DEPARTMENT OF BUSINESS AND SOCIAL STUDIES

COURSE TITLE: QUANTITATIVE TECHNIQUES

Instructional Material for BBM - distance learning

COURSE OUTLINE

QUANTITATIVE TECHNIQUES

Purpose: To equip students with skills in quantitative techniques necessary for decision making in business

Expected Learning Outcomes of the Course: By the end of the course unit the learners should be able to:-

i) Apply sets theory, probability and linear programming in decision making
ii) Summarize raw data into information fro decision making

Course Content:

Sets theory; Measures of central tendencies; Measures of dispersion; Measures of kurtosis; Probability- Axiom definition of probability; Sample space and events1st and 2nd laws, Bayes theorem; Conditional probability; Probability trees; Linear equations, inequalities and their applications; Utility functions and curves

Course Assessment

Examination - 70%; Continuous Assessment Test (CATS) - 20%; Assignments - 10%; Total - 100%

Recommended Text Books:

i) Sweeny Williams Anderson (2007), *Quantitative Methods For Business,* McGraw Hill, New York
ii) Oakshott Les (2006), *Essential Quantitative Methods For Business, Management And Finance* Routledge, London
iii) B S Sharma(2006); *Quantitative Methods*; Anmol Publications Pvt
iv) Anderson (2007); *Quantitative Methods;* Cengage Learning (Thompson)
v) Louise Swift and Sally Piff (2005), *Quantitative Methods For Business, Management And Finance* McGraw Hill, New York

TABLE OF CONTENT

TOPIC 1

SET THEORY

Objectives

By the end of the topic, the leaner should be able to;

i) Define a set

ii) Use correct set notation

iii) Describe various types of sets

iv) Classify numbers. ie number types

v) Express a worded set in set builder notation

vi) Define the complement of a set

vii) Use Venn diagrams to represent sets

1.1 Introduction

A set is a collection of numbers or objects with the same characteristics eg. the set of all digits we use to form numbers i.e. $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

Notice that the digits have been enclosed in brackets $\{\ \}$.

Sets do not have to contain numbers. For example, a set of vowels is $\{a, e, i, o, u\}$.

We usually use capital letters to represent sets. Every item within the bracket is called an **element** (or a **member**) of the set.

1.1.1 Notation

$\in$ Reads 'is an element of'

$\in$ Reads 'is an element of'

$\{\ \}$ or $\emptyset$ is called the empty set and contains no elements. It is also referred to as the **null set**.

1.1.2 Subsets

If P and Q are two sets, then:

P Q reads "P is a subset of Q" and means that every element in P is also an ele ent of Q.

For example:

Let P = $\{2, 3, 5\}$ and Q = $\{1, 2, 3, 4, 5, 6\}$.

As every element is set P is contained in set Q, then P Q.

1.1.3 Union and Intersection

(i) P $\cup$ Q is the union of sets P and Q and is made up of all elements which are in P or Q.

(ii) P $\cap$ Q is the intersection of sets P and Q and is made up of all elements which are both in P and Q.

For example, if P = $\{1, 2, 3\}$ and Q = $\{2, 3, 5\}$, then P $\cup$ Q = $\{1, 2, 3, 5\}$ and P $\cap$ Q = $\{2, 3\}$.

1.1.4 Disjoint Sets

If two sets have no common elements, then their intersection is an empty set and we say the sets are disjoint.

Which of these sets are disjoint?

(a) A = $\{3, 5, 7, 9\}$ and B = $\{2, 4, 6, 8\}$

(b) $\{3, 5, 6, 7, 8, 10\}$ and $\{4, 9\ 10\}$

1.1.5 Notation

N = $\{1, 2, 3, 4, 5, \dots\dots\}$ is the set of all natural numbers.

Z = $\{0, \pm 1, \pm 2, \pm 3, \}$ is the set of all integers.

Z_+ = $\{1, 2, 3, \dots\dots\}$ is the set of all positive integers.

Z_- = $\{-1, -2, -3, \dots\dots\}$ is the set of all negative integers.

$Q = \left\{\frac{p}{q} \text{ where p and q are integers of } q \neq 0\right\}$ is a set of all rational numbers.

R = $\{Real\ numbers\}$ is the set of all real numbers i.e. all numbers which can be pl ced on a number line.

Notice that Z = Z_- $\cup \{\ \} \cup$ Z_+ which simply means a set of all positive integers.

1.1.6 Irrational Numbers

These numbers which cannot be expressed in the form

$\frac{p}{q} \text{ where p and q are integers of } q \neq 0$.

Numbers such as $\sqrt{2}, \sqrt{3}, \sqrt{5}, \sqrt[3]{2}, \sqrt[3]{11}, \sqrt[3]{10}$ are all irrational.

This means that their decimal expansions do not terminate or recur.

Other irrational numbers include number like which is the ratio of a circle's circumference to its diameter and e = 2.718281…

Rational numbers are denoted by Q while irrational numbers are denoted by Q'.

1.1.7 Counting in a Set

The set A = $\{a, e, i, o, u\}$ has 5 elements, so we write n(A) = 5. The set $\{2, 4, 6, 8, \ldots\ldots, 98, 100\}$ has 50 elements, so we write n(A) = 50.

1.1.8 Finite & Infinite Sets

A set that has a finite number of elements is said to be a finite set.

Infinite sets are sets with infinitely many members. For example Z = $\{0, \pm 1, \pm 2, \pm 3, \pm 4 \ldots\ldots\}$ is an infinite set.

Note: R = $\{All\ numbers\ between\ 2.1\ and\ 2.2\}$ is also an infinite set..

1.2 Set Builder Notation

$A\ x\ x\ Z, 2\ x\ 4$ reads "The set of all x such that x is an integer between -2 and 4 including -2 and 4.

$A \quad x : x \quad Z , 2 \quad x \quad 4$ reads the same as above.

Example 1

Given , $A \quad x | x \quad Z, 3 \quad x \ 10$ }write down:

(a) The meaning of the set builder notation

(b) List the elements of set A

(c) Find n(A)

(a) The set of all x such that x is an integer between 3 and 10 including 3 and 10.

(b) A ={3,4,5,6,7,8,9,10}

(c) n(A) = 8 as these are 8 elements

1.3 Complement of a Set

If, for example, we are interested in natural numbers from 1 to 20 and we want to consider subsets of this set, then:

$\cup$ = $\{x: x \in N, 1 \leq x \leq 20\}$ is the **universal set** in this situation. The symbol $\cup$ is used to represent the universal set. In some cases, the Greek letter (epsilon) is used to represent the universal set.

Definition: The complement of a set A, denoted as A^I is the set of all elements of $\cup$ which are not in A.

If the universal set is ....= = and A = , then the complement of A, denoted as A_1 is $\{2,4,6\}$

n(A) is used to represent the number of elements in set A.

Three obvious relationships are observed concerning A and A_1. These are:

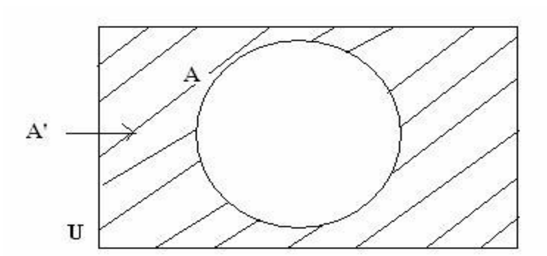
(i) A A^I U

(ii) $AA^{I}0$

(iii) $n(A^{I})\ n(U)\ n(A)$

1.4 Venn diagrams

Venn diagrams are used to represent sets of objects, numbers or things. The universal set is usually represented by a rectangle whereas sets within it are usually represented by circles or ellipses. For example:

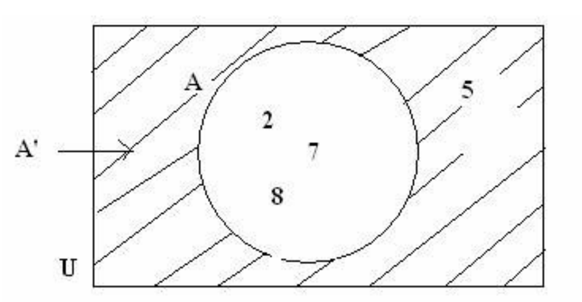


Is a Venn diagram which shows set A within the universal set U .A_1 the complement of A is shaded. In this case, U could be $\{2, 3, 5, 7, 8\}$.

A = $\{2, 7, 8\}$

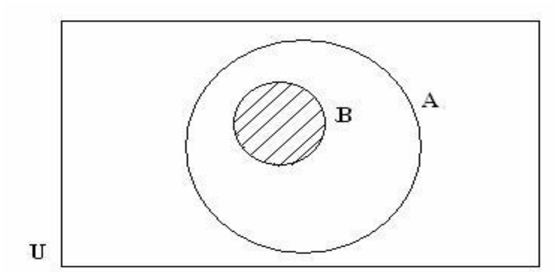
A_1 = $\{3, 5\}$

These elements can be shown as:



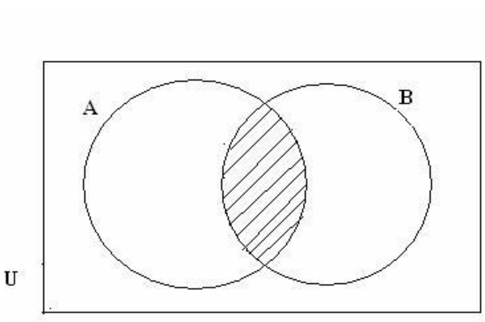
1.4.1 SUBSETS

As B A, the B is contained within A as every element of B is also in A.



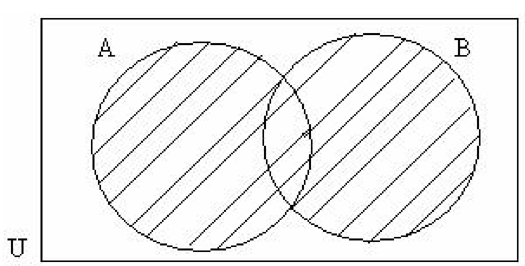
1.4.2 INTERSECTION

As A ∩ B consists of all elements common to both A and B then A∩B is the shaded region.

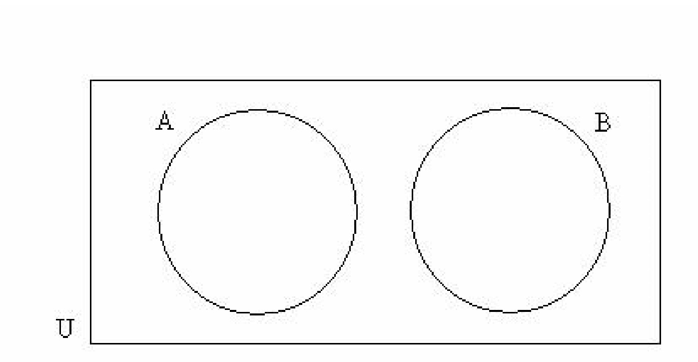


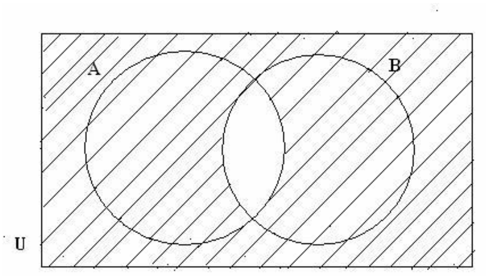
1.4.3 UNION

As A∪B consists of all elements in A or B (or both A and B), then A∪B is the shaded region.



If two sets A and B are disjoint and exhaustive, then $A \cap B = \emptyset$ and $U = A \cup B$. This can be shown as:



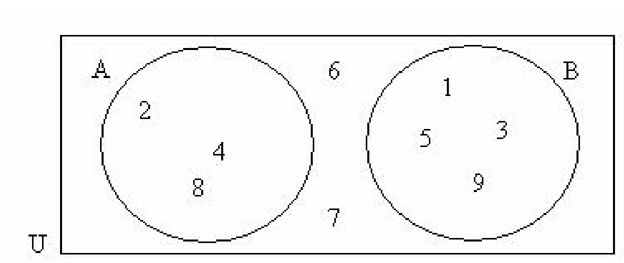


Example:

Given U = {1, 2, 3, 4, 5, 6, 7, 8, 9}, Illustrate on a Venn diagram the sets

(a) A = {2, 4, 8} and B = {1, 3, 5, 9}.

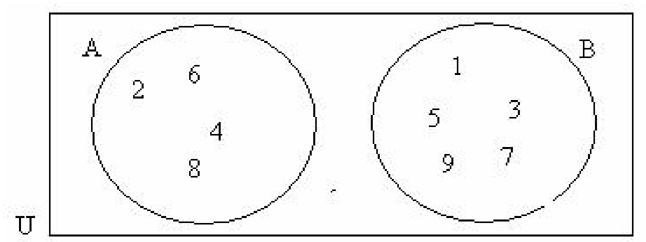
Solution



In this case $A \cup B \neq U$

(b) A= {2, 4, 6, 8} B= {1, 3, 5, 7, 9}

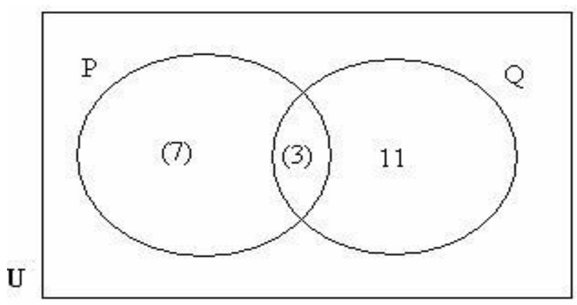
Solution



In this case, A∪B = ∩ i.e. A and B are disjoint and mutually exclusive.

1.5 Numbers in Regions

Consider the Venn diagram below.



If (3) means there are e elements in set P∩Q, how many elements are there in:

(a) P (b) Q_1 (c) P..........Q (d) P but not Q (e) Q but not P (f) neither P nor Q?

Solution

(a) n(P) = 7 + 3 = 10

(b) $n(Q_1)$ = 7 + 4 = 11

(c) n(P∪Q) = 7 + 3 + 11 = 21

(d) n(P but not Q) = 7

(e) n(Q but not P) = 11

(f) n(neither P nor Q) = 4

Example 3

(a) A squash club has 27 members. 19 have black hair, 14 have brown eyes and it have both black hair and brown eyes.

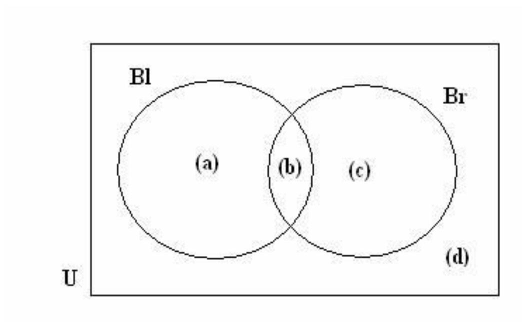
(i) Place this information on a Venn diagram.

(ii) Hence, find the number of members with:

IBlack or brown eyes

IIBlack hair but not brown eyes

(a) Let Bl represent black hair and Br represent brown eyes.



a = b + c + d = 27

a + b = 19

B + c = 14

But b = 11

∴ a = 19 – 11 = 8

c = 14 – 11 = 3

d = 27 – (11 + 8 + 3) = 27 – 22 = 5

∴ I) n(Bl∪Br) = 8 + 11 + 3 = 22

Or 27-d =27 – 5 = 22

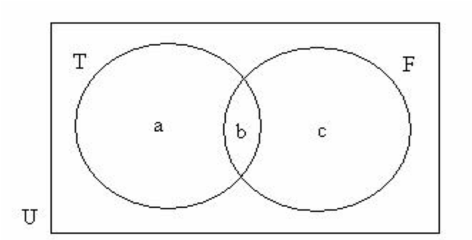
II) n(Bl but not Br) = 8

Example 4

A platform diving squad of 25 had 18 members who dive from 10m and 17 who dive from 4m. How many divide from both platforms?

Solution

Let T represent those who dive from 10m and F represent those who dive from 4m. Then



a + b = 18

b + c = 17

a + b + c = 25

a = 25 – 17

= 8

b = 18 – 8 = 10

8 + c = 17

c= 7

∴ c = 8, b = 10 and h = 1

(a) P (reads A only) = 18% (shaded)

(b) n(B or C) = n(B∪C)

= 100% - 1% - 18%

= 81%

(c) n(A or B but not C) = 18% + 5% + 24%

= 47%

Question

A city has three newspapers A, B and C . Of the adult population, 1% read none of these newspapers. 36% read A, 40% read B, 52% read C, 8% read A and B, 11% read B and C, 13%

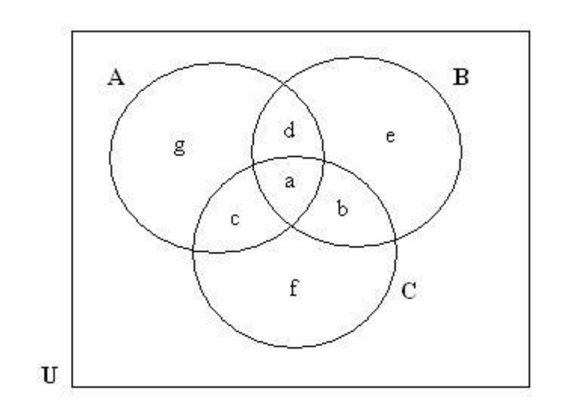
read A and C and 3% read all the three newspapers. What percent of the adult population read:

(a) Newspaper A only

(b) Newspaper B or Newspaper C

(c) Newspaper A or B but not C

Solution



From the given information:

a = 3

a + d = 8

a + b = 11

a + c = 13

∴ d = 5, b = 8, c = 10

g + 3 + 5 + 10 = 36 ∴ g = 18

e + 3 + 5 + 8 + 40 ∴ e = 24

f + 3 + 8 + 10 = 52 ∴ f = 31

1.6 Revision Exercise

1. Given M = $\{2, 3, 5, 7, 8, 9\}$ and N = $\{3, 4, 6, 9, 10\}$

   (a) Indicate True or False

   (i) $4 \in M$

   (ii) $6 \in M$

   (b) List:

   (i) $M \cap N$

   (ii) $M \cup N$

   (c) Is:

   (i) M N?

   (ii) $\{9, 6, 3\}$ N?

2. Write in set language.

   (a) 5 is an element of D

   (b) d is not an element of all vowels

   (c) $\{2, 5\}$ is a subset of $\{1, 2, 3, 4, 5, 6\}$

3. Find $A \cap B$ and $A \cup B$ for A = $\{6, 7, 9, 11, 12\}$ and B = [illegible].

4. State whether the following sets are finite or infinite:

   (a) $\{x: x \in Z, -2 \leq x \leq 1\}$,

   (b) $\{x: x \in R, -2 \leq x \leq 1\}$,

   (c) $\{x: x \in Q, 0 \leq x \leq 1\}$,

   (d) $\{x: x \in Z, x \geq 5\}$,

5. Write down the meaning of the following in set builder notation.

   (a) A = $\{x: x \in Z, -1 \leq x \leq 7\}$,

   $\{x: x \in N, -2 \leq x \leq 5\}$

   (c) A = $\{x: x \in R, 0 \leq x \leq 1\}$,

6. Find C^I given that:

   ∪ = {all intergers} and C = {all even numbers}

(b) $C = \{x: x \in Z, x \geq 2\}$ and $\cup = Z$.

1. If $\cup = \{positive\ intergers\}$, $P = \{multiple\ of\ 4\ less\ than\ 50\}$ and $Q = \{multiples\ pf\ 6\ less\ than\ 50\}$.

   (a) List P and Q

   (b) Find $P \cap Q$

   (c) Find $P \cup Q$

   (d) Verify that $n(P \cup Q) = n(P) + n(Q) - n(P \cap Q)$

8. Represent sets A and B on a Venn diagram.

   (a) $U = \{2, 3, 4, 5, 6, 7\}$, $A = \{2, 4, 6\}$, $B = \{5, 7\}$

   (b) $U = \{1, 2, 3, 4, 5, 6, 7\}$, $A = \{2, 4, 5, 6\}$, $B = \{1, 4, 6, 7\}$

   (c) $U = \{3, 4, 5, 7\}$, $A = \{3, 4, 5, 7\}$, $B = \{3, 5\}$

   (d) $U = \{1, 2, 3, 4, 5, 6, 7\}$, $A = \{2, 4, 6\}$, $B = \{3, 5, 7\}$

9. Given that $U = \{x: x \in Z, 1 \leq x \leq 10\}$

   $A = \{odd\ numbers < 10\}$, $B = \{prime\ numbers < 10\}$,

   (a) List sets A and B

   (b) Find $A \cap B$ and $A \cup B$

   (c) Represent sets A and B on a Venn diagram

10. Given that $U = \{x: x \in Z, x \leq 30\}$

    $R = \{primes < 30\}$

    $S = \{composites\ less\ than\ 30\}$

    (a) List sets R and S

(b) Find R∩S and R∪S

(c) Represent the sets R and S on a Venn diagram

11. If (Z) means that there are Z elements in the set A∩B, give the number of elements in:

(b) B (b) A (c) ……………

(d) neither B nor A (e) A but not B

12. Given $n(U) = 50,$ n (S) = 30, n(R) = 25 and $n(R \cap S)$ = 48, find

(a) $n(R \cap S)$

(b) n(S but not R)

Selected References

i. Basic set theory by A shen and N.K Verchagin

ii. Basic set theory by A Levy

iii. Quaantitative Techniques Simplifiedby N.A Saleemi

TOPIC 2

PROBABILITY

Objectives

By the end of the topic, the learner should be able to

i) Define probability

ii) Use the formula for finding the probability of an event. to find the probabilities of events with equally likely and non-equally likely outcomes.

iii) To determine the sample space of an experiment by examining each possible outcome.

iv) To understand the theory behind mutually exclusive and non-mutually exclusive events and to classify an experiments accordingly.

v) To find the probability of mutually exclusive events by applying the addition rule.

vi) To understand the theory behind independent events and to use the multiplication rule to compute related probabilities.

vii) To understand the theory behind conditional probability. To derive the formula for finding conditional probabilities, and

viii) To compute related probabilities using tree diagrams.

2.1 Introduction

In the study of chance, we need a mathematical method to describe the likelihood of events happening.

Probability is the study of chance (likelihood) of events happening.

The events are also referred to as outcomes. The study of probability has vitally important applications in physical and biological sciences, economics, politics, sport, insurance, quality control, planning in industry, and a host of other areas.

The range of probability

Probability is studied by assigning a number which lies between 0 and 1.

An event which a 0% chance of happening ("It did not rain on Tuesday" i.e., it is impossible) is assigned a probability of 0.

An event which has a 100% chance of happening is assigned a probability of 1

All other events can then be assigned a probability between 0 and 1.

2.2 Definitions

Probability Experiment

Process which leads to well-defined results called outcomes

Outcome

The result of a single trial of a probability experiment .For example, in tossing of a coin, a head or a tail constitute the two possible outcomes.

Sample Space

Set of all possible outcomes of a probability experiment It is denoted by 'S' and its number of elements are n(s). For example; In throwing a dice, the number that appears on top is any one of 1, 2, 3,4,5,6. So here,

S = {1, 2, 3, 4, 5, 6} and n(s) = 6

Event

One or more outcomes of a probability experiment. In other words every subset of a sample space is called an event and is denoted by 'E'.

Example; In throwing a dice S = {1, 2, 3, 4, 5, 6}, the appearance of an even number will be the event E = {2, 4, 6}.

Clearly, E is a subset of S.

Classical Probability

Uses the sample space to determine the numerical probability that an event will happen. Also called theoretical probability.

Equally Likely Events

Events which have the same probability of occurring. For example, in tossing of a fair coin, the appearance of a head or a tail are equally likely events.

Complement of an Event

All the events in the sample space except the given events.

Empirical Probability

Uses a frequency distribution to determine the numerical probability. An empirical probability is a relative frequency.

Subjective Probability

Uses probability values based on an educated guess or estimate. It employs opinions and inexact information.

Mutually Exclusive Events

Two events which cannot happen at the same time. For example in tossing a coin it is not possible to obtain a head and a tail in a single toss. Similarly, events such as "It rained on Tuesday" and "It did not rain on Tuesday" are mutually exclusive events. When calculating the probabilities of mutually exclusive events you add up the probabilities.

Disjoint Events

Another name for mutually exclusive events.

Independent Events

Two events are independent if the occurrence of one does not affect the probability of the other occurring. For example the events "It rained on Tuesday" and "My chair broke at work" are two independent events. When calculating the probabilities of independent events you multiply the probabilities. If a coin is tossed twice, the two outcomes will be unrelated Obtaining a head on the first toss does not give a bearing on what will show up in the next toss.

Dependent Events

Two events are dependent if the first event affects the outcome or occurrence of the second event in a way the probability is changed.

Conditional Probability

The probability of an event occurring given that another event has already occurred.

Bayes' Theorem

A formula which allows one to find the probability that an event occurred as the result of a particular previous event.

2.3 Axiomatic Approach to Probability

The modern theory of probability is based on the axiomatic approach introduced by the Russian mathematician A.N Kolmogorov in 1930's. In axiomatic approach, some concepts are laid down at certain properties or postulates commonly known as axioms are defined and from these axioms the entire theory is developed by logic of deduction.

2.3.1 Probability Axioms

Non Negativity

This means that $0 \leqq P(E) \leqq 1$ for every event in S.[S = sample space]

Proof: Let 'S' be the sample space and 'E' be the event. Then

$$0 \leqq n(E) \leqq n(S)$$

[The number of elements in a non zero outcome cannot exceed the sample space]

Dividing through by n(S)

$0/n(S) \leqq \quad n(E)/n(S) \leqq n(S)/n(S)$

i.e, $0 \leqq P(E) \leqq 1$

Sample Space refers to the total outcomes of a probability theory experiment e.g the above tossing of a coin experiment, etc

Universal Set:

This means P(S) = 1 [S = sample space]

Proof: In the sure event, n(E) = n(S)

[Since the number of elements in the event 'E' will be equal to the number of elements in the sample space 'S']

By definition of probability:

P(S) = n(S)/n(S) = 1

Which implies P(S) = 1

The addition axiom

This means that P {E1 + E2) = P (E1) + P (E2) for any two or even more mutually exclusive events in S.

$$P(E1 \cup E2) = P(E1) + P(E2)$$

Provided E1 and E2 are mutually exclusive.

i.e. $E1 \cap E2$ is empty.

2.4 Mutually exclusive events

Two or more events are called mutually exclusive if the occurrence of any one of them precludes the occurrence of any of the others. The probability of occurrence of two r more mutually exclusive events is the sum of the probabilities of the individual events

Sometimes when one event has occurred, the probability of another event is excluded (referring to the same given occasion or trial).

For example, throwing a die once can yield a 5 or 6, but not both, in the same toss. The probability that either a 5 or 6 occurs is the sum of their individual probabilities.

$$p = p_1 + p_2$$

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

$$= \frac{1}{3}$$

EXAMPLE: From a bag containing 5 white balls, 2 black balls, and 11 red balls, 1 ball is drawn. What is the probability that it is either black or red?

SOLUTION: The draw can be made in 18 ways. The choices are 2 black balls and 11 red balls, which are favorable, or a total of 13 favorable choices. Then, the probability of success is

$$p = \frac{13}{18}$$

Since drawing a red ball excludes the drawing of a black ball, and vice versa, the two events are mutually exclusive; so the probability of drawing a black ball is

$$p_1 = \frac{2}{18}$$

and the probability of drawing a red ball is

$$p_2 = \frac{11}{18}$$

Therefore, the probability of success is

$$p = p_1 + p_2$$

$$= \frac{2}{18} + \frac{11}{18} = \frac{13}{18}$$

EXAMPLE: What is the probability of drawing either a king, a queen, or a jack from a deck of playing cards?

SOLUTION: The individual probabilities are

$$\text{king} = \frac{4}{52}$$

$$\text{queen} = \frac{4}{52}$$

$$\text{jack} = \frac{4}{52}$$

Therefore, the probability of success is

$$p = \frac{4}{52} + \frac{4}{52} + \frac{4}{52}$$

$$= \frac{12}{52}$$

$$= \frac{3}{13}$$

EXAMPLE: What is the probability of rolling a die twice and having a 5 and then a 3 show or having a 2 and then a 4 show?

SOLUTION: The probability of having a 5 and then a 3 show is

$$p_1 = \frac{1}{6} \cdot \frac{1}{6}$$

$$= \frac{1}{36}$$

and the probability of having a 2 and then a 4 show is

$$p_2 = \frac{1}{6} \cdot \frac{1}{6}$$

$$= \frac{1}{36}$$

Then, the probability of either p_1 or p_2 is

$$p = p_1 + p_2$$

$$= \frac{1}{36} + \frac{1}{36}$$

$$= \frac{1}{18}$$

PRACTICE PROBLEMS:

1 When tossing a coin, you have what probability of getting either a head or a tail?

2. A bag contains 12 blue, 3 red, and 4 white marbles. What is the probability of

drawing a. in 1 draw, either a red or a white marble?

b. in 1 draw, either a red, white, or blue marble?

c. in 2 draws, either a red marble followed by a blue marble or a red marble followed by a red marble?

3. What is the probability of getting a total of at least 10 points in rolling two dice? (HINT: You want either a total of 10, 11, or 12.)

ANSWERS:

1. 1

2. a. $\frac{7}{19}$

   b. 1

   c. $\frac{7}{57}$

3. $\frac{1}{6}$

2.5 Dependent events

In some cases one event is dependent on another; that is, two or more events are said to be dependent if the occurrence or nonoccurrence of one of the events affects the probabilities of occurrence of any of the others.

Consider that two or more events are dependent. If p_1 is the probability of a first event; p_2 the probability that after the first happens, the second will occur; p_3 the probability that after the first and second have happened, the third will occur; etc., then the probability that all events will happen in the given order is the product $p_1 - p_2 - p_3$.

EXAMPLE: A box contains 3 white marbles and 4 black marbles. What is the probability of drawing 2 black marbles and 1 white marble in succession without replacement?

SOLUTION: On the first draw the probability of drawing a black marble is

$$p_1 = \frac{4}{7}$$

On the second draw the probability of drawing a black marble is

$$p_2 = \frac{3}{6}$$

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

On the third draw the probability of drawing a white marble is

$$p_3 = \frac{3}{5}$$

Therefore, the probability of drawing 2 black marbles and 1 white

marble is

$$p = p_1 \cdot p_2 \cdot p_3$$

$$= \frac{4}{7} \cdot \frac{1}{2} \cdot \frac{3}{5}$$

$$= \frac{6}{35}$$

EXAMPLE: Slips numbered 1 through 9 are placed in a box. If 2 slips are drawn, without replacement, what is the probability that

1. both are odd?

2. both are even?

SOLUTION:

1. The probability that the first is odd is

$$p_1 = \frac{5}{9}$$

and the probability that the second is odd is

$$p_2 = \frac{4}{8}$$

Therefore, the probability that both are odd is

$$p = p_1 \cdot p_2$$

$$= \frac{5}{9} \cdot \frac{4}{8}$$

$$= \frac{5}{18}$$

2. The probability that the first is even is

$$p_1 = \frac{4}{9}$$

and the probability that the second is even is

$$p_2 = \frac{3}{8}$$

Therefore, the probability that both are even is

$$p = p_1 \cdot p_2$$

$$= \frac{4}{9} \cdot \frac{3}{8}$$

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

A second method of solution involves the use of combinations.

1. A total of 9 slips are taken 2 at a time and 5 odd slips are taken 2 at a time; therefore,

$$p = \frac{{}_5C_2}{{}_9C_2} = \frac{5}{18}$$

2. A total of ${}_9C_2$ choices and 4 even slips are taken 2 at a time; therefore,

$$p = \frac{{}_4C_2}{{}_9C_2}$$

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

Practice Problems:

In the following problems assume that no replacement is made after each selection:

1. A box contains 5 white and 6 red marbles. What is the probability of successfully drawing, in order, a red marble and then a white marble?

2. A bag contains 3 red, 2 white, and 6 blue marbles. What is the probability of drawing, in order, 2 red, 1 blue, and 2 white marbles?

3. Fifteen airmen are in the line crew. They must take care of the coffee mess and line shack cleanup. They put slips numbered 1 through 15 in a hat and decide that anyone who draws a number divisible by 5 will be assigned the coffee mess and anyone who draws a number divisible by 4 will be assigned cleanup. The first person draws a 4, the second a 3, and the third an 11. What is the probability that the fourth person to draw will be assigned

a. the coffee mess?

2.6 Independent events

Two or more events are independent if the occurrence or nonoccurrence of one of the events has no affect on the probability of occurrence of any of the others.

When two coins are tossed at the same time or one after the other, whether one falls heads or tails has no affect on the way the second coin falls. Suppose we call the coins A and B. The coins may fall in the following four ways:

1. A and B may fall heads.

2. A and B may fall tails.

3. A may fall heads and B may fall tails.

4. A may fall tails and B may fall heads.

The probability of any one way for the coins to fall is calculated as follows:

$s=4$

and

n=4

therefore,

$$p = \frac{1}{4}$$

This probability may be determined by considering the product of the separate probabilities; that is,

the probability that A will fall heads is $\frac{1}{2}$

the probability that B will fall heads is $\frac{1}{2}$

and the probability that both will fall heads is

$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

In other words, when two events are independent, the probability that one and then the other will occur is the product of their separate probabilities.

EXAMPLE: A box contains 3 red marbles and 7 green marbles. If a marble is drawn, then replaced, and another marble is drawn, what is the probability that both marbles are red?

SOLUTION: Two solutions are offered. First, by the principle of choice, 2 marbles can be selected in 10 ways. The red marble may be selected on the first draw in three ways and on the second draw in three ways; and by the principle of choice, a red marble may be drawn on both trials in 3 ways. Then the required probability is

$$p = \frac{9}{100}$$

The second solution, using the product of independent events, follows: The probability of drawing a red marble on the first draw is $\frac{3}{10}$, and the probability of drawing a red marble on the second draw is $\frac{3}{10}$. Therefore, the probability of drawing a red marble on both draws is the product of the separate probabilities or

$$p = \frac{3}{10} \cdot \frac{3}{10} = \frac{9}{100}$$

2.7 Sample Space:

The set of all possible outcomes of a random experiment is called the sample space for that experiment. It is usually denoted by S.

Examples:

(i) When a coin is tossed either a head or a tail will come up. If H denotes the occurrence of head and T denotes the occurrence of tail, the

Sample space S = (H, T)

Note:

If a denotes the occurrence of head and b denotes the occurrence of tail, then

Sample space S = (a, b).

When two coins are tossed.

Sample Space S = {(H, H), (H, T), (T, H), (T, T)}

where (H, H) denotes the occurrence of head on the first coin and occurrence of head on the second coin. Similarly (H, T) denotes the occurrence of head on the first coin and occurrence of tail on the second coin.

When a die is thrown any one of the numbers 1, 2, 3, 4, 5 and 6 will come up. Therefore, sample space

S={1,2,3,4,5,6}

Here 1 denotes the occurrence of 1, 2 denotes the occurrence of 2 and so on.

Note:

If occurrence of 1, 2, 3, 4, 5 and 6 are denoted by *a, b, c, d, e,* f respectively then sample space S = *{a, b, c. d, e,f)*

(iv) When two balls are drawn from a bag containing 3 red and 2 black balls.

Sample space**,**
S = {(R_1, R_2), (R_1, R_3), (R_2, R_3), (B_1, B_2), (R_1, B_1), (R_1, B_2), (R_2, B_1) (R_2, B_2), (R_3, B_1), (R_3, B_2)}

Note:
Here R_1, R_2, R_3 have been used for the occurrence of the three red balls whether red balls are identical or not.

Example

When one ball is drawn at random from a bag containing 3 black and 4 red balls (balls of the same colour being identical or different), then sample space

S = {B_1, B_2, B_3, R_1, R_2, R_3, R_4}\ n(S) = 7

Here the three black balls may be denoted by B|, B_2 and B_3 even if they are identical because while finding probability only number of black and red balls are to be taken into account.

Let E = the event of occurrence of a red ball.

Then E = {R,, R_2, R_3, R,) .-. n(E) = 4

Example 2

When two coins are tossed, sample space S = (HH, HT, TH, TT}.

Let E = the event of occurrence of one head and one tail, then

E = {HT, TH}

Now P(E) = n(E) / n(S) = 2/4 =

1/2 **2.8 Conditional Probability**

The ***conditional probability*** of an event *B* is the probability that the event will occur given the knowledge that an event *A* has already occurred. This probability is written *P(B|A)*, notation for the *probability of B given A*. In the case where events *A* and *B* are *independent* (where event *A* has no effect on the probability of event *B*), the conditional probability of event *B* given event *A* is simply the probability of event *B*, that is *P(B)*.

If events *A* and *B* are not independent, then the probability of the *intersection of A and B* (the probability that both events occur) is defined by *P(A and B)* = *P(A)P(B|A)*.

From this definition, the conditional probability *P(B|A)* is easily obtained by dividing by *P(A)*:

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$$

Note: This expression is only valid when P(A) is greater than 0.

Example 1:

In a card game, suppose a player needs to draw two cards of the same suit in order to win. Of the 52 cards, there are 13 cards in each suit. Suppose first the player draws a heart. Now the player wishes to draw a second heart. Since one heart has already been chosen, there are now 12 hearts remaining in a deck of 51 cards. So the conditional probability *P(Draw second heart|First card a heart)* = 12/51.

Suppose an individual applying to a college determines that he has an 80% chance of being accepted, and he knows that dormitory housing will only be provided for 60% of all of the accepted students. The chance of the student being accepted *and* receiving dormitory housing is defined by *P(Accepted and Dormitory Housing)* = *P(Dormitory Housing|Accepted)P(Accepted)* = (0.60)*(0.80) = 0.48.

To calculate the probability of the intersection of more than two events, the conditional probabilities of *all* of the preceding events must be considered. In the case of three events, *A*, *B*, and *C*, the probability of the intersection *P(A and B and C)* = *P(A)P(B|A)P(C|A and B)*.

Example 2:

Consider the college applicant who has determined that he has 0.80 probability of acceptance and that only 60% of the accepted students will receive dormitory housing. Of the accepted students who receive dormitory housing, 80% will have at least one roommate. The probability of being accepted *and* receiving dormitory housing *and* having no roommates is calculated by: *P(Accepted and Dormitory Housing and No Roommates)* = *P(Accepted)P(Dormitory Housing|Accepted)P(No Roomates|Dormitory Housing and Accepted)* = (0.80)*(0.60)*(0.20) = 0.096. The student has about a 10% chance of receiving a single room at the college.

Example 3:

The question, "Do you smoke?" was asked of 100 people. Results are shown in the table.

	Yes	**No**	**Total**
Male	19	41	**60**
Female	12	28	**40**
Total	**31**	**69**	**100**

What is the probability of a randomly selected individual being a male who smokes? This is just a joint probability. The number of "Male and Smoke" divided by the total = 19/100 = 0.19

What is the probability of a randomly selected individual being a male? This is the total for male divided by the total = 60/100 = 0.60. Since no mention is made of smoking or not smoking, it includes all the cases.

What is the probability of a randomly selected individual smoking? Again, since no mention is made of gender, this is a marginal probability, the total who smoke divided by the total = 31/100 = 0.31.

What is the probability of a randomly selected male smoking? This time, you're told that you have a male - think of stratified sampling. What is the probability that the male smokes? Well, 19 males smoke out of 60 males, so 19/60 = 0.31666...

What is the probability that a randomly selected smoker is male? This time, you're told that you have a smoker and asked to find the probability that the smoker is also male. There are 19 male smokers out of 31 total smokers, so 19/31 = 0.6129 (approx)

2.9 Bayes' Theorem

Another important method for calculating conditional probabilities is given by ***Bayes's formula***. The formula is based on the expression $P(B) = P(B|A)P(A) + P(B|A_c)P(A_c)$, which simply states that the probability of event B is the sum of the conditional probabilities of event B given that event A has or has not occurred. For independent events A and B, this is equal to $P(B)P(A) + P(B)P(A_c) = P(B)(P(A) + P(A_c)) = P(B)(1) = P(B)$, since the probability of an event and its complement must always sum to 1. Bayes's theorem is defined as follows:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

Example

Suppose a voter poll is taken in three states. In state A, 50% of voters support the liberal candidate, in state B, 60% of the voters support the liberal candidate, and in state C, 35% of the voters support the liberal candidate. Of the total population of the three states, 40% live in state A, 25% live in state B, and 35% live in state C. Given that a voter supports the liberal candidate, what is the probability that she lives in state B?

By Bayes's formula,

P(Voter lives in state B|Voter supports liberal candidate) =

P(Voter supports liberal candidate|Voter lives in state B)P(Voter lives in state B)/

(P(Voter supports lib. cand.|Voter lives in state A)P(Voter lives in state A) +

P(Voter supports lib. cand.|Voter lives in state B)P(Voter lives in state B) +

P(Voter supports lib. cand.|Voter lives in state C)P(Voter lives in state C))

= (0.60)*(0.25)/((0.50)*(0.40) + (0.60)*(0.25) + (0.35)*(0.35))

= (0.15)/(0.20 + 0.15 + 0.1225) = 0.15/0.4725 = 0.3175.

The probability that the voter lives in state B is approximately 0.32.

2.10 Tree Diagrams

Tree diagrams, as the name suggests, look like a tree as they branch out symmetrically. They are used to help you visualize more complicated probability problems.

A favourite with maths examiners is to get you to use tree diagrams to show the probabilities of you picking a red then a white ball out of a bag of red and white balls etc. This is not very realistic, so here is an example of how tree diagrams can be used in real life:

A box of chocolates is randomly selected from a production line to check to see if any of the chocolates are faulty. Each box contains 12 soft-centres and 8 hard-centres. Two chocolates are randomly selected from the box and are tested to see if they have any faults.

What is the probability of selecting two soft-centred chocolates?

What is the probability of selecting a soft-centred and a hard-centred chocolate?

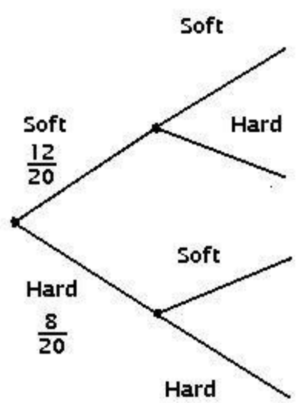
To answer these questions, we can draw a tree diagram. First you need to work out some probabilities to get the tree diagram started.

Example 1

If we have 12 soft-centred and 8 hard-centred chocolates in a box, we have a tot l of 20 to choose from.

When we select the first chocolate the probability of getting a soft-centre = $\frac{12}{20}$ and the probability of getting a hard-centre = $\frac{8}{20}$.

Now we can draw the first branches of the tree diagram:



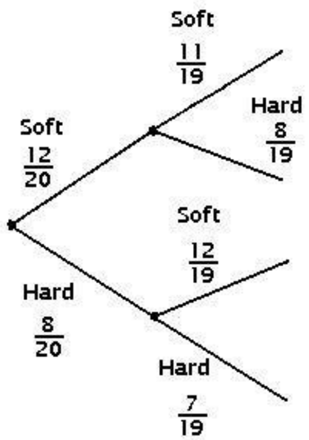
Note that when 2 branches come from a single point the total of the probabilities on each branch = 1 (this can make calculations quicker).

After first selecting a Soft-centred chocolate, the tree diagram indicates that there are two things that can happen. We can select another Soft-centre or we can select a hard-centre.

You now need to work out the probability of selecting another soft-centre if you've selected one already. Note that if you've already selected a chocolate you will only have 19 in total left in the box to choose from when you select the second chocolate. Note also that if you selected a soft centre first then you will only have 11 soft-centres left in the box to choose from. So the probability of choosing a second soft-centre = $\frac{11}{19}$.

Using the fact that total of probabilities on two branches = 1, we can say that the probability of getting a hard-centre as the second chocolate = $\frac{19}{19} - \frac{11}{19} = \frac{8}{19}$.

Using similar methods we work out the rest of the probabilities and put them on the tree as follows:



Now we can find the probability of selecting two soft-centred chocolates:

P(Soft AND then Soft) = $\frac{12}{20} \times \frac{11}{19} = \frac{132}{380}$

This fraction can be simplified a bit further (divide by two a couple of times) to give $\frac{33}{80}$

We can also find the probability of selecting a soft-centred and a hard-centred chocolate. Note that there are two ways to get this result: Select a soft-centre then a hard-centre or select a hard-centre then a soft-centre.

P(Hard AND then Soft) = $\frac{8}{20} \times \frac{12}{19} = \frac{96}{380}$

P(Soft AND then Hard) = $\frac{8}{20} \times \frac{12}{19} = \frac{96}{380}$

We add these together to get the answer:

P(Hard then Soft OR Soft then Hard) = $\frac{96}{380} + \frac{96}{380}$

This answer can be simplified to give $\frac{6}{100}$

Note that the tree diagram representation of this experiment involves two parts, 'the first toss of the coin' and 'the second toss of the coin'.

2.11 Revision Exercise

1. If a die is tossed twice, what is the probability of rolling a 2 followed by a 3?

2. A box contains 2 white, 3 red, and 4 blue marbles. If after each selection the marble is replaced, what is the probability of drawing, in order

a) a white then a blue marble?

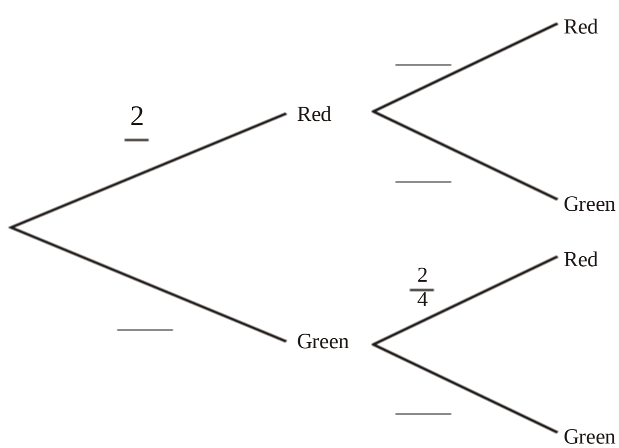
b) a blue then a red marble?

c) a white, a red, then a blue marble?

3. A bag contains two red sweets and three green sweets. Jacques takes one sweet from the bag, notes its colour, then eats it. He then takes another sweet from the bag.

Complete the tree diagram below to show all probabilities.

5



In the following problems assume that no replacement is made after each selection:

4. A box contains 5 white and 6 red marbles. What is the probability of successfully drawing, in order, a red marble and then a white marble?

5. A bag contains 3 red, 2 white, and 6 blue marbles. What is the probability of drawing, in order, 2 red, 1 blue, and 2 white marbles?

6. Fifteen airmen are in the line crew. They must take care of the coffee mess and line shack cleanup. They put slips numbered 1 through 15 in a hat and decide that anyone who draws a number divisible by 5 will be assigned the coffee mess and anyone who draws a number divisible by 4 will be assigned cleanup. The first person draws a 4, the second a 3, and the third an 11. What is the probability that the fourth person to draw will be assigned

a. the coffee mess?

b. the cleanup?

Suggested References

i. Probability and Statistics by R.S Pilaai

ii. Schaum's Outlines Probability-3rd edition

iii. Quaantitative Techniques Simplifiedby N.A Saleemi

TOPIC 3

MEASURES OF CENTRAL TENDENCY

Objectives

By the end of the topic, the learner should be able to

i) Define and calculate the various types of averages such as;

- Arithmetic mean
- Median
- Mode
- Weighted mean
- Geometric mean
- Harmonic mean

ii) Choose an appropriate measure of central tendency for a given situation using the characteristics of the above averages

iii) Describe the qualities of a good average.

iv) Define skewness and distinguish between the normal, the positively skewed and the negatively skewed distributions.

3.1 Averages

Measures central tendency or measures of location or simple averages are used for comparing frequency distributions. There are several different measures of central tendency and each has its own advantages and disadvantages. It is of importance to choose the appropriate measure for the particular comparison we wish to make.

3.1.1 Qualities of a good average

The measurement of the values around which the data is scattered is known as measures of central tendency or averages. The qualities of a good average are as follows: -

(i) It shall be rigidly defined

(ii) It should be based on all values

(iii) It should be easily understood and calculated

(iv) It should be least affected by the fluctuations of sampling.

(v) It should be capable of further algebraic or statistical treatment / or analysis.

(vi) It should be least affected by extreme values.

3.1.2 Types of averages

The following are the important types of averages: -

(i) Arithmetic mean or simple average

(ii) Median

(iii) Mode

(iv) Weighted Mean

(v) Geometric mean

(vi) Harmonic mean

3.2 Arithmetic Mean

3.2.1Arithmetic Mean of Ungrouped Data

The arithmetic mean, , or simply called mean, is obtained by adding together all of the measurements and dividing by the total number of measurements taken. Mathematically it is given as

$$\frac{X}{n} \quad \text{For ungrouped data}$$

Example 1

Given the following set of ungrouped data:

20, 18, 15, 15, 14, 12, 11, 9, 7, 6, 4, 1

Find the mean of the ungrouped data.

$$mean \quad \frac{20\ 18\ 2\ 15\ 14\ 12\ 11\ 9\ 7\ 6\ 4\ 1}{12}$$

$$\frac{132}{12}$$

$$11$$

3.2.2 Arithmetic Mean of Grouped Data

Arithmetic mean is quotient of sum of the given values by the number of the given values Mean of grouped data is calculated by dividing the sum of all observations by the number of observations.

$$ArithmeticMean \quad \frac{fx}{f} \text{ where}$$

x = Individual score and f =frequency

Example2

Find the mean of the following distribution:

x	2	4	7	8	13
f	3	8	8	5	6

Solution:

Calculation of arithmetic mean

x	*F*	*Fx*
2	3	6
4	8	32
7	8	56
8	5	40
13	6	78
	Σ f = 30	Σ fx = 212

$\bar{x} \quad \frac{fx}{f}$

$\frac{212}{31}\ 7.06$

Example 3

scores	x_i	f_i	$x_i \cdot f_i$
10- 20	15	1	15
20- 30	25	8	200
30-40	35	10	350
40 -50	45	9	405
50 -60	55	8	440
60-70	65	4	260
70-80	75	2	150
		42	1 820

The test scores of 42 students are shown in the table below. Calculate the mean

$$\bar{x} = \frac{1,820}{42} = 43.33$$

Arithmetic mean can be used to calculate any numerical data and it is always unique. It is obvious that extreme values affect the mean. Also, arithmetic mean ignores the degree of importance in different categories of data.

3.2.3 Characteristics of the Arithmetic Mean

Each measure of central tendency has its own particular characteristics. The arithmetic mean is fully representative of a frequency distribution as it is based on all and not merely some, of the observations. One weakness of the arithmetic mean is that it may give a wrong picture of the distribution e.g. take 10 boys' visits to the dispensary are tabled as follows:-

9,5,4,8,6,6,8,6,9,49 = 110 visits

Arithmetic mean = 11, gives a false picture of the distribution since all but the tenth boy visited the local dispensary less than eleven times in the year. Another disadvantage is that it is not a physically possible value of the variable e.g. you cannot have a mean number of 6.8 or 10.3 of girls or men.

3.3 The Median

Median is defined as the middle item of all given observations arranged in order.

3.3.1 Median of Ungrouped Data

For ungrouped data, the median is obvious. In case of the number of measurements is even, the median is obtained by taking the average of the middle.

Example 6

The median of the ungrouped data:: 20, 18, 15, 15, 14, 12, 11, 9, 7, 6, 4, 1 is

12 11 2

= 11.5

3.3.2 Median of Grouped Data

The median of a simple frequency distribution of a discrete variable such as the number of visits made by 100 mothers to a local dispensary is simple since we can identify it in the variable arranged in ascending order – see figures below:-

Number of visits by 100 mothers to a local dispensary

Number of visits (variables) (x)	**Number of mothers (frequency) (f)**	**"Less than" cumulative (frequency.)**
4	8	0
5	12	8
6	15	20
7	25	35
8	17	60
9	13	77
10	10	90
11	-	100

Median number of visits will lie between the 50th and 51st observations (100 is an even number). From the table, 35 mothers made less than 7 visits whereas 60 mothers had made less than 8 visits. The median is 7 visits. This shows half the mothers made 7 or fewer visits to the local dispensary and half the mothers made 7 or more visits to the local dispensary.

For grouped data, the median can be found by first identify the class containing the median, then apply the following formula:

$$median = l_1 + \frac{\frac{n}{2} - C}{f_m}(l_2 - l_1)$$

where: l_1 is the lower class boundary of the median class;

n is the total frequency;

C is the cumulative frequency just before the median class; f_m is the frequency of the median;

l_2 is the upper class boundary containing the median.

Example

The weight of 75 pigs was tabulated as follows:

Weight of pig	0-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100	Over 100
No. Of pigs	1	7	8	11	19	10	7	5	4	3

Calculate the median weight.

Solution

Weight (kg) variable	**Number of pigs (f)**	**cumulative frequency**
0-20	1	1
20-30	7	8
30-40	8	16
40-50	11	27
50-60	19	46
60-70	10	56
70-80	7	63
80-90	5	68
90-100	4	72
Over 100	3	75
	75	

We need in calculating the median the weight of the 38^{th} pig. This is the weight of

$\frac{n+1}{2}^{th}$ *pig*

Where n = total number of pigs.

38^{th} pig is somewhere in the 50 and under 60 kg class. 27 pigs weighed less than 50 kg.

46 pigs weighed less than 60kg.

Assumption made: weights of the pigs in the "50 and fewer than 60" classes are evenly spread across the class.

Again, 27 weighed less than 50 kg

We need to account for 38 pigs (to reach the median pig). Therefore, we need to consider, 11 out of 19 pigs in the class (50 an 60) to arrive at the 38^{th} pig.

Median = 11/19, thus, of the way across the 50 and under 60 kg class interval.

Median = 50 + 11/19 x 10 = 50 + 5.79

= 55.79

The median weight is 55.8 to one decimal place. Half of the pigs weighed less than 55.8 kg and half weighed 55.8 kg or more

NB

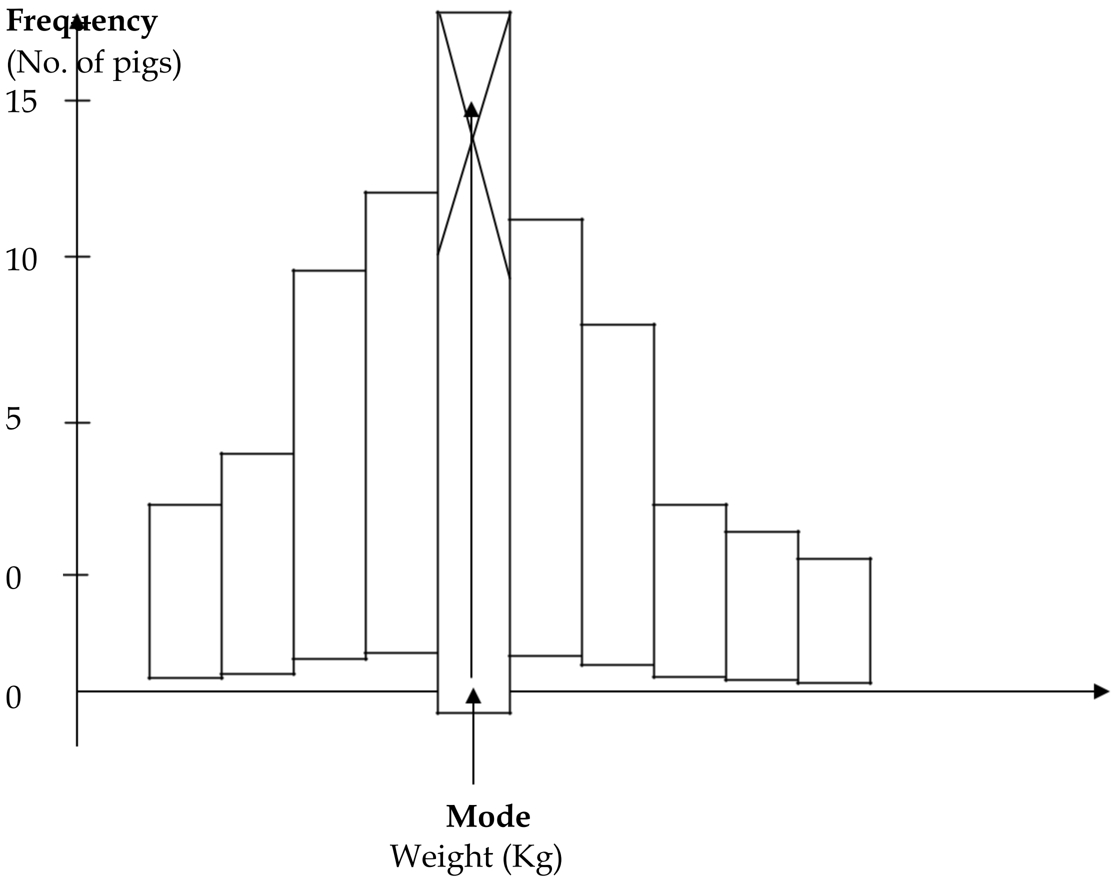
It is common practice when dealing with grouped data to calculate the median as the value of the n /2 item in the distribution , in this case the weight of the 37 ½ pig , although strictly speaking it should be $\frac{n \;\; 1}{2}^{th}$ *Item*

It is obvious that the median is affected by the total number of data but is independent of extreme values. However if the data is ungrouped and numerous, finding the median is tedious. Note that median may be applied in qualitative data if they can be ranked.

3.4 The Mode

Mode is the value of the variable with the highest frequency. It is at the highest peak of the frequency curve. The mode is easy to find in a discrete frequency distribution such as the numbers of visits made by 100 mothers to the local dispensary, where the mode is 7. Twenty five (25) mothers made 7 visits to the local dispensary whereas fewer mothers made any other number of visits (see table given earlier).

For a grouped frequency distribution with a continuous variable, such as weights of pigs, (see tables given earlier) the mode can be estimated using a histogram – see below. The estimate of the mode from the graph is less than 55 kg.



A histogram showing the weight of 75 pigs

Finding the Mode in a Grouped Frequency Distribution with a Continuous Variable

Note that the graphical procedures can be replaced by the following formula:-

$$L + \frac{fz - f_1}{(fz - f_1) + (fz - F_h)} \, i$$

Where:

L = lower class boundary of the modal class, (i.e. 50)

i = Class interval (i.e. 10)

fz = frequency of the modal class

F_1 = Frequency in the adjacent lower-class

F_h= Frequency in the adjacent higher class

Substituting in the formula, we have the following

$$50 \quad \frac{19 \quad 11}{19 \quad 11 \; 19 \quad 10} \quad 10$$

$$50 \; 17\frac{8}{} \; 10$$

$$50 \; 4.7$$

$$54.7(2d.p)$$

This is the same mode given in the graph. Advantage of the mode over the arithmetic mean: its value actually can occur e.g. the modal number of visits made by 100 mothers to the local dispensary were 7 whereas the arithmetic mean was 7.1. The mode is not affected by a few extreme values. The mode is easy to find for discrete variables since no calculation is required, but it can only be estimated for a continuous distribution. There can be more than one mode, e.g. the set numbers: 5,6,6,7,7,8,9,9,9,11,12 there are two modes: 6 and 9. Distributions with two modes are referred to as bimodal and their histograms have two distinct peaks. Not every set of numbers has a mode, e.g. in the set 5, 6, 7,8,9,11,12 each number occurs once only, so the mode does not exist. The mode is not used much in statistical work.

3.5 Weighted Mean

The weighted mean is a mean where there is some variation in the relative contribution of individual data values to the mean. Each data value (Xi) has a weight assigned to it (Wi). Data values with larger weights contribute more to the weighted mean and data values with smaller weights contribute less to the weighted mean. The formula is

$$\bar{X}_w = \frac{\sum W_i X_i}{\sum W_i}$$

There are several reasons why you might want to use a weighted mean.

1. Each individual data value might actually represent a value that is used by multiple people in your sample. The weight, then, is the number of people associated with that particular value.
2. Your sample might deliberately over represent or under represent certain segments of the population. To restore balance, you would place less weight on the over represented segments of the population and greater weight on the under represented segments of the population.
3. Some values in your data sample might be known to be more variable (less precise) than other values. You would place greater weight on those data values known to have greater precision.

Example

Joan gets quiz grades of 79, 82, and 69. She gets a 65 on her final exam. Find the weighted mean if the quizzes each count for 10% and the final exam counts for 70% of the final grade.

Solution

The information can be organized as follows:

X_i	W_i	$W_i X_i$
79	10	790
82	10	820
69	10	690
65	70	4550
	100	6850

[redacted]

$\bar{X}_W = \frac{6850}{100}$

68.5%

3.6 Geometric mean

The geometric mean is an average calculated by multiplying a set of numbers and taking the nth root, where n is the number of numbers.

A common example when the geometric mean is use is when averaging growth rates.

The formula for the geometric mean:-

$$G.M = \sqrt[n]{[x_1][x_2][x_3]...[x_n]}$$

Where n is the number of observations made of the variable x and X_1, X_2 …, X_n are the values of these observations

For example, the geometric mean of numbers: 3, 25 and 45

There are three observation, thus n = 3

$$G.M = \sqrt[3]{[3][25][45]}$$

$$= 15$$

The geometric mean cannot be calculated if we have negative or zero observations. The geometric mean of a set of readings is always less than the arithmetic mean (unless all readings are identical) and is less influenced by very large values / items.

Take the arithmetic mean of the following salaries: - in thousands of shillings per month

6, 8, 10, 10,10,12,16

Arithmetic mean = Kshs 10, 286 per month to the nearest shilling. The geometric men of the salaries are:-

$$G.M = \sqrt[7]{[6][8][10][10][10][12][16]}$$

7 [redacted]

$$= 9.884 \text{ (to 3 decimal places)}$$

The geometric mean of the salaries is Kshs.9884 per

month (To the nearest shilling)

Given the following salaries (i.e. in thousands of Kshs) in accompany per annum (p.a):-

6, 8, 10, 10,10,12,48

The arithmetic mean salary is Kshs 14,857 to the nearest shilling. The geometric mean is:-

$G.M \quad \sqrt[7]{[6][8][10][10][10][12][48}$

7[redacted]0

11.564

11.564 to three decimal places

The geometric mean salary is Kshs.11.564 per annum to the nearest shilling. The geometric mean is useful when only a few items in a distribution are changing: it's in the circumstances more stable than the arithmetic mean. It is useful in the calculation of share indices and also in such calculations where data grows in geometric progression i.e. the population of a country

Given population in a city 300,000 in 1980 and 400,000 in 1990, if we wanted to find out an estimate of the arithmetic mean of the population in 1985.

$\frac{300,000 + 400,000}{2} = \frac{700,000}{2}$

= 350,00

Here, we are making an assumption the population grows by the same number each year which is not correct. The same thing applies to money assuming its growing in a compound rate. The geometric mean for 1985 would be:-

= 2√ 300,000 x 400,000

= 371,080

3.7 Harmonic mean

Harmonic mean is another measure of central tendency and also based on mathematic footing like arithmetic mean and geometric mean. Like arithmetic mean and geometric mean, harmonic mean is also useful for quantitative data. Harmonic mean is defined in

following terms:
Harmonic mean is quotient of "number of the given values" and "sum of the reciprocals of the given values".

Harmonic mean in mathematical terms is defined as follows:

For Ungrouped Data	For Grouped Data
$H.M \quad \bar{X} \quad \frac{n}{\frac{1}{x}}$	$HM \quad \bar{X} \quad \frac{f}{\frac{f}{x}}$

<u>Example</u>:

Calculate the harmonic mean of the numbers: 13.5, 14.5, 14.8, 15.2 and 16.1

<u>Solution</u>:

The harmonic mean is calculated as below:

x	$\frac{1}{x}$
13.2	0.0758
14.2	0.0704
14.8	0.0676
15.2	0.0658
16.1	0.0621
Total	$\frac{1}{x}$ 0.3417

$$H.M = \bar{X} = \frac{n}{\sum \frac{1}{x}}$$

$$H.M = \bar{X} = \frac{5}{0.3417} = 14.63$$

Example:

Given the following frequency distribution of first year students of a particular college. Calculate the Harmonic Mean.

Age (Years)	13	14	15	16	17
Number of Students	2	5	13	7	3

Solution:

The given distribution belongs to a grouped data and the variable involved is ages of first year students. While the number of students Represent frequencies.

Ages (Years) x	**Number of Students** f	$\frac{1}{x}$
13	2	0.1538
14	5	0.3571
15	13	0.8667
16	7	0.4375
17	3	0.1765
Total	$\sum f = 30$	$\sum \frac{1}{x} = 1.9916$

Now we will find the Harmonic Mean as

$$HM = \overline{X} = \frac{\sum f}{\sum \frac{f}{x}} = \frac{30}{1.9916} = 15.0631 \approx 15$$

Example:

Calculate the harmonic mean for the given below:

Marks	30-39	40-49	50-59	60-69	70-79	80-89	90-99
f	2	3	11	20	32	25	7

Solution:

The necessary calculations are given below:

Marks	X	F	$\frac{f}{x}$
30-39	34.5	2	0.0580
40-49	44.5	3	0.0674
50-59	54.5	11	0.2018
60-69	64.5	20	0.3101
70-79	74.5	32	0.4295
80-89	84.5	25	0.2959
90-99	94.5	7	0.0741
Total		$\sum f = 100$	$\sum \frac{f}{x} = 1.4368$

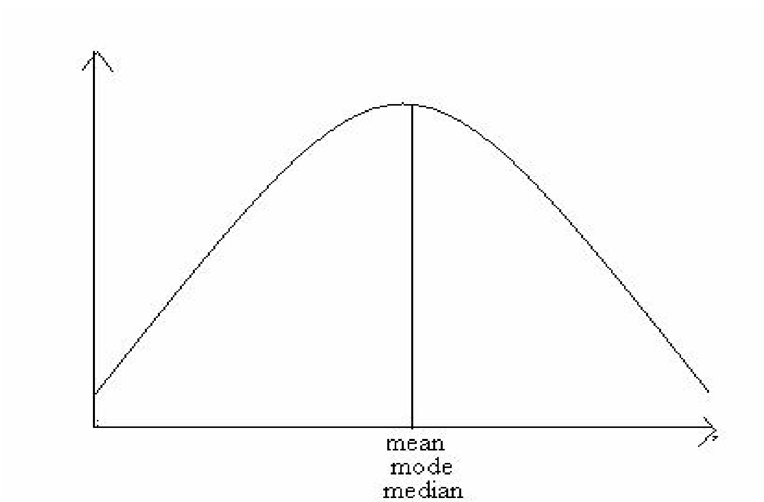
Now we will find the Harmonic Mean as

$$\overline{x} = \frac{\sum f}{\sum \frac{f}{x}} = \frac{100}{1.4368} = 69.60$$

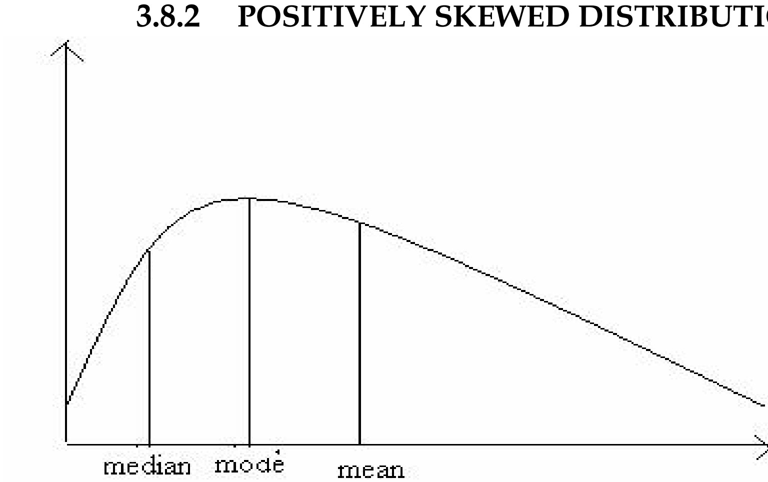
3.8 Relationship Between The Arithmetic Mean, The Median And The Mode

In a symmetrical distribution or normal distribution, that's peaked in the centre, the arithmetic mean = median = mode. Other features of a normal distribution are: - It is bell shaped and is divided into equal parts by the mean, mode and median

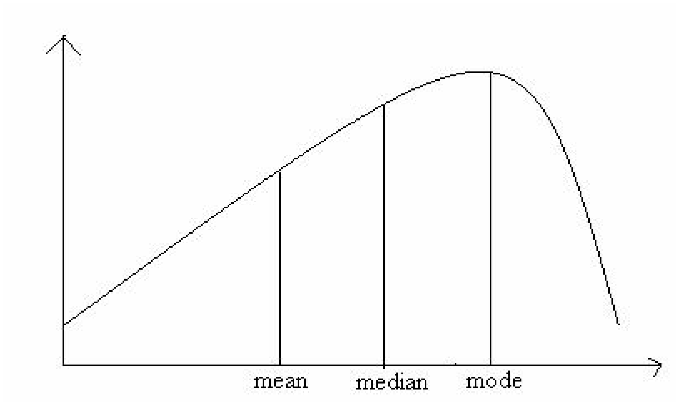
3.8.1 NORMAL DISTRIBUTION



3.8.2 POSITIVELY SKEWED DISTRIBUTION



3.8.3 NEGATIVELY SKEWED DISTRIBUTION



NB:

If we have a positively skewed distribution, the arithmetic mean is not at the centre. The mean dragged to the right of centre by few high values of the variable e.g., an arithmetic mean of salaries for comparison purposes, majority (higher frequency) will be earnings less than average. The median salary would be more typical.

In a negatively skewed distribution, the mean is reduced by the few small values of the variable frequency and hence will be left of the centre. The comparatively few old people features – of – a population in a developing country like Kenya make the mean of the population appear low, since in any case the majority in the population is young people. Again the median would be more representative.

In a moderately skewed distribution the following relationship holds approximately:-

(ARITHMETIC MEAN MINUS MODE = 3 (ARITHMETIC MEAN MINUS MEDIAN) OR

(ARITHMETIC MEAN – MODE = 3 (ARITHMETIC MEAN - MEDIAN)

This formula can be used for estimating the mode from the values of the mean and the median

3.9 Revision Exercise

1. For the set of {8, 4, 2, 10, 2, 5, 9, 12, 2, 6}

(a) calculate the mean;

(b) find the mode;

(c) find the median.

2. The table below shows the relative frequencies of the ages of the students at *Lisuvila High School.*

Age (in years)	Relative frequency
13	0.11
14	0.30
15	0.23
16	0.21
17	0.15
Total	1

(a) If a student is randomly selected from this school, find the probability that

(i) the student is 15 years old;

(ii) the student is 16 years of age or older.

There are 1200 students at *Lisuvila High School.*

3 The average tuition price of two colleges in Georgia is \$7,700. In New York, the average tuition price of three colleges is \$9,400. What is the weighted mean tuition price for all colleges together?

4. "Elaine gets quiz grades of 79, 82, and 69. She gets a 65 on her final exam. Find the weighted mean if the quizzes each count for 10% and the final exam counts for 70% of the final grade.

5. Average cost of 5 pencils and 4 rubbers is Rs. 36. The average cost of 7 apples and 8 mangoes is Rs. 48. What is the total cost of 24 apples and 24 mangoes?

6. What are the special uses of G.M. and H.M. ?

Calculate G.M. and H.M. for the following data :

Value :	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
Frequency :	8	12	20	6	4

(Ans.: 49)

What are the special uses of G.M. and H.M. ?

7. A person walks 8 km. at 4 km. an hour, 6 km. at 3 km. an hour and 2 kms. at 4 km. an hour. Find his average speed per hour

8. Draw a histogram for the following data and locate the mode :

Marks :	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
Frequency:	4	10	16	22	20	18	8	2

(Ans.: Mode = 36)

(a) Write down the value of w.

(b) Draw and label the **Cumulative Frequency** graph for this data

9. (a) A man travels first 900 kms. of his journey by train at an average speed of 80 kms. per hour, next 2,000 kms. by plane at an average speed of 300 kms. per hour and finally, 20 kms. by Taxi at an average speed of 30 kms. per hour. What is his average speed for the entire journey ?

(Ans.: 157.24)

(b) A man travelled by car for 3 days. He covered 480 miles each day, on the first day he drove for 10 hours at 48 miles an hour, on the second day he drove for 12 hours at 40 miles an hour and on the last day he drove 15 hours at 32 miles per hour. What was his average speed ?

(Ans.: 38.92 m.p.h.)

10. Compared to the previous year the overhead expenses went up by 32% in 1961 ; they increased by 44% in the next year and 50% in the following year. Calculate the average rate of increase in overhead expenses over the three years. Explain clearly the reason for the choice of average

(Ans.: 40.5%)

11. Calculate G.M. and H.M. for the following data :

Value :	0—10	10—20	20—30	30—40	40—50
Frequency :	8	12	20	6	4

(Ans.: 49)

12. The following table gives marks obtained by 70 students in Mathematics. Calculate the simple series: —

Marks more than	70	60	50	40	30	20
No. of students	7	18	40	40	63	70

(Ans.: 46.39)

13. Calculate distribution: Harmonic mean for the following frequency

Class	Frequency	Class	Frequency
0 – 10	5	40 – 50	7
10 – 20	8	50-60	6
20 – 30	10	60 – 70	3
30 – 40	12		

(Ans.: H.M.= 19.71)

14. From the results of two colleges A and B, state which of them is better.

Name of the	College A		College B	
examination	Appeared	Passed	Appeared	Passed
M.A.	30	25	100	80
M. Com.	50	45	120	95
B.A.	200	150	100	70
B. Com.	120	75	80	50
Total	400	295	400	295

Suggested References

i. Quantitative Techniques Simplifiedby N.A Saleemi

ii. Elementary Statistics by J.H Van Doorne

iii. Probability and Statistics by R.S Pilaai

TOPIC 4

MEASURES OF DISPERSION AND MEASURES OF KURTOSIS

OBJECTIVES

By then end of the topic, the learner should be able to:

i) Calculate measures of dispersion such as

Range

Quartiles

Deciles and Percentiles

Mean deviation

Variance and Standard deviation and their notations

ii) State the properties of the standard deviation

iii) Use the ogive to read the quartiles and percentiles

iv) Define kurtosis and classify curves as mesokurtic, leptokurtic or platykurtic.

4.1 Introduction

It is important to note that a measure of tendency alone does not give sufficient information about a population. Given that heights of men in Nairobi and Mombasa are 180cm, we may rush to conclude that the heights of the men are the same in the two cities because of the two equal means. We need, however, to have more information about the populations of men in Nairobi and Mombasa i.e. range of heights, how many short and tall men are there in the two cities, in brief: we need to know the measures of dispersion or variation to give a better answer!.

4.2 Range

The range is simply a measure of dispersion i.e. between the highest or biggest value of a variable and the lowest or smallest one respectively.

When we are dealing with a continuous variable the range can only be found accurately from the original data. Its value can only be estimated from a grouped frequency distribution because of its simplicity, it's used in quality control work.

4.3 The Quartile Deviation Or Semi-Interquartile Range

The quartile deviation of a distribution is the middle 50% of the distribution -in this case any extreme values of the distribution are left out. Quartiles are values of the variable that make up 25%, 50% or 75% of the population all the way through the distribution. There are three (3) quartiles, first as Q_1, second as Q_2 and the third is Q_3. The second quartile we've discussed: is the median, i.e., the value of the variable that belongs to the item half-way through the distribution. Given the following figures:-

WEIGHTS OF 75 PIGS

Weight (kg) (variable)	No. of pigs (frequency)	"Less than" cumulative frequency
Under 20	1	1
20 and under 30	7	8
30 and under 40	8	16
40 and under 50	11	27
50 and under 60	19	46
60 and under 70	10	56
70 and under 80	7	63
80 and under 90	5	68

90 and under 100	4	72
Over 100	3	75

Quarter $\frac{1}{4}$ through the distribution 75

$= \frac{n \; 1}{4}^{th}$ where n is the number of pigs

$\frac{75 \quad 1}{4} \quad \frac{76}{4}$

$= 19$

The 19th pig belongs to: "40 kg under 50 kg" class. Before this class 16 pigs have been accounted for, thus, less (19 -16 =3 pigs not accounted for out of 11 in the "40 and under 50kg"class before we reach the first quartile).

$Q_1 \quad 40 \quad \frac{3}{11} X10$

$= 40 \;\; 2.73$

$= 42.73$

We conclude one quarter of the pigs weighed less than 42.73kg.

Similarly the third quartile is the $\frac{3(n \;\; 1)}{4}^{th}$ i.e. the weight of the 57th pig. This is the 1st item in the "70 and under 80kg" class.ie

$$Q_3 = 70 + \frac{57 - 56}{10} X 10$$

= 70 + 1.43

= 71.43

Thus, three quarters of the pigs weighed less than 71.43 kg so that one quarter weighed 71.43 kg or more.

The interquartile range is the difference between Q_3 and Q_1

$Q_3 - Q_1 = 71.43 - 42.73$

= 28.70kg

$\frac{3(n+1)}{4}^{th}$

This tells us that the middle 50% of the data spans 28.70kg. The quartile deviation is the semi-inter quartile range:

$$\text{Quartile deviation} = \frac{Q_3 - Q_1}{2} = \frac{28.70}{2}$$

$= 14.35kg$

NB: We can use the ogive to find approximate values of the quartile – see graph below. Quartile deviation gives us an idea of the variability in the values of a population. We're, therefore, able to make comparisons and say how variable the values in one distribution are compared to another. Since the quartile deviation looks at the middle 50% of the data, its value is unaffected by extreme values and can be used when we're open-ended classes at the edges of a distribution. It is used as a measure of dispersion when the media is chosen as the approximate measure of location, for instance, when dealing with income statistics.

NB: take the following heights;

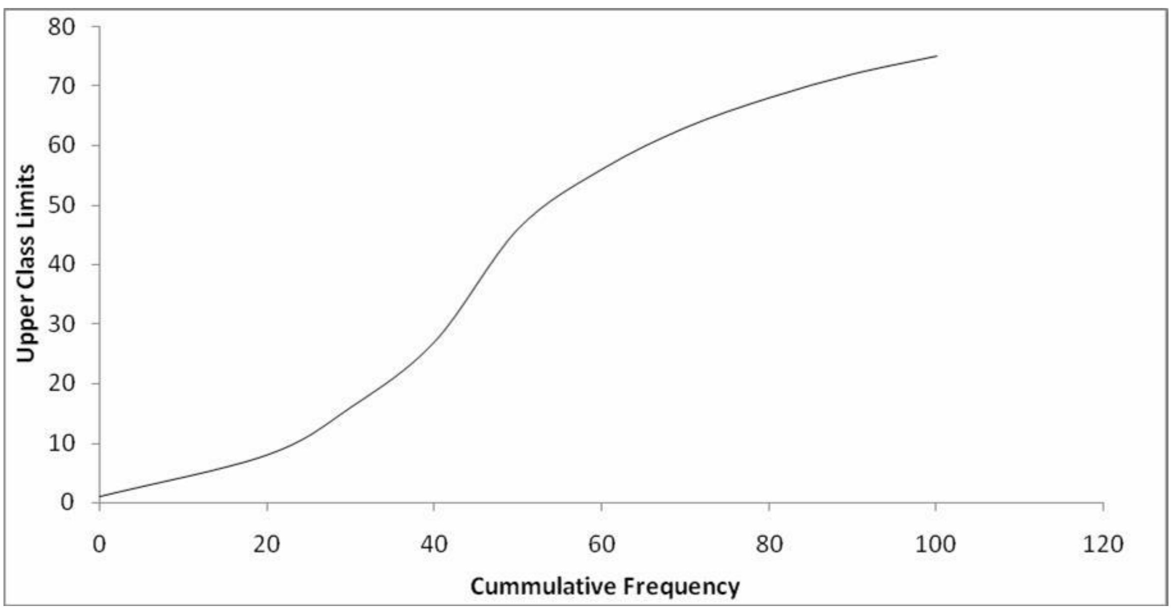
Median pig weight – 55.8 kg

quartile (¼) weight – 42.7kg

three quartiles (¾) weight – 71.4kg

The median weight is not the arithmetic mean of the quartile weights. Only when the distribution is symmetrical is the median is exactly half –way the two quartile values.

OGIVE OF WEIGHTS OF 75 PIGS



The quartiles and other measures can also be read from the ogive.

4.4 Deciles and Percentiles

Deciles and percentiles are used in the sub division of income and employment statistics in similar fashion to quartiles. The deciles are the values which divide the total frequency into tenths and the percentiles are the values which divide the total frequency into hundredths. Deciles and percentiles are useful when we are dealing with a very large number of observations, preferably thousands.

Deciles are denoted by D_1, D_2, D_3 …………D_9. The third decile D_3 e.g., is the value which 30% of the data lies. Method of calculation follows the same pattern as calculation of the median and quartiles e.g. 30% of 75 pigs, we get D_3 as the weight of the 22.5 pig or rather the 22_{nd} pig or 23_{rd} pig, the fraction is of no consequence.

NB: 22.5 pigs is in the "40 and 50kg" class (-see figures on page ??):

$$D_3 \quad 40 \quad \frac{22.5 \quad 16}{11} X10$$

$$40 \quad \frac{6.5}{11} X10$$

$$= 40 \quad 5.9$$

$$= 45.9 \text{ kg}$$

Therefore, 30% of the pigs weighed less than 45.9kg.

Percentiles are denoted by P_1, P_2, …, P_{99} and, for example, P_6 is the value below which 6% of the data lies and P_{66} is the value below which 66% of the data lies. U.K. income statistics use percentiles: P_1, P_2, P_5, P_{95} and P_{99}. These values give useful information about the lowest and highest incomes.

4.5 Mean Absolute Deviation

The Mean Absolute Deviation or simply Mean deviation measures how far on average the readings are from the arithmetic mean. The median is very occasionally used instead of the arithmetic mean. If the data have a small spread about the mean, the mean deviation has a lower value than for data which show large variations about the mean.

To calculate the mean deviation ,

Find the arithmetic mean, of the data

Find the deviation of each reading from i.e., work out the difference between each reading and

Find the arithmetic mean of the deviations, ignoring their signs, .e. the visits made to the local dispensary by 10 mothers is as follows:

8, 6, 5, 5, 7, 4, 5, 9, 7, 4 = 60 visits

Arithmetic mean = $10 \frac{60}{} \; 6$

CALCULATION OF THE MEAN DEVIATION FOR UNGROUPED DATA

NO. OF VISITS	MEAN $\bar{X}$	DEVIATIONS X - $\bar{X}$	DEVIATIONS IGNORING
8	6	2	2
6	6	0	0
5	6	-1	1
5	6	-1	1
7	6	1	1
4	6	-2	2
5	6	-1	1

9	6	3	3
7	6	1	1
4	6	-2	2
Totals 60		0	14

If we ignore the signs above, the sum of the deviations is 14

Mean Absolute deviation: $\frac{14}{10}$ 1.4

For ungrouped data the formula for the mean deviation is:-

Mean Absolute deviation (M.A.D) $= \frac{\sum |x - \bar{x}|}{\sum f}$

Where $|x - \bar{x}|$ stands for the absolute deviation, that's the deviation ignoring the sign.

In calculating the mean deviation of a grouped frequency distribution the formula is modified to include the class frequencies:-

$M.A.D = \frac{\sum f|x - \bar{x}|}{\sum f}$

Where: X in this formula is the class mid point. Given the earlier example of a grouped frequency distribution of the weights of 75 pigs, assumption is made that the open-ended classes are of the same length as adjacent classes.

Calculation of the Mean Absolute Deviation

WEIGHT (kg) (variable)	Mid point	No. of pigs (frequency) f	fx	$\|x - \bar{x}\|$	$f\|x - \bar{x}\|$

0-20	15	1	15	42.4	42.4
20–30	25	7	175	32.4	226.8
30–40	35	8	280	22.4	179.2
40–50	45	11	485	12.4	136.4
50–60	55	19	1045	2.4	45.6
60–70	65	10	650	7.6	76.0
70–80	75	7	525	17.6	123.2
80–90	85	5	425	27.6	138.0
90 – 100	95	4	380	37.6	150.4
Over 100	105	3	315	47.6	142.8
Totals		75	4305		1260.8

$$M.A.D \quad \frac{fx \quad | \quad x|}{f}$$

$\frac{4305}{75}$

$57.4kg$

$$M.A.D \quad fx.x \quad | \quad -|$$

$$\frac{1260.8}{75}$$

$16.8kg$ to one decimal place)

The mean absolute deviation can be used to compare the variation of different distributions buts its not useful in statistical work because negative signs in the deviations are ignored when its calculated. Mean absolute deviation suffers the same problem as the arithmetic mean, i.e. Its value is distorted by extreme values of the variable.

4.6 Population and sample

In statistics, a **sample** is a subset of a population. Typically, the population is very large, making a census or a complete enumeration of all the values in the population impractical or impossible. The sample represents a subset of manageable size. Samples are collected and statistics are calculated from the samples so that one can make inferences or extrapolations from the sample to the population. This process of collecting information from a sample is referred to as sampling.

Care must be taken when dealing with the signs involving population and sample parameters. The table below shows the different notations.

	Population	Sample
Mean	$\overline{X}$	$\bar{x}$
Variance	^{2}	s^2
Standard Deviation		s

While analyzing variance and standard deviation, we shall assume that all the data is taken from a sample.

4.7 Variance and standard deviation

Unlike range and quartiles, the variance combines all the values in a data set to produce a measure of spread. The variance (symbolized by s^2) and standard deviation (the square root of the variance, symbolized by s) are the most commonly used measures of spread.

We know that variance is a measure of how spread out a data set is. It is calculated as the average squared deviation of each number from the mean of a data set. For example, for the numbers 1, 2, and 3 the mean is 2 and the variance is 0.667.

$[(1 - 2)^2 + (2 - 2)^2 + (3 - 2)^2]$ &divide 3 = 0.667

[Squaring deviation from the mean and dividing by the number of observations = variance]

Variance (s^2) = average squared deviation of values from mean

Calculating variance involves squaring deviations, so it does not have the same unit of measurement as the original observations. For example, lengths measured in metres (m) have a variance measured in metres squared (m^2).

Taking the square root of the variance gives us the units used in the original scale and this is the standard deviation.

Standard deviation (s) = square root of the variance

Standard deviation is the measure of spread most commonly used in statistical practice when the mean is used to calculate central tendency. Thus, it measures spread around the mean. Because of its close links with the mean, standard deviation can be greatly affected if the mean gives a poor measure of central tendency.

Standard deviation is also influenced by outliers one value could contribute largely to the results of the standard deviation. In that sense, the standard deviation is a good indicator of the presence of outliers. This makes standard deviation a very useful measure of spread for symmetrical distributions with no outliers.

Standard deviation is also useful when comparing the spread of two separate data sets that have approximately the same mean. The data set with the smaller standard deviati n has a narrower spread of measurements around the mean and therefore usually has comparatively fewer high or low values. An item selected at random from a data set whose standard deviation is low has a better chance of being close to the mean than an item from a data set whose standard deviation is higher.

Generally, the more widely spread the values are, the larger the standard deviation is. For example, imagine that we have to separate two different sets of exam results from a class of 30 students the first exam has marks ranging from 31% to 98%, the other ranges from 82% to 93%. Given these ranges, the standard deviation would be larger for the results of the first exam.

Standard deviation might be difficult to interpret in terms of how big it has to be in order to consider the data widely spread. The size of the mean value of the data set depends on the size of the standard deviation. When you are measuring something that is in the millions, having measures that are "close" to the mean value does not have the same meaning as when you are measuring the weight of two individuals. For example, a measure of two large companies with a difference of $10,000 in annual revenues is considered pretty close, while the measure of two individuals with a weight difference of 30 kilograms is considered far apart. This is why, in most situations, it is useful to assess the size of the standard deviation relative to the mean of the data set.

Although standard deviation is less susceptible to extreme values than the range, standard deviation is still more sensitive than the semi-quartile range. If the possibility of high values (outliers) presents itself, then the standard deviation should be supplemented by the semi-quartile range.

4.7.1 Properties of standard deviation

When using standard deviation keep in mind the following properties.

Standard deviation is only used to measure spread or dispersion around the mean of a data set.

Standard deviation is never negative.

Standard deviation is sensitive to outliers. A single outlier can raise the standard deviation and in turn, distort the picture of spread.

For data with approximately the same mean, the greater the spread, the greater the standard deviation.

If all values of a data set are the same, the standard deviation is zero (because each value is equal to the mean).

When analyzing normally distributed data, standard deviation can be used in conjunction with the mean in order to calculate data intervals.

If $\bar{x}$= mean, **s** = standard deviation and **x** = a value in the data set, then

about 68% of the data lie in the interval: $\bar{x} - s < x < \bar{x} + s$.

about 95% of the data lie in the interval: $\bar{x} - 2s < x < \bar{x} + 2s$.

about 99% of the data lie in the interval: $\bar{x} - 3s < x < \bar{x} + 3s$.

4.7.2 The Variance of Discrete variables

The variance for a discrete variable made up of **n** observations is defined as:

$$s^2 = \frac{\Sigma (x - \bar{x})^2}{n}$$

The standard deviation for a discrete variable made up of **n** observations is the positive square root of the variance and is defined as:

$$s = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n}}$$

Use this step-by-step approach to find the standard deviation for a discrete variable.

1. Calculate the mean.
2. Subtract the mean from each observation.
3. Square each of the resulting observations.
4. Add these squared results together.
5. Divide this total by the number of observations (variance, $\mathbf{s^2}$).
6. Use the positive square root (standard deviation, **s**).

Example 1 – Standard deviation

A hen lays eight eggs. Each egg was weighed and recorded as follows:

60 g, 56 g, 61 g, 68 g, 51 g, 53 g, 69 g, 54 g.

a) First, calculate the mean:

$$\bar{x} = \frac{\Sigma x}{n} = \frac{472}{8} = 59$$

b) Now, find the standard deviation.

Table 1. Weight of eggs, in grams

Weight (x)	$(x - \bar{x})$	$(x - \bar{x})^2$
60	1	1
56	-3	9
61	2	4
68	9	81
51	-8	64
53	-6	36
69	10	100

54	-5	25
472		**320**

Using the information from the above table, we can see that

$$\Sigma (x - \bar{x})^2 = 320$$

In order to calculate the standard deviation, we must use the following formula:

$$s = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n}}$$

$$= \sqrt{\frac{320}{8}}$$

$$= 6.32 \text{ grams}$$

4.7.3 The Variance of Grouped variables

Example 2 – Standard deviation calculated using a frequency table

Thirty farmers were asked how many farm workers they hire during a typical harvest season. Their responses were:

4,5,6,5,3,2,8,0,4,6,7,8,4,5,7,9,8,6,7,5,5,4,2,1,9,3,3,4,6,4

Thirty farmers were asked how many farm workers they hire during a typical harvest season. Their responses were:

Workers (x)	**Tally**	**Frequency (f)**	**(xf)**	**$(x - \bar{x})$**	**$(x - \bar{x})^2$**	**$(x - \bar{x})^2 f$**
0	I	1	0	-5	25	25
1	I	1	1	-4	16	16
2	II	2	4	-3	9	18
3	III	3	9	-2	4	12
4	[illegible]	6	24	-1	1	6

5	▬	5	25	0	0	0
6	IIII	4	24	1	1	4
7	III	3	21	2	4	12
8	III	3	24	3	9	27
9	II	2	18	4	16	32
		30			**150**	**152**

To calculate the mean:

$$\bar{x} = \frac{\sum x\,f}{\sum f} = \frac{150}{30} = 5$$

To calculate the standard deviation:

$$S = \sqrt{\frac{\sum (x - \bar{x})^2 f}{n}} = \sqrt{\frac{152}{30}} = 2.25$$

Example 3 – Standard deviation using grouped variables (continuous or discrete)

information, calculate the mean and standard deviation of hours spent watching television by the 220 students.

Table 3. Number of hours per week spent watching television	
Hours	**Number of students**
10–14	2

15–19	12
20–24	23
25–29	60
30–34	77
35–39	38
40–44	8

a. First, using the number of students as the frequency, find the midpoint of time intervals.
b. Now calculate the mean using the midpoint (**x**) and the frequency (**f**).

Note: In this example, you are using a continuous variable that has been rounded to the nearest integer. The group of 10&ndash14 is actually 9.5 to 14.499 (as the 9.5 would be rounded up to 10 and the 14.499 would be rounded down to 14). The interval has a length of 5 but the midpoint is 12 (9.5 + 2.5 = 12).

$$\bar{x} = \frac{\Sigma x f}{\Sigma f} = \frac{6{,}560}{220} = 29.82$$

6,560=(2X12+12X17+23X22+60X27+77X32+38X37+8X42)

Then, calculate the numbers for the **xf**, **$(x - \bar{x})$**, **$(x - \bar{x})^2$** and **$(x - \bar{x})^2 f$** formulas.

Add them to the frequency table below..

Table 4. Number of hours spent watching television

Hours	Midpoint (x)	Frequency (f)	Xf	$(x - \bar{x})$	$(x - \bar{x})^2$	$(x - \bar{x})^2 f$
10–14	12	2	24	-17.82	317.6	635.2

15–19	17	12	204	-12.82	164.4	1,972.8
20–24	22	23	506	-7.82	61.2	1,407.6
25–29	27	60	1,620	-2.82	8.0	480.0
30–34	32	77	2,464	2.18	4.8	369.6
35–39	37	38	1,406	7.18	51.6	1,960.8
40–44	42	8	336	12.18	148.4	1,187.2
		220	**6,560**			**8,013.2**

Example 4

Use the information found in the table above to find the standard deviation.

$$S = \sqrt{\frac{\sum (x - \bar{x})^2 f}{n}}$$

$$= \sqrt{\frac{8{,}013.2}{220}}$$

$$= \sqrt{36.42}$$

$$= 6.03$$

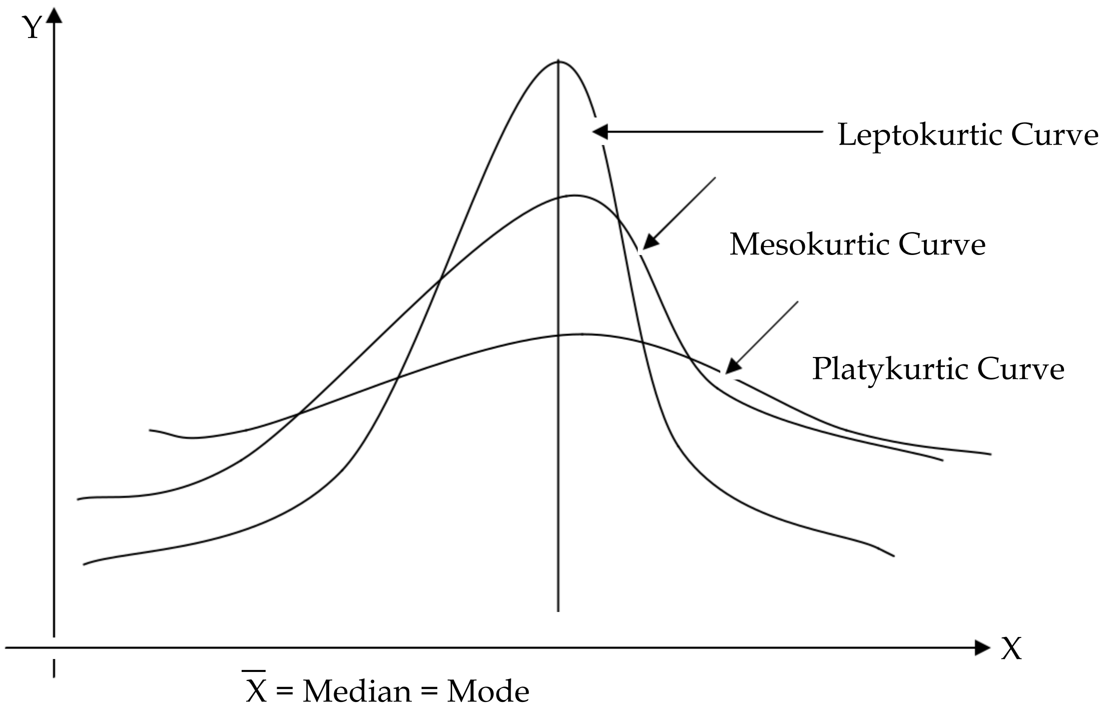
Note: During calculations, when a variable is grouped by class intervals, the midpoint of the interval is used in place of every other value in the interval. Thus, the spread of observations within each interval is ignored. This makes the standard deviation *always* less than the true value. It should, therefore, be regarded as an approximation.

4.8 Kurtosis

Kurtosis refers to the degree of flatness of a frequency curve. It helps to find out how a particular distribution conforms to the normal curve. It shows the extent to which the curve is more peaked or more flat topped than the normal curve. From the stand point of kurtosis, the normal curve is mesokurtic i.e. of "intermediate peakedness" When the curve of distribution is relatively flatter than the normal curve, it is said to have kurtosis. When the curve or polygon is relatively more peaked, it is said to "lack kurtosis" Thus the c ncept of kurtosis helps us in studying the peakedness of distribution.

The following diagram illustrates the scope of three different curves mentioned above: -

NATURE OF KURTOSIS



Where

$B_2 = 3$ Mesokurtic curve

$B_2 < 3$ Platykurtic curve

$B_2 > 3$ Leptokurtic curve

4.9 Revision Exercise

1.The cumulative frequency table below shows the ages of 200 students at a college.

Age	Number of Students	Cumulative Frequency
17	3	3
18	72	75
19	62	137
20	31	m
21	12	180
22	9	189
23	5	194
25	6	n

(a) What are the values of m and n?

(b) How many students are younger than 20?

(c) Find the value in years of the lower quartile.

(d) Calculate the standard deviation

2.The heights of 200 students are recorded in the following table.

Height (h) in cm	Frequency
$140 \leq h < 150$	2
$150 \leq h < 160$	28
$160 \leq h < 170$	63
$170 \leq h < 180$	74
$180 \leq h < 190$	20
$190 \leq h < 200$	11
$200 \leq h < 210$	2

(a) Write down the modal group.

(b) Calculate an estimate of the mean and standard deviation of the heights.

3. The table below shows the number and weight (w) of fish delivered to a local fish market one morning.

weight (kg)	**frequency**	**cumulative frequency**
$0.50 \le w < 0.70$	16	16
$0.70 \le w < 0.90$	37	53
$0.90 \le w < 1.10$	44	c
$1.10 \le w < 1.30$	23	120
$1.30 \le w < 1.50$	10	130

(a) (i) Write down the value of c.

(1)

(ii) On graph paper, draw the *cumulative frequency curve* for this data. Use a scale of 1 cm to represent 0.1 kg on the horizontal axis and 1 cm to represent 10 units on the vertical axis. Label the axes clearly.

(4)

(iii) Use the graph to show that the median weight of the fish is 0.95 kg.

(1)

(b) (i) The zoo buys all fish whose weights are above the 90^{th} percentile. How many fish does the zoo buy?

(2)

(ii) A pet food company buys all the fish in the lowest quartile. What is the maximum weight of a fish bought by the company?

(3)

(c) A restaurant buys all fish whose weights are within 10% of the median weight.

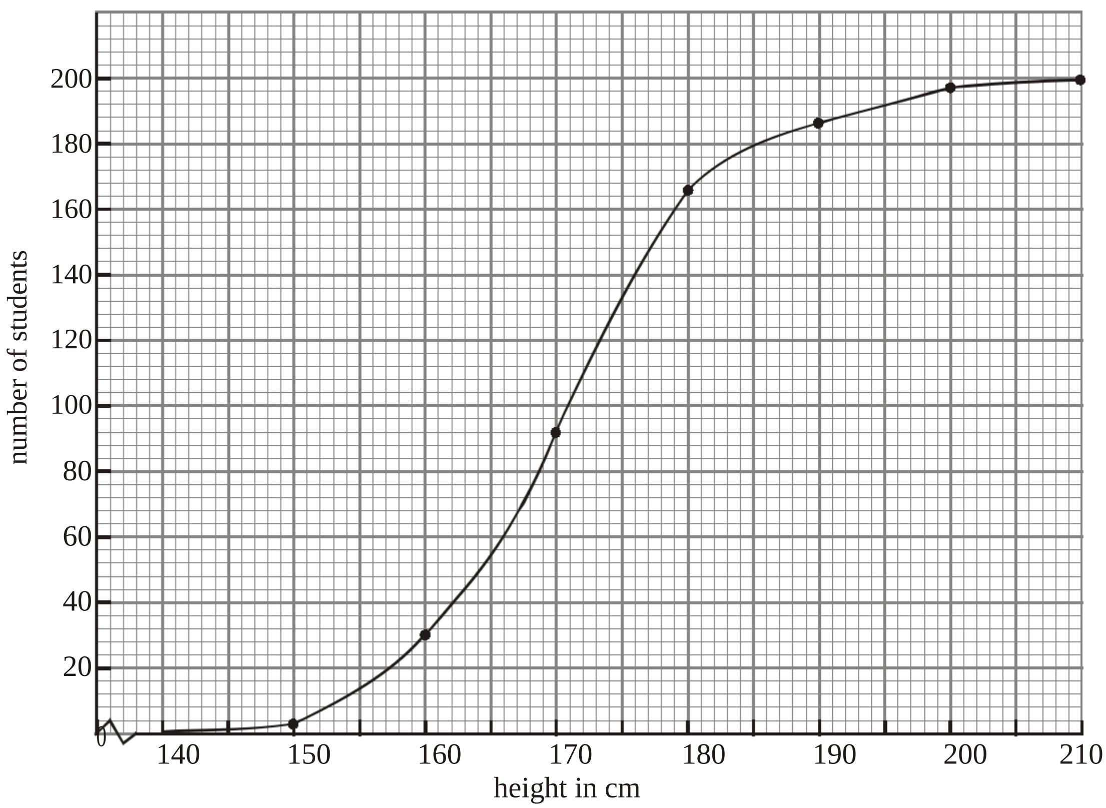
(i) Calculate the minimum and maximum weights for the fish bought by the restaurant.

4 **(2)**

(ii) Use your graph to determine how many fish will be bought by the restaurant.

5 **(3)**

4. The cumulative frequency curve for this data is drawn below.



(c) Write down the median height.

(d) The upper quartile is 177.3 cm. Calculate the interquartile range.

(e) Find the percentage of students with heights less than 165 cm.

Suggested References

i. Quantitative Techniques Simplified by N.A Saleemi

ii. Advanced Mathematics Statistics by SteveDobbs and Jane Miller

iii. Elementary Statistics by J.H Van Doorne

TOPIC 5

Equations and Inequalities

OBJECTIVES

By the end of the topic the learner should be able to

i) Identify different types of variables

ii) Solve linear equations

iii) Solve linear inequalities

iv) Represent linear inequalities graphically.

5.1 Definition of an equation

An equation is a mathematical statement that relates two algebraic expressions involving at least one variable.

It is a statement showing the equality of two expressions.

5.2 Properties of equality

For any real numbers a, b and c

1. **Addition property:** If a=b, then a+ c= b + c

2. **Subtraction property**: **:** If a=b, then a- c= b - c

3. **Multiplication property:** If a=b, then ac = bc

4. **Division property**: **:** If a=b, then a/c = b/c

5. **Substitution property:** If a=b, then either may replace the other in any statement without changing the truth or falsity of the statement.

5.3 Terms used in equations

(a) Unknown
These are the variables whose values are not given. They are found by solving the equation e.g. in the equation 4x + 5y =40 and are unknowns. *x and y* .

(b) Constants

They are fixed figures, which are shown on the left hand side of unknowns or separately. From the equation 4x + 5y =40, 4, 5 and 40 are constants.

(c) Coefficients

They are those figures, which are shown on the left hand side of unknowns. From the equation 4x + 5y =40 , 4 and 5 are coefficients of and respectively.

(d) Index

It is the power of an unknown. It is shown on the right hand side on top of any unknown e.g. in the equation, $4x_3$ +2x + 3 has an index or power 3.

5.4 Variables

5.4.1 Definition of a variable
It is a measurable characteristic that assumes different values among the subjects.

5.4.2 Types of variables
1. Independent variables / Predictor variables

It is a variable that a researcher manipulates in order to determine its effect or influence on another variable. They predict the amount of variation that occurs in another variables.

Types of independent variables

i. **Experimental variables**: They are variables which the researcher has manipulative control over them. Are commonly used in biological and physical sciences e.g. influence of amount of fertilizer on the yield of wheat, influence of alcohol on reaction time.

ii. **Measurement types of independent variables**: Are variables, which have already occurred. They have fixed manipulative and uninfluenceable properties. Most f the variables are either environmental or personal e.g. age, gender, marital status, race, colour, geographical location, nationality, soil type, altitude etc. (e.g. influence of nationality on choice of food).

2. Dependent variables / criterion variables

They attempt to indicate the total influence arising from the effects of the independent variable. It varies as a function of the independent variable e.g. influence of hours studied on performance in a statistical test, influence of distance from the supply center on cost of building materials.

3. Discrete variables

They are variables, which can only assume whole numbers. They are always counted e.g. number of cars, children, trees books etc.

4. Continuous variables

They are variables, which can assume any value in a specific range. They are always measured e.g. height, weight, weight, radius, distance etc.

5. Extraneous variables

They are those variables that affect the outcome of a research study either because the researcher is not aware of their existence or if the researcher is aware, she or he has no control over them.

6. Control variables / concomitant / covariate or blocking variables

They are extraneous variables that are built into the study. They are introduced to increases the validity of the data and to leads to more convincing generalizations.

5.5 Types of equations

There are three major types of equations. They include:-

(a) Linear / simple equations

(b) Quadratic equations

(c) Simultaneous equations

For this topic, we shall only deal with linear equations.

5.6 Linear / simple equations

They are equations with one unknown and the index of the unknown is one. It is of the form ax + b =0. Where *a* and *b* are constants and the solution is obtained by writing ax= -b

And getting a = $\frac{-b}{x}$ provided that a $\neq 0$

Example 1: Solve for x **in the following equation.**

x - 4 = 10
solution

Add 4 to both sides of the equation:

$$x = 14$$

The answer is **x = 14**

Check the solution by substituting **14** in the original equation for x. If the left side of the equation equals the right side of the equation after the substitution, you have found the correct answer.

Example 2: Solve for x **in the following equation.**

$$2x - 4 = 10$$

Add 4 to both sides of the equation:

$$2x = 14$$

Divide both sides by 2:

$$x = 7$$

The answer is x = 7.

Check the solution by substituting 7 in the original equation for x. If the left side of the equation equals the right side of the equation after the substitution, you have found the correct answer.

2(7)-4=14-4=10.

Example 3: Solve for x in the following equation $5x - 6 = 3x - 8$

Subtract 3x from both sides of the equation:

$$2x - 6 = -8$$

Add 6 to both sides of the equation:

$$2x = -2$$

Divide both sides by 2:

$$x = -1$$

The answer is x = - 1

Add 500 to both sides of the equation:

$$510 = 51x$$

Divide both sides by 51:

$$x = \frac{510}{51} = 10$$

The answer is x = 10.

Example 5: Solve for x in the following equation

$$\frac{6x - 7}{4} + \frac{3x - 5}{7} = \frac{5x + 78}{28}$$

Get rid of the denominators by multiplying both sides by 28, the smallest number that 4,7, and 28 will divide into evenly.

Recall the line that separates the numerator and the denominator also functions as a parenthesis. It instructs the reader to treat the numerator as one number and the denominator as one number.

$$28\left(\frac{6x-7}{4}+\frac{3x-5}{7}\right)=28\left(\frac{5x+78}{28}\right)$$

Simplify:

$$28\left(\frac{6x-7}{4}\right)+28\left(\frac{3x-5}{7}\right)=28\left(\frac{5x+78}{28}\right)$$

$$7(6x-7)+4(3x-5)=1(5x+78)$$

$$42x-49+12x-20=5x+78$$

Add the *x* terms and the constants on the left side of the equation.

$$54x-69=5x+78$$

Subtract 5*x* from both sides of the equation:

$$49x-69=78$$

Add 69 to both sides of the equation:

$$49x=147$$

Divide both sides by 49:

$$x=3$$

The answer is *x*=3.

Example6: Solve for *x* in the following equation 2(3*x* - 7) + 4 (3 *x* + 2) = 6 (5 *x* + 9) + 3

Complete multiplication.

$$6x - 14 + 12x + 8 = 30x + 54 + 3$$

Group like terms on each side of the equal sign.

$$18x - 6 = 30x + 57$$

Subtract 18 x from both sides of the equation.

$$-6 = 12x + 57$$

Subtract 57 from both sides of the equation.

$$-63 = 12x$$

Divide both sides of the equation by 12 and simplify.

$$x = -\frac{63}{12} = -\frac{21}{4}$$

$$x = -\frac{21}{4}$$

Example 5: Solve for *x* in the following equation

$$\sqrt{2}x - \sqrt{3} = \sqrt{5}$$

Solution

Add $\sqrt{3}$ to both sides of the equation: $\sqrt{2}x \; \sqrt{3} \; \sqrt{5}$

Divide both sides by $\sqrt{2}$,

$$x = \frac{\sqrt{3} + \sqrt{5}}{\sqrt{2}}$$

5.7 Linear inequalities

Definition:

An inequality is a mathematical statement related to another by the operation They can be solved using the graphical method. ≥ > ≤ or, <

5.7.1 Properties of inequalities

Let a, b and c be real numbers.

1. Transitive Property
   If $a < b$ and $b < c$ then $a < c$

2. Addition Property
   If $a < b$ then $a + c < b + c$

3. Subtraction Property If
   $a < b$ then $a - c < b - c$

4. Multiplication Property
   i. If $a < b$ and c is positive then $c*a < c*b$
   ii. If $a < b$ and c is negative $c*a > c*b$

   Note: If each inequality sign is reversed in the above properties, we obtain similar properties. If the inequality sign < is replaced by <= (less than or equal) or the sign > is replaced by >= (greater than or equal), we also obtain similar properties.

5.7.2 Solving Linear Inequalities

Example 1: Solve the inequality

$6x - 6 > 2x + 2$

Solution

Given

$6x - 6 > 2x + 2$

Add 6 to both sides and simplify (Property 2

above) $6x > 2x + 8$

Subtract 2x to both sides and simplify (Property 3 above)

$4x > 8$

Multiply both sides by 1/4; and simplify (Property 4-i above)

$x > 2$

Conclusion
The solution set consists of all real numbers in the interval (2 , + infinity).

Example 2: Solve the inequality

$2(3x + 2) - 20 > 8(x - 3)$

Solution

Given
$2(3x + 2) -20 > 8(x - 3)$

Multiply factors and group like terms
$6x + 4 -20 > 8x - 24$
$6x - 16 > 8x - 24$

Add 16 to both sides and simplify (Property 2
above) $6x > 8x - 8$

Subtract 8x to both sides and simplify (Property 3 above)
$-2x > -8$

Multiply both sides by -1/2 and **REVERSE** (-1/2 is negative) the inequality sign and simplify (Property 4-ii above)
$x < 4$

Conclusion The solution set consists of all real numbers in the interval (- infinity , 4)

Example 3: Solve the double inequality

$-3 < 4(x + 2) - 3 < 9$

Solution

Given

$-3 < 4(x + 2) - 3 < 9$

Multiply factors and group like terms

$-3 < 4x + 8 - 3 < 9$

$-3 < 4x + 5 < 9$

Subtract 5 to all three terms and simplify $-3 - 5 < 4x + 5 - 5 < 9 - 5$

$-8 < 4x < 4$

Divide all three terms by 4

$-2 < x < 1$

Conclusion

The solution set consists of all real numbers in the interval (- 2 , 1)

5.8 Graphing A Linear Inequality

Step 1: Graph the boundary line

Essentially you graph the boundary line the same as if the problem was a linear equation. Assume that there is an equal sign and use an appropriate method to graph the line. Unless the directions to a problem indicate otherwise, you can use any method to graph it. Two ways of graphing linear inequalities are

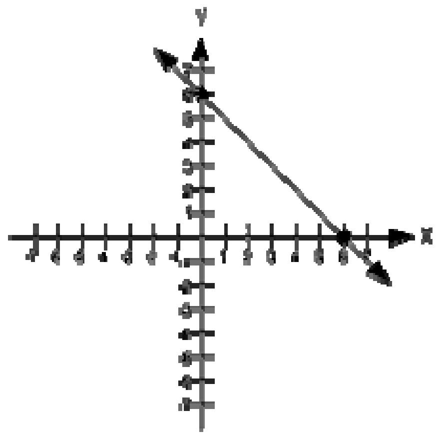
i) plotting any three points or
ii) Using the x and y intercepts.

When you draw the boundary line, you need to have a way to indicate if the line is included or not in the final answer.

For a solid boundary line, $\leq$ or $\geq$ are used in the inequality. Otherwise the dashed line is used.

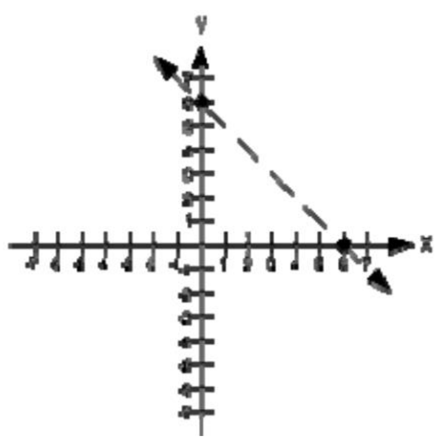
Example 1: Graph the inequality for $x + y \leq 6$

The graph below shows the boundary line for $\mathbf{x + y \leq 6}$**:**
(note that this does not show the inequality part)



If the problem does not include where it is equal, then you will use a dashed boundary line. In other words, if you have < or >, you will have a dashed line for your boundary line.

This graph below shows the boundary line for $x + y < \mathbf{6}$**:**
(note that this does not show the inequality part)



In either case, you still graph the line the same. You just have to decide if you are needing a solid line or a dashed line.

The boundary line separates the rectangular coordinate system into two parts. One of those parts will make the inequality true and be it's solution.

Step 2: Try out in a point that is not on the boundary line

Pick a test point on either side of the boundary line and plug it into the original problem. This will help determine which side of the boundary line is the solution.

Step 3: Shade in the answer to the inequality

If you get a true statement when you plug in the test point in step 2, then you have found a solution. Shade the region that the test point is in.

If you get a false statement when you plug in the test point in step 2, then you don't have a solution. Shade in the region that is on the other side of the test point.

It doesn't matter what you use for the test point as long as it is not on the boundary line. You want to keep it as simple as possible.

Example 2: Graph the inequality

$x + y > 2$.

Step 1: Graph the boundary line.

In this case the boundary line is x + y = 2

Using thee points for x as shown below, we obtain the corresponding values of y

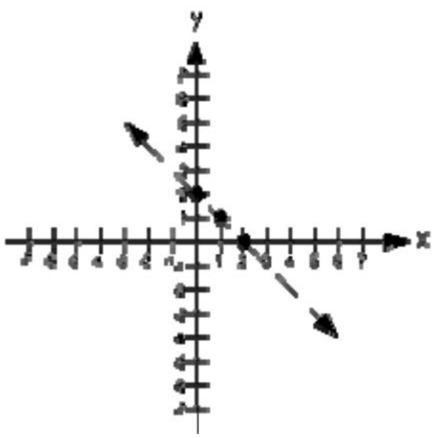
x	*y*	*(x, y)*
2	0	(2, 0)
0	2	(0, 2)

1	1	(1, 1)

Since the original problem has a >, this means it DOES NOT include the boundary line.

So are we going to draw a solid or a dashed line for this problem? It looks like it will have to be a dashed line.

Putting it all together, we get the following boundary line for this problem:



Step 2: Plug in a test point that is not on the boundary line.
Note how the boundary line separates it into two parts.

An easy test point would be (0, 0). Note that it is a point that is not on the boundary line. In fact, it is located below the boundary line.

Let's put (0, 0) into the original problem and see what happens:

$x + y > 2$

$0 + 0 > 2$ *Replacing *x* and *y* with 0

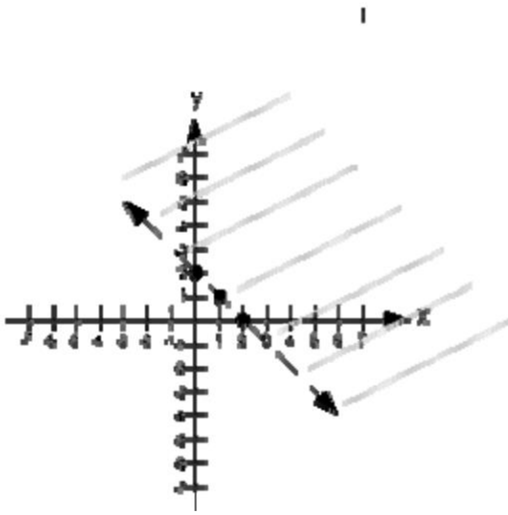
$0 > 2$ *False statement

Step 3: <u>Shade in the answer to the inequality.</u>

Since our **test** point (0, 0) made our inequality **FALSE**, this means it is not a solution.

Since it has to be on one side or the other of the boundary line, and it is not below it, then **our solution would lie above the boundary line**. This means we will shade in the part that is above it.

Note that the gray lines indicate where you would shad**e** your final answer.



Example 2: Graph the inequality
$2x - 3y \leq 6$.

Step 1: Graph the boundary line.
We are going to use the intercepts to help us graph the boundary line. Again, you can use any method that you want, unless the directions say otherwise.

When I'm working with only the boundary line, I will put an equal sign between the two sides to emphasize that we are working on the boundary line. That doesn't mean that I changed the problem. When we put it all together in the end, I will put the inequality back in.

What value is *y* on the *x*-intercept?
If you said 0, you are correct.

$2x - 3y = 6$ *Replace y with 0

$2x - 3(0) = 6$

$2x = 6$

$\frac{2x}{2} = \frac{6}{2}$ *Inverse of multiplication by 2 is div. by 2

$x = 3$

*x-intercept

***x*-intercept is (3, 0)**

What is the value of *x* on the *y*-intercept?

If you said 0, you are correct.

$2x - 3y = 6$ *Replace x with 0

$2(0) - 3y = 6$

$-3y = 6$ *Inverse of multiplication. by -3 is div. by -3

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

$y = -2$

*y-intercept

***y*-intercept is (0, -2).**

Plug in 1 for x to get a third solution:

$2x - 3y = 6$

$2(1) - 3y = 6$ *Replace x with 1

$2 - 3y = 6$

$2 - 3y - 2 = 6 - 2$

$-3y = 4$ *Inverse of add 2 is sub. 2

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

*Inverse of multiplication by -3 is div. by -3

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

1, -4/3) is another solution on the boundary line.

However, it advisable not to use fractional points that are not easy to graph like this one. In such cases, try another value of x that would yield a better solution

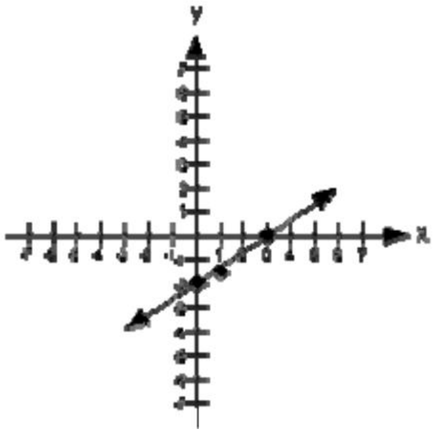
Solutions:

x	*y*	*(x, y)*
3	0	(3, 0)
0	-2	(0, -2)
1	-4/3	(1, -4/3)

Since the original problem has a ≤, this means it DOES include the boundary line.

So are we going to draw a solid or a dashed line for this problem? It looks like it will have to be a solid line.

Putting it all together, we get the following boundary line for this problem:



Step 2: Plug in a test point that is not on the boundary line.

Note how the boundary line separates it into two parts.

An easy test point would be (0, 0**)**. Note that it is a point that is not on the boundary line. In fact, it is located above the boundary line.

Let's put (0, 0) into the original problem and see what happens:

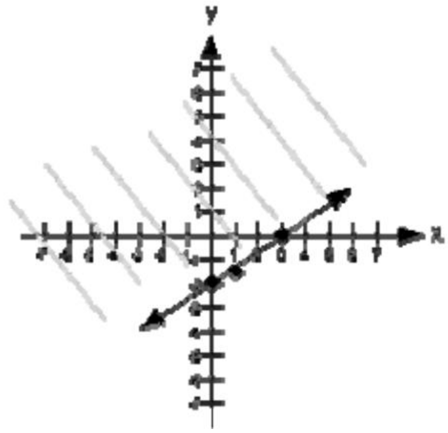
$2x - 3y \leq 6$	
$2(0) - 3(0) \leq 6$	*Replace *x* and *y* with 0
$0 \leq 6$	*True statement

Step 3: Shade in the answer to the inequality.

Since our test point (0, 0) made our inequality **TRUE**, this means it is a solution.

Our solution would lie above the boundary line. This means we will shade in the part that is above it.

Note that the gray lines indicate where you would shade your final answer.



Example 3: Graph the inequality $x < 4$.

Step 1: <u>Graph the boundary line.</u>

If we wrote this as an equation, it would be $x = 4$. This is in the form $x = c$, which is one of our "special" lines.

Remember what type of line $x = c$ graphs is.

It comes out to be a vertical line.

Every x's value on the boundary line would have to be 4.

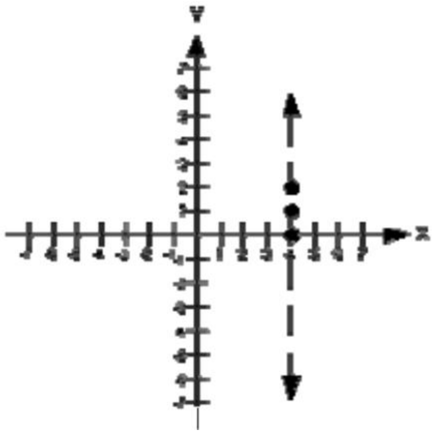
Solutions:

x	y	(x, y)
4	0	(4, 0)
4	1	(4, 1)
4	2	(4, 2)

Since the original problem has a <, this means it DOES NOT include the boundary line.

So are we going to draw a solid or a dashed line for this problem? It looks like it will have to be a dashed line.

Putting it all together, we get the following boundary line for this problem:



Step 2: Plug in a test point that is not on the boundary line.

Note how the boundary line separates it into two parts.

An easy test point would be (0, 0). Note that it is a point that is not on the boundary line. In fact, it is located to the left of the boundary line.

Let's put (0, 0) into the original problem and see what happens

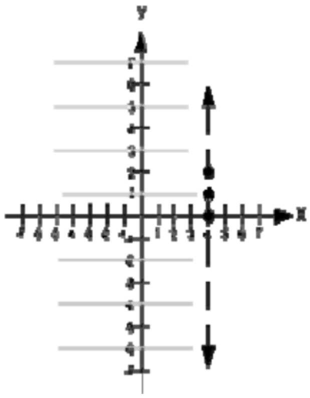
$x < 4$	*Replace x with 0
$0 < 4$	*True Statement

Step 3: Shade in the answer to the inequality.

Since our test point (0, 0) made our inequality **TRUE**, this means it is a solution.

Our solution would lie to the left of the boundary line. This means we will shade in the part that is to the left of it

Note that the gray lines indicate where you would shade your final answer.



Applications of linear inequalities are mostly found in linear programming.

1.7 Revision Exercise

1. Graph the following Inequalities:

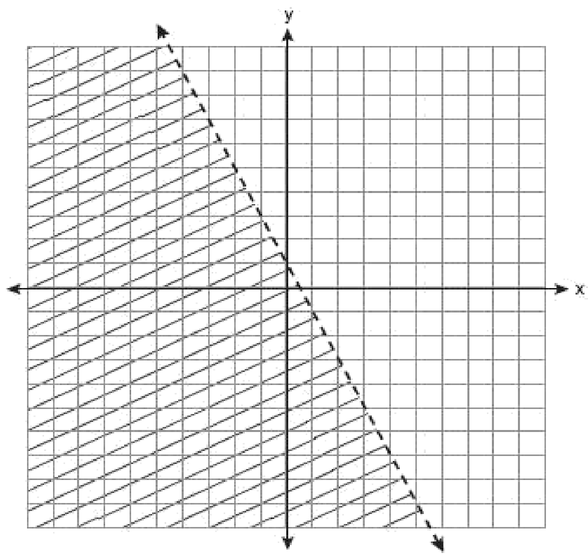
i) $y \leq x + 3$

ii) $y > 2x - 1$

iii) $2y \leq 4x + 6$

iv) $2x + y > 4$

2. Which inequality is represented by the graph below?



Suggested references

i. Linear Programming: Foundations and Extensions by Robert J. Venderbei

ii. Quantitative Techniques Simplified by N.A Saleemi

TOPIC 6

Linear programming

Objectives

By the end of the topic the learner should be able to:

i. Formulate an appropriate LP problem

ii. Graph the constraint inequalities

iii. Identify the solution space or feasible region which satisfies all the constraints simultaneously

iv. Locate the solution points on the feasible region.

v. Evaluate the objective function at each of the corner points.

vi. Identify the optimum value of the objective function.

6.1 Introduction

Definition

It is a mathematical technique that deals with the optimization of a linear function of variables known as objective function to a set of linear inequalities known as constraints.

t is a method of determining an optimum programme of inter-dependent activities in view of available resources.

6.1.1 Basic parts of a linear programming problem

i. **The objective function**: it describes the primary purpose of the formulation i.e. to maximize profit or to minimize cost.

ii. **The constraint set:** It is a system of equalities and / or inequalities describing the conditions under which optimization is to be accomplished e.g. machine hours, man-hours, materials etc.

6.1.2 Assumptions of Linear programming

i. Linearity: Costs, revenues or any physical properties, which form the basis of the problem, vary in direct proportion with the quantities or numbers of components produced.

ii. Divisibility: Quantities, revenues and costs are infinitely divisible i.e. any fraction or decimal answer is valid.

iii. Certainty: The technique makes no allowance for uncertainty in the estimate made.

iv. Positive solutions: Non-negativity constraints are introduced to ensure only positive values are considered.

v. Interdependence between demand products is ignored, products may be complementary or a substitute for one another.

vi. Time factors are ignored. All production is assumed to be instantaneous.

vii. Costs and benefits which cannot be quantified easily, such as liquidity, good will and labour stability are ignored.

6.2 Advantages of Linear Programming

i. Helps in attaining the optimum use of productive factors

ii. Improves the quality of decisions

iii. Improves the knowledge and skill of tomorrow’s executives

iv. It highlights the bottlenecks in the production process

v. It gives insight and perspective into problem situations

vi. Enables one to consider all possible solutions to problems.

vii. Enables one to come up with better and more successful decisions

viii. It's a better tool for adjusting to meet changing conditions.

6.3 Limitations of Linear Programming

i. It treats all relationships as linear

ii. It is assumed that any activity is infinitely divisible

iii. It takes into account single objective only i.e. profit maximization or cost minimization

iv. It can be adopted only under the condition of certainty i.e. that the resources available, per unit contribution, costs etc are known with certainty. This does not hold in real situations.

6.4 Applications of Linear Programming

Linear programming can find use in many areas where decisions need to be made based on scarcity of recourses. A few areas are mentioned below:.

i. Business and industry e.g. in the petroleum industry

ii. Food processing industry: to determine the optimal mix of feeds

iii. Paper and textile industry: To determine the optimal cutting method to minimize trim losses

iv. Transport industry: To determine the best route

v. Financial institutions – to determine investment plans.

vi. Advertising media: assigning advertising expenditures to different media plans

vii. Politics- campaign strategies

viii. Auditing – to find the number of financial audits

ix. Agriculture- amount of fertilizer to apply per acre

x. Hospital scheduling- number of nurses to employ

xi. Marketing – determining the best marketing strategy

xii. Crude oil refining

6.5 Mathematical formulation of Linear Programming problems

If there are n decision variables and m constraints in the problem, the mathematical formulation of the LP

is:-

Optimization (Max or Min) $Z=c_1x_1+c_2x_2+\ldots c_nx_n$

Subject to the constraints

$a_{11}x_1 + a_{12}x_2+\ldots a_{1n}x_n \leq b_1$

$a_{21}x_1 + a_{22}x_2+\ldots a_{2n}x_n \leq b_2$

$\vdots$

$a_{m1}x_1 + a_{m2}x_2+\ldots a_{mn}x_n \leq b_m$

and $x_1, x_2,\ldots x_n \geq 0$

Where

x_i- decision variable –

C_j- constant representing per unit contribution of the objective function of the

a_{ij}- -constant representing exchange coefficient of the j^{th} decision variable in the i^{th} constraint

b_i---constant representing i^{th} constraint requirement or availability

Solutions to LP problems

LP problems can be solved by the help of the following methods:

i. Graphical method

ii. Simplex method

In this study we shall concentrate on the graphical method.

6.6 Solving Linear Programming Problems Graphically.

In order to solve a LP problem graphically, the following procedure is adopted:-

i. Formulate the appropriate LP problem

ii. Graph the constraint inequalities as follows: treat each inequality as though it were an equality and for each equation, arbitrarily select two sets of points. Plot each of the points and connect them with appropriate lines.

iii. Identify the solution space or feasible region which satisfies all the constraints simultaneously. For ≤ constraints, this region is below the lines and for ≥ constraints; the region is above the lines.

iv. Locate the solution points on the feasible region. These points always occur at the corner points of the feasible region.

v. Evaluate the objective function at each of the corner points.

vi. Identify the optimum value of the objective function.

Example

At the start of the current week there are 30 units of X and 90 units of Y in stock. Available processing time on machine A is forecast to be 40 hours and on machine B is forecast to be 35 hours.

The demand for X in the current week is forecast to be 75 units and for Y is forecast to be 95 units. Company policy is to maximize the combined sum of the units of X and the units of Y in stock at the end of the week.

Formulate the problem of deciding how much of each product to make in the current week as a linear program.

Solve this linear program graphically.

Solution

Let

x be the number of units of X produced in the current week

y be the number of units of Y produced in the current week

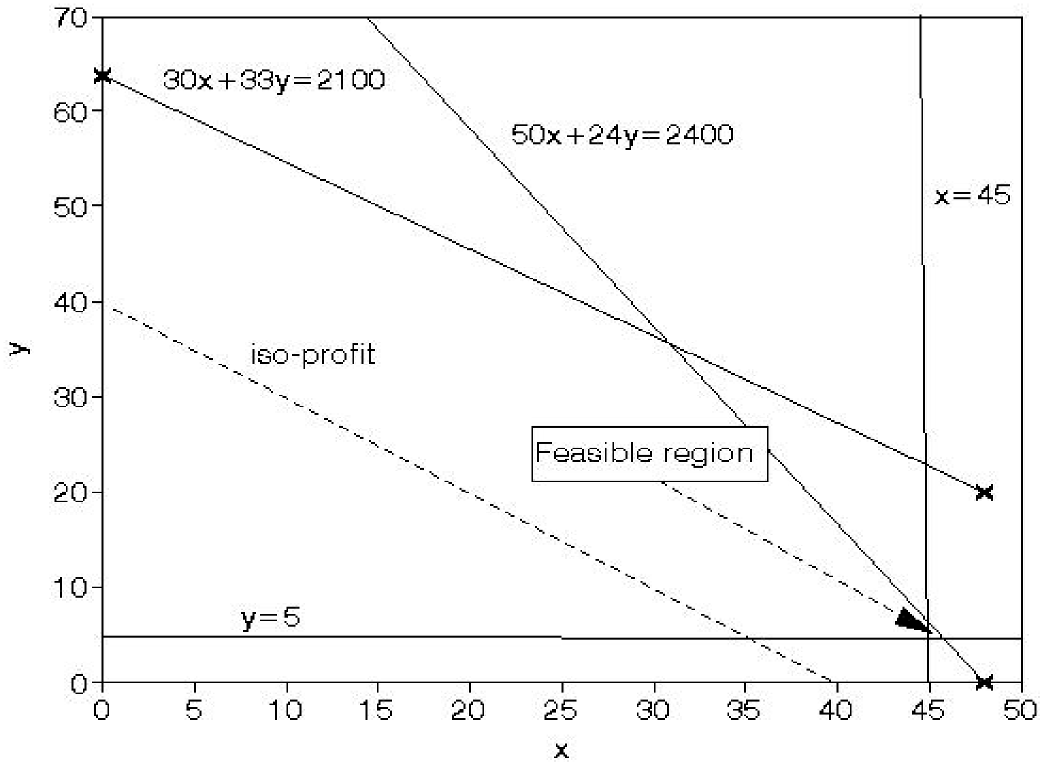
then the constraints are:

50x + 24y <= 40(60) machine A time
30x + 33y <= 35(60) machine B time
x >= 75 - 30
i.e. x >= 45 so production of X >= demand (75) - initial stock (30), which ensures we meet demand
y >= 95 - 90
i.e. y >= 5 so production of Y >= demand (95) - initial stock (90), which ensures we meet demand

The objective is: maximize (x+30-75) + (y+90-95) = (x+y-50)

i.e. to maximize the number of units left in stock at the end of the week

It is plain from the diagram below that the maximum occurs at the intersection of x=45 and 50x + 24y = 2400



6.7 Revision exercise

1. A firm is engaged in producing two products A and B. each unit of product A requires 2 kg of raw material and 4 labour hours for processing, whereas each unit of product B requires 3 kg of raw material and 3 hours of labour, of the same type. Every week, the firm has an availability of 60 kg of raw material and 96 labour hours. One unit of product A sold yields Sh. 40 and one unit of product sold gives Sh. 35 as profit. Formulate this problem as a linear programming problem to determine as how many units of each of the products should be produced per week so that the firm can earn the maximum profit. Assume that there is no marketing constraint so that all that is produced can be sold.

2. The Kenya Agricultural Research Institute suggested to a farmer to spread out at least 4800 kg of a special phosphate fertilizer and not less than 7200 kg of a nitrogen fertilizer to raise productivity of crops in his fields. There are two sources for obtaining these mixtures A and B. Both of these are available in bags weighing 100 kg each and they cost

Sh. 40 and Sh 24 respectively. Mixture A contains phosphate and nitrogen equivalent of 20 kg and 80 kg respectively, while mixture B contains these ingredients equivalent of 50 kg each. Determine how many bags of each type the farmer should buy in order to obtain the required fertilizer at minimum cost.

Suggested References

i. Linear Programming: Foundations and Extensions

ii. Quantitative Techniques Simplified by N.A Saleemi

Topic 7

UTILITY AND UTILITY FUNCTIONS

Objectives

By the end of the topic the learner should be able to

i) Define and relate utility to consumption

ii) Relate utility to demand

iii) Describe utility functions

iv) State the assumptions of utility

v) Explain the types of utility.

Utility theory provides a methodological framework for the evaluation of alternative choices made by individuals, firms and organizations. Utility refers to the satisfaction that each choice provides to the decision maker. Thus, utility theory assumes that any decision is made on the basis of the utility maximization principle, according to which the best choice is the one that provides the highest utility (satisfaction) to the decision maker.

7.1 Utility Theory in Consumer Behavior

Utility theory is often used to explain the behavior of individual consumers. In this case the consumer plays the role of the decision maker that must decide how much of each of the many different goods and services to consume so as to secure the highest possible level of total utility subject to his/her available income and the prices of the goods/services.

7.2 Utility Theory and Demand

In addition to providing an explanation of consumer disposition of income, utility theory is useful in establishing individual consumer demand curves for goods and services. A consumer's demand curve for a good or service shows the different quantities that consumers purchase at various alternative prices. Factors that are held constant are consumers' tastes and preferences, income, and price.

7.3 Utility Functions

In all cases the utility that the decision maker gets from selecting a specific choice is measure by a utility function U, which is a mathematical representation of the decision maker's system of preferences such that: $U(x) > U(y)$, where choice x is preferred over choice y or $U(x) = U(y)$, where choice x is indifferent from choice y—both choices are equally preferred.

Utility functions can be either cardinal or ordinal. In the former case, a utility function is used to derive a numerical score for each choice that represents the utility of this choice. In this setting the utilities (scores) assigned to different choices are directly comparable. For instance, a utility of 100 units towards a cup of tea is twice as desirable as a cup of coffee with a utility level of 50 units. In the ordinal case, the magnitude of the utilities (scores) are not important; only the ordering of the choices as implied by their utilities matters. For instance, a utility of 100 towards a cup of tea and a utility level of 50 units for a cup of coffee simply state that a cup of coffee is preferred to a cup of tea, but it cannot be argued that a cup of tea is twice as desirable as a cup of coffee. Within this setting, it is important to note that an ordinal utility function is not unique, since any monotonic increasing transformation of an ordinal utility function will still provide the same ordering for the choices.

7.4 Assumptions on Preferences

Irrespective of the type of utility function, utility theory assumes that preferences are complete, reflexive and transitive. The preferences are said to be complete if for any pair of choices x and y, one and only one of the following be stated: (1) x is preferred to y, (2) y is preferred to x, or (3) x and y are equally preferred. The preferences are said to be reflexive if for any pair of choices x and y such that x equally preferred to y, it is concluded that y is also equally preferred to x. Finally, the preferences are said to be transitive if for any three choices x, y, z such that x is preferred over y, and y is preferred over z, it is concluded that x is preferred over z. The hypotheses on reflexivity and transitivity imply that the decision maker is consistent (rational).

7.5 Marginal Rate of Substitution

A further assumption of utility theory is that decision makers are willing to trade one choice for another. The existing trade-offs define the marginal rate of substitution. As example suppose that two investment projects are considered by a decision maker. Project x has a return of 6 percent and a risk of 4 percent, whereas the return for project y is 5 percent and its risk is 2 percent. Furthermore assume that the decision maker considers both projects to be equally preferred. With this assumption it is clear that the decision maker is willing to increase the risk by 2 percent in order to improve return by 1 percent. Therefore, the marginal rate of substitution of risk for return is 2. In real world situations, the marginal rates of substitution are often decreasing. Such situations correspond to diminishing marginal utilities (marginal utility is defined as the change in total utility resulting from a one-unit change in consumption of the good or service). In the above example, we can assume that the decision maker is willing to take higher risks in order to get higher return, but only up to a specific point which is called saturation point. Once the risk has reached that point, the decision maker would not be willing to take any higher risk to increase return and therefore the marginal rate of substitution at this risk level would be zero

7.6 Multi-Attribute Utility Theory

The traditional framework of utility theory has been extended over the past three decades to the multi-attribute case, in which decisions are taken by multiple criteria. Multi-attribute utility theory has been evolved as one of the most important topics in multiple criteria decision making with many real world applications in complex real world problems.

The concept of utility can be used to analyze individual consumer behavior, to explain individual consumer demand curves as well as in modeling the decision makers' preferences. In all cases, it is assumed that some choices are evaluated and the best one is identified as the choice that maximizes the utility or satisfaction. The utility theory has been a research topic of major importance for the development of economics, decision theory, and management and it still attracts the interest of both practitioners and academic researchers.

7.7 Revision Questions

1. Define 'utility'

2. Explain the meaning

of i) Complete utility

ii) Transitive utility

iii) Reflexive utility

3. Distinguish between cardinal and ordinal utility.

Suggested References:

i. Quantitative Methods for Business and Economics

ii. Aleskerov, F., and B. Monjardet. *Utility Maximization, Choice and Preference.* Heidelberg: Springer Verlag, 2002.

Sample Papers

Sample Paper 1

UNIT TITLE : QUANTITATIVE TECHNIQUES   UNIT CODE : SBC 212

TIME: 2 HOURS

Instructions: Answer Question One and Any Other Two Questions

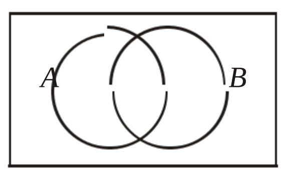
<u>**QUESTION ONE**</u> **(Compulsory)**   **(30 marks)**

a) Solve the inequality

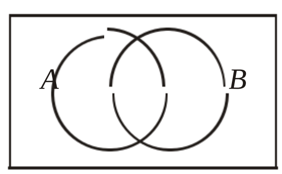
$2(x - 3) < 5x + 3$ (3 marks)

b) In each of the Venn diagrams, shade the region indicated.

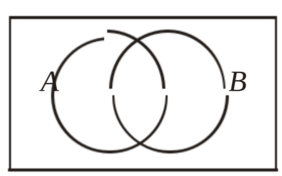
(i)   *A*   *B*



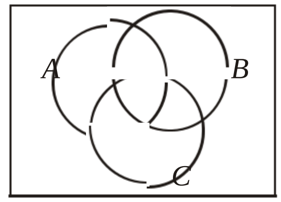
(ii)   The complement of (*A*   *B*)



iii)   The complement of (*A*   *B*)



(iv) A (B C)



(8 marks)

(c) Given a set of data; 2,9,8,4,7,6

i) Calculate the arithmetic mean (2 marks)

ii) Calculate the geometric mean (2 marks)

iii) Calculate the harmonic mean (2 marks)

iv) State the median (1 mark)

v) Calculate the standard deviation. (3 marks)

d) A publisher is planning to produce a new textbook. The fixed costs are Sh. 320,000 and the variable costs are Sh. 31.25 per book. The wholesale price will be Sh. 43.75 per book. How many books must the publisher sell to break-even? (5 marks)

e) A senior lecturer is set to give a series of four lectures. If he doesn't give any one lecture, the lecture is given by his assistant. He is certain to give the first lecture. The probability of giving the second lecture is 0.45. If he gives the second lecture, the probability of giving the third lecture is 0.7, otherwise it is 0.4. If he gives the third lecture, the probability of giving the fourth lecture is 0.35 otherwise it is 0.7.Calculate the probability that the lecturer

i) Gives all the four lectures

ii) Gives two lectures only

Iii) Gives lecture one and lecture four only. (5 marks)

QUESTION TWO (20 marks)

Let ξ = $\{x : 1 \leq x < 17, x \quad \mathbb{N}\}$.

a) P , Q and R are the subsets of ξ such that

P = {multiples of four};

Q = (factors of 36};

R = {square numbers}.

(a) List the elements of

(i) ξ

(ii) PQR.

(2)

(b) Describe in words the set $P \quad Q$.

(1)

(c) (i) Draw a Venn diagram to show the relationship between sets P, Q and R.

(2)

(ii) Write the elements of ξ in the appropriate places on the Venn diagram.

(1)

b) In a club with 60 members, everyone attends either on Tuesday for Drama (D) or on Thursday for Sports (S) or on both days for Drama and Sports.

One week it is found that 48 members attend for Drama and 44 members attend for Sports and x members attend for both Drama and Sports.

(i) Draw and **label fully** a Venn diagram to illustrate this information.

(3)

(ii) Find the number of members who attend for both Drama and Sports.

(2)

(iii) Describe, in words, the set represented by $(D \quad S)'$.

(2)

(iv) What is the probability that a member selected at random attends for Drama only or Sports only?

(3)

The club has 28 female members, 8 of whom attend for both Drama and Sports.

What is the probability that a member of the club selected at random

(i) is female and attends for Drama only or Sports on

(ii) is male and attends for both Drama and Sports? (2)

QUESTION THREE (20 marks)

a) A market researcher investigating consumers' reference for three brands of beverages namely: coffee, tea and cocoa, in Kisii Town gathered the following information. From a sample of 800 consumers, 230 took coffee, 245 took tea and 325 took cocoa, 30 took all the three beverages, 70 took coffee and cocoa, 110 took coffee only, 185 took cocoa only.

Required:

(i) Present the above information in a Venn diagram

(ii) The number of customers who took tea only

(iii) The number of customers who took coffee and tea only

(iv) The number of customers who took tea and cocoa only

(v) The number of customers who took none of the beverages. (10 marks)

b) A public transportation company has been experimenting on a possibility of developing a system of charging fares .The demand functions, which expresses the ridership as a function of fare charged is given below:

$$Q=10,000 - 125p$$

Where Q equals the average number of riders per hour and p equals the fare in shillings.

(i) Determine the fare, which should be charged in order to maximize hourly bus fare revenue.

(ii) What is the expected maximum revenue?

(iii) How many riders per hour are expected under this figure? (10 marks)

QUESTION FOUR (20 marks)

The age of each patient over the age of 14 visiting a doctor's practice in one day was recorded as

18	18	76	15	72	45	48	62
21	27	45	43	28	19	17	37
35	34	23	25	46	56	32	18
24	34	32	56	29	43		

Use a tally chart to produce a grouped frequency distribution with age groups 15-19, 20-24, 25-29 etc. (3marks)

a. What is the relative frequency of patients in the age groups 70-74? (2marks)

b. Express the number of patients in the age group 25-29 as a percentage (2marks)

c. Calculate the mean and the standard deviation (6marks)

d. Calculate the median (3marks)

e. Calculate the Pearson's coefficient of skewness. (4marks

QUESTION FIVE (20 marks)

a) State Four advantages and Three disadvantages of Linear programming. (7 marks)

b) A firm is engaged in producing two products A and B. each unit of product A requires 2 kg of raw material and 4 labour hours for processing, whereas each unit of product B requires 3 kg of raw material and 3 hours of labour, of the same type. Every week, the firm has an availability of 60 kg of raw material and 96 labour hours. One unit of product A sold yields Sh. 40 and one unit of product sold gives Sh. 35 as profit. Formulate this problem as a linear programming problem to determine as how many units of each of the products should be produced per week so that the firm can earn the maximum profit. Assume that there is no marketing constraint so that all that is produced can be sold. (13 marks)

Sample paper 2

UNIT TITLE : QUANTITATIVE TECHNIQUES UNIT CODE : SBC 212

TIME: 2 HOURS

Instructions: Answer Question One and Any Other Two Questions

__QUESTION ONE__ **(Compulsory) (30 marks)**

a) Given ℤ the set of integers, ℚ the set of rational numbers, ℝ the set of real numbers.

(a) Write down an element that belongs to ℝ Uℤ.

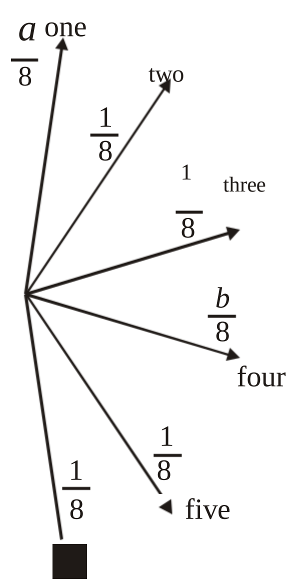
(b) Write down an element that belongs to ℚ Uℤ

(c) Write down an element that belongs to $\mathbb{Q}_I$

(d) Use a Venn diagram to represent the sets ℤ, ℚ and ℝ.

(**6 marks**)

b)



(a) Find the values of *a* and *b* in the diagram.

(b) Both dice are thrown. Calculate the probability that two fours appear on top.

(c) One of the dice is thrown once. The result is not a two or a three. What is the probability that it is a six?

(8 marks)

c) The universal set U is defined as the set of positive integers less than 10. The subsets A and B are defined as:

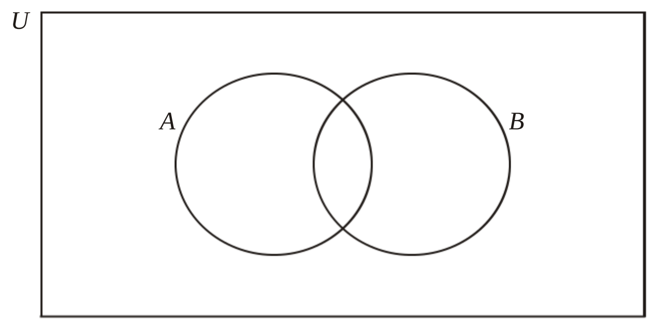
A = {integers that are multiples of 3}
B = {integers that are factors of 30}

(a) List the elements of

(i) A;

(ii) B.

(b) Place the elements of A and B in the appropriate region in the Venn diagram below.



(4 marks)

d) Given the set of data, 2,3,4,5

Calculate the

i) Harmonic mean

ii) Geometric mean

iii) Standard deviation **(7 marks)**

e) State three assumptions of linear programming**.** **(3marks)**

QUESTION TWO

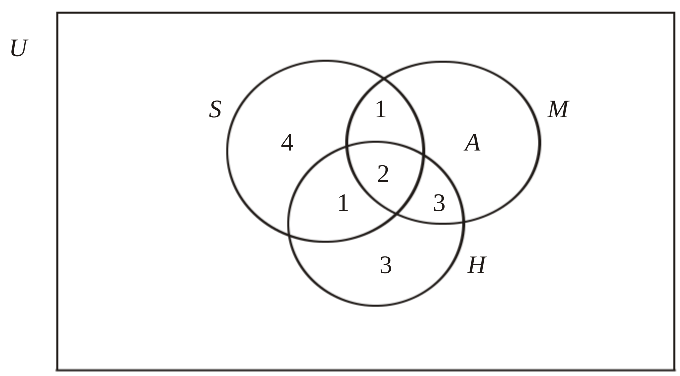
a) The following results were obtained from a survey concerning the reading habits of students.

60 % read magazine P
50 % read magazine Q
50 % read magazine R
30 % read magazines P and Q
20 % read magazines Q and R
30 % read magazines P and R
10 % read all three magazines

Represent all of this information on a Venn diagram.

(5)

b) The Venn diagram below shows the number of students studying Science (*S*), Mathematics (*M*) and History (*H*) out of a group of 20 college students. Some of the students do not study a ny of these subjects, 8 study Science, 10 study Mathematics and 9 study History.



(a) (i) How many students belong to the region labeled *A*?

(ii) Describe in words the region labeled *A*.

(iii) How many students do not study any of the three subjects?

(5)

(b) Draw a sketch of the Venn diagram above and shade the region which represents *S H*.

(1)

(c) Calculate $n(S \quad H)$.

(2)

This group of students is to compete in an annual quiz evening which tests knowledge of Mathematics, Science and History. The names of the twenty students are written on pieces of paper and then put into a bag.

(d) One name is randomly selected from the bag. Calculate the probability that the student selected studies

(i) all three subjects;

(ii) History or Science.

(2)

(e) A team of two students is to be randomly selected to compete in the quiz evening. The first student selected will be the captain of the team. Calculate the probability that

(i) the captain studies all three subjects and the other team member does not study any of the three subjects;

(ii) one student studies Science only and the other student studies History only;

(iii) the second student selected studies History, given that the captain studies History and Mathematics.

(5)
(Total 15 marks)

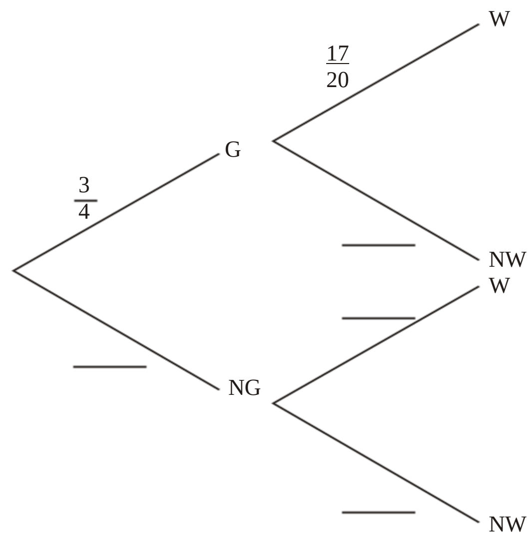
<u>QUESTION THREE</u>

Today Philip intends to go walking. The probability of good weather (G) is $\frac{3}{4}$. If the weather is good, the probability he will go walking (W) is $\frac{17}{20}$. If the weather forecast is not good (NG) the probability he will go walking is $\frac{1}{5}$.

(a) Complete the probability tree diagram to illustrate this information.

(b) What is the probability that Philip will go walking?

(Total 5 marks)



A box contains 10 coloured light bulbs, 5 green, 3 red and 2 yellow. One light bulb is selected at random and put into the light fitting of room A.

(a) What is the probability that the light bulb selected is

(i) green? **(1)**

(ii) not green? **(1)**

A second light bulb is selected at random and put into the light fitting in room B.

(b) What is the probability that

(i) the second light bulb is green given the first light bulb was green? **(l)**

(ii) both light bulbs were not green? **(2)**

(iii) one room had a green light bulb and the other room does not have a green light bulb? **(3)**

A third light bulb is selected at random and put in the light fitting of room C.

(c) What is the probability that

(i) all three rooms have green light bulbs? **(2)**

(ii) only one room has a green light bulb?

(3)

(iii) at least one room has a green light bulb?

(2)

(Total 15 marks)

QUESTION FOUR

a) In a club with 60 members, everyone attends either on Tuesday for Drama (*D*) or on Thursday for Sports (*S*) or on both days for Drama and Sports.

One week it is found that 48 members attend for Drama and 44 members attend for Sports and x members attend for both Drama and Sports.

(a) (i) Draw and **label fully** a Venn diagram to illustrate this information.

(3)

(ii) Find the number of members who attend for both Drama and Sports.

(2)

(iii) Describe, in words, the set represented by (*D* *S*)'.

(2)

(iv) What is the probability that a member selected at random attends for Drama only or Sports only?

(1)

(8 marks)

b) Let

ξ = {positive integers less than 15};

X= {multiples of 2};

Y = {multiples of 3}.

(a) Show, in a Venn diagram, the relationship between the **sets** ξ, *X* and *Y*.

(1)

(b) List the elements of:

(i) X Y

(1)

(ii) X Y.

(1)

(c) Find the **number of elements** in the complement of $(X \quad Y)$I

(1)

(4 marks)

c) The cost of producing a mathematics textbook is $ 15 (US dollars) and it is then sold for $ x.

(a) Find an expression for the profit made on each book sold.

(2)

A total of (100 000 – 4000x) books is sold.

(b) Show that the profit made on all the books sold is

$$P = 160\,000x - 4000x^2 - 1500\,000.$$

(c) (i) Find $\frac{dP}{dx}$.

(2)

(ii) Hence calculate the value of x to make a maximum profit

(2)

(d) Calculate the number of books sold to make this maximum profit.

(2)

(8 marks)

d) From the results of two colleges A and B, state which of them is better**.** **(5 marks)**

Name of the	College A		College B	
examination	Appeared	Passed	Appeared	Passed
M.A.	30	25	100	80
M. Com.	50	45	120	95
B.A.	200	150	100	70
B. Com.	120	75	80	50
Total	400	295	400	295

QUESTION FIVE

(a) The heights of 200 students are recorded in the following table.

Height (h) in cm	Frequency
$140 \le h < 150$	2
$150 \le h < 160$	28
$160 \le h < 170$	63
$170 \le h < 180$	74
$180 \le h < 190$	20
$190 \le h < 200$	11
$200 \le h < 210$	2

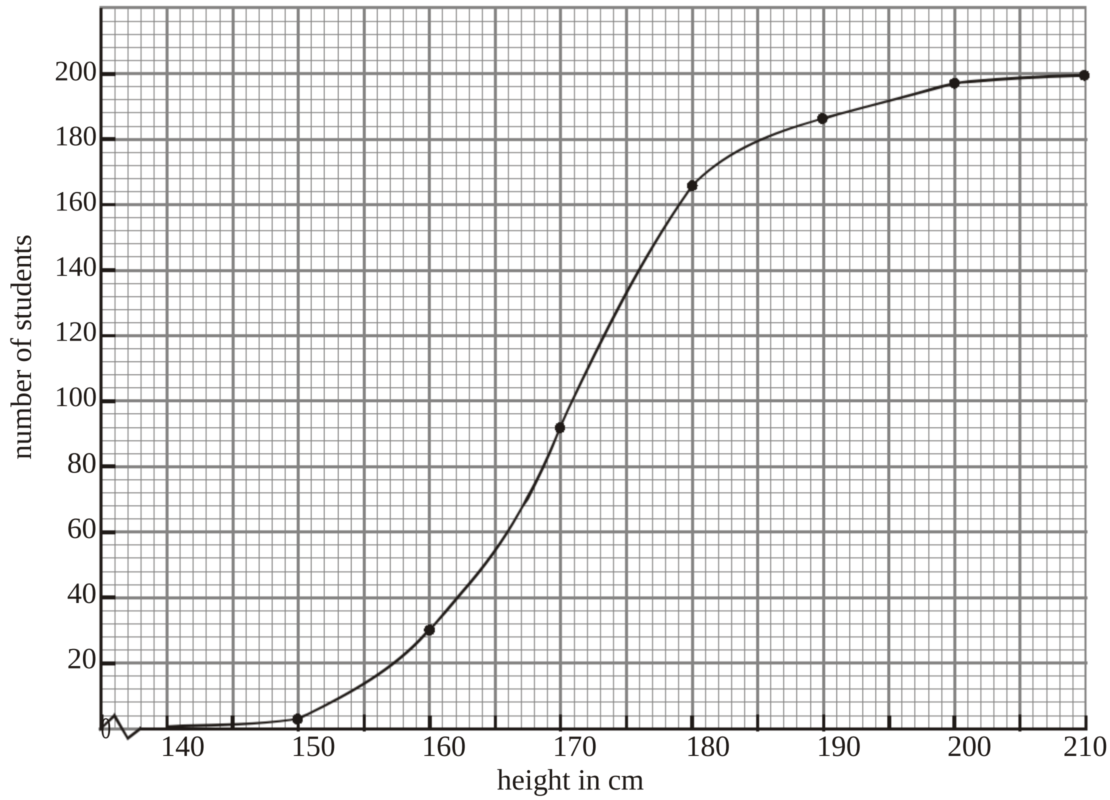
(a) Write down the modal group.

(1)

(b) Calculate an estimate of the mean and standard deviation of the heights.

(4)

The cumulative frequency curve for this data is drawn below.



(c) Write down the median height.

(1)

(d) The upper quartile is 177.3 cm. Calculate the interquartile range.

(2)

(e) Find the percentage of students with heights less than 165 cm.

(2)

(Total 10 marks)

(b) The table below shows the number of left and right handed tennis players in a sample of 50 males and females.

	Left handed	Right handed	Total
Male	3	29	32
Female	2	16	18
Total	5	45	50

If a tennis player was selected at random from the group, find the probability that the player is

(a) male and left handed;

(b) right handed;

(c) right handed, given that the player selected is female.

(4 marks)

(c)

i) What is linear programming?

Ii) State three assumptions of linear programming

iii) State tow limitations of linear programming **(6 marks)**