CM 1110 Fundamentals of Mathematics and Statistics

Dr. Priyanga D. Talagala

2020-09-02

Contents

Course Syllabus

Pre-requisites

None

Learning Outcomes

On successful completion of this module, students will be able to apply fundamental concepts in Mathematics and Statistics for real world problem solving.

Outline Syllabus

Fundamentals of Mathematics

- Number Systems
- Sequences and Series
- Introduction to Logic
- Boolean Algebra
- Differentiation and Integration

Fundamentals of Statistics

- Descriptive Statistics
- Sets and Relations
- Random Variables
- Probability
- Correlation and Regression

Method of Assessment

- Mid-semester examination
- End-semester examination

Lecturer

Dr. Priyanga D. Talagala

Schedule

Lectures:

- Monday [1.15 pm - 4.30 pm]

Tutorial:

- Thursday [1.15 pm - 4.30 pm]

Consultation time:

- Tuesday [11.30 am to 12.30 pm]

Copyright Notice

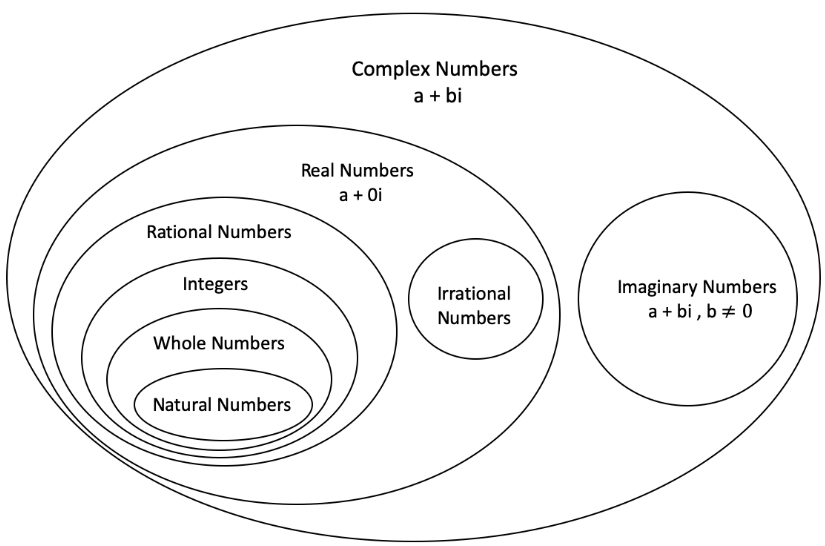
My lectures and course materials, including presentations, tests, exams, outlines, and similar materials, are protected by copyright. I am the exclusive owner of copyright in those materials I create. I encourage you to take notes and make copies of course materials for your own educational use. However, you may not, nor may you knowingly allow others to reproduce or distribute lecture notes and course materials publicly without my express written consent.

Chapter 1

Number Systems

Numbers can be classified according to how they are represented or according to the properties that they have.

1.1 Main types



1.1.1 Complex numbers

- Every number in number system is considered as a complex number

- A number of the form $a + ib$ is called a complex number when a and b are real numbers and $i = \sqrt{-1}$.
- For a given complex number, $a + ib$, 'a' is known as the real part and 'b' is known as the imaginary part.
- If $a = 0$, the number ib is said to be purely imaginary, if $b = 0$ the number a is real.
- A pair of complex number $a + ib$ and $a - ib$ are said to be conjugate of each other.

Show that the sum and product of a complex number and its conjugate complex are both real.

- Let $a + ib$ and $c + id$ be two complex numbers. Then

Addition. $(a + ib) + (c + id) = (a + c) + i(b + d)$

Subtraction $(a + ib) - (c + id) = (a - c) + i(b - d)$

Multiplication $(a + ib) \times (c + id) = ac - bd + i(ad + bc)$

Addition. $\frac{a+ib}{c+id} = \frac{a+ib}{c+id} . \frac{c-id}{c-id} = \frac{ac+bd}{c^2+d^2} + i\frac{bc-ad}{c^2+d^2}$

- Complex numbers are denoted by $\mathbb{C}$.

1.1.2 Imaginary numbers

- A number that does not exist in the number line is known as imaginary number.

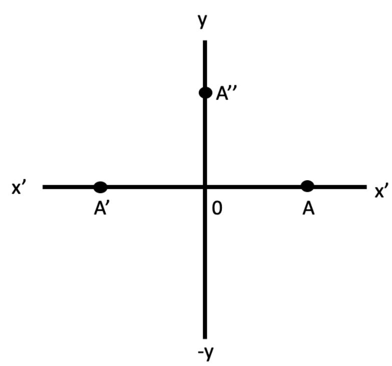
- For example, square root of negative numbers are imaginary numbers. It is denoted by i. i.e

$$\sqrt{-1} = i$$

$$i^2 = -1$$

- So there is no real number i that satisfies the above equation.
- The quantity, i is called the unit imaginary number.

Geometrical Representation of imaginary numbers



- Let OA be positive numbers which is represented by x and OA' by $-x$.
- And $-x = (i)^2x = i(ix)$ is on OX'.
- According to the above expression, the multiplication of the real number x by i twice amounts is equivalent to the rotation of OA through two right angles to reach OA'.
- Therefore, the multiplication of x by i is equivalent to the rotation of x through one right angle to reach OA''.
- Therefore, y-axis is known as imaginary axis.
- Multiplication by i rotates its direction through right angle.

1.1.3 Real numbers

- All numbers that can be represented on the number line are known as real numbers.
- The real numbers is the set of numbers containing all of the rational numbers and all of the irrational numbers.
- Real Numbers are denoted by $\mathbb{R}$.

1.1.4 Rational numbers

- Rational Numbers are denoted by $\mathbb{Q}$
- A rational number is defined as number of the form x/y where x and y are integers and $y \neq 0$.
- The set of rational numbers encloses the set of integers and fractions.
- The rational numbers that are not integral will have decimal values. These values can be of two types
  - Terminating decimal fractions (finite decimal factors): For example $1/5 = 0.5$, $13/5 = 2.6$.
  - Non Terminating decimal fractions. The non terminating decimal fractions having two types:
    i) Non terminating periodic fractions
    ii) Non terminating non periodic fractions

i) Non terminating periodic fractions

a. These are non terminating decimal fractions of the type $a.b1b2b3b4b5.....bmb1b2b3b4b5.....bm$

b. Examples

$19/6 = 3.16666666.....$

$18/7 = 2.57142857142857.......$

$21/9 = 2.3333.......$

ii) Non terminating non periodic fractions

a. These are non terminating and there is no periodic decimal places for that number.
b. i.e $a.b1b2b3b4b5.....bmc1c2.........$
c. for example 6.789542587436512……….

- **So from above terminating and non terminating periodic fraction numbers belongs to rational numbers.**

1.1.5 Irrational numbers

- Irrational numbers are denoted by $\mathbb{I}$
- Irrational numbers are consisted with **non terminating and non periodic fractions**.

- i.e irrational number is a number that cannot be written as a ratio x/y form (or fraction).
- In decimal form, it never ends or repeats.
- Examples for irrational numbers are $\sqrt{2} = 1.414213......$, $\pi = 3.14159265.......$, $\sqrt{3}$, $\sqrt{5}$ etc.

1.1.6 Integers

- All numbers that do not have the decimal places in them are called integers.
- $\mathbb{Z} = \{..., -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, ...\}$
- i.e it may be positive or negative or zero.
- Integers are denoted by $\mathbb{Z}$.
- Any integers are added, subtracted, or multiplied the result is always is an integer.
- When any integers multiplied , each of the multiplied integer is called a factor or divisor of the resulting product.

1.1.7 Whole numbers

- The set of whole numbers means narrator numbers and 0
- Whole numbers $= \mathbb{W} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, ...\}$

1.1.8 Natural numbers

- The counting numbers start with 1 and their end is not defined.
- i.e $\mathbb{N} = \{1, 2, 3, 4, ...\}$

Further Reading :

Dass, H. K. (2008). 'Complex Numbers', *Advanced Engineering Mathematics*. S. Chand Publishing. pp. 474-520.

1.2 Number representations

1.2.1 Glossary of terms used in the positional numeral systems

- There are various types of the number system in mathematics.
- The four most common number system types are:
  - Decimal number system (Base- 10)
  - Binary number system (Base- 2)
  - Octal number system (Base-8)
  - Hexadecimal number system (Base- 16)

System	Radix/ Base	Digits
Binary	2	0 1
Octal	8	0 1 2 3 4 5 6 7
Decimal	10	0 1 2 3 4 5 6 7 8 9
Hexadecimal	16	0 1 2 3 4 5 6 7 8 9 A B C D E F

1.2.2 Decimal Number System

- The Decimal number system is a number of base or radix equal to 10.
- To determine the actual number in each position, take the number that appears in the position and multiply it by 10^x, where x is the power representation.

Example 1: The value of the combination of symbols 453 is determined by adding the weights of each position as

$$4 \times 10^2 + 5 \times 10^1 + 3 \times 10^0$$

$$= 4 \times 100 + 5 \times 10 + 3 \times 1$$

Or

$$= 400 + 50 + 3 = 453$$

Example 2: The value of the combination of symbols 369.54 is determined by adding the weights of each position as

$$3 \times 10^2 + 6 \times 10^1 + 9 \times 10^0 + 5 \times 10^{-1} + 4 \times 10^{-2}$$

Or

$$= 300 + 60 + 9 + \frac{5}{10} + \frac{4}{100}$$

Or

$$= 300 + 60 + 9 + 0.5 + 0.04 = 369.54$$

1.2.3 Binary Number System

- The binary number system is a number system of base or radix equal to 2.
- In the binary number system, there are two symbols to represent number: 0 and 1
- When the symbols 0 an 1 are used to represent binary number, each symbol is called a binary digit or a bit.
- Therefore, the binary number 1011 is a four-digit number or a 4-bit binary number.

1.2.3.1 Binary-to-Decimal Conversion

- Multiply binary digit (1 or 0) in each position by the weight of the position and add the results.

Example 1: Convert the binary number 11010 to its decimal equivalent

Example 2: Convert the binary number 0.011 to its decimal equivalent

Example 3: Convert the binary number 110.011 to its decimal equivalent

1.2.3.2 Decimal-to-Binary Conversion

- To convert decimal numbers to their binary equivalent, the following procedures are employed:

1.2.3.2.1 Whole number conversion: Repeated division by 2

- The remainder resulting from each division forms binary number.
- The first remainder to be produced is called the least significant bit (LSB) and the last remainder is called most significant bit (MSB).

Example 1: Convert the decimal number 17 to binary

1.2.3.2.2 Fractional number conversion: Repeated multiplication by 2

- Multiply any factional part repeatedly by 2.
- The equivalent binary number is formed from the 1 or 0 in the units position (10^0 position).

Example 1: Convert the decimal number 0.625 to binary

- Sometimes it will be necessary to terminate the multiplication when an acceptable degree of accuracy is obtained. Then the resulted binary number will be an approximation.

Example 2: Convert the decimal number 0.6375 to binary

- We see that it continues in this way and does not terminate.
- So, $0.6375_{10} = (0.1010001101....)_2$.
- If we discard all the bits after the 7th bit, then we get the approximate representation $0.6375_{10} \approx (0.1010001)_2$, committing an error amounting to

$$0.6375_{10} - (0.1010001)_2 = 0.6375_{10} - 0.6328125_{10} = (.0046875)_{10},$$

which is known as ***round-off error***.

- Although, strictly, this is *chopping-off error*, it is generally, termed as *round-off error*.

Example 3: Convert the decimal number 49.683 to binary

1.2.4 Octal Number System

- The octal system is a number system of base or radix equal to 8.

1.2.4.1 Octal-to-Decimal Conversion

Example 1: Convert the following octal numbers to their decimal equivalent

(a) 35_8
(b) 100_8
(c) 0.24_8

1.2.4.2 Decimal-to-Octal Conversion

- To convert decimal numbers to their octal equivalent, the following procedures are employed:

1.2.4.2.1 Whole number conversion: Repeated division by 8

Example 1: Convert the decimal number 245 to octal equivalent

1.2.4.2.2 Fractional number conversion: Repeated multiplication by 8

Example 2: Convert the decimal fraction 0.432 to octal equivalent

- This conversion to octal is not precise as there is a remainder.
- If greater accuracy is required, continue multiplying by 8 to obtain more octal digits.

Example 3: Convert the decimal number 419.95 to octal equivalent

1.2.4.3 Octal-to-Binary Conversion

Octal and binary number correspondence

Octal	Binary
0	000
1	001
2	010
3	011
4	100
5	101
6	110
7	111

- To convert from octal to binary, simply replace each octal digit with the corresponding three-digit binary number.

Example 1: Convert the following octal numbers to their binary equivalent

(a) 247_8

(b) 124.375_8

1.2.4.4 Binary-to-Octal Conversion

- To convert from binary to octal, subdivide the number into groups of three bits, proceeding both left and right from the binary point, and if necessary padding the last group with zero. The octal representation of each group gives the required octal number.

Example 1: Convert the binary number 11010101.01101 to its octal equivalent

1.2.5 Hexadecimal Number System

- The Hexadecimal system is a number system of base or radix equal to 16.
- Hexadecimal is similar to the octal numeral system because each can be easily compared to the binary numeral system.
- Example values of hexadecimal numbers converted into binary, octal and decimal.

Hexadecimal	Binary	Octal	Decimal
0	0000	0	0
1	0001	1	1
2	0010	2	2
3	0011	3	3
4	0100	4	4
5	0101	5	5
6	0110	6	6
7	0111	7	7
8	1000	10	8
9	1001	11	9
A	1010	12	10
B	1011	13	11
C	1100	14	12
D	1101	15	13
E	1110	16	14
F	1111	17	15

1.2.5.1 Hexadecimal-to-Decimal Conversion

Example 1: Convert the following octal numbers to their decimal equivalent

(a) $C7_{16}$
(b) $2F_{16}$

1.2.5.2 Decimal-to-Hexadecimal Conversion

- To convert decimal numbers to their hexadecimal equivalent, the following procedures are employed:

1.2.5.2.1 Whole number conversion: Repeated division by 16

Example 1: Convert the decimal number 245 to Hexadecimal equivalent

1.2.5.2.2 Fractional number conversion: Repeated multiplication by 16

Example 2: Convert the decimal fraction 0.0738 to Hexadecimal equivalent

Example 3: Convert the decimal number 420.0095 to Hexadecimal equivalent

1.2.5.3 Hexadecimal-to-Binary Conversion

- To convert from Hexadecimal to binary, simply replace each Hexadecimal digit with the corresponding four-digit binary number.

Example 1: Convert the following Hexadecimal numbers to their binary equivalent

(a) $7B_{16}$
(b) $6D.8C_{16}$

1.2.5.4 Binary-to-Hexadecimal Conversion

- To convert from binary to Hexadecimal, subdivide the number into groups of four bits. The hexadecimal representation of each group then gives the hexadecimal representation.

Example 1: Convert the binary number 11010101.01101 to its Hexadecimal equivalent

1.2.5.5 Octal-to-Hexadecimal Conversion

- First convert octal to binary, and then convert binary to hexadecimal.

Example 1: Convert the octal number 365.52 to its Hexadecimal equivalent

1.2.5.6 Hexadecimal-to-octal conversion

- First convert Hexadecimal to binary, and then convert binary to octal.

Example 1: Convert the octal number D2B.284 to its Hexadecimal equivalent

1.2.6 Binary Arithmetic

The arithmetic operations: addition, subtraction, multiplication and division of binary numbers follow the rules as summarized below

Addition	Subtraction	Multiplication	Division
$0+0=0$	$0-0=0$	$0\times 0=0$	$\frac{0}{1}=0$
$0+1=1$	$1-0=1$	$0\times 1=0$	$\frac{1}{1}=1$
$1+0=1$	$1-1=0$	$1\times 0=0$	$\frac{0}{0}=undefined$
$1+1=10$	$1-0=10-1=1$	$1\times 1=1$	$\frac{1}{0}=undefined$

1.2.6.1 Binary Addition

Example 1: Add (a) 111 and 101 (b) 1010, 1001 and 1101

1.2.6.2 Binary Subtraction

Example 1: Perform the following subtractions

(a) $11 - 01$ (b) $11 - 10$ (c) $100 - 011$

- When subtracting a larger number from a smaller number, the output will be negative.

- To perform this subtraction, subtract the smaller number from the larger number and prefix the output with the sign of the larger number.

Example 2: Perform the following subtraction 101 - 111

1.2.6.3 Binary multiplication

Example 1: Multiply the following binary numbers

(a) 101×11 (b) 1101×10 (c) 1010×101
(b) 1011×1010

- Multiplication of factional numbers is perform in the same manner as with decimal factional numbers.

Example 2: Perform the binary multiplication 0.01×11

1.2.6.4 Binary Division

Example 1: Perform the following binary division

(a) $110 \div 11$ (b) $1100 \div 11$ (c) $1111 \div 110$ (d) $1100 \div 101$

References

Butenko, S., & Pardalos, P. M. (2014). Numerical methods and optimization: An introduction. CRC Press

Dass, H. K. (2008). ‘Complex Numbers’, Advanced Engineering Mathematics. S. Chand Publishing. pp. 474-520.

Chapter 2

Sequences and Series

2.1 Sequences

- A **Sequence** is a list of numbers written in a definite order:

$$a_1, a_2, a_3, a_4, ..., a_n$$

- An element of a sequence is called a term . The number a_1 is called the *first term*, a_2 is the *second term*, and in general a_n is called the n^{th} *term*.
- As can be seen in the above sequence, for every positive integer n there is a corresponding number a_n.Therefore a sequence can be defined as a function whose domain is the set of positive integers.
- But we usually write a_n instead of the function notation $f(n)$ for the value of the function at the number n.
- When dealing with infinite sequences, each term a_n will have a successor a_{n+1}
- **Notation:** The sequence $\{a_1, a_2, a_3, ...\}$ is also denoted by

$$\{a_n\}$$

or

$$\{a_n\}_{n=1}^{\infty}$$

- Some sequences can be defined by giving a formula for the nth term.
- In the following examples we give three descriptions of the sequences:

  i. by using the preceding notation
  ii. by using the defining formula
  iii. by writing out the terms of the sequence

- Notice that n doesn't have to start at 1.

a) $\{\frac{n}{n+1}\}_{n=1}^{\infty}$ $\quad a_n = \frac{n}{n+1}$ $\quad \{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, ..., \frac{n}{n+1}, ...\}$

b) $\{\frac{(-1)^n(n+1)}{3^n}\}_{n=1}^{\infty}$	$a_n = \frac{(-1)^n(n+1)}{3^n}$	$\{-\frac{2}{3}, \frac{3}{9}, -\frac{4}{27}, \frac{5}{81}, ..., \frac{(-1)^n(n+1)}{3^n}, ...\}$
c) $\{\sqrt{n-3}\}_{n=3}^{\infty}$	$a_n = \sqrt{n-3}, n \geq 3$	$\{0, 1, \sqrt{2}, \sqrt{3}, ..., \sqrt{n-3}, ...\}$
d) $\{cos\frac{n\pi}{6}\}_{n=0}^{\infty}$	$a_n = cos\frac{n\pi}{6}, n \geq 0$	$\{1, \frac{\sqrt{3}}{2}, \frac{1}{2}, 0, ..., cos\frac{n\pi}{6}, ...\}$

2.2 Series

- A series is the sum of a number of terms of a sequence.
- When writing series, the shorthand $\sum$ notation is used to represent the sum of a number of terms having a common form.
- The series $f(1) + f(2) + \cdots + f(n-1) + f(n)$ can be written as

$$\sum_{r=1}^{n} f(r).$$

2.2.1 Arithmetic sequences and series

- A sequence in which each term after the first term is obtained from the preceding term by adding a fixed number (Common difference), is called an arithmetic sequence or Arithmetic Progression.

- The sequence defined by

$$u_1 = a \text{ and } u_n = u_{n-1} + d \text{ for } n \geq 2$$

gives

$$a, a+d, a+2d, ...$$

- The nth term (i.e. the solution) is given by $u_n = a + (n-1)d$.

- This is the **arithmetic sequence** with first term a and common difference d

- The **arithmetic series** with n terms

$$a + (a+d) + (a+2d) + \cdots + [a + (n-1)d]$$

has sum

$$S_n = \frac{n}{2}(\text{first term} + \text{ last term})$$

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

2.2.2 Geometric sequences and series

- The sequence defined by

$$u_1 = a \text{ and } u_n = ru_{n-1} \text{ for } n \geq 2$$

gives

$$a, ar, ar^2, ...$$

- The nth term is given by $u_n = ar^{n-1}$.
- This is the **geometric sequence** with first term a and common ratio r.
- The **geometric series** with n terms

$$a + ar + ar^2 + \cdots + ar^{n-1}$$

has sum

$$S_n = \frac{a(1-r^n)}{1-r} \text{ or } \frac{a(r^n-1)}{r-1} \text{ for } r \neq 1$$

Important results relating to the $\sum$ notation

(1)

$$\begin{aligned}\sum_{r=1}^{n}\{f(r)+g(r)\} &= \{f(1)+g(1)\}+\{f(2)+g(2)\}+\cdots+\{f(n)+g(n)\}\\ &= \{f(1)+f(2)+\cdots+f(n)\}+\{g(1)+g(2)+\cdots+g(n)\}\\ &= \sum_{r=1}^{n} f(r)+\sum_{r=1}^{n} g(r)\end{aligned}$$

(2)

$$\sum_{r=1}^{n} af(r) = af(1)+af(2)+\cdots+af(n)$$

where a is a constant.

$$= a\{f(1)+f(2)+\cdots+f(n)\}$$

$$= a\sum_{r=1}^{n} f(r)$$

(3)

(a) $\sum_{r=1}^{n} r = \frac{n(n+1)}{2}$

(b) $\sum_{r=1}^{n} r^2 = \frac{n}{6}(n+1)(2n+1)$

(c) $\sum_{r=1}^{n} r^3 = \frac{n^2}{4}(n+1)^2$ or $\left[\frac{n(n+1)}{2}\right]^2$

(d) $\sum_{r=1}^{n} 1 = (1+1+\cdots+1) = n$

2.2.3 Methods of proof

1. Mathematical induction
2. The difference method

2.2.3.1 Mathematical induction

- Let n be a natural number. Then the aim is to show that some statement $P(n)$ involving n is true for any n.

- The following steps are used in Mathematical induction

  1. Let $P(n)$ be a statement
  2. Show that the statement is true for $P(1)$ and $P(2)$. (*i.e.* $P(n)$ is true for $n = 1$ and $n = 2$.)
  3. Assume that $P(k)$ is true (*i.e.* $P(n)$ is true for $n = k$).
  4. Show that $P(k+1)$ follows from $P(k)$.

2.2.3.2 The difference method

- The process of proof by induction is a powerful mathematical tool. However it has the disadvantage that, in order to employ the method it requires the answer.
- There are, however, direct methods of proof such as the method of differences, or the difference method.
- The difference method can be summarised as follows,

$$\sum_{r=1}^{n} \{f(r) - f(r-1)\} = f(n) - f(0)$$

where f is any function suitably defined on the non-negative integers.

- This is also know as the fundamental theorem of summation:

2.3 Infinite Sequences

- When dealing with infinite sequences, each term a_n will have a successor a_{n+1}

Definition: Convergence and Divergence

A sequence $\{a_n\}$ has the **limit** L and we write

$$\lim_{n\to\infty} a_n = L \text{ or } a_n \to L \text{ as } n \to \infty$$

if we can make the term a_n as close to L as n becomes sufficiently large.

If $\lim_{n\to\infty} a_n$ exists, we say the sequence **converges** (or is **convergent**).

If $\lim_{n\to\infty} a_n$ does not exist, we say the sequence **diverges** (or is **divergent**).

2.3.1 Limit Laws for Sequences

If $\{a_n\}$ and $\{b_n\}$ are convergent sequences and c is a constant, then

a) $\lim_{n\to\infty}(a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n$

b) $\lim_{n\to\infty}(a_n - b_n) = \lim_{n\to\infty} a_n - \lim_{n\to\infty} b_n$

c) $\lim_{n\to\infty} ca_n = c\ \lim_{n\to\infty} a_n$

d) $\lim_{n\to\infty} c = c$

e) $\lim_{n\to\infty}(a_n b_n) = \lim_{n\to\infty} a_n . \lim_{n\to\infty} b_n$

f) $\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{\lim_{n\to\infty} a_n}{\lim_{n\to\infty} b_n}$ if $\lim_{n\to\infty} b_n \neq 0$

g) $\lim_{n\to\infty}(a_n)^p = [\lim_{n\to\infty} a_n]^p$ if $p > 0$ and $a_n > 0$

Squeeze Theorem for Sequences

If $a_n \leq b_n \leq c_n$ for $n \geq n_0$ and $\lim_{n\to\infty} a_n = \lim_{n\to\infty} c_n = L$, then $\lim_{n\to\infty} b_n = L$

Theorem:

$$\text{If} \quad \lim_{n\to\infty} |a_n| = 0, \text{ then } \quad \lim_{n\to\infty} a_n = 0.$$

Theorem

The sequence $\{r^n\}$ is convergent if $-1 < r \leq 1$ and divergent for all other values of r.

$$\lim_{n\to\infty} r^n = \begin{cases} 0 & \text{if } -1 < r < 1 \\ 1 & \text{if } r = 1 \end{cases} \tag{2.1}$$

Definition

A sequence $\{a_n\}$ is called **increasing** if $a_n < a_{n+1}$ for all $n \geq 1$, that is, $a_1 < a_2 < a_3 < \dots$. It is called **decreasing** if $a_n > a_{n+1}$ for all $n \geq 1$. It is called **monotonic** if it is either increasing or decreasing.

Definition

- If there exists a number m such that $m \leq a_n$ for every n we say the sequence is **bounded below**. The number m is sometimes called a *lower bound* for the sequence.
- If there exists a number M such that $a_n \leq M$ for every n we say the sequence is **bounded above**. The number M is sometimes called an *upper bound* for the sequence.
- If the sequence is both bounded below and bounded above we call the sequence **bounded**.

2.4 Infinite Series

- Consider the infinite sequence $\{a_n\}_{n=1}^{\infty}$.
- Consider the *partial sums*

$$\begin{aligned} s_1 &= a_1 \\ s_2 &= a_1 + a_2 \\ s_3 &= a_1 + a_2 + a_3 \\ s_4 &= a_1 + a_2 + a_3 + a_4 \\ &\vdots \\ s_n = a_1 + a_2 + a_3 + a_4 + \cdots + a_n &= \sum_{i=1}^{n} a_i \end{aligned}$$

- These partial sums form a new sequence $\{s_n\}_{n=1}^{\infty}$, which may or may not have a limit.
- Consider the limit of the sequence of partial sums, $\{s_n\}_{n=1}^{\infty}$.

$$\lim_{n\to\infty} s_n = \lim_{n\to\infty} \sum_{i=1}^{n} a_i = \sum_{i=1}^{\infty} a_i$$

- This $\sum_{i=1}^{\infty} a_i$ $(= a_1 + a_2 + a_3 + a_4 + \cdots + a_n \, ...)$ is called an **infinite series**

Definition

Consider the series $\sum_{n=1}^{\infty} a_n == a_1 + a_2 + a_3 + \dots$. Let s_n denote its nth partial sum:

$$s_n = \sum_{i=1}^{n} a_i = a_1 + a_2 + \cdots + a_n$$

If the sequence $\{s_n\}$ is convergent and $\lim_{n\to\infty} s_n = s$ exist as a real number, then the series $\sum_{i=1}^{\infty} a_i$ is called **convergent** and we write,

$a_1 + a_2 + \cdots + a_n + \cdots = s$ or $\sum_{n=1}^{\infty} a_n = s$

The number s is called the **sum of the series**. If the sequence of partial sums is divergent then the infinite series is also called **divergent**.

Geometric Series

- An important example of an infinite series is the *geometric series*

$$\sum_{n=1}^{\infty} ar^{n-1} = a + ar + ar^2 + \dots$$

is convergent if $|r| < 1$ and its sum is

$$\sum_{n=1}^{\infty} ar^{n-1} = \frac{a}{1-r} \qquad |r| < 1$$

If $|r| \geq 1$, the geometric series is divergent.

Theorem

If the series $\sum_{n=1}^{\infty} a_n$ is convergent, then $\lim_{n\to\infty} a_n = 0$.

Proof

Note 1

- Consider any series $\sum a_n$.
- This associates two sequences
  - the sequence $\{s_n\}$ of its *partial sums*
  - the sequence $\{a_n\}$ of its *terms*.
- If $\sum a_n$ is convergent, then
  - the limit of the sequence $\{s_n\}$ is s (the sum of the series)

  – the limit of the sequence $\{a_n\}$ is 0.

Note 2 - The converse of the above theorem is not true in general.

The Test for Divergence

If $\lim_{n\to\infty} a_n$ does not exist or if $\lim_{n\to\infty} a_n \neq 0$, then the series $\sum_{n=1}^{\infty} a_n$ is divergent.

Theorem

If $\sum a_n$ and $\sum b_n$ are convergent series, then

a) $\sum_{n=1}^{\infty} ca_n = c\sum_{n=1}^{\infty} a_n \quad c$ is a constant.

b) $\sum_{n=1}^{\infty}(a_n + b_n) = \sum_{n=1}^{\infty} a_n + \sum_{n=1}^{\infty} b_n$

c) $\sum_{n=1}^{\infty}(a_n - b_n) = \sum_{n=1}^{\infty} a_n - \sum_{n=1}^{\infty} b_n$

Reading:

- Stewart, J., Clegg, D. K., & Watson, S. (2020). ‘Infinite Sequences and Series’, *Calculus: early transcendentals*. Cengage Learning.
- Acharjya, D. P. (2009). Fundamental approach to discrete mathematics. New Age International.

Tutorial

1. Find a formula for the general term a_n of the sequence, assuming that the pattern of the first few terms continues.

(a) $\left\{\frac{3}{5}, -\frac{4}{25}, \frac{5}{125}, -\frac{6}{625}, \frac{7}{3125}, \ldots\right\}$
(b) $\{1, 1, 2, 3, 5, 8, 13, 21, \ldots\}$
(c) $\left\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots\right\}$
(d) $\{2, 7, 12, 17, \ldots\}$
(e) $\left\{1, -\frac{2}{3}, \frac{4}{9}, -\frac{8}{27}, \ldots\right\}$

2. Show that

$$\sum_{r=1}^{n}(6r^2 + 4r - 1) = n(n+2)(2n+1)$$

3. Show by method of induction

(a) $\sum_{r=1}^{n} r = \frac{n(n+1)}{2}$
(b) $\sum_{r=1}^{n} r^2 = \frac{n}{6}(n+1)(2n+1)$
(c) $\sum_{r=1}^{n} r^3 = \left[\frac{n(n+1)}{2}\right]^2$
(d) $\sum_{r=1}^{n} \frac{1}{r(r+1)} = \frac{n}{n+1}$

4.

(a) By considering $n^3 - (n-1)^3$ and similar expressions, find the formula for $\sum_{r=1}^{n} r^2$ in terms of n, assuming the results for $\sum_{r=1}^{n} r$

(b) By considering $n^5 - (n-1)^5$ and similar expressions, find the formula for $\sum_{r=1}^{n} r^4$ in terms of n, assuming the results for $\sum_{r=1}^{n} r$, $\sum_{r=1}^{n} r^2$ and $\sum_{r=1}^{n} r^3$.

5. Determine whether the sequence $a_n = (-1)^n$ is convergent or divergent.

6. Determine whether the sequence converges or diverges. If it converges, find the limit

a) $a_n = n(n-1)$
b) $a_n = \frac{n}{n+1}$
c) $a_n = \frac{(-1)^n}{n}$
d) $a_n = \frac{4n^2+2}{n^2+n}$
e) $a_n = \frac{2^n}{3^{n+1}}$

f) $a_n = \frac{(-1)^{n-1}n}{n^2+1}$

7. Discuss the convergence of the sequence $a_n = n!/n^n$, where $n! = 1 \times 2 \times \cdots \times n$.

8. Determine if the following sequences are monotonic and/ or bounded

a) $\{-n^2\}_{n=0}^{\infty}$
b) $\{(-1)^{n+1}\}_{n=1}^{\infty}$
c) $\left\{\frac{2}{n^2}\right\}_{n=5}^{\infty}$

9. Determine if the following series converges or diverges. If it converges determine its value.

a) $\sum_{n=1}^{\infty} n$
b) $\sum_{n=1}^{\infty} (-1)^n$
c) $\sum_{n=1}^{\infty} \frac{1}{3^{n-1}}$
d) $5 - \frac{10}{3} + \frac{20}{9} - \frac{40}{27} + ...$
e) $\sum_{n=1}^{\infty} 2^{2n}3^{1-n}$
f) $\sum_{n=1}^{\infty} x^n$, where $|x| < 1$
g) $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$

10. Show that harmonic series

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + ...$$

is divergent.

11. Show that the series $\sum_{n=1}^{\infty} \frac{n^2}{5n^2+4}$ diverges.

12. Determine whether the series is convergent or divergent. If it is convergent, find its sum.

a) $\sum_{n=1}^{\infty} \left(\frac{3}{n(n+1)+} + \frac{1}{2^n}\right)$
b) $\sum_{n=1}^{\infty} \left(\frac{3}{n(n+3)+} + \frac{5}{4^n}\right)$
c) $\sum_{n=1}^{\infty} \frac{1}{n^2+3n+2}$
d) $\sum_{n=1}^{\infty} \frac{1}{n^2+4n+3}$
e) $\sum_{n=1}^{\infty} 9^{-n+2}4^{n+1}$
f) $\sum_{n=1}^{\infty} \frac{(-4)^{3n}}{5^{n-1}}$
g) $\sum_{n=1}^{\infty} \left(\frac{1}{n^2+4n+3} - 9^{-n+2}4^{n+1}\right)$

Chapter 3

Introduction to Logic

3.1 Statement (Proposition)

- A statement is a declarative sentence which is either true or false **but not both**
- A proposition is a collection of declarative statements that has either a truth value "true" or a truth value "false".
- The truth value True and False are denoted by the symbols **T** and **F**, respectively.
- Sometimes it is also denoted by **1** (stands for true) and **0** (stands for false).
- Since it depends on only two possible truth values, it is known as *two-values logic* of *bi-values logic*.

Example 1

Consider the following sentences and select the statements

(a) Man is Mortal
(b) 12 plus 9 is equal to 1
(c) Moon rises in the east
(d) 3 is less than four
(e) Please sit down
(f) y is a cat
(g) 14 is a composite number
(h) *Chaddy* is a nice dog.

3.2 Propositional Variables

- Every statement in propositional logic consists of **propositional variables** combined via **propositional connectives**.
- Propositional variables are usually denoted by capital English letters, such as P, Q, R etc.
- Each variable can take one of two values: **T** or **F**.

3.3 Logical Connectives

- We use **logical connectives** to connect the propositional variables (several statements into a single statement).
- The most basic and fundamental connectives are Negation, Conjunction and Disjunction.

3.3.1 Negation

- Negation is the simplest common example of a truth-functional operation.
- The negation of a statement is also a statement.
- We use the connective **Not** for negation.
- If P be a statement, then the negation of P is denoted by $\neg P$.
- P and $\neg P$ has opposite truth values.
- The relationship between the truth values of P and $\neg P$ can be made explicit by a diagram called a **truth table**

Truth Table (Negation)

P	$\neg P$
T	F
F	T

3.3.2 Conjunction

- The conjunction of two statements, P and Q is also a statement.
- We use the connective **And** ($\wedge$) for conjunction.
- It is denoted by $P \wedge Q$

Truth Table (Conjunction)

P	Q	$(P \wedge Q)$
T	T	T
T	F	F
F	T	F
F	F	F

Rule: $(P \wedge Q)$ is true if both P and Q are true, otherwise false.

- In order to make $(P \wedge Q)$ true, P and Q have to be **simultaneously** true

3.3.3 Disjunction

- The disjunction of two statements, P and Q is also a statement.
- We use the connective **Or** $(\vee)$ for disjunction
- It is denoted by $P \vee Q$

Truth Table (Disjunction)

P	Q	$(P \vee Q)$
T	T	T
T	F	T
F	T	T
F	F	F

Rule: $(P \vee Q)$ is true if either P or Q is true and it is false when both P and Q are false.

- $(P \vee Q)$ shall stand for "P or Q or both".

3.4 Conditional

- In mathematics, expressions of the form "IF P then Q" occur so often.
- This can be expressed in any one of the following forms.

  a) If P, then Q
  b) P only if Q
  c) P implies Q
  d) Q if P

- Therefore, it is important to understand the corresponding truth-functional operation.

- Let P and Q be any two statements.
- Then the statement $P \to Q$ is called a conditional statement.
- In an implication $P \to Q$, P is called the *antecedent* (hypothesis) and Q is called the *consequent* (conclusion).

Truth Table (Conditional)

P	Q	$(P \to Q)$
T	T	T
T	F	F
F	T	T
F	F	T

Rule: An implication (conditional) $(P \to Q)$ is False only when the hypothesis (P) is true and conclusion (Q) is false, otherwise True.

3.5 Bi-Conditional

- Let P and Q be any two statements.
- Then the statement $P \leftrightarrow Q$ is called a bi-conditional statement.
- This can be expressed in any one of the following forms.

  a) P if and only if Q
  b) P is necessary and sufficient of Q
  c) P is necessary and sufficient for Q
  d) P implies Q and P is implied by Q

- The bi-conditional (double implication) $P \leftrightarrow Q$ can be defined as

$$P \leftrightarrow Q : (P \to Q) \wedge (Q \to P)$$

Truth Table (Bi-Conditional)

P	Q	$(P \rightarrow Q)$	$(Q \rightarrow P)$	$(P \leftrightarrow Q)$
T	T	T	T	T
T	F	F	T	F
F	T	T	F	F
F	F	T	T	T

Rule: $(P \leftrightarrow Q)$ is True only when both P and Q have identical truth values, otherwise false.

3.6 Converse

- Let P and Q be any two statements.
- The converse statement of the conditional $P \rightarrow Q$ is given as $Q \rightarrow P$

3.7 Inverse

- Let P and Q be any two statements.
- The inverse statement of the conditional $(P \rightarrow Q)$ is given as $(\neg P \rightarrow \neg Q)$

3.8 Contra Positive

- Let P and Q be any two statements.
- The contra positive statement of the conditional $(P \to Q)$ is given as $(\neg Q \to \neg P)$

Truth Table (Contra Positive)

P	Q	$P \to Q$	$\neg Q$	$\neg P$	$(\neg Q \to \neg P)$
T	T	T	F	F	T
T	F	F	T	F	F

P	Q	$P \to Q$	$\neg Q$	$\neg P$	$(\neg Q \to \neg P)$
F	T	T	F	T	T
F	F	T	T	T	T

- From the truth table it can be seen that both conditional $P \to Q$ and contra positive $(\neg Q \to \neg P)$ have same truth values.

3.9 Exclusive OR

- Let P and Q be any two statements.
- The exclusive OR of two statements P and Q is denoted by $(P \overline{\vee} Q)$.
- We use the connective XOR for exclusive OR
- The exclusive OR $(P \overline{\vee} Q)$ is true if either P or Q is True but not both.
- The exclusive OR is also known as exclusive disjunction.

the conditional $(P \to Q)$ is given as $(\neg Q \to \neg P)$

Truth Table (Exclusive OR)

P	Q	$(P \overline{\vee} Q)$
T	T	F
T	F	T
F	T	T
F	F	F

Rule: $(P \overline{\vee} Q)$ is true if either P or Q is True but not both, otherwise false.

3.10 NAND

- The word NAND stands for NOT and AND.
- The connective NAND is denoted by the symbol $\uparrow$.
- Let P and Q be any two statements.
- The NAND of P and Q is denoted by $(P \uparrow Q)$.
- The NAND, $(P \uparrow Q)$ can be defined as

$$(P \uparrow Q) \equiv \neg(P \wedge Q)$$

Truth Table (NAND)

P	Q	$(P \uparrow Q)$
T	T	F
T	F	T
F	T	T
F	F	T

Rule: $(P \uparrow Q)$ is True if *either* P or Q is false, otherwise False

3.11 NOR

- The word NOR stands for NOT and OR.
- The connective NOR is denoted by the symbol $\downarrow$.
- Let P and Q be any two statements.
- The NOR of P and Q is denoted by $(P \downarrow Q)$.
- The NOR, $(P \downarrow Q)$ can be defined as

$$(P \downarrow Q) \equiv \neg(P \vee Q)$$

Truth Table (NOR)

P	Q	$(P \downarrow Q)$
T	T	F
T	F	F
F	T	F
F	F	T

Rule: $(P \downarrow Q)$ is True only when both P or Q are false, otherwise False

3.12 Tautology

- If the truth values of a composite statement are always true, irrespective of the truth values of the atomic (individual) statements, then it is a **tautology**.

Example

- The composite statement $(P \wedge (P \to Q)) \to Q$ is a tautology.

- Verify this using a truth table with composite statement as $(P \wedge (P \rightarrow Q)) \rightarrow Q$

3.13 Contradiction

- If the truth values of a composite statement are always false, irrespective of the truth values of the atomic (individual) statements, then it is a **contradiction** or **unsatisfiable**.

Example

- The composite statement $\neg(P \rightarrow (Q \rightarrow (P \wedge Q)))$ is a contradiction.
- Verify this using a truth table with composite statement as $\neg(P \rightarrow (Q \rightarrow (P \wedge Q)))$

3.14 Satisfiable

- If the truth values of a composite statement are sometimes true and sometimes false, irrespective of the truth values of the atomic statements, then it is called a satisfiable.

Example

- The composite statement $(P \rightarrow Q) \rightarrow (Q \rightarrow P)$ is satisfiable.
- Verify this using a truth table of $(P \rightarrow Q) \rightarrow (Q \rightarrow P)$

3.15 Duality Law

- Two formulae P and $p^*$are said to be duals of each other if either one can be obtained from the other by interchanging $\wedge$ by $\vee$ and $\vee$ by $\wedge$.
- The two connectives $\vee$ and $\wedge$ are called dual to each other.

Example

The formulae $P \equiv (P \vee Q) \wedge R$ and $P^* \equiv (P \wedge Q) \vee R$ are dual to each other.

3.16 Algebra of Propositions

- If P, Q and R be three statements, then the following laws hold.

Algebra of Propositions	
Commutative Laws	$P \wedge Q \equiv Q \wedge P$ and $P \vee Q \equiv Q \vee P$
Associative Laws	$P \wedge (Q \wedge R) \equiv (P \wedge Q) \wedge R$ and $P \vee (Q \vee R) \equiv (P \vee Q) \vee R$
Distributive Laws	$P \wedge (Q \vee R) \equiv (P \wedge Q) \vee (P \wedge R)$ and $P \vee (Q \wedge R) \equiv (P \vee Q) \wedge (P \vee R)$
Idempotent Laws	$P \wedge P \equiv P$ and $P \vee P \equiv P$
Absorption Laws	$P \vee (P \wedge Q) \equiv P$ and $P \wedge (P \vee Q) \equiv P$
De Morgan's Laws	$\neg(P \wedge Q) \Leftrightarrow (\neg P) \vee (\neg Q)$ and $\neg(P \vee Q) \Leftrightarrow (\neg P) \wedge (\neg Q)$

3.17 Logical Implication and Equivalence

3.18 Necessary and Sufficient Conditions

3.18.1 Necessary Condition

3.18.2 Sufficient Condition

3.18.3 Necessary and Sufficient Condition

References

Acharjya, D. P. (2009). Fundamental approach to discrete mathematics. New Age International.

Mendelson, E. (1970). Boolean algebra and switching circuits. McGraw-Hill Edition 2004.

Tutorial

1. Reduce the following sentences to statement forms

(a) A necessary condition for x to be prime is that x is odd or $x = 2$.

(b) A sufficient condition for f to be continuous is that f is differentiable.

(c) A necessary and sufficient condition for Mr. Perera to be elected is that Mr. Perera wins 75 votes.

(d) Calathea plant will stay healthy only if enough moisture is available.

(e) It is raining but the sun is still shining.

(f) She will die today unless medical aid is obtained.

(g) If taxes are increased or government spending decreases, then inflation will not occur this year.

2. Show that the composite statement $(P \wedge (P \to Q)) \to Q$ is a tautology using a truth table

3. Show that the composite statement $\neg(P \to (Q \to (P \wedge Q)))$ is a contradiction using a truth table

4. Show that $(A \to B) \to A$ logically implies A

5. Show that $(A \leftrightarrow B)$ is logically equivalent to $(A \to B) \wedge (B \to A)$

6. Find the negation of $A \to B$

7. Construct the truth table for $A \to (B \leftrightarrow A \wedge B)$.

8. Find the negation of the following statement “He is rich but unhappy”

9. Prove by constructing truth table

$$P \to (Q \vee R) \equiv (P \to Q) \vee (P \to R)$$

10. Find the negation of $P \leftrightarrow Q$.
11. With the help of the truth table prove that $\neg(P \wedge Q) \equiv \neg P \vee \neg Q$
12. Show that $(P \to Q) \leftrightarrow (\neg P \vee Q)$ is a tautology

13. Show that the following stamens are equivalent.

- *Statement 1:* Good food is not cheap
- *Statement 2:* Cheap food is not good

14. Express $P \rightarrow Q$ only using $\downarrow$ and $\uparrow$.
15. Prove that $(P \wedge Q) \wedge \neg(P \vee Q)$ is a contradiction.
16. Express $P \leftrightarrow Q$ only using $\downarrow$ and $\uparrow$.
17. Show by truth table the following statement are equivalent

- *Statement 1:* Rich men are unhappy
- *Statement 2:* Men are unhappy or poor

18. A constructor promises a client "We will fix it on Monday if it is not raining". When the constructor would be deemed to have broken his promise. Explain with the help of truth table.

19. Eliminate as many parentheses as possible from

a) $\{[(A \vee B) \rightarrow (\neg C)] \vee [(((\neg B) \wedge C) \wedge B)]\}$
b) $\{[A \wedge (\neg(\neg B))] \leftrightarrow [B \leftrightarrow (C \vee B)]\}$
c) $[(B \leftrightarrow (C \vee B)) \leftrightarrow (A \wedge (\neg(\neg B)))]$

20 Show the following are tautologies without using truth tables.

a) $((A \rightarrow B) \rightarrow C) \rightarrow ((C \rightarrow A) \rightarrow (D \rightarrow A))$.
b) $(A \rightarrow B) \rightarrow ((B \rightarrow C) \rightarrow (A \rightarrow C))$.

Chapter 4

Boolean Algebra

- George Boole approached logic in a new way reducing it to a simple algebra.
- He introduced symbolic logic known as *Boolean Algebra, Boolean function, Boolean expression, Boolean ring* etc.
- Each variable in Boolean Algebra has either of two values: true or false.
- The purpose of Boolean Algebra is to solve logic problems.
- C.E Shannon observed that Boolean Algebra could be used to analyze electronic circuits.

4.1 Gates

- In Chapter 3 we discussed about logic connectives $\neg$, $\wedge$ and $\vee$.
- The connectives $\wedge$ and $\vee$ can be considered as circuits connected in series and parallel, respectively.
- A circuit with one or more input signals but only one output signal is known as a gate.
- Gates are digital circuits because of input and output signals, which are either low or high.
- Gates are also known as logical circuits as they can be analyzed with Boolean Algebra.
- In gates, the connectives $\neg$, $\wedge$ and $\vee$ are usually denoted by the symbols $'$, $.$ and $+$, respectively.

4.1.1 NOT gate

A NOT gate receives input x, where x is a bit (binary digit) and produces output x' where

$$x' = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x = 1 \end{cases} \tag{4.1}$$

- x' is called the *complement* of x.
- 0 is called the *zero* element.
- 1 is called the *unit* element.
- The output state is always the opposite of the input state.
- The output is also known as the complement of the input.
- The block diagram and the logic table for the basic NOT gate:

4.1.2 AND gate

- An AND gate receives input x_1 and x_2, where x_1 and x_2 are bits, and produces output $(x_1 \wedge x_2)$ where

$$(x_1 \wedge x_2) = \begin{cases} 1 & \text{if } x_1 = x_2 = 1 \\ 0 & \text{Otherwise} \end{cases} \tag{4.2}$$

- $(x_1 \wedge x_2)$ is called the *meet* of x_1 and x_2.

- An AND gate can have more than two inputs, but only one output.
- The block diagram and the logic table for the basic AND gate:

4.1.3 OR gate

- An OR gate receives input x_1 and x_2, where x_1 and x_2 are bits, and produces output $(x_1 \vee x_2)$ where

$$(x_1 \vee x_2) = \begin{cases} 1 & \text{if } x_1 \text{ or } x_2 = 1 \\ 0 & \text{Otherwise} \end{cases} \tag{4.3}$$

- $(x_1 \vee x_2)$ is called the *join* of x_1 and x_2.
- An OR gate can have more than two inputs, but only one output.
- The block diagram and the logic table for the basic OR gate:

4.1.4 More logic gates

- There are some other types of gates that are widely used in Computer Science such as NAND, NOR, XOR, and XNOR gates

4.1.4.1 NOR gate

- A NOR gate receives input x_1 and x_2, where x_1 and x_2 are bits, and produces output $(x_1 \vee x_2)'$ where

$$(x_1 \vee x_2)' = \begin{cases} 1 & \text{if } x_1 = x_2 = 0 \\ 0 & \text{Otherwise} \end{cases} \tag{4.4}$$

- A NOR gate can have more than two inputs, but only one output.
- The block diagram and the logic table for the basic NOR gate:

4.1.4.2 NAND gate

- A NAND gate receives input x_1 and x_2, where x_1 and x_2 are bits, and produces output $(x_1 \wedge x_2)'$ where

$$(x_1 \wedge x_2)' = \begin{cases} 1 & \text{if } x_1 = 0 \text{ or } x_2 = 0 \\ 0 & \text{Otherwise} \end{cases} \tag{4.5}$$

- A NAND gate can have more than two inputs, but only one output.
- The block diagram and the logic table for the basic NAND gate:

4.1.4.3 XOR gate (Exclusive OR gate)

- A XOR gate receives input x_1 and x_2, where x_1 and x_2 are bits, and produces output $(x_1 \overline{\vee} x_2)$ or $(x_1 \oplus x_2)$, where

$$(x_1 \oplus x_2)' = \begin{cases} 1 & \text{if } x_1 = 1 \text{ or } x_2 = 1 \text{ but not both} \\ 0 & \text{Otherwise} \end{cases} \tag{4.6}$$

- Rule: XOR gate produces 1 that have an odd number of 1's.
- A XOR gate can have more than two inputs, but only one output.
- The block diagram and the logic table for the basic XOR gate:

4.1.4.4 XNOR gate (Exclusive NOR gate)

- A XNOR gate receives input x_1 and x_2, where x_1 and x_2 are bits, and produces output (x_1 XNOR x_2), where

$$(x_1 \text{ XNOR } x_2)' = \begin{cases} 1 & \text{if } x_1 \text{ and } x_2 \text{ are same bits} \\ 0 & \text{Otherwise} \end{cases} \tag{4.7}$$

- A XNOR gate can have more than two inputs, but only one output.
- It can recognize *even-parity* words *i.e* a word which has an even number of 1's.
- Example: 11001111 is even-parity as it contains six 1's, 1110 is an odd-parity as it has an odd number of 1's.
- The block diagram and the logic table for the basic XNOR gate:

4.2 Combinatorial Circuit

- A combinatorial circuit produces a unique output for every combination of inputs.
- A combinatorial circuit has no memory, previous inputs and the state of the system do not affect the output of a combinatorial circuit.
- These circuits can be constructed using gates which we have already discussed.

4.3 Boolean Expression

- Any expression built up from the variables $x_1, y_1, z_1, x_2, y_2, z_2, ...$ by applying the operations $\wedge$, $\vee$ and $'$ a finite number of times is called a Boolean expression.
- If X and Y are two Boolean expressions, then X', Y', $(X \wedge Y)$ and $(X \vee Y)$ are also Boolean expressions.
- The output of a combinatorial circuit is also a Boolean expression.

Example

4.3.1 Theorem

- If $\wedge$, $\vee$ and $'$ are connectives introduced earlier, then the following properties hold.

(i) Associative Law: For all $a, b, c \in \{0, 1\}$

$$(a \wedge b) \wedge c = a \wedge (b \wedge c) \text{ and } (a \vee b) \vee c = a \vee (b \vee c)$$

(ii) Identity Law: For all $a \in \{0, 1\}$

$$(a \wedge 1) = a \text{ and } (a \vee 0) = a$$

(iii) Commutative Law: For all $a, b \in \{0, 1\}$

$$(a \wedge b) = (b \wedge a) \text{ and } (a \vee b) = (b \vee a)$$

(iv) Complement Law: For all $a \in \{0, 1\}$

$$(a \wedge a') = 0 \text{ and } (a \vee a') = 1$$

(v) Distributive Law: For all $a, b, c \in \{0, 1\}$

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c) \text{ and } a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$$

(vi) De-Morgan's Law: If x_1 and x_2 are bits, *i.e.* $x_1, x_2 \in \{0, 1\}$, then

$$(x_1 \wedge x_2)' = x_1' \vee x_2' \text{ and } (x_1 \vee x_2)' = x_1' \wedge x_2'$$

4.4 Equivalent Combinatorial Circuits

- Two combinatorial circuits, each having inputs $x_1, x_2, \dots, x_n$ are said to be equivalent if they produce the same output for same inputs.

4.5 Boolean Algebra

- A *Boolean algebra* consists of a set S together with two binary operations $\wedge$ and $\vee$ on S, a singular operation $'$ on S and two specific elements 0 and 1 of S such that the following laws hold.

- A Boolean algebra will be designated by a hextuple $B = \langle S, \wedge, \vee, ', 0, 1\rangle$

- Sometimes one refers to the set S as a Boolean algebra, but this is just a loose misuse of language.

(a) Associative Law: For all $a, b, c \in S$

$$(a \wedge b) \wedge c = a \wedge (b \wedge c) \text{ and } (a \vee b) \vee c = a \vee (b \vee c)$$

(b) Identity Law: For all $a \in S$

$$(a \wedge 1) = a \text{ and } (a \vee 0) = a$$

(c) Commutative Law: For all $a, b \in S$

$$(a \wedge b) = (b \wedge a) \text{ and } (a \vee b) = (b \vee a)$$

(d) Complement Law: For all $a \in S$

$$(a \wedge a') = 0 \text{ and } (a \vee a') = 1$$

(e) Distributive Law: For all $a, b, c \in S$

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c) \text{ and } a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$$

4.5.1 Theorem

In a Boolean algebra: if $(a \vee b) = 1$ and $(a \wedge b) = 0$, then $b = a'$, *i.e.* the complement is unique

4.5.2 Theorem

In a Boolean algebra $B = \langle S, \wedge, \vee, ', 0, 1\rangle$; the following properties hold.

(a) Idempotent Law: For all $x \in S$

$$(x \vee x) = x \text{ and } (x \wedge x) = x$$

(b) Bound Law: For all $x \in S$

$$(x \vee 1) = 1 \text{ and } (x \wedge 0) = 0$$

(c) Absorption Law: For all $x, y \in S$

$$x \wedge (x \vee y) = x \text{ and } x \vee (x \wedge y) = x$$

(d) Involution Law: For all $x \in S$

$$(x')' = x$$

(e) 0 and 1 Law: $0' = 1$ and $1' = 0$
(f) De-Morgan's Law: For all $x, y \in S$

$$(x \wedge y)' = x' \vee y' \text{ and } (x \vee y)' = x' \wedge y'$$

4.6 Dual of a Statement

- The dual of a statement involving Boolean expressions is obtained by replacing 0 by 1, 1 by 0, $\wedge$ by $\vee$, and $\vee$ by $\wedge$.
- Two Boolean expressions are said to be dual of each other if one expression is obtained from other by replacing 0 by 1, 1 by 0, $\wedge$ by $\vee$, and $\vee$ by $\wedge$.
- In Boolean Algebra, the dual of a theorem is also a theorem.

Example

What the dual of the statement: $(x \wedge y)' = x' \vee y'$

4.7 Boolean Function

Let $B = \langle S, \wedge, \vee, ', 0, 1\rangle$ be a Boolean algebra and let $X(x_1, x_2, x_3, \dots, x_n)$ be a Boolean expression in n variables.

A function $f : B^n \to B$ is called a Boolean function if f is of the form

$f(x_1, x_2, x_3, \dots, x_n) = X(x_1, x_2, x_3, \dots, x_n)$

Example

Consider the Boolean function $f : B^3 \to B$; $B = \{0, 1\}$ defined by

$$f(x_!, x_2, x_3) = x_1 \wedge (x_2 \vee \bar{x_3})$$

4.7.1 Representation of Boolean Functions

- There are several ways of representing Boolean functions.

  (a) **Tabular representation**
  (b) n Space representation
  (c) Cube representation

Tabular representation

- A Boolean function is completely determined by its evaluation over any Boolean algebra.
- In tabular representation, we consider a row R of the table where the output is 1.
- We then form the combination $(x_1 \wedge x_2 \wedge x_3 \wedge ... \wedge x_n)$ and place a bar over each x_i whose value is 0 in row R.
- The combination formed is 1 if and only if x_i have the value given in row R.
- We thus join (OR) the terms to obtain the Boolean expression.

Example

x_1	x_2	x_3	$f(x_1, x_2, x_3)$
1	1	1	1
1	1	0	0
1	0	1	1
0	1	1	0
1	0	0	0
0	1	0	1
0	0	1	0
0	0	0	0

4.8 Various Normal Forms

4.8.1 Disjunctive normal form

- A Boolean function $f : B^n \to B$ which consists of a sum of elementary products is called the **disjunctive normal form** of the given function f
- Let $f : B^n \to B$ is a Boolean function.
- If f is not identically zero, let $A_1, A_2, A_3, \dots, A_k$ denote the elements A_i of B_2^n, for which $f(A_i) = 1$, where

$$A_i = (a_1, a_2, \dots, a_n).$$

- For each A_i set $m_i = (y_1 \wedge y_2 \wedge y_3 \wedge \dots \wedge y_n)$, where

$$y_i = \begin{cases} x_i & \text{if } a_i = 1 \\ x_i' & \text{if } a_i = 0. \end{cases} \tag{4.8}$$

- Then $f(x_1, x_2, x_3, \dots, x_n) = m_1 \vee m_2 \vee m_3 \vee \dots \vee m_k$.
- This representation of a Boolean function is called the **disjunctive normal form**

Example

Consider the Boolean function $(x_1 \oplus x_2)$. The truth table for this function is given below.

x_1	x_2	$(x_1 \oplus x_2)$
1	1	0
1	0	1
0	1	1
0	0	0

Write the disjunctive form of this function

4.8.2 Conjunctive normal form

- A Boolean function $f : B^n \to B$ which consists of a product of elementary sums is called the **conjunctive normal form** of the given function f
- Let $f : B^n \to B$ is a Boolean function.
- If f is not identically one, let $A_1, A_2, A_3, \dots, A_k$ denote the elements A_i of B_2^n, for which $f(A_i) = 0$, where

$$A_i = (a_1, a_2, \dots, a_n).$$

- For each A_i set $M_i = (y_1 \vee y_2 \vee y_3 \vee \dots \vee y_n)$, where

$$y_i = \begin{cases} x_i & \text{if } a_i = 0 \\ x_i' & \text{if } a_i = 1. \end{cases} \tag{4.9}$$

- Then $f(x_1, x_2, x_3, \dots, x_n) = M_1 \wedge M_2 \wedge M_3 \wedge \dots \wedge M_k$.
- This representation of a Boolean function is called the **conjunctive normal form**

Example

Consider the Boolean function $(x_1 \oplus x_2)$. The truth table for this function is given below.

x_1	x_2	$(x_1 \oplus x_2)$
1	1	0
1	0	1
0	1	1
0	0	0

Write the conjunctive form of this function

- A term of the form $(y_1 \wedge y_2 \wedge y3 \wedge \dots \wedge y_n)$, where each y_i is either x_i or $\bar{x}_i$ is called a *minterm*.

- A term of the form $(y_1 \vee y_2 \vee y3 \vee ... \vee y_n)$, where each y_i is either x_i or $\bar{x}_i$ is called a *maxterm*

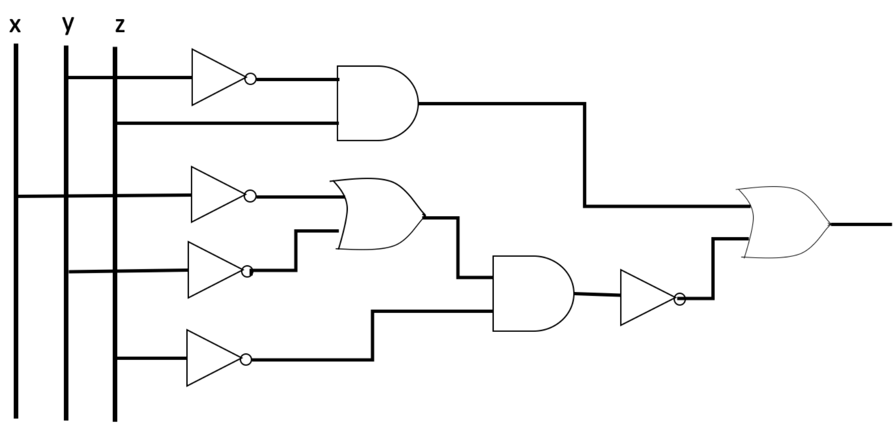
References

Acharjya, D. P. (2009). Fundamental approach to discrete mathematics. New Age International.

Mendelson, E. (1970). Boolean algebra and switching circuits. McGraw-Hill Edition 2004.

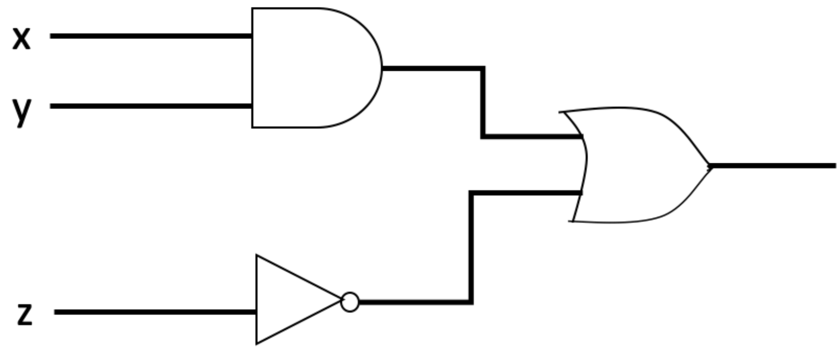
Tutorial

1. Construct an AND gate using three NOR gates
2. Construct an OR gate using three NAND gates
3. Draw a gating network to the statement $(x.y) + (y.z) + (z.x)$
4. Draw a gating network to the statement $(x + y)'(z.u) + (x.y)'(z + u)$
5. What is the output of the following gating network



6. Construct a gating network using inverter and OR gate to the statement $(x.y) + (y.z) + (z.x)$
7. Find the value of the Boolean expression given below for $x = 1$, $y = 1$ and $z = 0$.
$$(x \wedge (y \vee (x \wedge y'))) \vee ((x \wedge y') \vee (x \wedge z')')$$
8. Construct an AND gate using inverters and three NOR gates.

9. Write the Boolean expression that represents the following combinatorial circuit, construct the logic table with the output of each gate.



10. Show that $y = z$ when $(x + y) = (x + z)$ and $(x' + y) = (x' + z)$.

11. Given the Boolean function f, write f in its disjunctive normal form.

x	y	z	$f(x, y, z)$
1	1	1	1
1	1	0	1
1	0	1	0
1	0	0	0
0	1	1	0
0	1	0	1
0	0	1	0
0	0	0	1

12. Draw the logic circuit (Combinatorial circuit) with input x, y, z and output Y to the following Boolean expressions.

a) $Y = x'yz + x'yz' + xyz'$
b) $Y = xy'z + xz' + y'z$

13. Show that the combinatorial circuits (a) and (b) are equivalent.

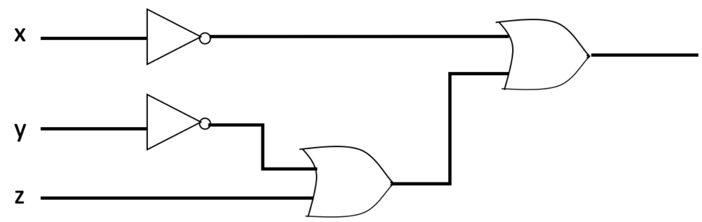


Figure 4.1: (a)

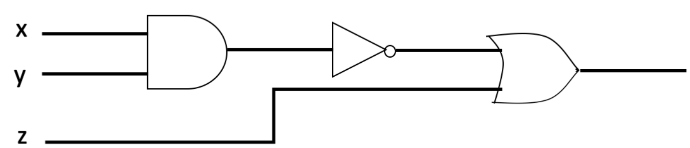


Figure 4.2: (b)

14. Reduce the following Boolean products to either 0 or a fundamental product

(a) $x.y.x'.z$
(b) $x.y.z'.y.x$

15. Given the Boolean function f, write f in its conjunctive normal form.

x	y	z	$f(x,y,z)$
1	1	1	1
1	1	0	1
1	0	1	0
1	0	0	0
0	1	1	0
0	1	0	1
0	0	1	0
0	0	0	1

16. Design a combinatorial circuit that computes exclusive OR (XOR) of x and y.

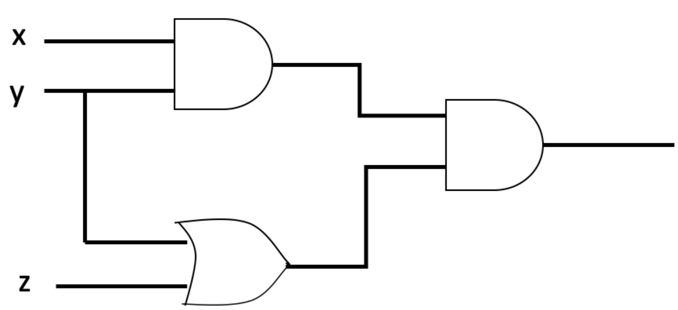
17. Given the Boolean function f, write f in its

(a) disjunctive normal form
(b) conjunctive normal form
(c) draw the combinatorial circuit to the disjunctive normal form.

x	y	z	$f(x,y,z)$
1	1	1	0
1	1	0	0
1	0	1	0
1	0	0	1
0	1	1	1
0	1	0	1
0	0	1	1
0	0	0	0

18. Find the disjunctive normal form of the function, $f(x,y) = (x+y).(x'+y')$ using algebraic technique

19. Find the disjunctive normal form for the following combinatorial circuit.



20. *Uniqueness of the complement*: Show that $y = x'$, when $x \vee y = 1$ antgvd $x \wedge y = 0$

Chapter 5

Differentiation and Integration

Limits (Summary)

The limit of a function

Definition 1

We write

$$lim_{x \to a} f(x) = L$$

and say "the limit of $f(x)$, as x approaches a, equals L"

if we can make the values of $f(x)$ arbitrarily close to L (as close to L as we like) by taking x to be sufficiently close to a (**on either side of** a) but not equal to a

One-sided limits

Definition 2

We write

$$lim_{x \to a^-} f(x) = L$$

and say the **left-hand limit of** $f(x)$ **as** x **approaches** a [or the **limit of** $f(x)$ **as** x **approaches** a **from the left**] is equal to L if we can make the values of $f(x)$ arbitrarily close to L by taking x to be sufficiently close to a and x less than a.

Notice that Definition 2 differs from Definition 1 only in that we require x to be less than a.

Similarly, if we require that x be grater than a, we get "the **right-hand limit of** $f(x)$ **as** x **approaches** a is equal to L" and we write

$$lim_{x \to a^+} f(x) = L$$

5.0.1 Relationship between the limit and one-side limits

$$lim_{x \to a} f(x) = L \quad \text{if and only if} \quad lim_{x \to a^-} f(x) = L \ \text{ and } \ lim_{x \to a^+} f(x) = L$$

That is,

$$lim_{x \to a} f(x) = L \Rightarrow lim_{x \to a^-} f(x) = lim_{x \to a^+} f(x) = L$$
$$lim_{x \to a^-} f(x) = lim_{x \to a^+} f(x) = L \Rightarrow lim_{x \to a} f(x) = L$$
$$lim_{x \to a^-} f(x) \neq lim_{x \to a^+} f(x) \Rightarrow lim_{x \to a} f(x) \text{ Does Not Exist}$$

Properties

Assume $lim_{x \to a} f(x)$ and $lim_{x \to a} g(x)$ both exist and c is any number then,

1. $lim_{x \to a}[cf(x)] = c\ lim_{x \to a} f(x)$
2. $lim_{x \to a}[f(x) \pm g(x)] = lim_{x \to a} f(x) \pm lim_{x \to a} g(x)$
3. $lim_{x \to a}[f(x)g(x)] = lim_{x \to a} f(x) lim_{x \to a} g(x)$
4. $lim_{x \to a}\left[\frac{f(x)}{g(x)}\right] = \frac{lim_{x \to a} f(x)}{lim_{x \to a} g(x)}$ provided $lim_{x \to a} g(x) \neq 0$
5. $lim_{x \to a}[f(x)]^n = [lim_{x \to a} f(x)]^n$

Continuous functions

Definition3

A function f is **continuous at a number** a if

$$lim_{x \to a} f(x) = f(a)$$

Notice that Definition 3 implicitly requires three things if f is continuous at a:

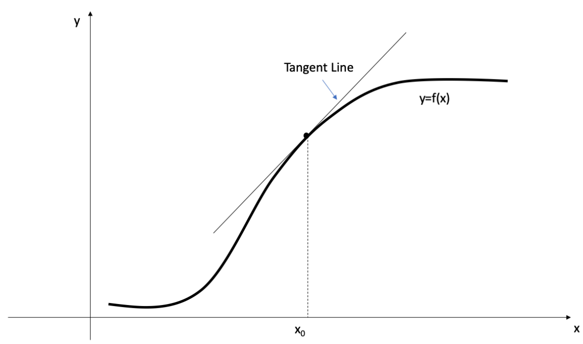
1. $f(a)$ is defined (that is, a is in the domain of f)
2. $lim_{x \to a} f(x)$ exits
3. $lim_{x \to a} f(x) = f(a)$

5.1 Differentiation

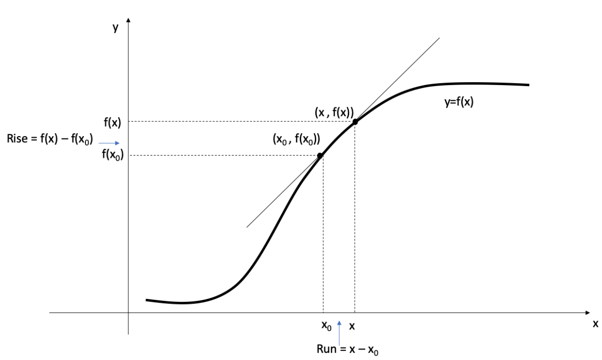
5.1.1 Derivative and tangent lines

5.1.1.1 Geometric meaning of the derivative

- The steepness of a line can be determined by its *slope*.
- The steepness of the graph of a function (f) at a given point on it (say $(x_0, f(x_0))$), can be approximated by its **tangent line** at that point.
- The slope of the tangent line at a point on the function is equal to the derivative of the function at that point
- *The derivative of f at x_0* is denoted by $f'(x_0)$

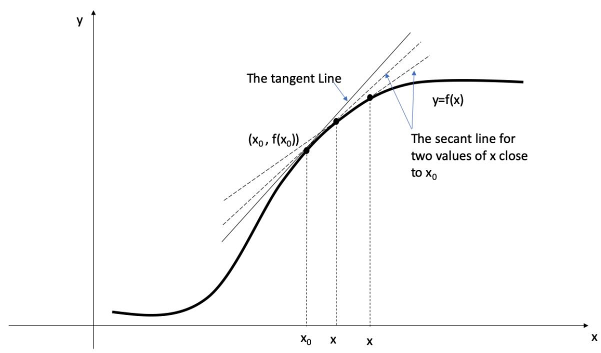


- The direct computation of the slope of the tangent line is not possible as we are given only one point $(x_0, f(x_0))$.
- Instead, we consider the **secant line** through two points: $(x_0, f(x_0))$ and $(x, f(x))$, on the graph of f.



- As x approaches x_0, the secant line is approaching to the tangent line.

- As x varies, the secant line turns about the fixed point $(x_0, f(x_0))$.
- We can describe the tangent line as the limiting line of the secant line when x approaches x_0 from either direction.



- As we move on the secant line from $(x_0, f(x_0))$ to $(x, f(x))$
  - the rise is $f(x) - f(x_0)$
  - the run is $x - x_0$.
- Therefore,

$$\text{The slope of the secant line} = \frac{f(x) - f(x_0)}{x - x_0} \text{ for } x \neq x_0.$$

- To have the derivative be the slope of the tangent line, we define it as the limit of the slope of the secant line.

Definition 1: Tangent line

The tangent line to the curve f at the point $(x_0, f(x_0))$ is the line through that point with slope

$$m = lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

provided that this limit exists

Definition 2

The derivative of a function f at x_0 is

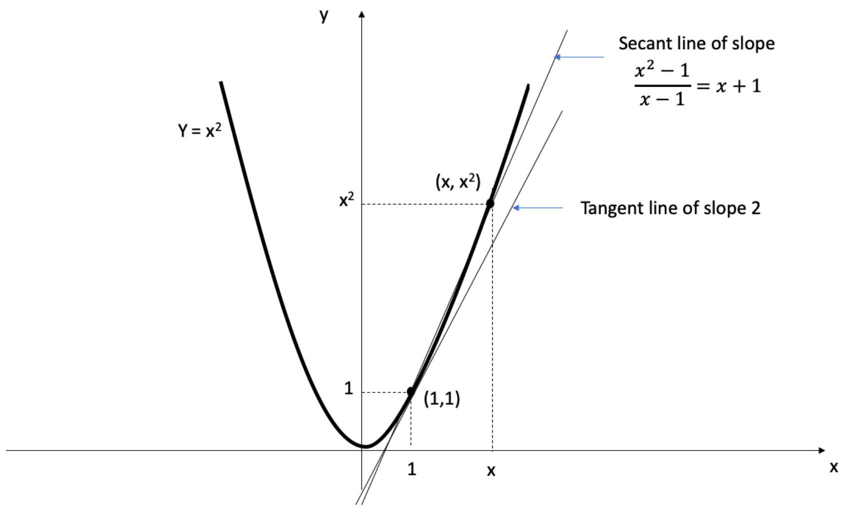
$$f'(x_0) = lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

provided this limit exists and is finite.

- If this limit does not exist or if it is infinite, we say that f **does not have a derivative at** x_0

Example 1 Compute the derivative of the function x^2 at $x = 1$, using the definition of derivative

The geometric interpretation of the above solution



- The derivative is the slope of the tangent line to the graph at the point $(1, 1)$. ($f'(1) = 2$)
- The secant line has slope $x + 1$, and the limit of that slope as x tends to 1 is the slope 2 of the tangent line

5.1.1.2 Equations of the tangent lines

- When we compute the derivative, $f'(x_0)$, we can give an equation for the tangent line at $(x_0, f(x_0))$ on the graph of f.
- The tangent line has the equation

$$y - f(x_0) = f'(x_0)(x - x_0)$$

Example 2

Give the equation for the tangent line to the parabola $y = x^2$ at $x = 1$.

5.1.1.3 Important properties

1. Differentiable functions are continuous

- A function must be continuous at any point where it has a derivative.

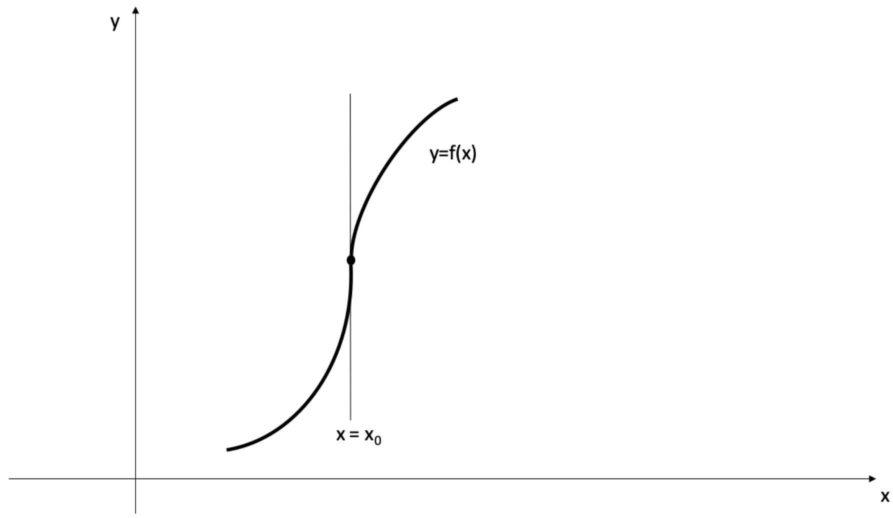
Theorem

If the derivative $f'(x_0)$ exists, then the function f is continuous at x_0.

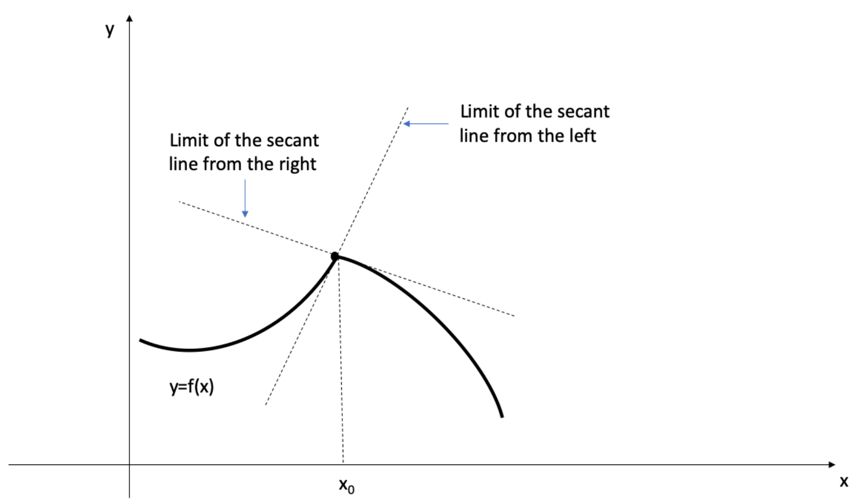
2. Continuous functions that do not have derivative

- The converse of the above theorem is false; that is, there are functions that are continuous but not differentiable.
- This usually occurs in one of the two ways:

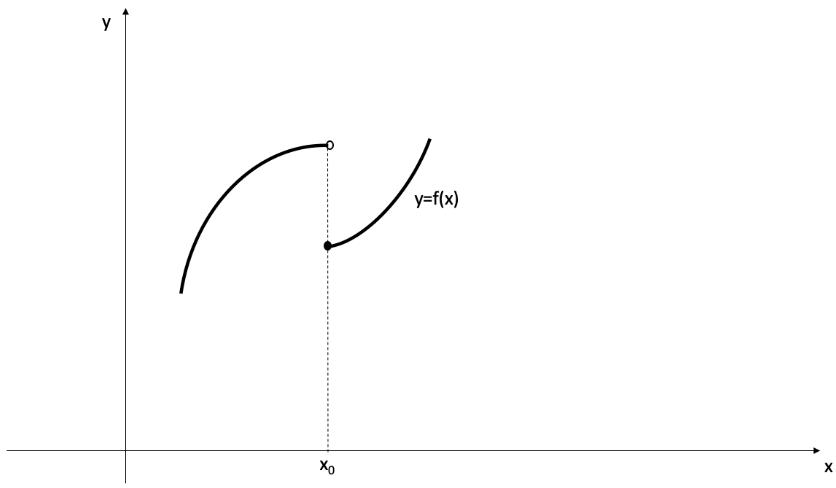
1. When the graph of f has a vertical tangent line at a given point. (The slope of the secant line $\frac{f(x)-f(x_0)}{x-x_0}$ tends to ∞ or to $-\infty$ as x tends to x_0)



2. When the secant line has different limits as x tends to x_0 from the left and from the right (i.e. when the graph has a sharp corner or a cusp at $x = x_0$).



The graph of a function will not have a tangent line at a point where the function is discontinuous.



- This graph does not have a tangent line at $x = x_0$
- Therefore, the function f does not have a derivative at x_0.

5.1.2 Leibniz notation and derivative as a function

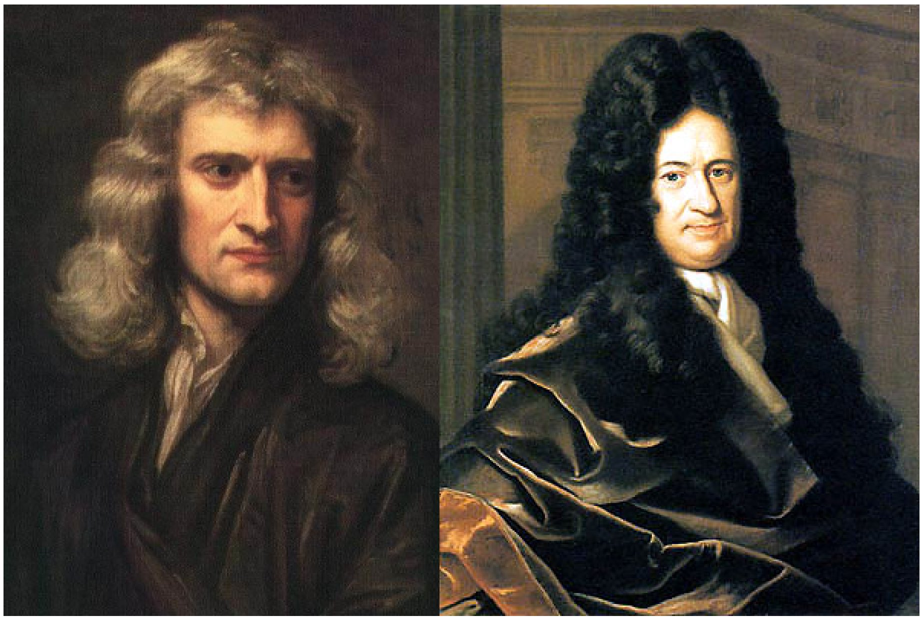


Figure 5.1: Isaac Newton and Gottfried Leibniz

5.1.2 Leibniz notation and derivative as a function

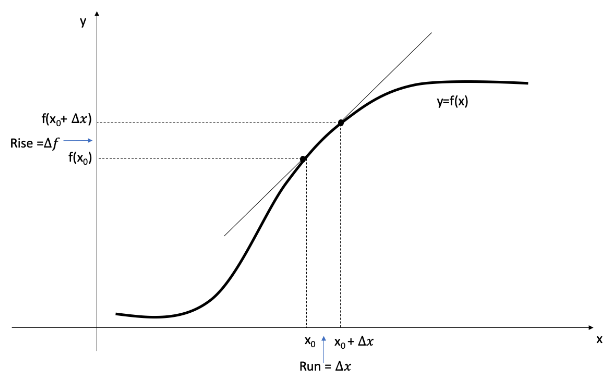
- In the Leibniz notation the derivative of f at x_0 is denoted by the symbol $\frac{df}{dx}(x_0)$
- With Leibniz notation the definition of the derivative reads

$$\frac{df}{dx}(x_0) = lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

We set $\Delta x = x - x_0$.

Then $x = x_0 + \Delta x$.

Let Δf for the change $f(x_0 + \Delta x) - f(x_0)$ in the value of f that occurs when we change x by Δx from x_0 to $x_0 + \Delta x$



Then the slope of the secant line takes the form

$$\frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = \frac{\Delta f}{\Delta x}.$$

Now the definition of the derivative reads

$$\frac{df}{dx}(x_0) = lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

Example 3

Use the above definition of the derivative to compute $\frac{df}{dx}(3)$ where $f(x) = x^2$

5.1.2.1 The derivative as a function

- So far we considered derivatives at a given value of x
- If we consider the derivative at a **variable** x, we get *the derivative function* $\frac{df}{dx}$ whose value $\frac{df}{dx}(x)$ is the slope of the tangent line to the graph of f at the point $(x, f(x))$.

$$\frac{df}{dx}(x) = lim_{\Delta x \to 0}\frac{f(x+\Delta x) - f(x)}{\Delta x}$$

5.1.3 Derivatives of polynomials and exponential functions

5.1.3.1 Derivative of a constant function

$$\frac{d}{dx}(c) = 0$$

5.1.3.2 Power functions

- We now look at the functions $f(x) = x^n$.

Pascal's Triangle and the Binomial Theorem

Binomial Expansions Using Pascal's Triangle

Consider the following expanded powers of $(a+b)^n$, where $a+b$ is any binomial and n is a whole number. Look for patterns.

$(a+b)^0 = 1$

$(a+b)^1 = a+b$

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

$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$

$(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$

$(a+b)^5 = a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + b^5$

Each expansion is a polynomial.

- Blaise Pascal found a numerical pattern, called **Pascal's Triangle**

Exponent	**Pascal's Triangle**
0	1
1	1 1
2	1 2 1
3	1 3 3 1
4	1 4 6 4 1
5	1 5 10 10 5 1
6	1 6 15 20 15 6 1

- Pascal's Triangle gives us the coefficients for an expanded binomial of the form $(a+b)^n$, where n is the row of the triangle.

Binomial Theorem

- According to the Binomial Theorem, if a and b are any real numbers and k is a positive integer, then

$$(a+b)^n = a^n + na^{n-1}b + \frac{n(n-1)}{2!}a^{n-2}b^2 + \frac{n(n-1)(n-2)}{3!}a^{n-3}b^3 +$$

$$\cdots + \frac{n(n-1)(n-2)\ldots(n-k+1)}{n!}a^{n-k}b^k + \cdots + nab^{n-1} + b^n$$

The traditional notation for the binomial coefficients is

$$\binom{n}{0} = 1$$

$$\binom{n}{k} = \frac{n(n-1)(n-2)\dots(n-k+1)}{k!}$$

$n = 1, 2, \dots, k$

which enables us to write the Binomial Theorem in the abbreviated form

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Example 1

$f(x) = x$

Example 2

$f(x) = x^2$

Example 3

$f(x) = x^3$

Example 4

$f(x) = x^4$

The Power Rule

If n is a positive integer, then

$$\frac{d}{dx}(x^n) = nx^{n-1}$$

Functions with negative integer exponents

Example 5

$f(x) = \frac{1}{x}$

When the exponent is a fraction

Example 6

$f(x) = \sqrt{x}$

The Power Rule (General version)

If n is any real number, then

$$\frac{d}{dx}(x^n) = nx^{n-1}$$

5.1.3.3 The constant multiple rule

If c is a constant and f is a differentiable function, then

$$\frac{d}{dx}[cf(x)] = c\frac{d}{dx}f(x)$$

Example 7

$f(x) = 3x^4$

5.1.3.4 The Sum Rule

If f and g are both differentiable, then

$$\frac{d}{dx}[f(x)+g(x)]=\frac{d}{dx}f(x)+\frac{d}{dx}g(x)$$

- The Sum Rule is applicable to the sum of any number of functions

5.1.3.5 The Difference Rule

If f and g are both differentiable, then

$$\frac{d}{dx}[f(x)-g(x)]=\frac{d}{dx}f(x)-\frac{d}{dx}g(x)$$

- The Constant Multiple Rule, the Sum Rule and the Difference Rule can be combined with the Power Rule to differentiate any polynomial.

Example 8

a) Differentiate $f(x) = x^4 - 6x^2 + 4$

b) Find the points on the curve $f(x) = x^4 - 6x^2 + 4$ where the tangent line is horizontal

5.1.3.6 Exponential Functions

Exercise:

Compute the derivative of the exponential function $f(x) = a^x$ using th definition of a derivative

Definition: Definition of the number e

e is the number such that

$$lim_{h\to 0}\frac{e^h - 1}{h} = 1$$

- If we put $a = e$ and , then, $f'(0) = 1$.

- Then **derivative of the Natural Exponential Function**

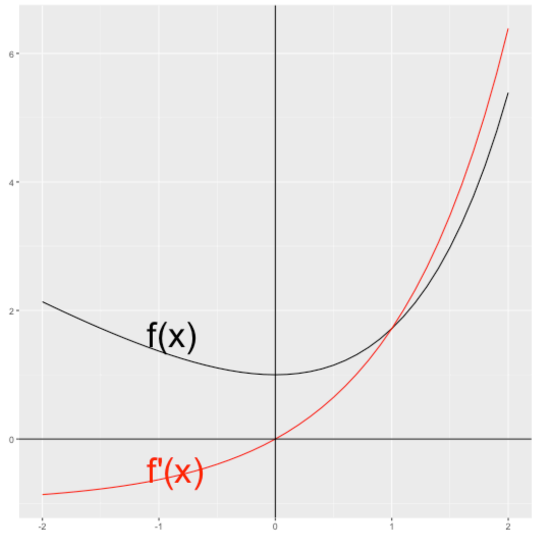
$$\frac{d}{dx}(e^x) = e^x$$

- Therefore, the exponential function, $f(x) = e^x$ has the property that it is its own derivative.

- Geometrically, this means that the slope of the tangent line to the curve $f(x) = e^x$ is equal to the y-coordinate of the point.

Example 9

If $f(x) = e^x - x$, find $f'(x)$



5.1.4 The product and quotient rules

5.1.4.1 The product rule

If $f$and g are both differentiable, then

$$\frac{d}{dx}[f(x)g(x)] = f(x)\frac{d}{dx}[g(x)] + g(x)\frac{d}{dx}[f(x)]$$

Example 10

If $f(x) = xe^x$, find $f'(x)$

5.1.4.2 The quotient rule

If $f$and g are both differentiable, then

$$\frac{d}{dx}[\frac{f(x)}{g(x)}] = \frac{g(x)\frac{d}{dx}[f(x)] - f(x)\frac{d}{dx}[g(x)]}{[g(x)]^2}$$

Example 11

If $f(x) = \frac{x^2+x-2}{x^3+6}$, find $f'(x)$

5.1.5 The chain rule

If $y = f(u)$ and $u = g(x)$ are both differentiable functions, then

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

Example 12

If $f(x) = \sqrt{x^2 + 1}$, find $f'(x)$ using chain rule.

5.1.5.1 The power rule combined with the chain rule

If n is any real number and $u = g(x)$ is differentiable, then

$$\frac{d}{dx}(u^n) = nu^{n-1}\frac{du}{dx}$$

Alternatively,

$$\frac{d}{dx}([g(x)]^n) = n[g(x)]^{n-1}\frac{d}{dx}g(x)$$

Example 13

If $f(x) = (x^3 - 2)^{100}$, find $f'(x)$

Example 14

If $f(x) = \frac{1}{(x^2+x+1)^{1/3}}$, find $f'(x)$

5.1.6 Partial Derivatives

If f is a function of two variables, its **partial derivatives** are the functions f_x and f_y defined by

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

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

- There are many alternative notations for partial derivatives.
- Instead of f_x, it can be written as f_1 or $D_1 f$ to indicate that the differentiation is with respect to the *first* variable.
- Another notation is $\frac{\partial f}{\partial x}$.
- But here $\frac{\partial f}{\partial x}$ can't be interpreted as a ratio of differentials.

Notations for Partial Derivatives

If $z = f(x, y)$, we write

$$f_x = f_x(x, y) = \frac{\partial f}{\partial x} = \frac{\partial}{\partial x} f(x, y) = \frac{\partial z}{\partial x} = f_1 = D_1 f = D_x f$$

$$f_y = f_y(x, y) = \frac{\partial f}{\partial y} = \frac{\partial}{\partial y} f(x, y) = \frac{\partial z}{\partial y} = f_2 = D_2 f = D_y f$$

- The partial derivative with respect to x is just the *ordinary* derivative of the function of a single variable that we get by keeping y fixed.

- Therefore, we have the following rule.

Rule for finding Partial Derivatives of $z = f(x, y)$

1. To find f_x, regard y as a constant and differentiate $f(x, y)$ with respect to x.

2. To find f_y, regard x as a constant and differentiate $f(x, y)$ with respect to y.

Example 15

If $f(x, y) = x^3 + x^2y^3 - 2y^2$, find $f_x(2, 1)$ and $f_y(2, 1)$

5.1.7 Higher Derivatives

- If f is a differentiable function, then its derivative f' is also a function.
- It is denoted by $(f')' = f''$.
- This f'' is called the **second derivative** of f as it is the derivative of the derivative of f.
- Using Leibniz notation we can write the second derivative of f as follows:

$$\frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{d^2y}{dx^2}$$

- Another notation is $f'' = D^2 f(x)$.
- We can interpret f'' as the slope of the curve $y = f'(x)$ at the point $(x, f'(x))$.

- In other words, it is the rate of change of the slope of the original curve $f(x)$.
- The **third derivative** f''' is the derivative of the second derivative: $f''' = (f'')'$.
- Therefore, $f'''(x)$ can be interpreted as the slope of the curve $y = f''(x)$ or as the rate of change of $f''(x)$
- Alternative notations for the third derivative:

$$y''' = f'''(x) = \frac{d}{dx}(\frac{d^2y}{dx^2}) = \frac{d^3y}{dx^3} = D^3f(x)$$

- In general, the nth derivative of f is denoted by $f^{(n)}$ and is obtained from f by differentiating n times.
- If $y = f(x)$,

$$y^{(n)} = f^{(n)}(x) = \frac{d^ny}{dx^n} = D^nf(x)$$

Example 16

If $f(x) = x^3 - 7x^2 - 3x + 5$

5.1.8 Derivatives of logarithmic functions

$$\frac{d}{dx}(a^x) = a^x ln(a)$$

$$\frac{d}{dx}(log_a x) = \frac{1}{xln(a)}$$

$$\frac{d}{dx}(lnx) = \frac{1}{x}$$

Example 17

If $y = ln(x^3 + 1)$

In general, if we combine derivatives of logarithmic functions with the chain rule. we get

$$\frac{d}{dx} ln\ u = \frac{1}{u}\frac{du}{dx}$$

or

$$\frac{d}{dx}[ln\ g(x)] = \frac{g'(x)}{g(x)}$$

5.1.9 Applications of differentiation

5.1.9.1 Maximum and minimum values

1. Definition

A function f has an **absolute maximum** (or **global maximum**) at c if $f(c) \geq f(x)$ for all x in D, where D is the domain of f. The number $f(c)$ is called the **maximum value** of f on D.

Similarly, f has an **absolute minimum** (or **global minimum**) at c if $f(c) \leq f(x)$ for all x in D and the number $f(c)$ is called the **minimum value** of f on D.

The maximum and minimum values of f are called the **extreme values** of f.

2. Definition

A function f has an **local maximum** (or **relative maximum**) at c if $f(c) \geq f(x)$ when x is near c. [This means that $f(c) \geq f(x)$ for all x in some open interval containing c]

Similarly, f has a **local minimum** (or **relative minimum**) at c if $f(c) \leq f(x)$ when x is near c.

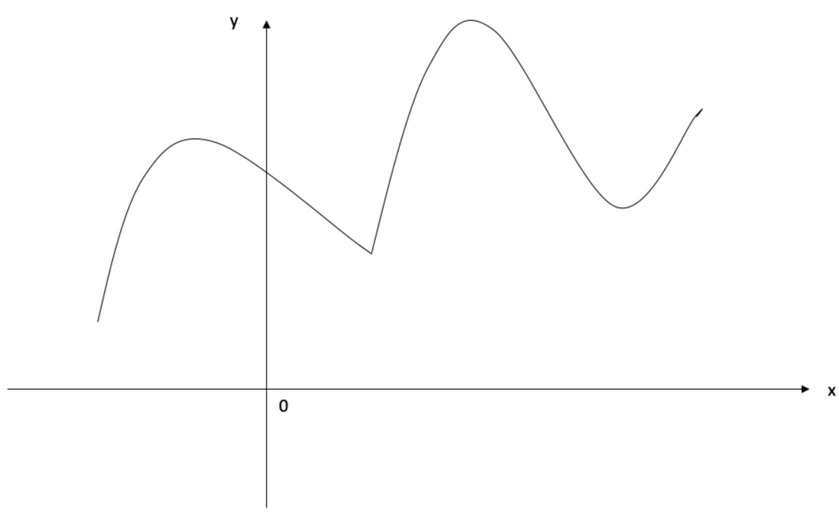


Figure 5.2: Isaac Newton and Gottfried Leibniz

3. The Extreme Value Theorem

If f is continuous on a closed interval $[a, b]$, then f attains an absolute maximum value $f(c)$ and an absolute minimum value $f(d)$ at some numbers c and d in $[a, b]$

- A function need not possess extreme values if either hypothesis (continuity or closed interval) is omitted from the Extreme Value Theorem

4. Fermat's Theorem

If f has a local maximum or minimum at c, and if $f'(c)$ exists, then $f'(c) = 0$.

Examples

5. Definition

A **critical number** of a function f is a number c in the domain of f such that either $f'(c) = 0$ or $f'(c)$ does not exist.

Example 18

Find the critical numbers of $f(x) = x^{3/5}(4 - x)$.

- In terms of critical numbers, Fermat's Theorem can be rephrased as,

6. Fermat's Theorem (modification)

If f has a local maximum or minimum at c, then c is a critical number of f.

7. The closed interval method

To find the *absolute* maximum and minimum values of a **continuous** function f on a closed interval $[a, b]$:

1. Find the values of f at the critical numbers of f in (a, b).

2. Find the values of f at the endpoints of the interval.
3. The largest of the values from Steps 1 and 2 is the absolute maximum value; the smallest of these values is the absolute minimum value.

Example 19

Find the absolute maximum and minimum values of the function $f(x) = x^3 - 3x^2 + 1 \; -\frac{1}{2} \leq x \leq 4$.

5.1.9.2 How derivatives affect the shape of a graph

5.1.9.2.1 What does f' say about f?

1. **Increasing/ Decreasing Test**

(a) If $f'(x) > 0$ on an interval, then f is increasing on that interval
(b) If $f'(x) < 0$ on an interval, then f is decreasing on that interval

Example 20

Find where the function $f(x) = 3x^4 - 4x^3 - 12x^2 + 5$ is increasing and where it is decreasing

2. **The First Derivative Test**

Suppose that c is a critical number of a continuous function f.

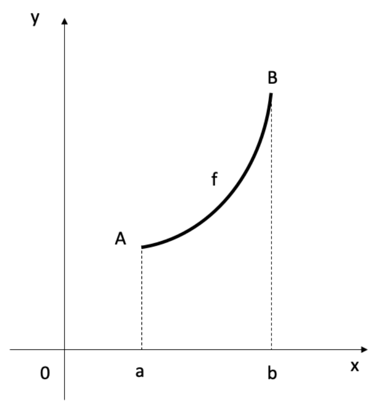
(a) If f' changes from positive to negative at c, then f has a local maximum at c.

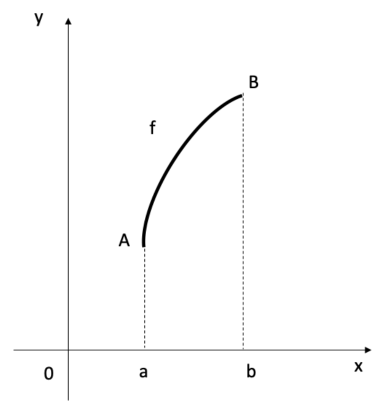
(b) If f' changes from negative to positive at c, then f has a local minimum at c.

(c) If f' does not change sign at c (for example, if f' is positive on both sides of c or negative on both sides), then f has no local maximum or minimum at c.

5.1.9.2.2 What does f'' say about f?

Concavity





3. **Definition**

If the graph of f lies above all of its tangents on an interval I, then it is called **concave upward** on I. If the graph of f lies below all of its tangents on I, it is called **concave downward** on I.

4. **Concavity Test**

(a) If $f''(x) > 0$ for all x in I, then the graph of f is concave upward on I.

(b) If $f''(x) < 0$ for all x in I, then the graph of f is concave downward on I.

5. **Definition: Inflection Point**

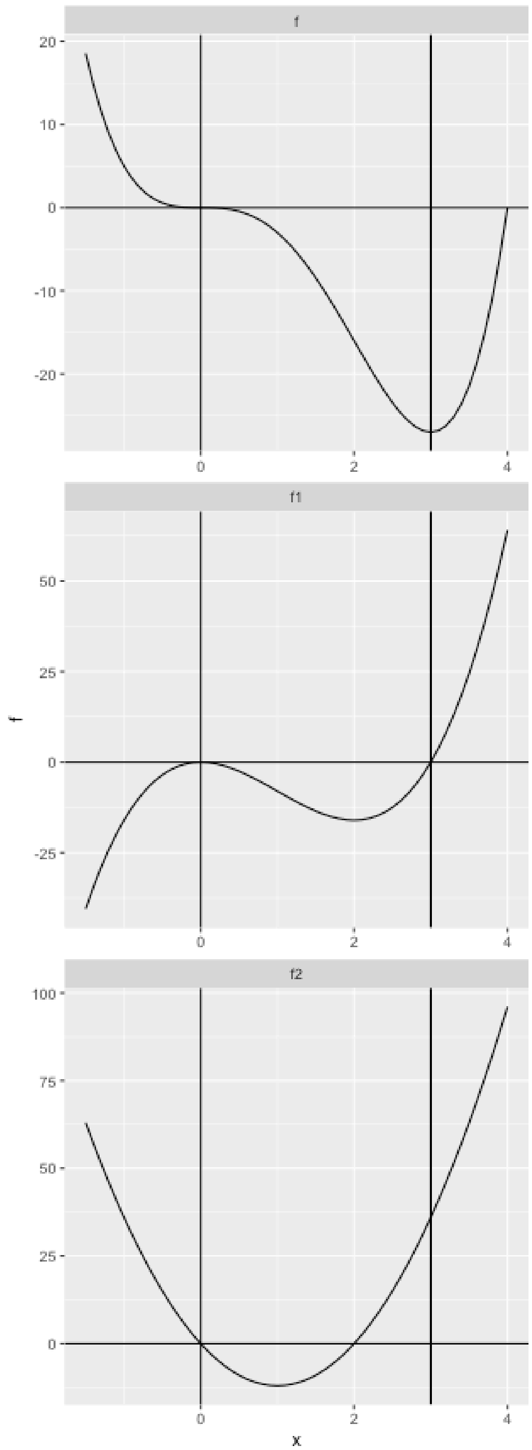
A point P in a curve $y = f(x)$ is called an **inflection point** if f is continuous there and the curve changes from concave upward to concave downward or from concave downward to concave upward at P.

6. **The Second Derivative Test**

Suppose f'' is continuous near c. (a) If $f'(c) = 0$ and $f''(c) > 0$, then f has a local minimum at c (b) If $f'(c) = 0$ and $f''(c) < 0$, then f has a local maximum at c

Example 21

Discuss the curve $y = x^4 - 4x^3$ with respect to concavity, points of inflection and local maxima and minima.


f
f1
f2
f
x
20
10
0
-10
-20
50
25
-25
100
75
2
4

5.2 Integration

5.2.1 The Definite Integral

Definition: Definite Integral

If f is a continuous function defined for $a \leq x \leq b$, we divide the interval $[a, b]$ into n sub-intervals of equal width $\Delta x = (b - a)/n$. We let $x_0(= a), x_1, x_2, ..., x_n(= b)$ be the endpoints of these sub-intervals and we let $x_1^*, x_2^*, ..., x_n^*$ be any **sample point** in these sub-intervals, so x_i^* lies in the ith sub-interval $[x_{i-1}, x_i]$. Then **definite integral of f from a to b** is

$$\int_a^b f(x)dx = lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\Delta x$$

- The **definite integral** $\int_a^b f(x)dx$ is a number and does not depend on x.
- Therefore, any letter can be used in place of x without changing the value of the integral

$$\int_a^b f(x)dx = \int_a^b f(t)dt = \int_a^b f(r)dr$$

- A definite integral can be interpreted as a **net area.**

5.2.1.1 Properties of the Definite Integral

1. $\int_a^b f(x)dx = -\int_b^a f(x)dx$
2. $\int_a^a f(x)dx = 0$
3. $\int_a^b [f(x) + g(x)]dx = \int_a^b f(x)dx + \int_a^b g(x)dx$
4. $\int_a^b cdx = c(b-a)$, where c is a constant.
5. $\int_a^b cf(x)dx = c\int_a^b f(x)dx$, where c is a constant.
6. $\int_a^b [f(x) - g(x)]dx = \int_a^b f(x)dx - \int_a^b g(x)dx$
7. $\int_a^b f(x)dx + \int_b^c f(x)dx = \int_a^c f(x)dx$

5.2.2 The Fundamental Theorem of Calculus

- The fundamental Theorem of Calculus establishes a connection between the two branches of calculus and integral calculus.

$$g(x) = \int_a^x f(t)dt,$$

where f is a continuous function on $[a, b]$ and x varies between a and b.

The Fundamental Theorem of Calculus, Part I

If f is continuous on $[a, b]$, then the function g defined by

$$g(x) = \int_a^x f(t)dt, \;\; a \leq x \leq b$$

is continuous on $[a, b]$ and differentiable on (a, b), and $g'(x) = f(x)$.

- Using Leibniz notation for derivatives, we can write the Fundamental Theorem of Calculus, Part I as

$$\frac{d}{dx}\int_a^x f(t)dt = f(x)$$

when f is continuous.

Example 22

Find the derivative of the function $g(x) = \int_0^x \sqrt{1+t+t^2}dt$

5.2.3 Indefinite Integrals

5.2.3.1 Antiderivatives

- Consider the problem of finding a function F whose derivative is a known function f.
- If such a function F exists, it is called **an antiderivative** of f.

Definition: Antiderivative

A function F is called an antiderivative of f on an interval I if $F'(x) = f(x)$ for all x in I

Indefinite integral

- The notation $\int f(x)dx$ is traditionally used for an antiderivative of f and is called an **indefinite integral.**

- Thus

$$\int f(x)dx = F(x) \text{ means } F'(x) = f(x)$$

- If f is continuous on $[a, b]$, then

$$\int_a^b f(x)dx = \left[\int f(x)dx\right]_a^b$$

Indefinite Integrals

1. $\int cf(x)dx = c\int f(x)dx$
2. $\int [f(x) + g(x)]dx = \int f(x)dx + \int g(x)dx$
3. $\int kdx = kx + C$
4. $\int x^n dx = \frac{x^{n+1}}{n+1} + C \ (n \neq -1)$
5. $\int \frac{1}{x}dx = ln|x| + C$
6. $\int e^x dx = e^x + C$
7. $\int a^x dx = \frac{a^x}{ln\ a} + C$

Example 23

Evaluate $\int_1^9 \frac{2t^2+t^2\sqrt{t}-1}{t^2} dt$

5.2.4 The Substitution Rule

The Substitution Rule

If $u = g(x)$ is a differentiable function whose range is and interval I and f is continuous on I, then

$$\int f(g(x))g'(x)dx = \int f(u)du$$

Example 24

Evaluate $\int \sqrt{2x+1}dx$

The Substitution Rule for Definite Integrals

If g' is continuous on $[a, b]$ and f is continuous on the range of $u = g(x)$, then

$$\int_a^b f(g(x))g'(x)dx = \int_{g(a)}^{g(b)} f(u)du$$

Example 25

Evaluate $\int_0^4 \sqrt{2x+1}dx$

Integrals of Symmetric Functions

Suppose f is continuous on $[-a, a]$.

a) If f is even $[f(-x) = f(x)]$, then $\int_{-a}^{a} f(x)dx = 2\int_{0}^{a} f(x)dx$
b) If f is odd $[f(-x) = -f(x)]$, then $\int_{-a}^{a} f(x)dx = 0$

5.2.5 The integral of the logarithmic and exponential function

1. $\int e^x dx = e^x + C$

2. $\int \frac{1}{x} dx = ln(x) + C$ if $x > 0$

3. $\int a^x dx = \frac{a^x}{ln(a)} + C$

Example 26

Evaluate $\int \frac{e^x - e^{-x}}{5} dx$

Example 27

Evaluate $\int \frac{3x^2}{x^3+1} dx \; x \neq -1$

5.2.6 Integration by Parts

$$\int f(x)g'(x)dx = f(x)g(x) - \int g(x)f'(x)dx$$

- Let $u = f(x)$ and $v = g(x)$. Then the differentials are $du = f'(x)dx$ and $dv = g'(x)dx$, so, by the Substitution Rule, the formula for integration by parts becomes

$$\int udv = uv - \int vdu$$

Example 28

Evaluate $\int ln\ xdx$

Example 29

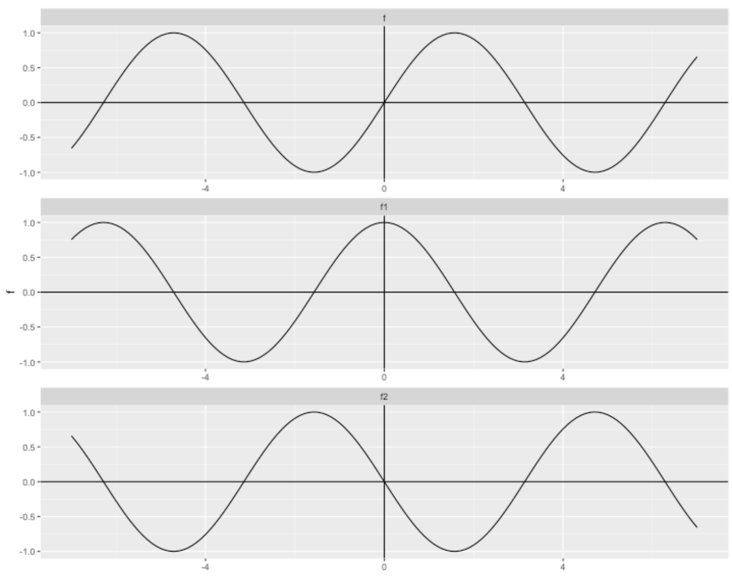
Evaluate $\int x^2 e^x dx$

$$\int_a^b f(x)g'(x)dx = [f(x)g(x)]_a^b - \int_a^b g(x)f'(x)dx$$

5.2.7 Trigonometric Functions

Derivatives of Trigonometric Functions

1. $\frac{d}{dx}sin(x) = cos(x)$
2. $\frac{d}{dx}cos(x) = -sin(x)$
3. $\frac{d}{dx}tan(x) = sec^2(x)$
4. $\frac{d}{dx}sec(x) = sec(x)tan(x)$



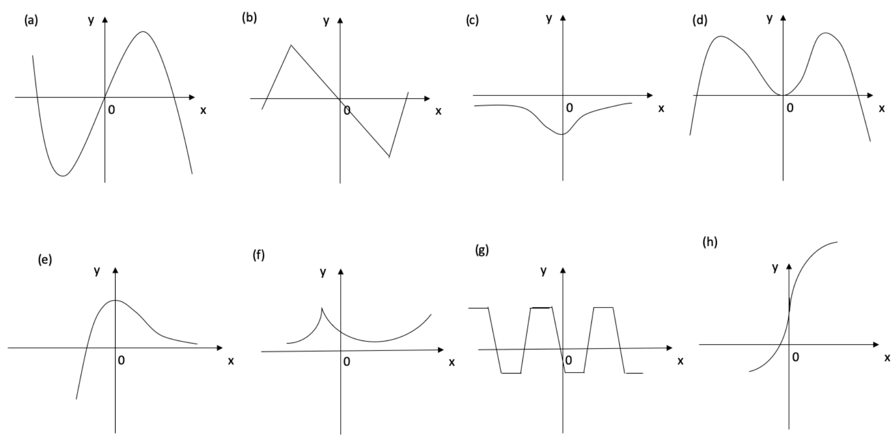
Integration of Trigonometric Functions

1. $\int cos(x)dx = sin(x) + C$
2. $\int sin(x)dx = -cos(x) + C$
3. $\int sec^2(x) = tan(x) + C$
4. $\int sec(x)tan(x) = sec(x) + C$

5.3 Tutorial

Chapter 5.1: Differentiation

1. The graph of function f is given. Use it to sketch the graph of the derivative f'



2. Find the derivative of the function using the definition of derivative. state the domain of the function and the domain of the derivative.

(a) $f(x) = 20$
(b) $f(x) = 1 - 3x^2$
(c) $f(x) = x^3 - 3x + 5$
(d) $f(x) = \sqrt{1 + 2x}$
(e) $f(x) = 12 + 5x$
(f) $g(x) = 5x^2 + 3x - 2$
(g) $g(x) = x + \sqrt{x}$
(h) $g(x) = \frac{3+x}{1-3x}$
(i) $G(t) = \frac{4t}{t+1}$
(j) $g(x) = \frac{1}{x^2}$
(k) $G(t) = t^4$

3. The graph of f is given.

(i) At what values is f discontinuous? Why?
(ii) At what values is f not differentiable? Why?

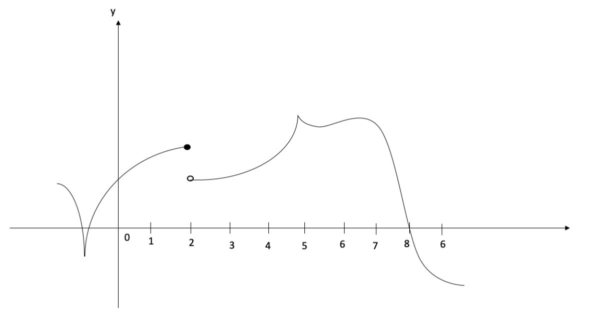


Figure 5.3: (a)

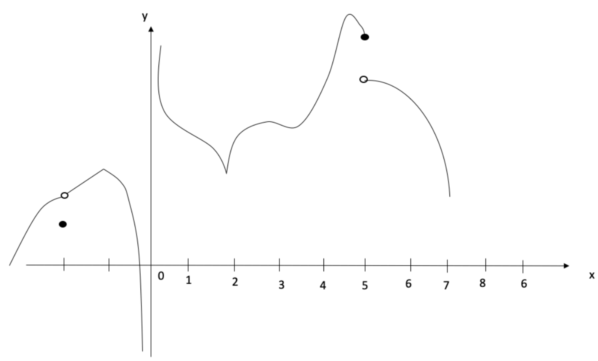


Figure 5.4: (b)

4. Where is the function $f(x) = |x|$ differentiable? find a formula for f' and sketch its graph.

5. Find an equation of the tangent line to the curve $y = x\sqrt{x}$ at the point $(1, 1)$. Illustrate by graphing the curve and its tangent line.

6. If $f(x) = e^x - x$, find f'. Compare the graphs of f and f'.

7. At what point on the curve $y = e^x$ is the tangent line parallel to the line $y = 2x$?

8. Differentiate the function

(a) $g(x) = \frac{3x-1}{2x+1}$
(b) $V(x) = (2x^3 + 3)(x^4 - 2x)$
(c) $f(y) = \left(\frac{1}{y^2} - \frac{3}{y^4}\right)(y + 5y^3)$
(d) $f(x) = \frac{1}{x^4+x^2+1}$

(e) $f(x) = \frac{x}{x+\frac{c}{x}}$
(f) $f(x) = \frac{3x^2+2\sqrt{x}}{x}$

9. Find an equation of the tangent line to the given curve at the specified point.

(a) $y = \frac{1}{1+x^2}$, $(-1, \frac{1}{2})$
(b) $y = \frac{e^x}{1+x^2}$, $(1, \frac{e}{2})$

10. Suppose that $f(5) = 1$, $f'(5) = 6$, $g(5) = -3$ and $g'(5) = 2$. Find the following values.

(a) $(fg)'(5)$
(b) $(f/g)'(5)$
(c) $(g/f)'(5)$

11. If $h(2) = 4$ and $h'(2) = -3$, find

$$\frac{d}{dx}\left(\frac{h(x)}{x}\right)\Big|_{x=2}$$

12.

(a) Use the Product Rule twice to prove that if f, g and h are differentiable, then

$$(fgh)' = f'gh + fg'h + fgh'$$

(b) Taking $f = g = h$ in part (a), show that

$$\frac{d}{dx}[f(x)]^3 = 3[f(x)]^2 f'(x)$$

(c) Use part (b) to differentiate $y = e^{3x}$

13. Write the composite function in the form $f(g(x))$. Then find the derivative of the function.

(a) $f(x) = \sqrt{x^2+1}$

(b) $f(x) = (x^3-1)^{100}$

(c) $f(x) = e^{-mx}$
(d) $g(t) = \frac{1}{(t^4+1)^3}$
(e) $f(x) = (1+4x)^5(3+x-x^2)^8$
(f) $f(t) = \left(\frac{t-2}{2t+1}\right)^9$

14. Suppose that $F(x) = f(g(x))$ and $g(3) = 6$, $g'(3) = 4$, $f'(3) = 2$ and $f'(6) = 7$. Find $F'(3)$.

15. A table of values for f, g, f', and g' is given.

x	$f(x)$	$g(x)$	$f'(x)$	$g'(x)$
1	3	2	4	6
2	1	8	5	7
3	7	2	7	9

(a) If $h(x) = f(g(x))$, Find $h'(1)$. (b) If $H(x) = g(f(x))$, Find $H'(1)$.

16. Suppose L is a function such that $L'(x) = 1/x$ for $x > 0$. Find an expression for the derivative of each function.

(a) $f(x) = L(x^4)$
(b) $g(x) = L(4x)$
(c) $F(x) = [L(x)]^4$
(d) $G(x) = L(1/x)$

17. Find the fist and second derivatives of the function

(a) $f(x) = x^5 + 6x^2 - 7x$
(b) $F(t) = t^8 - 7t^6 + 2t^4$
(c) $F(t) = (1-7t)^6$

18. (a) Show that $\frac{d}{dx}[ln\ g(x)] = \frac{g'(x)}{g(x)}$ (b) Differentiate $y = ln(x^3+1)$

19. Find $\frac{d}{dx} ln\left[\frac{x+1}{\sqrt{x-2}}\right]$

20. Sketch the graph of the function $f(x) = x^{\frac{2}{3}}(6-x)^{\frac{1}{3}}$.

21. Suppose you are given a formula for a function f.

(a) How do you determine where f is increasing or decreasing?
(b) How do you determine where the graph of f is concave upward or concave downward? (c)How do you locate inflection points?

22. If $x^2 + y^2 = 25$,

(a) Find $\frac{dy}{dx}$
(b) Fins an equation of the tangent to the circle $x^2 + y^2 = 25$ at the point $(3, 4)$.

23. If $f(x, y) = 4 - x^2 - 2y^2$, find $f_x(1, 1)$ and $f_y(1, 1)$ and interpret these numbers as slopes.

24. If $f(x, y) = sin\left(\frac{x}{1+y}\right)$, calculate $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$

25. Differentiate $y = \frac{x^{3/4}\sqrt{x^2+1}}{(3x+2)^5}$

26. Use the properties of integrals to evaluate $\int_0^1 (4 + 3x^2)dx$

27. If it is known that $\int_0^{10} f(x)dx = 17$ and $\int_0^8 f(x)dx = 12$, find $\int_8^{10} f(x)dx$

28. Use Part I of the Fundamental Theorem of Calculus to find the derivative of the function

(a) $g(x) = \int_o^x \sqrt{1+2t}dt$
(b) $g(x) = \int_o^x ln(t)dt$
(c) $g(y) = \int_2^y t^2 sin(t)dt$
(d) $g(y) = \int_3^y \frac{1}{x+x^2}dx$
(e) $g(x) = \int_x^2 cos(t^2)dt$
(f) $y = \int_{1-3x}^1 \frac{u^3}{1+u^2}du$

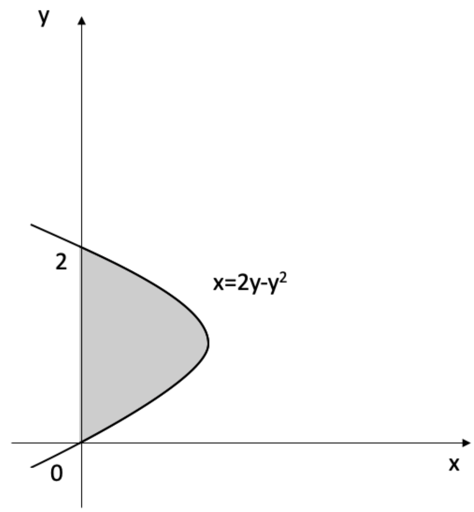
29. Find the general indefinite integral $\int(10x^4 - 2sec^2 x)dx$.

30. Evaluate the integral

(a) $\int_0^2 (6x^2 - 4x + 5)dx$
(b) $\int_{-1}^0 (2x - e^x)dx$
(c) $\int_{-2}^2 (3u + 1)^2 du$

(d) $\int_1^4 \sqrt{t}(1+t)dt$
(e) $\int_1^4 \sqrt{\frac{5}{x}}dx$

31. Find the area of the region that lies to the right of the y-axis and to the left of the parabola $x = 2y - y^2$.



32. Find

(a) $\int \frac{x}{\sqrt{1-4x^2}}dx$
(b) $\int \sqrt{1+x^2}x^5dx$

33. Find

(a) $\int tan\ x\ dx$
(b) $x^3cos(x^4+2)dx$

34. Calculate

(a) $\int_1^e \frac{ln\ x}{x}dx$
(b) $\int e^{5x}dx$

35. Show that $\frac{d}{dx}(log_a x) = \frac{1}{xln(a)}$

36. Find $\int x sin(x) dx$
37. Find $\int t^2 e^t dt$
38. Evaluate $\int e^x sin(x) dx$
39. If $f(0) = g(0) = 0$ and f'' and g'' are continuous, show that

$$\int_0^a f(x)g''(x)dx = f(a)g'(a) - f'(a)g(a) + \int_0^a f''(x)g(x)dx$$

40. Suppose that $f(1) = 2$, $f(4) = 7$, $f'(1) = 5$, $f'(4) = 3$ and f'' is continuous. Find the value of $\int_1^4 xf''(x)dx$

References

Shenk, A. (1988). Calculus and analytic geometry. Scott Foresman & Company.

Stewart, J. (2003). Calculus: Early Transcendentals. Thomson Learning. Inc., Belmont, CA.

Chapter 6

Descriptive Statistics

6.1 Introduction to Statistics

6.1.1 Some Basic Terminologies Used in Statistics

i Population

- The set of **all** possible elements in the universe of interest to the researcher

ii Sample

- A Sample is a **subset** (a portion or part) of the population of interest
- The sample must be a representative of the population of interest

iii Element

- Element is an **entity or object** which the information is collected.
- *Eg: Student, household, farm, company, tomato plant*

iv Variable

- A variable is **a feature characteristic which has different 'values' or categories for different elements** (items/subjects/individuals)
- *Eg: Gender of client, brand of mobile phones, risk level, number of emails received per day, age of client, income of client*

v Data

- Data are **measurements or facts** that are collected from a statistical unit/entity of interest
- We collect data on variables
- Data are raw numbers or facts that must be processed (analysed) to get useful information.
- We get information from data.
- *Eg:*

Variable: *Age (in years) of client*

Data: *21, 45, 18, 32, 30, 22, 23, 27*

Information:

The mean age is 27.25 years

The minimum age is 18 years

The range of ages is 18-45

The percentage of clients below 25 years of age: 50%

vi Statistic

- **Characteristic** of a **sample**
- The value which calculated based on sample data

vii Parameter

- **Characteristic** of a **population**
- The value which calculated based on population data

viii Census

- When a researcher **gathers data from the whole population for a given measurement,** it is called a census

ix Sampling

- When a researcher **gathers data from a sample of the population for a given measurement,** it is called sampling
- The process of selecting a sample is also called sampling

Why take a sample instead of studying every member of the population ?

- Prohibitive cost of census
- Destruction of item being studied may be required
- Not possible to test or inspect all members of a population being studied.

6.1.2 Branches of Statistics

i Descriptive Statistics

- Descriptive statistics consists of organizing, summarizing and presenting data in an informative way.
- The main purpose of descriptive statistics is to provide an overview of the data collected.
- Descriptive statistics describes the data collected through frequency tables, graphs and summary measures (mean, variance, quartiles, etc.).

ii Inferential Statistics

- In inferential statistics sample data are used to draw inferences (i.e. derive conclusions) or make predictions about the populations from which the sample has been taken.
- This includes methods used to make decisions, estimates, predictions or generalizations about a population based on a sample.
- This includes point estimations, interval estimation, test of hypotheses, regression analysis, time series analysis, multivariate analysis, etc.

6.1.3 Types of Variables

6.1.3.1 Qualitative / Quantitative Variables

i Qualitative variable (Categorical variable)

- The characteristic is a quality.
- The data are categories.
- They cannot be given numerical values.
- However, it may be given a numerical label
- Qualitative variables are sometimes referred as categorical variables.
- *Eg:*

Gender:

Age group:

Education level:

A/L stream:

Degree type:

Hair colour:

FIT student batch:

Undergraduate level:

Grade that you can obtain for CM 1110/ CM1130

ii Quantitative variable

- The characteristic is a quantity
- The data are numbers
- Quantitative data require numeric values that indicate how much or how many.
- They are obtained by counting or measuring with some scale
- *Eg:*

Number of family members:

Number of emails received per day:

Weight of a student:

Age:

Credit balance in the SIM card:

Time remaining in class:

Temperature:

Marks

6.1.3.2 Discrete/ Continuous Variables

- Quantitative variables can be classified as either discrete or continuous.

i Discrete Variables

- Quantitative
- Usually the data are obtained by counting
- There are impossible values between any two possible values
- *Eg:*

Number of family members:

Number of emails received per day:

ii Continuous Variables

- Quantitative
- Usually, the data are obtained by measuring with a scale

- There are no impossible values between any two possible values.(any value between any two possible values is also a possible value)
- i.e a continuous variable can take any value within a specified range.
- *Eg:*

Weight of a student:

Age:

Credit balance in the SIM card:

Time remaining in class:

Temperature:

Marks

6.1.4 Scales of Measurements

- There are four levels of measurements called, **nominal, ordinal, interval and ratio.**
- Each levels has its own rules and restrictions
- Different levels of measurement contains different amount of information with respect to whatever the data are measuring

i Nominal Scale

- Qualitative
- No order or ranking in categories.
- These categories have to be mutually exclusive, i.e. it should not be possible to place an individual or object in more than one category
- A name of a category can be substituted by a number, but it will be mere label and have no numerical meaning

ii Ordinal Scale

- Qualitative
- Categories can be ordered or ranked
- A name of a category can be substituted by a number, but such a sequence does not indicate absolute quantities.
- Difference between any two numbers on the scale does not have a numerical meaningful.
- It cannot be assumed that the differences between adjacent numbers on the scale are equal.

iii Interval Scale

- Quantitative
- Data can be ordered or ranked
- There is no absolute zero point. Zero is only an arbitrary point with which other values can compare
- Difference between two numbers is a meaningful numerical value
- Ration of two numbers is not a meaningful numerical value.

iv Ratio Scale

- Quantitative
- Highest level of measurement
- There exist an absolute zero point (It has a true zero point)
- Ratio between different measurements is meaningful

6.2 Presentation of Data

The sinking of the Titanic is one of the most infamous shipwrecks in history.

On April 15, 1912, during her maiden voyage, the widely considered "unsinkable" RMS Titanic sank after colliding with an iceberg. Unfortunately, there weren't enough lifeboats for everyone onboard, resulting in the death of 1502 out of 2224 passengers and crew

[1]

Here's a quick summary of our variables:

Variable Name	Description
PassengerID	Passenger ID (just a row number, so obviously not useful for prediction)
Survived	Survived (1) or died (0)
Pclass	Passenger class (first, second or third)
Name	Passenger name
Gender	Passenger Gender
Age	Passenger age
SibSp	Number of siblings/spouses aboard
Parch	Number of parents/children aboard
Ticket	Ticket number
Fare	Fare
Cabin	Cabin
Embarked	Port of embarkation (S = Southampton, C = Cherbourg, Q = Queenstown)

6.2.1 Tabular Presentations of Data

Raw Data

- Raw data are collected data that have not been organized numerically
- Eg: Passenger age

```
##   PassengerId Survived Pclass
## 1           1        0      3
## 2           2        1      1
```

[1] Data source: https://www.kaggle.com/varimp/a-mostly-tidyverse-tour-of-the-titanic

```
## 3           3        1      3
## 4           4        1      1
## 5           5        0      3
## 6           6        0      3
##                                                    Name    Sex Age SibSp Parch
## 1                               Braund, Mr. Owen Harris   male  22     1     0
## 2 Cumings, Mrs. John Bradley (Florence Briggs Thayer) female  38     1     0
## 3                                Heikkinen, Miss. Laina female  26     0     0
## 4        Futrelle, Mrs. Jacques Heath (Lily May Peel) female  35     1     0
## 5                              Allen, Mr. William Henry   male  35     0     0
## 6                                      Moran, Mr. James   male  NA     0     0
##             Ticket    Fare Cabin Embarked
## 1        A/5 21171  7.2500              S
## 2         PC 17599 71.2833   C85        C
## 3 STON/O2. 3101282  7.9250              S
## 4           113803 53.1000  C123        S
## 5           373450  8.0500              S
## 6           330877  8.4583              Q
```

```
##  [1] 22 38 26 35 35 NA 54  2 27 14  4 58 20 39 14 55  2 NA 31 NA 35 34 15 28  8
## [26] 38 NA 19 NA NA 40 NA NA 66 28 42 NA 21 18 14
```

An array

- An array is an arrangement of raw numerical data in ascending or descending order of magnitude.
- Eg: Passenger age

```
##  [1]  2  2  4  8 14 14 14 15 18 19 20 21 22 26 27 28 28 31 34 35 35 35 38 38 39
## [26] 40 42 54 55 58 66
```

Frequency Table (Frequency Distributions)

- A frequency table (frequency distribution) is a listing of the values a variable takes in a data set, along with how often (frequently) each value occurs
- frequency can be recorded as a
  - **frequency or count:** the number of times a value occurs, or
  - **percentage frequency:** the percentage of times a value occurs
- Percentage frequency can be calculated as,

$$Percentage frequency = \frac{a}{b} \times 100\%$$

- The objective of constructing a frequency table are as follows
  - to organize the data in a meaningful manner
  - to determine the nature or shape of the distribution
  - to draw charts and graphs for the presentation of data
  - to facilitate computational procedures for measures of average and spread
  - to make comparisons between different data sets
- There are two basic types of frequency tables
  1. Simple frequency tables (Ungrouped frequency distribution)
  2. Grouped frequency distribution

6.2.1.1 Simple frequency table (Ungrouped frequency distribution)

- Each possible value or category is taken as a class
- More suitable for
  - Qualitative variables
  - Discrete variables
- Sometimes construct for continuous variables when there is a small number of possible values between the minimum and maximum.

Examples:

CASE I:

Example 1

The native countries of 56 students from a certain education institute are as follows:

```
##  [1] "SL" "BD" "SL" "SL" "SL" "SL" "IN" "SL" "SL" "SL" "BD" "SL" "SL" "SL" "IN"
## [16] "SL" "SL" "BD" "SL" "SL" "SL" "SL" "SL" "SL" "SL" "SL" "SL" "MD" "SL" "SL"
## [31] "SL" "SL" "SL" "SL" "PK" "MD" "PK" "SL" "SL" "SL" "SL" "SL" "PK" "MD" "SL"
## [46] "SL" "SL" "SL" "SL" "SL" "SL" "SL" "SL" "SL" "MD" "MD"
```

BD- Bangladesh, IN-India, MD-Maldives, PK-Pakistan, SL- Sri Lanka

Construct a frequency table

```
##  Native Country Count Percentage (%)
##      Bangladesh     3          5.357
##           India     2          3.571
##        Maldives     5          8.929
##        Pakistan     3          5.357
##       Sri Lanka    43         76.786
##           Total    56        100.000
```

CASE II:

Example 2

The grades of 30 students for Statistics are as follows:

```
##  [1] "B" "C" "B" "D" "B" "C" "C" "A" "B" "C" "C" "B" "E" "B" "B" "D" "D" "F" "B"
## [20] "D" "D" "A" "B" "A" "B" "C" "E" "A" "A"
```

Construct a frequency table

```
##  Grade Count Percentage (%)
##      A     5         17.241
##      B    10         34.483
##      C     6         20.690
##      D     5         17.241
##      E     2          6.897
##      F     1          3.448
##  Total    29        100.000
```

CASE III:

Example 3

The number of family members of a sample of undergraduates of Batch 19 are as follows:

```
##  [1] 7 5 3 4 5 4 3 6 4 4 5 2 7 4 5 6 4 4 3 5
```

Construct a frequency table

```
## # A tibble: 6 x 3
##   `Number of family members` Count `Percentage (%)`
##                        <dbl> <int>            <dbl>
## 1                          2     1                5
## 2                          3     3               15
## 3                          4     7               35
## 4                          5     5               25
## 5                          6     2               10
## 6                          7     2               10
```

CASE IV:

Example 4

The ages (in years) of a sample of undergraduates of Batch 19 are as follows:

```
##  [1] 21 22 22 23 22 24 24 23 21 22 23 22 22 23 21 21 22 23 22 23
```

Construct a frequency table

```
## # A tibble: 4 x 3
##   `Age (years)` Count `Percentage (%)`
##           <dbl> <int>            <dbl>
## 1            21     4               20
## 2            22     8               40
## 3            23     6               30
## 4            24     2               10
```

6.2.1.2 Grouped frequency distribution

- A grouped frequency distribution (table) is obtained by constructing classes (or intervals) for the data and then listing the corresponding number of values in each interval.
- Suitable for quantitative variables with large number of possible values in the range of data.
- Note that when items have been grouped in this way, their individual values are lost.
- When studying about frequency distributions it is very important to know the meaning of the following terms

i Class intervals

- In a frequency distribution the total range of the observations are divided into a number of classes. Those are called *class intervals*
- Eg: Class intervals: 10-14, 15-19, 20-24, …, 40-44

ii Class limits

- Class limits are the smallest and largest piece of data value that can fall into a given class.
- In the class interval 10-14, the end numbers, 10 and 14, are called class limits
- The smaller number (10) is the *lower class limit*
- The larger number (14) is the *upper class limit*

iii Class boundaries

- Class boundaries are obtained by adding the upper limit of one class interval to the lower limit of the next-higher class interval and dividing by 2.

- Class boundaries are also called **True class limits**
- Class boundaries **should not** *coincide with actual observations*

Class interval	Class boundaries
10 - 14	9.5 – 14.5
15 - 19	14.5 – 19.5
20 - 24	19.5 – 24.5
25 - 29	24.5 – 29.5
30 - 34	29.5 – 34.5
35 - 39	34.5 – 39.5
40 - 44	39.5 – 44.5

iv The size or width of a class interval

- The size or width of a class interval is the difference between the *lower and upper class boundaries*
- It is also referred to as the *class width, class size, or class length*
- Eg: The class width for the class 10-14 is = 14.5-9.5 = 5

v The class mark (Midpoint of the class)

- Midpoint of the class
- Also called as *class midpoint*
- Midpoint of the class = $\frac{\text{Lower limit+Upper limit}}{2}$

or

- Midpoint of the class = $\frac{\text{Lower boundary+Upper boundary}}{2}$

vi Open class intervals

- A class interval that, at least theoretically, has either no upper class limit or no lower class limit indicated is called an *open class interval*
- For example, referring to age groups of individuals, the class interval "65 year and over" is an open class interval

Rules and Practices for constructing grouped frequency tables

- Every data value should be in an interval
- The intervals should be mutually exclusive
- The classes of the distribution must be arrayed in size order.
- The number of classes not less than 5 or not greater than 15 is recommended.

- The following formula is often used to determine the number of classes: If n is the number of observations, then

$$\text{Number of classes} = \sqrt{n}$$

$$\text{Width of the class interval} = \frac{Range}{\sqrt{n}} = \frac{Min - Max}{\sqrt{n}}$$

- Data should be represented within classes having limits which the data can attain
- Classes should be continuous
- By convention, the beginning of the interval is given the appropriate exact value, rather than the end.
  Eg: intervals of 0-49, 50-99,100-149 would be preferred over the intervals 1-50, 51-100, 101-150 etc.
- The number f observations falling into each category or class interval (class frequency) can be easily found using *tally marks*.

Examples:

In a grouped frequency distribution, class intervals can be constructed in different ways

Example 1

Class interval	Number of students
10 - 14	4
15 - 19	5
20 - 24	11
25 - 29	9
30 - 34	6
35 - 39	3
40 - 44	2

Example 2

Salary	Number of employees
0 – 1999	1
2000 – 3999	31
4000 – 5999	18
6000 – 7999	4
8000 – 9999	2

Salary	Number of employees
10000 - 11999	1
12000 – 13999	0
14000 – 15999	0
16000 – 17999	1
18000 -19999	1
20000-21999	1

Salary	Number of employees
0 – 1999	1
2000 – 3999	31
4000 – 5999	18
6000 – 7999	4
8000 – 9999	2
10000 - 15999	1
16000 – 21999	3
Total	60

Example 3

Salary	Number of employees
Less than 2000	1
2000 – 2999	11
3000 – 3999	20
4000 – 5999	18
6000 – 9999	6
Greater than or equal to 10000	4
Total	60

6.2.1.3 Two-way frequency table

- Cross tabulation, Cross classification table, Contingency table, Two-way table
- Display the relationship between two or more qualitative variables (categorical variables (nominal or ordinal))

```
## # A tibble: 2 x 4
##   Survived First Second Third
##   <chr>    <dbl>  <dbl> <dbl>
## 1 died        80     97   372
```

```
## 2 Survived   136     87   119
```

```
## # A tibble: 2 x 4
##   Survived First Second Third
##   <chr>    <dbl>  <dbl> <dbl>
## 1 died      0.37   0.53  0.76
## 2 Survived  0.63   0.47  0.24
```

6.2.2 Graphic Presentations of Data

- A diagram is a visual form for presentation of statistical data.
- The form of the diagram varies according to the nature of the data

6.2.2.1 Describing Qualitative Data

- Bar chart / Pie chart
- Suitable for
  - Qualitative variables (nominal or ordinal)
  - Discrete variables (when the number of bars or number of different values is small)

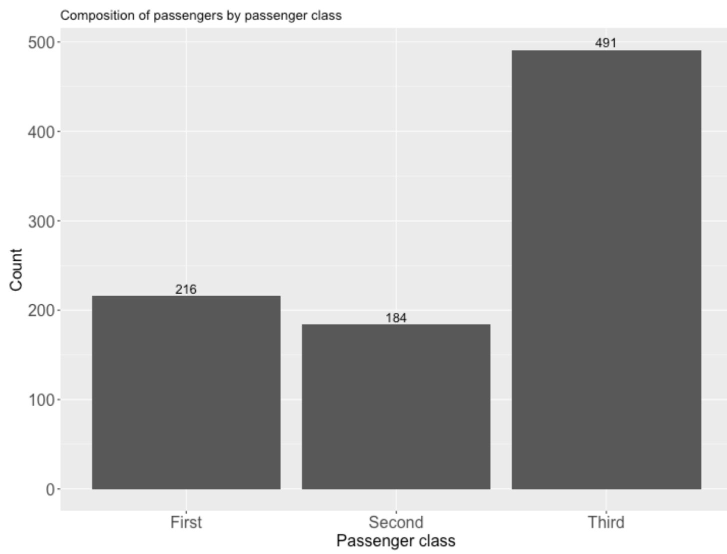
I Bar Chart

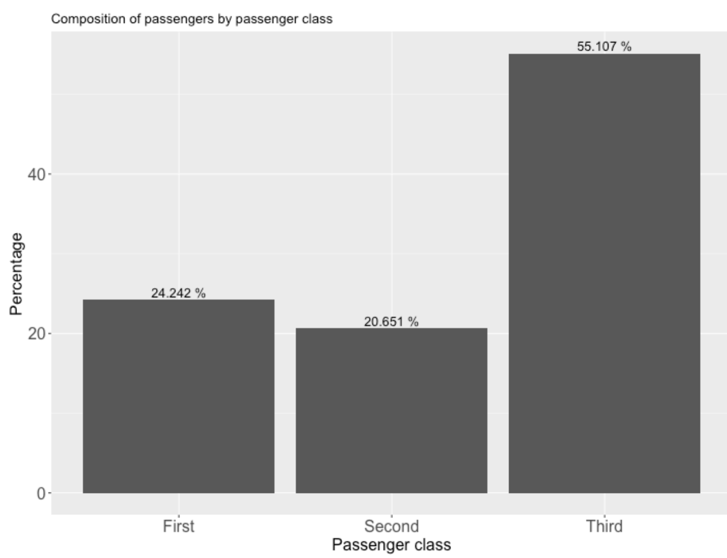
- A bar graph uses bars to represent discrete categories of data
- It can be drawn either on horizontal (more common) or vertical base
- A rectangle of equal width is drawn for each category
- The height (in vertical bar chart) or the length (in horizontal bar chart) of the rectangle is equal to the category's **frequency** or **percentage**.

i Simple Bar Chart

- Only one categorical variable can be presented
- Often used in conjunction with simple frequency tables
- The bars do not touch each other
- The gaps between adjacent bars are same in length

Passenger class	Count	Percentage
First	216	24.242
Second	184	20.651
Third	491	55.107



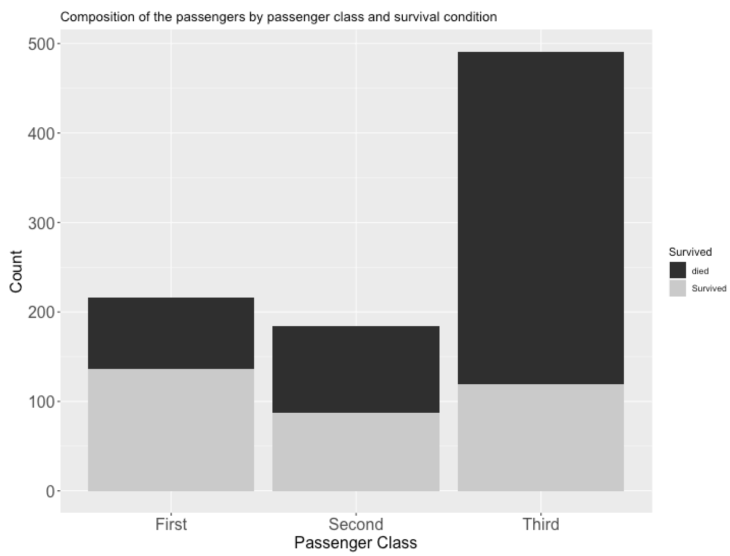


ii Component Bar Chart

- **Sub divided bar chart/ Stacked bar chart**
- Use to compare two or more qualitative variables (nominal or ordinal)
- Often used in conjunction with two way tables
- Start by drawing a simple bar chart with the total figures.
- The bars are then divided into the component parts
- Can be drawn on absolute figures or percentages

- The various components should be kept in the same order in each bar
- To distinguish different components from one another, different colours or shades can be used

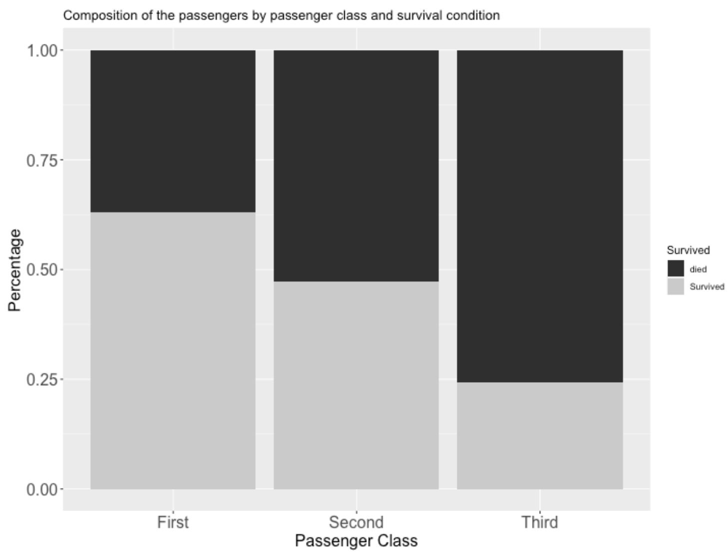
Survived	First	Second	Third
died	80	97	372
Survived	136	87	119



Percentage component bar chart

- When sub-divided bar chart is drawn on percentage basis it is called percentage bar chart
- The various components are expressed as percentage to the total
- All bars are equal in height

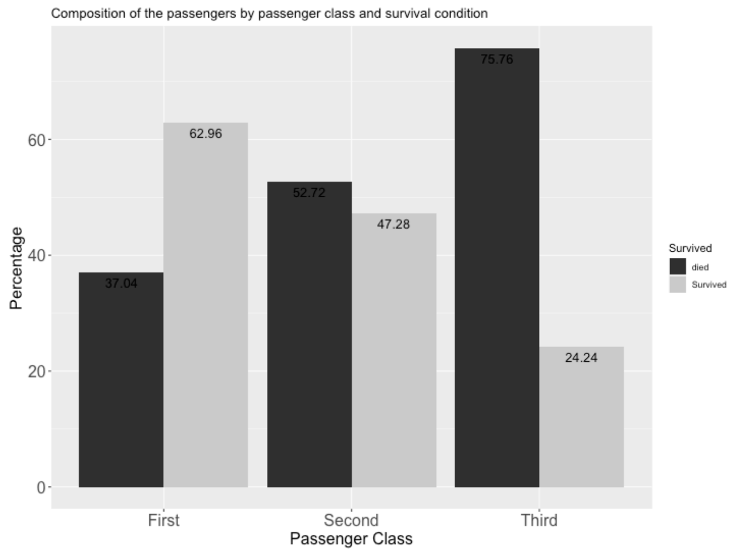
Survived	First	Second	Third
died	0.3703704	0.5271739	0.7576375
Survived	0.6296296	0.4728261	0.2423625



iii Multiple Bar Chart

- Compound bar chart/ Cluster bar chart
- Use to compare two or more qualitative variables (nominal or ordinal)
- Often used in conjunction with two way tables
- These bar charts are drawn side by side

Survived	First	Second	Third
died	37.04	52.72	75.76
Survived	62.96	47.28	24.24



6.2.2.2 Describing Quantitative Data

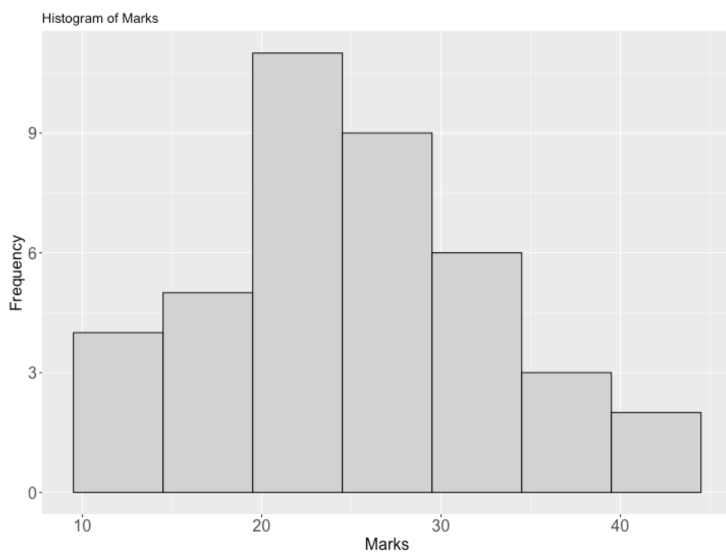
- Histogram/ Dot plot / Box plot/ Scatter plot

II Histogram

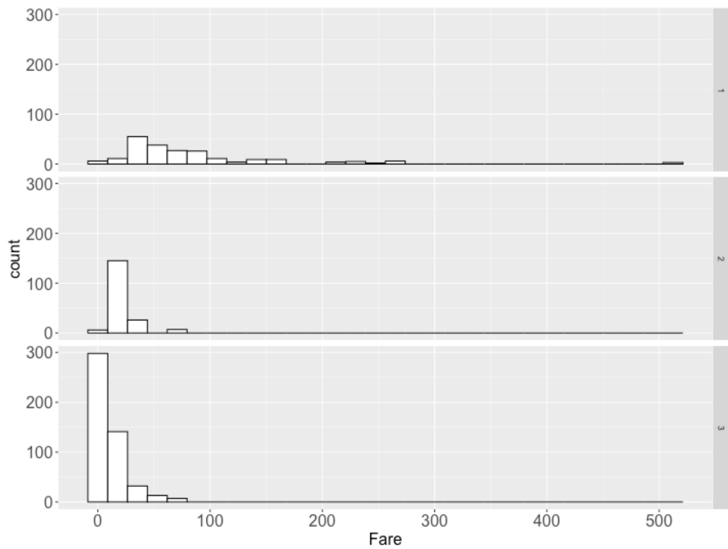
- Histogram looks similar to bar chart since it also has bars.
- However, it is different from a bar chart in a number of aspects.
- One main difference is that in the histogram, the bars are drawn attached to each other; there are no gaps between bars like in a bar chart.
- Histogram is used to show the shape of the distribution of a **continuous variable**.
- However, the histogram is also used for discrete variables when the data are grouped in to class intervals.
- In a histogram, **the area of a bar should be proportional to the frequency of the corresponding class.**
- If all the bars have the same width, then the height of a bar can represent the frequency.
- The bar corresponding to a class interval should be drawn from the lower class boundary to the upper class boundary. In this way there will be no gaps between the bars.

Example: The marks(out of 50) of a group of students are recorded in the accompanying table. Draw a histogram for the data

Marks	Number of students
10 - 14	4
15 - 19	5
20 - 24	11
25 - 29	9
30 - 34	6
35 - 39	3
40 - 44	2
Total	40



Example 2

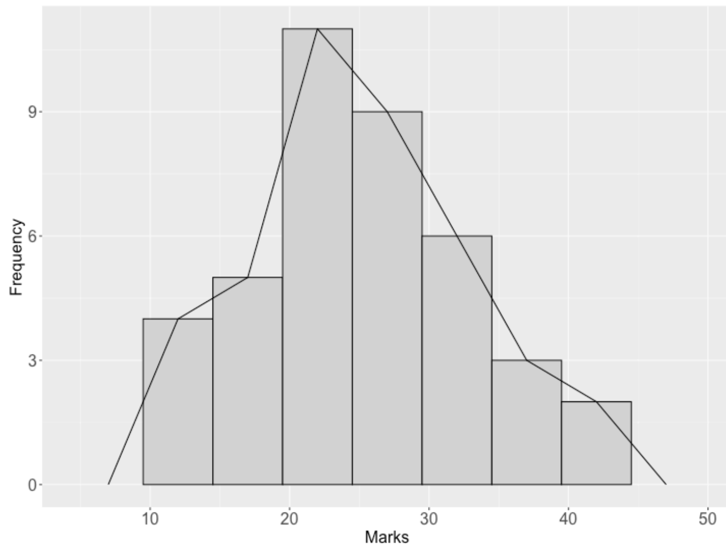


III Frequency polygon

- If the mid-point of the top of each block in a histogram is joined by a straight line, a frequency polygon is produced.
- This is done under the assumption that the frequencies in a class-interval are evenly distributed throughout the class

Example: The marks(out of 50) of a group of students are recorded in the accompanying table. Draw a frequency polygon for the data

Marks	Number of students
10 - 14	4
15 - 19	5
20 - 24	11
25 - 29	9
30 - 34	6
35 - 39	3
40 - 44	2
Total	40

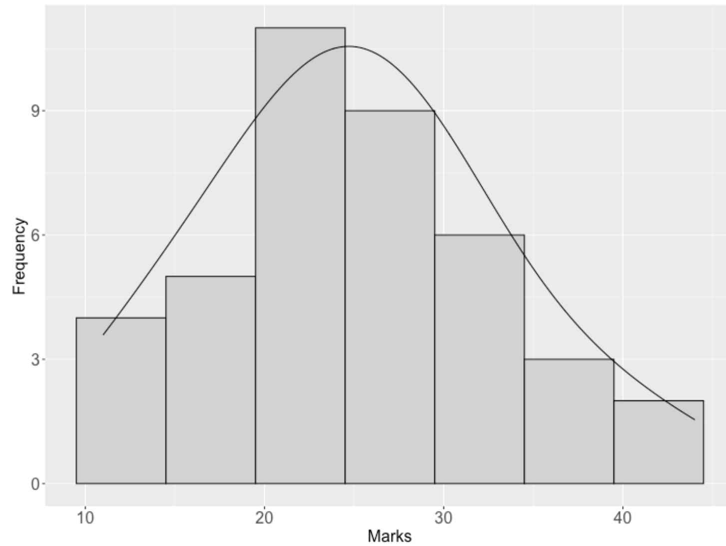


IV Frequency curve

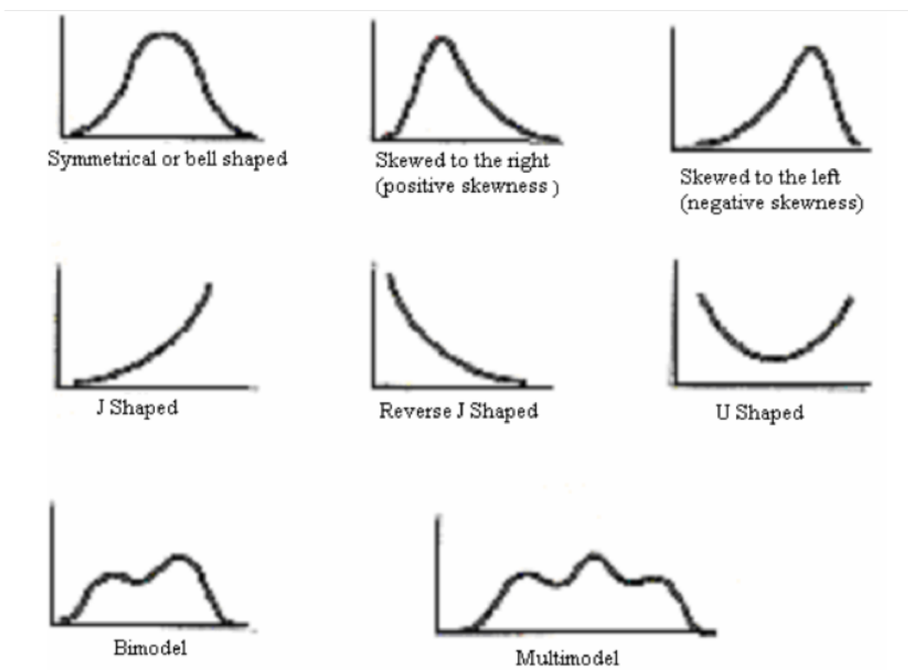
- A frequency curve is drawn by smoothing the frequency polygon.
- It is smooth in such a way that the sharp turns are avoided

Example: The marks(out of 50) of a group of students are recorded in the accompanying table. Draw a frequency curve for the data

Marks	Number of students
10 - 14	4
15 - 19	5
20 - 24	11
25 - 29	9
30 - 34	6
35 - 39	3
40 - 44	2
Total	40



frequency curves arising in practice take on certain characteristics shapes as shown bellow



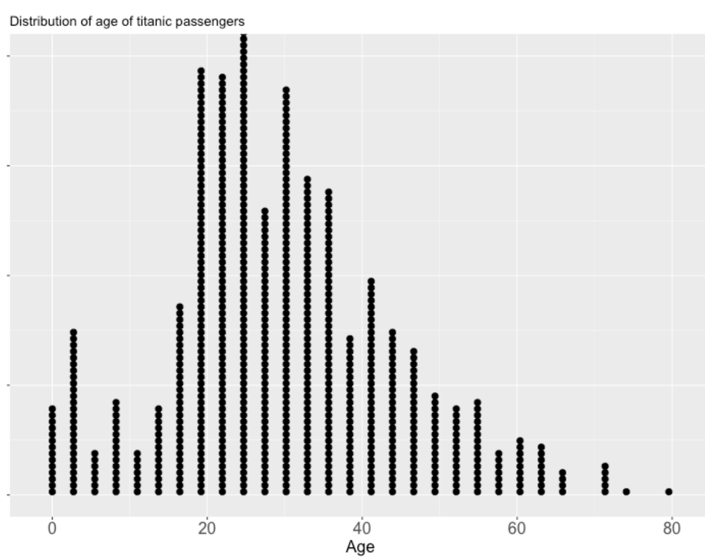
1. The **symmetrical** or **bell shaped** frequency curves are characterized by

the fact that observations equidistant from the central maximum have the same frequency. An important example is the normal curve.

2. In the **moderately asymmetrical** or **skewed** frequency curves the tail of the curve to one side of the central maximum is longer than that to the other. If the longer tail occurs to the right the curve is said to be **skewed to the right** or to have **positive skewness**.While if the reverse is true the curve is said to be **skewed to the left** or to have **negative skewness**.
3. In a **J shaped** or **reverse J shaped** curve a maximum occurs at one end.
4. A **U shaped** frequency curve has maxima at both ends.
5. A **bimodal** frequency curve has two maxima. These appear as two distinct peaks (local maxima) in the frequency curve.When the two modes are unequal the larger mode is known as the major mode and the other as the minor mode.
6. A **multimodal** frequency curve has more than two maxima.

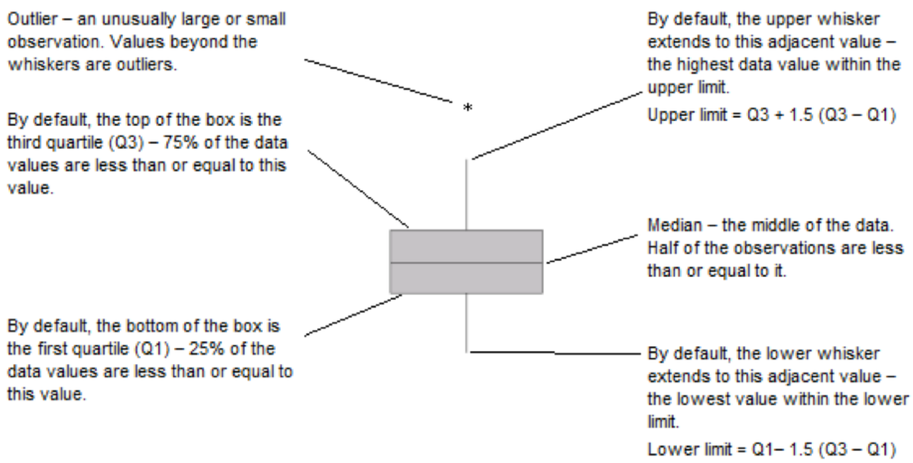
V Dot Plot

- A dot plot is a method of presenting data which gives a rough but rapid visual appreciation of the way in which the data are distributed
- It consists of a horizontal line marked out with divisions of the scale on which the variable is being measured -This graph can be used to represent only the numerical data.



VI Box plot (Box and whisker plot)

- Box plot is also a useful method of representing the behavior of a data set or comparing two or more data sets.
- Box plot is constructed by identifying five statistics from the data set as largest value, smallest values, median, Q1 and Q3.

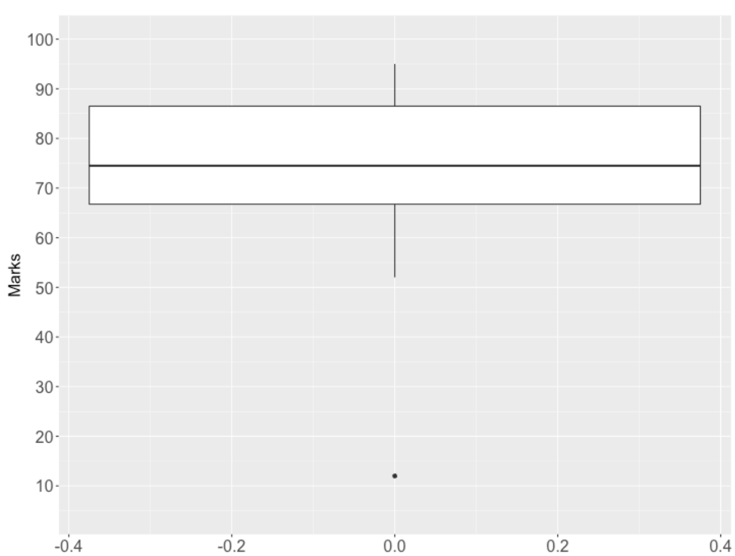


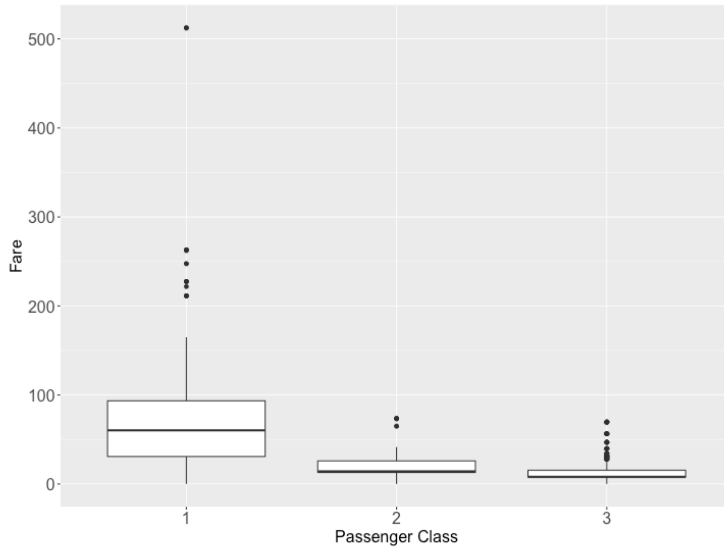
Example:

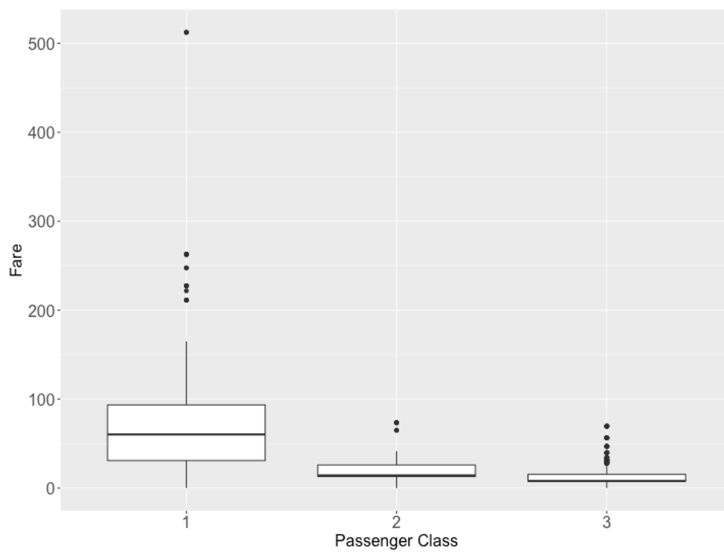
Construct a box plot for the following data set (Marks of students)

52, 88, 56, 79, 72, 91, 85, 88, 68, 63, 76, 73, 86, 95, 12, 69

Xmin = 12 Xmax = 95 Q1 =64.25 Q2 = Median = 74.5 Q3 = 87.5

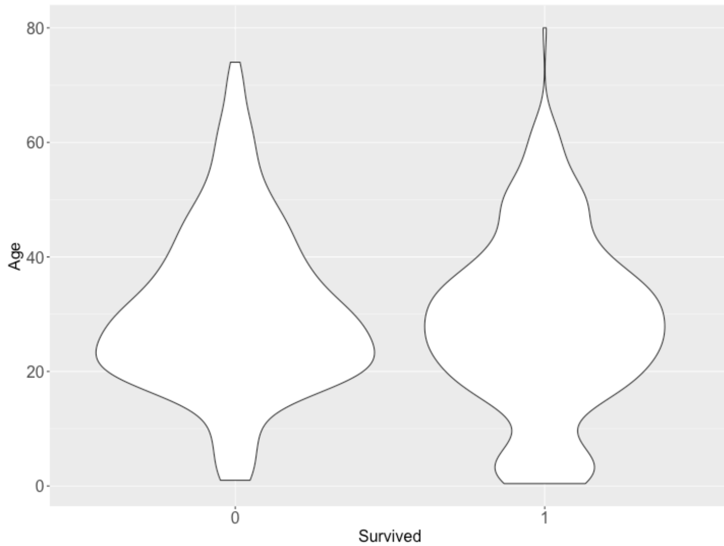

Marks
100
90
80
70
60
50
40
30
20
10
-0.4
-0.2
0.0
0.2
0.4

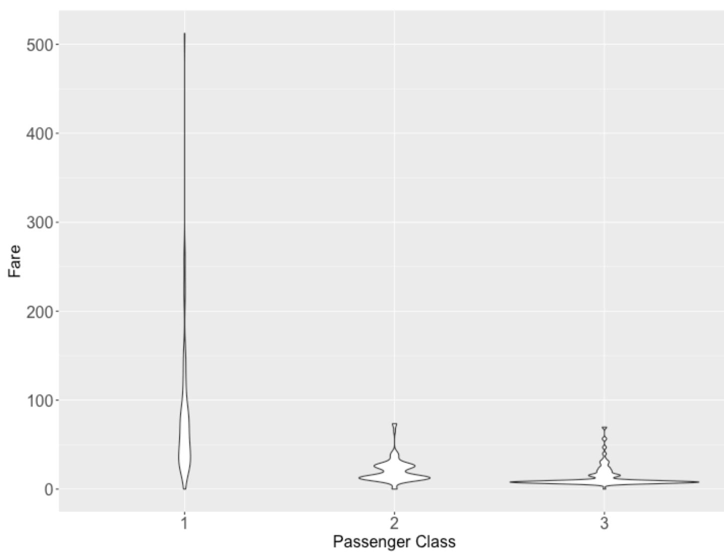




VII Violin plot

- A violin plot is a method of plotting quantitative data.
- It is similar to a box plot, with the addition of a rotated kernel density plot on each side.
- Violin plots are similar to box plots, except that they also show the probability density of the data at different values, usually smoothed by a kernel density estimator.


80
60
40
20
0
Age
0
1
Survived


500
400
300
200
100
0
Fare
1
2
3
Passenger Class

6.3 Summary Measures

- Although frequency distribution serves useful purpose, there are many situations that require other types of data summarization.
- What we need in many instances is the ability to summarize the data by means of a single number called a descriptive measure.
- Descriptive measures may be computed from the data of a sample or the data of a population. To distinguish between them we have the following definitions.

Definitions

- A descriptive measure computed from the data of a sample is called a **statistic**.
- A descriptive measure computed from the data of a population is called a **parameter**.

6.3.1 Measures of Central Tendency

- **Measure of central tendency** yield information about the center, or middle part, of a group of numbers.
- Eg: Mode, Median, Arithmetic Mean, Geometric mean, Harmonic Mean, Quadratic Mean, Quartiles, Deciles, and Percentiles

6.3.1.1 Mode

- The Mode is the most frequently occuring value in a set of data
- Organizing the data into an ordered array (an ordering of the numbers from smallest to largest) helps to locate the mode.
- A series having only one mode is called as **uni-modal**
- In the case of a tie for the most frequent occuring value, two modes are listed. Then the data are set to be **bimodal**
- If a set of data is not exactly bimodal but contains two values that are more dominant than others, some researchers take the liberty of referring to the data set as bimodal even without an *exact tie* for the mode.
- Data sets with more than two modes are referred to as **multimodal**.

- The mode is an appropriate measure of central tendency for nominal-level data.
- The mode can be used to determine which category occurs most frequently.

For ungrouped data

Example 01: Find mode of the following datasets

Dataset 1: 12, 14, 10, 8, 6, 8, 15, 8

Dataset 2: 40, 44, 57, 48, 78

Dataset 3: 42, 45, 55, 50, 45, 40, 55, 45, 52, 55, 54

For grouped frequency data

Example 02: Find mode of the following data

Marks	Number of students
20	8
30	10
40	16
50	8
60	5
70	3

- Advantages and disadvantages of mode

Advantages

- Easy to understand
- Easy to calculate
- Not affected by extreme values in the dataset
- Good for qualitative data

Disadvantages

- Not suitable for further mathematical calculations
- There may be more than one mode for a given dataset
- It is not based upon all the observations
- In some cases, we may not be able to find a mode for a given dataset

6.3.1.2 Median

- The median is the middle value in an ordered array of numbers.
- Median divides the series into equal parts
- The following steps are used to determine the median.
- STEP 1: Arrange the observations in an ordered data array.
- STEP 2: For an **odd number** of terms, find the middle term of the ordered array. It is the median.
- STEP 3: For an **even number** of terms, find the arithmetic mean of the middle two terms. This arithmetic mean is the median.

$$Median = \text{the } (\frac{n+1}{2})\text{th item in the data array}$$

- The level of data measurement must be at least ordinal for a median to be meaningful.

Example 1: Find the median of the dataset 1, 8, 6, 3, 2

Example 2: Find the median of the dataset 8, 9, 1, 2, 14, 12

- Advantages and disadvantages of median

Advantages

- Simple to understand
- Easy to calculate
- Not affected by extreme values in the dataset
- Can be calculated even for qualitative variables (ordinal scale data)

Disadvantages

- It is not based upon all the observations

6.3.1.3 Arithmetic Mean

- The arithmetic mean (usually called mean) is the sum of all observations divided by the total number of observations.

- Population Mean

- The poupulation mean us represented by the Greek letter *mu* (μ).
- Let, N is the number of terms in the population.

$$\mu = \frac{\sum x}{N} = \frac{x_1 + x_2 + x_3 + ... + x_N}{N}$$

- Sample Mean
  - The sample mean is represented by $\bar{x}$
  - Let, n is the number of terms in the sample

$$\bar{x} = \frac{\sum x}{n} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n}$$

- It is inappropriate to use the mean to analyse data that are not at least interval level in measurement.

Example 1: Calculate the mean from the following data

Student	1	2	3	4	5	6	7	8	9	10
Marks	40	50	53	78	58	60	73	35	43	48

- Advantages and disadvantages of arithmetic mean

Advantages

- Simple to understand
- Easy to calculate
- Based on all the observations
- Well defined
- Unique
- Can be used in further calculation

Disadvantages

- Can be affected by extreme values in the dataset
- May lead to false conclusions
- Only applicable to quantitative data (not applicable to qualitative data)

Empirical relationship between mean, mode, median

- In case of symmetrical distribution, mean, median and mode coincide ($mean = meadian = mode$)

- For a moderately asymmetrical distribution, the following relationship exists $Mean–Mode = 3(Mean - Median)$

Choice between mean and median

- Mean is very sensitive to outliers.Median is not sensitive to outliers
- When there are outliers in a data set, median is more appropriate than mean

6.3.1.4 Quartiles, Deciles and Percentiles

- Median divides the data set into two equal parts.
- There are other values which divide the data set into a number of equal parts
- Those are Quartiles, Deciles and Percentiles

(a) Quartiles (Q) – Quartiles divide an array into four equal parts

$$Q_i = \text{the } \frac{i}{4}(n+1)\text{th item in the data array}$$

(b) Deciles (D) – Deciles divide an array into ten equal parts

$$D_i = \text{the } \frac{i}{10}(n+1)\text{th item in the data array}$$

(c) Percentiles (P) – Percentiles divide an array into 100 equal parts

$$P_i = \text{the } \frac{i}{100}(n+1)\text{th item in the data array}$$

6.3.2 Measures of Variability

- Measure of central tendency yield information about particular points of a data set.

- However, business researchers can use another group of analytic tools to describe a set of data.

- These tools are measures of variability, which describe the spread or the dispersion of a set of data.

- Using measures of variability in conjunction with measures of central tendency makes possible a more complete numerical description of the data.

- This section focuses on seven measures of variability for ungrouped data: range, interquartile range, variance, standard deviation, z score and coefficient of variation.

6.3.2.1 Range

- The range is the difference between the largest value of a data set and the smallest value.

$$Range = Maximum - Minimum$$

- One important use of the range is in quality assurance, where the range is used to construct control charts
- Advantages and disadvantages of range

Advantages

- Easy to understand and calculate

Disadvantages

- Consider only the highest and lowest values of the data and fails to take account of any other observations in the dataset
- Heavily influenced by extreme values

6.3.2.2 Interquartile Range (IQR)

- We use the interquartile range (IQR) to measure the spread of a data around the median (M).
- The interquartile range is the range of values between the first and third quartile.
- Essentially it is the range of the middle 50% of the data and is determined by computing the value of $Q_3 - Q_1$.
- The interquartile range is especially useful in situations where data users are more interested in values towards the middle and less interested in extremes.
- The interquartile range is used in the construction of box and whisker plots.
- By eliminating the lowest 25% and the highest 25% of the items in a series, we are left with the central 50% , which are ordinarily free of extreme values.

Advantages

- Easy to understand and calculate
- Not influenced by extreme values

Disadvantages

- Ignore the first 25% and the last 25% in the dataset

6.3.2.3 Variance and Standard Deviation

- To measure the spread of data around the mean, we use the standard deviation (S).
- The variance and standard deviation are two very popular measures of dispersion.
- These measures are not meaningful unless the data are at least interval-level data.
- Their formulations are categorized into whether to evaluate from a population or from a sample.

NOTE

- Sum of deviations from the arithmetic mean is always zero.

$$\sum (x - \mu) = 0$$

- This property requires considering the alternative ways to obtain measure of variability.

6.3.2.3.1 Variance

- The **variance** is the average of the squared deviations about the mean for a set of numbers.
- The population variance is denoted by σ^2

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N}$$

- *The sum of the squared deviations about the mean of a set of values* - called the **sum of squares of** x and sometimes abbreviated as SS_x
- Because the variance is computed from squared deviations, the final result is expressed in terms of squared units of measurements.
- Statistics measured in squared units are problematic to interpret.

6.3.2.3.2 Standard Deviation

- The standard deviation is the square root of the variance.
- The population standard deviation is denoted by σ

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}} = \sqrt{\sigma^2}$$

- One feature of standard deviation that distinguishes it from a variance is that the standard deviation is expressed in the same units as the raw data, whereas the variance is expressed in those units squared.

Advantages

- Based on all the observations
- Since this is based on arithmetic mean, it has all the merits of it
- The most important and widely used measure of dispersion

Disadvantages

- Not easy to understand and difficult to calculate
- Gives more weight to extreme values, because the values are squared up

6.3.2.4 Empirical Rule

- The empirical rule is an important rule of thumb that is used to state the approximate percentage of values that lie within a given number of standard deviations from the mean of a set of data **if the data are normally distributed**
- The empirical rule is used only for three numbers of standard deviations: 1σ, 2σ, 3σ

Distance from the mean	Values within distance
$\mu \pm 1\sigma$	68%
$\mu \pm 2\sigma$	95%
$\mu \pm 3\sigma$	99.7%

- If a set of data is normally distributed, or bell shaped, approximately 68% of the data values are within one standard deviation of the mean, 95% are within two standard deviations, and almost 100% are within three standard deviations.

6.3.2.5 Population versus sample variance and standard deviations

- The sample variance is denoted by s^2 and the sample standard deviation by s.
- The main use for sample variances and standard deviations is as estimators of population variances and standard deviations.
- Thus, computation of the sample variance and standard deviation differs slightly from computation of the population variance and standard deviation.
- Both the sample variance and sample standard deviation use $n - 1$int he denominator instead of n because using n in the denominator of a sample variance results in a statistic that tends to underestimate the population variance.
- While discussion of the properties of *good estimator* is beyond the scope of this course, one of the properties of a good estimator is being *unbiased.*
- Whereas, using n in the denominator of the sample variance makes it a *biased* estimator, using $n-1$ allows it to be an *unbiased* estimator, which is a desirable property in inferential statistics.

Sample variance

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

Sample standard deviation

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = \sqrt{s^2}$$

6.3.2.6 Computational formulas for variance and standard deviation

- An alternative method of computing variance and standard deviation, sometimes referred to as the computational method or shortcut method, is available.
- Algebraically,

$$\sum (x - \mu)^2 = \sum x^2 - \frac{(\sum x)^2}{N}$$

and

$$\sum (x - \bar{x})^2 = \sum x^2 - \frac{(\sum x)^2}{n}$$

- Substituting these equivalent expressions into the original formulas for variance and standard deviation yields the following computational formulas.

Computational formula for population variance and standard deviation

$$\sigma^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N}$$

$$\sigma = \sqrt{\sigma^2}$$

Computational formula for sample variance and standard deviation

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n - 1}$$

$$s = \sqrt{s^2}$$

- For situations in which the mean is already computed or is given, alternative forms of these formulas are:

$$\sigma^2 = \frac{\sum x^2 - N\mu^2}{N}$$

$$s^2 = \frac{\sum x^2 - n(\bar{x})^2}{n-1}$$

6.3.2.7 Coefficient of variation

- In general, for two variables measured with he same units (eg: two groups of people both weighed in kg), the group with the larger variance and standard deviation has n=more variability among their scores.
- The unit of measure affects the size od the variance.
- The same population weights, expressed in 'grams' rather than kg would have a larger variance and standard deviation.
- The *coefficient of variation*, a measure of relative variability gets around this difficulty and makes it possible to compare variability across variables measured in different units.
- The coefficient of variation is the ratio of the standard deviation to the mean, expressed as a percentage and is denoted CV.

$$CV = \frac{\sigma}{\mu}(100)$$

Using median and quartile deviation

$$CV = \frac{\frac{Q_3 - Q_1}{2}}{Median}(100)$$

- The coefficient of variation essentially is a relative comparison of a standard deviation to its mean.
- The coefficient of variation can be useful in comparing standard deviations that have been computed from data with different means.

- The choice of whether to use a coefficient of variation or raw standard deviations to compare multiple standard deviations is a matter of preference
- The coefficient of variation also provides and optional method of interpreting the value of a standard deviation.

6.3.3 Measures of Shape

- Measures of shape are tools that can be used to describe the shape of a distribution of data.
- In this section, we examine two measures of shapes: skewness and Kurtosis.

6.3.3.1 Skewness

- A distribution of data in which the right half is a mirror image of the left half is said to be *symmetrical.*
- One example of a symmetrical distribution is the normal distribution, or bell curve.
- **Skewness** is a measure of symmetry, or more precisely, the lack of symmetry
- The measures of asymmetry are called as measures of skewness.
- The skewed portion is the long, thin part of the curve

Skewness and the relationship of the mean, median and mode

- The concept of skewness helps us to understand the relationship between the mean, median and mode.
- In a unimodal distribution (distribution with a single peak or mode) that is skewed, the mode is the apex (high point) of the curve and the median is the middle value.
- The mean tends to be located toward the tail of the distribution, because the mean is affected by all values, including the extreme ones.
- A bell-shaped or normal distribution with the mean, median and mode all at the centre of the distribution has no skewness.

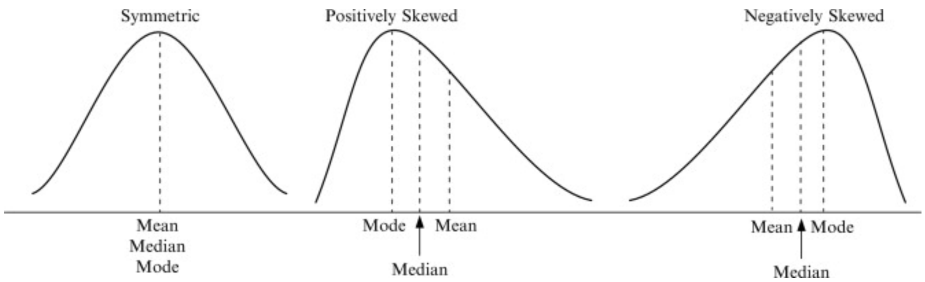


Figure 6.1: Relationship of mean, median and mode for different types of skewness

6.3.3.1.1 Pearsonian coefficient of skewness

- This coefficient compares the mean and median in light of the magnitude of the standard deviation

$$S_k = \frac{3(\mu - M_d)}{\sigma}$$

where S_k = coefficient of skewness, M_d = median

- Note that if the distribution is symmetrical, the mean and median are the same value and hence the coefficient of skewness is equal to zero.
- If the value of S_k is positive, the distribution is positively skewed.
- If the value of S_k is negative, the distribution is negatively skewed.
- The greater the magnitude of S_k, the more skewed is the distribution.

6.3.3.2 Kurtosis

- Kurtosis describes the amount of peakedness of a distribution.
- **Kurtosis** is a measure of whether the data are peaked or flat relative to a normal distribution
- Distributions that are high and thin are referred to as **leptokurtic** distributions.
- Distributions that are flat and spread out are referred to as **platykurtic** distributions.
- Between the above two types are distributions that are more ‘normal’ in shape, referred to as **mesokurtic** distributions

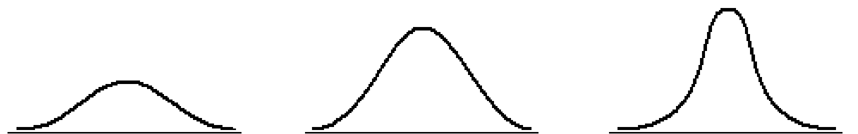


Figure 6.2: Types of kurtosis

6.3.4 Measures of Association

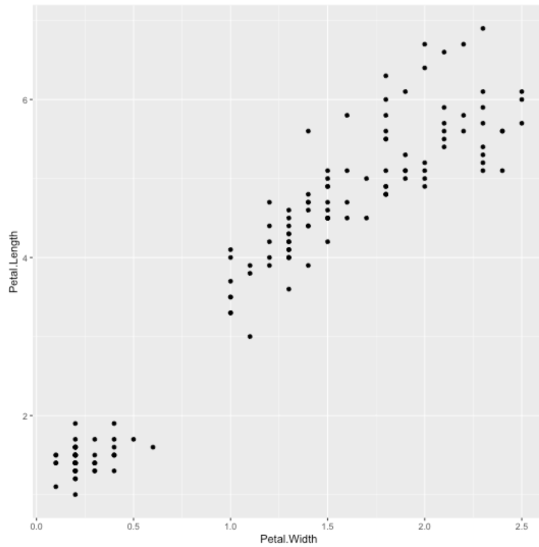
- Many times in business it is important to explore the relationship between two numerical variables
- Measures of association are statistics that yeild information about the relatedness of numerical variables.
- In this section, we discuss only one measure of association, correlation, and do so only for two numerical variables.

6.3.4.1 Scatter plot

- A scatter plot is a two dimensional graph of pairs of points from **two numerical** variables
- In a quantitative bi-variate dataset, we have a (x, y) pair for each sampling unit, where x denotes the independent variable and y denotes the dependent variable.
- Each (x, y) pair can be considered as a point on the Cartesian plan.
- Scatter plot is a plot of all the (x, y) pairs in the dataset.
- The purpose of scatter plot is to illustrate any relationship between two quantitative variables.
  - If the variables are related, what kind of relationship it is, linear or nonlinear?
  - If the relationship is linear, the scatter plot will show whether it is negative or positive.

Example: *Edgar Anderson's Iris Data*

This famous (Fisher's or Anderson's) iris data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are Iris setosa, versicolor, and virginica.



6.3.4.2 Correlation

- Correlation is a measure of the degree of relatedness of two or more variables.
- Several measures of correlation are available , the selection of which depends mostly on the level of data being analysed.
- Ideally, researchers would like to calculate ρ, the **population** coefficient of correlation.
- However, because researchers virtually always deal with sample data, this section introduces a widely used sample coefficient of correlation, r.
- This measure is applicable **only if both variables being analysed have at least an interval level of data**

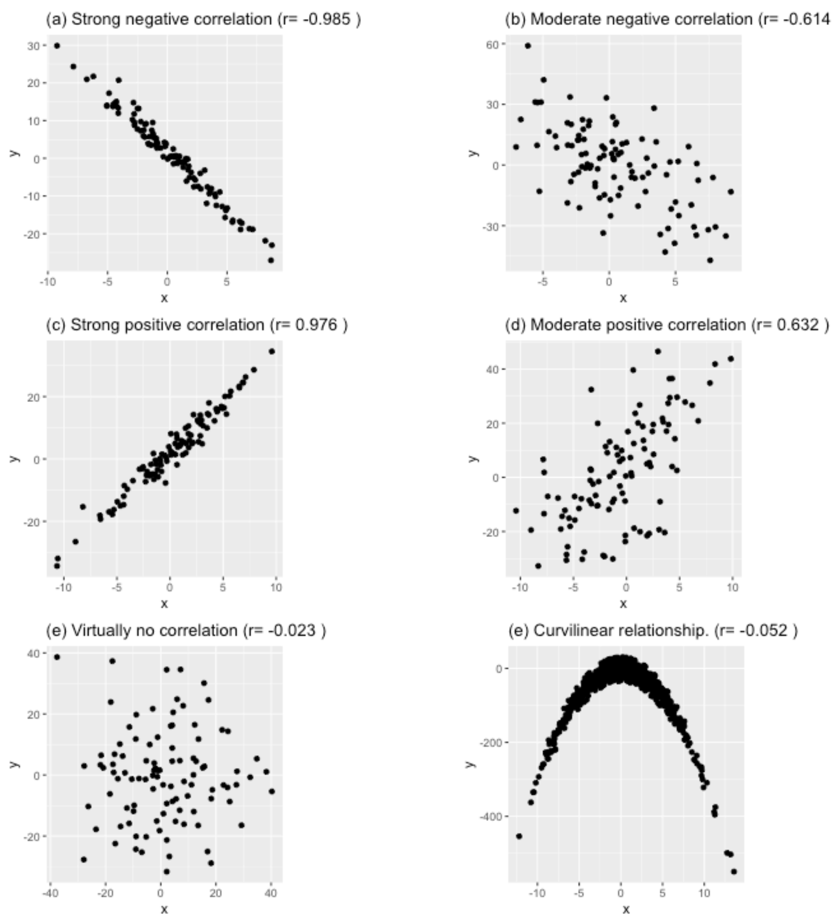
Pearson product-moment correlation coefficient (r)

- The statistic r is the Pearson product-moment correlation coefficient, named after Karl Pearson (1857 - 1936).
- The tern r is a measure of the **linear** correlation of two variables.
- It is a number that ranges from -1 to 0 to +1, representing th strength of the linear relationship between the variables.
- An r value of $+1$ denotes a perfect **linear** positive relationship between two variables.
- An r value of -1 denotes a perfect **linear** negative relationship between two variables, which indicates an inverse relationship between two variables: as one variable gets larger, the other gets smaller.
- An r value of 0 means no **linear** relationship is present between the two variables.

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}}$$

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

- Examples: Following figure shows five different degrees of correlation:



NOTE

- When $r = 0$, it signifies there is **no linear** relationship between the two variables. (There can be a non-linear relationship, Figure (e))
- Figure (e): There is a very strong curvilinear relationship. But there is **no linear** relationship.

6.4 Tutorial

Chapter 6: Descriptive Statistics

1. A power company in Sri Lanka designs and manufactures power distribution switchboards for hospitals, bridges, airports, highways and water treatment plants. Power company marketing director wants to determine client satisfaction with their products and services. He developed a questionnaire that yields a satisfaction score between 0 and 100 for participant responses. A random sample of 50 of the company's 1000 clients is asked to complete a satisfaction survey. The satisfaction scores for the 50 participants are averaged to produce a mean satisfaction score.

a. What is the objective of this study?
b. What is the population for this study?
c. What is the sample for this study?
d. What is the statistic for this study?
e. What would be a parameter for this study?

2. Determine the type and the scale of measurement of the following variables

a. The time required to produce an item on an assembly line
b. the number of litres of milk a family drinks in a week
c. The ranking of 50 students in your class after their overall performance have been designated as excellent, good, satisfactory or poor
d. The telephone area code of clients in Australia
e. The age of each of your batch mates
f. The sales at the local pizza restaurant each month
g. A student index number
h. The response time of emergency services
i. The number of tickets sold at a ticket counter on any given day
j. Monthly maximum air temperature

3. The grades of 30 students for Statistics are as follows:

B, C, B, D, B, C, C, A , B, C,

C, B, E, B, B, D, D, F, B, D,

D, A, B, A, B, C, E, A, A, E

a. Construct a frequency table with suitable values (eg: absolute frequency, relative frequency, cumulative frequency, cumulative relative frequency)

4. The gender type of a group of 10 students are as follows:

Male, Female, Female, Male, Female Female, Male, Male, Female, Female

a. Construct a frequency table with suitable values (eg: absolute frequency, relative frequency, cumulative frequency, cumulative relative frequency)

5. In a study conducted to investigate the relationship between delivery time and computer-assisted ordering, the following data were gathered from a sample of 40 firms, that 16 use computer-assisted ordering, while 24 do not. Furthermore, past data are used to categorize each firm's delivery times below the industry average, equal to the industry average, or above the industry average. The results obtained are given in the table below.

Computer Assisted Ordering	Delivery Time Below Industry Average	Delivery Time Equal to Industry Average	Delivery Time Above industry Average	Row Total
No	4	12	8	24
Yes	10	4	2	16
Total	14	16	10	40

a) For each row and column total, calculate the corresponding row or column percentage.
b) For each cell, calculate the corresponding cell, row, and column percentages.
c) Carry out graphical analysis to investigate the relationship between delivery-time performance and computer-assisted ordering.
d) What conclusions can be made about the nature of the relationship?

6.

a) Giving suitable examples distinguish between multiple bar chart and a component bar chart.
b) Write down advantages and disadvantages of each chart and when they should be used

7. In a study conducted to investigate the effect of wearing helmets in riding motorcycles to reduce the head injuries, the following data were gathered by investigating 151 motorcycle accidents reported during last 10 month period. Out of the total number of cases (151) investigated, only 102 people were wearing helmets properly at the time of the accidents. Out of them, only 9 got severe head injuries. The rest got either minor head injuries or no head injuries at all. Among 49 who were not wearing helmets at the time of the accident, 15 got server head injuries and the rest got minor head injuries or no head injuries at all.

a) Use an appropriate table to represent the above data and state what conclusions can be drawn from it

b) Carry out a suitable graphical analysis to investigate the effect of wearing helmets in riding motorcycles to reduce the head injuries.

8. The table below shows a frequency distribution of the weekly wages of 65 employees at the ABC Company. With reference to this table,

Wages (in $)	Number of Employees
250.00 - 259.99	8
260.00 - 269.99	10
270.00 - 279.99	16
280.00 - 289.99	14
290.00 - 299.99	10
300.00 - 309.99	5
310.00 - 319.99	2
Total	65

a) Construct a histogram
b) Construct a frequency polygon
c) Construct a frequency curve
d) State what conclusions can be drawn from the above graphical analysis

9. The contribution of the agriculture, industrial, and service sector to the Gross Domestic Product (GDP) for each province in Sri Lanka is given in the table below. It is required to investigate whether the contributions are varying from province to province. What kind of graph you would suggest for representing data to serve the purpose. Justify your answer. Sketch the proposed graph.

Sector	Western	Southern	Sabaragamuwa	Central	Uva	Eastern	North Western	North Central	Northern	Total
Agriculture	29767	58761	39209	60373	45689	27709	45118	40320	15851	362797
Industry	341636	41762	39163	34294	13183	25626	57483	10887	4986	569020
Services	695094	86593	54828	82575	33956	47122	84365	38485	43168	1166187
GDP	1066495	187115	133199	177241	92827	100457	186964	89691	64004	2098004

10. Marks of 16 students are given below.

52, 88, 56, 79, 72, 91, 85, 88, 68, 63, 76, 73, 86, 95, 12, 69

(a) Find the quartile of the distribution and interpret the values

(b) Construct a box plot for the data set
(c) Are there outliers in the data set?

11. Listed below, ordered from smallest to largest, is the time in days the customers take to pay their invoices.

$$13, 13, 13, 20, 26, 27, 31, 34, 34, 34, 35, 35, 36, 37, 38, 41, 41, 41, 45, 47, 47, 47, 50, 51, 53, 54, 56, 62, 67, 82$$

(a) Determine the median
(b) Determine the first and third quartiles
(c) Determine the 2nd decile and the 8th decile
(d) Determine the 67th percentile

12.

(a) What are outliers?
(b) Which of the following is/are unaffected by outliers? Underline the correct answer/ answers.

a. Mean
b. Median
c. Mode
d. Range
e. Standard deviation
f. Inter-quartile range

(c) Some summary measures of a variable is given below. Descriptive Statistics:

N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
50	5.389	0.997	3.150	4.658	5.350	6.020	8.600

(d) Are there outliers in this data set? YES/NO Justify your answer using a box plot

13. The mean and the standard deviation of 250 observations of a variable X are 62.1 and 4.3 respectively. However, in a re-scrutinizing process, it was found that the observations 72 and 81 were incorrectly recorded as 92 and 87. Find the correct mean and standard deviation of the data set.

14. The weekly sales from a sample of AB company were organized into a frequency distribution. The mean of weekly sales was computed to be \$105,900, the median \$105,000, and the mode \$104,500

(a) Sketch the sales in the form of a smoothed frequency polygon. Note the location of the mean, median and mode on the X-axis
(b) Is the distribution symmetrical, positively skewed, or negatively skewed? Explain.

15. Compare the variation of the annual incomes of executives with the variation of the incomes of unskilled employees. The sample information is given below.

Type	$\bar{x}$	SD
Executives	500,000	50,000
Unskilled employees	22,000	2,200

16. Marks for the course module AB and the corresponding lecture attendance of 10 students are given below

Student Number	1	2	3	4	5	6	7	8	9	10
Marks in course module AB	84	51	91	60	68	89	98	58	53	47
Lecture attendance	13	6	15	4	12	14	15	10	11	6

(a) Draw a scatter plot for the above data
(b) Calculate the coefficient of correlation
(c) Comment on the relationship between the marks and the lecture attendance

References

Black, K., Asafu-Adjaye, J., Khan, N., Perera, N., Edwards, P., & Harris, M. (2007). *Australasian business statistics.* John Wiley & Sons.

Spiegel, M. R., & Stephens, L. J. (2017). Schaum's outline of statistics. McGraw Hill Professional.

Chapter 7

Sets and Relations

7.1 Introduction

- Everyday we hear, or make statements to convey ideas about possibilities of occurring events.
  - Which lottery gives me the highest **chance** of winning?
  - What are **odds** of our cricket team winning the next world cup?
  - What is the **chance** of getting an A+ for CM1110?
- In order to make decisions or take actions, we need to **quantify the possibility** of occurrence of an event. For this, we use the term **probability**.
- Probability is a **quantitative measure** of uncertainty, a **number** that conveys the strength of our belief in the occurrence of an uncertain event.
- The likelihood of an outcome is quantified by assigning a number from the interval [0, 1] to the outcome (or a percentage from 0 to 100%).
- Higher numbers indicate that the outcome is more likely than lower numbers.
- A probability of 0 indicates an outcome will not occur.
- A probability of 1 indicates that an outcome will occur with certainty.

7.2 Sample Spaces and Events

7.2.1 Random Experiments

- Statisticians use the word **experiment** to describe any process that generates a set of data.
- An experiment usually involves observing or counting or measuring. For example;

- we might classify items coming off an assembly line as defective or non-defective
- we may record the number of accidents that occur monthly at the Katubedda junction, hoping to justify the installation of a traffic light;
- or we may be interested in the volume of gas released in a chemical reaction when the concentration of an acid is varied.

- An experiment that is repeatable in an identical fashion is called a **random experiment** or a **probability experiment**.
- Random experiment is an activity that leads to the occurrence of one of several possible outcomes which is not likely to be known until its completion, that is, the outcome is not perfectly predictable. This process has the properties that,
  - All possible outcomes can be specified in advance.
  - It can be repeated in an **identical manner.**
  - The same outcome may not occur on various repetitions so that the actual outcome is not known in advance.

Examples

- Tossing a coin.
- Throwing a die (or several dice)
- Drawing cards from a pack
- Launching of missile and observing its velocity at specified times.

7.2.2 Outcomes

- The result of a **single experiment** is called an outcome.
- It can be instrument reading, counts, "yes" or "no" answers, values obtained through extensive calculations etc.
- To model and analyze a random experiment, we must understand the set of possible outcomes from the experiment.

7.2.3 Sample space

- The set of all possible outcomes of a random experiment is called the **sample space** of the experiment.
- We denote the sample space by Ω.
- Each outcome in a sample space is called an **element** of the sample space or simply a **sample point**.

Activity 01 *Write down the corresponding sample space of the following experiments?*

Example 01: A player rolls a die.

Example 02: A player rolls two dice.

Example 03: A quality control engineer of a certain company takes a random sample of 4 products as the products come off its production line and counts the number of defective products in the sample.

Example 04: A quality control engineer of a certain company takes a random sample of 4 products as products come off its production line and determines whether each product is defective or non-defective

Example 05: A company produces a certain type of car batteries. If a battery has a voltage that is outside certain limits, that battery is characterized as bad (B); if the battery has a voltage within the prescribed limits it is characterized as good (G). A quality inspector selects batteries one by one as they come off the production line and tests them until he gets a bad battery. He uses this data to estimate the proportion of bad batteries produced by the company.

Example 06: An experiment consists of measuring the lifetime of a bulb.

- A sample space is often defined based on the objectives of the analysis.
- The following example illustrates several alternatives

Example:

Consider an experiment that selects a digital camera and records the recycle time of a flash (the time taken to ready the camera for another flash). The possible values for recycle time can vary according to the resolution of the timer and on the minimum and maximum recycle times.

However, since the time is positive we can define the sample space as simply the positive real line

Case I

Case II If it is known that all recycle times are between 1.0 and 5.5 seconds, the sample space can be written as

Case III If the objective of the analysis is to consider only whether the recycle time is low, medium or high, the sample space can be written as a set of three outcomes

Case IV If the objective is only to evaluate whether or not a particular camera confirms to a minimum recycle time specification, the sample space can be written as a set of two outcomes

- The number of outcomes in the sample space can be **finite** or **infinite**.
- Infinite sample spaces can be **countable** or **uncountable**.
- A sample space is countable if the outcomes can be associated with the integers $1, 2, 3, ...$
- If the sample space contains a finite number of elements or an infinite though countable number of elements, it is said to be **discrete**.
- The outcomes of some experiments are neither finite nor countably infinite, it is said to be **continuous**.

- In the above Example, the choice $\Omega = \{x|x \geq 0\}$ is an example of a continuous sample space,
- $\Omega = \{yes, no\}$ is a discrete sample space.
- As mentioned, the best choice of a sample space depends on the objectives of the study.

7.2.4 Events as subsets of sample spaces

- **An event is a subset of the sample space of a random experiment**
- Events are usually denoted by capital English letters A, B, C, D, E etc.
- An event containing only one outcome is called a **simple event**.
- An event containing more than one outcome is called a **compound event**.

Activity 02 *Express the following events as subsets of the sample spaces. Decide whether each event is simple or compound. (No need to calculate probabilities yet.)*

Example 07: A player rolls a die.

a) A = the event of getting 3
b) B = the event of getting an even number
c) C = the event of getting a number less than 4
d) D = the event of getting 2 or 5

Example 08: A player rolls two dice.

a) A = The event that one number is even.
b) B = The event that sum of the two numbers is odd
c) C = The event that the numbers are equal
d) D = The event of getting 2 and 5
e) E = The event of getting 2 or 5

Example 09: Consider Example 03 in Activity 01,

a) E = the event that sample contains more than 2 defective products.
b) F = the event that the sample contains only two defective products.

Example 10: Consider Example 04 in Activity 01

a) G = the event that sample contains more than 2 defective products.
b) H = the event that the sample contains two defective products.

Example 11: Consider Example 05 in Activity 01,

a) I = the event that number of good batteries before a bad battery is 3.
b) J = the event that the number of good batteries before a bad battery is less than 3.

Example 12: Consider Example 06 in Activity 01,

a) A = the event that bulb burns for at least 25 hours but burns out before 50 hours.

7.3 Event Operations

- We can describe new events from combinations of existing events.
- Because events are subsets, we can use basic set operations to form other events of interest.
- Some of the basic set operations are summarized here in terms of events:

Venn diagrams

- Diagrams are often used to portray relationships between sets, and these diagrams are also used to describe relationships between events.
- We can use Venn diagrams to represent a sample space and events in a sample space.

7.3.1 Union

- Let A and B be two events.
- The set of outcomes that belong to **either only A or only B or to both A and B,** is called the **union** of the two events.
- It is denoted by $A \cup B$.

$$A \cup B = \{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\}$$

- Shaded set in the Venn diagram represents the event $A \cup B$

7.3.2 Intersection

- Let A and B be two sets.
- The set of outcomes that belong to **sets A and B both** is called the intersection of the two events.

- It is denoted by $A \cap B$.

$$A \cap B = \{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}$$

- Shaded set in the Venn diagram represents the event $A \cap B$

7.3.2.1 Mutually exclusive events

- If the events A and B cannot occur together, in other words if $A \cap B$ cannot occur, then the events A and B are said to be **disjoint** or **mutually exclusive**.

- If A and B are mutually exclusive, then we denote the intersection as

$$A \cap B = \{\} \text{ or } A \cap B = \phi.$$

7.3.3 Complement

- The set of outcomes that are in sample space (Ω) but not in A is called the compliment of A.

- It is denoted by A' or A^c or $\bar{A}$

$$A' = \{\omega \in \Omega : \omega \notin A\}.$$

7.4 Tutorial

Chapter 7: Sets and Relations

1. Which of these sets are equal?

(a) $\{r, t, s\}$
(b) $\{s, t, r, s\}$
(c) $\{t, s, t, r\}$
(d) $\{s, r, s, t\}$)

i. Only b and d are equal
ii. Only b,c and d are equal
iii. They are all equal
iv. None of them are equal

2. Let $A = \{1, 3, 5\}$ and $B = \{2, 4, 6\}$

Then

(a) $A \cap B = 0$
(b) $A \cap B = \phi$
(c) $A \cap B = \{\}$
(d) $A \cap B = \{1, 2, 3, 4, 5, 6\}$)

i. Only d is correct
ii. Only b is correct
iii. Only b and c are correct
iv. Only a, b and c are correct

3. Find the elements of the set $A = \{\{1, 2, 3\}, \{4, 5\}, \{6, 7, 8\}\}$ and determine whether each of the following is true or false

Statement	True	False
a. $1 \in A$		
b. $\{1, 2, 3\} \subseteq A$		
c. $\{6, 7, 8\} \in A$		
d. $\{\{4, 5\}\} \subseteq A$		
e. $\phi \in A$		
f. $\phi \subseteq A$		
g. $(4, 5) \in A$		

4. A player rolls a die.

(a) What is the sample space of this experiment?
(b) Write down the following events as subsets of the sample space. Decide whether each event is simple or compound.

i. A = the event of getting 3
ii. B = the event of getting 7
iii. C = the event of getting a number less than 4
iv. D = the event of getting an even number
v. E = the event of getting 2 or 5

5. A player rolls two dies

(a) What is the sample space of this experiment?
(b) Write down the following events as subsets of the sample space. Decide whether each event is simple or compound.

i. A= the event that one number is even
ii. B= the event that sum of the two numbers is odd
iii. C= the event that the numbers are equal
iv. D= the event of getting 2 and 5
v. E= the event of getting 2 or 5

6. A player rolls three dies

(a) What is the sample space of this experiment? (Use set builder form of a set to represent the sample space)
(b) Write down the following events as subsets of the sample space. Decide whether each event is simple or compound

i. A = the event of getting all the numbers greater than 4
ii. B = the event that only two numbers are equal
iii. C = the event of getting two even numbers

7. A company produces a certain type of car batteries. If a battery has a voltage that is outside certain limits, that battery is characterized as bad (B); if the battery has a voltage within the prescribed limits it is characterized as good (G). A quality inspector selects batteries one by one as they come off the production line and tests them until he gets a bad battery. He uses this data to estimate the proportion of bad batteries produced by the company.

(a) What is the sample space of this experiment

(b) Write down the following events as subsets of the sample space. Decide whether each event is simple or compound.

i. I = the event that number of good batteries before a bad battery is 3.
ii. J = the event that the number of good batteries before a bad battery is less than 3.

8. A doctor observes the lifetime of leukemia patients after being treated with a newly developed treatment.

(a) What is the sample space of this experiment?
(b) Write down the following event as subsets of the sample space. Decide whether it is simple or compound.

i. K= the event that a randomly selected patient lives more than 10 years

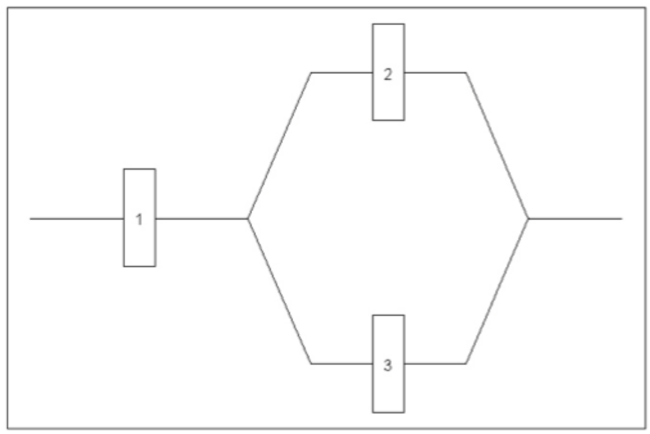
9. Three students are selected at random from the class. Let P1 = the event that the first student pass P2 = the event that the second student pass P3 = the event that the third student pass Write the following events using the above events.

(a) The event that all three students pass
(b) The event that all three students fails
(c) The event that only 2 of the three students pass after the treatments

10. Identify the events of interest as subsets of the sample space. A box contains one ticket with number 1 on it, 2 tickets with number 2, 3 tickets with number 3, 4 tickets with number 4, 5 tickets with number 5, and 6 tickets with number 6. A ticket is taken randomly from the box

(a) A = The event of getting an even number?
(b) B = The event of getting a number less than 5?
(c) C = The event of getting an even number less than 5?
(d) Express C in terms of A and B using event operations
(e)) D = The event of getting an even number or a number less than 5?
(f) Express D in terms of A and B using event operations
(g) Express, E= the event of getting an odd number, using the above defined event or events

11. Three components are connected to form a system as shown in figure (a). Because the components in the 2-3 subsystem are connected in parallel, that subsystem will function if at least one of the two individual components functions. For the entire system to function, component 1 must function and so must be 2-3 subsystem.



Each component is either in working condition (s) or not working condition (f) at any time. Let $W1$ = The event that component 1 is in working condition $W2$ = The event that component 2 is in working condition $W3$ = The event that component 3 is in working condition W = The event that the system is in working condition

(a) Write the above events as subsets of the sample space. (b) Write the following events.

i. $W3'$
ii. $W1 \cap W2$
iii. $W2 \cup W3$
iv. $W1 \cap (W2 \cup W3)$

(c) Express W in terms of $W1, W2$ and $W3$ using event operations.

12. Three products are selected at random, from a production line. Each product is either defective or non defective. Let $D1$ = The event that the first item is defective $D2$ = The event that the second item is defective $D3$ = The event that the third item is defective E = The event that two of the three products are defective Express E in terms of $D1, D2$ and $D3$ using event operations

References

Lipschutz, S. (2000). Schaum's Outline of Probability. McGraw Hill Professional.

[4] Prepared by Dr. Priyanga D. Talagala (Copyright 2020 Priyanga D. Talagala)

Chapter 8

Probability

8.1 Counting Techniques

- This section develops some techniques for determining, without direct enumeration, the number of possible outcomes of a particular experiment or event or the number of elements in a particular set.
- These methods are referred to as **counting techniques**.

8.1.1 Basic Counting Principles

- Sum Rule Principle
- Product Rule Principle / Multiplication Rule Principle

8.1.1.1 Sum Rule Principle

Example 1: Suppose there are 8 male professors (say A, B, C, D, E, F, G and H) and 5 female professors (say J, K, L, M, N) teaching a language class. How many ways are there, a student can choose a language professor?

Example 2: Suppose there are 3 different mystery novels, 5 different romance novels, and 4 different adventure novels on a bookshelf. How many different ways are there to choose one of the novels?

Sum Rule Principle: Suppose an event E_1 can occur in n_1 ways, a second event E_2 can occur in n_2 ways, a third event E_3 can occur in n_3 ways and so on, and suppose no two of the events can occur simultaneously (at the same time). Then one of the events can occur in $n_1 + n_2 + n_3 + ...$ ways

This principle can be stated in terms of sets as follows,

Sum Rule Principle: Suppose A and B are mutually exclusive events (disjoint events). Then:

$$n(A \cup B) = n(A) + n(B)$$

8.1.1.2 Product Rule Principle / Multiplication Rule

Example 3: The design for a Website is to consist of four colors, three fonts, and three positions for an image. Draw a tree diagram for the different types of designs.

- How many different designs are possible?
- The above tree diagram describes the sample space of all possible designs.
- The size of the sample space equals the number of branches in the last level of the tree, and this quantity equals $4 \times 3 \times 3 = 36$. - This leads to the following useful result.

Product Rule Principle / Multiplication Rule: Assume an operation can be described as a sequence of k steps, and - the number of ways of completing step 1 is n_1, and - the number of ways of completing step 2 is n_2 for each way of completing step 1, and - the number of ways of completing step 3 is n_3 for each way of completing step 2, and so forth. The total number of ways of completing the operation is $n_1 \times n_2 \times \cdots \times n_k$

Example 4: Suppose a password consists of 4 characters, the first 2 being simple letters in the English alphabet and the last 2 being digits. Find the number of

(a) Passwords

(b) Passwords beginning with a vowel

Example 5: There are 5 bus lines from city A to city B, 2 bus lines from city B to city C and 3 bus lines from city C to city D. Find the number of ways a person can travel by bus from city A to city D.

Example 6: There are five gates in a school. No student leaves the school from the gate he entered. Find the number of ways a student can enter and leave the school

8.1.2 Factorial Notation

- The product of the positive integers from 1 to n inclusive occurs very often in mathematics and hence it is denoted by the special symbol $n!$, read "n factorial". That is

  $$n! = 1\times2\times3\times\cdots\times(n-2)\times(n-1)\times n = n\times(n-1)\times(n-2)\times\cdots\times3\times2\times1$$

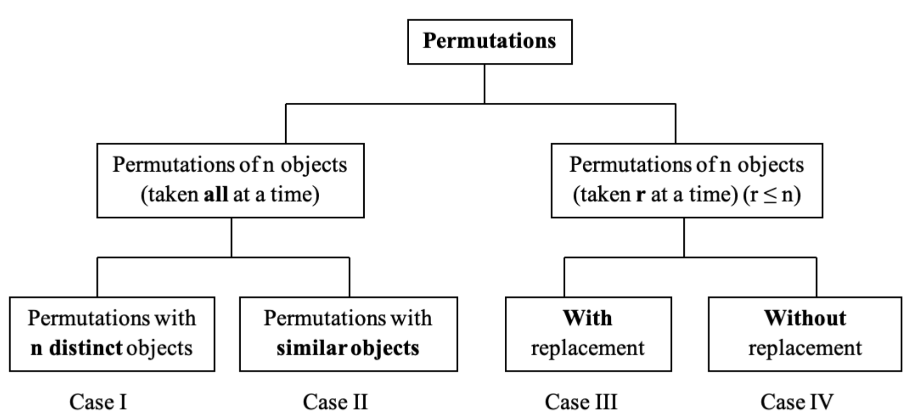
NOTE

$$0! = 1$$
$$1! = 1$$

Examples 7: (a) $2! =$
(b) $4! =$
(c) $\frac{8!}{6!}$ (d) Represent $\frac{12.11.10}{3.2.1}$ in terms of factorial notation.

8.1.3 Permutations

Permutation is an arrangement of all or part of a set of objects. Here the order is important



8.1.3.1 Case I: Permutations of n distinct objects (taken all at a time)

Example 8: Find the number of “three-letter words (not necessary to have a meaning)” that can be formed using the letters of the word “TWO”, without repetitions.

Using multiplication rule:

Permutations of n distinct objects (taken all at time):

The number of permutations of n different elements is $n!$ where

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$$

NOTE

- This result follows from the multiplication rule. - A permutation can be constructed by selecting the element to be placed in the first position of the sequence from the n elements, then selecting the element for the second position from the remaining $(n-1)$ elements, then selecting the element for the third position from the remaining $(n-2)$ elements, and so forth.

Example 9: Find the number of ways that 4 people can sit in a row of 4 seats

8.1.3.2 Case II: Permutations of similar objects (taken all at a time)

Example 10: Find the number of "three-letter words (not necessary to have a meaning)" that can be formed using the letters of the word "TOO".

Permutations of Similar Objects (taken all at a time):

The number of permutations of n $(= n_1 + n_2 + n_3 + \cdots + n_k)$ objects in which n_1 are alike, n_2 are alike, ..., n_k are alike, is

$$\frac{n!}{n_1!n_2!\dots n_k!}$$

Example 11: Find the number of "ten-letter words (not necessary to have a meaning)" that can be formed using the letters of the word "STATISTICS".

- In some situations, we are interested in the number of arrangements of only some of the elements of a set.
- Case III and Case IV address this particular situation.

8.1.3.3 Case III: Permutations of n objects taken r at a time (where $r \leq n$) (WITH Replacement)

- Here the element is replaced in the set before the next element is chosen.
- Since there are n different ways to choose each element (repetitions are allowed), the product rule principle tells us that there are

$$n.n.n \ldots n = n^r$$

different ordered possibilities of size r.

Example 12: How many "three digit numbers" can be made with the four digits 3, 5, 7 and 8 (A digit can be used as much as you can)

8.1.3.4 Case IV: Permutations of n objects taken r at a time (where $r \leq n$) (WITHOUT Replacement)

- Here the element is not replaced in the set before the next element is chosen.
- Thus, there are no repetitions in the ordered sample.
- Accordingly, an ordered sample of size $r4 without replacement is simply an r permutation of the elements in the set with n elements.
- In other words, by the product rule, the first element can be chosen in n ways, the second in $(n-1)$ ways, and so on

$$n.(n-1)(n-2) \ldots (n-r+1) = \frac{n.(n-1)(n-2) \ldots (n-r+1).(n-r)!}{(n-r)!} = \frac{n!}{(n-r)!} =^n P_r$$

NOTE

Consider the case that $n = r$. we get

$$^nP_r = \frac{n!}{(n-n)!} = \frac{n!}{0!} = \frac{n!}{1} = n!$$

Example 13: How many "three digit numbers" can be made with the four digits 3, 5, 7 and 8 (without repetition)

Example 14: A printed circuit board has eight different locations in which a component can be placed. If four different components are to be placed on the board, how many different designs are possible?

8.1.4 Combinations

- Another counting problem of interest is the number of subsets of r elements that can be selected from a set of n elements.
- These are called **combinations**.
- Here, **order is not important**.
- For example, the combinations of the letters a, b, c, d taken three at a time are $\{a, b, c\}, \{a, b, d\}, \{a, c, d\}, \{b, c, d\}$ or simply abc, abd, acd, bcd
- Observe that the following combinations are equal:

$$abc, acb, bac, bca, cab, cba$$

- That is, each denotes the same set $\{a, b, c\}$.
- The number of combinations of n objects taken r at a time will be denoted by ${}^nC_r$.

Before we derive the general formula for ${}^nC_r$, we consider a particular case.

Example: Find the number of combinations of four objects a, b, c, d taken three at a time.

Combinations	Permutations
$\{a, b, c\}$	
$\{a, b, d\}$	
$\{a, c, d\}$	
$\{b, c, d\}$	

- As can be seen in the above table, each combination consisting of three objects determines $3! = 6$ permutations of the objects in the combination.
- Thus, the number of combinations multiplied by 3! equals the number of permutations.
- That is,

$${}^4C_3 \times 3! = {}^4P_3$$

$${}^4C_3 = \frac{{}^4P_3}{3!}$$

But we know ${}^4p_3 = \frac{4!}{(4-3)!} = \frac{4!}{1!} = 2.3.4 = 24$ and $3! = 6$.

Thus

$$^4C_3 = \frac{24}{6} = 4$$

which is noted in the above table.

Formula for ^nC_r

Since any combination of n objects taken r at a time determines $r!$ permutations of the objects in the combination, we can conclude that

$$^nC_r \times r! =^n P_r.$$

Thus we obtain the following formula for ^nC_r as

$$^nC_r \times = \frac{^nP_r}{r!} = \frac{\frac{n!}{(n-r)!}}{r!} = \frac{n!}{r!(n-r)!}.$$

Example 15: In a small company there are 8 executive managers. How many ways are there to select 3 executive managers to form a new committee?

If Mr. Perera must be one of the 3, how many ways are there to form a new committee of 3?

Example 16: A printed circuit board has eight different locations in which a component can be placed. If five identical components are to be placed on the board, how many different designs are possible?

Example 17: In a small company there are 7 women and 5 men. A committee of 3 women and 2 men are to be selected. How many different possibilities are there to select a committee?

Summary

Sum Rule Principle:

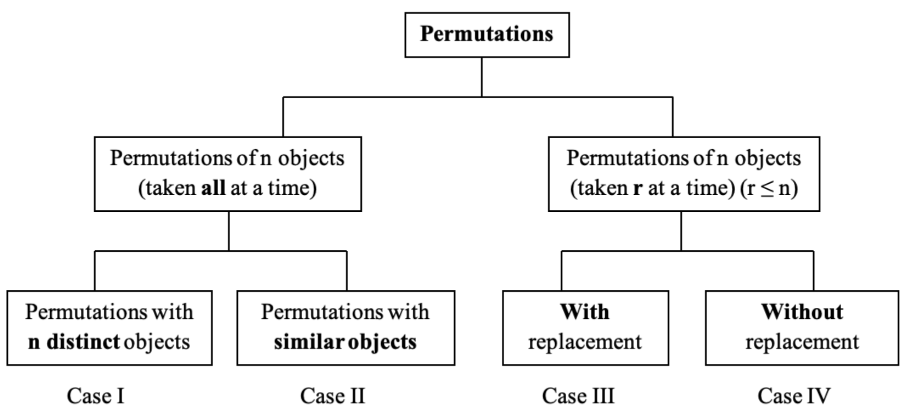
Suppose an event E_1 can occur in n_1 ways, a second event E_2 can occur in n_2 ways, a third event E_3 can occur in n_3 ways and so on, and suppose no two of the events can occur simultaneously (at the same time). Then one of the events can occur in $n_1 + n_2 + n_3 + \ldots$ ways

Product Rule Principle / Multiplication Rule

Assume an operation can be described as a sequence of k steps, and - the number of ways of completing step 1 is n_1, and - the number of ways of completing step 2 is n_2 for each way of completing step 1, and - the number of ways of completing step 3 is n_3 for each way of completing step 2, and so forth.

The total number of ways of completing the operation is $n_1 \times n_2 \times \cdots \times n_k$

Permutations

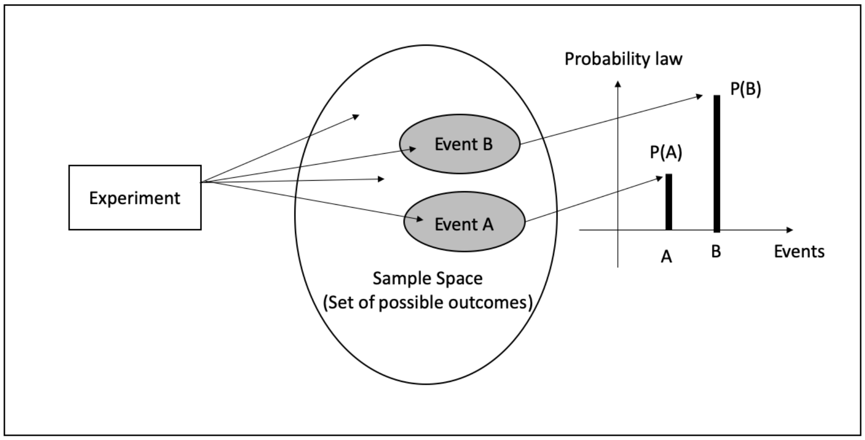


$n!$ $\qquad \dfrac{n!}{n_1!n_2!\ldots n_k!}$ $\qquad n^r$ $\qquad {}^nP_r = \dfrac{n!}{(n-r)!}$

Number of combinations of n objects taken r at a time $(r \leq n)$

$$ {}^nC_r = \frac{n!}{r!(n-r)!} $$

8.2 Axioms of probability



- **Probability** of an event quantifies the **uncertainty**, randomness, or the possibility of occurrence the event.
- The probability of event E is usually denoted by $P(E)$.
- Mathematically, the function $P(.)$ is a set function defined from sample space (Ω) to $[0, 1]$ interval, satisfying the following properties.
- These are called the **'axioms of probability'**.
- **Axiom 1:** For any event A, $P(A) \geq 0$
- **Axiom 2:** $P(\Omega) = 1$
- **Axiom 3:**
  - (a) If $A_1, A_2, \dots, A_k$ is a finite collection of mutually exclusive events, then
  $$P(A_1 \cup A_2 \cup \dots \cup A_k) = \sum_{i=1}^{k} P(A_i)$$
  (b) If $A_1, A_2, \dots$ is an infinite collection of mutually exclusive events, then
  $$P(A_1 \cup A_2 \cup \dots) = \sum_{i=1}^{\infty} P(A_i)$$

NOTE

- Axioms 1 and 2 imply that for any event E, $0 \leq P(E) \leq 1$.
- $P(E) = 1 \iff$ the event E is certain to occur.
- $P(E) = 0 \iff$ the event E cannot occur.

8.3 Methods for determining Probability

- There are several ways for determining the probability of events.
- Usually we use the following methods to obtain the probability of events.
  - Classical method
  - Relative frequency method (Empirical approach)
  - Subjective method
  - Using probability models

8.3.1 Classical Method

- In the classical method, the probability of an event is calculated based on the sample space.
- In this method no need to carry out any experiment. It is enough to know the sample space
- Let E be an event of the sample space Ω. Suppose that
  - the sample space is finite, and
  - all the outcomes in the sample space are equally likely.

Then, the probability of event E is given by

$$P(E) = \frac{\text{Number of outcomes in E}}{\text{Number of outcomes in } \Omega} = \frac{n(E)}{n(\Omega)}$$

NOTE:

- For calculating probability using the classical method, knowledge on counting techniques is helpful.

Activity 01

Consider the following problems. Is it possible to use the classical method to calculate the probabilities of the event interest in these problems? Calculate the probabilities using classical method if possible.

(a) A fair coin is tossed. What is the probability of getting a head?

(b) Two fair dice are rolled. What is the probability that the sum of two numbers is odd?

(c) A random sample of 4 products is taken from a production line. What is the probability that sample will contain more than 1 defective product?

(d) The carton of 12 eggs is randomly selected. What is the probability that carton will have more than 4 brown eggs?

(e) A one liter bottle of lemonade is selected randomly. What is the probability that the bottle will contain more than 900 ml?

(f) What is the probability that at least 1 call will come to my mobile phone during the lecture?

Activity 02

A box contains 3 white balls and 2 black balls. Two balls are taken from this box.

(a) Write down the sample space

(b) Are the outcomes of your sample space equally likely?

(c) Can you write the sample space so that outcomes will be equally likely?

(d) Find the probability of getting one white ball and one black ball

8.3.2 Relative frequency method (Empirical or objective approach)

- Consider an experiment that can be repeatedly performed in an identical and independent fashion, and let E be an event consisting of a fixed set of outcomes of the experiment.
- If the experiment is performed k times, on some of the replications the event E will occur, and on others, E will not occur.

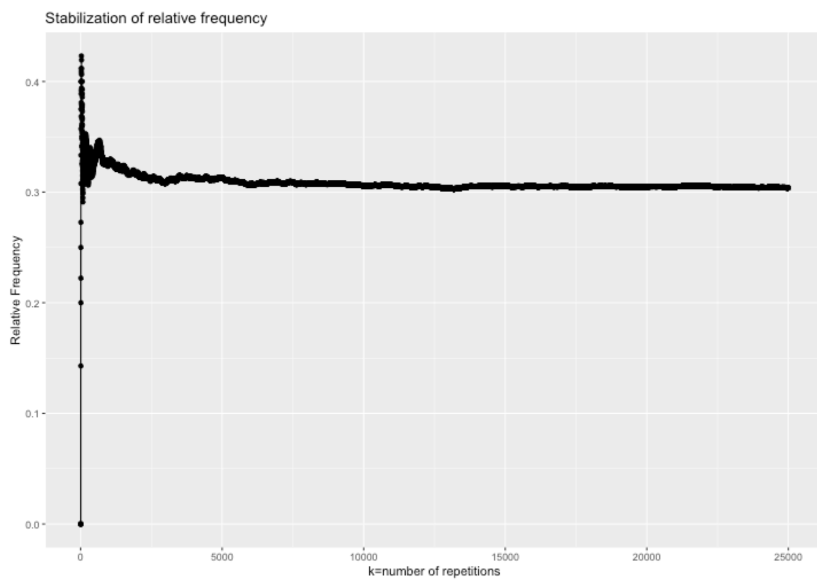
- Let $freq(E)$ be the number of replications on which E occurs.
- Then the ratio $\frac{freq(E)}{k}$ is called the **relative frequency** of event E in a sequence of k replications.
- Empirical evidence, based on the results of many such sequences of repeatable experiments, indicates that the relative frequency stabilizes as k is increased.
- This limiting relative frequency is equal to the probability of the event. That is,

$$P(E) = \lim_{k\to\infty} \frac{freq(E)}{k} \text{....................(1)}$$

Example

- A population contained an unknown proportion of defective items

- The following experiment was repeated: one item was selected randomly, and determined whether it was defective or not.
- An event E was defined to have occurred if the selected item was defective.
- The relative frequency of event E was calculated at each replication.
- The following figure shows that the relative frequency stabilizes around 0.3.
- Therefore, the probability of event E is 0.3.



- Therefore, the probability of an event E related to a random experiment can be interpreted as the **"approximate proportion** of times that E occurs if we repeat the experiment **very large number of times"**.
- For example, the quality control engineer can say that the probability that a randomly selected product from his production process is defective with probability 0.3.
- It means that if we examine **a very large number** of products from his production process, the proportion of defective items will be approximately 30%.
- The relative frequency notion of probability given by eq. (1), is not practically applicable directly because in order to determine the probability we need to do the experiment *infinitely many times*.
- However, as we can see from above graph the relative frequency is approximately equal to the limit when k is **large** enough.
- Therefore we can use the relative frequency notion to obtain an **'estimate'** for the probability of the event E.
- That is, if the number of trials, k, is sufficiently large, then the observed relative frequency becomes close to the probability $P(E)$, or

$$P(E) \approx \frac{\text{Number of times the event E occurs}}{\text{Number of trials}} = \frac{freq(E)}{k}$$

- This idea is used in practice, by using large data sets to estimate probability.
- Such probability is called **empirical probability**.
- This is also referred to as an **objective approach** because it depends on a property of the experiment rather than any particular individual concerned with the experiment.

Activity 03

In a large manufacturing company, there are 2500 employees. An opinion survey was conducted on 960 randomly selected employees regarding three types of bonus schemes. Employees were divided into four categories, namely, laborers, clerical, technical and executives. The results obtained by ways of opinion survey are presented in the following table.

Bonus scheme	Laborer	Clerk	Technician	Executive	Total
Type X	190	82	23	5	
Type Y	243	44	78	12	
Type Z	197	44	34	8	
Total					

Let

X be the event that a randomly selected employee with type X bonus scheme

Y be the event that a randomly selected employee with type Y bonus scheme

Z be the event that a randomly selected employee with type Z bonus scheme

Let

L be the event that a randomly selected employee is a Laborer

C be the event that a randomly selected employee is a Clerk

T be the event that a randomly selected employee is a Technician

E be the event that a randomly selected employee is an Executive

I. Then write the following events in words

(a) C'
(b) $C \cap X$
(c) $L \cup T$

(d) $C' \cap Z$

II. If a sample is selected at random, determine the following probabilities:

(a) $P(X)$
(b) $P(L)$
(c) $P(C')$
(d) $P(C \cap X)$
(e) $P(L \cup T)$
(f) $P(C' \cap Z)$

8.3.3 Subjective method

- A subjective probability reflects a person's opinion about the likelihood of an event.
- These probabilities are personal and they will differ between people.
- Assigning subjective probabilities to events is hard as these guesses are often based on their personal experiences and evaluation of related facts.
- Subjective interpretation of probability is used for problems in which it is difficult to imagine a repetition of an experiment.
- These are 'one shot' situations.
- For example,
  - A sports writer may say that the probability that Sri Lanka will win the next T20 World Cup is 0.6
  - A surgeon may guess that the probability of recovery after a heart surgery for certain person is about 0.5
  - A seismologist might say that there is an 80% probability that an earthquake will occur in a certain area
- One major shortcoming of subjective probability is that the probability of the same event can vary from person to person and the accuracy of these probabilities cannot be checked.

8.3.4 Using probability models

- In this method, the event of interest is expressed in terms of a "random variable" and the probability is calculated using a suitable mathematical function derived based on data. This method will be discussed later.

8.4 Conditional Probability

- Let A be an event of the sample space Ω of an experiment.

- Then $P(A)$ is called the unconditional probability or simply the probability of event A. - Then, $P(A)$ is equal to the sum of probability of each outcome in A.

- Now suppose that the experiment has been carried out and we are told that the outcome is an element of another event B (But we are not told the exact outcome).

- In other words, we are told that event B has occurred. Then, what is the probability that event A has occurred?

- The probability of an event A occurring when it is known that some event B has occurred is called a **conditional probability** and is denoted by $P(A|B)$.

- The symbol $P(A|B)$ is usually read **"the probability that A occurs given that B occurs"** or simply **"the probability of A, given B."**

Activity 01

A box contains 3 white balls and 2 black balls. Two balls are taken one after the other without replacement.

a) What is the probability that the first ball is white?
b) If the first ball is white, what is the probability that the second ball is black?
c) Show all the possible outcomes and probabilities of the above experiment in a tree diagram.
d) Identify the unconditional probabilities and conditional probabilities from the given information.
e) What is the probability that the second ball is black?

Definition - Conditional Probability

Let A and B be two events. The probability that event A occurs given that event B has occurred is called the conditional probability of A given B. It is denoted by $P(A|B)$ and defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ when } P(B) \neq 0$$

8.4.1 Independent events

- Two events A and B are said to be independent, if the occurrence of A does not depend on the occurrence of B.

$$\text{A and B are independent} \iff P(A|B) = P(A) \text{ and } P(B|A) = P(B)$$

Activity 02

A box contains 3 white balls and 2 black balls. Two balls are taken one after the other with replacement.

a) What is the probability that the first ball is white?
b) If the first ball is white, what is the probability that the second ball is black?
c) If the first ball is black, what is the probability that the second ball is black?
d) What is the probability that the second ball is black?

8.5 Rules of Probability

- Probability rules are used for calculating the probabilities of events that are expressed in terms of other events with set operations.

8.5.1 Complement rule

- Let E be an event, Then $P(E') = 1–P(E)$

8.5.2 Addition rule

- Let A and B are two events. Then,

$$P(A \cup B) = P(A) + P(B)–P(A \cap B)$$

- If A and B are mutually exclusive, then $P(A \cap B) = 0$ and, therefore,

$$P(A \cup B) = P(A) + P(B)$$

8.5.3 Multiplication rule

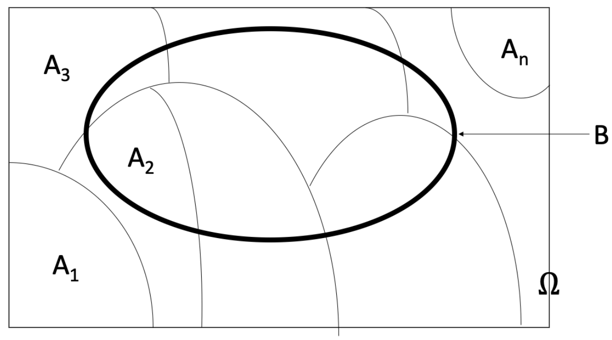
- Let A and B be two events. Then,

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

$$\text{A and B are independent} \iff P(A|B) = P(A) \text{ and } P(B|A) = P(B)$$

$$\text{A and B are independent} \iff P(A \cap B) = P(A)P(B)$$

8.5.4 The law of total probability



$$A_1 \cup A_2 \cup \cdots \cup A_n = \Omega, \text{ all mutually disjoint}$$

- Let $A_1, A_2, \ldots A_n$ be mutually exclusive and exhaustive events. Then, for any event B,

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_n)$$

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n)$$

8.6 Bayes' Theorem

- Let $A_1, A_2, \ldots, A_n$ be mutually exclusive and exhaustive events, and B be any event.

Then for $i = 1, 2, \ldots, n$.

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)}$$

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$$

8.7 Tutorial

Chapter 8: Probability

1. Three cards are chosen in succession from a deck with 52 cards. Find the number of ways this can be done

(a) with replacement
(b) without replacement

2. Find the number m of ways that 9 books can be divided between 4 students if the student with the highest marks is to receive 3 books and each of the others 2 books.

3. Find the number n of different signals, each consisting of 6 bulbs arranged in a vertical line, that can be formed from 4 identical red bulbs and 2 identical blue bulbs.

4. Suppose repetitions are not allowed.

(a) Find the number n of three-digit numbers that can be formed from the six digits: $2, 3, 5, 6, 7, 9$
(b) How many of them are even?
(c) How may of them exceed 400?

5. A student is to answer 8 out of 10 questions on an exam.

   (a) Find the number n of ways the student can choose the eight questions
   (b) Find n if the student must answer the first three questions

6. A random experiment can result in one of the outcomes $\{a, b, c, d\}$ with probabilities 0.1, 0.3, 0.5, and 0.1, respectively. Let A denotes the event $\{a, b\}$, B the event $\{b, c, d\}$, and C the event $\{d\}$. Then calculate the following probabilities

(a) $P(A)$
(b) $P(B)$
(c) $P(C)$
(d) $P(A')$
(e) $P(B')$
(f) $P(C')$
(g) $P(A \cap B)$
(h) $P(A \cup B)$
(i) $P(A \cap C)$

7. A box contains one ticket with number 1 on it, 2 tickets with number 2, 3 tickets with number 3, 4 tickets with number 4, 5 tickets with number 5, and 6 tickets with number 6. A ticket is taken randomly from the box.

   (a) What is the probability of getting an even number?
   (b) What is the probability of getting a number less than 5?
   (c) What is the probability of getting an even number less than 5?
   (d) What is the probability of getting an even number or a number less than 4?

8. A credit card contains 16 digits between 0 and 9. However, only 100 million numbers are valid. If a number is entered randomly, what is the probability that it is a valid number?

9. In a group of 200 college students, 134 are enrolled in a course in Statistics, 68 are enrolled in a course in Computer Science and 43 are enrolled in both. What is the probability that a randomly selected student is not enrolled in either course? What is the method you use to calculate the above probability value? Justify

10. In an acid-base titration, a base or acid is gradually added to the other until they have completely neutralized each other. Because acids and bases are usually colorless (as are the water and salt produced in the neutralization reaction), pH is measured to monitor the reaction. Suppose that the equivalence point is reached at approximately 100 ml of an NaOH solution has been added (enough to react with all the acetic acid present) but that replicates are equally likely to indicate from 95 to 104 ml to the nearest ml. Assume that volumes are measured to the nearest ml and describe the sample space.

(a) What is the probability that equivalence is indicated at 100 ml?
(b) What is the probability that equivalence is indicated at less than 100 ml?
(c) What is the probability that equivalence is indicated between 98 and 102 ml (inclusive)?

11. Suppose A, B, and C are events with $P(A) = 1/2$, $P(B) = 1/2$, $P(C) = 1/3$, $P(A \cup B) = 3/4$,
$P(A \cap C) = 1/6$, $P(B \cap C) = 1/6$, and $Pr(A \cap B \cap C) = 1/12$.

(a) Determine whether or not A and B are independent.
(b) Calculate $P(A \cup B \cup C)$.

12. A lot of 100 semiconductor chips contains 20 that are defective. Two are selected randomly, without replacement, from the lot.

(a) What is the probability that the first one selected is defective?

(b) What is the probability that the second one selected is defective given that the first one was defective?
(c) What is the probability that both are defective?
(d) How does the answer to part (b) change if chips selected were replaced prior to the next selection?

13. Disks of polycarbonate plastic from a supplier are analyzed for scratch and shock resistance. The results from 100 disks are summarized as follows:

		Shock Resistance	
		High	Low
Scratch Resistance	High	70	9
	Low	16	5

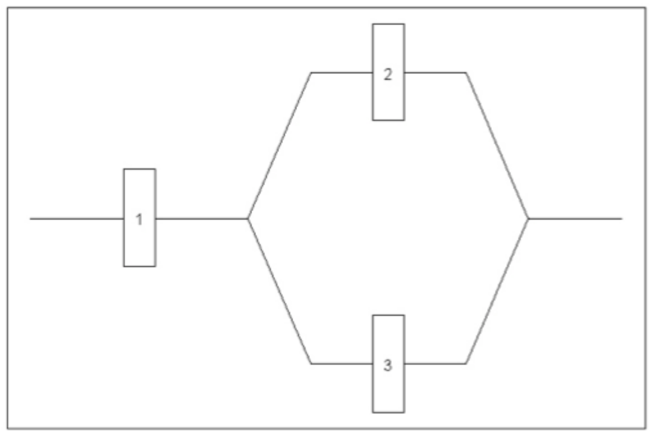
Let A denote the event that a disk has high shock resistance, and let B denote the event that a disk has high scratch resistance.

If a disk is selected at random, determine the following probabilities:

(a) $P(A)$
(b) $P(B)$
(c) $P(A')$
(d) $P(A' \cup B)$
(e) If a disk is selected at random, what is the probability that its scratch resistance is high and its shock resistance is high?
(f) If a disk is selected at random, what is the probability that its scratch resistance is high or its shock resistance is high?
(g) Consider the event that a disk has high scratch resistance and the event that a disk has high shock resistance. Are these two events mutually exclusive?
(h) $P(A|B)$
(i) $P(B|A)$

14. Two pumps connected serially fail independently of one another on any given day. The probability that the older pump will fail is on any given day is 0.05. The probability that the newer pump will fail on any given day is is 0.01. The system fails if at least one pump fails. What is the probability that the pumping system will fail on any given day?

15. Three components are connected to form a system as shown in figure (a). Because the components in the 2-3 subsystem are connected in parallel, that subsystem will function if at least one of the two individual components functions. For the entire system to function, component 1 must function and so must be 2-3 subsystem.



Each component is either in working condition (s) or not working condition (f) at any time. Let

$W1 =$ The event that component 1 is in working condition

$W2 =$ The event that component 2 is in working condition

$W3 =$ The event that component 3 is in working condition

$W =$ The event that the system is in working condition

If the probability that any component work is 0.9, calculate the probability that system works.

16. Three students are selected at random from a class. According to past data, 90% of students pass. Calculate the probability that 2 of the three students pass.

17. A company employing 1000 workers offers deluxe medical coverage (D), standard medical coverage (S), and economy medical coverage (E) to its employees. Of the employees, 30% have D, 60% have S, and 10% have E. From past experience, the probability that an employee with D, will submit no claims during next year is 0.1. The corresponding probabilities for employees with S and E are 0.4 and 0.7 respectively. If an employee is selected at random,

(a) What is the probability that the selected employee will not submit a claim during next year?
(b) If an employee has not submitted any claim during the last year, what is the probability that the employee has S

18. Two students are selected at random, one after the other and without replacement, from a group of 10 students of whom 6 are male. What

is the probability that first selected student is a male given that second selected student is a male?

19. In a lotto game, a person may pick any 6 numbers from the numbers $1, 2, \dots, 44$ for his/her ticket. Suppose one number can be chosen only once. What is the probability of getting

(a) all the winning numbers in the same order as those were drawn in the draw.
(b) all the winning numbers in some order
(c) What is the method you used to calculate the above probabilities? Justify you method.

20. A chain of video stores sells three different brands of video cassette recorders (VCR). 50% of its VCR sales are brand A, 30% are brand B and 20% are brand C. Each manufacturer offers a one-year warranty on parts and labour. It is known that 25% brand A require warranty repair work, whereas the corresponding percentages for brands B and C are 20% and 10% respectively.

(a) What is the probability that a randomly selected buyer will buy a brand A VCR and it will need repair while under warranty?

(b) What is the probability that a randomly selected buyer will buy a VCR that will need repair under warranty?

(c) If a customer returns to the store with a VCR that needs warranty repair, what is the probability that it is a

  (i) brand A VCR?
  (ii) brand B VCR?
  (iii) brand C VCR?

21. Consider the experiment of rolling a weighted dice with numbers from 1 to 6. The weights have been applied to sides so that the probability of getting a number is directly proportional to that number.

(a) What is the probability of getting an even number in any single rolling?
(b) What is the probability of getting a number greater than 2 in any single rolling?
(c) What is the probability of getting an even number or a number greater than 2 in any single rolling?
(d) If the number that turns up in any single rolling is even, what is the probability that the number is greater than 2?

22. A particle moves on a circle through points which have been marked $0, 1, 2, 3, 4$ (in a clockwise order). At each step it has a probability p of moving to right (clock-wise) immediate value and $(1 - p)$ to the left (counter-clock wise) immediate value. Let X_n denote its location on the circle in the nth step and $P_{ij} = P(X_{n+1} = j|X_n = i)$.

a) Calculate the P_{ij}'s where $i = 0, 1, 2, 3, 4$ and $j = 0, 1, 2, 3, 4$
b) display the calculated values in a) in the following format as a matrix.

$$\begin{pmatrix} P_{00} & P_{01} & P_{02} & P_{03} & P_{04} \\ P_{10} & P_{11} & P_{12} & P_{13} & P_{14} \\ P_{20} & P_{21} & P_{22} & P_{23} & P_{24} \\ P_{30} & P_{31} & P_{32} & P_{33} & P_{34} \\ P_{40} & P_{41} & P_{42} & P_{43} & P_{44} \end{pmatrix}$$

23. Consider the experiment of tossing two fair dice. Let A denote the event of an odd total, B the event of 1 on the first die, and C the event of a total of seven.

(a) Are A and B independent?
(b) Are A and C independent?
(c) Are B and C independent?

24. Verify the following relationships (De Morgan's laws) using Venn diagrams

(a) $A' \cap B' = (A \cup B)'$
(b) $A' \cup B' = (A \cap B)'$

25. Suppose A and B are two independent events. Prove the followings

(a) $P(A' \cap B) = P(A')P(B)$
(b) $P(A \cap B') = P(A)P(B')$
(c) $P(A' \cap B') = P(A')P(B')$

References

Casella, G., & Berger, R. L. (2002). Statistical inference (Vol. 2, pp. 337-472).Pacific Grove, CA: Duxbury

Lipschutz, S. (2000). Schaum's Outline of Probability. McGraw Hill Professional.

Mood, A.M., Graybill, F.A. and Boes, D.C. (2007): Introduction to the Theory of Statistics, 3rd Edn. (Reprint). Tata McGraw-Hill Pub. Co. Ltd.

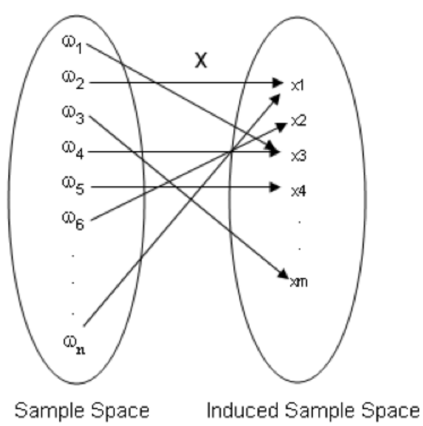
Chapter 9

Random Variables

- Some sample spaces contain quantitative (numerical) outcomes, others contain qualitative outcomes.
- Often it is convenient to work with sample spaces containing numerical outcomes.
- A function that maps the original sample space into the real numbers is called a 'random variable'.
- This is more useful when the original sample space contains qualitative outcomes.

Definition 1: Random Variable

Let Ω be a sample space. Let X be a function from Ω to $\Re$ (*i.e.* $X : \Omega \rightarrow \Re$). Then X is called a random variable.



- A random variable assigns a real number to each outcome of a sample space.

- In other words, to each outcome of an experiment or a sample point ω_i, of the sample spaces, there is a unique real number x_i, known as the value of the random variable X.
- The range of the random variable is called the *induced sample space.*
- *A note on notation:* Random variables will always denoted with uppercase letters and the realized values of the random variable (or its range) will be denoted by the corresponding lowercase letters. Thus, the random variable X can take the value x.
- Each outcome of a sample space occurs with a certain probability. Therefore, each possible value of a random variable is associated with a probability.
- Any events of a sample space can be written in terms of a suitably defined random variable.

9.1 Types of Random Variables

- A random variable is of two types
  - Discrete Random Variable
  - Continuous Random Variable

9.2 Discrete Random Variable

- If the induced sample space is discrete, then the random variable is called a **discrete random variable**.

Example 01 Consider the experiment of tossing a coin. Express the following events using a suitably defined random variable

$H = \textit{The event of getting a head}$

$T = \textit{The event of getting a tail}$

Example 02

Consider the experiment of rolling of a die. Express the following events using a suitably defined random variable

$A =$ *The event that the number faced up is less than 5*

$B =$ *The event that the number faced up is even*

$C =$ *The event that the number faced up is 2 or 5*

Example 03

Consider the experiment of tossing a coin 10 times. Then the sample space Ω contains $2^{10} = 1024$ outcomes. Each outcome is a sequence of 10 H's and T's.

Express the following events in terms of a suitably defined random variable.

$D =$ *The event that the number of heads is 5*

$E =$ *The event that the number of tails is less than 4*

9.3 Continuous Random Variable

- If the induced sample space is continuous, then the random variable is called a **continuous random variable.**

Example 04

Consider the experiment of measuring the lifetime (in hours) of a randomly selected bulb. Express the following events in terms of a suitably defined random variable.

$F =$ *The event that the lifetime is less than 300 hours*

$G =$ *The event that the lifetime is 1000 hours*

References

Casella, G., & Berger, R. L. (2002). Statistical inference (Vol. 2, pp. 337-472).Pacific Grove, CA: Duxbury

Mood, A.M., Graybill, F.A. and Boes, D.C. (2007): Introduction to the Theory of Statistics, 3rd Edn. (Reprint). Tata McGraw-Hill Pub. Co. Ltd.

Chapter 10

Correlation and Regression

10.1 Introduction to Correlation and Regression

- In many practical situations we want to identify various types of relationships between variables.
- Sometimes we want to estimate how one variable is related to or affected by some other variables.
- **Regression analysis** is a statistical technique for investigating and **modeling the relationship between variables*.
- There are numerous applications of regression in many fields.

Examples - A production company may need to determine how its sales related to advertising - How the growth of the bacteria is related to moisture level of the environment. - The relationship between blood pressure and the age of a person. - the relationship between transaction time and transaction amount in fraud detection.

- Usually, the first step in regression analysis is to construct a scatter plot (or scatter matrix).
- Graphing the data in a scatter plot yields preliminary information about the *shape* and *spread* if the data.

10.1.0.1 Scatter plot

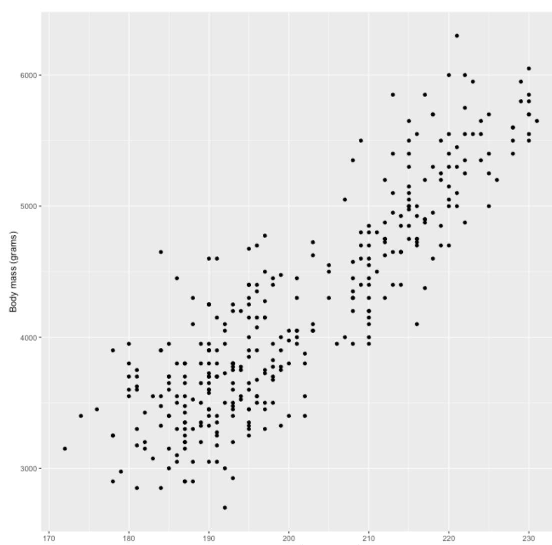
- A scatter plot is a two dimensional graph of pairs of points from **two numerical** variables
- In a quantitative bi-variate dataset, we have a (x, y) pair for each sampling unit, where x denotes the independent variable and y denotes the dependent variable.

CHAPTER 10. CORRELATION AND REGRESSION

- Each (x, y) pair can be considered as a point on the Cartesian plan.
- Scatter plot is a plot of all the (x, y) pairs in the dataset.
- The purpose of scatter plot is to illustrate any relationship between two quantitative variables.
  - If the variables are related, what kind of relationship it is, linear or nonlinear?
  - If the relationship is linear, the scatter plot will show whether it is negative or positive. - The scatter plot gives some idea of how well a regression line fits the data.

Example: *Palmer Archipelago (Antarctica) Penguin Data*

The palmerpenguins data (available through `palmerpenguins` R package) contains size measurements for three penguin species observed on three islands in the Palmer Archipelago, Antarctica.



Penguin flipper length and body mass show a positive association for the 3 species.

10.1.0.2 Correlation

- Correlation is a measure of the degree of relatedness of two or more variables.
- Several measures of correlation are available , the selection of which depends mostly on the level of data being analysed.

- Ideally, researchers would like to calculate ρ, the **population** coefficient of correlation.
- However, because researchers virtually always deal with sample data, this section introduces a widely used sample coefficient of correlation, r.
- This measure is applicable **only if both variables being analysed have at least an interval level of data**

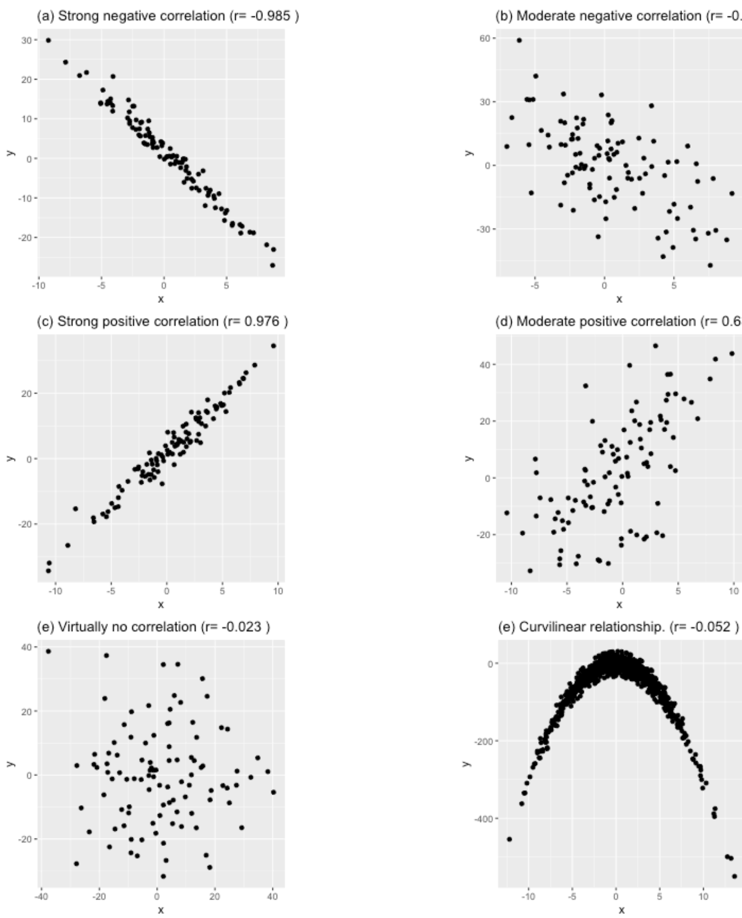
Pearson product-moment correlation coefficient (r)

- The statistic r is the Pearson product-moment correlation coefficient, named after Karl Pearson (1857 - 1936).
- The tern r is a measure of the **linear** correlation of two variables.
- It is a number that ranges from -1 to 0 to +1, representing th strength of the linear relationship between the variables.
- An r value of $+1$ denotes a perfect **linear** positive relationship between two variables.
- An r value of -1 denotes a perfect **linear** negative relationship between two variables, which indicates an inverse relationship between two variables: as one variable gets larger, the other gets smaller.
- An r value of 0 means no **linear** relationship is present between the two variables.

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}$$

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

- Examples: Following figure shows five different degrees of correlation:



NOTE

- When $r = 0$, it signifies there is **no linear** relationship between the two variables. (There can be a non-linear relationship, Figure (e))
- Figure (e): There is a very strong curvilinear relationship. But there is **no linear** relationship.

10.2 Simple Linear Regression

- The most elementary regression model is called **simple linear regression**.
- Is is also known as *bivariate linear regression*, which means that it involves only two variables.
- On variable is predicted by another variable.

- The variable to to be predicted is called the *independent variable* an is denoted by y.
- The *predictor* is called the *independent variable* or *explanatory variable* and is denoted by x
- In simple *linear* regression analysis, only a strait-line relationship between two variables is examined.
- Nonlinear relationships and regression models with more than one independent variable can be explored by using multiple regression models.

10.2.1 Determining the equation of the regression line

- The first step in determining the equation of the regression line that passes through he sample data is to establish the equation's form.

- In mathematics, the equation of a line can be written as

$$y = mx + c$$

where: m = slope of the line
$c = y$ intercept of the line.

- In statistics, the slope-intercept form of the equation of the regression line through the population points is:

$$\hat{y} = \beta_0 + \beta_1 x$$

where: $\hat{y}$ = the predicted value of y
β_0 = the population y intercept
β_1 = the population slope.

- For any specific dependent variable value, y_i:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

where: x_i = the value of the independent variable for the ith value
y_i = the value of the dependent variable for the ith value
β_0 = the population y intercept
β_1 = the population slope
ϵ_i = the error of prediction for the ith value.

- Unless the points being fitted by the regression equation are in perfect alignment, the regression line will miss at least some of the points.

- In the above equation, ϵ_i represents the error of the regression line in fitting these points. If a point is on the regression line, $\epsilon_i = 0$

10.2.2 Deterministic models vs probabilistic models

- These mathematical models can be either deterministic models or probabilistic models.

Deterministic models

- Deterministic models are *mathematical models that produces an* **exact** *output for a given input*
- For example, consider the equation of a regression line is:

$$\hat{y} = 1.68 + 2.40x$$

.

For a value of $x = 5$, the exact predicted value of y is

$$\hat{y} = 1.68 + 2.40 \times 5 = 13.68$$

. - However, the most of the time the values of y will not equal exactly the values yields by the equations. - Random error will occur in the prediction of the y values for values of x, because it is likely that the variable x doe bot explain all the variability of the variable y.

Example

- Suppose we want to predict the sales volume (y) for a mobile phone company through regression analysis by using the annual amount of advertising (in Rupees) (x) as the predictor.
- Although sales are often related to advertising, there can be other factors related to sales that are not accounted for by the amount of advertising.
- Therefore, a regression model to predict sales volume by the amount of advertising probably involves some error.
- For this reason, in regression, we present the general model as a probabilistic model.

Probabilistic models

- A probabilistic model is *one that includes an error term that allows for the y values to vary for any given value of x.*

- The deterministic regression model is

$$y = \beta_0 + \beta_1 x$$

- The probabilistic regression model is

$$y = \beta_0 + \beta_1 x + \epsilon.$$

- $\beta_0+\beta_1 x$ is the deterministic portion of the probabilistic model, $\beta_0+\beta_1 x+\epsilon$.
- In deterministic mode, all points are assumed to be on the line and in all cases ϵ is zero.

10.2.3 Least squares estimation of the parameters

- The parameters β_0 and β_1 are unknown and need to be estimated using *sample data*.
- The equation of the regression line contains the sample y intercept, $\hat{\beta}_0$, and the sample slope, $\hat{\beta}_1$.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

where: $\hat{y}$ = the predicted value of y
$\hat{\beta}_0$ = the sample y intercept
$\hat{\beta}_1$ = the sample slope

- To determine the equation of the regression line for a sample of data, the researcher must determine the values of $\hat{\beta}_0$ and $\hat{\beta}_1$.
- This process is sometimes referred to as **least squares analysis**.
- Least squares analysis is a process whereby a regression model is developed by *producing the minimum sum of the squared error values.*
- The least squares regression line is the **regression line that results in the smallest sum of error squared**.

Least square analysis contd.

10.2.4 Residual analysis

- Each difference between the actual y values and the predicted y values is the error of the regression line at a given point, $y - \hat{y}$, and is referred to as the **residual**.

Example

Number of CD sales ('000) x	Concert attendance ('000) y	Predicted value $\hat{y}$	Residual $y - \hat{y}$
61	4.280		
63	4.080		
67	4.420		
69	4.170		
70	4.480		
74	4.300		
76	4.820		
81	4.700		
86	5.110		
91	5.130		
95	5.640		
97	5.560		

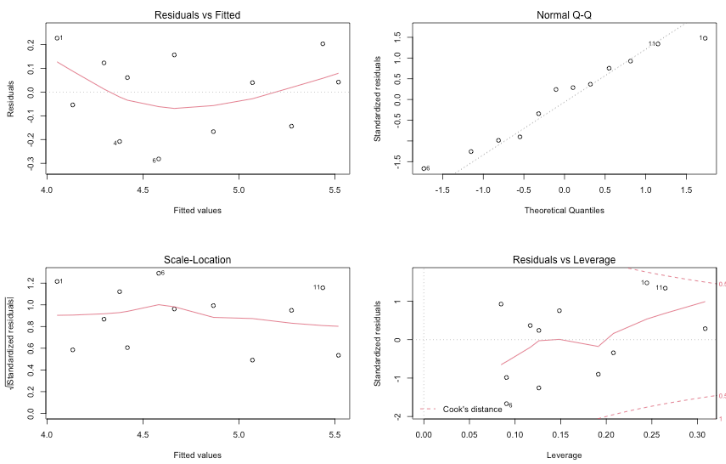
- Except for rounding error, the sum of the residuals is approximately zero
- The analysis of residuals plays an important role in validating the regression model.
- Residual analysis plots are a very useful tool for assessing aspects of accuracy of a linear regression model on a particular dataset and testing that the attributes of a dataset meet the requirements for linear regression.
- The following are the assumptions of simple linear regression analysis

1. The model is linear.
2. The error terms have constant variances. (*The assumption of constant variance is called homoskedasticity. If the error variances are not constant, it is called heteroskedasticity.*)
3. The error terms are independent.
4. The error terms are normally distributed.

- Four standard plots can be accessed using the `plot()` function in R with the fit variable once the model is generated.

- These can be used to show if there are problems with the dataset and the model produced that need to be considered in looking at the validity of the model.

```
##
## Call:
## lm(formula = concert_attendance ~ record_sales)
##
## Coefficients:
##  (Intercept)  record_sales
##       1.5698        0.0407
```



Residual analysis

10.2.5 Coefficient of determination

- A widely used measure of fit for regression ,models is the **coefficient of determination** (R^2)
- The coefficient of determination is the proportion of variability of the dependent variable (y) accounted for, or explained by, the independent variable (x)
- The coefficient of determination ranges from 0 to 1.
- An R^2 of zero means that the predictor accounts for none of the variability of the dependent variable and that there is no regression prediction of y by x.
- An R^2 of 1 means perfect prediction of y by x and that 100% of the variability of y is accounted for by x
- The dependent variable, y, being predicted in a regression model has a variation that is measured by the total sum of squares of y (SS_{yy}):

$$SS_{yy} = \sum(y - \bar{y})^2 = \sum y^2 - \frac{(\sum y)^2}{n}$$

and is the sum of the squared deviations of the y values from the mean value of y.

- This variation can be broken into two additive variations: the *explained variation*, measure by the sum of squares of regression (SSR), and the *unexplained variation*, measured by the sum of squares of error (SSE).
- The relationship can be expressed in equation from as:

$$SS_{yy} = SSR + SSE$$

  where $SSR = \sum(\hat{y} - \bar{y})^2$ and $SSE = \sum(y - \hat{y})^2$.
- If each term in the equation is divided by SS_{yy}, the resulting equation is:

$$1 = \frac{SSR}{SSyy} + \frac{SSE}{SSyy}.$$

- The term R^2 is the proportion of the y variability that is explained by the regression model and represented here as :

$$R^2 = \frac{SSR}{SSyy}.$$

- Substituting this equation into the preceding relationship gives:

$$1 = R^2 + \frac{SSE}{SSyy}.$$

- Solving for R^2 gives

$$R^2 = 1 - \frac{SSE}{SSyy} = 1 - \frac{SSE}{\sum y^2 - \frac{(\sum y)^2}{n}}.$$

```
reg <- lm(concert_attendance ~ record_sales)
summary(reg)
```

```
##
## Call:
## lm(formula = concert_attendance ~ record_sales)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.28171 -0.14938  0.04101  0.13162  0.22741
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.569793   0.338083   4.643 0.000917 ***
## record_sales 0.040702   0.004312   9.439 2.69e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1772 on 10 degrees of freedom
## Multiple R-squared:  0.8991, Adjusted R-squared:  0.889
## F-statistic: 89.09 on 1 and 10 DF,  p-value: 2.692e-06
```

10.2.6 Relationship between r and R^2

- Let r be the coefficient of correlation.
- The coefficient of determination (R^2) is the square of the coefficient of correlation

$$r = \sqrt{R^2}$$

- The researcher must examine the sign of the slope of the regression line to determine whether a positive or negative relationship exists between the variables and then assign the same sign to the correlation value.

```
r <- cor(concert_attendance, record_sales)
r^2
```

```
## [1] 0.8990839
```

10.3 Multiple Linear Regression

- Multiple regression is an extension of ordinary least-squares (OLS) regression that involves more than one explanatory variable.
- Multiple linear regression (MLR) is a statistical technique that uses several explanatory variables to predict the outcome of a response variable.
- The goal of multiple linear regression (MLR) is to model the **linear** relationship between the explanatory (independent) variables and response (dependent) variable.

Formula of Multiple Linear Regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k + \epsilon$$

where

- y = the value of the dependent variable
- β_0 = the regression constant, or intercept
- β_1 = the partial regression coefficient for independent variable 1
- β_2 = the partial regression coefficient for independent variable 2
- β_k = the partial regression coefficient for independent variable k
- k = the number of independent variables.

More about regression analysis:

References

- Black, K., Asafu-Adjaye, J., Khan, N., Perera, N., Edwards, P., & Harris, M. (2007). *Australasian business statistics.* John Wiley & Sons.

- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to linear regression analysis (Vol. 821). John Wiley & Sons.

- Salvatore, D., & Reagle, D. (2002). Statistics and Econometrics, Schaum's Outline Series.

Tutorial

1. The data in the table are the twelve observations corresponding to concert attendance (in thousand) and total worldwide CD sales by the performing artist (or band) in the previous year (also in thousand). Draw a scatter diagram for the data and determine by inspection if there exists an approximate linear relationship between X and Y

Number of CD sales ('000)	Concert attendance ('000)
61	4.280
63	4.080
67	4.420
69	4.170
70	4.480
74	4.300
76	4.820
81	4.700
86	5.110
91	5.130
95	5.640
97	5.560

2. Use the data in Question 1 to develop a regression model to predict concert attendance by CD sales.

R code

```
record_sales <- c(61,63,67,69,70,74,76,81, 86,91,95,97)
concert_attendance <- c(4.28,4.08,4.42,4.17,4.48,4.3,4.82,4.7,5.11,5.13,5.64,5.56)

# plot(x, y)
plot(concert_attendance, record_sales)

# lm(y ~ x)
reg <- lm(concert_attendance ~ record_sales)
reg
```

```
##
## Call:
## lm(formula = concert_attendance ~ record_sales)
##
## Coefficients:
```

```
##  (Intercept)  record_sales
##       1.5698        0.0407
```

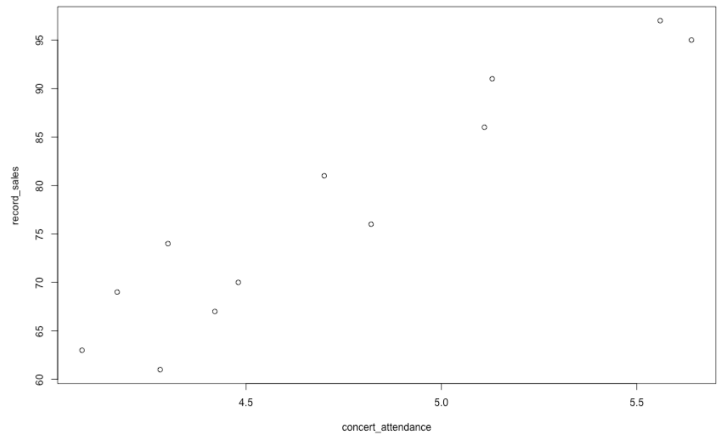
```
summary(reg)
```

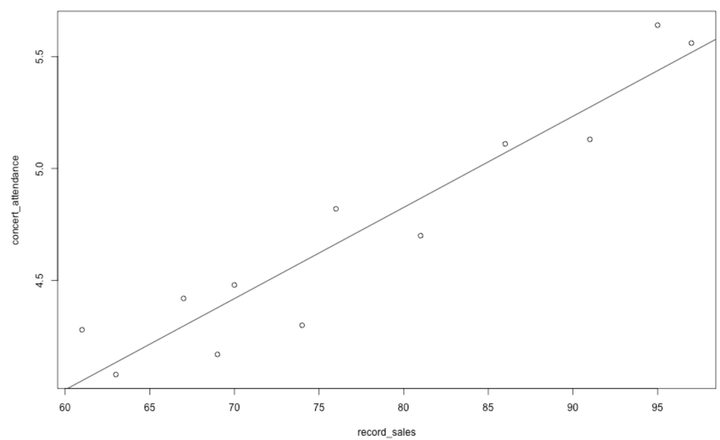
```
##
## Call:
## lm(formula = concert_attendance ~ record_sales)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.28171 -0.14938  0.04101  0.13162  0.22741
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.569793   0.338083   4.643 0.000917 ***
## record_sales 0.040702   0.004312   9.439 2.69e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1772 on 10 degrees of freedom
## Multiple R-squared:  0.8991, Adjusted R-squared:  0.889
## F-statistic: 89.09 on 1 and 10 DF,  p-value: 2.692e-06
```

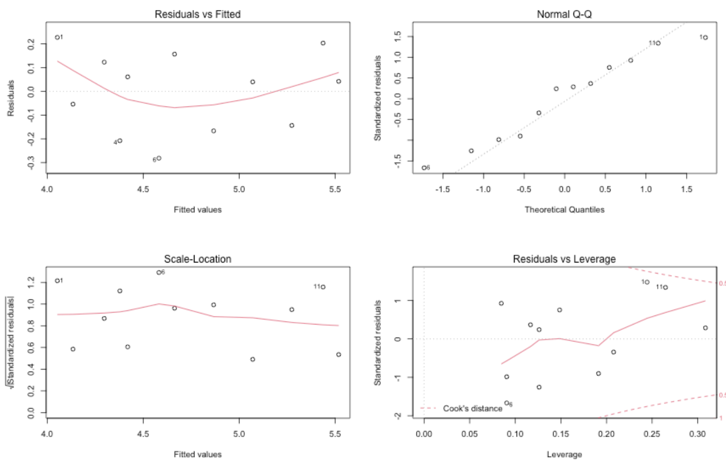
```
#Add a regression line
plot(concert_attendance ~ record_sales) + abline(reg)
```

```
## integer(0)
```

```
par(mfrow = c(2,2))
plot(reg)
```


record_sales
60
65
70
75
80
85
90
95
4.5
5.0
5.5
concert_attendance


concert_attendance
4.5
5.0
5.5
60
65
70
75
80
85
90
95
record_sales



3. An executive at a telecommunications company is interested in the relationship between an individual's income and their mobile-phone usage. In particular, to help him in pricing and marketing strategies, he is interested in ascertaining whether he can us an individual's gross annual income to predict how much time they will spend on making National Direct Calls from their mobile phone per week. He surveyed 12 mobile-phone users and recorded their annual income and time (in minutes) spent each week making National Direct Calls.

Annual income ('000)	Weekly time on National Direct Calls (minutes)
23	69
29	95
29	102
35	118
42	126
46	125
50	138
54	178
64	156
66	184
76	176
78	225

4. Compute coefficient of determination (R^2) for Question 2 (CD-concert question). Discuss the value of R^2 obtained.

5. Compute the coefficient of determination (R^2)for question 3 in which a

regression model was developed to predict weekly mobile phone call times by a person's income. Discuss the value of R^2 obtained.

6. What is meant by and what is the function of

a) Simple regression analysis
b) Linear regression analysis
c) A scatter diagram
d) An error term

References

Acharjya, D. P. (2009). Fundamental approach to discrete mathematics. New Age International.

Black, K., Asafu-Adjaye, J., Khan, N., Perera, N., Edwards, P., & Harris, M. (2007). *Australasian business statistics*. John Wiley & Sons.

Butenko, S., & Pardalos, P. M. (2014). Numerical methods and optimization: An introduction. CRC Press

Casella, G., & Berger, R. L. (2002). Statistical inference (Vol. 2, pp. 337-472).Pacific Grove, CA: Duxbury

Dass, H. K. (2008). 'Complex Numbers', Advanced Engineering Mathematics. S. Chand Publishing. pp. 474-520.

Lipschutz, S. (2000). Schaum's Outline of Probability. McGraw Hill Professional.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to linear regression analysis (Vol. 821). John Wiley & Sons.

Mendelson, E. (1970). Boolean algebra and switching circuits. McGraw-Hill Edition 2004.

Mood, A.M., Graybill, F.A. and Boes, D.C. (2007): Introduction to the Theory of Statistics, 3rd Edn. (Reprint). Tata McGraw-Hill Pub. Co. Ltd.

Salvatore, D., & Reagle, D. (2002). Statistics and Econometrics, Schaum's Outline Series.

Shenk, A. (1988). Calculus and analytic geometry. Scott Foresman & Company.

Spiegel, M. R., & Stephens, L. J. (2017). Schaum's outline of statistics. McGraw Hill Professional.

Stewart, J. (2003). Calculus: Early Transcendentals. Thomson Learning. Inc., Belmont, CA.