KALIKADEVI ART’S, COMMERCE & SCIENCE COLLEGE, SHIRUR(KA)

Department of Mathematics

MR.GHADGE R.B

Numerical analysis
unit 1

The image part with relationship ID rId2 was not found in the file.

The method is shown graphically in Fig. 2.1.

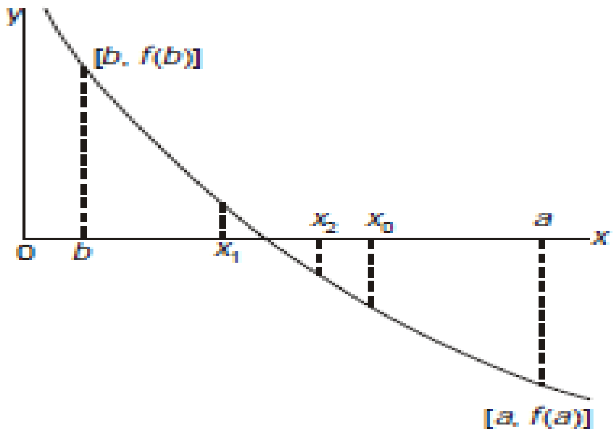


Figure 2.1 Graphical representation of the bisection method.

It should be noted that this method always succeeds. If there are more roots than one in the interval, bisection method finds one of the roots. It can be easily programmed using the following computational steps:

1. Choose two real numbers a and b such that $f(a)\, f(b) < 0$.
2. Set $x_r = (a + b)/2$.
3. (a) If $f(a)\, f(x_r) < 0$, the root lies in the interval (a, x_r). Then, set $b = x_r$ and go to step 2 above.
   (b) If $f(a)\, f(x_r) > 0$, the root lies in the interval (x_r, b). Then, set $a = x_r$ and go to step 2.
   (c) If $f(a)\, f(x_r) = 0$, it means that x_r is a root of the equation $f(x) = 0$ and the computation may be terminated.

In practical problems, the roots may not be exact so that condition (c) above is never satisfied. In such a case, we need to adopt a criterion for deciding when to terminate the computations.

A convenient criterion is to compute the percentage error ε_r defined by

$$\varepsilon_r = \left| \frac{x_r' - x_r}{x_r'} \right| \times 100\%. \qquad (2.5)$$

where x_r' is the new value of x_r. The computations can be terminated when ε_r becomes less than a prescribed tolerance, say ε_p. In addition, the maximum number of iterations may also be specified in advance.

Example 2.2 Find a real root of the equation $x^3 - 2x - 5 = 0$.

Let $f(x) = x^3 - 2x - 5$. Then

$$f(2) = -1 \quad \text{and} \quad f(3) = 16.$$

Hence a root lies between 2 and 3 and we take

$$x_1 = \frac{2+3}{2} = 2.5$$

Since $f(x_1) = f(2.5) = 5.6250$, the root lies between 2 and 2.25. Hence

$$x_2 = \frac{2+2.5}{2} = 2.25$$

Now, $f(x_2) = 1.890625$, the root lies between 2 and 2.25. Therefore,

$$x_3 = \frac{2+2.25}{2} = 2.125$$

Since $f(x_3) = 0.3457$, the root lies between 2 and 2.125. Therefore,

$$x_4 = \frac{2+2.125}{2} = 2.0625$$

Proceeding in this way, we obtain the successive approximations:

$$x_5 = 2.09375, \quad x_6 = 2.10938, \quad x_7 = 2.10156,$$
$$x_8 = 2.09766, \quad x_9 = 2.09570, \quad x_{10} = 2.09473,$$
$$x_{11} = 2.09424, \ldots$$

We find

$$x_{11} - x_{10} = -0.0005,$$

and

$$\left|\frac{x_{11} - x_{10}}{x_{11}}\right| \times 100 = \frac{0.0005}{2.09424} \times 100 = 0.02\%$$

Hence a root, correct to three decimal places, is 2.094.

2.3 METHOD OF FALSE POSITION

This is the oldest method for finding the real root of a nonlinear equation $f(x) = 0$ and closely resembles the bisection method. In this method, also known as *regula–falsi* or the *method of chords*, we choose two points a and b such that $f(a)$ and $f(b)$ are of opposite signs. Hence, a root must lie in between these points. Now, the equation of the chord joining the two points $[a, f(a)]$ and $[b, f(b)]$ is given by

$$\frac{y - f(a)}{x - a} = \frac{f(b) - f(a)}{b - a}. \qquad (2.6)$$

The method consists in replacing the part of the curve between the points $[a, f(a)]$ and $[b, f(b)]$ by means of the *chord* joining these points, and taking the point of intersection of the chord with the x-axis as an *approximation* to the root. The point of intersection in the present case is obtained by putting $y = 0$ in Eq. (2.6). Thus, we obtain

$$x_1 = a - \frac{f(a)}{f(b) - f(a)}(b - a) = \frac{af(b) - bf(a)}{f(b) - f(a)}, \qquad (2.7)$$

which is the *first approximation* to the root of $f(x) = 0$. If now $f(x_1)$ and $f(a)$ are of opposite signs, then the root lies between a and x_1, and we replace b by x_1 in Eq. (2.7), and obtain the *next* approximation. Otherwise, we replace a by x_1 and generate the next approximation. The procedure is repeated till the root is obtained to the desired accuracy. Figure 2.2 gives a graphical representation of the method. The error criterion Eq. (2.5) can be used in this case also.

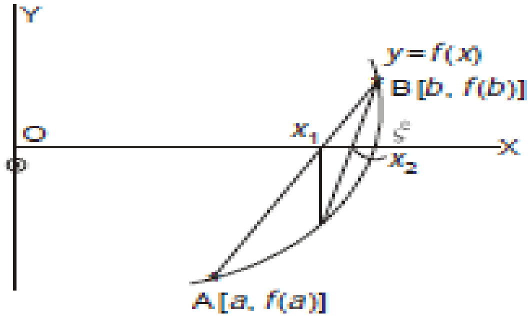


Figure 2.2 Method of false position.

Example 2.7 Given that the equation $x^{2.2} = 69$ has a root between 5 and 8. Use the method of regula–falsi to determine it.

Let $f(x) = x^{2.2} - 69$. We find

$$f(5) = -34.50675846 \quad \text{and} \quad f(8) = 28.00586026.$$

Hence

$$x_1 = \frac{5(28.00586026) - 8(-34.50675846)}{28.00586026 + 34.50675846} = 6.655990062.$$

Now, $f(x_1) = -4.275625415$ and therefore, the root lies between 6.655990062 and 8.0. We obtain

$$x_2 = 6.83400179, \qquad x_3 = 6.850669653.$$

The correct root is 6.8523651..., so that x_3 is correct to three significant figures.

The image part with relationship ID rId2 was not found in the file.

The image part with relationship ID rId2 was not found in the file.

2.5 NEWTON–RAPHSON METHOD

This method is generally used to improve the result obtained by one of the previous methods. Let x_0 be an approximate root of $f(x) = 0$ and let $x_1 = x_0 + h$ be the correct root so that $f(x_1) = 0$. Expanding $f(x_0 + h)$ by Taylor's series, we obtain

$$f(x_0) + hf'(x_0) + \frac{h^2}{2!} f''(x_0) + \cdots = 0.$$

Geometrically, the method consists in replacing the part of the curve between the point $[x_0, f(x_0)]$ and the x-axis by means of the tangent to the curve at the point, and is described graphically in Fig. 2.3. It can be used for solving both algebraic and transcendental equations and it can also be used when the roots are complex.

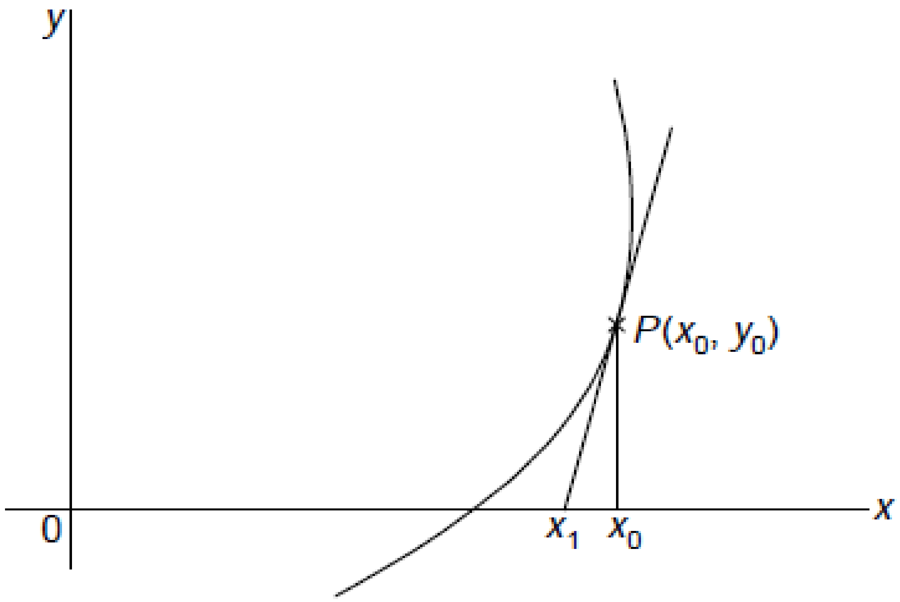


Figure 2.3 Newton–Raphson method.

Newton Raphson problem

Example 2.16 Find a root of the equation $x \sin x + \cos x = 0$.

We have

$$f(x) = x \sin x + \cos x \quad \text{and} \quad f'(x) = x \cos x.$$

The iteration formula is, therefore,

$$x_{n+1} = x_n - \frac{x_n \sin x_n + \cos x_n}{x_n \cos x_n}.$$

With $x_0 = \pi$, the successive iterates are given below

n	x_n	$f(x_n)$	x_{n+1}
0	3.1416	–1.0	2.8233
1	2.8233	–0.0662	2.7986
2	2.7986	–0.0006	2.7984
3	2.7984	0.0	2.7984

3.3.1 Forward Differences

If $y_0, y_1, y_2, \ldots, y_n$ denote a set of values of y, then $y_1 - y_0, y_2 - y_1, \ldots, y_n - y_{n-1}$ are called the *differences* of y. Denoting these differences by $\Delta y_0, \Delta y_1, \ldots, \Delta y_{n-1}$ respectively, we have

$$\Delta y_0 = y_1 - y_0, \qquad \Delta y_1 = y_2 - y_1, \ldots, \qquad \Delta y_{n-1} = y_n - y_{n-1},$$

where Δ is called the *forward difference operator* and $\Delta y_0, \Delta y_1, \ldots,$ are called *first forward differences*. The differences of the first forward differences are called *second forward differences* and are denoted by $\Delta^2 y_0, \Delta^2 y_1, \ldots$ Similarly, one can define *third forward differences*, *fourth forward differences*,

Table 3.1 Forward Difference Table

x	y_0	Δ	Δ^2	Δ^3	Δ^4	Δ^5	Δ^6
x_0	y_0						
		Δy_0					
x_1	y_1		$\Delta^2 y_0$				
		Δy_1		$\Delta^3 y_0$			
x_2	y_2		$\Delta^2 y_1$		$\Delta^4 y_0$		
		Δy_2		$\Delta^3 y_1$		$\Delta^5 y_0$	
x_3	y_3		$\Delta^2 y_2$		$\Delta^4 y_1$		$\Delta^6 y_0$
		Δy_3		$\Delta^3 y_2$		$\Delta^5 y_1$	
x_4	y_4		$\Delta^2 y_3$		$\Delta^4 y_2$		
		Δy_4		$\Delta^3 y_3$			
x_5	y_5		$\Delta^2 y_4$				
		Δy_5					
x_6	y_6						

In practical computations, the forward difference table can be formed in the following way. For the data points (x_i, y_i), $i = 0, 1, 2, \ldots, n$ and $x_i = x_0 + ih$, we have

$$\Delta y_j = y_{j+1} - y_i, j = 0, 1, \ldots, n - 1.$$

3.3.2 Backward Differences

The differences $y_1 - y_0, y_2 - y_1, \ldots, y_n - y_{n-1}$ are called first *backward differences* if they are denoted by $\nabla y_1, \nabla y_2, \ldots, \nabla y_n$ respectively, so that

$$\begin{array}{l} \nabla y_1 = y_1 - y_0, \quad \nabla y_2 = y_2 - y_1, \\ \vdots \qquad \vdots \qquad \vdots \qquad \vdots \\ \nabla y_n = y_n - y_{n-1}, \end{array}$$

where ∇ is called the *backward difference operator*. In a similar way, one can define backward differences of higher orders.

Table 3.3 Backward Difference Table

x	y	∇	∇^2	∇^3	∇^4	∇^5	∇^6
x_0	y_0						
x_1	y_1	∇y_1					
x_2	y_2	∇y_2	$\nabla^2 y_2$				
x_3	y_3	∇y_3	$\nabla^2 y_3$	$\nabla^3 y_3$			
x_4	y_4	∇y_4	$\nabla^2 y_4$	$\nabla^3 y_4$	$\nabla^4 y_4$		
x_5	y_5	∇y_5	$\nabla^2 y_5$	$\nabla^3 y_5$	$\nabla^4 y_5$	$\nabla^5 y_5$	
x_6	y_6	∇y_6	$\nabla^2 y_6$	$\nabla^3 y_6$	$\nabla^4 y_6$	$\nabla^5 y_6$	$\nabla^6 y_6$

Unit 2

6.4.1 Trapezoidal Rule

Setting $n = 1$ in the general formula (6.29), all differences higher than the first will become zero and we obtain

$$\int_{x_0}^{x_1} y\,dx = h\left(y_0 + \frac{1}{2}\Delta y_0\right) = h\left[y_0 + \frac{1}{2}(y_1 - y_0)\right] = \frac{h}{2}(y_0 + y_1). \quad (6.30)$$

For the next interval $[x_1, x_2]$, we deduce similarly

$$\int_{x_1}^{x_2} y\,dx = \frac{h}{2}(y_1 + y_2) \quad (6.31)$$

6.4.2 Simpson's 1/3-Rule

This rule is obtained by putting $n = 2$ in Eq. (6.29), i.e. by replacing the curve by $n/2$ arcs of second-degree polynomials or parabolas. We have then

$$\int_{x_0}^{x_2} y\,dx = 2h\left(y_0 + \Delta y_0 + \frac{1}{6}\Delta^2 y_0\right) = \frac{h}{3}(y_0 + 4y_1 + y_2).$$

Similarly,

$$\int_{x_2}^{x_4} y\,dx = \frac{h}{3}(y_2 + 4y_3 + y_4)$$

$$\vdots$$

and finally

$$\int_{x_{n-2}}^{x_n} y\,dx = \frac{h}{3}(y_{n-2} + 4y_{n-1} + y_n).$$

Summing up, we obtain

$$\int_{x_0}^{x_n} y\,dx = \frac{h}{3}[y_0 + 4(y_1 + y_3 + y_5 + \cdots + y_{n-1}) + 2(y_2 + y_4 + y_6 + \cdots + y_{n-2}) + y_n], \qquad (6.39)$$

6.4.3 Simpson's 3/8-Rule

Setting $n = 3$ in Eq. (6.29), we observe that all the differences higher than the third will become zero and we obtain

$$\int_{x_0}^{x_3} y\,dx = 3h\left(y_0 + \frac{3}{2}\Delta y_0 + \frac{3}{4}\Delta^2 y_0 + \frac{1}{8}\Delta^3 y_0\right)$$

$$= 3h\left[y_0 + \frac{3}{2}(y_1 - y_0) + \frac{3}{4}(y_2 - 2y_1 + y_0) + \frac{1}{8}(y_3 - 3y_2 + 3y_1 - y_0)\right]$$

$$= \frac{3h}{8}(y_0 + 3y_1 + 3y_2 + y_3).$$

Similarly

$$\int_{x_3}^{x_6} y\,dx = \frac{3h}{8}(y_3 + 3y_4 + 3y_5 + y_6)$$

and so on. Summing up all these, we obtain

$$\int_{x_0}^{x_n} y\,dx = \frac{3h}{8}[(y_0 + 3y_1 + 3y_2 + y_3) + (y_3 + 3y_4 + 3y_5 + y_6) + \cdots$$

$$+ (y_{n-3} + 3y_{n-2} + 3y_{n-1} + y_n)]$$

$$= \frac{3h}{8}(y_0 + 3y_1 + 3y_2 + 2y_3 + 3y_4 + 3y_5 + 2y_6 + \cdots$$

$$+ 2y_{n-3} + 3y_{n-2} + 3y_{n-1} + y_n) \tag{6.41}$$

7.5.1 Gauss Elimination

This is the elementary elimination method and it reduces the system of equations to an equivalent upper-triangular system, which can be solved by *back substitution*.

Let the system of n linear equations in n unknowns be given by

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + a_{n3}x_3 + \cdots + a_{nn}x_n &= b_n. \end{aligned}\right\} \tag{7.27}$$

There are two steps in the solution of the system given in Eq. (7.27), viz., the elimination of unknowns and back substitution.

Step 1: The unknowns are eliminated to obtain an upper-triangular system.

To eliminate x_1 from the second equation, we multiply the first equation by $(-a_{21}/a_{11})$ and obtain

$$-a_{21}x_1 - a_{12}\frac{a_{21}}{a_{11}}x_2 - a_{13}\frac{a_{21}}{a_{11}}x_3 - \cdots - a_{1n}\frac{a_{21}}{a_{11}}x_n = -b_1\frac{a_{21}}{a_{11}}.$$

Adding the above equation to the second equation of Eq. (7.27), we obtain

$$\left(a_{22} - a_{12}\frac{a_{21}}{a_{11}}\right)x_2 + \left(a_{23} - a_{13}\frac{a_{21}}{a_{11}}\right)x_3 + \cdots + \left(a_{2n} - a_{1n}\frac{a_{21}}{a_{11}}\right)x_n = b_2 - b_1\frac{a_{21}}{a_{11}}, \tag{7.28}$$

which can be written as

$$a'_{22}x_2 + a'_{23}x_3 + \cdots + a'_{2n}x_n = b'_2,$$

where $a'_{22} = a_{22} - a_{12}(a_{21}/a_{11})$, etc. Thus the primes indicate that the original element has changed its value. Similarly, we can multiply the first equation by $-a_{31}/a_{11}$ and add it to the third equation of the system (7.27). This eliminates the unknown x_1 from the third equation of Eq. (7.27) and we obtain

$$a'_{32}x_2 + a'_{33}x_3 + \cdots + a'_{3n}x_n = b'_3. \tag{7.29}$$

In a similar fashion, we can eliminate x_1 from the remaining equations and after eliminating x_1 from the last equation of Eq. (7.27), we obtain the system

$$
\left.
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= b_1 \\
a'_{22}x_2 + a'_{23}x_3 + \cdots + a'_{2n}x_n &= b'_2 \\
a'_{32}x_2 + a'_{33}x_3 + \cdots + a'_{3n}x_n &= b'_3 \\
&\vdots \\
a'_{n2}x_2 + a'_{n3}x_3 + \cdots + a'_{nn}x_n &= b'_n.
\end{aligned}
\right\} \qquad (7.30)
$$

We next eliminate x_2 from the last $(n-2)$ equations of Eq. (7.30). Before this, it is important to notice that in the process of obtaining the above system, we have multiplied the first row by $(-a_{21}/a_{11})$, i.e. we have divided it by a_{11} which is therefore assumed to be nonzero. For this reason, the first equation in the system (7.30) is called the *pivot equation*, and a_{11} is called the *pivot* or *pivotal element*. The method obviously fails if $a_{11} = 0$. We shall discuss this important point after completing the description of the elimination method. Now, to eliminate x_2 from the third equation of Eq. (7.30), we multiply the second equation by $(-a'_{32}/a'_{22})$ and add it to the third equation. Repeating this process with the remaining equations, we obtain the system

$$
\left.
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= b_1 \\
a'_{22}x_2 + a'_{23}x_3 + \cdots + a'_{2n}x_n &= b'_2 \\
a''_{33}x_3 + \cdots + a''_{3n}x_n &= b''_3 \\
&\vdots \\
a''_{n3}x_3 + \cdots + a''_{nn}x_n &= b''_n.
\end{aligned}
\right\} \qquad (7.31)
$$

In Eq. (7.31), the 'double primes' indicate that the *elements have changed twice*. It is easily seen that this procedure can be continued to eliminate x_3 from the fourth equation onwards, x_4 from the fifth equation onwards, etc., till we finally obtain the upper-triangular form:

$$
\left.
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= b_1 \\
a'_{22}x_2 + a'_{23}x_3 + \cdots + a'_{2n}x_n &= b'_2 \\
a''_{33}x_3 + \cdots + a''_{3n}x_n &= b''_3 \\
&\vdots \\
a_{nn}^{(n-1)}x_n &= b_n^{(n-1)},
\end{aligned}
\right\} \qquad (7.32)
$$

where $a_{nn}^{(n-1)}$ indicates that the element a_{nn} has changed $(n-1)$ times. We thus have completed the first step of elimination of unknowns and reduction to the upper-triangular form.

Step 2: We now have to obtain the required solution from the system (7.32). From the last equation of this system, we obtain

$$x_n = \frac{b_n^{(n-1)}}{a_{nn}^{(n-1)}}. \tag{7.33}$$

This is then substituted in the $(n-1)$th equation to obtain x_{n-1} and the process is repeated to compute the other unknowns. We have therefore first computed x_n, then $x_{n-1}, x_{n-2}, \ldots, x_2, x_1$, in that order. Due to this reason, the process is called *back substitution.*

Example 7.4 Use Gauss elimination to solve the system

$$2x + y + z = 10$$

$$3x + 2y + 3z = 18$$

$$x + 4y + 9z = 16.$$

We first eliminate x from the second and third equations. For this we multiply the first equation by (–3/2) and add to the second to get

$$y + 3z = 6. \qquad \text{(i)}$$

Similarly, we multiply the first equation by (–1/2) and add it to the third to get

$$7y + 17z = 22. \qquad \text{(ii)}$$

We thus have eliminated x from the second and third equations. Next, we have to eliminate y from (i) and (ii). For this we multiply (i) by –7 and add to (ii). This gives

$$-4z = -20 \quad \text{or} \quad z = 5.$$

The upper-triangular form is therefore given by

$$2x + y + z = 10$$
$$y + 3z = 6$$
$$z = 5.$$

It follows that the required solution is $x = 7$, $y = -9$ and $z = 5$.

The next example demonstrates the necessity of pivoting in the elimination method.

Example 7.5 Solve the system

$$0.0003120x_1 + 0.006032x_2 = 0.003328$$
$$0.5000x_1 + 0.8942x_2 = 0.9471$$

The exact solution is $x_1 = 1$ and $x_2 = 0.5$.

We first solve the system with pivoting. We write the given system as

$$0.5000x_1 + 0.8942x_2 = 0.9471$$
$$0.000312x_1 + 0.006032x_2 = 0.003328$$

using Gaussian elimination, the above system reduces to

$$0.5000x_1 + 0.8942x_2 = 0.9471$$
$$0.005474x_2 = 0.002737$$

Back substitution gives: $x_2 = 0.5$ and $x_1 = 1.0$.

Without pivoting, Gaussion elimination gives the system

$$0.000312x_1 + 0.006032x_2 = 0.003328$$
$$-8.7725x_2 = -5.3300$$

The back substitution process gives

$$x_2 = 0.6076 \text{ and } x_1 = -1.0803$$

Without pivoting, Gaussion elimination gives the system

$$0.000312x_1 + 0.006032x_2 = 0.003328$$
$$-8.7725x_2 = -5.3300$$

The back substitution process gives

$$x_2 = 0.6076 \text{ and } x_1 = -1.0803$$

The effect of pivoting is clearly seen.

7.5.3 Gauss–Jordan Method

This is a modification of the Gauss elimination method, the essential difference being that when an unknown is eliminated, it is eliminated from all equations. The method does not require back substitution to obtain the solution and is best illustrated by the following example.

Example 7.6 Solve the system (*see* Example 7.4)

$$2x + y + z = 10$$
$$3x + 2y + 3z = 18$$
$$x + 4y + 9z = 16.$$

by the Gauss–Jordan method.

Elimination of x from the second and third equations is done as in 'Gauss elimination' and we obtain the system

$$2x + y + z = 10$$
$$(1/2)y + (3/2)z = 3$$
$$(7/2)y + (17/2)z = 11.$$

Next, the unknown y is eliminated from *both* the first and third equations. This gives us

$$x - z = 2 \quad \text{and} \quad z = 5.$$

Hence the system becomes:

$$x - z = 2$$
$$y + 3z = 6$$
$$z = 5.$$

Evaluation of y and z is trivial and the result is the same as before.

Unit 3

8.2 SOLUTION BY TAYLOR'S SERIES

We consider the differential equation

$$y' = f(x, y) \tag{8.1a}$$

with the initial condition

$$y(x_0) = y_0. \tag{8.1b}$$

If $y(x)$ is the exact solution of Eq. (8.1), then the Taylor's series for $y(x)$ around $x = x_0$ is given by

$$y(x) = y_0 + (x - x_0)\, y_0' + \frac{(x - x_0)^2}{2!} y_0'' + \cdots \tag{8.2}$$

If the values of $y_0', y_0'', \ldots$ are known, then Eq. (8.2) gives a power series for y. Using the formula for total derivatives, we can write

$$y'' = f' = f_x + y' f_y = f_x + f f_y,$$

where the suffixes denote partial derivatives with respect to the variable concerned. Similarly, we obtain

$$\begin{aligned} y''' = f'' &= f_{xx} + f_{xy} f + f(f_{yx} + f_{yy} f) + f_y (f_x + f_y f) \\ &= f_{xx} + 2 f f_{xy} + f^2 f_{yy} + f_x f_y + f f_y^2 \end{aligned}$$

and other higher derivatives of y. The method can easily be extended to simultaneous and higher-order differential equations.

Example 8.1 From the Taylor series for $y(x)$, find $y(0.1)$ correct to four decimal places if $y(x)$ satisfies

$$y' = x - y^2 \quad \text{and} \quad y(0) = 1.$$

The Taylor series for $y(x)$ is given by

$$y(x) = 1 + xy'_0 + \frac{x^2}{2}y''_0 + \frac{x^3}{6}y'''_0 + \frac{x^4}{24}y_0^{\text{iv}} + \frac{x^5}{120}y_0^{\text{v}} + \cdots$$

The derivatives $y'_0, y''_0, \ldots$ etc. are obtained thus:

$$\begin{aligned}
y'(x) &= x - y^2 & y'_0 &= -1\\
y''(x) &= 1 - 2yy' & y''_0 &= 3\\
y'''(x) &= -2yy'' - 2y'^2 & y'''_0 &= -8\\
y^{\text{iv}}(x) &= -2yy''' - 6y'y'' & y_0^{\text{iv}} &= 34\\
y^{\text{v}}(x) &= -2yy^{\text{iv}} - 8y'y''' - 6y''^2 & y_0^{\text{v}} &= -186
\end{aligned}$$

Using these values, the Taylor series becomes

$$y(x) = 1 - x + \frac{3}{2}x^2 - \frac{4}{3}x^3 + \frac{17}{12}x^4 - \frac{31}{20}x^5 + \cdots$$

To obtain the value of $y(0.1)$ correct to four decimal places, it is found that the terms up to x^4 should be considered, and we have $y(0.1) = 0.9138$.

Suppose that we wish to find the range of values of x for which the above series, truncated after the term containing x^4, can be used to compute the values of y correct to four decimal places. We need only to write

$$\frac{31}{20}x^5 \le 0.00005 \quad \text{or} \quad x \le 0.126.$$

Example 8.2 Given the differential equation

$$y'' - xy' - y = 0$$

with the conditions $y(0) = 1$ and $y'(0) = 0$, use Taylor's series method to determine the value of $y\,(0.1)$.

We have $y\,(x) = 1$ and $y'(x) = 0$ when $x = 0$. The given differential equation is

$$y''(x) = xy'(x) + y(x) \tag{i}$$

Hence $y''(0) = y(0) = 1$. Successive differentiation of (i) gives

$$y'''(x) = xy''(x) + y'(x) + y'(x) = xy''(x) + 2y'(x), \tag{ii}$$

$$y^{iv}(x) = xy'''(x) + y''(x) + 2y''(x) = xy'''(x) + 3y''(x), \tag{iii}$$

$$y^{v}(x) = xy^{iv}(x) + y'''(x) + 3y'''(x) = xy^{iv}(x) + 4y'''(x), \tag{iv}$$

$$y^{vi}(x) = xy^{v}(x) + y^{iv}(x) + 4y^{iv}(x) = xy^{v}(x) + 5y^{iv}(x), \tag{v}$$

and similarly for higher derivatives. Putting $x=0$ in (ii) to (v), we obtain

$$y'''(0)=2y'(0)=0, \quad y^{iv}(0)=3y''(0)=3, \quad y^{v}(0)=0, \quad y^{vi}(0)=5.$$

By Taylor's series, we have

$$y(x)=y(0)+xy'(0)+\frac{x^2}{2}y''(0)+\frac{x^3}{6}y'''(0)+\frac{x^4}{24}y^{iv}(0)$$

$$+\frac{x^5}{120}y^{v}(0)+\frac{x^6}{720}y^{vi}(0)+\cdots$$

Hence

$$y(0.1)=1+\frac{(0.1)^2}{2}+\frac{(0.1)^4}{24}(3)+\frac{(0.1)^6}{720}(5)+\cdots$$

$$=1+0.005+0.0000125, \text{ neglecting the last term}$$

$$=1.0050125, \text{ correct to seven decimal places.}$$

Example 8.5 To illustrate Euler's method, we consider the differential equation $y' = -y$ with the condition $y(0) = 1$.

Successive application of Eq. (8.8) with $h = 0.01$ gives

$$y(0.01) = 1 + 0.1(-1) = 0.99$$

$$y(0.02) = 0.99 + 0.01\,(-0.99) = 0.9801$$

$$y(0.03) = 0.9801 + 0.01\,(-0.9801) = 0.9703$$

$$y(0.04) = 0.9703 + 0.01\,(-0.9703) = 0.9606.$$

The exact solution is $y = e^{-x}$ and from this the value at $x = 0.04$ is 0.9608.

8.4.2 Modified Euler's Method

Instead of approximating $f(x, y)$ by $f(x_0, y_0)$ in Eq. (8.6), we now approximate the integral given in Eq. (8.6) by means of trapezoidal rule to obtain

$$y_1 = y_0 + \frac{h}{2}[f(x_0, y_0) + f(x_1, y_1)] \tag{8.13}$$

We thus obtain the iteration formula

$$y_1^{(n+1)} = y_0 + \frac{h}{2}[f(x_0, y_0) + f(x_1, y_1^{(n)})], \quad n = 0, 1, 2, \ldots \tag{8.14}$$

where $y_1^{(n)}$ is the nth approximation to y_1. The iteration formula (8.14) can be started by choosing $y_1^{(0)}$ from Euler's formula:

$$y_1^{(0)} = y_0 + hf(x_0, y_0).$$

Example 8.7 Determine the value of y when $x = 0.1$ given that

$$y(0) = 1 \quad \text{and} \quad y' = x^2 + y$$

We take $h=0.05$. With $x_0 = 0$ and $y_0 = 1.0$, we have $f(x_0, y_0) = 1.0$. Hence Euler's formula gives

$$y_1^{(0)} = 1 + 0.05(1) = 1.05$$

Further, $x_1 = 0.05$ and $f(x_1, y_1^{(0)}) = 1.0525$. The average of $f(x_0, y_0)$ and $f(x_1, y_1^{(0)})$ is 1.0262. The value of $y_1^{(1)}$ can therefore be computed by using Eq. (8.14) and we obtain

$$y_1^{(1)} = 1.0513.$$

Repeating the procedure, we obtain $y_1^{(2)} = 1.0513$. Hence we take $y_1 = 1.0513$, which is correct to four decimal places.

Next, with $x_1 = 0.05$, $y_1 = 1.0513$ and $h = 0.05$, we continue the procedure to obtain y_2, i.e., the value of y when $x = 0.1$. The results are

$$y_2^{(0)} = 1.1040, \quad y_2^{(1)} = 1.1055, \quad y_2^{(2)} = 1.1055.$$

Hence we conclude that the value of y when $x = 0.1$ is 1.1055.

Example 8.8 Given $dy/dx = y - x$ where $y(0) = 2$, find $y(0.1)$ and $y(0.2)$ correct to four decimal places.

(i) *Runge–Kutta second-order formula:* With $h = 0.1$, we find $k_1 = 0.2$ and $k_2 = 0.21$. Hence

$$y_1 = y(0.1) = 2 + \frac{1}{2}(0.41) = 2.2050.$$

To determine $y_2 = y(0.2)$, we note that $x_0 = 0.1$ and $y_0 = 2.2050$. Hence, $k_1 = 0.1(2.105) = 0.2105$ and $k_2 = 0.1(2.4155 - 0.2) = 0.22155$.

It follows that

$$y_2 = 2.2050 + \frac{1}{2}(0.2105 + 0.22155) = 2.4210.$$

Proceeding in a similar way, we obtain

$$y_3 = y(0.3) = 2.6492 \quad \text{and} \quad y_4 = y(0.4) = 2.8909$$

We next choose $h = 0.2$ and compute $y(0.2)$ and $y(0.4)$ directly. With $h = 0.2$. $x_0 = 0$ and $y_0 = 2$, we obtain $k_1 = 0.4$ and $k_2 = 0.44$ and hence $y(0.2) = 2.4200$. Similarly, we obtain $y(0.4) = 2.8880$.

From the analytical solution $y = x + 1 + e^x$, the exact values of $y(0.2)$ and $y(0.4)$ are respectively 2.4214 and 2.8918. To study the order of convergence of this method, we tabulate the values as follows:

x	Computed y	Exact y	Difference	Ratio
0.2	$h = 0.1$: 2.4210	2.4214	0.0004	3.5
	$h = 0.2$: 2.4200		0.0014	
0.4	$h = 0.1$: 2.8909	2.8918	0.0009	4.2
	$h = 0.2$: 2.8880		0.0038	

It follows that the method has an h^2-order of convergence.

(ii) *Runge–Kutta fourth-order formula:* To determine $y(0.1)$, we have $x_0 = 0$, $y_0 = 2$ and $h = 0.1$. We then obtain

$$k_1 = 0.2,$$
$$k_2 = 0.205$$
$$k_3 = 0.20525$$
$$k_4 = 0.21053.$$

Hence

$$y(0.1) = 2 + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) = 2.2052.$$

Proceeding similarly, we obtain $y(0.2) = 2.4214$.

8.6 PREDICTOR–CORRECTOR METHODS

In the methods described so far, to solve a differential equation over a single interval, say from $x = x_n$ to $x = x_{n+1}$, we required information only at the beginning of the interval, i.e. at $x = x_n$. *Predictor–corrector* methods are the ones which require function values at x_n, x_{n-1}, x_{n-2}, … for the computation of the function value at x_{n+1} A *predictor* formula is used to predict the value of y at x_{n+1} and then a *corrector* formula is used to improve the value of $y_{n+1.}$

In Section 8.6.1 we derive Predictor–corrector formulae which use backward differences and in Section 8.6.2 we describe Milne's method which uses forward differences.

8.6.1 Adams–Moulton Method

Newton's backward difference interpolation formula can be written as

$$f(x, y) = f_0 + n\nabla f_0 + \frac{n(n+1)}{2}\nabla^2 f_0 + \frac{n(n+1)(n+2)}{6}\nabla^3 f_0 + \cdots \tag{8.22}$$

where

$$n = \frac{x - x_0}{h} \quad \text{and} \quad f_0 = f(x_0, y_0).$$

If this formula is substituted in

$$y_1 = y_0 + \int_{x_0}^{x_1} f(x, y)\, dx, \tag{8.23}$$

we get

$$y_1 = y_0 + \int_{x_0}^{x_1} \left[f_0 + n\nabla f_0 + \frac{n(n+1)}{2}\nabla^2 f_0 + \cdots \right] dx$$

$$= y_0 + h\int_0^1 \left[f_0 + n\nabla f_0 + \frac{n(n+1)}{2}\nabla^2 f_0 + \cdots \right] dn$$

$$= y_0 + h\left(1 + \frac{1}{2}\nabla + \frac{5}{12}\nabla^2 + \frac{3}{8}\nabla^3 + \frac{251}{720}\nabla^4 + \cdots\right) f_0.$$

It can be seen that the right side of the above relation depends only on y_0, y_{-1}, y_{-2}, ..., all of which are known. Hence this formula can be used to compute y_1. We therefore write it as

$$y_1^p = y_0 + h\left(1 + \frac{1}{2}\nabla + \frac{5}{12}\nabla^2 + \frac{3}{8}\nabla^3 + \frac{251}{720}\nabla^4 + \cdots\right) f_0 \tag{8.24}$$

This is called *Adams–Bashforth* formula and is used as a *predictor* formula (the superscript p indicating that it is a predicted value).

A corrector formula can be derived in a similar manner by using Newton's backward difference formula at f_1:

$$f(x, y) = f_1 + n\nabla f_1 + \frac{n(n+1)}{2}\nabla^2 f_1 + \frac{n(n+1)(n+2)}{6}\nabla^3 f_1 + \cdots \tag{8.25}$$

Example 8.12 We consider again the differential equation discussed in Examples 8.9 and 8.10, viz., to solve $y' = 1 + y^2$ with $y(0) = 0$ and we wish to compute $y(0.8)$ and $y(1.0)$.

With $h = 0.2$, the values of $y(0.2)$, $y(0.4)$ and $y(0.6)$ are computed in Example 8.9 and these values are given in the table below:

x	y	$y' = 1 + y^2$
0	0	1.0
0.2	0.2027	1.0411
0.4	0.4228	1.1787
0.6	0.6841	1.4681

To obtain $y(0.8)$, we use Eq. (8.32) and obtain

$$y(0.8) = 0 + \frac{0.8}{3}[2(1.0411) - 1.1787 + 2(1.4681)] = 1.0239$$

This gives

$$y'(0.8) = 2.0480.$$

To correct this value of $y(0.8)$, we use formula (8.34) and obtain

$$y(0.8) = 0.4228 + \frac{0.2}{3}[1.1787 + 4(1.4681) + 2.0480] = 1.0294.$$

Proceeding similarly, we obtain $y(1.0) = 1.5549$. The accuracy in the values of $y(0.8)$ and $y(1.0)$ can, of course, be improved by repeatedly using formula (8.34).

Example 8.13 The differential equation $y' = x^2 + y^2 - 2$ satisfies the following data:

x	y
–0.1	1.0900
0	1.0000
0.1	0.8900
0.2	0.7605

Use Milne's method to obtain the value of $y(0.3)$.

We first form the following table:

x	y	$y' = x^2 + y^2 - 2$
–0.1	1.0900	–0.80190
0	1.0	–1.0
0.1	0.8900	–1.19790
0.2	0.7605	–1.38164

Using Eq. (8.32), we obtain

$$y(0.3) = 1.09 + \frac{4(0.1)}{3}[2(-1) - (-1.19790) + 2(-1.38164)] = 0.614616.$$

In order to apply Eq. (8.34), we need to compute $y'(0.3)$. We have

$$y'(0.3) = (0.3)^2 + (0.614616)^2 - 2 = -1.532247.$$

Now, Eq. (8.34) gives the corrected value of $y(0.3)$:

$$y(0.3) = 0.89 + \frac{0.1}{3}[-1.197900 + 4(-1.38164) + (-1.532247)] = 0.614776.$$

Unit 4

(a) **Range**The measure of dispersion which is easiest to understand and easiest to calculate is the **range**. Range is defined as:

Range = Largest observation – Smallest observation

(b) **Mean Deviation**

(i) Mean deviation for ungrouped data:

For n observation $x_1, x_2, ..., x_n$, the **mean deviation about their mean** $\bar{x}$ is given by

$$\text{M.D}(\bar{x}) = \frac{|x_i - \bar{x}|}{n} \quad (1)$$

Mean deviation about their median M is given by

$$\text{M.D}(\text{M}) = \frac{|x_i - \text{M}|}{n} \quad (2)$$

(ii) Mean deviation for discrete frequency distribution

Let the given data consist of discrete observations $x_1, x_2, \ldots, x_n$ occurring with frequencies $f_1, f_2, \ldots, f_n$, respectively. In this case

$$\text{M.D}\,(\bar{x}) = \frac{f_i\,|x_i - \bar{x}|}{f_i} = \frac{f_i\,|x_i - \bar{x}|}{\text{N}} \qquad (3)$$

$$\text{M.D (M)} = \frac{f_i\,|x_i - \text{M}|}{\text{N}} \qquad (4)$$

where $\text{N} = \quad f_i$.

(iii) **Mean deviation for continuous frequency distribution (Grouped data).**

$$\text{M.D}(\bar{x}) = \frac{f_i\,|x_i - \bar{x}|}{\text{N}} \qquad (5)$$

$$\text{M.D}(\text{M}) = \frac{f_i\,|x_i - \text{M}|}{\text{N}} \qquad (6)$$

where x_i are the midpoints of the classes, $\bar{x}$ and M are, respectively, the mean and median of the distribution.

(c) **Variance :** Let $x_1, x_2, ..., x_n$ be n observations with $\bar{x}$ as the mean. The variance, denoted by σ^2, is given by

$$\sigma^2 = \frac{1}{n} \quad (x_i - \bar{x})^2 \qquad (7)$$

(d) **Standard Deviation:** If σ^2 is the variance, then σ, is called the standard deviation, is given by

$$\sigma = \sqrt{\frac{1}{n} \quad (x_i - \bar{x})^2} \qquad (8)$$

(e) **Standard deviation for a discrete frequency distribution is given by**

$$\sigma = \sqrt{\frac{1}{\text{N}} \quad f_i (x_i - \bar{x})^2} \qquad (9)$$

where f_i's are the frequencies of x_i' s and $\text{N} = \sum_{i=1}^{n} f_i$.

(f) **Standard deviation of a continuous frequency distribution (grouped data) is given by**

$$\sigma = \sqrt{\frac{1}{N} \quad f_i(x_i - \overline{x})^2} \qquad (10)$$

where x_i are the midpoints of the classes and f_i their respective frequencies. Formula (10) is same as

$$\sigma = \frac{1}{N}\sqrt{N \quad f_i x_i^2 - (\quad f_i x_i)^2} \qquad (11)$$

(g) Another formula for standard deviation :

$$\sigma_x = \frac{h}{N}\sqrt{N \quad f_i y_i^2 - (\quad f_i y_i)^2} \qquad (12)$$

where h is the width of class intervals and $y_i = \frac{x_i - A}{h}$ and A is the assumed mean.

15.1.2 ***Coefficient of variation*** It is sometimes useful to describe **variability** by expressing the standard deviation as a proportion of mean, usually a percentage. The formula for it as a percentage is

$$\text{Coefficient of variation} = \frac{\text{Standard deviation}}{\text{Mean}} \times 100$$

Example 1 Find the mean deviation about the mean of the following data:

Size (x):	1	3	5	7	9	11	13	15
Frequency (f):	3	3	4	14	7	4	3	4

Solution Mean = $\bar{x} = \dfrac{f_i x_i}{f_i} = \dfrac{3+9+20+98+63+44+39+60}{42} = \dfrac{336}{42} = 8$

$$\text{M.D.}\,(\bar{x}) = \frac{f_i\,|x_i - \bar{x}|}{f_i} = \frac{3(7)+3(5)+4(3)+14(1)+7(1)+4(3)+3(5)+4(7)}{42}$$

$$= \frac{21+15+12+14+7+12+15+28}{42} = \frac{62}{21} = 2.95$$

Example 2 Find the variance and standard deviation for the following data:

57, 64, 43, 67, 49, 59, 44, 47, 61, 59

Solution Mean ($\bar{x}$) $= \frac{57+64+43+67+49+59+61+59+44+47}{10} = \frac{550}{10} = 55$

$$\text{Variance } (\sigma^2) = \frac{(x_i - \bar{x})^2}{n}$$

$$= \frac{2^2 + 9^2 + 12^2 + 12^2 + 6^2 + 4^2 + 6^2 + 4^2 + 11^2 + 8^2}{10}$$

$$= \frac{662}{10} = 66.2$$

Standard deviation $(\sigma) = \sqrt{\sigma^2} = \sqrt{66.2} = 8.13$

Example 3 Show that the two formulae for the standard deviation of ungrouped data.

$$\sigma = \sqrt{\frac{(x_i - \overline{x})^2}{n}} \quad \text{and} \quad \sigma' = \sqrt{\frac{x_i^2}{n} - \overline{x}^2}$$

are equivalent.

Solution We have

$$(x_i - \overline{x})^2 = (x_i^2 - 2\overline{x}\, x_i + \overline{x}^2)$$

$$= x_i^2 + \quad - 2\overline{x}\, x_i + \quad \overline{x}^2$$

$$= x_i^2 - 2\overline{x} \quad x_i + (\overline{x})^2 \quad 1$$

$$= x_i^2 - 2\overline{x}\,(n\overline{x}) + n\,\overline{x}^2$$

$$= x_i^2 - n\overline{x}^2$$

Example 4 Calculate **variance** of the following data :

Class interval	Frequency
4 - 8	3
8 - 12	6
12 - 16	4
16 - 20	7

$$\text{Mean } (\bar{x}) = \frac{f_i x_i}{f_i} = \frac{3 \times 6 + 6 \times 10 + 4 \times 14 + 7 \times 18}{20} = 13$$

Solution Variance $(\sigma^2) = \dfrac{f_i\,(x_i - \bar{x})^2}{f_i} = \dfrac{3(-7)^2 + 6(-3)^2 + 4(1)^2 + 7(5)^2}{20}$

$$= \frac{147 + 54 + 4 + 175}{20} = 19$$

Example 5 Calculate mean, variation and standard deviation of the following frequency distribution:

Classes	Frequency
1 - 10	11
10 - 20	29
20 - 30	18
30 - 40	4
40 - 50	5
50 - 60	3

Solution Let A, the assumed mean, be 25.5. Here $h = 10$

Classes	x_i	$y_i = \frac{x_i - 25.5}{10}$	f_i	$f_i y_i$	$f_i y_i^2$
1 - 10	5.5	–2	11	–22	44
10 - 20	15.5	–1	29	–29	29
20 - 30	25.5	0	18	0	0
30 - 40	35.5	1	4	4	4
40 - 50	45.5	2	5	10	20
50 - 60	55.5	3	3	9	27
			70	–28	124

$$x' = \frac{f_i y_i}{f_i} = \frac{-28}{70} = -0.4$$

Mean $= \bar{x} = 25.5 + (-10)(0.4) = 21.5$

Variance $$(\sigma^2) = \frac{h}{N}\sqrt{N \quad f_i y_i^2 - (\quad f_i y_i)^2}^{\,2}$$

$$= \frac{10 \times 10}{70 \times 70}\,[70(124) - (-28)^2]$$

$$= \frac{70\,(124)}{7 \times 7} - \frac{28 \times 28}{7 \times 7} = \frac{1240}{7} - 16 = 161$$

S.D. $(\sigma) = \sqrt{161} = 12.7$

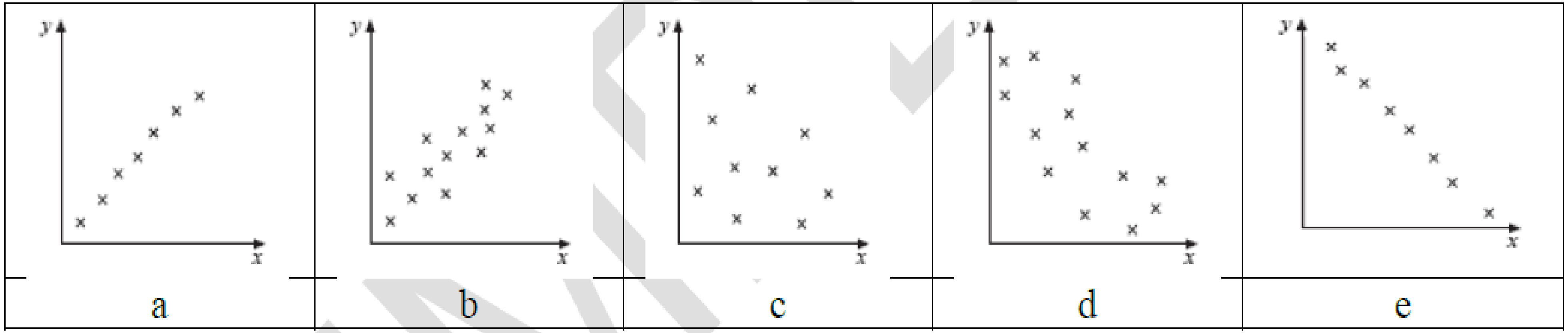
Unit 5

Correlation is a statistical measure that indicates the extent to which two or more variables fluctuate together. A positive correlation indicates the extent to which those variables increase or decrease in parallel; a negative correlation indicates the extent to which one variable increases as the other decreases.

When the fluctuation of one variable reliably predicts a similar fluctuation in another variable, there's often a tendency to think that means that the change in one causes the change in the other. However, correlation does not imply causation. There may be an unknown factor that influences both variables similarly.

Correlation is a statistical technique that can show whether and how strongly pairs of variables are related. Although this correlation is fairly obvious your data may contain unsuspected correlations. You may also suspect there are correlations, but don't know which are the strongest. An intelligent correlation analysis can lead to a greater understanding of your data.

- Correlation is **Positive** or direct when the values **increase** together, and
- Correlation is **Negative** when one value **decreases** as the other increases, and so called inverse or contrary correlation.



If the points plotted were all on a straight line we would have perfect correlation, but it could be positive or negative as shown in the diagrams above,

a. Strong positive correlation between x and y. The points lie close to a straight line with y increasing as x increases.
b. Weak, positive correlation between x and y. The trend shown is that y increases as x increases but the points are not close to a straight line
c. No correlation between x and y; the points are distributed randomly on the graph.
d. Weak, negative correlation between x and y. The trend shown is that y decreases as x increases but the points do not lie close to a straight line
e. Strong, negative correlation. The points lie close to a straight line, with y decreasing as x increases

2.2. Assumption of Correlation

Employing of correlation rely on some underlying assumptions. The variables are assumed to be independent, assume that they have been randomly selected from the population; the two variables are normal distribution; association of data is homoscedastic (homogeneous), homoscedastic data have the same standard deviation in different groups where data are heteroscedastic have different standard deviations in different groups and assumes that the relationship between the two variables is linear. The correlation coefficient is not satisfactory and difficult to interpret the associations between the variables in case if data have outliers.

An inspection of a scatterplot can give an impression of whether two variables are related and the direction of their relationship. But it alone is not sufficient to determine whether there is an association between two variables. The relationship depicted in the scatterplot needs to be described qualitatively. Descriptive statistics that express the degree of relation between two variables are called correlation coefficients. A commonly employed correlation coefficient are Pearson correlation, Kendall rank correlation and Spearman correlation.

Correlation used to examine the presence of a linear relationship between two variables providing certain assumptions about the data are satisfied. The results of the analysis, however, need to be interpreted with care, particularly when looking for a causal relationship.

2.3. Bivariate Correlation

Bivariate correlation is a measure of the relationship between the two variables; it measures the strength and direction of their relationship, the strength can range from absolute value 1 to 0. The stronger the relationship, the closer the value is to 1. Direction of The relationship can be positive (direct) or negative (inverse or contrary); correlation generally describes the effect that two or more phenomena occur together and therefore they are linked For example, the positive relationship of .71 can represent positive correlation between the statistics degrees and the science degrees. The student who has high degree in statistics has also high degree in science and vice versa.

The Pearson correlation coefficient is given by the following equation:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Where $\bar{x}$ is the mean of variable x values, and $\bar{y}$ is the mean of variable y values.

Example – Correlation of statistics and science tests

A study is conducted involving 10 students to investigate the association between statistics and science tests. The question arises here; is there a relationship between the degrees gained by the 10 students in statistics and science tests?

Table (2.1) Student degree in Statistic and science

Students	1	2	3	4	5	6	7	8	9	10
Statistics	20	23	8	29	14	12	11	20	17	18
Science	20	25	11	24	23	16	12	21	22	26

Notes: the marks out of 30

Suppose that (x) denotes for statistics degrees and (y) for science degree

Calculating the mean $(\bar{x}\ ,\ \bar{y})$;

$$\bar{x}=\frac{\sum x}{n}=\frac{173}{10}=17.3\ ,\ \bar{y}=\frac{\sum y}{n}=\frac{200}{10}=20$$

Where the mean of statistics degrees $\bar{x}=17.3$ and the mean of science degrees $\bar{y}=20$

Table (2.2) Calculating the equation parameters

Statistics	Science					
x	y	$x-\bar{x}$	$(x-\bar{x})^2$	$y-\bar{y}$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
20	20	2.7	7.29	0	0	0
23	25	5.7	32.49	5	25	28
8	11	-9.3	86.49	-9	81	83
29	24	11.7	136.89	4	16	46
14	23	-3.3	10.89	3	9	-9.9
12	16	-5.3	28.09	-4	16	21.2
11	12	-6.3	39.69	-8	64	50.4
21	21	3.7	13.69	1	1	3.7
17	22	-0.3	0.09	2	4	-0.6
18	26	0.7	0.49	6	36	4.2
173	200	0	356.1	0	252	228

$$\sum(x-\bar{x})^2=356.1\ ,\ \sum(y-\bar{y})^2=252\ ,$$
$$\sum(x-\bar{x})(y-\bar{y})=228$$

Calculating the Pearson correlation coefficient;

$$r=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}}=\frac{228}{\sqrt{356.1}\sqrt{252}}$$
$$=\frac{228}{(18.8706)(15.8745)}=\frac{228}{299.5614}=0.761$$

Other solution

Also; the Pearson correlation coefficient is given by the following equation:

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

Table (2.3) Calculating the equation parameters

x	y	xy	x^2	y^2
20	20	400	400	400
23	25	575	529	625
8	11	88	64	121
29	24	696	841	576
14	23	322	196	529
12	16	192	144	256
11	12	132	121	144
21	21	441	441	441
17	22	374	289	484
18	26	468	324	676
173	200	3688	3349	4252

Required calculation
$\sum x = 173$, $\sum y = 200$
$\sum xy = 3688$
$\sum x^2 = 3349$
$\sum y^2 = 4252$

Calculating the Pearson correlation coefficient by substitute in the aforementioned equation;

$$r = \frac{3688 - \frac{(173)(200)}{10}}{\sqrt{\left(3349 - \frac{(173)^2}{n10}\right)\left(4252 - \frac{(200)^2}{10n}\right)}} = \frac{228}{\sqrt{(356.1)(252)}} = \frac{228}{299.5614} = 0.761$$

Pearson Correlation coefficient $r = 0.761$ exactly the same output of the first equation.

The calculation shows a strong positive correlation (0.761) between the student's statistics and science degrees. This means that as degrees of statistics increases the degrees of science increase also. Generally the student who has a high degree in statistics has high degree in science and vice versa.

3.1. Definition

Regression analysis is one of the most commonly used statistical techniques in social and behavioral sciences as well as in physical sciences which involves identifying and evaluating the relationship between a dependent variable and one or more independent variables, which are also called predictor or explanatory variables. It is particularly useful for assess and adjusting for confounding. Model of the relationship is hypothesized and estimates of the parameter values are used to develop an estimated regression equation. Various tests are then employed to determine if the model is satisfactory. If the model is deemed satisfactory, the estimated regression equation can be used to predict the value of the dependent variable given values for the independent variables.

Linear regression explores relationships that can be readily described by straight lines or their generalization to many dimensions. A surprisingly large number of problems can be solved by linear regression, and even more by means of transformation of the original variables that result in linear relationships among the transformed variables.

When there is a single continuous dependent variable and a single independent variable, the analysis is called a **simple linear regression analysis**. This analysis assumes that there is a linear association between the two variables. **Multiple regression** is to learn more about the relationship between several independent or predictor variables and a dependent or criterion variable.

Independent variables are characteristics that can be measured directly; these variables are also called predictor or explanatory variables used to predict or to explain the behavior of the dependent variable.

Dependent variable is a characteristic whose value depends on the values of independent variables.

3.3. Assumption of Regression Analysis

The regression model is based on the following assumptions.

- The relationship between independent variable and dependent is linear.
- The expected value of the error term is zero
- The variance of the error term is constant for all the values of the independent variable, the assumption of homoscedasticity.
- There is no autocorrelation.
- The independent variable is uncorrelated with the error term.
- The error term is normally distributed.
- On an average difference between the observed value (yi) and the predicted value (ˆyi) is zero.
- On an average the estimated values of errors and values of independent variables are not related to each other.
- The squared differences between the observed value and the predicted value are similar.
- There is some variation in independent variable. If there are more than one variable in the equation, then two variables should not be perfectly correlated.

Intercept or Constant

- Intercept is the point at which the regression intercepts y-axis.
- Intercept provides a measure about the mean of dependent variable when slope(s) are zero.
- If slope(s) are not zero then intercept is equal to the mean of dependent variable minus slope × mean of independent variable.

Slope

- Change is dependent variable as we change independent variable.
- Zero Slope means that independent variable does not have any influence on dependent variable.
- For a linear model, slope is not equal to elasticity. That is because; elasticity is percent change in dependent variable, as a result one percent change in independent variable.

Example – linear Regression of patient's age and their blood pressure

A study is conducted involving 10 patients to investigate the relationship and effects of patient's age and their blood pressure.

Table (3.1) calculating the linear regression of patient's age and blood pressure

Obs	Age	BP		
	x	y	xy	x^2
1	35	112	3920	1225
2	40	128	5120	1600
3	38	130	4940	1444
4	44	138	6072	1936
5	67	158	10586	4489
6	64	162	10368	4096
7	59	140	8260	3481
8	69	175	12075	4761
9	25	125	3125	625
10	50	142	7100	2500
Total	491	1410	71566	26157

Required calculation
$\sum x = 491$ $\sum y = 1410$ $\sum xy = 71566$ $\sum x^2 = 26157$

Calculating the mean $(\bar{x}\ ,\ \bar{y})$;

$$\bar{x} = \frac{\sum x}{n} = \frac{491}{10} = 49.1\ ,\ \bar{y} = \frac{\sum y}{n} = \frac{1410}{10} = 141$$

Calculating the regression coefficient;

$$\beta_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \qquad \beta_1 = \frac{10 * 71566 - 491 * 1410}{10 * 26157 - (491)^2}$$

$$\beta_1 = \frac{715660 - 692310}{261570 - 241081} \qquad \beta_1 = \frac{23350}{20489} = 1.140$$

$$\beta_0 = \bar{y} - \beta_1\bar{x} \qquad \beta_0 = 141 - 1.140 * 49.1$$

$$\beta_0 = 141 - 55.974 \qquad \beta_0 = 85.026$$

Then substitute the regression coefficient into the regression model

$$\textbf{Estimated blood pressure } (\hat{Y}) = 85.026 + 1.140\ age$$

Interpretation of the equation;

Constant (intercept) value β_0 = 85.026 indicates that blood pressure at age zero.

Regression coefficient β_1 **=** 1.140 indicates that as age increase by one year the blood pressure increase by 1.140

THANK

YOU