b) 50% for both

(c) 13% and 50% respectively (d) 50% and 13% respectively

9.13. _____________ tends to understate inflation.

(a) Laspeyer's Index Number (b) Paasche's Index Number

(c) Fisher's Index Number (d) All of these.

9.14. If Laspeyer's Index Number is of 113 and Paasche's Index Number of 118, then the Fisher's Price Index would be

(a) 115.4 (b) 112.34

(c) 154.7 (d) 123.4

9.15. Construct price index using Laspeyre formula for the following data:

Product	Price in 2020	Price in 2021	Quantity in 2020	Quantity in 2021
A	5	7	10	8
B	12	13	12	10
C	15	20	14	9

(a) 125.25 (b) 124.07

(c) 124.66 (d) 73.02

9.16. Construct price index using Paasche formula for the following data:

Product	Price in 2020	Price in 2021	Quantity in 2020	Quantity in 2021
A	5	7	10	8
B	12	13	12	10
C	15	20	14	9

(a) 125.25 (b) 124.07

(c) 124.66 (d) 73.02

9.17. Construct price index using Fisher formula for the following data:

Product	Price in 2020	Price in 2021	Quantity in 2020	Quantity in 2021
A	5	7	10	8
B	12	13	12	10
C	15	20	14	9

(a) 125.25 (b) 124.07

(c) 124.66 (d) 73.02

9.18. Construct quantity index using Laspeyre formula for the following data:

Product	Price in 2020	Price in 2021	Quantity in 2020	Quantity in 2021
A	5	7	10	8
B	12	13	12	10
C	15	20	14	9

(a) 73.02 (b) 72.33

(c) 72.68 (d) 125.25

9.19. Construct quantity index using Paasches formula for the following data:

Product	Price in 2020	Price in 2021	Quantity in 2020	Quantity in 2021
A	5	7	10	8
B	12	13	12	10
C	15	20	14	9

(a) 73.02 (b) 72.33

(c) 72.68 (d) 125.25

9.20. Construct quantity index using Fishers formula for the following data:

Product	Price in 2020	Price in 2021	Quantity in 2020	Quantity in 2021
A	5	7	10	8
B	12	13	12	10
C	15	20	14	9

(a) 73.02 (b) 72.33

(c) 72.68 (d) 125.25

9.21. Provided there is no change in prices between base year and current year but the quantity consumed for each type of product has increased by 5 units each. What will be the Laspeyre price index.

(a) 105 (b) 100

(c) 95 (d) Cannot be determined

9.22. Provided there is a 20% rise in prices from base year to current year of each of the products under consideration. Which of the following will increase by 20% as well?

(a)	Laspeyre price index	(b)	Paasche price index
(c)	Fisher price index	(d)	All of these

9.23. If Paasche price index is 20% higher than Laspeyre price index. What will be the value of Fisher's price index?

(a)	Cannot be determined	(b)	120
(c)	109.55	(d)	20% higher than Laspeyre price index

9.24. A machine costs Rs. 5,000 in 2015. The price indices are as follows:

2015: 45

2021: 63

How much will the machine cost in 2021?

(a)	8,000	(b)	7,000
(c)	6,000	(d)	10,000

9.25. Four years ago the price index of a particular product was 60. If the same index is now 108. Compute percentage increase in the price of a product.

(a)	80%	(b)	48%
(c)	58%	(d)	60%

9.26. Five years ago the price of a particular product was Rs. 5,000 and the relevant index was 90. Compute the price index now if the price of the product today is Rs. 7,500.

(a)	135	(b)	140
(c)	40	(d)	125

9.27. The price index linked to a particular product has increased from 80 to 96 in one years' time and is expected to increase by same percentage every year. Which of the following statements is true?

(a)	Price of the product has increased by 16%	(b)	Price of the product has increased by 20%
(c)	Price of the product will increase by 16% every year	(d)	Index will be 112 in next year

9.28. Select one or more correct options from below:

(a)	Index numbers are used to measure changes over time	(b)	A chain index series is an index series where each new value in the index is calculated based on the previous year
(c)	Consumer price index measures the cost of a consumer basket purchased by a "typical" urban family at a point in time as compared to the base year	(d)	A chain index series uses same base year

9.29. In Laspeyre price index number, quantities belonging to ____ year are taken as weight. Therefore, sometimes this method is called ___ year quantity weight method.

9.30. The percentage change in the Consumer Price Index can be used to determine rate of _________. If the CPI increases there is _____ and if the CPI ____ there is deflation

9.31. Fisher price index is a price index, computed for a given period by taking the _____ root of the product of the Paasche index value and the Laspeyre index value

Square

cube

9.32. In the calculation of the _____ price index (P_0Q_0), the denominator does not change from year to year. The only information that has to be collected each year is the prices of items in the index

Laspeyre

Paasche

9.33. The Paasche price index measures ___ changes with reference to current quantities of goods in the basket.

Price

Quantity

9.34. _____price index tends to overstate inflation whereas the ______price index tends to understate it. This is because consumers react to price increases by changing what they buy.

Laspeyre	Laspeyre
Paasche	Paasche

ANSWERS TO SELF-TEST QUESTIONS

9.1	**9.2**	**9.3**	**9.4**	**9.5**	**9.6**
(b)	(a)	(b)	(c)	(a)	(c)
9.7	**9.8**	**9.9**	**9.10**	**9.11**	**9.12**
(c)	(c)	(c)	(c)	(d)	(c)
9.13	**9.14**	**9.15**	**9.16**	**9.17**	**9.18**
(b)	(a)	(a)	(b)	(c)	(a)
9.19	**9.20**	**9.21**	**9.22**	**9.23**	**9.24**
(b)	(c)	(b)	(d)	(c)	(b)
9.25	**9.26**	**9.27**	**9.28**	**9.29**	**9.30**
(a)	(a)	(b)	(a), (b) & (c)	base	Inflation/deflation, inflation, decreases
9.31	**9.32**	**9.33**	**9.34**		
Square	Laspeyre,	Price	Laspeyre, Paasche		

CHAPTER 10

CORRELATION AND REGRESSION

IN THIS CHAPTER

AT A GLANCE

Relationship between two variables is established by means of statistical tools. These tools may differ when variables are categorical or are quantitative. An association exists between two variables if particular values for one variable are dependent on values of other variable. For Example, children with greater heights are likely to be of more weight, students with higher GPA are likely to be opting for higher education and people who smoke more have greater chances of cancer.

In identifying the nature of these associations between two variables, this chapter deals with contingency tables, scatter plots and statistical measures of correlation and regression analysis.

1 ASSOCIATION OR RELATIONSHIP BETWEEN VARIABLES

In understanding the relationship between two variables or association between them, variable that is independent and the one which is dependent on that independent variable is critical. No doubt, association can be both ways, but usually, possible change in dependent variable with the change in independent variable are analysed.

Contingency Tables:

A contingency table displays two categorical variables in rows and columns respectively. Frequency of observations in each intersection (combination) is then identified. It is also referred to as cross-tabulation.

- *For example:*

Consider the following data

S. No	Qualification	Preference for Learning
1	Bachelors	Online
2	Masters	Online
3	Masters	In person
4	Bachelors	In person
5	Bachelors	In person
6	Bachelors	Online

There are two variables: Qualification and preference for learning. Contingency table for two variables is as follows:

Qualification	Preference for Learning		Total
	Online	In person	
Bachelors	2	2	4
Masters	1	1	2
Total	3	3	

The relative frequency can help determine any sort of relationship between two variables. Whether preference for online or inperson learning is associated with a particular level of qualification.

Further analysis of association between categorical variables can be done using probabilities and ratios which is discussed later in the text.

Scatter plots:

A scatterplot is a graphical display for two quantitative variables using the horizontal (x) axis for the independent variable x and the vertical (y) axis for the dependent variable y[1].

Observations are plotted as points on the graph. These plotted points then generate a pattern. The patterns on the graph can be positive, negative, weak, strong, linear or non linear.

[1] Agresti, A., Franklin, C., & Klingenberg, B., 2018. Statistics – the science and art of learning from Data. Pearson. Fourth Edition. pg. 129

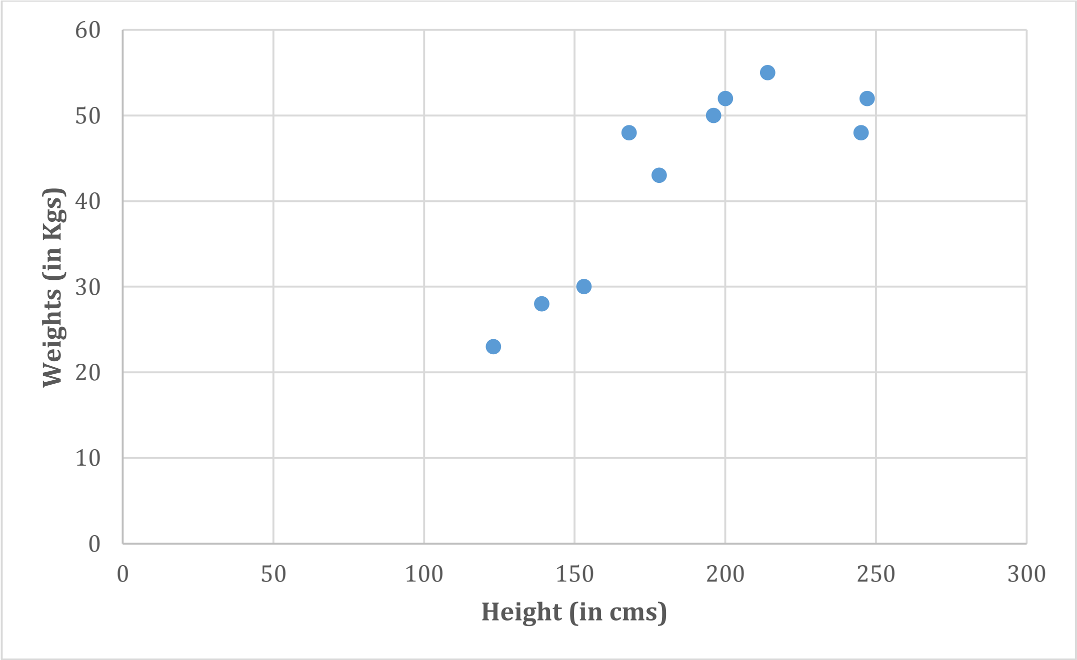
AT A GLANCE SPOTLIGHT STICKY NOTES

► *For example:*

Consider the following data

S. No	Height (in cms)	Weight (in kgs)
1	123	23
2	245	48
3	247	52
4	153	30
5	178	43
6	196	50
7	200	52
8	168	48
9	214	55
10	139	28

Scatter plot with Height on x-axis and Weight on y-axis is given below:



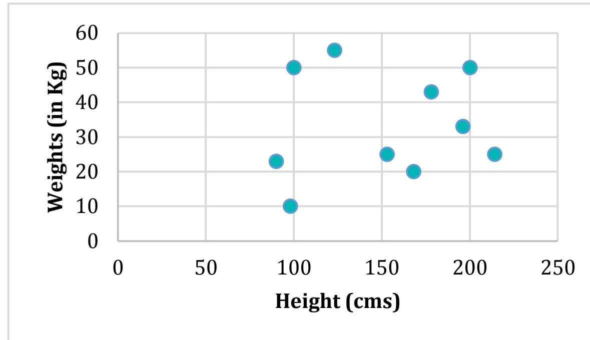
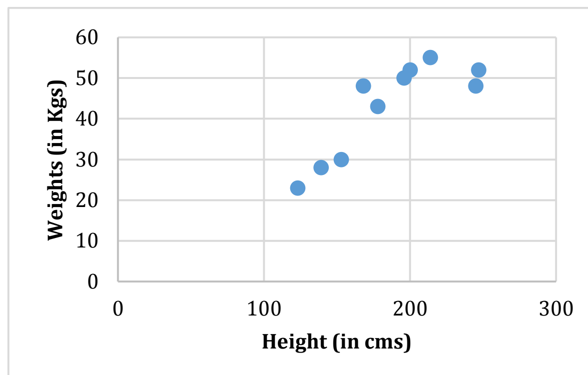
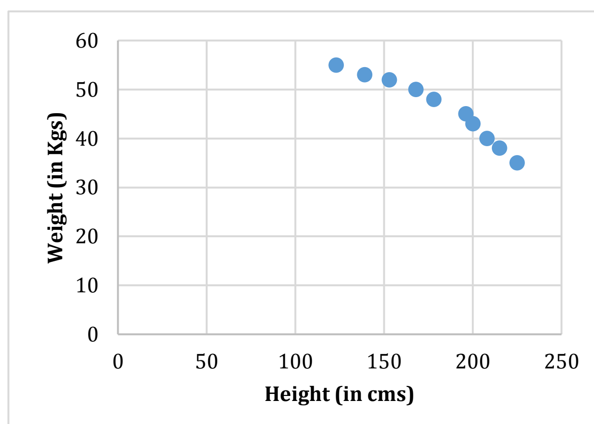
Looking at the scatter plot, it can be inferred that the people with greater heights are weighing more. The pattern on the plot seems to be increasing linearly as we are increasing the heights in most of the cases.

Positive, negative or no associations:

Positive association is when value of y goes up with the increasing value of x. Negative association is when value of y goes down with the increasing value of x. There seems to be no association between variables if values of y fails to follow any pattern with changing values of x.

AT A GLANCE
SPOTLIGHT
STICKY NOTES

- *For example:*

No Association	
Positive	
Negative	

Limitations

A scatter diagram can be a useful in giving a visual impression of the relationship between variables and is a useful starting point for a deeper analysis. However, they suffer from the following limitations:

- they might indicate a relationship where there is none;
- they can just show relationship between only two variables at a time but the issue under study might be more complexed than this; and
- they might lead to incorrect conclusions if there are only few data points available or the data collected is atypical for some reason.

2 CORRELATION AND CORRELATION COEFFICIENT:

Measure of the degree of association between two quantitative variables is Correlation. It determines the strength or direction of the relationship between variables.

It is usually denoted by *r* and takes values between -1 and +1.

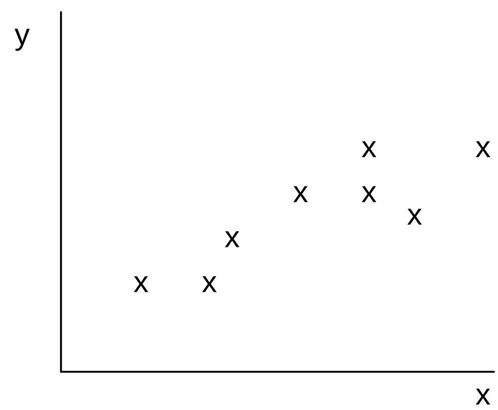
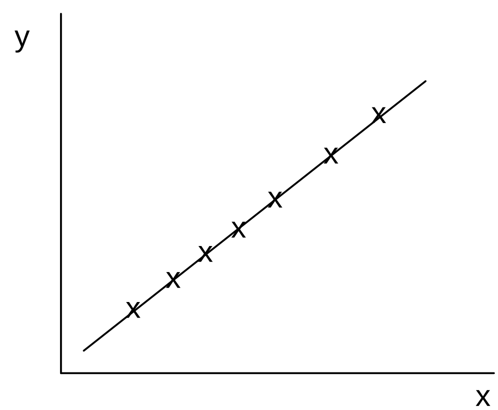
- A positive value for *r* indicates a positive (linear) association, and
- A negative value for *r* indicates a negative (linear) association.
- Association (linear) is stronger when the value for *r* is closer to +1, and
- Association (linear) is weaker when the value of r is closer to 0.
- No association when the value of *r* is 0.

As a general guide, a value for r between + 0.90 and +1 indicates good linear correlation between the values of *x* and *y*.

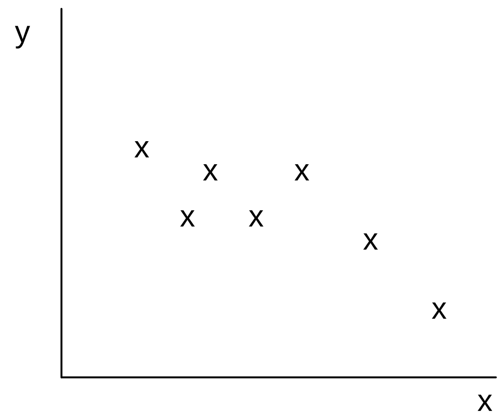
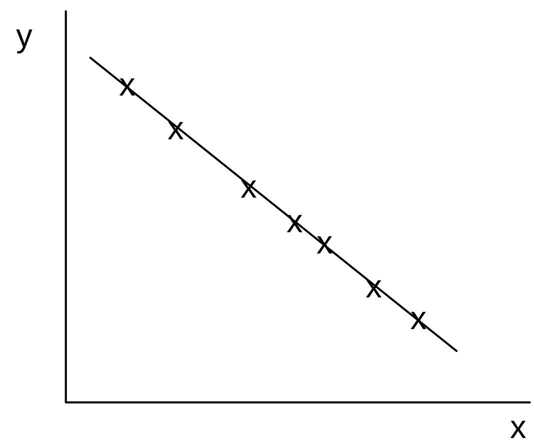
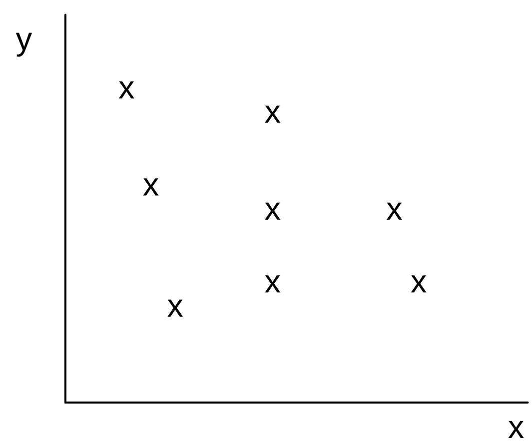
A positive correlation indicates a positive association, and a negative correlation indicates a negative association[2]. However, it is to note that correlation does not indicate causation.

Units of the variables make no impact on the value of the correlation for example, if weights are changed to pounds instead of kgs, the correlation would be same. Moreover, it is not dependent on which variable is taken as independent or dependent.

► *Illustration:*

Positive correlation means that the value of y increases as the value of x increases (and vice versa). (r closer to 1)	
Perfect positive correlation is when all the data points lie in an exact straight line and a linear relationship exists between the two variables. (r equal to 1)	

[2] Agresti, A., Franklin, C., & Klingenberg, B., 2018. Statistics – the science and art of learning from Data. Pearson. Fourth Edition. pg. 134

Negative correlation is where the value of y declines as the value of x increases (and vice versa). (r closer to -1)	
Perfect negative correlation is when all the data points plotted lie in an exact straight line. (r is equal to -1)	
'Uncorrelated' means that no correlation exists between the variables (r is equal to 0)	

Correlation coefficient r

Correlation between different variables can be measured as a correlation coefficient

- *Formula:*

Method 1:

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$$

Where:

x, y = values of pairs of data.

n = the number of pairs of values for x and y

Method 2:

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{n(s_x s_y)}$$

Where:

x, y = values of pairs of data.

$\bar{x}$ = variable mean

n = the number of pairs of values for x and y

s = standard deviation of variables = $s_x = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$

The Expression: $\frac{\sum(x-\bar{x})(y-\bar{y})}{n}$ is called Covariance and is denoted by s_{xy}

- *For example:*

A commercial tea producing business employs professional tea tasters for buying decisions.

The two tasters are ranked in six different brands of tea in their order of preference.

The business is interested to know whether there is a relationship between the choices that they make.

	Taster 1's preference	Taster 2's preference
	x	**y**
Brand A	1	2
Brand B	2	3
Brand C	3	1
Brand D	4	4
Brand E	5	6
Brand F	6	5
	21	21
	= ∑x	= ∑y

AT A GLANCE SPOTLIGHT STICKY NOTES

AT A GLANCE
SPOTLIGHT
STICKY NOTES

Method 1:

	x	y	x^2	xy	y^2
Brand A	1	2	1	2	4
Brand B	2	3	4	6	9
Brand C	3	1	9	3	1
Brand D	4	4	16	16	16
Brand E	5	6	25	30	36
Brand F	6	5	36	30	25
	21	21	91	87	91
	$=\sum x$	$=\sum y$	$=\sum x^2$	$=\sum xy$	$=\sum y^2$

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$$

$$r = \frac{6(87) - (21)(21)}{\sqrt{(105)(105)}}$$

$$\mathrm{r} = \frac{81}{105}$$

$$\mathrm{r} = 0.7714$$

Method 2:

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{n(s_x s_y)}$$

$$s_x = \sqrt{\frac{\sum(x-\bar{x})^2}{\mathrm{n}}} \text{ and } s_y = \sqrt{\frac{\sum(y-\bar{y})^2}{\mathrm{n}}}$$

			$\bar{x} = 3.5$	$\bar{y} = 3.5$			
	x	y	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
Brand A	1	2	-2.5	-1.5	3.75	6.25	2.25
Brand B	2	3	-1.5	-0.5	0.75	2.25	0.25
Brand C	3	1	-0.5	-2.5	1.25	0.25	6.25
Brand D	4	4	0.5	0.5	0.25	0.25	0.25
Brand E	5	6	1.5	2.5	3.75	2.25	6.25
Brand F	6	5	2.5	1.5	3.75	6.25	2.25
	21	**21**			**13.5**	**17.5**	**17.5**
	$=\sum x$	$=\sum y$			$\sum(x-\bar{x})(y-\bar{y})$	$\sum(x-\bar{x})^2$	$\sum(y-\bar{y})^2$

$$s_x = \sqrt{\frac{17.5}{6}} \text{ and } s_y = \sqrt{\frac{17.5}{6}}$$

$$s_x = 1.7078 \text{ and } s_y = 1.7078$$

$$r = \frac{\sum(x-\overline{x})(y-\overline{y})}{n(s_x\, s_y)}$$

$$r = \frac{13.5}{6(1.7078)(1.7078)}$$

$$r = \frac{13.5}{17.5}$$

$$r = 0.7714$$

- *For example:*

The following table represents output and total cost.

Output	Total cost
17	63
15	61
12	52
22	74
18	68

The correlation coefficient of the following data can be calculated as follows:

x	y	x^2	xy	y^2
17	63	289	1,071	3,969
15	61	225	915	3,721
12	52	144	624	2,704
22	74	484	1,628	5,476
18	68	324	1,224	4,624
84	318	1,466	5,462	20,494
$= \sum x$	$= \sum y$	$= \sum x^2$	$= \sum xy$	$= \sum y^2$

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$$

$$r = \frac{5(5{,}462) - (84)(318)}{\sqrt{(5(1{,}466) - 84^2)(5(20{,}494) - 318^2)}}$$

$$\mathrm{r} = \frac{598}{\sqrt{(274)(1{,}346)}} = \frac{598}{607} = +0.985$$

The correlation coefficient r is + 0.985.

Coefficient of determination r^2

The coefficient of determination (***r^2***) is the square of the correlation coefficient.

The value of ***r^2*** shows how much variation in the value of y is explained by variations in the value of x.

The value of the coefficient of determination must always be in the range ***0*** to ***+1***.

- *For example:*

  If the value of ***r*** is + 0.70 then ***r^2*** = 0.49. This means that on the basis of the data used 0.49 or 49% of variations in the value of y are explained by variations in the value of x.

  Similarly, where ***r*** = + 0.97, ***r^2*** = 0.940 (0.97 × 0.97) meaning that 94.09% of the variations in one variable can be explained by variations in the other.

Spearman's rank correlation coefficient

Rank correlation coefficient is calculated when variables are based on rank or order rather than their magnitude. For example, measures for attributes like intelligence, beauty, ability to work and their correlation can be calculated using Spearman's formula.

- *Formula:*

$$r = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

Where:

d = Difference between rank values in a pair of observations.

n = number of pairs of data

Note that the number 6 in the numerator is a constant and does not relate to the number of observations.

- *For example:*

  Using variables x and y in the above example, lets calculate rank correlation coefficient:

x	y	d	d^2
1	2	–1	1
2	3	–1	1
3	1	2	4
4	4	0	0
5	6	–1	1
6	5	1	1
21	21		8
= $\sum x$	= $\sum y$		= $\sum d^2$

$$r = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

$$r = 1 - \frac{6(8)}{6(36 - 1)} = 1 - \frac{48}{210} = 0.77$$

The rank correlation coefficient r is + 0.77.

Also note that some organizations might develop a method of quantifying qualitative attributes. For example, banks develop credit scores to give them an indication of the risk of lending to a particular party, and companies often develop scoring systems to measure the overall performance of employees across key areas. Often these "quantifications" will be based in part on subjective judgements.

3 LINEAR REGRESSION ANALYSIS:

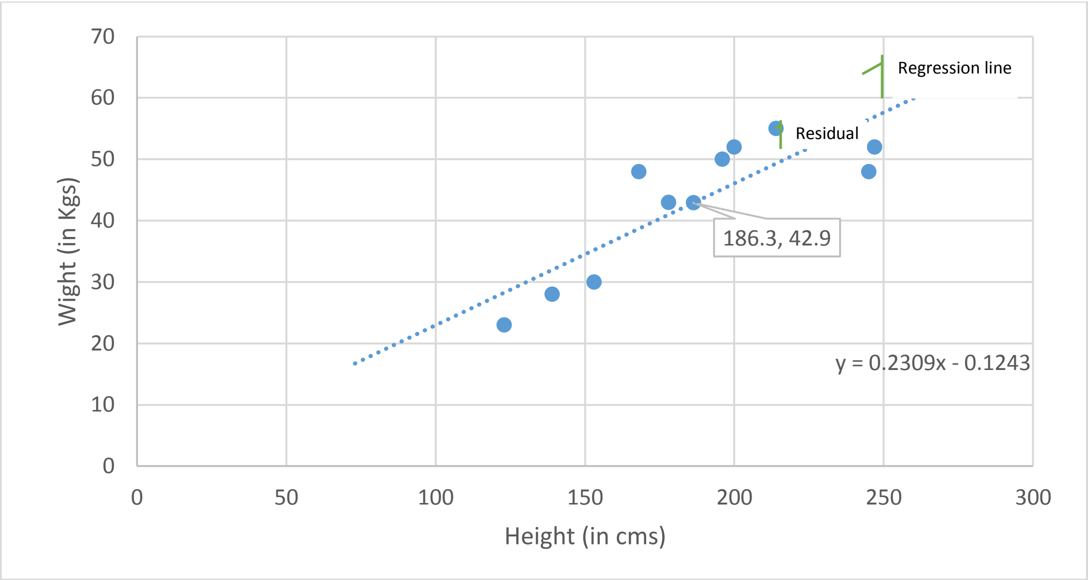
Relationship or association between two variables can be identified using correlation. However, to what extent these variables are related functionally (or algebraically) with each other may be identified using regression analysis.

Linear regression analysis is a statistical technique for calculating a line that best fits the points in the scatter diagram. It is the line that minimized the sum of the squared deviations or residuals between line and actual values. In other words, it is a line that makes the vertical distances from the points to this line as small as possible[3]. The method used is call **least square method.** The best fit or the link (if established) can support forecasting and planning.

Linear regression analysis is widely used in economics and business. One application is that it can be used to estimate fixed costs and variable cost per unit (or number of units) from historical total cost.

- *For example:*

The regression line for the heights and weights, in the example discussed before, is represented as follows:



- *Formula:*

Regression line is written as simple algebraic equation of a straight line with $\hat{y}$ as predicted values of dependent variable in the straight line.

$\hat{y} = a + bx.$

The linear regression formulae for calculating ***a*** and ***b*** are shown below.

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \quad \text{or} \quad b = \frac{\sum(x-\overline{x})(y-\overline{y})}{\sum(x-\overline{x})^2}$$

$$a = \frac{\sum y}{n} - \frac{b\sum x}{n} \quad \text{or} \quad a = \overline{y} - b\overline{x}$$

[3] Albright, S.C., & Winston, W. L. 2015. Business Analytics: Data Analysis and Decision Making. 5th Edition. Cengage Learning. Pg. 477

Where:

x, y = values of pairs of data.

n = the number of pairs of values for x and y.

$\sum$ = a sign meaning the sum of. (The capital of the Greek letter sigma).

$\bar{x}, \bar{y}$= means of variables x and y

Note: the term b must be calculated first as it is used in calculating a.

Forecasting

Once the equation of the line of best fit is derived it can be used to make forecasts of the impact of changes in x on the value of y. This is denoted by ŷ.

- *For example (method 1):*

A company has recorded the following output levels and associated costs in the past six months:

Month	Output (000 of units)	Total cost (Rs m)
January	5.8	40.3
February	7.7	47.1
March	8.2	48.7
April	6.1	40.6
May	6.5	44.5
June	7.5	47.1

In constructing the regression line (line of best fit), following working is required:

	x	y	x^2	xy
January	5.8	40.3	33.64	233.74
February	7.7	47.1	59.29	362.67
March	8.2	48.7	67.24	399.34
April	6.1	40.6	37.21	247.66
May	6.5	44.5	42.25	289.25
June	7.5	47.1	56.25	353.25
	41.8	268.3	295.88	1,885.91
	= $\sum x$	= $\sum$y	= $\sum x^2$	= $\sum xy$

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

$$b = \frac{6(1{,}885.91) - (41.8)(268.3)}{6(295.88) - (41.8)^2}$$

$$b = \frac{11{,}315.46 - 11{,}214.94}{1{,}775.28 - 1{,}747.24} = \frac{100.52}{28.04} = 3.585$$

AT A GLANCE SPOTLIGHT STICKY NOTES

This is the cost in millions of rupees of making 1,000 units

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

$$a = \frac{268.3}{6} - \frac{3.585(41.8)}{6}$$

$$a = 44.72 - 24.98 = 19.74$$

$$y = a + bx$$

$$y = 19.74 + 3.585x$$

The cost of 3,000 and 10,000 units of output can be calculated using the above regression line as follows:

$$y = 19.74 + 3.585(3) = 30.5 \text{ (for 3000 units)}$$

$$y = 19.74 + 3.585(10) = 55.6 \text{ (for 10,000 units)}$$

- *For example (method 2):*

For the following information, the line of best fit can be determined as:

Output (000s)	Total cost (Rs m)
17	63
15	61
12	52
22	74
18	68

We have to extend the table to find relevant values as follows:

		$\bar{x}$ =16.8	$\bar{y}$ =63.6		
x	**y**	**(x-$\bar{x}$)**	**(y-$\bar{y}$)**	**(x-$\bar{x}$) (y-$\bar{y}$)**	**(x-$\bar{x}$)^2**
17	63	0.2	-0.6	-0.12	0.04
15	61	-1.8	-2.6	4.68	3.24
12	52	-4.8	-11.6	55.68	23.04
22	74	5.2	10.4	54.08	27.04
18	68	1.2	4.4	5.28	1.44
84	**318**			**119.6**	**54.8**
= $\sum x$	= $\sum$y			=$\sum (x - \bar{x})$ (y-$\bar{y}$)	=$\sum (x - \bar{x})^2$

AT A GLANCE
SPOTLIGHT
STICKY NOTES

$$b = \frac{\sum(x-\overline{x})(y-\overline{y})}{\sum(x-\overline{x})^2}$$

$$b = \frac{119.6}{54.8}$$

$$b = 2.18$$

considering regression line passes through the mean values then taking x = 16.8 and y = 63.6

a = $\overline{y}$ - b$\overline{x}$a = 63.6 – (2.18)(16.8)

a = 63.6 – 36.624

a = 26.976

The line of best fit:
$y = a + bx$

$y = 26.98 + 2.18x$

In order to estimate the total costs when output is 15,000 units, we may put the values to find the estimated cost.

$y = 26.98 + 2.18(15) = 59.68$

Limitations

The analysis is only based on a pair of variables. There might be other variables which affect the outcome but the analysis cannot identify these outcomes.

A regression line should only be used for forecasting if there is a good fit between the line and the data.

It might not be valid to extrapolate the line beyond the range of observed data. In the example above the cost associated with 10,000 units was identified as 55.6. However, the data does not cover volumes of this size and it is possible that the linear relationship between costs and output may not be the same at this level of output.

Difference between Correlation and Regression:

Correlation identifies the degree of relationship or association between two variables. It does not provide for causation or prediction. Whereas regression provide for a functional relationship between two variables where one variable is dependent and the other an independent variable. The purpose for identifying the two variables as dependent and independent is to predict values of a dependent variable on the basis of independent variable.

Once an equation is derived, correlation analysis can help determine the extent of relationship between two variables.

Relationship between correlation coefficient and b of regression is written as: $r = b\frac{S_x}{S_y}$ or $b = r\frac{S_y}{S_x}$

STICKY NOTES

Association between two variables x and y can be positive association when value of one variable 'y' goes up with the increasing value of another variable 'x'; negative when value of y goes down with the increasing value of x. There seems to be no association between variables if values of y fail to follow any pattern with changing values of x.

Measure of the degree of association between two quantitative variables is Correlation. It determines the strength or direction of the relationship between variables. It is usually denoted by r and takes values between -1 and +1.

Correlation between different variables can be measured as a correlation coefficient

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}} \text{ or } r = \frac{\sum(x-\bar{x})(y-\bar{y})}{n(s_x s_y)}$$

Where:

x, y = values of pairs of data.

n = the number of pairs of values for x and y

$\bar{x}$ = variable mean

s = standard deviation of variables = $s_x = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$

The Expression: $\frac{\sum(x-\bar{x})(y-\bar{y})}{n}$ is called Covariance and is denoted by s_{xy}

Spearman's Rank correlation coefficient is calculated when variables are based on rank or order rather than their magnitude.

$$r = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

Linear regression analysis is a statistical technique for calculating a line that best fits the points in the scatter diagram. It is the line that minimized the sum of the squared deviations or residuals between line and actual values.

Regression line is written as simple algebraic equation of a straight line with ŷ as predicted values of dependent variable in the straight line.

$$\hat{y} = a + bx.$$

The linear regression formulae for calculating a and b are shown below.

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \quad \text{or} \quad b = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$$

$$a = \frac{\sum y}{n} - \frac{b\sum x}{n} \quad \text{or} \quad a = \bar{y} - b\bar{x}$$

Where:

x, y =	values of pairs of data.
n =	the number of pairs of values for x and y.
$\sum$ =	A sign meaning the sum of. (The capital of the Greek letter sigma).
$\bar{x}, \bar{y}$ =	Variable means

Note: the term b must be calculated first as it is used in calculating a.

AT A GLANCE

SPOTLIGHT

STICKY NOTES

SELF-TEST

10.1. If the values of two different variables (say *x* and *y*) are plotted on a rectangular axes, such a plot is referred to as a:

(a) Frequency diagram (b) Value diagram

(c) Scatter diagram (d) None of these

10.2. From the inspection of scatter diagram if it is seen that the points follow closely a straight line, it indicates that the two variables are to some extent:

(a) Unrelated (b) Related

(c) Linearly related (d) None of these

10.3. In a scatter diagram, if the points follow closely a straight line of positive slope, the two variables are said to have:

(a) No correlation (b) High positive correction

(c) Negative correlation (d) None of these

10.4. In a scatter diagram, if the points follow clearly a straight line of negative slope, the two variables are said to have:

(a) No correlation (b) High positive correlation

(c) High negative correlation (d) None of these

10.5. In a scatter diagram, if the points follow a strictly random pattern, the two variables are said to have:

(a) No linear relationship (b) Low positive relationship

(c) Low negative relationship (d) None of these

10.6. A measure of the strength or degree of relationship or the interdependence is called:

(a) Correlation (b) Regression

(c) Least square estimate (d) None of these

10.7. The phenomenon that investigates the dependence of one variable on one or more independent variables is called:

(a) Correlation (b) Regression

(c) Least square estimate (d) None of these

10.8. The linear relation between a dependent and an independent variable is called:

(a) Regression line (b) Regression co-efficient

(c) Co-efficient of correlation (d) None of these

10.9. Slope of the regression line is called:

(a) Regression parameter (b) Sample parameter

(c) Regression co-efficient (d) None of these

10.10. In regression analysis, if the value of a is positive the value of b:

(a) Must be positive (b) May take any value

(c) Must be negative (d) Less than -1or more than 1

10.11. The procedure which selects that particular line for which the sum of the squares of the vertical distances from the observed points to the line is as small as possible, is called:

(a) Sum of squares method (b) Sum of squares of errors method

(c) Least square method (d) None of these

10.12. The numerical values of regression co-efficients must be:

(a) Both positive (b) Both negative

(c) Both positive or both negative (d) None of these

10.13. The dependent variable is also called response or:

(a) The explained variable (b) Unexplained variable

(c) The explanatory variable (d) None of these

10.14. The explained variable or response is also called:

(a) The independent variable (b) The dependent variable

(c) Non-random variable (d) None of these

10.15. The predictor or unexplained variable is also called:

(a) The independent variable (b) The dependent variable

(c) Random variable (d) None of these

10.16. In regression analysis, *b = 2.8*, indicates that the value of dependent variable:

(a) Increases by 2.8 units at per unit increase in independent variable

(b) Decreases by 2.8 units at per unit increase in independent variable

(c) Increases by 2.8 units at per unit decrease in independent variable

(d) None of these

10.17. If $\bar{x}$ = 11.33; $\bar{y}$ = 33.56 and b_{yx}=2.832 then a is equal to:

(a) 0.96 (b) 1.47

(c) 11.85 (d) 4.00

10.18. If a random sample of 9 observations yielded the values $\sum x = 102, \sum y = 302, \sum xy = 3583 \text{ and } \sum x^2 = 1308$ then the value of b is:

(a) 1.47 (b) 2.831

(c) Cannot be determined (d) None of these

10.19. If Y is the observed value and $\hat{Y}$ is the estimated value (estimated by using the regression line) then $\sum(Y - \hat{Y})$:

(a) Should be zero

(b) Is likely to be close to zero

(c) In majority of the cases would be equal to zero

(d) None of these

10.20. If two variables tends to vary simultaneously in some direction, they are said to be:

(a) Dependent (b) Independent

(c) Correlated (d) None of these

10.21. If two variable tends to increase (or decrease) together, the correlation is said to be:

(a) Zero (b) Direct or positive

(c) 1 (d) None of these

10.22. If one variable tends to increase as the other variable decreases, the correlation is said to be:

(a) Zero (b) Inverse or negative

(c) -1 (d) None of these

10.23. While calculating "r" if x and y are interchanged i.e. instead of calculating r_{xy} if r_{yx} is calculated then:

(a) $r_{xy} = r_{yx}$ (b) $r_{xy} > r_{yx}$

(c) $r_{xy} < r_{yx}$ (d) None of these

10.24. Limits of the co-efficient of Correlation are:

(a) -1 to 0 (b) 0 to 1

(c) -1 to +1 (d) None of these

10.25. If $r = 0.9$ and if 5 is subtracted from each observation of x, then r will:

(a) Decrease by 5 units (b) Decreases by less than 5 units

(c) Remain unchanged (d) None of these

10.26. If $r = 0.9$ and if 5 is added to each observation of x, then r will:

(a) Increase by 5 units (b) Increase by more than 5 units

(c) Remain unchanged (d) None of these

10.27. If $r = 0.9$ and if 3 is subtracted from each observation of Y, then r will:

(a) Decrease by 3 units (b) Decrease by less than 3 units

(c) Remain unchanged (d) None of these

10.28. If $r = 0.9$ and if 3 is added to each observation of y, then r will:

(a) Increase by 3 units (b) Increase by more than 3 units

(c) Remain unchanged (d) None of these

10.29. If $r = 0.9$ and if 3 is subtracted from each observation of x and 5 is added to each observation of y, then r will:

(a) Decrease by 2 units (b) Increase by 2 units

(c) Remain unchanged (d) None of these

10.30. If r = 0.9 and each observation of x is multiplied by 100, then r will:

(a) Increase by 100 times (b) Less than 100 times

(c) Remain unchanged (d) None of these

10.31. If r = 0.9 and each observation of Y is divided by 10, then r will:

(a) Decrease by 10 times (b) Decrease by less than 10 times

(c) Remain unchanged (d) None of these

10.32. If $r = 0.9$ and each observation of x and y is divided by 10, then r will:

(a) Decrease by 10 times (b) Decrease by 100 times

(c) Remain unchanged (d) None of these

10.33. The co-efficient of correlation is independent of:

(a) Only origin (b) Only scale

(c) Origin and scale (d) None of these

10.34. The geometric mean of two regression co-efficient is equal to:

(a) Co-efficient of determination (b) Co-efficient of correlation

(c) Co-efficient of rank correlation (d) None of these

10.35. If b_{xy} = -0.78 and b_{yx} = -0.45, then r is equal to:

(a) +0.351 (b) -0.351

(c) Cannot be determined (d) None of these

10.36. If b_{xy} = -0.78 and b_{yx} = 0.45, then r is equal to:

(a) +0.351 (b) -0.351

(c) Cannot be determined (d) None of these

10.37. If b_{xy} = +1.93 and b_{yx} = 0.6, then r is equal to:

(a) 1.158 (b) 1.0761

(c) Data is fictitious (d) None of these

10.38. If b_{xy} = 1.93 and b_{yx} = 0.51, then r is equal to:

(a) 0.9843 (b) 0.992

(c) Data is fictitious (d) None of these

10.39. If b_{xy} = -1.93 and b_{yx} = 0.51, then r is equal to:

(a) -0.9843 (b) -0.992

(c) Data is fictitious (d) None of these

10.40. If $N = 6, \sum x = 68, \sum y = 112, \sum xy = 1292, \quad \sum x^2 = 786, \sum y^2 = 2128$ then r is equal to:

(a) 0.947 (b) 0.8968

(c) Cannot be determined (d) None of these

10.41. The co-efficient of correlation can never be:

(a) Negative (b) Positive

(c) Zero (d) Can assume any value

10.42. The square of r is known as:

(a) Co-efficient of correlation (b) Co-efficient of regression

(c) Co-efficient of determination (d) None of these

10.43. The lower and upper limits of r^2 are:

(a) -1 to + 1 (b) 0 to 1

(c) $-\infty$ to $+\infty$ (d) None of these

10.44. The quantity which describes that the proportion (or percentage) of variation in the dependent variable explained (or reduced) by the independent variable is called:

(a) Co-efficient of determination (b) Co-efficient of regression

(c) Co-efficient of correlation (d) None of these

10.45. If r = 0.8, then the variation in the dependent variable y due to independent variable x is about:

(a) 80% (b) 64%

(c) 64% to 80% (d) None of these

10.46. If r =1 0.8 and b_{yx}= 1.04 then b_{xy} is equal to:

(a) 0.769 (b) 0.615

(c) Cannot be determined (d) None of these

10.47. If r^2 = 0.796 and b_{xy} = -1.04 then b_{yx} is equal to:

(a) 0.765 (b) -0.765

(c) Cannot be determined (d) None of these

AT A GLANCE

SPOTLIGHT

STICKY NOTES

10.48. For the following pair of values compute correlation coefficient

(1,3), (4,8), (8,17), (10,18)

(a)	0.9887	(b)	0.9775
(c)	1.7743	(d)	0.5510

10.49. For the following pair of values compute coefficient of determination

(1,3), (4,8), (8,17), (10,18)

(a)	0.9887	(b)	0.9775
(c)	1.7743	(d)	0.5510

10.50. For the following pair of values compute equation of least squares line of Y on X

(1,3), (4,8), (8,17), (10,18)

(a)	y =1.297 + 1.774x	(b)	x = - 0.5860+0.55y
(c)	y =1.774 + 1.297x	(d)	x = 0.55 - 0.5860y

10.51. For the following pair of values compute equation of least squares line of X on Y

(1,3), (4,8), (8,17), (10,18)

(a)	y = 1.297 + 1.774x	(b)	x = - 0.5860+0.55y
(c)	y = 1.774 + 1.297x	(d)	x = 0.55 - 0.5860y

10.52. For the following pair of values compute regression coefficient of Y on X

(1,3), (4,8), (8,17), (10,18)

(a)	0.9887	(b)	0.9775
(c)	1.7743	(d)	0.5510

10.53. For the following pair of values compute regression coefficient of X on Y

(1,3), (4,8), (8,17), (10,18)

(a)	0.9887	(b)	0.9775
(c)	1.7743	(d)	0.5510

10.54. Select one or more of the possible correct statements:

1. Correlation may be direct or positive.
2. Correlation must be direct or positive.
3. Correlation may be inverse or negative.
4. Correlation coefficient ranges from 0 to 1

10.55. Select one or more of the possible correct statements:

1. A positive relationship between two variables means that high values of one variable are paired with high values of the other.
2. Correlation coefficient is not expressed in the units of measurement from which it is obtained.
3. Regression coefficient ranges between -2 and +1
4. Regression coefficient cannot be 1

10.56. Select one or more of the possible correct statements:

1. Coefficient of determination is obtained by squaring the regression coefficient.
2. Coefficient of determination is obtained by squaring the coefficient of correlation.
3. Coefficient of determination can be negative.
4. Coefficient of determination cannot be negative.

10.57. Given N=4 and $\sum x$ =23. Calculate $\overline{x}$.

(a) 5.75 (b) 11

(c) 0.0782 (d) 0.25

10.58. Given b=1.89, $\overline{x}$=5.75 and $\bar{y}$=11. Calculate a.

(a) 5.75 (b) 11

(c) 0.0782 (d) 0.25

10.59. Given r= 0.9745, regression coefficient y on x = 1.8994. Calculate regression coefficient x on y

(a) 0.5 (b) 11

(c) 0.0782 (d) 0.25

10.60. Using y = 1.297 + 1.774x. Compute y when x = 3

(a) 6.619 (b) 7.719

(c) 8.819 (d) cannot be determined

10.61. Using x = - 0.55 + 2y. Compute x when y=4

(a) 7.45 (b) 8.45

(c) 9.45 (d) 6.45

10.62. Identify regression coefficient from the following equation: Y=3+4x

(a) 3 (b) 4

(c) 3/4 (d) 4/3

10.63. The regression coefficient of a perfectly positively correlated data will be:

(a) -1 (b) 0

(c) +1 (d) It can take any value

10.64. The coefficient of correlation of a perfectly positively correlated data will be:

(a) -1 (b) 0

(c) +1 (d) It can take any value

10.65. The coefficient of determination of a perfectly positively correlated data will be:

(a) -1 (b) 0

(c) +1 (d) It can take any value

10.66. From its past experience a company has observed that it is able to achieve sales revenue of around Rs. 500,000 in years when there is no marketing expenditure and also that the sales revenue increases by around Rs. 150,000 for every Rs. 75,000 spending on marketing expenditure. Identify the best possible equation linking sales and marketing expenditure from the data provided

(a) Sales = 500,000 + 75,000 × marketing expenditure (b) Marketing expenditure = 500,000 + 2 × Sales

(d) Sales = 2 × marketing expenditure (d) Sales = 500,000 + 2 × marketing expenditure

10.67. Identify the limitation(s) of a scatter diagram:

1. Useful aid to give a visual impression of the relationship between variables.
2. Might indicate a relationship where there is none.
3. Might lead to incorrect conclusions if there are only few data points available or the data collected is atypical for some reason.

ANSWERS TO SELF-TEST QUESTIONS

10.1	**10.2**	**10.3**	**10.4**	**10.5**	**10.6**
(c)	(c)	(b)	(c)	(a)	(a)
10.7	**10.8**	**10.9**	**10.10**	**10.11**	**10.12**
(b)	(a)	(c)	(b)	(c)	(c)
10.13	**10.14**	**10.15**	**10.16**	**10.17**	**10.18**
(a)	(b)	(a)	(a)	(b)	(d)
10.19	**10.20**	**10.21**	**10.22**	**10.23**	**10.24**
(a)	(c)	(b)	(b)	(a)	(c)
10.25	**10.26**	**10.27**	**10.28**	**10.29**	**10.30**
(c)	(c)	(c)	(c)	(c)	(c)
10.31	**10.32**	**10.33**	**10.34**	**10.35**	**10.36**
(c)	(c)	(c)	(b)	(d)	(c)
10.37	**10.38**	**10.39**	**10.40**	**10.41**	**10.42**
(b)	(b)	(c)	(a)	(d)	(c)
10.43	**10.44**	**10.45**	**10.46**	**10.47**	**10.48**
(b)	(a)	(b)	(b)	(b)	(a)
10.49	**10.50**	**10.51**	**10.52**	**10.53**	**10.54**
(b)	(a)	(b)	(c)	(d)	1, 3
10.55	**10.56**	**10.57**	**10.58**	**10.59**	**10.60**
1,2	2,4	(a)	(c)	(a)	(a)
10.61	**10.62**	**10.63**	**10.64**	**10.65**	**10.66**
(a)	(b)	(d)	(c)	(c)	(d)
10.67					
2,3					

CHAPTER 11

PROBABILITY CONCEPTS

IN THIS CHAPTER

AT A GLANCE

In our daily events, often number of ways to perform a certain task are determined. Similarly, in statistics we often need to be able to estimate the number of possible outcomes given a set of circumstances.

To determine the possible outcomes and possible ways number of things can be ordered or arranged various methods including permutations and combinations and simply tree diagrams can be used.

This chapter also discusses probability and related concepts. Probability is the measure of the likelihood of occurrence of event. It is normally expressed as a number between 0 and 1 (or as percentage between 0 and 100%). The estimate for an event occurring helps in various decisions including prediction. Its business applications extends to assessment and risk management in business.

1. POSSIBLE OUTCOMES

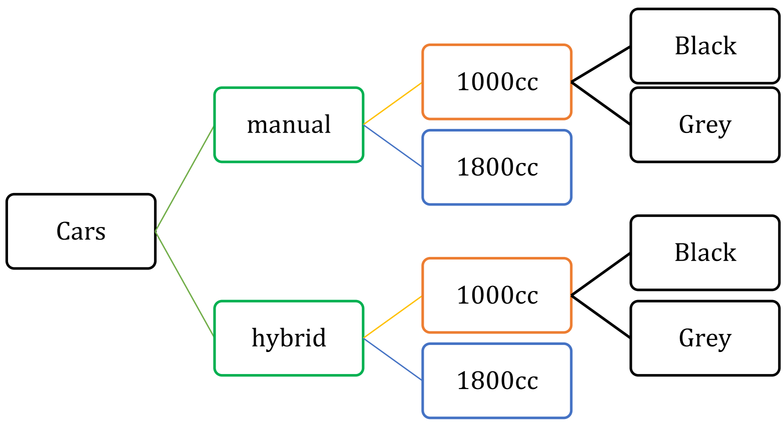
The logical possibilities or outcomes that address a given circumstance can be as simple as counting the available options to often difficult and error prone calculations.

Trees diagrams:

One of the ways to determine the possible outcomes is to use tree diagram to reach out to possible outcomes. Here, each possible outcome is denoted by a leaf or node at the end of the corresponding branches.

- *For example:*

In choosing a suitable car, options can be as outlined as below:



In this example, available options for cars can be Manual 1000cc grey, Manual 1800cc, Hybrid 1000cc black or Hybrid 1800cc

Counting Principle:

Another way to reach to the possible outcome is by way of fundamental counting principle or counting rule. It says that when there are **m** ways of doing something and **n** ways of doing another and these are independent of each other then there are **m × n** ways of doing both.

Similarly, if there are more than two events **m, n** and **p**, then number of ways that all events can occur is **m × n × p.**

Two events being independent of each other means that one event does not impact the other.

- *For example:*

A lady has 10 pairs of shoes and 12 dresses.

There are 10 × 12 = 120 possible combinations.

A car manufacturer sells a model of car with the following options:

Body style	Sedan, hatchback, soft top	3
Colours	Choice of 15	15
Internal trim	3 colours of leather	3
Engine size	1.8l, 2.2l and 2.5l turbo	3

The car manufacturer is offering to supply 405 (3 × 15 × 3 × 3) variations of the car

Furthermore, additive principle of counting says that if **m** and **n** are number of ways two mutually exclusive events occur, then the possibility that either **m** or **n** occur is given by **m + n.** The same principle applies to two or more events.

- *For example:*

A lady has 6 flowers in a basket, 2 red, 2 white and 2 pink.

Two possible events that can occur are

3 chances that lady picks up either of the colored flower from her basket and

9 chances that either of the two colored flowers are being picked.

Hence, the possible number of time when either one or two colored flowers are picked is 3 + 9 = 12.

Permutations

Permutation is a set of several possible ways in which a set or number of things can be ordered or arranged. It is referred to as an ordered combination or a group of items where order does matter.

- *Formula:*

$$n!$$

Where:

n = number of items

! = A mathematical operator expressed as "factorial" (e.g. 6! is pronounced 6 factorial).

It is an instruction to multiply all the numbers counting down from the number given. i.e. 6! = 6 × 5 × 4 × 3 × 2 × 1.

Remember this: 0! = 1

- *For example:*

There are six ways of arranging three letters (A, B, C):

ABC, ACB, BAC, BCA, CAB, ABA

Using the term above:

Number of ways of arranging 3 items =3! = 3×2×1 = 6

How many ways can five different ornaments be arranged on a shelf?

n! = 5! = 5×4×3×2×1 = 120

Qadir has 4 sisters who ring him once every Saturday. How many possible arrangements of phone calls could there be?

$n! = 4! = 4 \times 3 \times 2 \times 1 = 24$

How many ways can 8 objects be arranged?

$n! = 8! = 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 40{,}320$

Permutations where some items are identical or repeated

The formula for computing permutations must be modified where some items are identical.

- *Formula:*

$$\frac{n!}{x!}$$

Where:

n = number of items

x = number of identical items or repetitions.

- *For example:*

How many possible arrangements are there of three letters ABB:

Since there are three letters and two of them are identical, the possible arrangement can be calculated as:

$$\frac{n!}{x!} = \frac{3!}{2!} = \frac{3 \times 2 \times 1}{2 \times 1} = \frac{6}{2} = 3$$

How many different ways can 6 objects be arranged if 3 of them are identical?

n = 6 and x= 3

$$\frac{n!}{x!} = \frac{6!}{3!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1} = 6 \times 5 \times 4 = 120$$

How many different ways can 7 objects be arranged if 4 of them are identical?

n = 7 and x= 4

$$\frac{n!}{x!} = \frac{7!}{4!} = \frac{7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{4 \times 3 \times 2 \times 1} = 7 \times 6 \times 5 = 210$$

How many ways can the letters in PAKISTAN be arranged?

In the word, PAKISTAN, letter A is repeated twice. Which makes n = 8 and x= 2

$$\frac{n!}{x!} = \frac{8!}{2!} = \frac{8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{2 \times 1} = 20{,}160$$

How many ways can the letters in COMMITMENT be arranged?

In the word, COMMITMENT, letter m is repeated thrice and t is repeated twice. Which makes n = 10 and x_i= 3 and x_{ii}= 2

$$\frac{n!}{x_i!.x_{ii}!} = \frac{10!}{3!.2!} = \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{(3 \times 2 \times 1).(2 \times 1)} = \frac{3{,}628{,}800}{12} = 302{,}400$$

How many different permutations are possible of 7 identical red books and 5 identical green books on a shelf?

In this example, n = 12, x_i= 5 and x_{ii}= 7

$$\frac{n!}{x_i!.x_{ii}!} = \frac{12!}{5!.7!} = \frac{12 \times 11 \times 10 \times 9 \times 8}{5 \times 4 \times 3 \times 2 \times 1} = \frac{95{,}040}{120} = 792$$

Permutations of size r from n items:

For a number of permutation of n items or elements taken r at a time or for size r, and when repetitions are not allowed, following formula would be used. This means that item picked next would be randomly selected from the remaining number of items.

- *Formula:*

$$_n\mathrm{P}_r = \frac{n!}{(n-r)!}$$

Where:

n = number of items

r = size of permutations.

- *For example:*

There are 9 books on a shelf. The librarian will log three of these first in order. How many ways does librarian have?

There are 9 books, but spots are 3.

$${}_{9}P_{3} = \frac{9!}{(9-3)!}$$

$${}_{9}P_{3} = \frac{9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{6 \times 5 \times 4 \times 3 \times 2 \times 1}$$

$${}_{9}P_{3} = 9 \times 8 \times 7 = 504$$

There are 10 balls in a bag. Each ball carries a different number from 1 to 10.

3 balls are selected at random from the bag and not replaced before the next draw is made.

Any one of 10 balls could be selected in the first draw followed by any one of the remaining 9 on the second draw and any one of the remaining 8 on the third draw.

$${}_{10}P_{3} = \frac{10!}{(10-3)!}$$

The number of possible permutations = 10 × 9 × 8 = 720

6 friends visit a cinema but can only buy 4 seats together. How many ways could the 4 seats be filled by the six friends?

$${}_{6}P_{4} = \frac{6!}{(6-4)!}$$

The number of possible permutations = 6 × 5 × 4 × 3 = 360

There are 10 balls in a bag. Each ball carries a different number from 1 to 10.

The number on each ball selected is noted and then the ball is replaced in the bag. How many different permutations are possible?

This example is different from the above scenarios. Here, any one of 10 balls could be selected in the first, second and third draws.

The number of possible permutations = 10 × 10 × 10 = 10^3 =1,000

A military code is constructed where each letter is represented by 4 randomly selected numbers in the sequence 1 to 50. How many permutations are possible for a single letter?

This example too, any one of numbers from 1-50 balls could be selected in the first, second, third and fourth number for the code.

The number of possible permutations = 50 × 50 × 50 × 50 = 50^4 =6,250,000

Combinations

The number of possible combinations of a number of items selected from a larger group is the number of possible permutations of the number of items selected from the group divided by the number of possible permutations of each sample.

Combination is referred to as a group of items chosen from a larger number without regard for their order. It is a group of items where order does not matter.

The formula below is also referred to as binomial coefficient.

- *Formula:*

$$nCx = \binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

Where:

$\binom{n}{x}$= a term indicating the number of permutations of n items of only two types where x is the number of one type.

n = total number of items from which a selection is made.

⍰ = number of identical objects of one type.

n – ⍰ = number of second type Identical object

- *For example:*

There are 10 balls in a bag. Each ball carries a different number from 1 to 10.

3 balls are selected at random from the bag and not replaced before the next draw is made. Where order does not matter, the number of possible combinations can be calculated as follows:

$$_{10}C_3 = \binom{10}{3} = \frac{10!}{3!\,(10-3)!}$$

$$= \frac{3{,}628{,}800}{6\ \times 5040} = \frac{3{,}628{,}800}{30{,}240} = 120$$

A university hockey team has a squad of 15 players. The coach carries out a strict policy that every player is chosen randomly. How many possible teams of 11 can be picked?

$$_{15}C_{11} = \binom{15}{11} = \frac{15!}{11!\,(15-11)!}$$

$$= \frac{15!}{11!\,(4!)} = \frac{15 \times 14 \times 13 \times 12}{4 \times 3 \times 2 \times 1} = \frac{32{,}760}{24} = 1{,}365$$

Five students from an honours class of 10 are to be chosen to attend a competition. One must be the head boy. How many possible teams are there?

Head boy must be included with 4 out of the remaining 9

$$_{9}C_4 = \binom{9}{4} = \frac{9!}{4!\,(9-4)!}$$

$$= \frac{9 \times 8 \times 7 \times 6}{4 \times 3 \times 2 \times 1} = \frac{3{,}024}{24} = 126$$

2. PROBABILITY

Probability is a measure of the likelihood of something happening. It is normally expressed as a number between 0 and 1 (or a percentage between 0% and 100%) where 0 indicates something cannot occur and 1 that it is certain to occur.

The probability of an event occurring plus the probability of that event not occurring must equal 1 (as it is certain that one or the other must happen).

In using the probability, an outcome is something that happens as a result of an event. An event is a group of one or more outcomes. A single attempt to obtain a given event is referred to as a trial.

- *For example:*

  If a dice is rolled each of the six numbers are outcomes.

  If a dice is rolled a possible event might be rolling an even number and this corresponds to the outcomes 2, 4 and 6.

When all possible outcomes are equally likely the probability an event can be estimated as follows:

- *Formula:*

$$\text{P (event)} = \frac{\text{Number of outcomes where the event occurs}}{\text{Total number of possible outcomes}}$$

  Where:

  P (event) = probability of an event occurring.

- *For example:*

  In tossing a coin, the number of possible outcomes are two head or a tail.

  Probability of rolling a head would be:

$$\text{P(head)} = \frac{1}{2} = 0.5 \text{ (or 50\%)}$$

  Similarly, in rolling a dice, the number of possible outcomes are 6 (1 to 6 numbers)

  Probability of rolling a 5 would be:

$$\text{P(5)} = \frac{1}{6} = 0.1667\text{or } 16.67\%$$

  Probability of rolling an even number would be:

$$\text{P(3)} = \frac{3}{6} = 0.5 \text{ or } 50\%$$

  A box contains 10 pens (4 Blue, 2 Red, 3 Black and 1 Green).

  Probability of picking a Blue pen would be:

$$\boldsymbol{P(blue\ pen) = \frac{4}{10} = 0.4\ or\ 40\%}$$

  Probability of not picking a black pen would be:

$$\boldsymbol{P(not\ black\ pen) = \frac{7}{10} = 0.7\ or\ 70\%}$$

Relative frequency

The circumstance under investigation might be more complexed than rolling a dice, all outcomes might not be equally likely.

Relative frequency is the probability of an event happening found by experiment or observation.

- *Formula:*

$$\text{Relative frequency} = \frac{\text{Number of trials where the event occured}}{\text{Total number of trials carried out}}$$

- *For example:*

The following table has been constructed from observing daily sales over a period

Daily sales	Number of days	
0	20	$P(\text{sales} = 0) = \frac{20}{100} = 0.2$ or 20%
1	30	$P(\text{sales} = 1) = \frac{30}{100} = 0.3$ or 30%
2	25	$P(\text{sales} = 2) = \frac{25}{100} = 0.25$ or 25%
3	15	$P(\text{sales} = 3) = \frac{15}{100} = 0.15$ or 15%
4	10	$P(\text{sales} = 4) = \frac{10}{100} = 0.1$ or 10%
	100	

Events that are mutually exclusive or independent

Two events are independent, if the outcome of one does not affect the outcome of the other. Two events are mutually exclusive, if the occurrence of one prevents the occurrence of the other in a single observation and there are no common outcomes. If there are one or more common outcomes, then the events are overlapping.

- *For example:*

If a dice is rolled twice, the number on the second roll is independent of the number on the first roll.

If a dice is rolled than an outcome being an even or odd is mutually exclusive.

Probability for either one or the other event to occur can be measured as follows:

- *Formula:*

If A and B are any two (non-mutually exclusive) events, then the probability of A or B is:

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$or

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

If A and B are two mutually exclusive events, then the probability of either A or B is

$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B)$

If A and B are two independent events, then the probability of both occurring in succession is

$P(A \text{ and } B) = P(A \cap B) = P(A) \times P(B)$

If A and B are two dependent, then the probability of both occurring in succession is

$P(A \text{ and } B) = P(A \cap B) = P(A) \times P(B \text{ I } A)$

If A is an event, then probability for all outcomes not in A is

$P(\overline{A}) = 1 - P(A)$

Here, $\overline{A}$ is called Complement of event A,

- *For example:*

What is the probability of rolling a 3 or a 4 when rolling a dice once?

$$P(\text{A or B}) = P(A) + P(B)$$

$$P(\text{3 or 4}) = P(3) + P(4) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3} = 0.333 \text{ or } 33.33\%$$

If two dice are rolled what is the probability of rolling two 6s?

$$P(\text{A and B}) = P(A) \times P(B)$$

$$P(\text{6 and 6}) = P(6) \times P(6) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} = 0.0278 \text{ or } 2.78\%$$

What is the probability of selecting a jack or a queen from a normal pack of 52 cards?

P (Jack or Queen) = P (J or Q) = $4/52 + 4/52 = 8/52 = 2/13$

What is the probability of selecting a club or a spade from a normal pack of 52 cards?

P (Club or Spade) = P (C or S) = $13/52 + 13/52 = 26/52 = 1/2$

What is the probability of selecting an ace or a heart from a normal pack of 52 cards?

P (Ace or Heart) = P (A or H) – P(Ace of hearts) = 4/52 + 13/52 – 1/52 = 16/52

2 Dice are rolled - 1 red and 1 blue. What is the probability of rolling a 6 on the red or blue dice?

P (6 on red or 6 on blue) = P (6R or 6B) = $1/6 + 1/6 - [1/6 \times 1/6] = 6/36 + 6/36 - 1/36 = 11/36$

A box contains 10 (4 Blue, 2 Red, 3 Black and 1 Green), what is the probability that a pen is blue or black?

$$P(\text{blue or black}) = \frac{4}{10} + \frac{3}{10} = \frac{7}{10} = 0.7 \text{ or } 70\%$$

Daily sales of a firm and probability of achieving the sales is provided below

Daily sales (number of items)	P
0	0.2
1	0.3
2	0.25
3	0.15
4	0.10
	1.0

The probability of daily sales of more than 2 would be:

$$P(\text{sales more than 2}) = P(\text{3 or 4}) = 0.15 + 0.1 = 0.25 \text{ or } 25\%$$

The probability of daily sales being at least 1 would be:

$$P(\text{at least 1}) = P(1, 2, 3 \text{ or } 4) = 0.3 + 0.25 + 0.15 + 0.1 = 0.8 \text{ or } 80\%$$

Another approach is to rearrange the statement to the probability of sales not being zero.

$$P(\text{not } 0) = 1 - P(0) = 1 - 0.2 = 0.8 \text{ or } 80\%$$

AT A GLANCE

SPOTLIGHT

STICKY NOTES

The probability of total sales of 6 in two consecutive days would require combination of techniques:

There are three combinations of days which would result in sales of 6 over two days. The probability of each is found using the multiplication law and then these are added

	Day 1		Day 2				
	4	and	2		0.10×0.25	=	0.0250
or	2	and	4	+	0.25×0.10	=	0.0250
or	3	and	3	+	0.15×0.15	=	0.0225
							0.0725

An item is made in three stages. At the first stage, it is formed on one of four machines, A, B, C, or D with equal probability. At the second stage it is trimmed on one of three machines, E, F, or G, with equal probability. Finally, it is polished on one of two polishers, H and I, and is twice as likely to be polished on the former as this machine works twice as quickly as the other.

First lets re-write the steps and probabilities in tabular form:

Formed	Trimmed	Polished
A ¼	E ⅓	H ⅔
B ¼	F ⅓	I ⅓
C ¼	G ⅓	
D ¼		

a) The probability that an item is polished on H is:

P(H) = 2/3

b) The probability that the item is trimmed on either F or G is:

P(F or G) = 1/3 + 1/3 = 2/3

c) The probability that the item is formed on either A or B, trimmed on F and polished on H is:

P(A or B) and P(F) and P(H) = (¼ + ¼) × 1/3 × 2/3 = 1/9

d) The probability that the item is either formed on A and polished on I, or formed on B and polished on H is:

P(A and I) or P(B and H) = (¼ × 1/3) + (¼ × 2/3) = ¼

e) The probability that the item is either formed on A or trimmed on F is:

Since A, F are not mutually exclusive:

P(A or F) = P(A) + P(F) – (P(A) × P(F)) = ¼ + 1/3 – (¼ × 1/3) = ½

The chances of survival of two accident victims are 0.3 and 0.6. What is the probability that both of them will survive?

$$\mathbf{P(A\ and\ B) = P(A) \times P(B) = 0.3 \times 0.6 = 0.18\ or\ 18\%}$$

Three dice are rolled. What is the probability of getting at least one six?

$$\mathbf{P(at\ least\ one\ six) = 1 - P(no\ sixes)}$$

$$\mathbf{P(no\ sixes) = \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} = \frac{125}{216}}$$

$$\mathbf{P(at\ least\ one\ six) = 1 - \frac{125}{216} = \frac{91}{216} = 0.4213\ or\ 42.13\%}$$

Three people A, B and C get into a lift on the ground floor of a building. The building has 6 floors, comprising of the ground floor and floors 1 to 5. The probability that a person entering the lift at the ground floor to a required floor is as follows:

Floor required	Probability
1	0.1
2	0.2
3	0.2
4	0.3
5	0.2

a) The probability that A requires either floor 3, 4 or 5.

P(3, 4 or 5) = 0.2 + 0.3 + 0.2 = 0.7

b) The probability that neither A, B, nor C require floor 1.

P(A does not require first floor, B does not require first floor and C does not require first floor) = 0.9 × 0.9 × 0.9 = 0.729

c) The probability that A and B require the same floor.

A	B		
1 and 1	0.1^2	=	0.01
2 and 2	$+0.2^2$		0.04
3 and 3	$+0.2^2$		0.04
4 and 4	$+0.3^2$		0.09
5 and 5	$+0.2^2$		0.04
			0.22

d) The probability that A and B require different floors.

P (different floors) = 1 – P(same floors)

= 1 – 0.22

= 0.78

e) The probability that only one of A, B, and C requires floor 2.

A	B	C		
2	x	x	0.2 × 0.8 × 0.8 =	0.128
x	2	x	0.8 × 0.2 × 0.8 =	0.128
x	x	2	0.8 × 0.8 × 0.2 =	0.128
				0.384

f) The probability that A travels three floors further than B.

	A	B			
	5 and	2	0.2 × 0.2	=	0.04
or	4 and	1	0.3 × 0.1	=	0.03
					0.07

AT A GLANCE

SPOTLIGHT

STICKY NOTES

AT A GLANCE

SPOTLIGHT

STICKY NOTES

Conditional probability:

Conditional probability is the probability of an event occurring given that another event has occurred (probability of one event is dependent on another).

If A and B are two events such that the probability of B is dependent on the probability of A then the probability of A and B occurring is the product of the probability of A and the conditional probability of B.

- *Formula:*

$$P(B|A) = \frac{P\ (A\ and\ B)}{P(A)} = \frac{P\ (A\ \cap\ B)}{P(A)} \quad or \qquad P\ (A\ and\ B) = P\ (A\ \cap\ B) = P(A) \times\ P(B|A) \times$$

Where:

P (B|A) = probability of B occurring given that A has occurred.

- *For example:*

There are 4 white and 5 blue pens. What is the probability that two white pens are drawn respectively if none of the pens are replaced?

Let A and B be the events of drawing pens in two simultaneous pick.

$$\text{P (A and B)} = \text{P(A)}\ \times\ \text{P(B|A)} \times$$

$$\text{P (A and B)} = \frac{4}{9} \times \frac{3}{8} = = \frac{1}{6}$$

A person is dealt two cards from a pack. What is the probability that both are aces?

There are 4 aces in a pack of 52 cards.

The probability of the first card being an ace is 4/52.

If the first card is an ace there are 3 aces left in 51 cards so the probability of being dealt a second ace if the first card was an ace is 3/51.

Therefore, the probability that both are aces is:

$$\text{P (A and B)} = \text{P(A)} \times \text{P(B|A)} = \frac{4}{52} \times \frac{3}{51} = \frac{12}{2{,}652} = 0.0045 \text{ or } 0.45\%$$

A box contains 4 Blue, 2 Red, 3 Black and 1 Green pen. What is the probability that both pens are blue (a)if the first pen is replaced prior to selecting the second? (b) The first pen is not replaced prior to selecting the second?

(a) If the pens are replaced, the probability of blue pens being drawn both the time is:

$$\text{P (A and B)} = \text{P(A)} \times \text{P(B|}\quad) = \frac{4}{10} \times \frac{4}{10} = \frac{16}{100} = 0.16 \text{ or } 16\%$$

(b) If the pens are not replaced, the probability of blue pens being drawn both the time is:

$$\text{P (A and B)} = \text{P(A)} \times \text{P(B|A)} = \frac{4}{10} \times \frac{3}{9} = \frac{12}{90} = 0.1333 \text{ or } 13.33\%$$

Sindh Export Limited – Employee report is given as follows:

	Male		Female		
Department	**Under 18**	**18 and over**	**Under 18**	**18 and over**	**Total**
Senior staff	0	15	0	5	20
Junior staff	6	4	8	12	30
Transport	2	34	2	2	40
Stores	1	20	2	17	40
Maintenance	1	3	4	12	20
Production	0	4	24	72	100
Total	10	80	40	120	250

(a) All employees are included in a weekly draw for a prize. What is the probability that in one particular week the winner will be:

i. Male:

$$P\text{ (male)} = \frac{10+80}{250} = \frac{90}{250} = 0.36 \text{ or } 36\%$$

ii. a member of junior staff

$$P\text{ (junior staff)} = \frac{30}{250} = \frac{3}{25} = 0.12 \text{ or } 12\%$$

iii. a female member of staff

$$P\text{ (Female)} = \frac{40+120}{250} = \frac{160}{250} = 0.64 \text{ or } 64\%$$

iv. from the maintenance or production departments

Since the events are mutually exclusive:

$$P\text{ (maintenance or production)} = \frac{20}{250} + \frac{100}{250} = \frac{120}{250} = 0.48 \text{ or } 48\%$$

v. a male or transport worker?

Since events are not mutually exclusive

$$P\text{ (male or transport)} = \frac{90}{250} + \frac{40}{250} - \frac{2+34}{250} = \frac{94}{250} = 0.376 \text{ or } 37.6\%$$

(b) Two employees are to be selected at random to represent the company at the annual Packaging Exhibition in London. What is the probability that:

i. both are female

$$P\text{ (F, F)} = \frac{160}{250} \times \frac{159}{249} = \frac{25{,}440}{62{,}250} = 0.409 \text{ or } 40.9\%$$

ii. both are production workers

$$P\text{ (P, P)} = \frac{100}{250} \times \frac{99}{249} = \frac{99{,}00}{62{,}250} = 0.159 \text{ or } 15.9\%$$

iii. one is a male under 18 and the other a female 18 and over?

$$P\text{ (M} < 18, \text{F} > 18) = \frac{10}{250} \times \frac{120}{249} + \frac{120}{250} \times \frac{10}{249} = \frac{1{,}200 + 1{,}200}{62{,}250} = 0.038 \text{ or } 3.8\%$$

(c) If three young employees under 18 are chosen at random for the opportunity to go on a residential course, what is the probability that:

i. all are male

$$P\ (M, M, M) = \frac{10}{50} \times \frac{9}{49} \times \frac{8}{48} = \frac{720}{117{,}600} = 0.006 \text{ or } 0.6\%$$

ii. all are female

$$P\ (F, F, F) = \frac{40}{50} \times \frac{39}{49} \times \frac{38}{48} = \frac{59{,}280}{117{,}600} = 0.504 \text{ or } 50.4\%$$

iii. only one is male?

$$P\ (M, F, F \text{ or } F, M, F \text{ or } F, F, M) = \frac{10}{50} \times \frac{40}{49} \times \frac{39}{48} + \frac{40}{50} \times \frac{10}{49} \times \frac{39}{48} + \frac{40}{50} \times \frac{39}{49} \times \frac{10}{48}$$

$$= \frac{15{,}600}{117{,}600} + \frac{15{,}600}{117{,}600} + \frac{15{,}600}{117{,}600} = 0.398 \text{ or } 39.8\%$$

Alternate methods for conditional probability:

Problems involving conditional probability can be quite complex so methods are needed to keep track of the information and understand the links between the probabilities.

Two such methods are:

- Table of probabilities – A table which sets out the probabilities of different combinations of possibilities.
- Contingency tables – Similar to a probability table but multiplying the probabilities by a convenient number to help visualise the outcomes.
- Probability trees – A diagram of links between different possible probabilities.

► *For example:*

A large batch of items comprises some manufactured by process X and some by process Y.

There are twice as many items from X as from Y in a batch.

Items from process X contain 9% defectives and those from Y contain 12% defectives.

(a) Calculate the proportion of defective items in the batch.

(b) If an item is taken at random from the batch and found to be defective. What is the probability that it came from Y?

► *Method 1: Probability table*

	Process X ($^2/_3$)	**Process Y ($^1/_3$)**
Defective items (X = 0.09/Y =0.12)	$\frac{2}{3} \times 0.09 = 0.06$	$\frac{1}{3} \times 0.12 = 0.04$
Good items (X = 0.91/Y =0.88)	$\frac{2}{3} \times 0.91 = 0.61$	$\frac{1}{3} \times 0.88 = 0.29$

(a) Proportion of defective items in a batch

P(defective) = 0.06+ 0.04 =0.10 or 10%

(b) Probability that a defective item is from process Y

$$P\ (\text{event}) = \frac{\text{Number of outcomes where the event occurs}}{\text{Total number of possible outcomes}}$$

$$P\ (event) = \frac{0.04}{0.06 + 0.04} = \frac{0.04}{0.1} = 0.4 \ or\ 40\%$$

- *Method 2: Contingency table*

Let the total batch size = 300

	Process X ($^2/_3$)	Process Y ($^1/_3$)	Total
Defective items (X = 0.09/Y =0.12)	18 (0.06 × 300)	12 (0.04 × 300)	30
Good items (balance)	182	88	270
	200	100	300

(a) Proportion of defective items in a batch

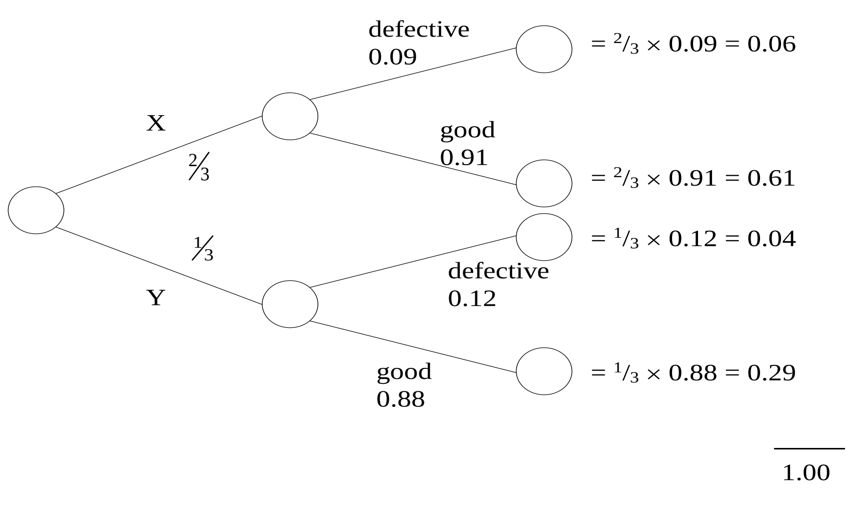
$$P(defective) = \frac{30}{300} = 0.10 \text{ } or \text{ } 10\%$$

(b) Probability that a defective item is from process Y

$$P\ (event) = \frac{Number\ of\ outcomes\ where\ the\ event\ occurs}{Total\ number\ of\ possible\ outcomes}$$

$$P\ (event) = \frac{12}{30} = 0.4 \text{ } or \text{ } 40\%$$

- *Method 3: Probability tree*



(a) Proportion of defective items in a batch

From the tree = 0.06 + 0.04 = 0.1

(b) Probability that a defective item is from process Y

$$\text{P (event)} = \frac{\text{Number of outcomes where the event occurs}}{\text{Total number of possible outcomes}}$$

$$P\ (event) = \frac{0.04}{0.06 + 0.04} = \frac{0.04}{0.1} = 0.4 \text{ } or \text{ } 40\%$$

STICKY NOTES

Fundamental counting principle or counting rule says that when there are *m* ways of doing something and *n* ways of doing another and these are independent of each other then there are m × n ways of doing both.

Permutation is a set of several possible ways in which a set or number of things can be ordered or arranged. It is referred to as an ordered combination or a group of items where order does matter. Mathematically, $n!$

The formula for computing permutations must be modified where some items are identical is $\frac{n!}{x!}$

For a number of permutation of n items or elements taken r at a time or for size r, following formula would be used. ${}_{n}P_{r} = \frac{n!}{(n-r)!}$

Combination is referred to as a group of items chosen from a larger number without regard for their order. It is a group of items where order does not matter. $nCx = \binom{n}{x} = \frac{n!}{x!(n-x)!}$

Probability is a measure of the likelihood of something happening. It is normally expressed as a number between 0 and 1 (or a percentage between 0% and 100%) where 0 indicates something cannot occur and 1 that it is certain to occur.

$$P\ (\text{event}) = \frac{\text{Number of outcomes where the event occurs}}{\text{Total number of possible outcomes}}$$

If A and B are any two (non-mutually exclusive) events, then the probability of A or B is:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) \text{or}$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If A and B are two mutually exclusive events, then the probability of either A or B is

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B)$$

AT A GLANCE

SPOTLIGHT

STICKY NOTES

If A and B are two independent events, then the probability of both occurring in succession is

$$P(\text{A and B}) = P(A \cap B) = P(A) \times P(B)$$

If A and B are two dependents, then the probability of both occurring in succession is

$$P(\text{A and B}) = P(A \cap B) = P(A) \times P(B\ I\ A)$$

If A is an event, then probability for all outcomes not in A is

$$P(\overline{A}) = 1 - P(A)$$

Conditional probability is the probability of an event occurring given that another event has occurred (probability of one event is dependent on another). If A and B are two events such that the probability of B is dependent on the probability of A then the probability of A and B occurring is the product of the probability of A and the conditional probability of B.

$$P(B|A) = \frac{P\ (\text{A and B})}{P(A)} = \frac{P\ (A \cap B)}{P(A)}$$

AT A GLANCE

SPOTLIGHT

STICKY NOTES

SELF-TEST

11.1. The set of all possible outcomes of a statistical experiment is called:

(a) Data (b) Statistical data

(c) Sample space (d) None of these

11.2. Something which can take different values with different probabilities is called:

(a) A variable (b) A random variable

(c) A discreet variable (d) A continuous variable

11.3. If two events have no elements in common then they are known as:

(a) Union of two events (b) Intersection of two events

(c) Mutually exclusive events (d) None of these

11.4. If two events have at least one element in common then the events are said to be:

(a) Union of two events (b) Intersection of two events

(c) Not mutually exclusive events (d) None of these

11.5. If three coins are tossed, then the number of outcomes are:

(a) 6 (b) 8

(c) 9 (d) None of these

11.6. If three dice are rolled; then the number of outcomes are:

(a) 18 (b) 216

(c) 36 (d) None of these

11.7. If an experiment consists of rolling a dice and then flipping a coin, the number of outcomes are:

(a) 36 (b) 8

(c) 12 (d) None of these

11.8. Which of the following applies to discreet as well as continuous variables:

(a) Binomial distribution (b) Normal Distribution

(c) Chi Square distribution (d) (b) and (c)

11.9. The arrangement of some or all of the objects without considering the order of arrangement is called:

(a) Selection (b) Permutation

(c) Combination (d) None of these

AT A GLANCE SPOTLIGHT STICKY NOTES

11.10. If ${}^{10}P_3 = 720$ then ${}^{10}C_3 = ?$

(a) 120 (b) 2,160

(c) 720 (d) None of these

11.11. If ${}^{n}P_2 = 20$ then the value of n = ?

(a) 10 (b) 5

(c) 8 (d) None of these

11.12. If 4 coins are tossed simultaneously the number of outcomes containing at least one head are:

(a) 15 (b) 1

(c) 8 (d) None of these

11.13. If 4 coins are tossed simultaneously, the number of outcomes containing at least two tails are:

(a) 11 (b) 15

(c) 9 (d) None of these

11.14. If ${}^{5}C_r = 10$, then r = ?

(a) 2 (b) 5

(c) 4 (d) None of these

11.15. If the experiment, tossing a pair of dice, is repeated over and over again in a very long series of trials, what proportion of outcomes do you think would result in a sum less than 7?

(a) $\frac{5}{12}$ (b) $\frac{6}{36}$

(c) $\frac{6}{12}$ (d) None of these

11.16. If the experiment, tossing a pair of dice, is repeated over and over again in a very long series of trials, what proportion of outcomes do you think would result in a sum equal to 7?

(a) $\frac{5}{12}$ (b) $\frac{6}{36}$

(c) $\frac{6}{12}$ (d) None of these

11.17. If the experiment, tossing a pair of dice, is repeated over and over again in a very long series of trials, what proportion of outcomes do you think would result in a sum more than 7?

(a) $\frac{5}{12}$ (b) $\frac{6}{36}$

(c) $\frac{6}{12}$ (d) None of these

11.18. If probability of an event A is 0.2 and the probability of an event B is 0.3 and the probability of either A or B is 0.5, then A and B are:

(a) Mutually exclusive
(b) Independent
(c) Dependent
(d) None of these

11.19. If a coin is tossed four times, the probability of four heads is equal to:

(a) 0.25
(b) 0.0625
(c) 0.625
(d) Zero

11.20. When all events have the same chance of occurrence, the events are said to be:

(a) Equally likely
(b) Independent
(c) Dependent
(d) None of these

11.21. If two or more events cannot occur simultaneously then they are called:

(a) Independent
(b) Mutually exclusive
(c) Not equally likely
(d) None of these

11.22. If the probability of an event is 0.01, which of the following statements is correct:

(a) The event is unlikely to occur
(b) The event is expected to occur about 10% of the time
(c) The event cannot occur
(d) None of the above

11.23. If a coin is tossed five times, the probability of:

(a) Five heads is equal to the probability of five tails
(b) Five heads is equal to the probability of zero tails
(c) Five heads is equal to the probability of zero heads
(d) All of the above

11.24. The probability of two mutually exclusive events is:

(a) $P(A \cup B)$
(b) $P(\mathrm{A} \cap \mathrm{B})$
(c) $P(A) + P(B)$
(d) None of these

11.25. The probability of two non-mutually exclusive events is:

(a) $P(A \cup B)$
(b) $P(\mathrm{A} \cap \mathrm{B})$
(c) $P(A) + P(B) - P(A \cap B)$
(d) None of these

11.26. The probability of two independent events is:

(a)	$P(A \cap B)$	(b)	$P(A) \times P(B)$
(c)	$P(A) \times P(B/A)$	(d)	None of these

11.27. The probability of two non-independent events if the event A occurs first is:

(a)	$P(A \cap B)$	(b)	$P(A) \times P(B)$
(c)	$P(A) \times P(B/A)$	(d)	None of these

11.28. The sum of the probabilities of two compliment events is always:

(a)	Zero	(b)	1
(c)	Does not exist	(d)	None of these

11.29. The probability of an impossible event is always:

(a)	Equal to zero	(b)	Less than zero
(c)	More than zero	(d)	None of these

11.30. The probability of a certain event is always:

(a)	Equal to zero	(b)	Equal to one
(c)	More than one	(d)	None of these

11.31. If P(A) = 0.3 and P(B) = +1.3 then $P(A \cup B)$ is, when A and B are mutually exclusive events:

(a)	1	(b)	-1.6
(c)	+1.6	(d)	None of these

11.32. Three horses A, B and C are in a race; A is twice as likely to wins as B and B is three times as likely to wins as C. The probability that A wins is:

(a)	$\frac{3}{5}$	(b)	$\frac{3}{7}$
(c)	$\frac{6}{7}$	(d)	None of these

11.33. Three horses A, B and C are in a race; A is twice as likely to wins as B and B is three times as likely to wins as C. The probability that B or C wins is:

(a)	$\frac{3}{5}$	(b)	$\frac{2}{5}$
(c)	$\frac{3}{10}$	(d)	None of these

AT A GLANCE

SPOTLIGHT

STICKY NOTES

AT A GLANCE SPOTLIGHT STICKY NOTES

11.34. Three horses A, B and C are in a race; A is twice as likely to wins as B and B is three times as likely to wins as C. The probability that A or B wins is:

(a) $\frac{3}{5}$ (b) $\frac{9}{10}$

(c) $\frac{7}{10}$ (d) None of these

11.35. If P(A) = 1 – P(B); then A and B are:

(a) Independent events (b) Compliment events

(c) Mutually exclusive events (d) None of these

11.36. A person invests in three stocks. After 12 months, he records the gain or loss in price as follows:

A: All three stocks rise in price

B: Stock1 rise in price

C: stock 2 experiences a Rs. 5 drop in price.

Therefore A and B are:

(a) Not mutually exclusive events (b) Mutually exclusive events

(c) Independent (d) None of these

11.37. A person invests in three stocks and after 12 months, he records the gain or loss in price as:

A: all three stocks rise in price

B: Stock 1 rises in price

C: stock 2 experiences a Rs. 5 Drop in price.

Therefore A and C are:

(a) Not mutually exclusive events (b) Mutually exclusive events

(c) Independent (d) None of these

11.38. The probability of an event A, given that an event B has occurred, is denoted as:

(a) $P(A \cap B)$ (b) $P(A\backslash B)$

(c) $P(B\backslash A)$ (d) None of these

11.39. From a deck of 52 cards, two cards are drawn in succession. The probability that both cards are spades is:

(a) 0.0588 (b) 0.0625

(c) 0.25 (d) None of these

11.40. From a deck of 52 cards, two cards are drawn in succession. The number of elements the sample space contains is:

(a) 2,704 (b) 1,326

(c) 2,652 (d) None of these

11.41. From a deck of 52 cards, two cards are drawn in succession. All the outcomes of the experiment are:

(a) Independent (b) Dependent

(c) Mutually exclusive (d) None of these

11.42. If one coin is tossed 7 times the number of possible outcomes would be:

(a) 128 (b) 14

(c) 49 (d) None of these

11.43. If seven coins are tossed simultaneously. All possible outcomes of the experiment are:

(a) Not mutually exclusive (b) Independent

(c) Not independent (d) None of these

11.44. In the game of craps, two dice are rolled. If the sum on the dice is 7 or 11 he wins, and if the sum is 2, 3 or 12, he loses. The probability of his winning is:

(a) $\frac{2}{9}$ (b) $\frac{1}{18}$

(c) $\frac{1}{6}$ (d) None of these

11.45. In the game of craps, two dice are rolled. If the sum on the dice is 7 or 11 he wins, and if the sum is 2, 3 or 12, he loses. The probability that he loses is:

(a) $\frac{1}{9}$ (b) $\frac{1}{36}$

(c) $\frac{1}{12}$ (d) None of these

11.46. Given P(A) = 0.24; P(B) = 0.39 And $P(A \cup B)$=0.49.

If A and B are not mutually exclusive events, then $P(A \cap B)$ is equal to:

(a) 0.49 (b) 0.63

(c) 0.14 (d) None of these

11.47. If P(A)= 0.5; P(B)=0.6 and $P(A \cap B) = 0.4$

then $P(A \cup B)$ is:

(a) 0.4 (b) 1.1

(c) 0.7 (d) None of these

AT A GLANCE
SPOTLIGHT
STICKY NOTES

11.48. If P(A) = 0.20; P(B) = 0.08 and P(C) = 0.03

where A = A person views a T.V. show

B = A person reads a magazine

C = A person views a T.V. show and reads a magazine

Then $P(A \cup B)$ is:

(a)	0.25	(b)	0.28
(c)	0.11	(d)	None of these

11.49. Events are said to be _____ if they have no elements in common

11.50. _____ is the set of all possible outcomes of an experiment

11.51. Select the correct statement(s) from the following

(a) Sample space is the set of all possible outcomes of an experiment

(b) The outcomes of an experiment are referred to as events

(c) Events are said to be non-mutually exclusive if they have no elements in common

(d) Events are said to be mutually exclusive if they have no elements in common

11.52. A pair of fair dice is tossed one time. What is the probability that the sum of two faces is a prime number?

(a)	A. 16/36	(b)	B. 15/36
(c)	C. 20/36	(d)	D. 21/36

11.53. A pair of fair dice is tossed one time. What is the probability that the sum of two faces is a multiple of 3 number?

(a)	A. 15/36	(b)	B. 12/36
(c)	C. 20/36	(d)	D. 10/36

11.54. Three dice are rolled once. What is the probability that the product of the faces is more than 215?

(a)	A. 3/216	(b)	B. 3
(c)	C. 6	(d)	D. 1/216

11.55. A coin is tossed twice and a die is rolled once, then the number of outcomes is:

(a)	A. 20	(b)	B. 10
(c)	C. 3	(d)	D. 24

11.56. A and B play four games of chess. A's chances of winning the game is 7/10. What is the probability that B wins at least one game?

(a)	A. 0.76	(b)	B. 0.80
(c)	C. 0.86	(d)	D. 0.66

11.57. A and B play four games of chess. A's chances of winning the game is 7/10. What is the probability that A wins first three games?

(a)	A. 0.25	(b)	B. 0.3
(c)	C. 0.32	(d)	D. 0.34

11.58. A and B play four games of chess. A's chances of winning the game is 7/10. What is the probability that A wins three games consecutively only and losses the remaining one?

(a)	A. 0.1	(b)	B. 0.2
(c)	C. 0.14	(d)	D. 0.24

11.59. Three students appeared in an examination with the following probability to pass the exam:

A: 9/10
B: 8/10
C: 7/10

What is the probability that none of the students will pass the exam?

(a)	A. 0.005	(b)	B. 0.006
(c)	C. 0.007	(d)	D. 0.009

11.60. Three students appeared in an examination with the following probability to pass the exam:

A: 9/10
B: 8/10
C: 7/10

What is the probability that at least one of them will pass the exam?

(a)	A. 0.884	(b)	B. 0.774
(c)	C. 0.994	(d)	D. 0.664

11.61. Three students appeared in an examination with the following probability to pass the exam:

A: 9/10
B: 8/10
C: 7/10

What is the probability that A and B will pass the exam only?

(a)	A. 0.046	(b)	B. 0.056
(c)	C. 0.036	(d)	D. 0.026

11.62. Three students appeared in an examination with the following probability to pass the exam:

A: 9/10
B: 8/10
C: 7/10

What is the probability that A and B will pass the exam?

(a)	A. 0.46	(b)	B. 0.56
(c)	C. 0.66	(d)	D. 0.76

11.63. The probability of rain on first three days of August is as follows:

Day	Probability of rain
1	3/10
2	If rain on Day 1: 5/10 If no rain on Day 1: 2/10
3	If rain on Day 1 and Day 2: 9/10 If rain on Day 1 only: 5/10 If rain on Day 2 only: 3/10 If no rain on Day 1 and Day 2: 5/10

What is the probability of rain on all three days

(a)	A. 0.145	(b)	B. 0.135
(c)	C. 0.155	(d)	D. 0.125

11.64. The probability of rain on first three days of August is as follows:

Day	Probability of rain
1	3/10
2	If rain on Day 1: 5/10 If no rain on Day 1: 2/10
3	If rain on Day 1 and Day 2: 9/10 If rain on Day 1 only: 5/10 If rain on Day 2 only: 3/10 If no rain on Day 1 and Day 2: 5/10

What is the probability of rain on second day only?

(a)	A. 0.088	(b)	B. 0.078
(c)	C. 0.098	(d)	D. 0.068

11.65. The probability of rain on first three days of August is as follows:

Day	Probability of rain
1	3/10
2	If rain on Day 1: 5/10 If no rain on Day 1: 2/10
3	If rain on Day 1 and Day 2: 9/10 If rain on Day 1 only: 5/10 If rain on Day 2 only: 3/10 If no rain on Day 1 and Day 2: 5/10

What is the probability of rain on first and third day?

(a) A. 0.065 (b) B. 0.075

(c) C. 0.085 (d) D. 0.095

11.66. The probability of rain on first three days of August is as follows:

Day	Probability of rain
1	3/10
2	If rain on Day 1: 5/10 If no rain on Day 1: 2/10
3	If rain on Day 1 and Day 2: 9/10 If rain on Day 1 only: 5/10 If rain on Day 2 only: 3/10 If no rain on Day 1 and Day 2: 5/10

What is the probability of rain on at least 2 out of 3 days?

(a) A. 0.367 (b) B. 0.267

(c) C. 0.467 (d) D. 0.567

11.67. A bag contains 6 red balls, 3 black balls and 2 blue balls. Two balls are selected with replacement. Which TWO of the following statements are correct?

(a) The probability that both balls are red is 36/121

(b) The probability that both balls are green is 0

(c) The probability that both balls are black is 6/121

(d) The probability that both balls are blue is 4/121

11.68. A bag contains 6 red balls, 3 black balls and 2 blue balls. Two balls are selected without replacement. Which TWO of the following statements are correct?

(a) The probability that first ball is red and second ball is black is 9/55

(b) The probability that first ball is red and second ball is black is 18/121

(c) The probability that both balls are red is 36/121

(d) The probability that neither ball is red is 25/121

ANSWERS TO SELF-TEST QUESTIONS

11.1	**11.2**	**11.3**	**11.4**	**11.5**	**11.6**
(c)	(a)	(c)	(c)	(b)	(b)
11.7	**11.8**	**11.9**	**11.10**	**11.11**	**11.12**
(c)	(b)	(c)	(a)	(b)	(a)
11.13	**11.14**	**11.15**	**11.16**	**11.17**	**11.18**
(a)	(a)	(a)	(b)	(a)	(a)
11.19	**11.20**	**11.21**	**11.22**	**11.23**	**11.24**
(b)	(a)	(b)	(a)	(d)	(c)
11.25	**11.26**	**11.27**	**11.28**	**11.29**	**11.30**
(c)	(b)	(c)	(b)	(a)	(b)
11.31	**11.32**	**11.33**	**11.34**	**11.35**	**11.36**
(d)	(a)	(b)	(b)	(b)	(a)
11.37	**11.38**	**11.39**	**11.40**	**11.41**	**11.42**
(b)	(b)	(a)	(c)	(b)	(a)
11.43	**11.44**	**11.45**	**11.46**	**11.47**	**11.48**
(b)	(a)	(a)	(c)	(c)	(a)
11.49	**11.50**	**11.51**	**11.52**	**11.53**	**11.54**
Mutually exclusive	Sample space	(a),(b), (d)	(b)	(b)	(d)
11.55	**11.56**	**11.57**	**11.58**	**11.59**	**11.60**
(d)	(a)	(d)	(c)	(b)	(c)
11.61	**11.62**	**11.63**	**11.64**	**11.65**	**11.66**
(b)	(b)	(b)	(c)	(b)	(b)
11.67	**11.68**				
(a),(b)	(a),(b)				

CHAPTER 12

PROBABILITY DISTRIBUTIONS

IN THIS CHAPTER

AT A GLANCE

There are number of distributions which can be applied to represent wider classes of problems. These are sometimes described as theoretical probability distributions. The term theoretical makes it sound as if these might not be of much use in practice, but nothing is further from the truth as they are very useful indeed.

This chapter discusses, various probability distributions to estimate the probability of a given result with known characteristics including Binomial Distribution (when there are fixed number of trials); Hyper-Geometric Distribution (when items are randomly selected and not replaced); Poisson Distribution (when events occur at a constant rate) and Normal Distribution (when variable are continuous).

1 PROBABILITY DISTRIBUTIONS

A probability distribution is similar to frequency distributions but it uses probabilities instead of frequencies. In probability frequency distribution, each frequency of each observation is divided by total frequency observations. Such that the total probability distribution equals 1.

► *For example:*

A sample of 300 employees has been taken from a large business.

Their monthly salaries were found to be as follows:

Monthly salary (Rs.000)	Frequency distribution		Probability distribution
25 to under 30	15	$^{15}/_{300}$ =	0.05
30 to under 35	24	$^{24}/_{300}$ =	0.08
35 to under 40	54	$^{54}/_{300}$ =	0.18
40 to under 45	90	$^{90}/_{300}$ =	0.30
45 to under 50	48	$^{48}/_{300}$ =	0.16
50 to under 55	42	$^{42}/_{300}$ =	0.14
55 to under 60	18	$^{18}/_{300}$ =	0.06
60 to under 65	9	$^{9}/_{300}$ =	0.03
	300		1.00

It can be used to give information like what percentage of employees earn under Rs. 40,000 per month. The answer to this is given by taking the sum of the probabilities of the first three groups. This is 31% (or 0.31 = 0.05 + 0.08 + 0.18).

Characteristics of probability distribution:

- A probability distribution for a random variable gives every possible value for the variable and its associated probability.
- A probability function of a random variable is a formula which gives the probability of a variable taking a specific value.
- The probability distribution and function depend on the different arrangement of outcomes that correspond to the same event in any one trial and the probability of success in that trial.
- Probability distributions are used to estimate the probability of a given result if a sample is taken from population with known characteristics.

Theoretical or experimental distributions

Theoretical probability distributions provide the expected probability of events when the observations are equally likely to occur. However, theoretical probability distributions distinguish from ***experimental*** distributions. Experimental probability distributions result from actual occurrences of events not just in theory.

► *For example:*

Theoretically, probability of picking a red ball from a bag which has 6 white and 6 red balls is 6/12 or 50%. However, in conducting an experiment, when an individual was asked to pick a ball 50 times from the same bag, the probability of picking a red ball every time may be different. It can be 6/50 times, 10/50 or even 25/50.

Discrete or Continuous distributions

Outcome for events can be finite or infinite. Recording finite number of outcomes in frequency distributions is easier than recording infinite number of outcomes.

Infinite number of possibilities are recorded using continuous distributions; whereas finite number of possibilities can be recorded using discrete distributions.

- *For example:*

  In recording time and distance for instance there can be continuous intervals between a period therefore continuous distributions are used.

  In recording events for rolling a dice or drawing a card from a deck there are finite possibilities therefore, discrete distributions are used.

Distributions usually are defined by two characteristics mean and their variance.

A number of these very important probability distributions which can be used to estimate probability, in given circumstances, are provided in detail in the following sections.

2 BINOMIAL DISTRIBUTION

The binomial distribution applies to discrete random variables. A random variable follows a binomial distribution if all of the following conditions are met:

- There is a fixed number of trials;
- Each trial is independent;
- Each trial has two possible results (success or failure); and
- The probability of success or failure is known and fixed for each trial.

Condition that the probabilities are constant implies that any item tested must be replaced (and could therefore be selected again).

▶ *For example:*

The probability associated with the colour of the first card is 26/52.

Selection of a card without replacement changes the probabilities of the remaining cards. If the first card was black the probability of the second card being black would be 25/51 and being red would be 26/51. In this case, binomial distribution does not apply because the probabilities are not constant.

If the card is replaced the probability of the second card being black would be 26/52 and being red would be 26/52. In this case, binomial distribution could apply (subject to meeting the other conditions) because the probabilities are constant.

However, if the population is very large and an item is not replaced, the impact on probability becomes negligible. In this case the binomial may be used.

▶ *For example:*

The probability of the first disc being black is 60,000/100,000.

If the disc is not replaced the probability of the second disc being black would be 59,999/99,999 and being red would be 40,000/99,999.

The probabilities are not constant but the binomial distribution could be used (subject to meeting the other conditions) because the change in probabilities is negligible.

Note that the word success is used in a specific way. If we were interested in the number of faulty goods in a trial a success might be defined as finding a faulty good (success simply refers to as something occurring).

▶ *For example:*

A factory produces 10,000 items a day. The probability of an item being faulty is 0.01. This situation fulfills all the conditions of binomial distribution: There are fixed number of trials (10,000), there are two possible outcomes (faulty and without fault), outcomes are independent (one faulty item does not affect faults in other items), and the probability is constant (0.01).

In a situation where number of times a dice is rolled until a 6 is obtained, the conditions of binomial distribution are not fulfilled. Although the trials are independent, trials have two possible outcomes (6 and non-6s), constant probability (1/6), but there are no fixed number of trials.

The distribution

In the binomial distribution, the probability associated with each combination is a function of the probabilities associated with the number of possible combinations that are possible for a given success.

These two factors (probability and number of possible combinations) can be modelled to give a general expression which can be used to estimate probabilities of outcomes for different sample sizes where there are two events that might occur with known probability.

- *Formula:*

If an experiment is performed *n* times the probability of x successes is given by

$$P(x) = \frac{\text{n!}}{x!\,(\text{n}-x)!}p^x q^{n-x}$$

and

mean = np

Standard deviation = $\sqrt{\text{npq}}$

Where:

p = probability of success

q = probability of failure (must = 1 – p as this distribution applies to circumstances with only two possible outcomes).

n = number of items in the sample

P(x) = the probability of x successes.

$$\frac{\text{n!}}{x!\,(\text{n}-x)!} = binomial\ coeffient$$

The **binomial coefficient** tells you how many success-failure sequences, of the set of all possible sequences, will result in exactly x successes. The probability of each of those individual sequences happening is just $p^x q^{n-x}$.

- *For examples:*

A bag contains 60,000 black disks and 40,000 white disks from which 4 disks are taken at random.

What is the probability that the sample will contain 0, 1, 2, 3 and 4 white disks?

Outcomes	Probabilities		
BBBB	0.6 × 0.6 ×0.6 ×0.6		= 0.1296
WBBB	0.4 × 0.6 ×0.6 ×0.6	= 0.0864	
BWBB	0.6 × 0.4 ×0.6 ×0.6	= 0.0864	= 0.3456
BBWB	0.6 × 0.6 ×0.4 ×0.6	= 0.0864	
BBBW	0.6 × 0.6 ×0.6 ×0.4	= 0.0864	
WWBB	0.4 × 0.4 ×0.6 ×0.6	= 0.0576	
WBWB	0.4 × 0.6 ×0.4 ×0.6	= 0.0576	
WBBW	0.4 × 0.6 ×0.6 ×0.4	= 0.0576	= 0.3456
BWWB	0.6 × 0.4 ×0.4 ×0.6	= 0.0576	
BWBW	0.6 × 0.4 ×0.6 ×0.4	= 0.0576	
BBWW	0.6 × 0.6 ×0.4 ×0.4	= 0.0576	
WWWB	0.4 × 0.4 ×0.4 ×0.6	= 0.0384	
WWBW	0.4 × 0.4 ×0.6 ×0.4	= 0.0384	0.1536
WBWW	0.4 × 0.6 ×0.4 ×0.4	= 0.0384	
BWWW	0.6 × 0.4 ×0.4 ×0.4	= 0.0384	
WWWW	0.4 × 0.4 ×0.4 ×0.4		0.0256
			1.0000

In using the formula:
$P(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$

Let p = 0.4, q = 1 – 0.4 = 0.6, n = 4, and x = 0, 1, 2, 3, 4

	P(x white discs in sample)
$\mathbf{P(0)} = \frac{4!}{0!\,(4)!} 0.4^0 \times 0.6^4 = 1 \times 0.4^0 \times 0.6^4$	0.1296
$\mathbf{P(1)} = \frac{4!}{1!\,(3)!} 0.4^1 \times 0.6^3 = 4 \times 0.4^1 \times 0.6^3$	0.3456
$\mathbf{P(2)} = \frac{4!}{2!\,(2)!} 0.4^2 \times 0.6^2 = 6 \times 0.4^2 \times 0.6^2$	0.3456
$\mathbf{P(3)} = \frac{4!}{3!\,(1)!} 0.4^3 \times 0.6^1 = 4 \times 0.4^3 \times 0.6^1$	0.1536
$\mathbf{P(4)} = \frac{4!}{4!\,(0)!} 0.4^4 \times 0.6^0 = 1 \times 0.4^4 \times 0.6^0$	0.0256
	$\mathbf{1.0000}$

10% of homes in a large town have one or more family members of 90 years or more.

Use the binomial distribution to estimate the probability that a random sample of 10 homes will contain 2 homes with an occupant or occupants in that age range?

In using the formula:
$P(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$

Let p = 0.1, q = 1 – 0.1 = 0.9, n = 10, $x = 2$

$$\mathrm{P(2)} = \frac{10!}{2!\,(8)!} 0.1^2 \times 0.9^8 = 45 \times 0.1^2 \times 0.9^8$$

$$\mathrm{P(2)} = 0.1937$$

It is estimated that 9% of items produced by a process are defective.

Use the binomial distribution to estimate the probability that a random sample of 20 items will contain 0,1 or 2 or more defective units.

In using the formula:
$P(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$

Let p = 0.09, q = 1 – 0.09 = 0.91, n = 20, $x = 0, 1, 2$ or more

$$\mathbf{P(0)} = \frac{20!}{0!\,(20)!} \times 0.09^0 \times 0.91^{20} = 1 \times 1 \times 0.1516 \qquad \mathbf{0.1516}$$

$$\mathbf{P(1)} = \frac{20!}{1!\,(19)!} \times 0.09^1 \times 0.91^{19} = 20 \times 0.09 \times 0.1666 \qquad \mathbf{0.2999}$$

$$\mathbf{P(2\ or\ more)} = 1 - (\mathrm{P(1)} + \mathrm{P(0)})$$

$$\mathbf{P(2\ or\ more)} = 1 - (0.1516 + 0.2999) = \qquad \mathbf{0.5485}$$

3 HYPER-GEOMETRIC DISTRIBUTION

This distribution applies when a sample of n items is randomly selected without replacement from a population (N) in which there are only two types of items (designated success or failure). This also assumes that outcomes are not independent.

- *Formula:*

$$P(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

OR

$$P(x) = \frac{\left(\frac{k!}{x!\,(k-x)!}\right) \times \left(\frac{(N-k)!}{(n-x)!\,[(N-k)-(n-x)]!}\right)}{\left(\frac{N!}{n!\,(N-n)!}\right)}$$

Where:

N = The size of the population from which a number of items is being selected.

n = The size of the sample selected.

k = The number of possible successes in the population.

x = The number of items in the sample classified as a success (the number of items in the sample for which we are trying to find the probability).

$\binom{k}{x} =$ Terms representing combinations of items from a larger group (binomial coefficients).

- *For example:*

What is the probability of there being exactly two red cards in the set of five cards selected from a deck?

Since the cards are not replaced

In this case, N = 52, n = 5, k = 26, x = 2

$$P(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

therefore $P(2) = \frac{\binom{26}{2}\binom{52-26}{5-2}}{\binom{52}{5}}$

Expanding the binomial coefficients

$$P(2) = \frac{\left(\frac{26!}{2!\,(26-2)!}\right) \times \left(\frac{26!}{3!\,(26-3)!}\right)}{\left(\frac{52!}{5!\,(52-5)!}\right)}$$

$$P(2) = \frac{\left(\frac{26!}{2!\,(24)!}\right) \times \left(\frac{26!}{3!\,(23)!}\right)}{\left(\frac{52!}{5!\,(47)!}\right)}$$

$$P(2) = \frac{\left(\frac{26 \times 25}{2 \times 1}\right) \times \left(\frac{26 \times 25 \times 24}{3 \times 2 \times 1}\right)}{\left(\frac{52 \times 51 \times 50 \times 49 \times 48}{5 \times 4 \times 3 \times 2 \times 1}\right)}$$

$$P(2) = \frac{\left(\frac{650}{2}\right) \times \left(\frac{15{,}600}{6}\right)}{\left(\frac{311{,}875{,}200}{120}\right)} = \frac{325 \times 2{,}600}{2{,}598{,}960} = \frac{845{,}000}{2{,}598{,}960} = 0.325$$

AT A GLANCE
SPOTLIGHT
STICKY NOTES

There are 8 blue pens and 7 black pens in a box. If the pens are drawn six times without replacing the pens, what is the probability that exactly three blue pens are drawn?

Since the pens are not replaced, we can use hypergeometric distribution.

In this case, N = 15, n = 6, k = 8, x = 3

$$P(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

therefore $P(3) = \frac{\binom{8}{3}\binom{7}{3}}{\binom{15}{6}}$

Expanding the binomial coefficients

$$P(3) = \frac{\left(\frac{8!}{3!\,(8-3)!}\right) \times \left(\frac{7!}{3!\,(7-3)!}\right)}{\left(\frac{15!}{6!\,(15-6)!}\right)}$$

$$P(2) = \frac{\left(\frac{8 \times 7 \times 6}{3 \times 2 \times 1}\right) \times \left(\frac{7 \times 6 \times 5}{3 \times 2 \times 1}\right)}{\left(\frac{15 \times 14 \times 13 \times 12 \times 11 \times 10}{6 \times 5 \times 4 \times 3 \times 2 \times 1}\right)}$$

$$P(2) = \frac{\left(\frac{336}{6}\right) \times \left(\frac{210}{6}\right)}{\left(\frac{3{,}603{,}600}{720}\right)} = \frac{56 \times 35}{5{,}005} = \frac{1960}{5{,}005} = 0.391$$

AT A GLANCE

SPOTLIGHT

STICKY NOTES

4 POISSON DISTRIBUTION

The Poisson distribution is a discrete probability distribution. It occurs in all kinds of situations where things happen at random.

The following conditions must be met for a random variable to follow a Poisson distribution:

- The events occur randomly;
- The events are independent of each other;
- The events occur singly (i.e. at one point in time or space);
- The events happen on average at a constant rate. (This rate is a function of the probability of an event occurring in relation to all possible outcomes in the period of time or space as defined).

This suggest that Poisson distribution does not work on the probability of the event but how often it occur during a specific time period.

▶ *For example:*

- number of accidents on a given stretch of road if there have been 3 accidents on average in the past period;
- number of telephone calls in a given period of time if usually it is 120;
- number of times a machine breaks down in a given time frame if historical performance is known.

Each of the above cases concern rate where a number of events occur or are presented in a given space of time or place. They tell how often something happens or is present in a given time period or location.

▶ *Formula:*

$$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$\text{mean} = \lambda = \text{np}$$

$$\text{Standard deviation} = \sqrt{\lambda} = \sqrt{\text{np}}$$

Where:

P(x) = The probability of x successes (occurrences) in a specified time frame or space.

λ = The rate that an event occurs (the average number of events that occur or are present in a particular period of time or space).

This is described as being the parameter of the Poisson distribution.

(λ is a Greek letter pronounced lamda)

e = Euler's (Napier's) constant =2.71828.

n = The number of items in the sample.

The probability formula above can look a little daunting. However, there are shortcuts available including the following.

- The term, ***e*** is a constant. If a question specifies a rate and then asks for probabilities of different numbers of events $e^{-\lambda}$ need only be calculated once and the total is used in each calculation.

  (Note that the probability of zero events (***x = 0***) for any value of λ is always equal to $e^{-\lambda}$ as the other terms in the expression solve to 1 – λ^x becomes λ^0 which is equal to 1 and ***x!*** becomes ***0!*** which is equal to 1).

- There are only two variables that change from one level of probability to another. These are ***x*** and ***λ***. This means that it is possible to construct tables showing solution to the probability expression for different combinations of these variables. (This will be revisited later).
- In the absence of tables there is a simple mathematical relationship which links consecutive probabilities. The probability of any number (say ***n***) of events is the probability of the previous number multiplied by λ/n. (This is demonstrated by the ***Simplified Poisson Distribution*** example later in the chapter).

Following must be noted regarding poisson distribution:

- The standard deviation of any distribution is the square root of its variance.
- The standard deviation of the Poisson is the square root of the mean.
- Therefore, the variance of the Poisson distribution is equal to its mean.

▶ *For example:*

If during a live chat 5 queries are responded in a day, what is the probability of responding to 8 queries in a particular day?

In this case, λ = 5, x = 8

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

$$P(x) = \frac{e^{-5}5^8}{8!}$$

$$P(x) = \frac{2.718^{-5} \times 390{,}625}{40{,}320}$$

$$P(x) = \frac{1}{2.718^5} \times \frac{390{,}625}{40{,}320}$$

$$P(x) = \frac{1}{148.336} \times 9.688$$

$$P(x) = 0.0653$$

A business has a large number of machines. These machines require minor adjustments on a regular basis. The need for the adjustments occurs at random at an average rate of 7 per hour.

What is the probability that the machines would require 0, 1, 2, 3 or 4 adjustments in an hour?

In this case, λ = 7, x = 0, 1, 2, 3 or 4

And
$e^{-\lambda} = e^{-7} = 0.000912$

$$P(0) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{0.000912 \times 7^0}{0!} = \frac{0.000912 \times 1}{1} = 0.0009$$

$$P(1) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{0.000912 \times 7^1}{1!} = \frac{0.000912 \times 7}{1} = 0.0064$$

$$P(2) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{0.000912 \times 7^2}{2!} = \frac{0.000912 \times 49}{2} = 0.0223$$

$$P(3) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{0.000912 \times 7^3}{3!} = \frac{0.000912 \times 343}{6} = 0.0521$$

$$P(4) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{0.000912 \times 7^4}{4!} = \frac{0.000912 \times 2{,}401}{24} = 0.0912$$

AT A GLANCE SPOTLIGHT STICKY NOTES

Simplified Poisson computation

$P(0) = 0.0009$

$P(1) = P(0) \times \frac{7}{1} = 0.0064$

$P(2) = P(1) \times \frac{7}{2} = 0.0064 \times \frac{7}{2} = 0.0223$

$P(3) = P(2) \times \frac{7}{3} = 0.0223 \times \frac{7}{3} = 0.0521$

$P(4) = P(2) \times \frac{7}{4} = 0.0521 \times \frac{7}{4} = 0.0912$

A roll of fabric contains 2.5 slight defects per square metre. What is the probability that a square metre will contain 0 defects? What is the probability that a square metre will contain 5 defects?

In this case, λ = 2.5, x = 0, or 5

And

$e^{-\lambda} = e^{-2.5} = 0.0821$

$$P(0) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{0.0821 \times 2.5^0}{0!} = 0.0821 \times \frac{1}{1} \qquad = 0.0821$$

$$P(5) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{0.0821 \times 2.5^5}{5!} = 0.0821 \times \frac{97.656}{120} \qquad = 0.0668$$

A factory has an average of 3 accidents per week. What is the probability that exactly 5 accidents will occur in a week?

In this case, λ = 3, x = 5

And

$e^{-\lambda} = e^{-3} = 0.0498$

$$P(5) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{0.0498 \times 3^5}{5!} = 0.0498 \times \frac{243}{120} \qquad = 0.1008$$

An office receives emailed sales orders at 6 every 20 minutes. What is the probability that no emails will be received during a 20 minute break?

In this case, λ = 6, x = 0

$e^{-\lambda} = e^{-6} = 0.0025$

$$P(0) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{0.0025 \times 6^0}{0!} \qquad = 0.0025$$

AT A GLANCE

SPOTLIGHT

STICKY NOTES

AT A GLANCE
SPOTLIGHT
STICKY NOTES

Poisson distribution tables

There are two sorts of Poisson distribution tables.

- The first table gives the probabilities of different values of x associated with different values of λ.
- The second is a cumulative table. It gives the probabilities of different values of x and fewer occurrences.

The following illustration demonstrates both types of table. More complete versions are set out in an appendix.

▶ *Illustration 01:*

	λ					
x	**1**	**2**	**3**	**4**	**5**	**6**
0	0.3679	0.1353	0.0498	0.0183	0.0067	0.0025
1	0.3679	0.2707	0.1494	0.0733	0.0337	0.0149
2	0.1839	0.2707	0.2240	0.1465	0.0842	0.0446
3	0.0613	0.1804	0.2240	0.1954	0.1404	0.0892
4	0.0153	0.0902	0.1680	0.1954	0.1755	0.1339
5	0.0031	0.0361	0.1008	0.1563	0.1755	0.1606

In this table an entry indicates the probability of a specific value of x for an average rate of λ.

If an event occurs at an average rate of 1 per hour there is only a low probability of 5 events in an hour occurring (0.0031 or 0.31 %) but a bigger chance of 1 occurrence (0.3679 or 36.79%). This can be seen in the first column above.

Similarly, if the rate is 6 per hour there is only a small chance of there being 0 occurrences in an hour (0.0025 or 0.25%) but a bigger chance of there being 5 occurrences (0.1606 or 16.06%). This can be seen in the right hand column above.

▶ *Illustration 02:*

Cumulative Poisson Distribution table

	λ						
x	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0	0.3679	0.1353	0.0498	0.0183	0.0067	0.0025	0.0009
1	0.7358	0.4060	0.1991	0.0916	0.0404	0.0174	0.0073
2	0.9197	0.6767	0.4232	0.2381	0.1247	0.0620	0.0296
3	0.9810	0.8571	0.6472	0.4335	0.2650	0.1512	0.0818
4	0.9963	0.9473	0.8153	0.6288	0.4405	0.2851	0.1730
5	0.9994	0.9834	0.9161	0.7851	0.6160	0.4457	0.3007

In this table, an entry indicates the probability of a specific value being x or fewer for an average rate of λ

If an event occurs at an average rate of 1 per hour it is almost certain (0.9994 or 99.94%) that there will be 5 or less in any one hour.

Problems of a more complex nature

For probabilities associated with a range of occurrences, for example, 4 or fewer, a sum of probabilities associated with 0, 1, 2, 3 and 4 or from the cumulative tables can be used.

However, where probability of at least 5 (which is the same as saying 5 or more), is required then calculating probabilities associated with any number greater than 5 and add those together could get complex. The probabilities would drop off as bigger numbers were used but we would have to decide when to stop.

The easier way to deal with this is to use the fact that the sum of all probabilities must = 1. We can therefore identify the probabilities associated with those numbers which would not satisfy the test, sum them and then simply subtract this number from 1.

- *For example:*

A business has a large number of machines. These machines require minor adjustments on a regular basis. The need for the adjustments occur at random at an average rate of 7 per hour.

What is the probability that 4 or fewer adjustments will be required in an hour?

The probabilities of individual values of x = 0, 1, 2, 3 and 4 is used as calculated before and then sum together.

P(0)	= 0.0009
P(1)	= 0.0064
P(2)	= 0.0223
P(3)	= 0.0521
P(4)	= 0.0912
Probability of 4 or fewer adjustments	**= 0.1729**

There is a 17.29% chance that 4 or fewer adjustments will be required in any one hour.

As an alternative approach, this value is easily obtained from the cumulative tables (as 0.1730, the difference being due to rounding)

What is the probability that at least 5 adjustments will be required in an hour?

Something must happen so the total probability associated with all possible outcomes is one

At least 5 adjustments is the same as 5 or more.

The probability of this happening is 1 less the sum of the probabilities of 0, 1, 2, 3 or 4 occurring.

Probability of at least 5 adjustments= 1 - 0.1729 = 0.8271

On average 5 cars an hour pass a given point on a road. A researcher is trying to find out about probabilities of 8 cars passing in a two-hour period.

In this case, first we have to find the average rate for that period (λ). Since 5 cars pass in an hour, 10 will pass in 2 hours (λ = np).

Possibility of passing 8 cars in 2 hours' period, we use formula:

In this case, $\lambda = 10$, $x = 8$

And

AT A GLANCE
SPOTLIGHT
STICKY NOTES

$$P(8) = \frac{e^{-\lambda}\lambda^{x}}{x!} = \frac{e^{-10}10^{8}}{8!}$$

$$P(8) = \frac{0.00004544 \times 100{,}000{,}000}{40320}$$

$$P(8) = \frac{4544.703}{40320}$$

$$P(8) = 0.1127$$

A shop sells 4 computers a week on average. The owner wishes to know about the probability associated with selling 2 computers within a day.

First we will calculate the average rate of selling computer for that period (λ).

This can be 4/6 or 2/3 per day (taking 6 days a week).

In this case, $\lambda = 2/3$, $x = 2$

And

$$P(2) = \frac{e^{-\lambda}\lambda^{x}}{x!} = \frac{e^{-2/3}2/3^{2}}{2!}$$

$$P(2) = \frac{0.5135 \times 0.4444}{2}$$

$$P(2) = \frac{0.2282}{2}$$

$$P(2) = 0.1141$$

A shop sells wedding dresses at a rate of 2.5 per week.

(a) What is the probability that it will sell at least 8 dresses in a randomly selected two-week period?

mean $(\lambda) = np = 2 \times 2.5 = 5$ dresses per 2 week period

The probability of at least 8 is the same as the probability of 8 or more.

This is equal to 1 – the probability of 7 or fewer.

From tables the probability of 7 or fewer = 0.8666.

Therefore, the probability of at least 8 = 1 – 0.8666 = 0.1334.

OR

Formula can be used as below:

In this case, λ = 5, x = 0, 1, 2, 3, 4, 6, 7

e^{-5}= 0.00674

$$P(0) = \frac{e^{-5}5^{0}}{0!} = \frac{0.00674 \times 1}{1} \qquad = 0.00674$$

$$P(1) = \frac{e^{-5}5^{1}}{1!} = \frac{0.00674 \times 5}{1} = \frac{0.0337}{1} \qquad = 0.0337$$

$$P(2) = \frac{e^{-5}5^{2}}{2!} = \frac{0.00674 \times 25}{2} = \frac{0.1685}{2} \qquad = 0.08425$$

$$\mathrm{P}(3)=\frac{e^{-5}5^{3}}{3!}=\frac{0.00674\ \times 125}{6}=\frac{0.8425}{6} \qquad =0.1404$$

$$\mathrm{P}(4)=\frac{e^{-5}5^{4}}{4!}=\frac{0.00674\times 625}{24}=\frac{4.2125}{24} \qquad =0.1755$$

$$\mathrm{P}(5)=\frac{e^{-5}5^{5}}{5!}=\frac{0.00674\ \times 3125}{120}=\frac{21.0625}{120} \qquad =0.1755$$

$$\mathrm{P}(6)=\frac{e^{-5}5^{6}}{6!}=\frac{0.00674\ \times 15625}{720}=\frac{105.3125}{720} \qquad =0.14626$$

$$\mathrm{P}(7)=\frac{e^{-5}5^{7}}{7!}=\frac{0.00674\ \times 78125}{5040}=\frac{526.5625}{5040} \qquad =0.1044$$

0.8667

1-0.8667 = 0.13325

Poisson as an approximation for the binomial distribution

The binomial distribution can be difficult to use when sample sizes are large and probability of success is small. The Poisson distribution is used instead of the binomial in such circumstances.

- *For example:*

A bag contains 99,940 black disks and 60 white disks from which 1,000 disks are taken at random. What is the probability that the sample will contain 5 white disks?

n = 1,000, p = 0.0006

using the binomial distribution would be difficult to calculate:

p = 0.0006

q = 1 – 0.0006 = 0.9994

n = 1,000

$$P(x)=\frac{\mathrm{n}!}{x!\,(\mathrm{n}-x)!}p^{x}q^{n-x}$$

$$P(5)=\frac{1{,}000!}{5!\,(1{,}000-5)!}0.0006^{5}0.9994^{1{,}000-5}$$

Lets use, Poisson distribution:

$\text{mean}=\lambda=\mathrm{np}=1{,}000\times 0.0006=0.6$

$e^{-\lambda}=e^{-0.6}=0.5488$

$$\mathrm{P}(5)=\frac{e^{-\lambda}\lambda^{x}}{x!}=\frac{0.5488\times 0.6^{5}}{5!}=\frac{0.5488\times 0.07776}{120}=0.000356$$

AT A GLANCE
SPOTLIGHT
STICKY NOTES

5 NORMAL DISTRIBUTION

The normal distribution is perhaps the most widely used frequency distribution.

It is a probability distribution of a continuous variable that fits many naturally occurring distributions (for example, heights of people, crop yields, petrol consumption, dimensions of mass produced articles, rate of defectives in a manufacturing process etc.).

In this section, normal distribution is only explored using information from normal distribution tables or plots rather than equation.

The plot of a normal distribution (normal curve) has the following characteristics.

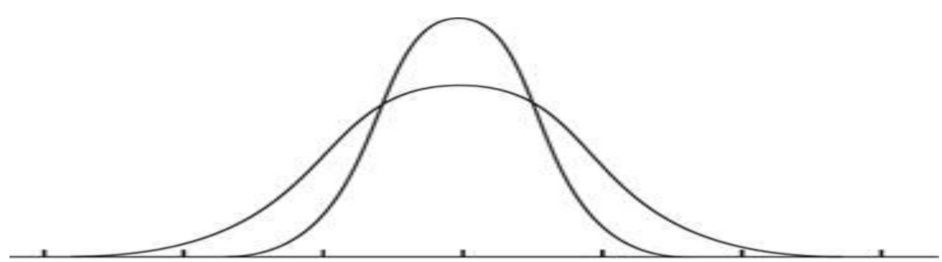
- It is symmetrical and bell shaped;
- Both tails of the distribution approach but never meet the horizontal axis.
- It is described by its mean (μ) (which is the same as the median and the mode) and standard deviation (σ).
- The area under the normal curve represents probability and so totals to 1 (or 100%).
- The same area, of all normal distributions, lie within the same number of standard deviations from the mean.

Normal curve

All normal curves are symmetrical about the mean and are bell shaped. However, the exact shape depends on the standard deviation (variance) of the distribution.

Higher standard deviation implies a more dispersed distribution leading to a flatter and broader curve.

▶ *Illustration:*



These two curves have the same mean but the curve with the higher peak has less dispersion. It would have a lower standard deviation than the lower peak curve.

Properties of a normal distribution

The area under the normal curve represents probability and totals to 1 (or 100%). The distribution is symmetrical so 50% of the area under the curve is to the left and 50% to the right of the mean.

For any normal curve, the area under the curve between the mean and a given number of standard deviations from the mean is always the same proportion of the total area.

Every normal curve can be fitted to a standard curve for which tables have been constructed to show the percentage contained in an area between the mean and a number of standard deviations from the mean (the Z value).

The tables show the probabilities for one half of the distribution only.

- *Illustration:*

Area under the normal curve (Z = number of standard deviations)

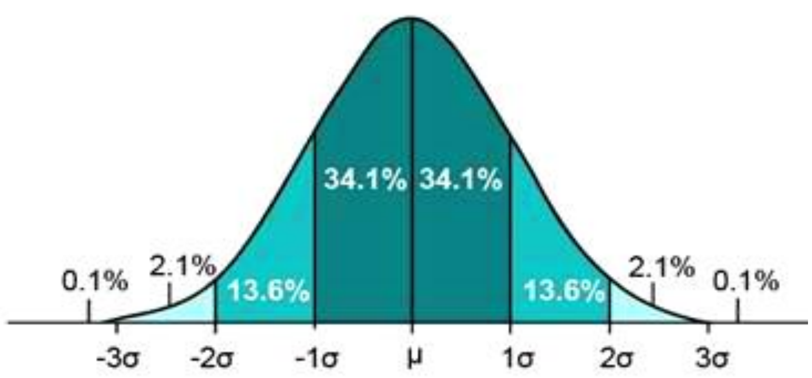
Z	0.00	0.01	0.02	0.03	0.04	0.05
1.0	.3413	.3438	.3461	.3485	.3508	.3531
1.1	.3643	.3665	.3686	.3708	.3729	.3749
1.2	.3849	.3869	.3888	.3907	.3925	.3944
1.3	.4032	.4049	.4066	.4082	4099	.4115
1.4	.4192	.4207	.4222	.4236	.4251	.4265
1.5	.4332	.4345	.4357	.4370	.4358	**.4394**
↓	↓					
2.0	**.4772**					
↓	↓					
3.0	**.4987**					

The table shows by looking down the vertical axis, that 0.4332 (43.32%) of the area under the curve is within 1.5 standard deviations of the mean (in each half of the curve).

Similarly, 0.4772 (47.72%) of the area under the curve is within 2 standard deviations of the mean (in each half of the curve).

Using the horizontal axis as well, .4394 (43.94%) of the area under the curve is within 1.55 standard deviations of the mean (in each half of the curve).

- *Illustration:*



34.16% of the area within each half of all normal curves lies between the mean and one standard deviation from the mean.

Therefore, 68.2% (2 × 34.16%) of the area of all normal curves lie in between one standard deviation on either side of the mean.

13.6% of the area within each half of all normal curves lies between one and two standard deviations from the mean.

Therefore, 47.76% (34.16% + 13.6%) of the area within each half of all normal curves lies between the mean and two standard deviations from the mean.

Therefore, 95.52% (2 × 47.76%) of the area of all normal curves lie in between two standard deviations either side of the mean.

It is very useful to know the following:

- Looking at both tails of the distribution:
  - A line drawn 1.96 standard deviations from the mean will enclose 47.5% (or exclude 2.5%) of the area under half of the distribution.
  - Two lines drawn 1.96 standard deviations above and below the mean will enclose 95% (or exclude 5%, 2.5% in each tail) of the area under the total distribution.
  - A line drawn 2.58 standard deviations from the mean will enclose 49.5% (or exclude 0.5%) of the area under half of the distribution.
  - Two lines drawn 2.58 standard deviations above and below the mean will enclose 99% (or exclude 1%, 0.5% in each tail) of the area under the total distribution.
- Looking at one tail of the distribution:
  - A line drawn 1.645 standard deviations above the mean will enclose 45% (or exclude 5%) of the area under half of the distribution. If concern is only in this direction 95% (50% + 45%) is below the line.
  - A line drawn 2.325 standard deviations above the mean will enclose 49% (or exclude 1%) of the area under half of the distribution. If concern is only in this direction 99% (50% + 49%) is below the line.

Using the normal distribution

The probability of any value in a set of data that fits the normal distribution can be estimated as long as the mean and standard deviation of the data are known.

The probability of a value within or outside a certain value of the mean can be estimated by calculating how many standard deviations that value is from the mean. This is known as its Z score. The normal distribution tables can then be used to derive the probability.

► *Formula:*

Z score (number of standard deviations from the mean)

$$Z = \frac{x - \mu}{\sigma}$$

Where:

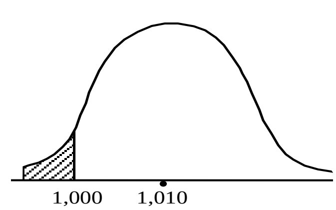
x = the value or limit under investigation

σ = the standard deviation of the distribution

μ = the mean of the distribution

► *For example:*

The weights of bags of sugar produced on a machine are normally distributed with mean = 1,010g and standard deviation 5g. A bag is picked at random; what is the probability it weighs less than 1,0



let x = 1000, μ = 1010 and σ = 5

$$Z = \frac{x - \mu}{\sigma} = \frac{1{,}000 - 1{,}010}{5} = -2$$

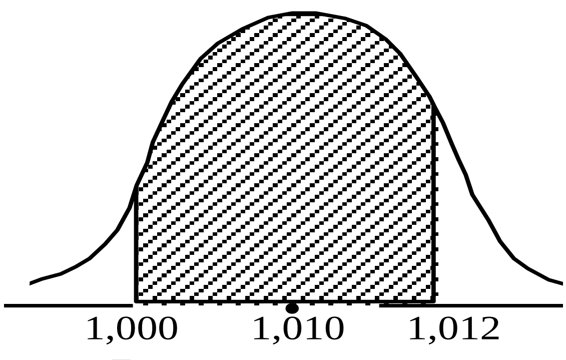
In other words, 1,000g is two standard deviations below the mean.

From the table, 0.4772 (47.72%) of the distribution is within two standard deviations below the mean so:

P(z <-2)=0.5-0.4772=0.0288

Therefore, the probability of a bag of sugar being less than 1000g is 0.0228 or 2.28%

The weights of bags of sugar produced on a machine are normally distributed with mean = 1,010g and standard deviation 5g. A bag is picked at random; what is the probability it weighs between 1,000g and 1,012g?



let x = 1000, 1012, $\mu = 1010$ and $\sigma = 5$

1,000g is below the mean $\quad Z = \frac{x-\mu}{\sigma} = \frac{1{,}000 - 1{,}010}{5} = -2$

1,012 is above the mean $\quad Z = \frac{x-\mu}{\sigma} = \frac{1{,}012 - 1{,}010}{5} = 0.4$

We are interested in the area of the graph between 2 standard deviations below the mean to 0.4 standard deviations above the mean.

From the table:	P
Z = 2 (below the mean)	0.4772
Z = 0.4 (above the mean)	0.1554
	0.6326

This is expressed more formally as:

$$P(1{,}000 < x < 1{,}012) = ?$$
$$= P(-2 < z < 0.4)$$
$$= P(-2 < z < 0) + P(0 < x < 0.4)$$
$$= 0.4772 + 0.1544 = 0.6326$$

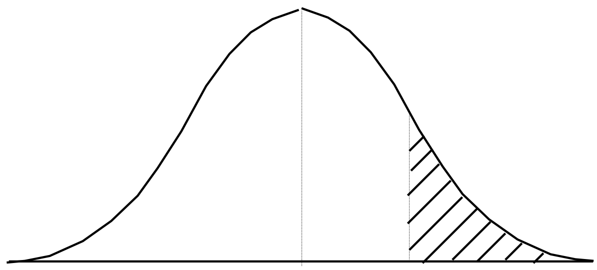
Conclusion: There is a 0.6326 (63.26%) chance that a bag picked at random will be between 1,000 and 1,012g.

Daily demand is normally distributed with a mean of 68 items per day and standard deviation of 4 items.

a) In what proportion of days was demand:

(i) more than 74?

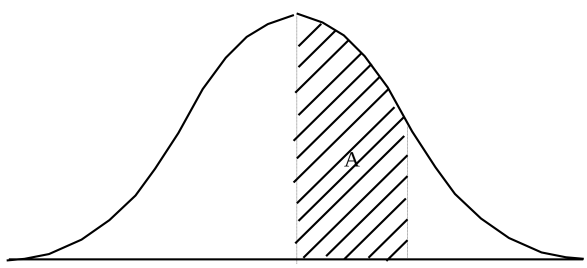
Let x = 74, μ = 68 and σ = 4



$$Z = \frac{x - \mu}{\sigma} = \frac{74 - 68}{4} = 1.5$$

P (x > 74) = 0.0668

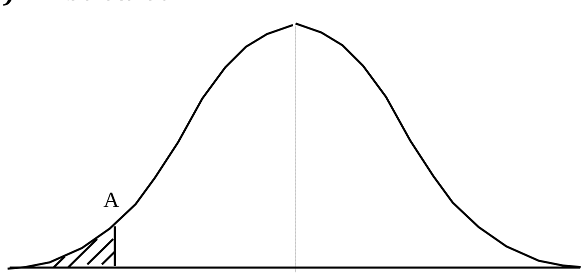
(ii) between 68 and 74?



μ = 68 x = 68, 74

P (between 68 and 74) = 0.5 - 0.0668 = 0.4332

(iii) below 60?

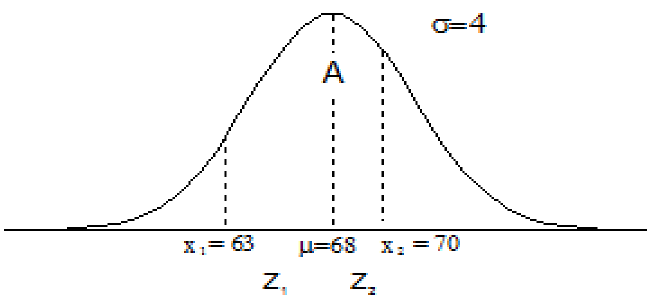


x = 60 μ = 68

$$Z = \frac{x - \mu}{\sigma} = \frac{60 - 68}{4} = -2$$

P(x < 60) = 0.5 – 0.4772 = 0.0228

(iv) between 63 and 70?

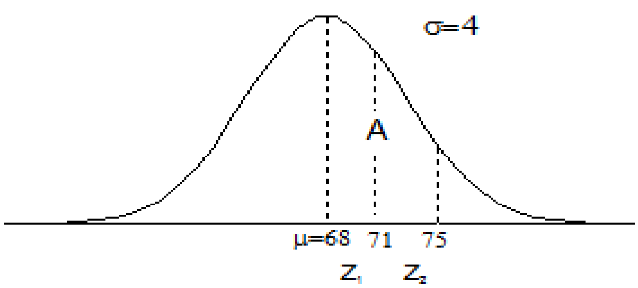


$$Z_1 = \frac{x-\mu}{\sigma} = \frac{63-68}{4} = -1.25$$

$$Z_2 = \frac{x-\mu}{\sigma} = \frac{70-68}{4} = -0.5$$

From tables, proportion = 1 - (0.1056 + 0.3085) = 0.5859

(v) between 71 and 75?

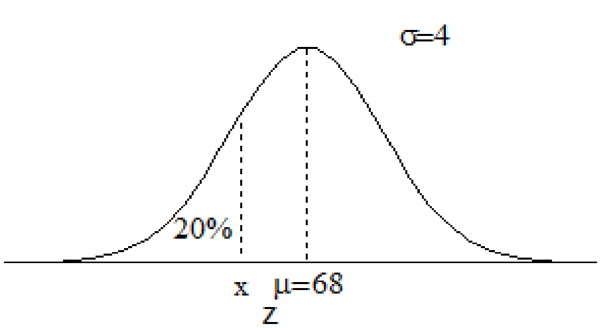


$$Z_1 = \frac{x-\mu}{\sigma} = \frac{71-68}{4} = 0.75$$

$$Z_2 = \frac{x-\mu}{\sigma} = \frac{75-68}{4} = 1.75$$

From tables, proportion = 0.2266 - 0.0401 = 0.1865

b) Below what sales level was the lowest 20% of days ?

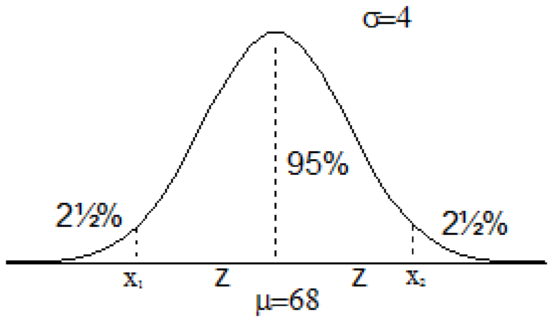


From tables z = 0.84

x = 68 - 0.84 × 4 = 64.64

AT A GLANCE
SPOTLIGHT
STICKY NOTES

c) ***Demand was between what 2 values on the middle 95% of days?***



From tables

z = ± 1.96

x = 68 ± 1.96 × 4

= 60.16 – 75.84

d) ***What about 99% of days?***

x = 68 ± 2.58 × 4

= 57.68 – 78.32

The specification for the length of an engine part is a minimum of 99 mm and a maximum of 104.4 mm. A batch of parts is produced that is normally distributed with a mean of 102 mm and a standard deviation of 2mm.

Find the percentage of parts which are:

(i) ***undersize;***

The differences between the tolerance limits of 99 mm and 104.4 mm and the mean of 102 mm must be expressed in terms of numbers of standard deviations.

$$Z = \frac{x - \mu}{\sigma} = \frac{102 - 99}{2} = 1.5$$

From tables, the proportion of parts below 99 mm and therefore undersize = 0.0668 = 6.68%.

(ii) ***oversize***

$$Z = \frac{x - \mu}{\sigma} = \frac{104.4 - 102}{2} = 1.2 \text{ standard deviations}$$

From tables, the proportion that are oversize = 0.1151 = 11.51%

6 NORMAL APPROXIMATIONS

The normal distribution can be used as an approximation to the binomial distribution and Poisson distribution subject to satisfying certain criteria and the application of a continuity correction factor.

Continuity correction factor

The normal distribution is a continuous probability distribution whereas both the binomial and Poisson are discrete probability distributions. This must be taken into account when using the normal as an approximation to these.

The correction involves either adding or subtracting 0.5 of a unit from the discrete value when applying the normal as an approximation. This fills in the gaps in the discrete distribution making it continuous.

Note that this is similar to the adjustments to form class boundaries for grouped frequency distributions.

- *Illustration:*

Discrete event under investigation		Corrected for normal approximation
Probability of 10 successes	P(x = 10)	P(9.5< x < 10.5)
Probability of more than 10 successes	P(x > 10)	P(x > 10.5)
Probability of 10 or more successes	P(x ≥ 10)	P(x > 9.5)
Probability of fewer than 10 successes	P(x < 10)	P(x < 9.5)
Probability of 10 or fewer successes	P(x ≤ 10)	P(x < 10.5)

Normal approximation to the binomial distribution

The normal distribution provides a good approximation to the binomial when the sample size (n) is large and the probability of success (p) is near to 0.5.

- *Formula:*

$$\text{Standard deviation} = \sqrt{np(1-p)}$$

$$\text{mean} = \text{np}$$

The probability of success can be further away from 0.5 than one might expect and there is no firm rule about what constitutes a large sample.

The normal distribution provides a good approximation to the binomial in the following conditions.

$$0.1 < \text{p} < 0.9$$

$$\text{mean} = \text{np} > 5$$

- *For example:*

The probability of a sales visit to a client leading to a sale is 0.4. A business makes 100 visits per week. What is the probability that 35 or more sales will be made in a week?

$$\text{mean} = \text{np} = 100 \times 0.4 = 40$$

$$\text{Standard deviation} = \sqrt{np(1-p)} = \sqrt{40(1-0.4)} = 4.899$$

Continuity correction: P(x ≥ 35) restated to P(x > 34.5)

$$Z = \frac{x - \mu}{\sigma} = \frac{34.5 - 40}{4.899} = -1.12$$

From the table, 0.3686 (36.86%) of the distribution is within 1.12 standard deviations below the mean so:

$P(x > 34.5) = 0.3686 + 0.5 = 0.8686$

Therefore, the probability of making 35 or more sales in a week is 0.8686 (86.86%).

Normal approximation to the Poisson distribution

The normal distribution can be used as an approximation to Poisson when the rate that an event occurs (λ) is greater than 30.

- *For example:*

A poultry farmer collects eggs every morning. There are 50 broken eggs per collection on average. What is the probability that more than 60 eggs might be broken in any one collection?

mean $= \lambda = 50$ eggs

Standard deviation $= \sqrt{\lambda} = \sqrt{50} = 7.071$

Continuity correction: $P(x > 60)$ restated to $P(x > 60.5)$

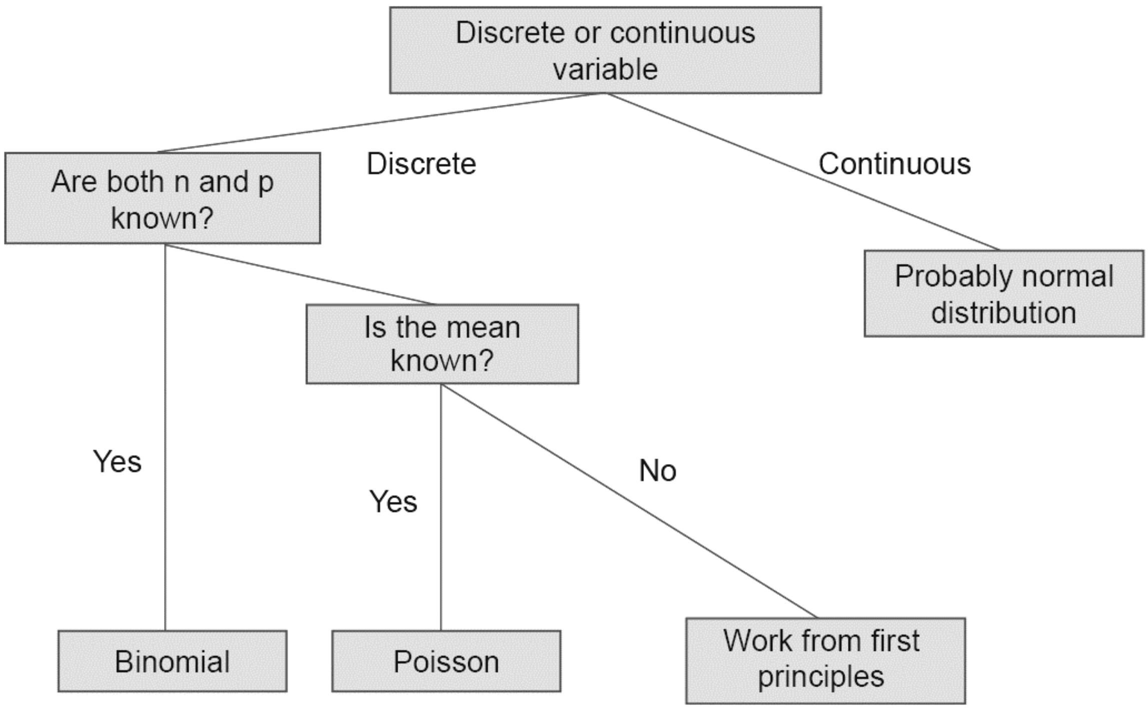
$$Z = \frac{x - \mu}{\sigma} = \frac{60.5 - 50}{7.071} = 1.48$$

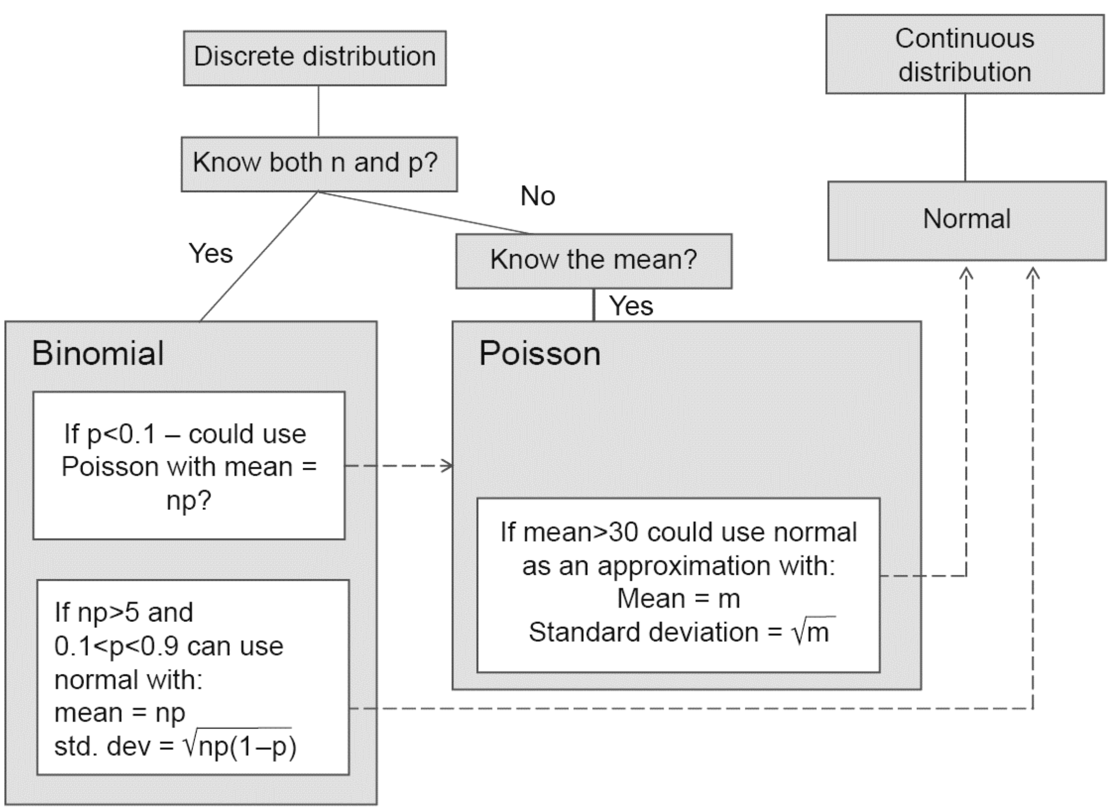
From the table, 0.4306 (43.06%) of the distribution is within 1.48 standard deviations above the mean so:

$P(x > 60.5) = 0.5 - 0.4306 = 0.0694$

Therefore, the probability that more than 60 eggs might be broken in any one collection is 0.0694 (6.94%).

STICKY NOTES





SELF-TEST

12.1. If n=5; p = 0.2; then using binomial distribution P(x =2) is:

(a)	0.409	(b)	0.2048
(c)	0.0512	(d)	None of these

12.2. If n=5; p = 0.2; then using binomial distribution P(x >3) is:

(a)	0.0643	(b)	0.576
(c)	0.0064	(d)	None of these

12.3. If mean of Binomial Distribution is 1.0 and q = 0.98 then n = ?

(a)	50	(b)	1
(c)	10	(d)	None of these

12.4. If mean of Binomial distribution is 1 and variance is 0.98 then p =

(a)	0.98	(b)	0.02
(c)	Does not exist	(d)	None of these

12.5. If mean of Binomial distribution is 1 and variance is 0.98, then q is:

(a)	0.98	(b)	0.2
(c)	5	(d)	None of these

12.6. If n = 10 and p = 0.9, then mean and variance of Binomial Distribution is:

(a)	9 and 0.9	(b)	0.9 and 0.09
(c)	9 and 100	(d)	None of these

12.7. The probability distribution showing the probability of x occurrences of an event over a specified time, distance or space is known as:

(a)	Binomial probability Distribution	(b)	Poisson Probability Distribution
(c)	Chi-square Distribution	(d)	None of these

12.8. The number of days in a given year in which a 130-point change occurs in the Karachi Stock Exchange index.

The above statement may describe a random variable that may posses:

(a)	Binomial Distribution	(b)	Poisson Distribution
(c)	Normal Distribution	(d)	None of these

12.9. The statement "The number of fatal accidents that occur per month in a large manufacturing plant" may describe a random variable that may possess:

(a) Binomial Distribution (b) Poisson Distribution
(c) Normal Distribution (d) None of these

12.10. The number of coloured eye babies born per month in the city of Karachi.
The above statement may describe a random variable that may possess:

(a) Binomial Distribution (b) Poisson Distribution
(c) Normal Distribution (d) None of these

12.11. If μ = 1.2 then using poisson distribution P(x = 0) is:

(a) 0.301 (b) 0.361
(c) 0.338 (d) None of these

12.12. If μ = 1.2 then using poisson distribution P(x > 0) is:

(a) 0.699 (b) 0.639
(c) 0.662 (d) None of these

12.13. If $e^{-4.5}$ = 0.0111 then the mean of Poisson distribution is:

(a) 4.5 (b) 45
(c) Does not exist (d) None of these

12.14. If $e^{-6.0}$ = 0.002478; then P (x = 0):

(a) Cannot be determined (b) 0.002478
(c) 0.5488 (d) None of these

12.15. If e^{-6} = 0.002478; then mean of Poisson distribution is:

(a) Incomplete information (b) 6
(c) 5.488 (d) None of these

12.16. If e^{-6} = 0.002478; then variance of Poisson distribution is:

(a) Can't be determined (b) 6
(c) $\sqrt{6}$ (d) None of these

12.17. If P (x = 4) = $\frac{(1.5)^4 e^{-y}}{4!}$=0.04706 then y =

(a) 4 (b) 1.5
(c) Can't be determined (d) None of these

12.18. If P (x = y) = $\frac{(3.0)^2 e^{-3}}{y!}$=0.07468 then y =

(a) 4 (b) 2

(c) Can't be determined (d) None of these

12.19. A complete list that gives the probabilities associated with each value of a random variable x is called:

(a) Frequency distribution (b) Probability distribution

(c) Expected value (d) None of these

12.20. A variable that assumes the numerical values associated with events of an experiment is called:

(a) Parameter (b) Statistic

(c) Random variable (d) None of these

12.21. A table or formula listing all possible values that a random variable can take on, along with the associated probabilities is called:

(a) Probability distribution (b) Frequency distribution

(c) Expected value (d) None of these

12.22. If x is a random variable belonging to a continuous probability distribution, then P(x = a) is:

(a) Equal to zero (b) Less than 1

(c) More than zero (d) None of these

12.23. Let x be a random variable with probability distribution
P(X= x), then Σ x p(x) is known as:

(a) Sum of probabilities (b) Mean or expected value

(c) Mean of Binomial Distribution (d) None of these

12.24. In a game, a man is paid Rs. 50 if he gets all heads or all tails when 4 coins are tossed and he pays out Rs. 30 if 1, 2 or 3 heads appear. His expected gain is:

(a) Rs. 20 (b) Rs. 6.26

(c) Rs. - 20 (d) None of these

12.25. In terms of mathematical expectation the formula E(x^2)-[E(x)]^2 represents :

(a) Standard deviation of distribution (b) Variance of distribution

(c) Difference of squared deviation (d) None of these

12.26. Let x assumes the value 0, 1, 2, and 3 with the respective probabilities 0.51, 0.38, 0.10 and 0.01 the mean of distribution is:

(a) 0.38 (b) 0.61

(c) 0.4979 (d) None of these

12.27. Let x assumes the value 0, 1, 2 and 3 with the respective probabilities 0.51, 0.38, 0.10 and 0.01 the variance of distribution is:

(a) 0.87 (b) 0.61

(c) 0.4979 (d) None of these

12.28. By investing into a particular stock, a person can make a profit in 1 year of Rs. 5000 with probability 0.4 or take a loss of Rs. 1500 with probability 0.8. the person's expected gain is:

(a) 1200 (b) 800

(c) Can't be determined due to wrong values of probabilities (d) None of these

12.29. If an experiment is repeated for n trials, each trial results in an outcome classified as success or failure, with constant probability of success, and trials are independent, the experiment is known as:

(a) Binomial Experiment (b) Poisson experiment

(c) Hyper geometric experiment (d) None of these

12.30. (i) A random sample of size n is selected from a population of N items.

(ii) K on N items may be classified as success and N-K classified as failures

The above properties are related to a

(a) Binomial experiment (b) Poisson experiment

(c) Hypergeometric experiment (d) None of these

12.31. If a random sample of size n is drawn with replacement the events are said to be:

(a) Independent (b) Not independent

(c) Not mutually exclusive (d) None of these

12.32. If a random sample of size n is drawn without replacement the events are said to be:

(a) Independent (b) Not independent

(c) Not mutually exclusive (d) None of these

12.33. Whenever we measure time intervals, weights, heights, volumes and so forth, our underlying population is described by a:

(a) Discrete distribution (b) Continuous distribution

(c) Frequency distribution (d) None of these

12.34. A distribution whose graph is a symmetric bell shaped curve extending indefinitely in both directions, with equal measures of central tendency is known as:

(a)	Non-skewed distribution	(b)	Normal distribution
(c)	Binomial distribution	(d)	None of these

12.35. The normal probability distribution is a continuous distribution with parameter(s):

(a)	Mean	(b)	Mean and variance
(c)	Mean and mean deviation	(d)	None of these

12.36. A standard normal distribution is one that has mean and variance equal to:

(a)	Zero and one	(b)	One and one
(c)	One and zero	(d)	None of these

12.37. If μ = 50, σ = 10 and z = 1.2 then the corresponding x value must be:

(a)	54	(b)	62
(c)	42	(d)	None of these

12.38. If μ = 50, z = -0.5and x = 45 the variance must be equal to:

(a)	10	(b)	100
(c)	3.1622	(d)	None of these

12.39. The percentage of measurements of a normal random variable fall within the interval is $\mu \pm \sigma$ is:

(a)	68.26%	(b)	A least 75%
(c)	95.44%	(d)	None of these

12.40. The percentage of measurements of a normal random variable fall within the interval $\mu \pm 2\ \sigma$ is:

(a)	At least 75%	(b)	95.44%
(c)	99.74%	(d)	None of these

12.41. The percentage of measurement of a normal random variable fall within the interval is $\mu \pm 3\ \sigma$ is:

(a)	At least 88.9%	(b)	95.44%
(c)	99.74%	(d)	None of these

12.42. On an examination the average grade was 74 and the standard deviation was 7. If 10% of the class are given A's and the distribution of grades is to follow a normal distribution. Then the lowest possible A and highest possible B if $z_{.10}$ = 1.28 is:

(a)	83 and 82	(b)	84 and 83
(c)	85 and 84	(d)	82 and 81

12.43. Given a normal distribution with μ = 200 and σ^2= 100 then the two x values containing the middle 75% of the area if $z_{.125}$ = 1.15 are:

(a) 188.5 and 211.5 (b) 85 and 315
(c) 187.5 and 212.5 (d) None of these

12.44. If a set of observations is normally distributed, then the percentage of observations differs from mean by more than 1.3σ is:

(a) 19.36% (b) 90.32%
(c) 9.68% (d) None of these

12.45. If a set of observations is normally distributed, then the percentage of observations differs from mean by less than 0.52 σ is:

(a) 69.85 (b) 30.15
(c) 39.7% (d) None of these

12.46. The IQ of 300 applicants to a certain college are approximately normally distributed with a mean of 115 and a standard deviation of 12. If the college required an IQ of at least 95, then the number of students those will be rejected on this basis regardless of their other qualifications are:

(a) 26 (b) 14
(c) Non3e will be rejected (d) None of these

12.47. A z-score measures how many standard deviations an observation is:

(a) Above the mean (b) Below the mean
(c) Above or below the mean (d) None of these

12.48. If z=-2 then it is correct to say that:

(a) Observation is less than mean (b) Observation is less than standard deviation
(c) Observation is more than standard deviation (d) None of these

12.49. If z= -2, μ = 17 and σ^2 = 64 then x =______:

(a) +1 (b) - 17
(c) +17 (d) None of these

12.50. It is possible to compare two observations measured in completely different units by z-score because:

(a) z-score has its own units (b) z-score is a unit less quantity
(c) z-score is a standard score (d) None of these

12.51. Which of the following statements are correct?

1. Random variables are either discrete or continuous.

2. A random variable is considered discrete if the values it assumes can be counted.

3. A random variable is considered continuous if they can assume any value within a given range.

(a) 1 (b) 1 and 3

(c) 1 and 2 (d) All statements are correct

12.52. Which of the following assumptions about binomial distribution are correct?

1. The experiment consists of n repeated trials.

2. Each trial has two possible outcomes, one called "success" and the other "failure".

3. The repeated trials are independent of each other.

(a) 1 (b) All statements are correct

(c) 2 (d) 1 and 3

12.53. Which of the following properties about Poisson experiment are correct?

1. The expected number of occurrences in an interval is proportional to the size of the interval.

2. The number of occurrences in one interval is independent of the number of occurrences in another interval.

3. The expected number of occurrences in an interval is not proportional to the size of the interval.

(a) 1 and 2 (b) 1

(c) All statements are correct (d) None

12.54. The sum of all probabilities in a probability distribution table must be equal to ____

(a) 1 (b) 0

(c) Mean of data (d) Standard deviation of data

12.55. Considering binomial distribution, variance divided by mean results in ____

(a) Standard deviation (b) Probability of failure

(c) Probability of success (d) Number of trials

12.56. Considering binomial distribution, the sum of probability of success and probability of failure is always ____

(a) 1 (b) 0

(c) 2 (d) Cannot be determined with certainty

12.57. Considering Poisson distribution, if mean is 4 compute value of standard deviation

(a) 4 (b) 2

(c) 16 (d) Cannot be determined with available data

12.58. Which of the following statements are correct about normal distribution?

1. The total area under the normal curve is 1.
2. Probabilities are always given for range of values.
3. A distribution with a smaller standard deviation would define a flatter curve.

(a) 1 (b) All three statements are correct

(c) 1 and 2 (d) 2 and 3

12.59. Considering a normal distribution, the normal distribution curve is ____ therefore ___, median and mode all coincide at the same point.

(a) Symmetric and variance (b) Symmetric and mean

(c) Dispersed and mean (d) Flat and variance

12.60. A ____ variable has a discrete number of occurrences in a continuous interval.

(a) Poisson (b) Hyper geometric

(c) Binomial (d) Normal

12.61. Find the probability of obtaining exactly three 6s in five throws of a fair dice.

(a) 0.042 (b) 0.032

(c) 0.022 (d) 0.052

12.62. Find the probability of obtaining exactly three 6s in five throws of a dice with probability of having a 6 on one throw being 2/6.

(a) 0.17 (b) 0.16

(c) 0.11 (d) 0.18

12.63. Trains arrive randomly at a railway station. The average number of trains arriving every three hours is five.
Find the probability that exactly four trains arrive in three hours

(a) 0.075 (b) 0.175

(c) 0.0275 (d) 0.185

12.64. Trains arrive randomly at a railway station. The average number of trains arriving every three hours is five.
Find the probability that exactly four trains arrive in four hours

(a) 0.1047 (b) 0.2047

(c) 0.3047 (d) 0.0047

AT A GLANCE

SPOTLIGHT

STICKY NOTES

12.65. Trains arrive randomly at a railway station. The average number of trains arriving every three hours is five.

Find the probability that less than two trains arrive in three hours

(a) 0.0304 (b) 0.0404
(c) 0.0004 (d) 0.0504

12.66. Trains arrive randomly at a railway station. The average number of trains arriving every three hours is five.
Find the probability that more than one train arrive in three hours

(a) 0.9596 (b) 0.8596
(c) 0.1596 (d) 0.3596

12.67. A company manufactures a component with a mean weight of 5kg and standard deviation of 1kg.
If the weights follow a normal distribution.
What is the probability that a component picked up at random will be more than 6kg

(a) 0.1785 (b) 0.1587
(c) 0.8715 (d) 0.1577

12.68. A company manufactures a component with a mean weight of 5kg and standard deviation of 1kg.

If the weights follow a normal distribution.

What is the probability that a component picked up at random will be less than 6kg

(a) 0.8413 (b) 0.1587
(c) 0.6 (d) 0.7

12.69. A company manufactures a component with a mean weight of 5kg and standard deviation of 1kg.

If the weights follow a normal distribution.

What is the probability that a component picked up at random will be more than 5kg but less than 6kg

(a) 0.1587 (b) 0.3413
(c) 0.8413 (d) 0.2783

12.70. A company manufactures a component with a mean weight of 5kg and standard deviation of 1kg.
If the weights follow a normal distribution.
What is the probability that a component picked up at random will be between 4kg and 6kg

(a) 0.6826 (b) 0.3413
(c) 0.8413 (d) 0.1587

ANSWERS TO SELF-TEST QUESTIONS

12.1	**12.2**	**12.3**	**12.4**	**12.5**	**12.6**
(b)	(a)	(a)	(b)	(a)	(a)
12.7	**12.8**	**12.9**	**12.10**	**12.11**	**12.12**
(b)	(b)	(b)	(b)	(a)	(a)
12.13	**12.14**	**12.15**	**12.16**	**12.17**	**12.18**
(a)	(b)	(b)	(b)	(b)	(b)
12.19	**12.20**	**12.21**	**12.22**	**12.23**	**12.24**
(c)	(a)	(a)	(a)	(b)	(a)
12.25	**12.26**	**12.27**	**12.28**	**12.29**	**12.30**
(b)	(b)	(c)	(b)	(a)	(c)
12.31	**12.32**	**12.33**	**12.34**	**12.35**	**12.36**
(a)	(b)	(b)	(b)	(b)	(a)
12.37	**12.38**	**12.39**	**12.40**	**12.41**	**12.42**
(b)	(b)	(a)	(b)	(c)	(a)
12.43	**12.44**	**12.45**	**12.46**	**12.47**	**12.48**
(a)	(a)	(c)	(b)	(c)	(a)
12.49	**12.50**	**12.51**	**12.52**	**12.53**	**12.54**
(a)	(b)	(d)	(b)	(a)	(a)
12.55	**12.56**	**12.57**	**12.58**	**12.59**	**12.60**
(b)	(a)	(b)	(c)	(b)	(a)
12.61	**12.62**	**12.63**	**12.64**	**12.65**	**12.66**
(b)	(b)	(b)	(a)	(b)	(a)
12.67	**12.68**	**12.69**	**12.70**		
(b)	(a)	(b)	(a)		

CHAPTER 13

SAMPLING AND SAMPLING DISTRIBUTIONS

IN THIS CHAPTER

AT A GLANCE

Often a researcher selects a sample of a population to test with the aim of extrapolating the results to the population as a whole. In order to determine an appropriate sample that best represents the population, various sampling methods are used. Of many sampling methods, this chapter discusses random, systematic, stratified, cluster, quota and multi-stage sampling methods.

Statistical inference refers to drawing a conclusion about a population from a sample. Different symbols are used to distinguish characteristics of a sample from those of the population when samples are used to estimate population parameters.

A sampling distribution in which a single attribute or statistic taken from many samples, for example, mean is referred to as sampling distribution of mean. Standard deviation of a sampling distribution of mean is called standard error. In estimating population mean from sample means, this chapter discusses confidence intervals, degrees of freedom and distribution tables.

AT A GLANCE
SPOTLIGHT
STICKY NOTES

AT A GLANCE
SPOTLIGHT
STICKY NOTES

1 SAMPLING AND SAMPLING DISTRIBUTIONS

Sample is a group of items or observations taken from the population for examination or experiment.

A population include all people or items with a characteristic that a researcher wishes to understand. This means that the term "population" means every single member of a particular group. Whereas a sample is a representative group from that population that is often taken or studied because of:

- cost – testing the whole population might be prohibitively expensive;
- feasibility – it might not be possible to test the whole population due to its size or location or the fact that the population might be dynamic (changing);
- perishability – testing might destroy the item (for example testing a bullet by firing it).

▶ *For example:*

A population would include the residents of Lahore or all Toyota buyers or all units of production from a specific machine.

In gathering information about a population, it is rarely possible or desirable to test (measure an attribute of each member of that population. Therefore, a group representative of all population is taken that is a sample of that population.

All efforts should be taken to make sure that the sample fairly represents the population at large.

▶ *For example:*

(a) Suppose a clothing retailer wanted to sell garments to male adults.

It would not be practical to measure the waist size of all men in the country. A sample would be selected and the waist size of each member of the sample would be measured. This could then be used as a basis of sizes ordered from the manufacturer for sale.

(b) Selecting a sample of men based entirely on students at a university would not result in a sample that was representative of the population as a whole.

This sample would be from a narrow age group which, by and large would be from the younger and fitter part of the population and would tend to have smaller waist sizes than the population as a whole.

If the retailer based its plans on such a sample it would be unable to sell to other age groups.

The three main advantages of sampling are:

- the cost is lower;
- data collection is faster; and
- since the data set is smaller it is possible to ensure homogeneity and to improve the accuracy and quality of the data.

Sampling frame

The sampling frame is a list of all the members of the population used as a basis for sampling. The frame, in effect, defines the population.

Suppose that an energy company wishes to obtain information on fuel consumption of cars in a country. The population would be all cars in the country. However, this is not specific enough for the researcher to draw a sample. The researcher would need a list of cars. Such a list might be all cars registered with the licensing authority. This then gives a specific population from which a sample can be drawn.

Sampling methods

The objective of any of these techniques is to extract a sample which is representative of the population. This sample is then tested and the results treated are representative of the full population.

There are different methods by which a sample might be extracted.

Random sampling

A random sample is one where every member of the population has an equal chance of being selected as a member of the sample. A random sample is a biased free sample.

▶ *For example:*

List of employees selected from the database using a random matrix.

Picking a unit of production randomly from the machine where the units are not labelled.

Possible problems with random sampling:

- It can be cumbersome when sampling from an unusually large target population.
- It can be prone to sampling error because the randomness of the selection may result in a sample that does not reflect the population.
- It is not suitable for investigators who are interested in issues related to subgroups of a population.

Systematic sampling

Systematic sampling relies on arranging the target population according to some ordering scheme and then selecting elements at regular intervals through that ordered list. The first item is selected at random and then every n^{th} member of the population is selected. The value for n is the population size divided by the sample size.

▶ *Formula:*

$$f = \frac{n}{s_n}$$

Where:

f = frequency interval

n = the total number of the wider population

s_n = the required number in the sample.

▶ *For example:*

Selecting every 10^{th} name from the telephone directory (an 'every 10^{th}' sample, also referred to as 'sampling with a skip of 10').

One problem with systematic sampling is that it is vulnerable to periodicities in the list and this might result in the sample being unrepresentative of the overall population.

Suppose a researcher is interested in household income. There might be a street where the odd-numbered houses are all on the north (expensive) side of the road, and the even-numbered houses are all on the south (cheap) side. If the sample started with say number 7 and then every 10^{th} house all houses sampled would be from the expensive side of the street and the data collected would not give a fair reflection of the household incomes of the street.

Stratified sampling

Where the population is split into relevant strata, the sample may also be split into strata in proportion to the population. Within each stratum, the sample is selected using random sampling methods.

▶ *For example:*

Suppose a clothing retailer wanted to sell garments to male and female adults.

The population would be split (stratified) into male and female for sampling purposes. The population could be further stratified by age groups.

One problem with stratified sampling is that the approach can increase the cost and complexity of sample selection, as well as can lead to increased complexity of population estimates.

Advantages of stratified sampling include

- dividing the population into distinct, independent strata which enable researchers to draw inferences about specific subgroups that may be lost in a random sample.
- stratification can lead to more efficient statistical estimates (provided that strata are selected based upon relevance to the criterion in question, instead of availability of the samples).
- data might be more readily available for individual, pre-existing strata within a population than for the overall population;
- different sampling approaches can be applied to different strata, potentially enabling researchers to use the approach best suited (or most cost-effective) for each identified subgroup within the population.

Note that even if a stratified sampling approach does not lead to increased statistical efficiency, it will not result in less efficiency than simple random sampling. Provided that each stratum is proportional to the group's size in the population.

Cluster sampling

This is very similar to multi-stage sampling and is likely to be used when the population is large. Population is divided into smaller segments, groups or clusters based on their similar characteristics and sample is selected at random or systematically from each group.

▶ *For example:*

Interviewers are sent into the areas to interview every person who fits a given definition (for example, mothers with children under 5).

Multi-stage sampling

This technique involves taking random samples of preceding random samples. It is used in nationwide surveys to cut down on travelling. It is an extension to cluster sampling.

The approach involves:

- dividing the country into a series of areas;
- picking a random sample of these areas;
- within each of these, picking a random sample of towns;
- within each of these pick a random sample of people.

All people selected at the last stage are then used for data collection.

It is another way of saving time and money but still having an element of random selection. It is very useful when a complete list of all members of a population does not exist.

▶ *For example:*

For survey of teachers' satisfaction in the city, sample is selected as follows

1. Levels are selected to focus on (for example, ECD, middle, high or secondary teachers)
2. Within each level, grades are selected (for example, 1-2, 3-5, 8 or above)
3. Within each levels, schools are selected randomly
4. Within each school teachers are selected randomly as per the level and grades

Quota sampling

A sampling method where population can be divided into group (of homogenous characteristic) and samples are taken from each group (quota).

- *For example:*

  An interviewer may be told to sample 200 females and 300 males between the age of 45 and 60.

Convenience sampling

As the name suggests, convenience sampling is based on the availability of respondents or observations. It is also called accidental or opportunity sampling. Ease of access and respondent willingness is the key for convenience sampling.

One problem with convenience sampling is lack of generalizability of the results since the data may be biased or inclined towards a segment and may not be representative of the whole population.

Sampling theory

Sampling is used to investigate the characteristics of a population by studying small groups (samples) taken from that population.

There are two main types of problem that can be addressed:

- statistical inference; and
- hypothesis testing.

A statistic (measure relating to a sample) is used as the basis of an estimate of a parameter (measure relating to a population).

Hypothesis testing refers to testing whether an assertion about a population is likely or not supported by the sample.

- *Illustration:*

	Sample	Population
Mean	$\bar{x}$	μ
Standard deviation	s	σ
	$s = \sqrt{\dfrac{\sum(x-\bar{x})^2}{n-1}}$	$\sigma = \sqrt{\dfrac{\sum(x-\mu)^2}{N}}$
	Where: $\bar{x}$ = Sample mean n= number in the sample	Where: μ = Population mean N = Number in the population

Note: It is unlikely to know the actual mean for a population. That is why statistical inference is used to estimate it. Therefore, we would not be able to calculate the population standard deviation.

The overall approach involves the application of knowledge of probability distributions.

- The normal distribution is used in statistical inference and hypothesis testing using large (30 or more observations) samples (z tests).
- The t distribution is used in statistical inference and hypothesis testing when only small sample sizes are available from a normally distributed population (t tests).
- The Chi squared (χ^2) distribution for hypothesis testing in certain circumstances. (***Chi*** is pronounced as ***ki*** as in kite).

Sampling distributions

A sampling distribution is a distribution of a sample characteristic, selected from many samples and is used to determine how close to the population parameter a sample statistic is likely to fall.

In understanding a population, a sample statistic may be calculated to conclude or estimate a characteristic of a population. With every distinct sample, the statistic may differ and each would have said to be a representative of a population.

The sampling distribution is a probability distribution of a sample statistic.

- *For Example:*

  An investigator is interested in average height of men in a population.

  He constructs samples of all possible combinations of 100 men in the population and calculates the mean of each sample.

  The investigator then has a list of means of the samples.

  The means form a distribution.

  The distribution described would be called a sampling distribution of the mean.

Sampling distributions have certain characteristics which make them very useful.

The mean of a sampling distribution is the same as the mean of the population.

A sampling distribution is a normal distribution (even if the population itself does not follow a normal distribution).

- It is symmetrical about its mean (even if the distribution within the population being sampled is not symmetrical).
- It has the same mathematical characteristics of all other normal distributions.

A sampling distribution will rarely, if ever, exists in practice. It is a theoretical distribution. However, the facts that it could theoretically exist, and that we know its mathematical characteristics, are very useful.

2 SAMPLING DISTRIBUTION OF THE MEAN

A sampling distribution of the mean is a distribution made up of the means of many samples.

This is the sampling distribution obtained by taking all possible samples of a fixed size n from a population, noting the mean of each sample and classifying the means into a distribution.

- *For example:*

  Suppose the variable under investigation is the height of men in Pakistan in the 25 to 54 age range. (suppose there are 34 million men in this age range).

  Theoretically, a researcher could construct every possible sample of 1,000 men, measure their height and take the mean of each sample. (Note that this would be impossible to do. All possible combinations of 1,000 at a time taken from a population of 34 million would be an astronomical figure).

  The means of these samples would constitute the sampling distribution of the mean height of 24 to 54 year old men in Pakistan. This would be a normal distribution with the same mean as the population mean.

The sampling distribution of the mean is a normal distribution with the same mean as that of the population. The standard deviation of a sampling distribution is called a standard error. (Sometimes people become confused by this term but it is just a name. All distributions have a standard deviation but the standard deviation of a sampling distribution is called standard error).

This means that 95% of the means of the samples collected fall within 2 (actually 1.96) standard errors of the mean of the sampling distribution.

Therefore, the mean of any single sample has a 95% chance of being within 2 standard errors of the mean of the population.

Standard error

As discussed earlier, the standard deviation of a sampling distribution of the mean is called the standard error of the mean.

A standard error could be calculated from the sampling distribution of the mean just like any other standard deviation. However, in most cases the sampling distribution of the mean is theoretical and its variance and standard deviation (standard error) are unknown.

Another method is needed to find the standard error. There is a mathematical relationship between the standard deviation of the population and the standard error of the mean.

- *Formula:*

$$Standard\ error\ of\ the\ mean = \frac{Standard\ deviation\ of\ the\ population}{\sqrt{Sample\ size}}$$

$$Standard\ error\ of\ the\ mean = \sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Where:

$\sigma_{\bar{x}}$ = standard error (standard deviation of the sampling distribution of the mean)

s = standard deviation of the sample

n = sample size

Please note that:

This formula is used to estimate the standard deviation earlier.

$$\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$$

When the standard deviation of a sample is used as a proxy for the standard deviation of a population it must be calculated using this formula.

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

At first sight this does not seem to make us any better off as we do not usually know the population standard deviation either. However, the standard deviation of a sample is usually very similar to the standard deviation of the population from which the sample is taken and this is used as a proxy.

The above box shows a slightly different way of calculating the standard deviation. This adjustment is made to remove bias from the sample. The sample standard deviation, as calculated in the usual way, is a biased predictor of the standard deviation of the population. This adjustment is made to remove bias from the sample.

Note: Adjustment has little effect when the samples are large so is not always necessary. However, in questions where it is asked to calculate a standard deviation of a sample, this approach can be used to estimate the standard error of the mean.

- *For example:*

A random sample of 100 items has a standard deviation of 25. Calculate the standard error of the mean

Let s = 25, n=100

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{25}{\sqrt{100}} = \frac{25}{10} = 2.5$$

A random sample of 50 items has a standard deviation of 35. Calculate the standard error of the mean

Let s = 35, n=50

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{35}{\sqrt{50}} = \frac{35}{7.07} = 4.95$$

A random sample of 50 items has been taken and the mean measured. The standard deviation of the sample has been calculated as:

$$\boldsymbol{Standard\ deviation} = \sqrt{\frac{\boldsymbol{\Sigma(\bar{x}-x)}}{\boldsymbol{n}}} = \sqrt{\frac{\mathbf{500}}{\mathbf{50}}} = \mathbf{3.16}$$

Calculate the standard error of the mean.

The first step is to recalculate the standard deviation to remove bias

$$\text{Standard deviation} = s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{\frac{500}{50-1}} = \sqrt{\frac{500}{49}} = 3.19$$

The sample standard deviation can then be used to find the standard error.

Will put s =3.19 and n=50

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{3.19}{\sqrt{50}} = \frac{3.19}{7.07} = 0.451$$

Estimating population means:

A sample can be taken and a variable (from that sample) can be used to give an estimate of the value of the variable in the population at large.

When a single figure is used as an estimate of a parameter of a population is called point estimate.

- *For example*

  A sample of 100 male ICAP students had a mean height of 170cm.

  This can be used as a point estimate for the mean height of all male ICAP students.

  A sample mean of height could be used as a point estimate of the population mean height.

However, it is unlikely that a point estimate of a mean taken from a single sample would be exactly right. In other words, it is unlikely that the mean from a sample would be exactly the same as the mean of the population. It would be useful to provide information about the margin of error in the point estimate.

Therefore, instead of asking for a point estimate usually a calculation of a range (confidence interval) within which the mean would be expected to lie at a given level of confidence is required.

Confidence levels

Information about the margin of error is given by presenting a range within which the population mean is expected to lie at a specified confidence level.

We can do this by using our knowledge of the sampling distribution of the mean.

- 95% of all samples will lie within 2 (1.96) standard errors of the population mean.
- 99% of all samples lie within 2.58 standard errors of the population mean.

This information can be used to express a measure of confidence in the point estimate.

- *Illustration:*

  A sample of 100 male ICAP students had a mean height of 170cm with a standard deviation of 8cm.

  $$Standard\ error = \sigma = \frac{s}{\sqrt{n}} = \frac{8}{\sqrt{100}} = \frac{8}{10} = 0.8$$

  We can be 95% sure that the population mean:

  $$\mu = \bar{x} \pm 1.96\sigma = 170 \pm (1.96 \times 0.8) = 168.43\ to\ 171.57$$

  Conclusion: We are 95% confident that the mean height of all male ICAP students is between 168.43 and 171.57 cm

The values at each end of the range are called confidence limits. The range itself is called a confidence interval (or an interval estimate).

The population mean cannot be measured by this process. The process is only able to specify a range within which the population mean is likely to be at a given level of confidence.

Following terms may be useful:

- **Confidence level** refers to the percentage of all possible samples that can be expected to include the true population parameter. For example, suppose all possible samples were selected from the same population, and a confidence interval were computed for each sample. A 95% confidence level implies that 95% of the confidence intervals would include the true population parameter.
- **Confidence interval** gives an estimated range of values which is likely to include an unknown population parameter, an estimated range being calculated from a given set of sample data.

If independent samples are taken repeatedly from the same population, and a confidence interval calculated for each sample, then a certain percentage (confidence level) of the intervals will include the unknown population parameter.

- **Confidence limits** are the values at each end of a confidence interval.

Confidence intervals are usually calculated so that this percentage is 95% (although any percentage can be used).

If many samples are taken, we would expect each sample to identify a range containing the mean on 95% or 19 out of 20 times. This means that there is a 5% (1 in 20) chance that the population mean is identified as being in a range when in fact it is not.

If the chance of being wrong 1 time in 20 is too great a risk to take, a higher confidence level should be used.

▶ *Illustration:*

In the example above where a sample of 100 male ICAP students had a mean height of 170cm with a standard deviation of 8cm.

Where,

$$Standard\ error = \sigma = 0.8$$

We can be 99% sure that the population mean:

$\mu = \bar{x} \pm 2.58\sigma = 170 \pm (2.58 \times 0.8) = 167.94\ to\ 172.06$

Conclusion: We are 99% confident that the mean height of all male ICAP students is between 167.94 and 172.06cm.

▶ *For examples:*

On a random stock check of 40 items, the mean value per item was found to be Rs 380 and the standard deviation Rs 100. Calculate 95% confidence limits for the overall mean value per item.

From the given example we have, n=40, $\bar{x} = 380$, s =100

Standard error of the mean $= \sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{100}{\sqrt{40}} = \frac{100}{6.325} = 15.8$

We can be 95% sure that the population mean

$\mu = \bar{x} \pm 1.96\sigma_{\bar{x}} = 380 \pm (1.96 \times 15.8) = 380 \pm 31$

Conclusion: We are 95% confident that the mean value per items is between Rs 349 and Rs 411.

A dock uses a mechanical shovel to load grain. A random sample of 50 scoops of the shovel has a mean weight of 510 kg and a standard deviation of 30 kg. Construct a 95% confidence interval for the mean of the population (being all scoops)

From the given example we have, n=50, $\bar{x} = 510$, s =30

Standard error of the mean $= \sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{30}{\sqrt{50}} = \frac{30}{7.07} = 4.24$

We can be 95% sure that the population mean:

$\mu = \bar{x} \pm 1.96\sigma = 510 \pm (1.96 \times 4.24) = 510 \pm 8.3$

Conclusion: We are 95% confident that the mean value weight is between 501.7 kg and 518.3 kg.

AT A GLANCE
SPOTLIGHT
STICKY NOTES

A company is testing the breaking strengths of cables measured in Newtons per square millimetre (N/mm2). A random sample of 50 cables had a mean breaking strength of 1,600 N/mm2 with a standard deviation of 60 N/mm2. Construct a 99% confidence interval for the mean breaking strength.

From the given example we have, n=50, $\bar{x} = 1600$, s =60

$$\text{Standard error of the mean} = \sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{60}{\sqrt{50}} = \frac{60}{7.07} = 8.48$$

We can be 99% sure that the population mean:

$$\mu = \bar{x} \pm 2.58\ \sigma = 1{,}600 \pm (2.58 \times 8.48) = 1{,}600 \pm 21.9$$

Conclusion: We are 99% confident that the mean breaking strength is between 1,578.1 N/mm2 and 1,621.9 N/mm2

Construction of Confidence Intervals:

Confidence intervals are constructed as follows

- Step 1: Identify the parameter being investigated (say μ)
- Step 2: Determine the confidence level.
- Step 3: Identify the formula for the required confidence level

$$\bar{x} \pm Z\ value\ (e.g. 1.96)\ standard\ errors$$

If the population standard deviation is known use it in the standard error expression	$\bar{x} \pm Z\ value\ \times \frac{\sigma}{\sqrt{n}}$
If the population standard deviation is unknown use the sample standard deviation in the standard error expression	$\bar{x} \pm Z\ value\ \times \frac{s}{\sqrt{n}}$

Improving accuracy

The above method for estimating the population mean from that of a sample mean results in a range of values (interval estimate or confidence interval) in which there is a level of confidence that the mean lies.

The range is a function of the standard error and number of standard errors appropriate to the confidence level chosen.

The standard error is a function of sample size. If the sample size is increased, the standard error falls.

If an investigator wants to narrow the range of possible values, he can increase the sample size. A sample size can be calculated which results in a range of a required size.

▶ *For example:*

A sample of 100 male ICAP students had a mean height of 170cm with a standard deviation of 8cm. Calculate the sample size that will result in a confidence interval of the mean ± 1cm for this data.

From the give question we have, n=100, $\bar{x} = 170$, s =8

We know that

$$Standard\ error = \sigma = \frac{s}{\sqrt{n}}$$

And the confidence limits are:

$$\mu = \bar{x} \pm (1.96 \times Standard\ error)$$

However, the required range is where 1.96 × Standard error = 1.

Therefore, using the formula for standard error and using the values we have:

$$1.96 \times \frac{8}{\sqrt{n}} = 1$$

Rearranging

$$1.96 \times 8 = 1 \times \sqrt{n}$$

$$15.68 = \sqrt{n}$$

Square both sides of the equation

$$15.68^2 = n$$

Therefore

$$246 = n$$

If the sample size is set at 246 it will result in confidence interval of the mean ± 1cm for this data. this can be verified as follow:

$$Standard\ error = \sigma = \frac{s}{\sqrt{n}} = \frac{8}{\sqrt{246}} = 0.51$$

$$Proof: \mu = \bar{x} \pm 1.96\sigma = 170 \pm (1.96 \times 0.51) = 170 \pm 1$$

On a random stock check of 40 items, the mean value per item was found to be Rs 380 and the standard deviation Rs 100. Calculate the sample size that will result in a confidence interval of the mean ± Rs.20 for this data at the 95% confidence level.

For the given question,

Required range = 1.96 × Standard error = 20

Which means

$$1.96 \times \frac{\mathrm{s}}{\sqrt{\mathrm{n}}} = \text{Confidence interval}$$

$$1.96 \times \frac{100}{\sqrt{\mathrm{n}}} = 20$$

$$\frac{196}{20} = \sqrt{n}$$

$$\left(\frac{196}{20}\right)^2 = \mathrm{n} = 96.04$$

Conclusion: A sample size of 96 must be taken to produce the required confidence interval.

3 SAMPLING DISTRIBUTION OF A PROPORTION

Suppose the statistic under investigation is the proportion of men to women in an area. Theoretically, a researcher could construct every possible sample of 1,000 people in the area and measure the proportion of men in each sample.

The proportions from the samples would constitute the sampling distribution of the proportion of men in the area. This would be a normal distribution.

It can be used in similar ways to the sampling distribution of the mean

- *Formula:*

$$Standard\ error\ of\ a\ proportion = \sigma_p = \sqrt{\left(\frac{pq}{n}\right)}$$

Where:

p = population proportion (the sample proportion is used as a proxy for this in cases when it is not known - usually the case)

q = 1- p

n = sample size

- *For example:*

In a sample of 500 people 215 of them were males.

(a) Estimate the population proportion at the 95% confidence level.

$$Sample\ proportion = \frac{215}{500} = 0.43$$

$$\sigma_p = \sqrt{\left(\frac{pq}{n}\right)} = \sqrt{\left(\frac{0.43(1-0.43)}{500}\right)} = \sqrt{\frac{0.43 \times 0.57}{500}} = 0.0221$$

We can be 95% sure that the population proportion is:

$$\mu = \bar{x} \pm 1.96\sigma = 0.43 \pm (1.96 \times 0.0221) = 0.387\ to\ 0.473$$

Conclusion: We are 95% confident that the population proportion of men is between 0.387 (38.7%) and 0.473 (47.3%).

(b) Calculate the sample size that will result in a confidence interval of the sample proportion ± 2% at the 95% confidence level.

$$Sample\ proportion = \frac{215}{500} = 0.43$$

$$\sigma_p = \sqrt{\left(\frac{pq}{n}\right)} = \sqrt{\left(\frac{0.43(1-0.43)}{n}\right)} = \sqrt{\frac{0.2451}{n}}$$

The confidence limits are:

$$\mu = \bar{x} \pm (1.96 \times Standard\ error)$$

AT A GLANCE

SPOTLIGHT

STICKY NOTES

The required range is where 1.96 × Standard error = 2%:

$$1.96 \times \sqrt{\frac{0.2451}{n}} = 0.02$$

Divide both sides by 1.96

$$\sqrt{\frac{0.2451}{n}} = \frac{0.02}{1.96}$$

Square both sides of the equation

$$\frac{0.2451}{n} = \left(\frac{0.02}{1.96}\right)^2$$

Rearranging

$$\frac{0.2451}{\left(\frac{0.02}{1.96}\right)^2} = n$$

Therefore

$$2{,}354 = n$$

If the sample size is set at 2,354 it will result in a confidence interval of the sample proportion ± 2% at the 95% confidence level.

4 ESTIMATION OF POPULATION MEAN – SMALL SAMPLE SIZE

Standard errors of a distribution with large sample were discussed in the earlier sections.

However, as sample sizes reduce in size this approach becomes less accurate. Following two modifications can be adopted:

Modification 1

The first modification has already been discussed. Using the standard deviation as a proxy for that of the population leads to slight bias as it is usually slightly smaller than that of the population.

The bias is not important as long as sample sizes are large but it increases as sample size falls. In order to compensate, an adjustment must be made in calculating the standard deviation of the sample for the purpose of inclusion in the standard error formula.

- *Formula:*

$$Standard\ error = \frac{Standard\ deviation\ of\ the\ sample}{\sqrt{Sample\ size}} = \sigma = \frac{s}{\sqrt{n}}$$

Where

$$\mathrm{s} = \sqrt{\frac{\sum(x-\overline{x})^2}{\mathrm{n}-1}}$$

Earlier, the text recommended that you always calculate the standard deviation of a sample. In this way when the sample is to be used in statistical inference even though it has only a small effect with large samples.

For small samples the effect cannot be ignored and must be used to calculate standard deviation.

Modification 2

Another distribution is used instead of the normal distribution. This is the ***t* distribution**. It is used in a similar way as the normal distribution by looking at proportions of the area of the distribution within a number of standard deviations from the mean but the numbers are slightly different.

T distribution tables are published with areas related to a selected number of levels of confidence.

A peculiarity of the ***t*** distribution is that it varies with sample size. In fact, it is more accurate to refer to ***t*** distributions rather than a single ***t*** distribution.

As the sample size increases the ***t*** distribution approaches the normal distribution. This is reflected in the number of degrees of freedom in the data.

Degrees of freedom

This is a concept found a lot in advanced statistics. You do not need to understand it but you do need to be able to calculate the number of degrees of freedom that relate to a problem. Luckily, this is straightforward as the number of degrees of freedom in this case is ***n* – *1***.

T distribution tables

The approach adopted in using the t distribution is the same as when using the normal distribution. Instead of reading a ***z* value** of a normal distribution table a ***t* value** is read off a t distribution table.

The ***t*** distribution table is structured differently to the normal distribution table.

The normal curve is structured to show the area of the curve within a number of standard deviations (***z* value**) from the mean given by values in the left hand column and the top row.

The t distribution is structured to the ***t value*** that includes (or excludes) a percentage of the curve at a number of degrees of freedom. The percentage (confidence levels) chosen is on the top row and the number of degrees of freedom in the left hand row.

However, once the values are in hand they are used in the same way.

- *Illustration:*

Full tables are given as an appendix to this text.

	Confidence levels	
df	**95%**	**99%**
7	2.3646	2.9980
8	2.3060	2.8965
9	2.2621	2.8214

At 9 degrees of freedom, 95% of the area under the curve is within 2.2621 standard errors of the mean.

Where:

df = degrees of freedom (n – 1)

Steps to estimate population mean:

1. Find the sample mean.
2. Estimate the sample standard deviation.
3. Estimate the standard error using the sample standard deviation.
4. Estimate the number of degrees of freedom.
5. Decide on the confidence level and use the table to find the required level for ***t*** at this confidence level and for the number of degrees of freedom.
6. Apply the formula to estimate the population mean.

- *For example:*

Estimate the population mean at the 95% confidence level from the following sample:

5.5, 6.1, 5.4, 5.8, 5,3, 4.8, 5.2, 4.6, 6.1, and 5.2

In order to calculate the mean:

$$\bar{x} = \frac{\sum x}{n} = \frac{54}{10} = 5.4$$

In order to calculate the Standard Deviation of the sample:

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

AT A GLANCE SPOTLIGHT STICKY NOTES

x	$(x-\bar{x})$	$(x-\bar{x})^2$
5.5	0.1	0.01
6.1	0.7	0.49
5.4	0.0	0.00
5.8	0.4	0.16
5.3	−0.1	0.01
4.8	−0.6	0.36
5.2	−0.2	0.04
4.6	−0.8	0.64
6.1	0.7	0.49
5.2	−0.2	0.04
54		2.24

$$s = \sqrt{\frac{2.24}{10-1}} = 0.5$$

Standard error

$$\sigma = \frac{s}{\sqrt{n}} = \frac{0.5}{\sqrt{10}} = 0.158$$

Number of degrees of freedom (df)

$$n - 1 = 10 - 1 = 9$$

From table: t score at 95% confidence level with 9 df (t0.95/9) =2.262

We can be 95% sure that the population mean:

$$\mu = \bar{x} \pm t_{\frac{0.95}{9}} \times \sigma = 5.4 \pm (2.262 \times 0.158) = 5.4 \pm 0.36 = 5.04 \text{ to } 5.76$$

Conclusion: We are 95% confident that the mean is between 5.04 and 5.76.

AT A GLANCE

SPOTLIGHT

STICKY NOTES

STICKY NOTES

Sample is a group of items or observations taken from the population for examination or experiment.

A random sample is one where every member of the population has an equal chance of being selected as a member of the sample. A random sample is a bias free sample.

Systematic sampling relies on arranging the target population according to some ordering scheme and then selecting elements at regular intervals through that ordered list.

A sampling distribution is a distribution of a sample characteristic, selected from many samples and is used to determine how close to the population parameter a sample statistic is likely to fall.

A sampling distribution of the mean is a distribution made up of the means of many samples. The standard deviation of a sampling distribution of the mean is called the standard error of the mean.

$$\text{Standard error of the mean} = \sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

A sample can be taken and a variable from that sample and can be used to give an estimate of the value of the variable in the population at large.

Confidence level refers to the percentage of all possible samples that can be expected to include the true population parameter. A 95% confidence level implies that 95% of the confidence intervals would include the true population parameter.

Confidence interval gives an estimated range of values which is likely to include an unknown population parameter. Estimated range being calculated from a given set of sample data and Confidence limits are the values at each end of a confidence interval

SELF-TEST

13.1 The probability distribution of a statistic is called a:

(a) Probability distribution (b) Sampling distribution

(c) Frequency distribution (d) None of these

13.2 The probability distribution of $\bar{x}$ is called:

(a) Probability distribution of mean (b) Sampling distribution of mean

(c) Frequency distribution of mean (d) None of these

13.3 It is customary to refer to the standard deviation of the sampling distribution as the:

(a) Variance (b) Standard deviation of mean

(c) Standard error (d) Mean deviation

13.4 If all possible samples of size n are drawn, without replacement, from a finite population of size N with mean μ and standard deviation σ, then the sampling distribution of the sample mean $\bar{x}$ will be approximately normally distributed with a mean and standard deviation given by:

(a) $\mu_{\bar{x}} = \bar{x} \text{ and } \sigma_{\bar{x}} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ (b) $\mu_{\bar{x}} = \mu \text{ and } \sigma_{\bar{x}} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

(c) $\mu_{\bar{x}} = \bar{x} \text{ and } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ (d) None of these

13.5 The quantity $\sqrt{\frac{N-n}{N-1}}$ is known as:

(a) Sampling fraction (b) Population correction

(c) Finite population correction (d) None of these

13.6 If the population size is infinite or N is large as compared to n, then the finite population correction factor will be approximately equal to:

(a) Zero (b) One

(c) Infinite (d) None of these

13.7 If n = 10 and N = 1000 the value of finite population correction factor (fpc) is:

(a) 0.9909 (b) 1

(c) 0.9954 (d) None of these

13.8 If n = 2 and N = 5 the value of fpc is:

(a) 0.75 (b) 0.86
(c) 0.866 (d) None of these

13.9 If n = 40 and N = 10,000 the value of fpc is:

(a) 0.99 (b) 0.996
(c) 0.998 (d) None of these

13.10 If N = 4, n = 2, μ = 5.25; σ^2 = 2.1875 then the values of $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$ if sampling is done without replacement:

(a) 5.25 and 1.2076 (b) 5.25 and 0.8539
(c) 5.25 and 0.729 (d) None of these

13.11 If n = 2, μ = 5.25 and σ^2 = 2.1875 then the values of $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$ if sampling is done with replacement, are:

(a) 5.25 and 1.09375 (b) 5.25 and 1.0458
(c) Cannot be determined (d) None of these

13.12 From a given population with $\sigma = 5.6$ kilograms samples were drawn with replacement, how may the standard error of the mean change, when the sample size is increased from 64 to 196:

(a) Reduced from 0.7 to 0.4 (b) Increased from 0.4 to 0.7
(c) Reduced from .0875 to .0286 (d) Increased from .0286 to .0875

13.13 From a given population with $\sigma = 5.6$ kilograms samples were drawn with replacement, how may the standard error of the mean change, when the sample size is decreased from 784 to 49:

(a) Decreased from 0.8 to 0.2 (b) Increased from 0.2 to 0.8
(c) Increased from 0.0071 to 0.114 (d) Decreased from 0.114 to 0.0071

13.15 A sampling process that selects every k^{th} element in the population for the sample, with the starting point determined at random from the first k elements is known as:

(a) Sample random sampling (b) Stratified random sampling
(c) Systematic random sampling (d) None of these

13.16 A sampling process that selects simple random samples from mutually exclusive sub populations, of population is called:

(a) Simple random sampling (b) Stratified random sampling
(c) Systematic random sampling (d) None of these

13.17 A sampling process which selects samples from the given population with equal chance of selection to each unit is called:

(a) Systematic Random sampling (b) Stratified random sampling

(c) Simple random sampling (d) None of these

13.18 A Sampling distribution is:

(a) an exclusive distribution (b) mean distribution

(c) a-symmetric distribution (d) normal distribution

13.19 Any sampling procedure that produces inferences that consistently over estimate or consistently under estimate some characteristics is said to be:

(a) Simple random sampling (b) Systematic sampling

(c) Biased (d) None of these

13.20 In simple random sampling each unit of the population has __________ chance of selection:

(a) equal (b) Unequal

(c) maximum (d) None of these

13.21 ____ is a group of items or observations taken from the population for examination or experiment.

(a) Sampling frame (b) Sampling distribution

(c) Sample (d) Units

13.22 Which of the following statements is true?

1. Sample is a group of items or observations taken from the population for examination or experiment.
2. Sampling frame is a list of all the members of the population used as a basis for sampling.
3. Cost saving is one of the advantages of sampling.
4. Time saving is not an advantage of sampling.

13.23 ____ sampling relies on arranging the target population according to some ordering scheme and then selecting elements at regular intervals through that ordered list.

(a) Systematic (b) Random

(c) Stratified (d) Cluster

13.24 A statistic is used as the basis of an estimate of a ____.

(a) Sample (b) Population

(c) Parameter (d) Mean

AT A GLANCE

SPOTLIGHT

STICKY NOTES

13.25 A sampling distribution of the mean is a distribution made up of the means of many ___.

(a) Populations (b) Samples

(c) Standard deviations (d) Standard errors

13.26 Compute standard error of the mean from following data:

1. standard deviation of the population: 500
2. Sample size: 100

(a) 50 (b) 10

(c) 500 (d) 2

13.27 Compute standard deviation of the population from following data:

1. Standard error of the mean: 50
2. Sample size: 100

(a) 50 (b) 500

(c) 100 (d) 5

13.28 A population consists of five numbers. How many samples of size two can be drawn from this population with replacement?

(a) 25 (b) 2

(c) 10 (d) 20

13.29 By how much will standard error change if sample size increases from 100 to 121

(a) Decrease by 9.09% (b) Increase by 21%

(c) Cannot be determined with available data (d) Increase by 9.09%

13.30 A machine produces certain component, sample of 40 such components was found to have mean of 380 and standard deviation of 100. Calculate 99% confidence limits for the overall production.

(a) 339.52 to 420.48 (b) 380 to 420

(c) 334.6 to 415.7 (d) 333.25 to 416.2

13.31 Confidence ___ are the values at each end of a confidence interval.

(a) Limits (b) Interval

(c) Level (d) region

13.32 ___ is obtained if both confidence limits are added together and divided by two.

(a) Standard error (b) Standard deviation

(c) Mean (d) Median

13.33 If the sample size increases, what will happen to standard error

(a) Increases (b) Decreases

(c) Remains unchanged (d) Cannot be determined with limited data

13.34 Compute confidence interval for the following data:

1. Confidence level: 95%

2. Standard error: 0.8

3. Sample mean: 5

(a) 3.3 to 4.3 (b) 3.43 to 6.57

(c) 3.5 to 7.5 (d) 3.2 to 6.2

13.35 Compute confidence interval for the following data:

1. Confidence level: 95%
2. Standard error: 0.8
3. Sample mean: thrice of standard error

(a) 0.832 to 3.968 (b) 0.5 to 3.5

(c) 1.832 to 2.832 (d) 1.832 to 4.832

13.36 Compute confidence interval for the following data:

1. Confidence level: 95%

2. Standard deviation of population: 15

3. Sample size: 9

3. Sample mean: 15

(a) 5.2 to 24.8 (b) 4.2 to 25.8

(c) 5.3 to 24.8 (d) 5.3 to 24.9

13.37 What happens to confidence interval if confidence level is increased?

(a) Increases (b) Decreases

(c) Remains unchanged (d) Cannot be determined

13.38 What happens to confidence interval if confidence level is decreased?

(a) Increases (b) Decreases

(c) Remains unchanged (d) Cannot be determined

13.39 What will happen to standard error of mean if sample size is decreased?

(a) Increases (b) Decreases

(c) Remains unchanged (d) Cannot be determined

13.40 What will happen to standard error of mean if sample size is increased?

(a) Increases (b) Decreases

(c) Remains unchanged (d) Cannot be determined

ANSWERS TO SELF-TEST QUESTIONS

13.1	13.2	13.3	13.4	13.5	13.6
(b)	(b)	(c)	(c)	(c)	(b)
13.7	13.8	13.9	13.10	13.11	13.12
(c)	(c)	(c)	(b)	(b)	(a)
13.13	13.14	13.15	13.16	13.17	13.18
(b)	(c)	(b)	(c)	(d)	(d)
13.19	13.20	13.21	13.22	13.23	13.24
(c)	(a)	(c)	1,2,3	(a)	(c)
13.25	13.26	13.27	13.28	13.29	13.30
(b)	(a)	(b)	(a)	(a)	(a)
13.31	13.32	13.33	13.34	13.35	13.36
(a)	(c)	(b)	(b)	(a)	(a)
13.37	13.38	13.39	13.40		
(a)	(b)	(a)	(b)		

CHAPTER 14

HYPOTHESIS TESTING

IN THIS CHAPTER

AT A GLANCE

Significance testing is concerned with testing claims or assertions based on a random sample. It closely resembles the construction of confidence intervals and the approach involve developing hypothesis and testing it. hypothesis is usually stated according to population parameter. At a given significance level, hypothesis are accepted or rejected using a z-score.

Z-Score tells the number of standard deviations between sample statistic and population statistic.

The chapter discusses significance test of means, difference between means, proportions, difference between proportions and small samples.

1 SIGNIFICANCE TESTING

Significance testing is a term given to any test devised to check if a difference between a sample characteristic and that of the population are only due to chance or not. If they are not solely due to chance they are said to be statistically significant. This means that they signify a difference between the sample and the population.

- *For example:*

  A machine is programmed to fill bags with 2 kgs of flour. A sample of bags shows a mean weight of 1.95 kg.

  There is bound to be some variation in any process so not all bags would be expected to receive 2 kgs exactly.

  The management of the business would be concerned to know whether the weights observed in this sample is due to chance or whether the weights indicate that the machine is not operating as it should be. If this is found to be the case the difference is said to be statistically significant.

Steps for significance testing:

The approach involves setting up a theory about a population characteristic and then taking a sample to see whether the theory is supported or not. Note that the test cannot prove the hypothesis but can only provide evidence to support or reject it.

- STEP 1: Set up the null hypothesis (H_0)
  - This is what is being tested.
  - It is usually phrased in terms of there being no difference between a sample and the population and that any difference is due to chance.
- STEP 2: Set up the alternate hypothesis (H_1) – This is the hypothesis that is accepted if the null hypothesis is rejected.

- *For example:*

  If the average bag of flour is 2kg, for a sample null and alternate hypothesis could be stated as:

  H_0= The weight of a bag of flour is equal to 2 kgs

  H_1= The weight of a bag of flour is different to or less than 2 kgs

  Hypothesis are usually stated according to the population parameter, such as mean, proportion.

- STEP 3: Determine a significance level – This is similar to the concept of confidence level explained earlier. This time though we are picking a limit beyond which we would conclude that any difference is significant or not.

  Significance levels are often stated as the inverse of confidence levels. Thus a confidence level of 95% might be stated as the 5% significance level.

  Confidence level + significance level = 100%; or

  Significance level = 100% – Confidence level;

- *For example:*

  At the 95% confidence level, 1 in 20 samples would be expected to display a statistic which fell outside 1.96 standard deviations from the mean in a two tailed test or above or below 1.65 standard deviations from the mean in a one-tailed test.

- STEP 4: Calculate the standard error or standard deviation of sampling distribution.

$$\text{Standard error} = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \; or \; \frac{s}{\sqrt{n}}$$

Where

$\sigma_{\bar{x}} = standard\ error\ or\ standard\ deviation\ of\ sampling\ distribution$

σ = population standard deviation

n= sample size

s= sample standard deviation

Usually, population standard deviation is not known, instead sample standard deviation (s) is used that is estimated to be representative of the population. This is true only where sample size is large.

- Select an appropriate test statistic.

  This usually tells the number of standard deviations between the sample statistic and the population statistic (z – score).

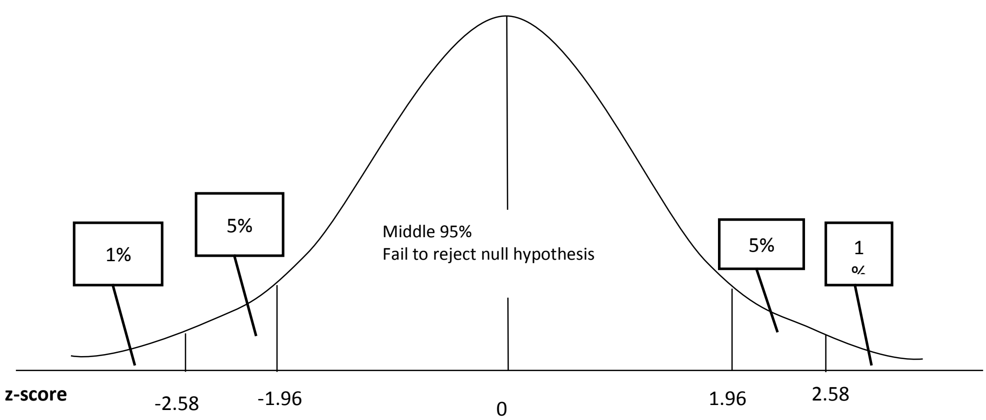
$$z = \frac{sample\ mean\ (\bar{x}) - population\ mean\ (\mu)}{standard\ error\ (\sigma_{\bar{x}})}$$

  t-value is another test statistic where population standard deviations are not known and sample size is smaller and therefore sample standard deviation cannot be used to estimate the standard error.

- Accept or reject the null hypothesis.

  The significance limit 1.96 for 5% of the significance level or 2.58 for 1% of significance level can be compared with the **z-score** for any decision for hypothesis. If the z score is within the limits, the null hypothesis is accepted and when it is beyond then the null hypothesis is rejected.

  For smaller sample size, **t value** is used to determine the range for acceptance or rejection of null hypothesis. relevant t-distribution value helps determine the critical region.



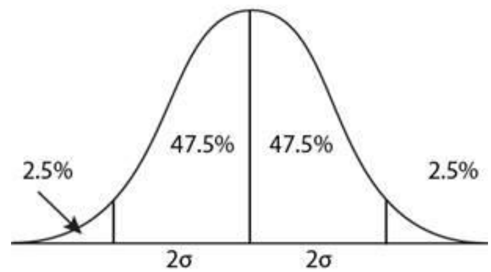
The nature of the alternate hypothesis determines how the test is carried out.

Two tail test

When the alternate hypothesis is stated to be different from the null hypothesis.

Critical region is from lower than -1.96 and greater than +1.96, beyond which is a rejection region. This means that any value of z-score falling in the critical region, the null hypothesis would be rejected.

In any normal distribution 95% of the area under the curve is within 2 standard deviations of the mean. This means 47.5% either side so that 2.5 % either side is excluded. A two tail test focuses on this 5% (2 parts of 2.5%).

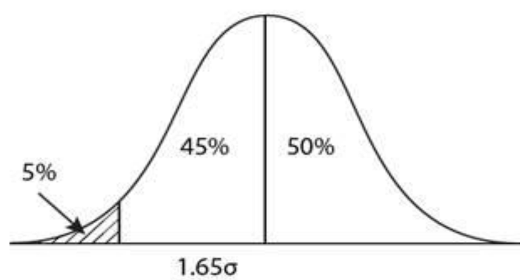


One tail test

A one tail test focuses on one half of the normal curve only. When the alternate hypothesis is stated to be smaller or larger than the null hypothesis.

Critical region is above -1.65 or below +1.65, smaller to or greater than the values respectively would mean rejection of null hypothesis.

Thus, 5% is outside and 45% within 1.65 standard deviations in one half (say the bottom half) of the distribution. This means that 97.5 % of the whole normal curve lies above this point. This is the 45% in the bottom half and the 50% in the top half.



Types of error

The tests above provide answers at a significance level. There is a chance that the tests would lead a researcher to an incorrect conclusion.

There are two types of error.

- Type I error: Rejecting the null hypothesis when it is true.
- Type II error: Accepting the null hypothesis when it is untrue. Or fails to reject the null hypothesis that is really untrue.

▶ *For example*

Type I error: The null hypothesis is that the person is innocent, while the alternative is guilty. A Type I error in this case would mean that the person is not found innocent and is sent to jail, despite actually being innocent.

Type II error: The null hypothesis is that a person has coronavirus while the alternative is that he does not have it. A Type II error in this case would mean that the person as per test result is found to have coronavirus, but in actual the person does not have it

Actual state	Accept H_0	Reject H_0
H_0 is correct	Correct decision	Type I error
H_0 is not correct	Type II error	Correct decision

AT A GLANCE
SPOTLIGHT
STICKY NOTES

2 SIGNIFICANCE TESTS OF MEANS

The overall approach is based on the assumption that the variables under examination are normally distributed. This means that most values fall close to the mean and a value would have to be quite different from the mean in terms of the shape of the curve in order to be statistically significant.

- *Illustration:*

The average distance spent by each person commuting to the office each month in a large city is asserted to be 460 km. A sample of 100 commuters showed a mean of 450 km and had a standard deviation of 25 km. Does the sample support the assertion about the population at the 5% significance level?

STEP 1:

H_0 (Null hypothesis): There is no significant difference between the sample mean and the asserted mean. Any observed difference is due to chance.

This means that H_0 = sample mean is equal to Population mean of 460 km

Type of test would be a two-tailed test since the H_1 is stated to be different from the H_0

STEP 2:

H_1 (Alternate hypothesis): There is a significant difference between the sample mean and the asserted mean. Any observed difference is due to factors other than chance and indicates that the assertion is incorrect.

In this case, H_1 = Sample mean is not equal to population mean of 460 km

STEP 3:

A significance level of 5% is already mentioned in the question which sets the Significance limits = 1.96

STEP 4:

Standard error can be calculated as:

Let sample standard deviation (s) = 25, sample size (n) = 100

$$\text{Standard error} = \sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{25}{\sqrt{100}} = 2.5$$

STEP 5:

Number of standard errors between the sample mean and the asserted mean (Z value):

$$z\ score = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{450 - 460}{2.5} = -4$$

STEP 6:

The sample mean is beyond -1.96 standard errors of the mean.

Therefore, the null hypothesis should be rejected.

The management of a company maintained that the average time taken to make a component was 50 minutes. A sample of 100 components showed that the mean time taken to make each was 52 minutes with a standard deviation of 15 minutes. Does the sample support the assertion about the population at the 95% confidence level??

STEP 1:

H_0 (Null hypothesis): There is no significant difference between the sample mean and the asserted mean. Any observed difference is due to chance.

This means that H_0 = sample mean is equal to Population mean of 50 minutes

Type of test would be a two-tailed test since the H_1 is stated to be different from the H_0

STEP 2:

H_1 (Alternate hypothesis): There is a significant difference between the sample mean and the asserted mean. Any observed difference is due to factors other than chance and indicates that the assertion is incorrect.

In this case, H_1 = Sample mean is not equal to population mean of 50 minutes

STEP 3:

A significance level of 5% is already mentioned in the question which sets the Significance limits = 1.96

STEP 4:

Standard error can be calculated as:

Let standard deviation (s) = 15, sample size (n) = 100

$$\text{Standard error} = \sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{15}{\sqrt{100}} = 1.5$$

STEP 5:

Number of standard errors between the sample mean and the asserted mean (Z value):

$$z\ score = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{52 - 50}{1.5} = 1.33$$

STEP 6:

The sample mean is within 1.96 standard errors of the mean.

There is no evidence that the null hypothesis should be rejected.

Please note that there is no proof that the null hypothesis is incorrect. There is absence of proof that it is wrong

A transport manager asserts that the mean journey time is 14 hours a day. A random sample of 64 journeys showed a mean of 13 hours and 20 minutes with a standard deviation of 3 hours. Test this assertion at the 5% significance level.

STEP 1:

H_0 (Null hypothesis): There is no significant difference between the asserted mean of 14 hours and the sample mean.

This means that H_0 = sample mean is equal to Population mean of 14 hours a day.

Type of test would be a two-tailed test since the H_1 is stated to be different from the H_0

STEP 2:

H_1 (Alternate hypothesis): There is a significant difference asserted mean of 14 hours and the sample mean.

In this case, H_1 = Sample mean is not equal to population mean of 14 hours a day.

STEP 3:

A significance level of 5% is already mentioned in the question which sets the Significance limits = 1.96

STEP 4:

Standard error can be calculated as:

Let standard deviation (s) = 3, sample size (n) = 64

$$\text{Standard error} = \sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{3}{\sqrt{64}} = \frac{3}{8} = 0.375$$

STEP 5:

Number of standard errors between the sample mean and the asserted mean (Z value):

$\bar{x}$ = 13 hours and 20 minutes (in terms of hours it is 13.33)

$$z\ score = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}\sigma} = \frac{13.3333 - 14}{0.375} = -1.78$$

STEP 6:

The sample mean is within -1.96 standard errors of the mean.

The null hypothesis should be accepted. There is no evidence that the average journey time differs from the transport manager's assertion.

A cable manufacturer makes cable with a breaking strength of 700 kg with a standard deviation of 13 kg. The manufacturer has invented a different technique that it claims results in stronger cables. A random sample of 40 cables has a mean strength of 725 kg.

STEP 1:

H_0 (Null hypothesis): There is no significant difference between the mean breaking strength of the new cables and the old cables.

This means that H_0 = sample mean is equal to Population mean of 700kg.

STEP 2:

H_1 (Alternate hypothesis): There is a significant difference between the mean breaking strength of the new cables and the old cables.

In this case, H_1 = Sample mean is greater than population mean of 700 kg.

Type of test would be a one-tailed test since the H_1 is stated to be greater than H_0

STEP 3:

A significance level of 5% is already mentioned in the question which sets the Significance limits = 1.65 for one tailed test.

STEP 4:

In the above question, standard deviation for sample is not provided, therefore standard error would be calculated using population standard deviation of 13.

STEP 5:

Number of standard errors between the sample mean and the asserted mean (Z value):

$$z\ score = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{725 - 700}{\frac{13}{\sqrt{64}}} = 15.38$$

STEP 6:

The sample mean is beyond 1.65 standard errors of the mean.

The null hypothesis should be rejected. There is evidence that the new technique produces stronger cable.

3 SIGNIFICANCE TESTS OF THE DIFFERENCE BETWEEN TWO MEANS

The means of two samples might be compared to see if they are from the same population or from two populations believed to be the same.

Two samples are constructed from two populations. The differences between the means of each pair of samples would be normally distributed (sampling distribution of differences between means) and mean of this distribution be zero.

It would not be possible to construct this distribution in practice but the theoretical notion of it is a useful tool.

- *Formula:*

For standard error of means:

$$Standard\ error = \sigma_{\bar{x}1-\bar{x}2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Where:

$\sigma_{\bar{x}1-\bar{x}2}$ = standard error of the difference between two means

s1,2 = standard deviation of samples one and two respectively

n1,2 = sample size

For z-score:

$$z = \frac{difference\ between\ sample\ means\ (\overline{x}_1 - \overline{x}_2)}{standard\ error\ (\sigma_{\bar{x}1-\bar{x}2})}$$

Where:

$\bar{x}_{1,2}$ = sample means

- *For example:*

Regular sugar tests for women and men in a city are provided in the table below:

Gender	Men	Women
Sample	1600	1900
Sugar level on average	130	125
Standard deviation of sample	17.5	20

Can we statistically verify that sugar level for women is lower than men.

STEP 1:

H_0 (Null hypothesis): There is no significant difference between the means in the two populations ($\mu_1 = \mu_2$)

This means that H_0 = Difference in sample means is equal to population means for men and women.

STEP 2:

H_1 (Alternate hypothesis): There is a significant difference between the means in the two populations ($\mu_1 < \mu_2$).

In this case, H_1 = Population mean for men is less than women.

Type of test would be a one-tailed test since the H_1 is stated to be less from the H_0

STEP 3:

A significance level of 5% is already mentioned in the question which sets the Significance limits = 1.65

STEP 4:

Standard error can be calculated as:

Let standard deviation women (s_1) = 20 standard deviation men (s_2) = 17.5; sample size women (n_1) = 1900; sample size men (n_2) = 1600

$$\sigma_{\bar{x}1-\bar{x}2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{20^2}{1900} + \frac{17.5^2}{1600}} = \sqrt{\frac{400}{1900} + \frac{306.25}{1600}} = \sqrt{0.21 + 0.19} = 0.63$$

STEP 5:

Number of standard errors between the sample mean and the asserted mean (Z value):

$$z\ score = \frac{125 - 130}{0.63} = -7.93$$

STEP 6:

The sample mean is lower than -1.65 standard errors of the mean.

Null hypothesis should be rejected. Population means for women is lower than men.

Height of men in two cities are provided in the table below:

City	Karachi	Lahore
Sample (number of men)	160	196
Height	170cm	171cm
Standard deviation of sample	8cm	7cm

STEP 1:

H_0 (Null hypothesis): There is no significant difference between the means in the two populations ($\mu_1 = \mu_2$)

This means that H_0 = Difference in sample means is equal to population means for Lahore and Karachi.

STEP 2:

H_1 (Alternate hypothesis): There is a significant difference between the means in the two populations ($\mu_1 \neq \mu_2$).

In this case, H_1 = Difference in sample means is not equal to population means for Lahore and Karachi.

Type of test would be a two-tailed test since the H_1 is stated to be different from the H_0

STEP 3:

A significance level of 5% is already mentioned in the question which sets the Significance limits = 1.96

STEP 4:

Standard error can be calculated as:

Let standard deviation Karachi (s_1) = 8, standard deviation Lahore (s_2) = 7; sample size Karachi (n_1) = 160; sample size Lahore (n_2) = 196

$$\sigma_{\bar{x}1-\bar{x}2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{8^2}{160} + \frac{7^2}{196}} = \sqrt{\frac{64}{160} + \frac{49}{196}} = \sqrt{0.4 + 0.25} = 0.81$$

STEP 5:

Number of standard errors between the sample mean and the asserted mean (Z value):

$$z\ score = \frac{170 - 171}{0.81} = -1.23$$

STEP 6:

The sample mean is within 1.96 standard errors of the mean.

There is no evidence that the null hypothesis should be rejected. Null hypothesis is accepted. There is no difference between population means for Lahore and Karachi.

However, there is no proof that the null hypothesis is correct. There is absence of proof that it is wrong.

4 SIGNIFICANCE TESTS OF PROPORTIONS

The tests for proportions follows the similar 6 step approach. Proportion of successful outcomes is often compared and used for setting up hypothesis.

- *Formula*

For standard error of means:

$$Standard\ error = \sigma_p = \sqrt{\left(\frac{pq}{n}\right)}$$

Where:

p = population proportion (the sample proportion is used as a proxy for this in cases when it is not known - usually the case)

q = 1- p

n = sample size

For z-score:

$$z = \frac{sample\ proportion\ (\hat{p}) - Population\ porportion\ (p)}{standard\ error\ (\sigma_p)}$$

- *For example:*

An election candidate claims that 60% of the voters support him. A random sample of 2,500 voters showed that 1,410 supported him (0.564 or 56.4%). Test the candidates claim 95% confidence level?

STEP 1:

H_0 (Null hypothesis): There is no significant difference between the proportion in the sample and population proportion. ($\hat{p}$=p)

This means that H_0 = sample proportion is equal to population proportion of 60%.

STEP 2:

H_1 (Alternate hypothesis): There is a significant difference between the proportion in the sample and population.

In this case, H_1 = Sample proportion is not equal to population proportion of 60% . ($\hat{p}$≠p)

Two tail – the level of support might be more or less than 60% (It is not asserted at least 60% or this would be a one tail test).

STEP 3:

A significance level of 5% is already mentioned in the question which sets the Significance limits = 1.96

STEP 4:

Standard error can be calculated as:

Let p = 0.6; q = 1-0.6 = 0.4; n = 2500

$$\sigma_p = \sqrt{\left(\frac{pq}{n}\right)} = \sqrt{\left(\frac{0.6(1-0.6)}{2{,}500}\right)} = \sqrt{\frac{0.6 \times 0.4}{2{,}500}} = 0.0098$$

STEP 5:

Number of standard errors between the sample mean and the asserted mean (Z value):

$$z\ score = \frac{0.564 - 0.6}{0.0098} = -3.7$$

STEP 6:

The sample proportion is beyond 1.96 standard errors of the mean.

The null hypothesis should be rejected.

Using the above example, now an election candidate claims that at least 60% of the voters support him.

STEP 1:

H_0 (Null hypothesis): There is no significant difference between the proportion in the sample and population proportion. ($\hat{p}$=p)

This means that H_0 = sample proportion is equal to population proportion of 60%.

STEP 2:

H_1 (Alternate hypothesis): There is a significant difference between the proportion in the sample and population.

In this case, H_1 = Sample proportion is less than population proportion of 60%. $\hat{p}$<p

One tail – the level of support might be less than 60%.

STEP 3:

A significance level of 5% is already mentioned in the question which sets the Significance limits = -1.65

STEP 4:

Standard error can be calculated as:

Let p = 0.6; q = 1-0.6 = 0.4; n = 2500

$$\sigma_p = \sqrt{\left(\frac{pq}{n}\right)} = \sqrt{\left(\frac{0.6(1-0.6)}{2{,}500}\right)} = \sqrt{\frac{0.6 \times 0.4}{2{,}500}} = 0.0098$$

STEP 5:

Number of standard errors between the sample mean and the asserted mean (Z value):

$$z\ score = \frac{0.564 - 0.6}{0.0098} = -3.7$$

STEP 6:

The sample proportion is below the critical region of -1.65. The null hypothesis would be rejected. There is evidence that there are at less than 60% voters.

5 SIGNIFICANCE TESTS INVOLVING DIFFERENCES OF PROPORTIONS

The proportions within two samples might be compared to see if they are from the same population or from two populations believed to be the same. Another sampling distribution can be used. This is the sampling distribution of proportions.

- *Formula:*

$$\text{Standard error} = \sigma_{\bar{p}1-\bar{p}2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

Where:

$p_{1,2}$ = population proportion (the sample proportion is used as a proxy for this in cases when it is not known - usually the case) of sample 1 and 2 respectively

q = 1- p

$n_{1,2}$ = sample size of sample 1 and 2 respectively

For z-score:

$$z = \frac{difference\ between\ proportions\ (\hat{p}_1 - \hat{p}_2)}{standard\ error\ (\sigma_{\bar{p}1-\bar{p}2})}$$

- *For example:*

A health official claims that citizens of Pesharwar are fitter than citizens of Gujrat. 96 out of 200 (48%) citizens in Pesharwar (selected at random) passed a standard fitness test. 84 out of 200 (42%) citizens in Gujrat (selected at random) passed the same test. Test the official's claim at the 95% confidence level.

STEP 1:

H_0 (Null hypothesis): There is no significant difference between fitness of citizens in two cities.

This means that H_0 = fitness of citizens of Pesharwar and citizens of Gujrat are equal (p_1=p_2)

STEP 2:

H_1 (Alternate hypothesis): There is a significant difference between the proportions of population between two cities.

In this case, H_1 = citizens of Peshawar are fitter than citizens of Gujrat. (p_1>p_2)

One tail – fitness of one city is claimed to be greater than the other.

STEP 3:

A significance level of 5% is already mentioned in the question which sets the Significance limits = 1.65

STEP 4:

Standard error can be calculated as:

Let p_1 = 048; q_1 = 1-0.48 = 0.52; n_1 = 200 and p_2 = 0.42; q_2 = 1-0.42 = 0.58, n_2 = 200

$$\sigma = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} = \sqrt{\frac{0.48 \times 0.52}{200} + \frac{0.42 \times 0.58}{200}} = 0.04966\ (= 0.05)$$

STEP 5:

Number of standard errors between the sample mean and the asserted mean (Z value):

$$z\ score = \frac{0.48 - 0.42}{0.05} = 1.2$$

STEP 6:

The sample proportion is below the critical region of 1.65. There is no evidence that null hypothesis would be rejected. Null hypothesis is accepted.

A sample of 300 smokers and 70 non-smokers were taken for a study of likelihood of getting cancer. Following table summarizes the findings:

	Cancer	Non-Cancer	Total
Smoker	270	30	300
Nonsmoker	8	62	70
	278	92	370

It is asserted that chances for cancer for those who smoke and do not smoke are same. Test this assertion at the 5% significance level.

STEP 1:

H_0 (Null hypothesis): There is no significant difference between population proportions.

This means that H_0 = likelihood of cancer in both the population (smoker and non-smoker are equal ($p_1=p_2$).

STEP 2:

H_1 (Alternate hypothesis): There is a significant difference between the proportions of two populations.

In this case, H_1 = likelihood for cancer in smokers is different than non-smokers. ($p_1 \neq p_2$)

Two tailed test - the likelihood of cancer might be more or less than each population (It is not asserted or this would be a one tail test).

STEP 3:

A significance level of 5% is already mentioned in the question which sets the Significance limits = 1.96

STEP 4:

Standard error can be calculated as:

Let p_1 = 270/300 =0.9; q_1 = 1-0.9 = 0.1; n_1 = 300 and p_2 = 0.12; q_2 = 1-0.12 = 0.88, n_2 = 70

$$\sigma = \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}} = \sqrt{\frac{0.9 \times 0.1}{300} + \frac{0.12 \times 0.88}{70}} = \sqrt{0.003 + 0.001} = 0.063$$

STEP 5:

Number of standard errors between the sample mean and the asserted mean (Z value):

$$z\ score = \frac{0.9 - 0.12}{0.063} = 12.38$$

STEP 6:

The sample proportion is beyond the critical region of 1.96. The null hypothesis would be rejected.

There is a difference between likelihood of cancer amongst smokers and non-smokers.

6 SIGNIFICANCE TESTS OF SMALL SAMPLES

For sample size smaller (less than 30), the **t distribution** can be used in a significance test (**t test**) to investigate whether two groups are the same.

The t-value or distribution like z-score follows normal distribution and is also helpful when population standard deviation is not known.

Degrees of freedom are determined based on the sample size which is equal to (n-1)

- *Formula:*

$$t\ value = \frac{sample\ mean\ (\overline{x}) - assumed\ population\ mean\ (\mu)}{standard\ error}$$

- *t-distribution table:*

α (1 tail)	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
α (2 tail)	**0.1**	**0.05**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
df							
1	6.3138	12.7065	31.8193	63.6551	127.3447	318.4930	636.0450
2	2.9200	4.3026	6.9646	9.9247	14.0887	22.3276	31.5989
3	2.3534	3.1824	4.5407	5.8408	7.4534	10.2145	12.9242
4	2.1319	2.7764	3.7470	4.6041	5.5976	7.1732	8.6103
5	2.0150	2.5706	3.3650	4.0322	4.7734	5.8934	6.8688

- *For example:*

An investigator suspects that the mean IQ level for a certain type of test is 110. He designs an experiment where 5 volunteers each took the test and achieve the following scores. Do the scores support his suspicion at the 95% confidence level?

	Test A
Volunteer 1	135
Volunteer 2	103
Volunteer 3	129
Volunteer 4	96
Volunteer 5	121

STEP 1:

H_0 (Null hypothesis): There is no significant difference between the population mean and sample mean.

This means that H_0 = sample mean is equal to the population mean of 110. (μ =110)

STEP 2:

H_1 (Alternate hypothesis): There is a significant difference between the population mean and sample mean.

In this case, H_1 = sample mean is not equal to population mean of 110. ($\mu \neq 110$)

Two tailed test -

AT A GLANCE SPOTLIGHT STICKY NOTES

STEP 3:

A significance level of 5% is already mentioned in the question which sets the Significance limits = 2.7764

In order to achieve the significance limit, we must know degrees of freedom:

Number of degrees of freedom (df) $= n - 1 = 5 - 1 = 4$

From the table at 0.05 two-tailed, with df = 4 we have value assigned as 2.7764

STEP 4:

In order to calculate standard error, we must first calculate standard deviation of the sample:

	Test A	$(x - \overline{x})$	$(x - \overline{x})^2$
Volunteer 1	135	18.2	331.24
Volunteer 2	103	-13.8	190.44
Volunteer 3	129	12.2	148.84
Volunteer 4	96	-20.8	432.64
Volunteer 5	121	4.2	17.64
$\sum x$	**584**		1120.8

Sample mean can be calculated as:

$$\bar{x} = \frac{\sum x}{n} = \frac{584}{5} = 116.8$$

Sample standard deviation can be calculated as:

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}} \sqrt{\frac{1120.8}{5 - 1}} = 16.74$$

For standard error or standard deviation of

Let s =16.74, n = 5

$$\sigma = \frac{s}{\sqrt{n}} = \frac{16.74}{\sqrt{5}} = 7.49$$

STEP 5:

Number of standard errors between the sample mean and the asserted mean (Z value):

$$t = \frac{sample\ mean\ (\overline{x}) - assumed\ population\ mean\ (\mu)}{standard\ error}$$

$$t = \frac{116.8 - 110}{7.49} = 0.91$$

STEP 6:

The sample proportion is within the critical region of 2.776. There is no evidence that the null hypothesis would be rejected.

There is no difference between the estimated test score and the sample test score.

An investigator suspects that there is a difference between two ways of assessing IQ. He designs an experiment where 5 volunteers each take both tests and achieve the following scores. Do the scores support his suspicion at the 95% confidence level?

	Test A	Test B
Volunteer 1	135	125
Volunteer 2	103	102
Volunteer 3	129	117
Volunteer 4	96	94
Volunteer 5	121	121

STEP 1:

H_0 (Null hypothesis): There is no significant difference between the two tests.

This means that H_0 = two IQ tests are likely to be same (the way test A measure IQ is equal to and B) (p_1=p_2).

STEP 2:

H_1 (Alternate hypothesis): There is a significant difference between the two tests.

In this case, H_1 = Test A score is less than test score for B. (p_1<p_2)

One tailed test -

STEP 3:

A significance level of 5% is already mentioned in the question which sets the Significance limits = 2.1319

In order to achieve the significance limit, we must know degrees of freedom:

$$\text{Number of degrees of freedom (df)} = n - 1 = 5 - 1 = 4$$

From the table at 0.05 One-tailed, with df = 4 we have value assigned as 2.1319

STEP 4:

In order to calculate standard error, we must first calculate standard deviation of the sample difference first we need to calculate the mean difference:

	Test A	Test B	Difference (d)	$(d - \overline{d})$	$(d - \overline{d})^2$
Volunteer 1	135	125	10	5	25
Volunteer 2	103	102	1	–4	16
Volunteer 3	129	117	12	7	49
Volunteer 4	96	94	2	–3	9
Volunteer 5	121	121	0	–5	25
∑d			25		124

$$mean\ of\ the\ difference\ between\ two\ samples\ \overline{d} = \frac{\sum d}{n} = \frac{25}{5} = 5$$

Standard deviation for the differences:

$$s = \sqrt{\frac{\sum(d - \bar{d})^2}{n - 1}} = \sqrt{\frac{124}{5 - 1}} = \sqrt{\frac{124}{4}} = 5.57$$

then standard error:

$$\sigma = \frac{s}{\sqrt{n}} = \frac{5.57}{\sqrt{5}} = 2.49$$

STEP 5:

Number of standard errors between the sample mean and the asserted mean (Z value):

$$t\ score = \frac{sample\ mean\ difference\ (\bar{d}) - assumed\ population\ mean\ difference\ (D)}{standard\ error}$$

$$t\ score = \frac{5 - 0}{2.49} = 2$$

STEP 6:

The t score is less than the t-score, there is no reason to reject the null hypothesis. hence null hypothesis will be accepted.

STICKY NOTES

Significance testing is a term given to any test devised to check if a difference between a sample characteristic and that of the population are only due to chance or not.

Steps for significance testing:

- **STEP 1: Set up the null hypothesis (H_0)**
- **STEP 2: Set up the alternate hypothesis (H_1) – This is the hypothesis that is accepted if the null hypothesis is rejected.**
- **STEP 3: Determine a significance level – This is similar to the concept of confidence level explained earlier. This time though we are picking a limit beyond which we would conclude that any difference is significant or not.**
- **STEP 4: Calculate the standard error or standard deviation of sampling distribution.**

$$\text{Standard error} = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \text{ or } \frac{s}{\sqrt{n}}$$

- **Select an appropriate test statistic.**

$$z = \frac{\textit{sample mean } (\bar{x}) - \textit{population mean } (\mu)}{\textit{standard error } (\sigma_{\bar{x}})}$$

In any normal distribution 95% of the area under the curve is within 2 standard deviations of the mean. This means 47.5% either side so that 2.5 % either side is excluded. A two tail test focuses on this 5% (2 parts of 2.5%).

A one tail test focuses on one half of the normal curve only. When the alternate hypothesis is stated to be smaller or larger than the null hypothesis.

Actual state	Accept H_0	Reject H_0
H_0 is correct	Correct decision	Type I error
H_0 is not correct	Type II error	Correct decision

The means of two samples might be compared to see if they are from the same population or from two populations believed to be the same. The differences between the means of each pair of samples would be normally distributed and mean of this distribution be zero. For standard error of means:

$$\textit{Standard error} = \sigma_{\bar{x}1-\bar{x}2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Proportion of successful outcomes is often compared and used for setting up hypothesis. For standard error of means:

$$Standard\ error = \sigma_p = \sqrt{\left(\frac{pq}{n}\right)}$$

For z-score:

$$z = \frac{sample\ proportion\ (\hat{p}) - Population\ porportion\ (p)}{standard\ error\ (\sigma_p)}$$

The proportions within two samples might be compared to see if they are from the same population or from two populations believed to be the same. Another sampling distribution can be used. This is the sampling distribution of proportions.

$$\text{Standard error} = \sigma_{\bar{p}1-\bar{p}2} = \sqrt{\frac{\mathrm{p_1 q_1}}{\mathrm{n_1}} + \frac{\mathrm{p_2 q_2}}{\mathrm{n_2}}}$$

For z-score:

$$z = \frac{difference\ between\ proportions\ (\hat{p}_1 - \hat{p}_2)}{standard\ error\ (\sigma_{\bar{p}1-\bar{p}2})}$$

For sample size smaller (less than 30), the t distribution can be used in a significance test (t test) to investigate whether two groups are the same. The t-value or distribution like z-score follows normal distribution and is also helpful when population standard deviation is not known.

$$t\ value = \frac{sample\ mean\ (\overline{x}) - assumed\ population\ mean\ (\mu)}{standard\ error}$$

SELF-TEST

14.1. The simplest form of inferential statistics, which uses known sample evidence (statistic) to draw conclusions regarding unknown population characteristics (parameter) is known as:

(a) Test of hypothesis (b) Estimation

(c) Inferential statistics (d) None of these

14.2. A numerical value assigned to the unknown population parameter is called:

(a) Statistic (b) Parameter

(c) An estimate (d) None of these

14.3. The statistic is referred to as __________ of the unknown parameter.

(a) Estimator (b) Characteristic

(c) Unbiased (d) None of these

14.4. A single numerical quantity used to estimate the population parameter is called:

(a) Point estimate (b) Interval estimate

(c) Unbiased estimate (d) None of these

14.5. If it is desirable to determine an interval within which one would expect to find the value of a parameter, such an interval is called:

(a) Point estimate (b) Interval estimate

(c) Unbiased estimate (d) None of these

14.6. The procedure of establishing a set of rules that lead to the acceptance or rejection of some kind of statements about population parameters is called:

(a) Statistical hypothesis (b) Null hypothesis

(c) Testing of hypothesis (d) None of these

14.7. The hypothesis against which we hope to gather evidence is called:

(a) Statistical hypothesis (b) Null hypothesis

(c) Composite of hypothesis (d) None of these

14.8. The hypothesis for which we wish to gather supporting evidence is called:

(a) Null hypothesis (b) Alternate hypothesis

(c) Composite hypothesis (d) None of these

AT A GLANCE
SPOTLIGHT
STICKY NOTES

14.9. Rejection of the null hypothesis when it is true is called:

(a)	Type-I error	(b)	Type-II error
(c)	Level of confidence	(d)	None of these

14.10. Acceptance of null hypothesis when it is false is called a:

(a)	Type-I error	(b)	Type-II error
(c)	Level of confidence	(d)	None of these

14.11. To prove that one teaching method is superior to another, the null hypothesis would be; that there is no difference in the two methods.

Is the stated hypothesis:

(a)	Correct	(b)	Incorrect
(c)	Does not exist	(d)	None of these

14.12. A function obtained from sample data which provides the means of testing a statistical hypothesis, is called:

(a)	Test of significance	(b)	Test-statistic
(c)	One-tailed test	(d)	None of these

14.13. The area specified for the values which are significantly different from Null hypothesis value is called:

(a)	Acceptance region	(b)	Critical region
(c)	Level of significance	(d)	None of these

14.14. Given H_o : student A and B are equal in ability.

The given hypothesis was rejected and the conclusion was drawn that student A is more capable than student B.

In the coming years it was found that the conclusion drawn was incorrect and there was no difference in their capabilities. The type of error was committed by the statistician was:

(a)	Type-I error	(b)	Type-II error
(c)	Error in drawing conclusion	(d)	None of these

14.15. If n is large i.e. n>30, the variance of population is unknown, then the test-statistic for testing mean of population must be:

(a)	t-test	(b)	z-test
(c)	χ^2 test	(d)	None of these

14.16. If the sample size is small (i.e. n<30) and variance of population is known, then to test the population mean, the test statistic to be selected is:

(a) Z-test (b) t-test

(c) χ^2 test (d) None of these

14.17. While testing the hypothesis about population mean, t-test is only selected if:

(a) n < 30 (b) σ^2 is unknown

(c) If (a) and (b) both are valid (d) None of these

14.18. While testing the hypothesis about population mean, if variance of population is known, then:

(a) Only z-test is valid (b) Only t-test is valid

(c) Either can be selected (d) None of these

14.19. While testing the hypothesis about population mean, if variance of population is unknown then the selection of test-statistic depends upon:

(a) Size of population (b) Size of sample

(c) The value of $\bar{x}$ and μ (d) None of these

14.20. While testing the hypothesis about population mean, if size of sample is large i.e. n> 30, then the selection of test-statistic depend upon:

(a) Variance of population (b) Variance of sample

(c) Does not depend upon anything (d) None of these

14.21. A tyre manufacturer claims that the tyres produced by them have a mean life of 25,000 km. Construct appropriate null and alternate hypothesis.

(a) $H_0 : \mu \geq 25,000$
$H_1 : \mu < 25,000$

(b) $H_0 : \mu \geq 25,000$
$H_1 : \mu = 25,000$

(c) $H_0 : \mu \geq 25,000$
$H_1 : \mu > 25,000$

(d) $H_0 : \mu < 25,000$
$H_1 : \mu > 25,000$

14.22. A school claims that their students have a passing ratio of 75%.
Construct appropriate null and alternate hypothesis.

(a) $H_0 : \mu \geq 75\%$
$H_1 : \mu < 75\%$

(b) $H_0 : \mu \geq 75\%$
$H_1 : \mu = 75\%$

(c) $H_0 : \mu \geq 75\%$
$H_1 : \mu > 75\%$

(d) $H_0 : \mu < 75\%$
$H_1 : \mu < 75\%$

14.23. A jeans manufacturer claims that only 2% of its production has defects. Construct appropriate null and alternate hypothesis.

(a) $H_0: \mu \geq 2\%$
$H_1: \mu < 2\%$

(b) $H_0: \mu \leq 2\%$
$H_1: \mu > 2\%$

(c) $H_0: \mu = 2\%$
$H_1: \mu \neq 2\%$

(d) $H_0: \mu \geq 2\%$
$H_1: \mu > 2\%$

14.24. Which of the following cannot be a possible value of significance level?

(a) 1% (b) 95%

(c) 5% (d) 101%

14.25. Which of the following cannot be a possible value of confidence level?

(a) 1% (b) 99%

(c) 95% (d) 105%

14.26. What is the sum of confidence level and significance level percentage?

(a) 100% (b) 105%

(c) Not a fixed value (d) Can take any value

14.27. If confidence level percentage is given as 95% what will be the value of significance level?

(a) 5% (b) 95%

(c) 5% (d) 0%

14.28. Identify the correct critical values for z test from following if significance level is 5% for a one tail test.

(a) ±1.65 (b) ±2.33

(c) ±1.96 (d) ±2.58

14.29. Identify the correct critical values for z test from following if significance level is 5% for a two tail test.

(a) ±1.65 (b) ±2.33

(c) ±1.96 (d) ±2.58

14.30. Identify the correct critical values for z test from following if significance level is 1% for a two tail test.

(a) ±1.65 (b) ±2.33

(c) ±1.96 (d) ±2.58

14.31. Identify the correct critical values for z test from following if significance level is 1% for a one tail test.

(a) ±1.65 (b) ±2.33

(c) ±1.96 (d) ±2.58

14.32. Perform hypothesis testing on following data and conclude whether to accept or reject null hypothesis:

1. Claim: $\mu = 8$
2. Population standard deviation = 0.2
3. Sample size = 49
4. Sample mean = 7.95
5. Significance level = 5%

(a) Accept (b) Reject

(c) Cannot be determined as confidence level is not given (d) Cannot be determined as sample standard deviation is not given

14.33. Perform hypothesis testing on following data and conclude whether to accept or reject null hypothesis:

1. Claim: $\mu = 8$
2. Population standard deviation = 0.2
3. Sample size = 49
4. Sample mean = 7.95
5. Significance level = 1%

(a) Accept (b) Reject

(c) Cannot be determined as confidence level is not given (d) Cannot be determined as sample standard deviation is not given

14.34. Perform hypothesis testing on following data and conclude whether to accept or reject null hypothesis:

1. Claim: $\mu \leq 8$
2. Population standard deviation = 0.2
3. Sample size = 49
4. Sample mean = 7.95
5. Significance level = 1%

(a) Accept (b) Reject

(c) Cannot be determined as confidence level is not given (d) Cannot be determined as sample standard deviation is not given

AT A GLANCE

SPOTLIGHT

STICKY NOTES

14.35. Perform hypothesis testing on following data and conclude whether to accept or reject null hypothesis:

1. Claim: $\mu \geq 8$
2. Population standard deviation = 0.2
3. Sample size = 49
4. Sample mean = 7.95
5. Significance level = 1%

(a) Accept
(b) Reject
(c) Cannot be determined as confidence level is not given
(d) Cannot be determined as sample standard deviation is not given

14.36. Perform hypothesis testing on following data and conclude whether to accept or reject null hypothesis:

1. Claim: $\mu = 11$
2. Sample standard deviation = 0.2
3. Sample size = 10
4. Sample mean = 11.4
5. Significance level = 1%

(a) Accept
(b) Reject
(c) Cannot be determined as confidence level is not given
(d) Cannot be determined as sample standard deviation is not given

14.37. If the sample size is 12, what will be the value of degrees of freedom in case of t-distribution.

(a) 11
(b) 12
(c) 13
(d) 10

14.38. What is the correct order of following steps for testing a hypothesis

1. Establish the hypothesis and its alternative.
2. Identify an appropriate sampling distribution.
3. Set the critical values.
4. Compute the test statistic.
5. Compare the test statistic with critical value and accept or reject the hypothesis

(a) 1, 2, 3, 4 and 5
(b) 1, 4, 2, 3 and 5
(c) 5, 4, 3, 2 and 1
(d) 5, 3, 2, 1 and 4

14.39. The hypothesis which we want to test is called the ____ hypothesis and is denoted by H_0

(a) Null
(b) Alternate
(c) Significant
(d) Confidence

14.40. In which of the following case is the level of significance divided into two tails equally.

(a) A manufacturer claims that his products last for more than 24 hours.

(b) A milk bottling plant is expected to fill 250ml of liquid in one bottle.

(c) A manufacturer claims that defects in its production are less than 5%

(d) A school claims that their passing ratio is 75% or more

ANSWERS TO SELF-TEST QUESTIONS

14.1	**14.2**	**14.3**	**14.4**	**14.5**	**14.6**
(b)	(c)	(a)	(a)	(b)	(c)
14.7	**14.8**	**14.9**	**14.10**	**14.11**	**14.12**
(a)	(b)	(a)	(b)	(a)	(b)
14.13	**14.14**	**14.15**	**14.16**	**14.17**	**14.18**
(b)	(a)	(a)	(a)	(c)	(a)
14.19	**14.20**	**14.21**	**14.22**	**14.23**	**14.24**
(b)	(c)	(a)	(a)	(b)	(d)
14.25	**14.26**	**14.27**	**14.28**	**14.29**	**14.30**
(d)	(a)	(c)	(a)	(c)	(d)
14.31	**14.32**	**14.33**	**14.34**	**14.35**	**14.36**
(b)	(a)	(a)	(a)	(a)	(b)
14.37	**14.38**	**14.39**	**14.40**		
(a)	(a)	(a)	(b)		

CHAPTER 15

CHI-SQUARE TESTING

IN THIS CHAPTER

AT A GLANCE

The chi-square test ($\chi 2$) is used to determine whether there is a significant difference between observed frequencies of results and the expected frequencies of results.

It is a test statistic for measuring independence. It asks whether the number of items in each category differs significantly from the number expected and whether any difference between the expected and observed is due to chance or is it a real difference.

The chapter also discusses test of association and goodness of fit

1. CHI-SQUARE TESTING

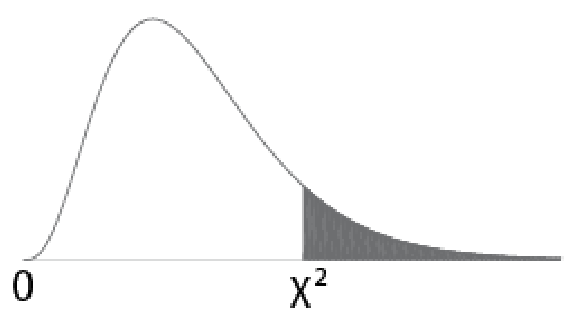
The Chi-Square test compares observed frequencies in data compared to those expected if no factor other than chance was operating (the expected frequencies).

Chi-square (χ^2) is a one tailed distribution since it squares the negative values as well.

- *Formula:*

$$\chi^2 = \sum \frac{(observed\ frequency - expected\ frequency)^2}{expected\ frequency}$$

- *Illustration:*



A cut off value for chi-square (χ^2) is calculated for a given confidence level. This value is a number of units from the source (0).

The null hypothesis is rejected if the chi-square (χ^2) value calculated for the data is greater than the cut off value (i.e. the calculated value falls in the shaded area).

- *Illustration:*

	Cut off point:	
(df)	**0.05**	**0.01**
1	3.841	6.635
2	5.991	9.210
3	7.815	11.345
4	9.488	13.277
5	11.070	15.086

The value of 0.05 refers to the area of the curve beyond the critical point. In other words, 95% of values are within this value.

Degrees of freedom

The cut off value at a given level of confidence depends on the number of degrees of freedom associated with the data.

If the data is a simple array this is usually found as n – 1 where n is the number of observations.

This is not the case if data is provided (or has to be arranged) into a contingency table.

AT A GLANCE
SPOTLIGHT
STICKY NOTES

- *Formula:*

For an array of data: $n - 1$

Where:

n = number of observations

Types of test

There are two types of chi-square test:

- The goodness of fit test: a test for comparing a theoretical distribution with the observed data from a sample; and
- The test of association: a test which allows the comparison of two attributes in a sample of data to determine if there is any relationship between them.

In each case the test compares an observed series of data to an expected one. A chi-square (χ^2) value is calculated for comparison to a critical value taken from tables.

Steps for significance testing:

Chi-square testing closely resembles the significance testing elaborated in the earlier chapter.

- STEP 1: Set up the null hypothesis (H_0) - There is no significant difference between the observed frequencies and the expected frequencies.
- STEP 2: Set up the alternate hypothesis (H_1) – There is a significant difference between the observed frequencies and the expected frequencies.
- STEP 3: Determine a significance level – Usually 95% or cut-off at 0.05 or 99% or cut off at 0.01 are being used. Values are already determined in a table of Chi-Square.
- STEP 4: Compute χ^2 for the data. Formula given above.
- STEP 5: Estimate the number of degrees of freedom.
- STEP 6: Use the table to find the cut-off point at the required confidence level for the estimated number of degrees of freedom.
- STEP 7: Accept or reject the null hypothesis by comparing the cut-off point with the calculated χ^2 value.

2. GOODNESS OF FIT

This is a test that compares frequencies of observed data from a sample against what those frequencies were expected to be. It is used where data is categorical.

- *For example:*

  A company prepares ball points from Machine A. It is claimed that 10% of the pens are defected, 25% are partially defected and 60% are perfect.

  Quality inspector collected a random sample of ball points prepared by machine A. He can use a chi-square goodness of fit test to see whether the sample distribution is different significantly from what is claimed by the company.

- *Illustration:*

A car retailer sells a model of car with a choice of 5 colours. He expects demand this year to follow demand last year in terms of choice of colour? The first 150 cars sold this year do not seem to have followed this pattern and he is worried that tastes have changed. Perform a χ^2 at the 5% level of significance to determine whether the customers' tastes as regard to the same as the last year.

	Actual sales (O)	Expected sales (based on last year's pattern)
Yellow	35	30
Red	50	45
Green	30	15
Blue	10	15
White	25	45

STEP 1:

H_0 (Null hypothesis): There is no evidence of change in taste from last year.

STEP 2:

H_1 (Alternate hypothesis): Taste has changed.

STEP 3:

A significance level of 5% is already mentioned in the question

STEP 4:

Table to calculate $\chi 2$:

Observed (O)	Expected (E)	O – E	(O – E)^2	$\frac{(O-E)^2}{E}$
35	30	5	25	0.83
50	45	5	25	0.56
30	15	15	225	15.00
10	15	–5	25	1.67
25	45	–20	400	8.89
150	150			26.95

STEP 5:

Df = n-1 = 5-1 = 4

STEP 6: As per the table 95% cut off value for 4 degrees of freedom is 9.488

STEP 7: The null hypothesis should be rejected since the chi-square value calculated (26.95) is more than this.

There is a significant difference between customer taste this year compared to last year.

3. TESTS OF ASSOCIATION (INDEPENDENCE)

A test of association or independence compares the occurrence of two attributes in a sample of data to determine if there is any relationship between them.

Data to be tested is usually presented in a contingency table. In this context, a contingency table is one that sets out data subject to two descriptions. In a question involving contingency tables you may need to compute expected values based on the actual data.

- *For example:*

  Following table provides grades and time of study data (extracted from 75 students).

	Grades				
		High	Average	Low	Total
Time of study	Long	3	14	5	**21**
	Short	10	30	14	**54**
		13	**44**	**18**	**75**
	Column percentage (long)	23%	32%	28%	

  If a researcher is interested in identifying association between high school grades and time of study, the chi-square test can be helpful.

  From the table, it can be observed that Average graders study for longer hours. Does this mean that there is an association between grades and time of study? We may need to study further to reach to surer conclusions. However, it is to say that, the test would not be helpful in determining a value if one of the observation is known.

- *Illustration:*

  Using the Example above, perform Chi-Square test of association (or independence) at 5% confidence interval to find any relationship between two variables.

  STEP 1:

  H_0 (Null hypothesis): There is no association between grades and time of study.

  STEP 2:

  H_1 (Alternate hypothesis): There is an association between grades and time of study.

  STEP 3:

  A significance level of 5% is already mentioned in the question

  STEP 4:

  The expected values can be calculated as:

	High	Average	Low	Total
Long	17% × 21 = 3.57	59% × 21 = 12.39	24% × 21 = 5.04	**21**
Short	17% × 54 = 9.18	59% × 54 = 31.86	24% × 54 = 12.96	**54**
	12	52	36	75

AT A GLANCE
SPOTLIGHT
STICKY NOTES

Table to calculate χ^2:

Observed (O)	Expected (E)	O – E	$(O – E)^2$	$\frac{(O-E)^2}{E}$
3	3.57	-0.57	0.3249	0.091
10	9.18	0.82	0.6724	0.0732
14	12.39	1.61	2.5921	0.209
30	31.86	-1.86	3.4596	0.108
5	5.04	-0.04	0.0016	0.00032
14	12.96	1.04	1.0816	0.083
75	75			0.56452

STEP 5:

Df = (r-1)(c-1) = (2-1)(3-1) = (1)(2)= 2

STEP 6:

95% cut off value for 2 degrees of freedom is 5.991

STEP 7:

There is no evidence that the null hypothesis should be rejected since the chi-square value calculated is less than the cut off value.

There is no evidence to link grades with time of study.

A manufacturing company produces items at two different factories. A survey is carried out to determine if there is an association between the quality of items and the factory that they are made in. A sample of 100 items gave the following results:

	Unsatisfactory quality	*Satisfactory quality*	*Good quality*	*Total*
Factory A	***8***	***25***	***21***	***54***
Factory B	***4***	***27***	***15***	***46***
	12	***52***	***36***	***100***

The bottom line in this table also shows that 12% of items were unsatisfactory, 52% were satisfactory and 36% were good. Perform a χ^2 at the 1% level of significance to determine whether there is an association between the quality of items and the factory that produced them.

STEP 1:

H_0 (Null hypothesis): There is no association between quality and the factory.

STEP 2:

H_1 (Alternate hypothesis): There is an association between quality and the factory.

STEP 3:

A significance level of 1% is already mentioned in the question

STEP 4:

If there is no association between quality of an item and the factory that made the item. Each factory would expect to have achieved these percentages across their production

The expected values can be calculated as:

	Unsatisfactory quality	Satisfactory quality	Good quality	Total
Factory A	12% × 54 = 6.48	52% × 54 = 28.08	36% × 54 = 19.44	54
Factory B	12% × 46 = 5.52	52% × 46 = 23.92	36% × 46 = 16.56	46
	12	52	36	100

Table to calculate χ^2:

Observed (O)	Expected (E)	O – E	$(O – E)^2$	$\frac{(O-E)^2}{E}$
8	6.48	1.52	2.3104	0.357
4	5.52	-1.52	2.3104	0.419
25	28.08	-3.08	9.4864	0.338
27	23.92	3.08	9.4864	0.397
21	19.44	1.56	2.4336	0.125
15	16.56	-1.56	2.4336	0.147
100	100			1.783

STEP 5:

Df = (r-1)(c-1) = (2-1)(3-1) = (1)(2)= 2

STEP 6:

99% cut off value for 2 degrees of freedom is 9.210

STEP 7:

There is no evidence that the null hypothesis should be rejected since the chi-square value calculated is less than the cut off value.

There is no evidence to link quality of items to the factory that they are made in.

2 × 2 Contingency Table:

There may be further complication, if data is presented in two rows and two columns or contingency table.

In this case there is an adjustment made during the calculation of the chi-square value of the data. specifically, for 2 × 2 table, any numerical value for O - E must be reduced by 0.5 regardless of sign.

In such cases formula for Chi-square with Yates' continuity correction is used.

- *Formula:*

$$\chi^2 = \sum \frac{(|\text{Actual} - \text{Expected value}| - 0.5)^2}{\text{Expected value}} = \sum \frac{(|O - E| - 0.5)^2}{E}$$

Where:

| | = Modulus (means ignore the sign)

The correction involves reducing all values of O -E by 0.5 regardless of the sign of O - E.

|10| - 0.5 = 9.5

|-10| - 0.5 = 9.5

Degrees of freedom: $(r-1)(c-1)$

Where:

r = number of rows

c = number of columns

- *Illustration:*

A survey of 50 students at high school showed that they had made the following choices to study a second European language. Carry out a χ^2 test at the 5% level of significance to determine whether there is an association between gender and language chosen:

	French	Italian	Total
Male	14	6	20
Female	8	22	30
	22	28	50
	44%	56%	

STEP 1:

H_0 (Null hypothesis): There is no association between gender and language chosen. Or language preference in both the genders are same.

STEP 2:

H_1 (Alternate hypothesis): There is an association between gender and language chosen.

STEP 3:

A significance level of 5% is already mentioned in the question

STEP 4:

The expected values can be calculated as:

	French	Italian	Total
Male	44% × 20 = 8.8	56% × 20 = 11.2	20
Female	44% × 30= 13.2	56% × 30 = 16.8	30
	22	28	50

Table to calculate χ^2:

Observed (O)	Expected (E)	$\lvert O - E \rvert - 0.5$	$\frac{(\lvert O-E\rvert - 0.5)^2}{E}$
14	8.8	|14-8.8| - 0.5 = 4.7	$\frac{(4.7)^2}{8.8} = 2.51$
8	13.2	|8-13.2| - 0.5 = 4.7	$\frac{(4.7)^2}{13.2} = 1.673$
6	11.2	|6-11.2| - 0.5 = 4.7	$\frac{(4.7)^2}{11.2} = 1.972$
22	16.8	|22-16.8| - 0.5 = 4.7	$\frac{(4.7)^2}{16.8} = 1.315$
100	50		7.47

STEP 5:

Df = (r-1)(c-1) = (2-1)(2-1) = (1)(1)= 1

STEP 6:

95% cut off value for 1 degree of freedom is 3.841

STEP 7:

The null hypothesis should be rejected since the chi-square value calculated is more than the cut off value.

There is evidence of association between gender and the language chosen.

STICKY NOTES

The Chi-Square test compares observed frequencies in data compared to those expected if no factor other than chance was operating (the expected frequencies).
Chi-square (χ^2) is a one tailed distribution since it squares the negative values as well.

$$\chi^2 = \sum \frac{(observed\ frequency - expected\ frequency)^2}{expected\ frequency}$$

Chi-square testing closely resembles the significance testing elaborated in the earlier chapter.

- **STEP 1: Set up the null hypothesis (H_0) - There is no significant difference between the observed frequencies and the expected frequencies.**
- **STEP 2: Set up the alternate hypothesis (H_1) – There is a significant difference between the observed frequencies and the expected frequencies.**
- **STEP 3: Determine a significance level – Usually 95% or cut-off at 0.05 or 99% or cut off at 0.01 are being used. Values are already determined in a table of Chi-Square.**
- **STEP 4: Compute χ^2 for the data. Formula given above.**
- **STEP 5: Estimate the number of degrees of freedom.**
- **STEP 6: Use the table to find the cut-off point at the required confidence level for the estimated number of degrees of freedom.**
- **STEP 7: Accept or reject the null hypothesis by comparing the cut-off point with the calculated $\chi 2$ value.**

Goodness of fit is a test that compares frequencies of observed data from a sample against what those frequencies were expected to be. It is used where data is categorical.

A test of association or independence compares the occurrence of two attributes in a sample of data to determine if there is any relationship between them. Data to be tested is usually presented in a contingency table. In this context, a contingency table is one that sets out data subject to two descriptions.

AT A GLANCE
SPOTLIGHT
STICKY NOTES

SELF-TEST

15.1. Chi-square values ranges from:

(a) $-\infty$ to $+\infty$ (b) -1 to +1

(c) 0 to ∞ (d) None of these

15.2. If we want to test whether or not two samples are independent, the test-statistic to be selected is:

(a) t-test (b) z-test

(c) Chi-square test (d) None of these

15.3. It is assumed (H_0) that the political affiliation of a person does not depend upon the level of education. The hypothesis was tested using:

(a) Normal distribution (b) Chi square distribution

(c) Binomial distribution (d) None of these

15.4. A CA Foundation course instructor believes that the grades of students in Foundation examination depend upon the college. The conjuncture will be tested by using:

(a) Chi-square distribution (b) t-distribution

(c) Normal distribution (d) None of these

15.5. A table containing the data classified according to characteristics of population (attributes) is known as:

(a) Frequency table (b) Contingency table

(c) Correlation table (d) None of these

15.6. A technique by means of which we test the hypothesis whether the sample distribution is in agreement with the theoretical distribution is called:

(a) Parametric test (b) Non-parametric test

(c) Goodness of-fit-test (d) None of these

15.7. A goodness of-fit-test can be applied by using:

(a) Normal distribution (b) t-distribution

(c) Chi-square distribution (d) None of these

15.8. For a test of association at 95% confidence interval, calculated chi-square value is 8.431 whereas the table value is of 9.488. This means that

(a) null hypothesis is rejected (b) null hypothesis is accepted

(c) more information is needed (d) none of the above

15.9. 2x2 contingency table means:

(a) 2 values in each rows and column (b) 2 cells in each row

(c) 2 rows and 2 columns (d) 2 cells in reach column

15.10. Chi-square is ___________

(a) a two-tailed distribution (b) not a tailed distribution

(c) a one-tailed distribution (d) none of the above

15.11. For the following set of data, compute the chi-square value:

Observed frequency	Expected frequency
11	10
11	11
5	4
8	8

(a) 0.35 (b) 0.45

(c) 0.55 (d) 0.30

15.12. Chi-square distribution is a continuous probability distribution having range from ___ to plus infinity.

(a) zero (b) one

(c) negative infinity (d) negative one

15.13. A random variable with a Chi-square distribution is always ___.

(a) negative (b) non-negative

(c) positive (d) Infinity

15.14. Total area under the Chi-square curve is ___.

(a) A. one (b) B. zero

(c) C. infinity (d) D. large

15.15. The Chi-square distribution is single peaked and has ___ skewness.

(a) negative (b) positive

(c) small (d) large

15.16. In goodness-of-fit test, the rejection region lies in the ___ tail of the distribution.

(a) right (b) left

(c) right and left (d) Depends upon data

15.17. A ____ test is a technique by which we test whether the observed frequencies in a sampling distribution are in agreement with the theoretical distribution.

(a) test of association (b) hyper geometric
(c) goodness-of-fit (d) chi-square curve

15.18. What will be the table value of chi-square for:
95% confidence level and 4 degrees of freedom

(a) 0.711 (b) 9.488
(c) 11.070 (d) 13.277

15.19. What will be the table value of chi-square for:
99% confidence level and 4 degrees of freedom

(a) 0.711 (b) 9.488
(c) 11.070 (d) 13.277

15.20. What will be the table value of chi-square for:
95% confidence level and 5 degrees of freedom

(a) 0.711 (b) 9.488
(c) 11.070 (d) 13.277

15.21. What will be the table value of chi-square for:
5% level of significance and 4 degrees of freedom

(a) 0.711 (b) 9.488
(c) 11.070 (d) 13.277

15.22. What will be the table value of chi-square for:
5% level of significance and 5 degrees of freedom

(a) 0.711 (b) 9.488
(c) 11.070 (d) 13.277

15.23. Which of the following cannot be a possible table value of chi-square:

(a) -0.1 (b) 0
(c) 53 (d) 54

15.24. What will be the value of degrees of freedom for a contingency table with five rows and four columns?

(a) 13 (b) 12
(c) 8 (d) 20

15.25. What will be the value of degrees of freedom for a contingency table with three rows and two columns?

(a) 5
(b) 6
(c) 2
(d) 1

15.26. What adjustment is made in the calculated value of chi-square in case of a 2 x 2 contingency table?

(a) Any numerical value for observed minus expected value must be reduced by 0.5 regardless of sign
(b) Any numerical value for observed minus expected value must be increased by 0.5 regardless of sign
(c) Any numerical value for observed vale must be reduced by 0.5 regardless of sign
(d) Nothing needs to be done

15.27. A ___ compares the occurrence of two attributes in a sample of data to determine if there is any relationship between them.

(a) test of association
(b) goodness-of-fit
(c) chi square curve
(d) frequency polygon

15.28. The null hypothesis is rejected if the chi-square value ___ for the data is greater than the cut off value.

(a) calculated
(b) observed
(c) expected
(d) judged

15.29. A dice was tossed 144 times and following outcomes were recorded:

Faces	1	2	3	4	5	6
Observed occurrences	22	23	27	25	26	21

Using chi-square test at 5% level of significance, assess the hypothesis that the dice is fair.

(a) Dice is fair
(b) Dice is not fair
(c) Insufficient data
(d) Impossible to compute even if more data is available

15.30. A group of people was surveyed about their favourite car. The following results were obtained:

Gender	Frequency		
	Civic	**Corolla**	**Liana**
Male	27	37	11
Female	26	14	5

At 5% level of significance, test the hypothesis that the choice of favourite car is independent of one's gender

(a) Favourite car is independent of one's gender
(b) Favourite car is dependent of one's gender
(c) Insufficient data
(d) Impossible to compute even if more data is available

ANSWERS TO SELF-TEST QUESTIONS

15.1	**15.2**	**15.3**	**15.4**	**15.5**	**15.6**
(b)	(c)	(b)	(a)	(b)	(c)
15.7	**15.8**	**15.9**	**15.10**	**15.11**	**15.12**
(c)	(b)	(c)	(c)	(a)	(a)
15.13	**15.14**	**15.15**	**15.16**	**15.17**	**15.18**
(b)	(a)	(b)	(a)	(c)	(b)
15.19	**15.20**	**15.21**	**15.22**	**15.23**	**15.24**
(d)	(c)	(b)	(c)	(a)	(b)
15.25	**15.26**	**15.27**	**15.28**	**15.29**	**15.30**
(c)	(a)	(a)	(a)	(a)	(a)

APPENDIX

IN THIS CHAPTER

AT A GLANCE

SPOTLIGHT

STICKY NOTES

AT A GLANCE
SPOTLIGHT
STICKY NOTES

PRESENT VALUE TABLE

This table shows the discount factor for an amount at the end of *n* periods at r%.

	Interest rates (r)									
Periods										
(n)	**1%**	**2%**	**3%**	**4%**	**5%**	**6%**	**7%**	**8%**	**9%**	**10%**
1	.990	.980	.971	.962	.962	.943	.935	.926	.917	.909
2	.980	.961	.943	.925	.907	.890	.873	.857	.842	.826
3	.971	.942	.915	.889	.864	.840	.816	.794	.772	.751
4	.961	.924	.888	.855	.823	.792	.763	.735	.708	.683
5	.951	.906	.863	.822	.784	.747	.713	.681	.650	.621
6	.942	.888	.837	.790	.746	.705	.666	.630	.596	.564
7	.933	.871	.813	.760	.711	.665	.623	.583	.547	.513
8	.923	.853	.789	.731	.677	.627	.582	.540	.502	.467
9	.914	.837	.766	.703	.645	.592	.544	.500	.460	.424
10	.905	.820	.744	.676	.614	.558	.508	.463	.422	.386
11	.896	.804	.722	.650	.585	.527	.475	.429	.388	.350
12	.887	.788	.701	.625	.557	.497	.444	.397	.356	.319
13	.879	.773	.681	.601	.530	.469	.415	.368	.326	.290
14	.870	.758	.661	.577	.505	.442	.388	.340	.299	.263
15	.861	.743	.642	.555	.481	.417	.362	.315	.275	.239
16	.853	.728	.623	.534	.458	.394	.339	.292	.252	.218
17	.844	.714	.605	.513	.436	.371	.317	.270	.231	.198
18	.836	.700	.587	.494	.416	.350	.296	.250	.212	.180
19	.828	.686	.570	.475	.396	.331	.277	.232	.194	.164
20	.820	.673	.554	.456	.377	.312	.258	.215	.178	.149

	Interest rates (r)									
Periods										
(n)	**11%**	**12%**	**13%**	**14%**	**15%**	**16%**	**17%**	**18%**	**19%**	**20%**
1	.901	.893	.885	.877	.870	.862	.855	.847	.840	.833
2	.812	.797	.783	.769	.756	.743	.731	.718	.706	.694
3	.731	.712	.693	.675	.658	.641	.624	.609	.593	.579
4	.659	.636	.613	.592	.572	.552	.534	.516	.499	.482
5	.593	.567	.543	.519	.497	.476	.456	.437	.419	.402
6	.535	.507	.480	.456	.432	.410	.390	.370	.352	.335
7	.482	.452	.425	.400	.376	.354	.333	.314	.296	.279
8	.434	.404	.376	.351	.327	.305	.285	.266	.249	.233
9	.391	.361	.333	.308	.284	.263	.243	.225	.209	.194
10	.352	.322	.295	.270	.247	.227	.208	.191	.176	.162
11	.317	.287	.261	.237	.215	.195	.178	.162	.148	.135
12	.286	.257	.231	.208	.187	.168	.152	.137	.124	.112
13	.258	.229	.204	.182	.163	.145	.130	.116	.104	.093
14	.232	.205	.181	.160	.141	.125	.111	.099	.088	.078
15	.209	.183	.160	.140	.123	.108	.095	.084	.074	.065
16	.188	.163	.141	.123	.107	.093	.081	.071	.062	.054
17	.170	.146	.125	.108	.093	.080	.069	.060	.052	.045
18	.153	.130	.111	.095	.081	.069	.059	.051	.044	.038
19	.138	.116	.098	.083	.070	.060	.051	.043	.037	.031
20	.124	.104	.087	.073	.061	.051	.043	.037	.031	.026

CUMULATIVE PRESENT VALUE

This table shows the annuity factor for an amount at the end of each year for *n* years at r%.

	Interest rates (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
2	1.970	1.942	1.913	1.886	1.859	1.833	1.808	1.783	1.759	1.736
3	2.941	2.884	2.829	2.775	2.723	2.673	2.624	2.577	2.531	2.487
4	3.902	3.808	3.717	3.630	3.546	3.465	3.387	3.312	3.240	3.170
5	4.853	4.713	4.580	4.452	4.329	4.212	4.100	3.993	3.890	3.791
6	5.795	5.601	5.417	5.242	5.076	4.917	4.767	4.623	4.486	4.355
7	6.728	6.472	6.230	6.002	5.786	5.582	5.389	5.206	5.033	4.868
8	7.652	7.325	7.020	6.733	6.463	6.210	5.971	5.747	5.535	5.335
9	8.566	8.162	7.786	7.435	7.108	6.802	6.515	6.247	5.995	5.759
10	9.471	8.983	8.530	8.111	7.722	7.360	7.024	6.710	6.418	6.145
11	10.368	9.787	9.253	8.760	8.306	7.887	7.499	7.139	6.805	8.495
12	11.255	10.575	9.954	9.385	8.863	8.384	7.943	7.536	7.161	6.814
13	12.134	11.348	10.635	9.986	9.394	8.853	8.358	7.904	7.487	7.103
14	13.004	12.106	11.296	10.563	9.899	9.295	8.745	8.244	7.786	7.367
15	13.865	12.849	11.938	11.118	10.380	9.712	9.108	8.559	8.061	7.606
16	14.718	13.578	12.561	11.652	10.838	10.106	9.447	8.851	8.313	7.824
17	15.562	14.292	13.166	12.166	11.274	10.477	9.763	9.122	8.544	8.022
18	16.398	14.992	13.754	12.659	11.690	10.828	10.059	9.372	8.756	8.201
19	17.226	15.679	14.324	13.134	12.085	11.158	10.336	9.604	8.950	8.365
20	18.046	16.351	14.878	13.590	12.462	11.470	10.594	9.818	9.129	8.514

	Interest rates (r)									
Periods										
(n)	**11%**	**12%**	**13%**	**14%**	**15%**	**16%**	**17%**	**18%**	**19%**	**20%**
1	0.901	0.893	0.885	0.877	0.870	0.862	0685	0.847	0.840	0.833
2	1.713	1.690	1.668	1.647	1.626	1.605	1.585	1.566	1.547	1.528
3	2.444	2.402	2.361	2.322	2.283	2.246	2.210	2.174	2.140	2.106
4	3.102	3.037	2.974	2.914	2.855	2.798	2.743	2.690	2.639	2.589
5	3.696	3.605	3.517	3.433	3.352	3.274	3.199	3.127	3.058	2.991
6	4.231	4.111	3.998	3.889	3.784	3.685	3.589	3.498	3.410	3.326
7	4.712	4.564	4.423	4.288	4.160	4.039	3.922	3.812	3.706	3.605
8	5.146	4.968	4.799	4.639	4.487	4.344	4.207	4.078	3.954	3.837
9	5.537	5.328	5.132	4.946	4.772	4.607	4.451	4.303	4.163	4.031
10	5.889	5.650	5.426	5.216	5.019	4.833	4.659	4.494	4.339	4.192
11	6.207	5.938	5.687	5.453	5.234	5.029	4.836	4.656	4.486	4.327
12	6.492	6.194	5.918	5.660	5.421	5.197	4.968	4.793	4.611	4.439
13	6.750	6.424	6.122	5.842	5.583	5.342	5.118	4.910	4.715	4.533
14	6.982	6.628	6.302	6.002	5.724	5.468	5.229	5.008	4.802	4.611
15	7.191	6.811	6.462	6.142	5.847	5.575	5.324	5.092	4.876	4.675
16	7.379	6.974	6.604	6.265	5.954	5.668	5.405	5.162	4.938	4.730
17	7.549	7.120	6.729	6.373	6.047	5.749	5.475	5.222	4.990	4.775
18	7.702	7.250	6.840	6.467	6.128	5.818	5.534	5.273	5.033	4.812
19	7.839	7.366	6.938	6.550	6.198	5.877	5.584	5.316	5.070	4.843
20	7.963	7.469	7.025	6.623	6.259	5.929	5.628	5.353	5.101	4.870

POISSON DISTRIBUTION

$$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

For a given value of λ an entry indicates the probability of a specific value of x.

	λ									
x	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	0.3679	0.1353	0.0498	0.0183	0.0067	0.0025	0.0009	0.0003	0.0001	0.0000
1	0.3679	0.2707	0.1494	0.0733	0.0337	0.0149	0.0064	0.0027	0.0011	0.0005
2	0.1839	0.2707	0.2240	0.1465	0.0842	0.0446	0.0223	0.0107	0.0050	0.0023
3	0.0613	0.1804	0.2240	0.1954	0.1404	0.0892	0.0521	0.0286	0.0150	0.0076
4	0.0153	0.0902	0.1680	0.1954	0.1755	0.1339	0.0912	0.0573	0.0337	0.0189
5	0.0031	0.0361	0.1008	0.1563	0.1755	0.1606	0.1277	0.0916	0.0607	0.0378
6	0.0005	0.0120	0.0504	0.1042	0.1462	0.1606	0.1490	0.1221	0.0911	0.0631
7	0.0001	0.0034	0.0216	0.0595	0.1044	0.1377	0.1490	0.1396	0.1171	0.0901
8	0.0000	0.0009	0.0081	0.0298	0.0653	0.1033	0.1304	0.1396	0.1318	0.1126
9		0.0002	0.0027	0.0132	0.0363	0.0688	0.1014	0.1241	0.1318	0.1251
10		0.0000	0.0008	0.0053	0.0181	0.0413	0.0710	0.0993	0.1186	0.1251
11			0.0002	0.0019	0.0082	0.0225	0.0452	0.0722	0.0970	0.1137
12			0.0001	0.0006	0.0034	0.0113	0.0263	0.0481	0.0728	0.0948
13			0.0000	0.0002	0.0013	0.0052	0.0142	0.0296	0.0504	0.0729
14				0.0001	0.0005	0.0022	0.0071	0.0169	0.0324	0.0521
15				0.0000	0.0002	0.0009	0.0033	0.0090	0.0194	0.0347
16					0.0000	0.0003	0.0014	0.0045	0.0109	0.0217
17						0.0001	0.0006	0.0021	0.0058	0.0128
18						0.0000	0.0002	0.0009	0.0029	0.0071
19							0.0001	0.0004	0.0014	0.0037
20							0.0000	0.0002	0.0006	0.0019
21								0.0001	0.0003	0.0009
22								0.0000	0.0001	0.0004
23									0.0000	0.0002
24										0.0001
25										0.0000

	λ									
x	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0010	0.0004	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0037	0.0018	0.0008	0.0004	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000
4	0.0102	0.0053	0.0027	0.0013	0.0006	0.0003	0.0001	0.0001	0.0000	0.0000
5	0.0224	0.0127	0.0070	0.0037	0.0019	0.0010	0.0005	0.0002	0.0001	0.0001
6	0.0411	0.0255	0.0152	0.0087	0.0048	0.0026	0.0014	0.0007	0.0004	0.0002
7	0.0646	0.0437	0.0281	0.0174	0.0104	0.0060	0.0034	0.0019	0.0010	0.0005
8	0.0888	0.0655	0.0457	0.0304	0.0194	0.0120	0.0072	0.0042	0.0024	0.0013
9	0.1085	0.0874	0.0661	0.0473	0.0324	0.0213	0.0135	0.0083	0.0050	0.0029
10	0.1194	0.1048	0.0859	0.0663	0.0486	0.0341	0.0230	0.0150	0.0095	0.0058
11	0.1194	0.1144	0.1015	0.0844	0.0663	0.0496	0.0355	0.0245	0.0164	0.0106
12	0.1094	0.1144	0.1099	0.0984	0.0829	0.0661	0.0504	0.0368	0.0259	0.0176
13	0.0926	0.1056	0.1099	0.1060	0.0956	0.0814	0.0658	0.0509	0.0378	0.0271
14	0.0728	0.0905	0.1021	0.1060	0.1024	0.0930	0.0800	0.0655	0.0514	0.0387
15	0.0534	0.0724	0.0885	0.0989	0.1024	0.0992	0.0906	0.0786	0.0650	0.0516
16	0.0367	0.0543	0.0719	0.0866	0.0960	0.0992	0.0963	0.0884	0.0772	0.0646
17	0.0237	0.0383	0.0550	0.0713	0.0847	0.0934	0.0963	0.0936	0.0863	0.0760
18	0.0145	0.0255	0.0397	0.0554	0.0706	0.0830	0.0909	0.0936	0.0911	0.0844
19	0.0084	0.0161	0.0272	0.0409	0.0557	0.0699	0.0814	0.0887	0.0911	0.0888
20	0.0046	0.0097	0.0177	0.0286	0.0418	0.0559	0.0692	0.0798	0.0866	0.0888
21	0.0024	0.0055	0.0109	0.0191	0.0299	0.0426	0.0560	0.0684	0.0783	0.0846
22	0.0012	0.0030	0.0065	0.0121	0.0204	0.0310	0.0433	0.0560	0.0676	0.0769
23	0.0006	0.0016	0.0037	0.0074	0.0133	0.0216	0.0320	0.0438	0.0559	0.0669
24	0.0003	0.0008	0.0020	0.0043	0.0083	0.0144	0.0226	0.0328	0.0442	0.0557
25	0.0001	0.0004	0.0010	0.0024	0.0050	0.0092	0.0154	0.0237	0.0336	0.0446

CUMULATIVE POISSON DISTRIBUTION

For a given value of λ an entry indicates the probability of a equal to or less than the specific value of x

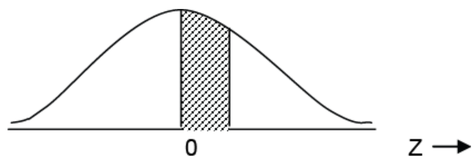
For example if λ = 5, the probability of x being equal to or less than 6 is 0.7622 (in bold below).

	λ									
x	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	0.3679	0.1353	0.0498	0.0183	0.0067	0.0025	0.0009	0.0003	0.0001	0.0000
1	0.7358	0.4060	0.1991	0.0916	0.0404	0.0174	0.0073	0.0030	0.0012	0.0005
2	0.9197	0.6767	0.4232	0.2381	0.1247	0.0620	0.0296	0.0138	0.0062	0.0028
3	0.9810	0.8571	0.6472	0.4335	0.2650	0.1512	0.0818	0.0424	0.0212	0.0103
4	0.9963	0.9473	0.8153	0.6288	0.4405	0.2851	0.1730	0.0996	0.0550	0.0293
5	0.9994	0.9834	0.9161	0.7851	0.6160	0.4457	0.3007	0.1912	0.1157	0.0671
6	0.9999	0.9955	0.9665	0.8893	**0.7622**	0.6063	0.4497	0.3134	0.2068	0.1301
7	1.0000	0.9989	0.9881	0.9489	0.8666	0.7440	0.5987	0.4530	0.3239	0.2202
8		0.9998	0.9962	0.9786	0.9319	0.8472	0.7291	0.5925	0.4557	0.3328
9		1.0000	0.9989	0.9919	0.9682	0.9161	0.8305	0.7166	0.5874	0.4579
10			0.9997	0.9972	0.9863	0.9574	0.9015	0.8159	0.7060	0.5830
11			0.9999	0.9991	0.9945	0.9799	0.9467	0.8881	0.8030	0.6968
12			1.0000	0.9997	0.9980	0.9912	0.9730	0.9362	0.8758	0.7916
13				0.9999	0.9993	0.9964	0.9872	0.9658	0.9261	0.8645
14				1.0000	0.9998	0.9986	0.9943	0.9827	0.9585	0.9165
15					0.9999	0.9995	0.9976	0.9918	0.9780	0.9513
16					1.0000	0.9998	0.9990	0.9963	0.9889	0.9730
17						0.9999	0.9996	0.9984	0.9947	0.9857
18						1.0000	0.9999	0.9993	0.9976	0.9928
19							1.0000	0.9997	0.9989	0.9965
20								0.9999	0.9996	0.9984
21								1.0000	0.9998	0.9993
22									0.9999	0.9997
23									1.0000	0.9999
24										1.0000

	λ									
x	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0012	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0049	0.0023	0.0011	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000
4	0.0151	0.0076	0.0037	0.0018	0.0009	0.0004	0.0002	0.0001	0.0000	0.0000
5	0.0375	0.0203	0.0107	0.0055	0.0028	0.0014	0.0007	0.0003	0.0002	0.0001
6	0.0786	0.0458	0.0259	0.0142	0.0076	0.0040	0.0021	0.0010	0.0005	0.0003
7	0.1432	0.0895	0.0540	0.0316	0.0180	0.0100	0.0054	0.0029	0.0015	0.0008
8	0.2320	0.1550	0.0998	0.0621	0.0374	0.0220	0.0126	0.0071	0.0039	0.0021
9	0.3405	0.2424	0.1658	0.1094	0.0699	0.0433	0.0261	0.0154	0.0089	0.0050
10	0.4599	0.3472	0.2517	0.1757	0.1185	0.0774	0.0491	0.0304	0.0183	0.0108
11	0.5793	0.4616	0.3532	0.2600	0.1848	0.1270	0.0847	0.0549	0.0347	0.0214
12	0.6887	0.5760	0.4631	0.3585	0.2676	0.1931	0.1350	0.0917	0.0606	0.0390
13	0.7813	0.6815	0.5730	0.4644	0.3632	0.2745	0.2009	0.1426	0.0984	0.0661
14	0.8540	0.7720	0.6751	0.5704	0.4657	0.3675	0.2808	0.2081	0.1497	0.1049
15	0.9074	0.8444	0.7636	0.6694	0.5681	0.4667	0.3715	0.2867	0.2148	0.1565
16	0.9441	0.8987	0.8355	0.7559	0.6641	0.5660	0.4677	0.3751	0.2920	0.2211
17	0.9678	0.9370	0.8905	0.8272	0.7489	0.6593	0.5640	0.4686	0.3784	0.2970
18	0.9823	0.9626	0.9302	0.8826	0.8195	0.7423	0.6550	0.5622	0.4695	0.3814
19	0.9907	0.9787	0.9573	0.9235	0.8752	0.8122	0.7363	0.6509	0.5606	0.4703
20	0.9953	0.9884	0.9750	0.9521	0.9170	0.8682	0.8055	0.7307	0.6472	0.5591
21	0.9977	0.9939	0.9859	0.9712	0.9469	0.9108	0.8615	0.7991	0.7255	0.6437
22	0.9990	0.9970	0.9924	0.9833	0.9673	0.9418	0.9047	0.8551	0.7931	0.7206
23	0.9995	0.9985	0.9960	0.9907	0.9805	0.9633	0.9367	0.8989	0.8490	0.7875
24	0.9998	0.9993	0.9980	0.9950	0.9888	0.9777	0.9594	0.9317	0.8933	0.8432
25	0.9999	0.9997	0.9990	0.9974	0.9938	0.9869	0.9748	0.9554	0.9269	0.8878

NORMAL DISTRIBUTION

This table gives the area under the normal curve between the mean and a point Z standard deviations above the mean. The corresponding area for deviations below the mean can be found by symmetry.



$Z=\dfrac{(x-\mu)}{\sigma}$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.0000	.0040	.0080	.0120	.0159	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1408	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4430	.4441
1.6	.4452	.4463	.4474	.4485	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4762	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4865	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4980	.4980	.4981
2.9	.4981	.4982	.4983	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	**.49865**	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	**.49903**	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	**.49931**	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	**.49952**	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	**.49966**	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	**.49977**									

T DISTRIBUTION TABLE

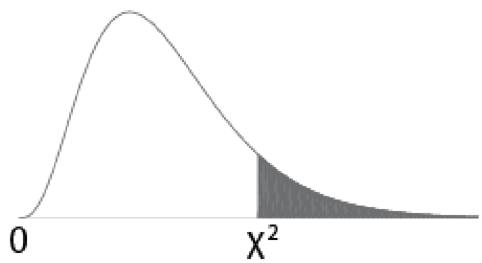
The critical values of t distribution are calculated according to the probabilities of two alpha values and the degrees of freedom. The Alpha(α) values 0.05 one tailed and 0.1 two tailed are the two columns to be compared with the degrees of freedom in the row of the table.

α (1 tail)	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
α (2 tail)	0.1	0.05	0.02	0.01	0.005	0.002	0.001
df							
1	6.3138	12.7065	31.8193	63.6551	127.3447	318.4930	636.0450
2	2.9200	4.3026	6.9646	9.9247	14.0887	22.3276	31.5989
3	2.3534	3.1824	4.5407	5.8408	7.4534	10.2145	12.9242
4	2.1319	2.7764	3.7470	4.6041	5.5976	7.1732	8.6103
5	2.0150	2.5706	3.3650	4.0322	4.7734	5.8934	6.8688
6	1.9432	2.4469	3.1426	3.7074	4.3168	5.2076	5.9589
7	1.8946	2.3646	2.9980	3.4995	4.0294	4.7852	5.4079
8	1.8595	2.3060	2.8965	3.3554	3.8325	4.5008	5.0414
9	1.8331	2.2621	2.8214	3.2498	3.6896	4.2969	4.7809
10	1.8124	2.2282	2.7638	3.1693	3.5814	4.1437	4.5869
11	1.7959	2.2010	2.7181	3.1058	3.4966	4.0247	4.4369
12	1.7823	2.1788	2.6810	3.0545	3.4284	3.9296	4.3178
13	1.7709	2.1604	2.6503	3.0123	3.3725	3.8520	4.2208
14	1.7613	2.1448	2.6245	2.9768	3.3257	3.7874	4.1404
15	1.7530	2.1314	2.6025	2.9467	3.2860	3.7328	4.0728
16	1.7459	2.1199	2.5835	2.9208	3.2520	3.6861	4.0150
17	1.7396	2.1098	2.5669	2.8983	3.2224	3.6458	3.9651
18	1.7341	2.1009	2.5524	2.8784	3.1966	3.6105	3.9216
19	1.7291	2.0930	2.5395	2.8609	3.1737	3.5794	3.8834
20	1.7247	2.0860	2.5280	2.8454	3.1534	3.5518	3.8495
21	1.7207	2.0796	2.5176	2.8314	3.1352	3.5272	3.8193
22	1.7172	2.0739	2.5083	2.8188	3.1188	3.5050	3.7921
23	1.7139	2.0686	2.4998	2.8073	3.1040	3.4850	3.7676
24	1.7109	2.0639	2.4922	2.7970	3.0905	3.4668	3.7454
25	1.7081	2.0596	2.4851	2.7874	3.0782	3.4502	3.7251

α (1 tail)	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
α (2 tail)	**0.1**	**0.05**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
df							
26	1.7056	2.0555	2.4786	2.7787	3.0669	3.4350	3.7067
27	1.7033	2.0518	2.4727	2.7707	3.0565	3.4211	3.6896
28	1.7011	2.0484	2.4671	2.7633	3.0469	3.4082	3.6739
29	1.6991	2.0452	2.4620	2.7564	3.0380	3.3962	3.6594
30	1.6973	2.0423	2.4572	2.7500	3.0298	3.3852	3.6459
31	1.6955	2.0395	2.4528	2.7440	3.0221	3.3749	3.6334
32	1.6939	2.0369	2.4487	2.7385	3.0150	3.3653	3.6218
33	1.6924	2.0345	2.4448	2.7333	3.0082	3.3563	3.6109
34	1.6909	2.0322	2.4411	2.7284	3.0019	3.3479	3.6008
35	1.6896	2.0301	2.4377	2.7238	2.9961	3.3400	3.5912
36	1.6883	2.0281	2.4345	2.7195	2.9905	3.3326	3.5822
37	1.6871	2.0262	2.4315	2.7154	2.9853	3.3256	3.5737
38	1.6859	2.0244	2.4286	2.7115	2.9803	3.3190	3.5657
39	1.6849	2.0227	2.4258	2.7079	2.9756	3.3128	3.5581
40	1.6839	2.0211	2.4233	2.7045	2.9712	3.3069	3.5510

CHI-SQUARE PROBABILITIES

The areas given across the top are the areas to the right of the critical value. To look up an area on the left, subtract it from one, and then look it up (i.e.: 0.05 on the left is 0.95 on the right)



The shaded area is equal to α for $X^2 = X^2_\alpha$

df	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	---	---	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997

df	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169