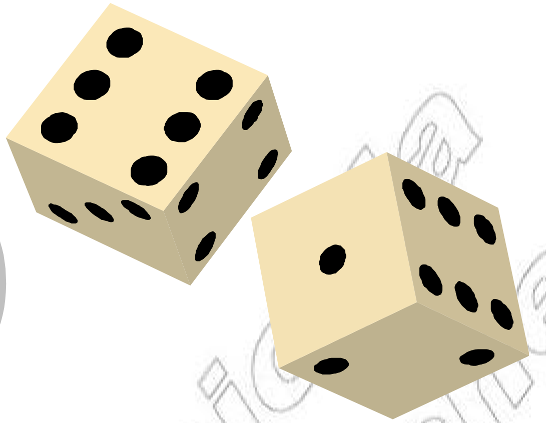
Unit 6



STATISTICS AND PROBABILITY

Unit Outcomes:

After completing this unit, you should be able to:

- *know methods and procedures in collecting and presenting simple statistical data.*
- *know basic concepts about statistical measures.*
- *understand facts and basic principles about probability.*
- *solve simple mathematical problems on statistics and probability.*

Main Contents

6.1 Statistical data

6.2 Probability

Key Terms

Summary

Review Exercises

INTRODUCTION

YOU HAVE SOME KNOWLEDGE ABOUT STATISTICS AND ITS BASICS, SUCH AS COLLECTI
PRESENTATION OF DATA, ETC., FROM YOUR PRIMARY GRADE MATHEMATICS. IN T
FORMALLY DEFINE 'STATISTICS' AS A BRANCH OF APPLIED MATHEMATICS AND LEAR
COLLECTION, PRESENTATION AND ANALYSIS OF NUMERICAL DATA. THE UNIT ALSO
CONCEPT OF PROBABILITY, WHICH WAS INTRODUCED IN GRADE 8 AND TEACHES YOU MORE ABOUT
EXPERIMENTAL AND THEORETICAL APPROACHES TO PROBABILITY AND HELPS YOU
PROBLEMS BASED ON THESE APPROACHES.

6.1 STATISTICAL DATA

6.1.1 Collection and Tabulation of Statistical Data

Group Work 6.1

1 SPLIT THE CLASS INTO THREE GROUPS. LET THE CLA
YEAR'S MATHEMATICS RESULT OF THE SCHOOL I
THE SCHOOL OFFICE'S RECORDS.

LET GROUP B COLLECT INFORMATION ABOUT THE DISEASES TREATED IN YOUR N
CENTRE, HOSPITAL, OR HEALTH POST. LET GROUP C MEASURE THE HEIGHT OF EAC
YOUR CLASS AND CONSIDER ITS DISTRIBUTION BY AGE AND SEX.

ANSWER THE FOLLOWING QUESTIONS USING THE INFORMATION GATHERED BY EAC

A HOW MANY STUDENTS APPEARED FOR THE EXAM?

B HOW MANY STUDENTS SCORED A IN THE NATIONAL EXAM?

C WHAT WAS THE SCORE OBTAINED BY MOST OF THE STUDENTS?

D WHICH DISEASES ARE TREATED MOST FREQUENTLY?

E WHAT IS THE AVERAGE HEIGHT OF THE CLASS?

F ARE MALES OR FEMALES TALLER?

2 DISCUSS MORE ABOUT THE IMPORTANCE AND PURPOSE OF STATISTICS.

3 WHAT IS THE ANNUAL BIRTH RATE AND DEATH RATE IN ETHIOPIA? WHICH GOV
AGENCY IS RESPONSIBLE FOR THE PREPARATION OF SUCH RECORDS?

4 WHY DOES THE GOVERNMENT OF ETHIOPIA CARRY OUT A CENSUS EVERY TEN YEA

THERE ARE MANY DEFINITIONS OF THE TERM statistics GIVEN BY DIFFERENT SCHOLARS. HOWEVER,
FOR THE PURPOSE OF THIS UNIT, WE WILL CONFINE OURSELVES TO THE FOLLOWING:

Definition 6.1

Statistics is the science of *collecting*, *organizing*, *presenting*, *analyzing* and *interpreting data* (quantitative information) in order to draw conclusions.

BASICALLY, STATISTICS IS A PROCEDURAL PROCESS CONCERNING FIVE LOGICAL STEPS OF NUMERICAL DATA. THESE ARE:

collection of data → organization of data → presentation of data → analysis of data → interpretation of data

1 Collection of data

THE FIRST STEP OF STATISTICS IS COLLECTION OF DATA. THIS IS THE PROCESS OF OBTAINING MEASUREMENTS OR COUNTS. FOR EXAMPLE, MEASURING THE HEIGHTS OF STUDENTS I OR COUNTING THE NUMBER OF PERSONS ADMITTED TO A CERTAIN HOSPITAL ARE EXA data collection.

2 Organization of data

THE SECOND STEP OF STATISTICS REFERS TO THE ORGANIZATION OF DATA. COLLECTE ORGANIZED IN A SUITABLE FORM TO UNDERSTAND THE INFORMATION GATHERED. THE MUST BE edited, classified AND tabulated.

3 Presentation of data

THE MAIN PURPOSE OF DATA PRESENTATION IS TO FACILITATE STATISTICAL ANALYS DONE BY ILLUSTRATING THE DATA USING GRAPHS AND DIAGRAMS LIKE BAR GRAPH, H CHARTS, PICTOGRAMS, FREQUENCY POLYGONS, ETC.

4 Analysis of data

IN ORDER TO MEET THE DESIRED PURPOSE OF INVESTIGATION, DATA HAS TO BE AN PURPOSE OF ANALYZING DATA IS TO HIGHLIGHT INFORMATION USEFUL FOR DECISION M

5 Interpretation of data

BASED ON ANALYZED DATA, CONCLUSIONS HAVE TO BE DRAWN. THIS STEP USUAL DECISION MAKING ABOUT A LARGE COLLECTION OF OBJECTS (THE POPULATION INFORMATION GATHERED FROM A SMALL COLLECTION OF SIMILAR OBJECTS (THE SAMP

THE DECISION MAKING PROCESSES USED BY THE MANAGERS OF MODERN BUSINESSES A IS GOVERNED BY STATISTICAL APPLICATION. STATISTICAL METHODS CAN BE APPLIED T WHERE NUMERICAL INFORMATION IS GATHERED WITH THE OBJECTIVE OF MAKING RAT IN THE FACE OF UNCERTAINTY.

THE FOLLOWING EXAMPLES SHOW US HOW STATISTICS PLAYS A MAJOR ROLE IN DECISIO
DIFFERENT SECTORS.

EXAMPLE 1 INFORMATION GATHERED ABOUT THE INCIDENCE OR PREVALENCE OF DIS COMMUNITY PROVIDES USEFUL INFORMATION ON CHANGING TRENDS I STATUS, MORTALITY, NUTRITIONAL STATUS OR ENVIRONMENTAL HAZARDS

EXAMPLE 2 STATISTICS IS USED TO STUDY EXISTING CONDITIONS AND THE RATE OF PREVALEN HIV/AIDS IN ORDER TO DESIGN NEW PROGRAMS OR TO STUDY THE MERI DIFFERENT METHODS ADOPTED TO CONTROL HIV/AIDS. IT ASSISTS IN DETE THE EFFECTIVENESS OF NEW MEDICATION AND THE IMPORTANCE OF COUNS

EXAMPLE 3 DEMOGRAPHIC DATA ABOUT POPULATION SIZE, ITS DISTRIBUTION BY AGE AND THE RATE OF POPULATION GROWTH, ETC., ALL HELP POLICY MA DETERMINING FUTURE NEEDS SUCH AS FOOD, CLOTHING, HOUSING, ED HEALTH FACILITIES, WATER, ELECTRICITY AND TRANSPORTATION SYSTEMS

EXAMPLE 4 RECORDING ANNUAL TEMPERATURES IN A COUNTRY PROVIDES THE COMMU TIMELY WARNING OF ENVIRONMENTAL HAZARDS.

EXAMPLE 5 STATISTICAL DATA COLLECTED ON CUSTOMER SERVICES PROVIDES FEEDB HELP TO REFORM POLICIES AND SYSTEMS.

IN THE ABSENCE OF ACCURATE AND TIMELY DATA, IT IS IMPOSSIBLE TO FORM SUITA STATISTICS ALSO PLAYS A VITAL ROLE IN MONITORING THE PROPER IMPLEMENTATIO AND POLICIES.

IN ITS ORDINARY USAGE, POPULATION REFERS TO THE NUMBER OF PEOPLE LIVING I COUNTRY. IN STATISTICS, HOWEVER, **population** REFERS TO THE COMPLETE COLLECTION OF INDIVIDUALS, OBJECTS OR MEASUREMENTS THAT HAVE A COMMON CHARACTERISTIC.

GAINING ACCESS TO AN ENTIRE GROUP (OR POPULATION) IS OFTEN DIFFICULT, EXP SOMETIMES DESTRUCTIVE. THEREFORE, INSTEAD OF EXAMINING THE ENTIRE GROUP, EXAMINES A SMALL PART OF THE GROUP, CALLED A **sample**

DATA CAN BE CLASSIFIED AS EITHER **qualitative** OR **quantitative**. HOWEVER, STATISTICS DEALS MAINLY WITH QUANTITATIVE DATA.

EXAMPLE 6 DATA COLLECTED FROM THE POPULATION OF STUDENTS IN ETHIOPIA COUL

I **Qualitative** IF THE DATA IS BASED ON SOME CHARACTERISTIC WHOSE VALUES A NUMBERS, SUCH AS THEIR EYE COLOUR, SEX, RELIGION OR NATIONALITY.

II **Quantitative** IF THE DATA IS NUMERICAL SUCH AS HEIGHT, WEIGHT, AGE OR SCO TESTS.

A RULE WHICH GIVES A CORRESPONDING VALUE TO EACH MEMBER OF A POPULATIO **population function**.

EXAMPLE 7 HERE IS A TABLE SHOWING THE APPROXIMATE SIZES OF MAJOR LAKES IN ET

TABLE 6.1: SIZE OF MAJOR LAKES IN ETHIOPIA

Name of Lake	Length (km)	Width (km)	Area (km^2)
ABAYA	60	20	1160
ABAYATA	17	15	205
ASHENGE	5	4	20
HAWASSA	16	9	229
CHAMO	26	22	551
HAYK	7	5	35
KOKA	20	15	205
LANGANO	18	16	230
SHALLA	28	12	409
TANA	70	60	3600
ZIWAY	25	20	434

WE CAN THINK OF THE SET OF THE ELEVEN LAKES AS THE POPULATION AND THE FUNCT L: LENGTH, W: WIDTH, A: AREA, ETC AS FUNCTIONS ON THIS POPULATION.

EXAMPLE 8 THE FOLLOWING TABLE SHOWS THE AGE OF 10 STUDENTS IN A CERTAIN CL

TABLE 6.2: AGE OF STUDENTS

Name of student x	Age (in years) A(x)
ABEBE	18
ABDU	17
BAYISSA	16
FATUMA	17
HIWOT	15
KIDANE	14
LEMLEM	18
MESERET	17
OMOD	15
ZEHARA	16

THE STUDENTS ARE MEMBERS OF THE POPULATION AND THEIR AGE, A IS THE POPULAT
STATISTICS CAN BE CLASSIFIED INTO TWO TYPES: DESCRIPTIVE STATISTICS AND INFERE

Definition 6.2

Descriptive statistics is a branch of statistics concerned with summarizing and describing a large amount of data without drawing any conclusion about a particular bit of data.

DESCRIPTIVE STATISTICS DESCRIBES INFORMATION COLLECTED THROUGH NUMERICA
CHARTS, GRAPHS AND TABLES. THE MAIN PURPOSE OF DESCRIPTIVE STATISTICS IS T
OVERVIEW OF THE INFORMATION COLLECTED.

Definition 6.3

Inferential statistics is a branch of statistics concerned with interpreting data and drawing conclusions.

WE CAN CLASSIFY DATA AS primary data AND secondary data.

1 Primary data

DATA IS SAID TO BE primary, IF IT IS OBTAINED FIRST HAND FOR THE PARTICULAR PURPO
WHICH ONE IS CURRENTLY WORKING. PRIMARY DATA IS ORIGINAL DATA, OBTAINED PE
PRIMARY SOURCES BY OBSERVATION, INTERVIEW OR DIRECT MEASUREMENT.

EXAMPLE 9 IF YOU MEASURE THE HEIGHTS OF STUDENTS IN YOUR CLASS, THIS IS PRIMARY DATA.

EXAMPLE 10 THE DATA GATHERED BY THE MINISTRY OF EDUCATION ABOUT THE NU
STUDENTS ENROLLED IN DIFFERENT UNIVERSITIES OF ETHIOPIA IS PRIMAR
THE MINISTRY ITSELF. (*If you were to use this data, it would be secondary data for you.*)

2 Secondary data

DATA WHICH HAS BEEN COLLECTED PREVIOUSLY (FOR SIMILAR OR DIFFERENT PURPOS
secondary data. SECONDARY DATA REFERS TO THAT DATA WHICH IS NOT ORIGINATE
RESEARCHER HIMSELF/HERSELF, BUT WHICH HE/SHE OBTAINS FROM SOMEONE ELSE'S F
SOURCES OF SECONDARY DATA ARE OFFICIAL PUBLICATIONS, JOURNALS, NEWSPAP
STUDIES, NATIONAL STATISTICAL ABSTRACTS, ETC.

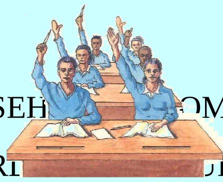
EXAMPLE 11 REPORTS ON THE NUMBER OF VICTIMS OF HIV/AIDS BY THE MINISTRY HEALTH IS SECONDARY DATA FOR ANYONE OTHER THAN THE MINISTRY.

EXAMPLE 12 THE 2007 CENSUS OF POPULATION SIZE OF REGIONS BY SEX REPORTED B CENTRAL STATISTICAL AGENCY (CSA) IS SECONDARY DATA FOR THE GOVE

INFORMATION EXPRESSED IN QUANTITATIVE FORM CAN RESULT IN SUCH A LARGE AMO UNLESS THESE FIGURES ARE PRESENTED IN SOME ORGANIZED MANNER, THEIR SIGNIFIC LOST. ONE OF THE BASIC METHODS OF PRESENTING STATISTICAL DATA IS PUTTING IT I tables. TO DO THIS, OFTEN THE DATA NEEDS TO BE CLASSIFIED.

Classification IS THE PROCESS OF ARRANGING THINGS INTO GROUPS OR CLASSES.

ACTIVITY 6.1



1 CLASSIFY THE EMPLOYEES IN YOUR SCHOOL BY HOUSEH OME.

2 GROUP THE NUMBER OF HIV/AIDS VICTIMS RECOR R NEAREST HEALTH CENTRE ACCORDING TO THEIR AGE GROUP.

3 COLLECT DATA ON AGE, HEIGHT AND MATHEMATICS EXAM SCORE OF THE STUD CLASS. CLASSIFY OR TABULATE THE DATA COLLECTED.

DIFFERENT PEOPLE OR ORGANIZATIONS COLLECT DATA FOR DIFFERENT REASONS A CLASSIFICATION THEY USE IS ALSO DIFFERENT ACCORDINGLY. TO SEE THIS, CONSIDE EXAMPLES.

EXAMPLE 13 AN ECONOMIST IN THE LABOUR DEPARTMENT OF A REGIONAL SOCIAL A BUREAU MAY CLASSIFY THE HOUSEHOLDS IN A CERTAIN LOCALITY BY I INCOME AS SHOWN IN THE TABLE BELOW.

TABLE 6.3: MONTHLY INCOME OF 300 HOUSEHOLDS

Income (in Birr)	Number of households
UNDER 350	85
BETWEEN 350 AND 650	72
BETWEEN 651 AND 950	64
BETWEEN 951 AND 1250	48
BETWEEN 1251 AND 155	21
ABOVE 1550	10
TOTAL	300

EXAMPLE 14 ACCORDING TO THE 2007 ETHIOPIAN CENSUS, THE ETHIOPIAN CENTRAL STA AGENCY (CSA) HAS CLASSIFIED THE POPULATION BY SEX AS FOLLOWS.

TABLE 6.4: POPULATION BY SEX (2007 ETHIOPIAN CENSUS)

Region	Male (in 1000)	Female (in 1000)	Both sexes (in 1000)
TIGRAY	2124.8	2189.6	4314.4
AFFAR	786.3	624.7	1411.0
AMHARA	8636.9	8577.2	17214.1
OROMIYA	13676.2	13482.3	27158.5
SOMALE	2468.8	1970.4	4439.2
BENSHANGU	340.4	330.5	670.9
SNNP	7482.0	7560.5	15042.5
GAMBELA	159.7	147.2	306.9
HARARI	92.3	91.1	183.4
ADDIS ABAE	1304.5	1433.7	2738.2
DIRE DAWA	171.9	170.9	342.8
Total	37243.8	36577.4	73821.2

A STATISTICAL TABLE IS A SYSTEMATIC PRESENTATION OR ORGANIZATION OF NUM COLUMNS AND ROWS. COLUMNS ARE VERTICAL ARRANGEMENTS AND ROWS ARE HO MAIN PURPOSE OF A STATISTICAL TABLE IS TO ALLOW THE READER TO QUICKLY A INFORMATION. A TITLE AND ROW AND COLUMN HEADERS ARE IMPORTANT.

Exercise 6.1

1. WHAT ARE THE STEPS USED IN DOING A STATISTICAL STUDY?
2. WHAT DO WE MEAN BY ORGANIZING OR PRESENTING DATA?
3. EXPLAIN EACH OF THE FOLLOWING STATISTICAL TERMS BY GIVING EXAMPLES.
   - A QUALITATIVE DATA
   - B QUANTITATIVE DATA
   - C POPULATION
   - D POPULATION FUNCTION
   - E SAMPLE
4. MENTION FOUR USES OF STATISTICS.
5. WHAT IS DESCRIPTIVE STATISTICS?
6. DESCRIBE IN YOUR OWN WORDS THE DIFFERENCE BETWEEN A POPULATION AND A
7. DETERMINE WHETHER THE FOLLOWING DATA IS QUALITATIVE OR QUANTITATIVE.
   - A GENDER
   - B TEMPERATURE
   - C ZIP CODE
   - D NUMBER OF DAYS
   - E RELIGIONS
   - F OCCUPATIONS
   - G AGES
   - H COLOURS
   - I NATIONALITY

8 MENTION SOME ADVANTAGES OF TABULAR PRESENTATION OF DATA.

9 WHY IS IT NECESSARY TO ORGANIZE DATA IN A SYSTEMATIC MANNER AFTER IT HAS BEEN COLLECTED?

10 DRAFT A TABLE TO SHOW THE FOLLOWING DATA, COLLECTED FROM EMPLOYEES I

- **A** SEX
- **B** THREE RANKS: SUPERVISORS, ASSISTANTS AND CLERKS
- **C** YEARS: 2000 AND 2001
- **D** AGE GROUP: 18 YEARS AND UNDER, OVER 18 BUT LESS THAN 50 YEARS, OVER 5

6.1.2 Distributions and Histograms

INFORMATION (DATA) IS OBTAINED FROM A CENSUS, EXISTING DATA SOURCES, SURVEY EXPERIMENTS. AFTER DATA IS COLLECTED, IT MUST BE ORGANIZED INTO A MANAGEAB THAT IS NOT ORGANIZED IS REFERRED TO AS raw data.

Definition 6.3

A quantity that we measure from observation is called a variate or variable denoted by *V*. The distribution of a population function is the function which associates with each variate of the population function a corresponding frequency denoted by *f*.

METHODS FOR ORGANIZING RAW DATA INCLUDE THE DRAWING OF TABLES OR GRAPH QUICK OVERVIEW OF THE INFORMATION COLLECTED.

EXAMPLE 1 SUPPOSE THERE ARE 10 STUDENTS IN A GROUP WHOSE SCORES IN A MATH QUIZ WERE AS FOLLOWS:

13, 12, 14, 13, 12, 12, 13, 14, 15, 12

ORGANIZE THE DATA IN TABULAR FORM. WHAT ARE THE VARIATES? GIVE THE FREQU VARIATE.

SOLUTION: THE DATA GIVEN ABOVE IS RAW DATA.

WE MAY NOW TABULATE THE GIVEN DATA IN THE FORM GIVEN BELOW.

Score (*V*)	12	13	14	15
Number of students (*f*)	4	3	2	1

THE TABLE GIVEN ABOVE IS CALLED THE frequency distribution table. THE SCORES ARE THE VARIATE AND THE NUMBER OF STUDENTS GETTING A PARTICULAR SCORE IS THE frequency OF THE VARIATE.

Definition 6.4

A frequency distribution is a tabular or graphical representation of a data showing the frequency associated with each data value.

EXAMPLE 2 ORGANIZE THE DATA BELOW INTO A FREQUENCY DISTRIBUTION TABLE.

8,	9,	8,	7,	10,	9,	6,	4,	9,	8,
7,	8,	10,	9,	8,	6,	9,	7,	8,	8

SOLUTION: (WRITE THE VALUES IN ASCENDING ORDER.)

Value(V)	4	5	6	7	8	9	10	TOTAL
Frequency(f)	1	0	2	3	7	5	2	20

QUANTITATIVE DATA CAN ALSO BE REPRESENTED GRAPHICALLY, THROUGH A histogram

Definition 6.5

A histogram is a graphical representation of a frequency distribution in which the variate (V) is plotted on the x-axis (horizontal axis) and the frequency (f) is plotted on the y-axis (vertical axis).

WHEN DRAWING A HISTOGRAM:

I CONSTRUCT A FREQUENCY DISTRIBUTION TABLE OF THE GIVEN DATA.

II THE x-AXIS

- **A** DETERMINE A SUITABLE SCALE FOR THE HORIZONTAL AXIS TO DETERMINE THE NUMBER OF RECTANGLES NEEDED TO REPRESENT EACH VARIATE OR GROUP OF VARIATES AS
- **B** TRY NOT TO BREAK THE x-AXIS.

III THE y-AXIS

- **A** DISPLAY INFORMATION ABOUT FREQUENCY ON THE VERTICAL (y) AXIS.
- **B** DETERMINE THE LENGTH OF THE y-AXIS.

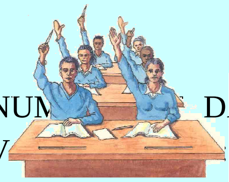
IV DRAW BARS FOR EACH VARIATE (V).

V LABEL THE HISTOGRAM WITH A TITLE, AND LABEL THE AXES.

Note: **I** THE HEIGHT OF EACH RECTANGLE IS THE FREQUENCY.

II THE WIDTH OF EACH RECTANGLE SHOULD BE THE SAME.

ACTIVITY 6.2



CONSIDER THE FOLLOWING DATA THAT SHOWS THE NUM DAYS 25 INDIVIDUALS PARTICIPATED IN SOIL AND WATER CONSERV

3	8	7	4	8
5	9	8	5	9
7	8	3	7	5
8	5	6	8	8
10	7	4	4	7

CONSTRUCT A FREQUENCY DISTRIBUTION TABLE AND A HISTOGRAM FOR THE ABOVE D

EXAMPLE 3 THE TEMPERATURE IN °C FOR THE FIRST 14 DAYS OF SEPTEMBER IN A CERTAI TOWN WERE RECORDED AS

22,	27,	19,	23,	19,	18,	27,
27,	25,	23,	26,	27,	28,	23

CONSTRUCT A FREQUENCY DISTRIBUTION TABLE AND A HISTOGRAM FO DATA.

SOLUTION: NOW CONSTRUCT THE FREQUENCY DISTRIBUTION TABLE FROM THE RAW D

Temperature (in °C) (*V*)	18	19	20	21	22	23	24	25	26	27	28
frequency (*f*)	1	2	0	0	1	3	0	1	1	4	1

USING THE ABOVE TABLE, WE DRAW A HISTOGRAM AS SHOWN BELOW.

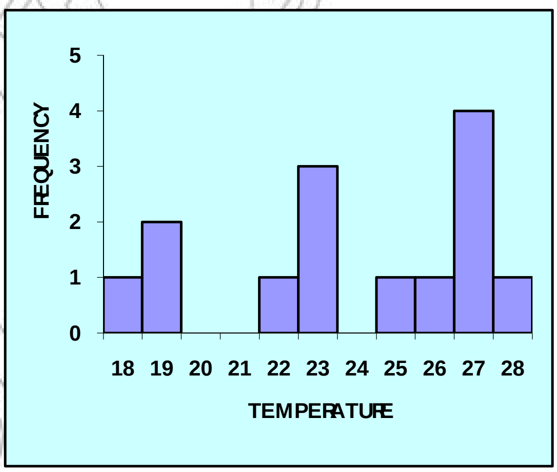


Figure 6.1

EXAMPLE 4 THE FOLLOWING HISTOGRAM SHOWS THE DAILY INCOME (IN BIRR) O
EMPLOYEES IN A FACTORY.

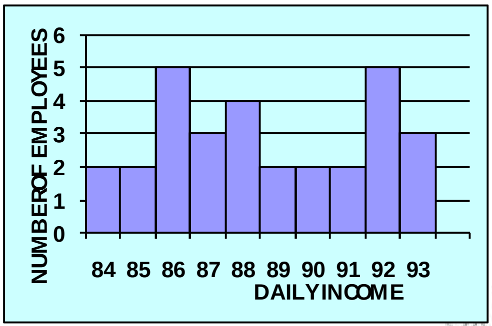


Figure 6.2

FROM THE HISTOGRAM, ANSWER THE FOLLOWING QUESTIONS:

A HOW MANY EMPLOYEES HAVE A DAILY INCOME OF BIRR 92?

B HOW MANY EMPLOYEES COLLECT A DAILY INCOME OF MORE THAN BIRR 90?

C WHAT IS THE HIGHEST FREQUENCY?

D WHAT PERCENT OF THE EMPLOYEES EARN A DAILY INCOME OF MORE THAN BI

SOLUTION:

A 5 EMPLOYEES HAVE A DAILY INCOME OF BIRR 92.

B 10 EMPLOYEES EARN A DAILY INCOME OF MORE THAN BIRR 90.

C THE HIGHEST FREQUENCY IS 5.

D PERCENTAGE

$$= \frac{\text{SUM OF THE FREQUENCIES OF EMPLOYEES EARNING MORE THAN 89}}{\text{TOTAL FREQUENCY}} \times 100\%$$

$$= \frac{2+2+5+3}{30} \times 100\% = \frac{12}{30} \times 100\% = 40\%$$

I.E., 40% OF THE EMPLOYEES EARN A DAILY INCOME OF MORE THAN BIRR 89.

Exercise 6.2

1 GIVE TWO REASONS WHY RAW DATA SHOULD BE SUMMARIZED INTO A FREQUENCY I

2 WHAT IS THE DIFFERENCE BETWEEN A FREQUENCY DISTRIBUTION TABLE AND A HIST

3 THE AGES (TO THE NEAREST YEAR) OF 40 CHILDREN IN A CERTAIN VILLAGE ARE AS FOLLOWS:

10	7	4	5	1	9	3	6	5	4
2	7	5	3	2	5	6	2	8	9
5	8	9	9	5	2	1	3	9	4
3	5	7	9	6	3	6	8	1	2

PREPARE A FREQUENCY DISTRIBUTION TABLE AND A HISTOGRAM FOR THE GIVEN I

4 COLLECT THE SCORE THE STUDENTS IN YOUR CLASS OBTAINED IN THEIR MID-SEMESTER MATHEMATICS EXAM AND

A PREPARE A FREQUENCY DISTRIBUTION TABLE.

B DRAW A HISTOGRAM.

C WHAT SCORE IS MOST FREQUENT?

D WHAT IS THE LEAST SCORE OBTAINED?

5 A SAMPLE OF 50 COUPLES MARRIED FOR 10 YEARS WERE ASKED HOW MANY CHILDREN THEY HAD. THE RESULT OF THE SURVEY IS AS FOLLOWS:

0	4	2	2	1	3	0	3	2	4
3	3	1	3	3	3	3	3	2	2
1	3	3	2	4	3	1	5	2	2
2	0	0	2	1	2	2	2	3	2
3	3	3	4	3	1	3	0	3	2

A CONSTRUCT A FREQUENCY DISTRIBUTION.

B CONSTRUCT A HISTOGRAM.

C WHAT PERCENTAGE OF COUPLES HAVE TWO CHILDREN?

D WHAT PERCENTAGE OF COUPLES HAVE AT LEAST TWO CHILDREN?

6 HERE ARE QUINTALS OF FERTILIZER DISTRIBUTED TO 50 FARMERS.

20	24	22	19	20	10	18	24	10	15
21	20	20	19	20	10	14	22	10	18
18	15	14	18	20	15	14	22	14	20
15	14	15	20	21	10	20	20	15	24
10	10	15	22	14	21	20	14	15	10

A CONSTRUCT A FREQUENCY DISTRIBUTION.

B CONSTRUCT A HISTOGRAM.

7 SUPPOSE THE FOLLOWING DATA REPRESENTS THE NUMBER OF PERSONS W COUNSELLING ON HIV/AIDS ON 40 CONSECUTIVE DAYS:

10	5	10	3	4	5	12	9	11	13
10	9	6	10	8	7	3	7	9	10
4	6	8	6	7	6	4	4	11	8
10	9	5	8	8	7	8	8	6	12

A CONSTRUCT A FREQUENCY DISTRIBUTION TABLE FROM THE DATA.

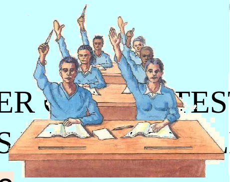
B CONSTRUCT A HISTOGRAM.

C ON WHAT PERCENT OF DAYS DID MORE THAN 10 PEOPLE TAKE COUNSELLING?

6.1.3 Measures of Location (Mean, Median and Mode(s))

QUANTITATIVE VARIABLES CONTAINED IN RAW DATA OR IN FREQUENCY TABLES SUMMARIZED BY MEANS OF A FEW NUMERICAL VALUES. A KEY ELEMENT OF THIS S CALLED THE measure of average OR measure of location. THE THREE COMMONLY USED MEASURES OF LOCATION ARE THE arithmetic mean (OR THE MEAN), the median AND THE mode(s).

ACTIVITY 6.3



1 AFTER COMPLETING A UNIT, A MATHEMATICS TEACHER … TEST MARKED OUT OF 10, AND THE SCORES OF 22 STUDENTS … LOWS:

6, 5, 8, 10, 6, 7, 3, 9, 3, 2, 9, 6, 7, 2, 6, 5, 4, 8, 6, 4,8, 3

- A DID THE GROUP DO WELL IN THE TEST?
- B PREPARE A FREQUENCY DISTRIBUTION TABLE FROM THE GIVEN DATA.
- C WHAT IS THE AVERAGE SCORE OF THE GROUP?
- D HOW MANY STUDENTS SCORE ABOVE AVERAGE?
- E FROM THE AVERAGE OBTAINED, CAN WE SAY SOMETHING ABOUT THE PERFOR THE GROUP?
- F WHAT RELATION CAN WE SEE BETWEEN THE SINGLE VALUE OBTAINED IN C AN MARKS OF THE STUDENTS? CAN THE SINGLE VALUE SUMMARIZE THE DATA?

2 RECORD THE HEIGHT AND AGE OF EACH STUDENT IN YOUR CLASS.

- A WHAT IS THE AVERAGE HEIGHT AND AGE OF THE STUDENTS?
- B WHAT IS THE MIDDLE VALUE OF HEIGHT AND AGE OF THE STUDENTS?
- C WHAT VALUE OF HEIGHT AND AGE IS MOST FREQUENT (OR HAS THE HIGHEST FREQUENCY)?

3 SUPPOSE A STUDENT SCORED THE FOLLOWING MARKS IN FIVE SUBJECTS:

85, 89, 78, 92, 91

- A WHAT IS THE AVERAGE SCORE OF THE STUDENT?
- B WHAT IS THE MIDDLE VALUE OF THE SCORE?

4 CONSIDERING THE FOLLOWING DATA

20, 21, 21, 22, 23, 23, 25, 27, 27, 27, 29, 98, 98

- A FIND THE MEAN, MEDIAN AND MODE.
- B WHICH MEASURE OF LOCATION DOES NOT GIVE A CLEAR INDICATION OF THE CENTRE OF THE DISTRIBUTION?

5 COULD YOU FIND THE ARITHMETIC MEAN OF QUALITATIVE DATA? WHAT ABOUT MODE?

1 The arithmetic mean

WHENUSEDINEVERYDAYLANGUAGETHEWORD"AVERAGE" OFTENSTANDS FORTHE ARITHMETIC MEAN

Definition 6.6

The **arithmetic mean** (or the **mean**) of a variable is the sum of all the data values, divided by the total frequency (number of observations).

If x_1, x_2, x_3,....., x_n are the n observations of a variable, then the mean, $\bar{x}$, is given by

$$\text{Mean}: \bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\text{SUM OF VALUES}}{\text{TOTAL NUMBER OF VAL}}$$

EXAMPLE 1 FINDTHE MEANOF THEFOLLOWINGDATA:

7, 21, 2, 17, 3, 13, 7, 4, 9, 7, 9

SOLUTION:

$$\bar{x} = \frac{7+21+2+17+3+13+7+4+9+7+9}{11} = \frac{99}{11} = 9$$

Note: THE MEAN OF A POPULATION FUNCTION CAN ALSO BE CALCULATED FROMITS FREQUE DISTRIBUTION. SO, IF THE VALUES x_1, x_2, x_3, …, x_n OCCUR f_1, f_2, f_3, …, f_n TIMES, RESPECTIVELY, THENTHE MEAN ($\bar{x}$) IS GIVENBY

$$\text{MEAN: } \bar{x} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_n f_n}{f_1 + f_2 + \ldots + f_n}$$

EXAMPLE 2 THEFOLLOWINGTABLESHOWS THE AGE OF 14 STUDENTS INACERTAINCLASS:

Age in years (V)	12	13	16	18
Number of students (f)	3	4	2	5

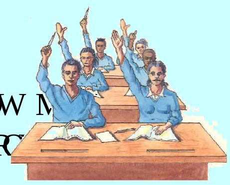
COMPUTETHE MEANAGE OF THESTUDENTS.

SOLUTION:

$$\bar{x} = \frac{12\times3+13\times4+16\times2+18\times5}{3+4+2+5} = \frac{36+52+32+90}{14} = \frac{210}{14} = 15 \text{ YEARS}$$

Properties of the mean

ACTIVITY 6.4



THERE ARE FIVE STUDENTS IN A GROUP. LEMLEM WANTS TOKNOW HOW M MONEY EACH STUDENT HAS AND ASKED ALL THE MEMBERS OF THE GR FOUNDTHE FOLLOWING AMOUNTS:

BIRR6, BIRR9, BIRR8, BIRR4 ANDBIRR3.

A WHAT IS THE MEANOF THE AMOUNT OF MONEYWITHINTHE GROUP?

B IF LEMLEM GIVES BIRR 2 TO EACH MEMBER OF THE GROUP, WHAT WILL BE THE NEW MEAN?

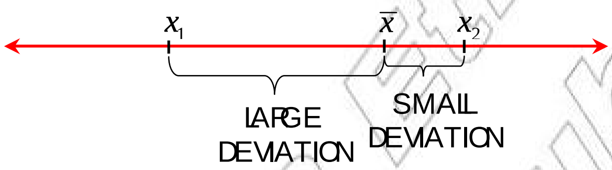
C IF THE AMOUNT OF MONEY IN THE POCKET OF EACH MEMBER IS MULTIPLIED BY 2, WHAT WILL BE THE NEW MEAN?

D IF YOU SUBTRACT THE MEAN OF THE DATA OBTAINED FROM EACH VALUE, WHAT IS THE SUM OF THE DIFFERENCES OBTAINED?

E DISCUSS, WHAT YOU OBSERVED FROM YOUR ANSWERS TO **A**, **B**, **C** AND **D**.

THE ABOVE ACTIVITY SHOULD HELP YOU TO OBSERVE DIFFERENT PROPERTIES OF THE MEAN.

THE DIFFERENCE BETWEEN A SINGLE DATA VALUE x AND THE MEAN IS CALLED THE DEVIATION FROM THE MEAN (OR SIMPLY THE DEVIATION) AND IS GIVEN BY $(x-\bar{x})$. A DATA POINT THAT IS CLOSE TO THE MEAN WILL HAVE A SMALL DEVIATION, WHEREAS DATA POINTS FAR FROM THE MEAN HAVE LARGE DEVIATIONS AS SHOWN IN THE FIGURE BELOW.



1 THE SUM OF THE DEVIATIONS OF INDIVIDUAL OBSERVATIONS FROM THE MEAN ($\bar{x}$) IS ZERO. THAT IS, LET $x_1, x_2, x_3, \ldots, x_n$ BE n OBSERVATIONS WITH MEAN $\bar{x}$. THEN THE SUM OF THE DEVIATIONS OF THE OBSERVATIONS FROM THE MEAN IS GIVEN BY

$$(x_1-\bar{x})+(x_2-\bar{x})+(x_3-\bar{x})+\ldots+(x_n-\bar{x})=0$$

Proof:-

SINCE THE MEAN OF n OBSERVATIONS $x_1, x_2, x_3, \ldots, x_n$ IS GIVEN BY $\bar{x}$,

$$\bar{x}=\frac{x_1+x_2+x_3+\ldots+x_n}{n} \text{ WHICH SHOWS } x_1+x_2+x_3+\ldots+x_n=n\bar{x}$$

NOW, $(x_1-\bar{x})+(x_2-\bar{x})+(x_3-\bar{x})+\ldots+(x_n-\bar{x})$

$$=(x_1+x_2+x_3+\ldots+x_n)-\underbrace{(\bar{x}+\bar{x}+\bar{x}\ldots+\bar{x})}_{n \text{ TIMES}}$$

$$=(x_1+x_2+x_3+\ldots+x_n)-n\bar{x}$$

$$=n\bar{x}-n\bar{x}=0 \text{ AS REQUIRED.}$$

EXAMPLE 3 LET THE AGES OF 5 CHILDREN BE 2, 3, 6, 9, 10. THEN, THE MEAN AGE

$$\bar{x}=\frac{2+3+6+9+10}{5}=\frac{30}{5}=6$$

THE SUM OF THE DEVIATIONS FROM THE MEAN IS:

(2 – 6) + (3 – 6) + (6 – 6) + (9 – 6) + (10 – 6) = –4 – 3 + 0 + 3 + 4 = 0

2 IF A CONSTANT K IS ADDED TO (OR SUBTRACTED FROM) EACH DATA VALUE, THEN THE NEW MEAN IS THE SUM (OR THE DIFFERENCE) OF THE OLD MEAN AND THE CONSTANT k

Proof:- LET $\bar{x}$ BE THE MEAN OF THE DATA VALUES AND k BE THE CONSTANT.

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

ADDING k TO EACH DATA VALUE, THE NEW MEAN IS THEN

$$\frac{(x_1+k)+(x_2+k)+(x_3+k)+\ldots+(x_n+k)}{n} =$$

$$\frac{x_1 + x_2 + x_3 + \ldots + x_n + k + k + k + \ldots + k}{n}$$

$$= \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} + \frac{nk}{n}$$

$= \bar{x} + k$ (THE OLD MEAN PLUS k).

A SIMILAR PROOF CAN BE DONE FOR THE CASE WHEN k IS SUBTRACTED FROM EACH DATA VALUE.

3 THE MEAN OF THE SUM OR DIFFERENCE OF TWO POPULATION FUNCTIONS (OF EQUAL NUMBERS OF OBSERVATIONS) IS EQUAL TO THE SUM OR DIFFERENCE OF THE MEANS OF THE POPULATION FUNCTIONS.

Proof:-

LET $\frac{x_1+x_2+x_3+\ldots+x_n}{n} = \bar{x}$ AND $\frac{y_1+y_2+y_3+\ldots+y_n}{n} = \bar{y}$

THEN THE MEAN OF THEIR SUM,

$$\text{MEAN}(x+y) = \frac{(x_1+y_1)+(x_2+y_2)+\ldots+(x_n+y_n)}{n}$$

$$= \frac{(x_1+x_2+x_3+\ldots+x_n)+(y_1+y_2+y_3+\ldots+y_n)}{n}$$

$$= \frac{(x_1+x_2+x_3+\ldots+x_n)}{n} + \frac{(y_1+y_2+y_3+\ldots+y_n)}{n}$$

$= \bar{x} + \bar{y}$ (THE SUM OF THE MEANS)

EXAMPLE 4 THE MEAN OF 2, 4, 6, 8 IS 5 AND THE MEAN OF 5, 7, 9, 7 IS 7. THEN, THE MEAN OF THE SUM 7, 11, 15, 15 IS 5+7 = 12.

4 THE MEAN OF A CONSTANT TIMES A POPULATION FUNCTION IS EQUAL TO THE CONSTANT TIMES THE MEAN OF THE POPULATION FUNCTION. THAT IS,

IF $\bar{x}$ IS THE MEAN OF THE POPULATION FUNCTION $x_1, x_2, x_3, \ldots, x_n$ AND IF k IS A CONSTANT, THEN THE MEAN OF $kx_1, kx_2, kx_3, \ldots, kx_n$ IS EQUAL TO $k\bar{x}$

Proof:-

$$\frac{kx_1 + kx_2 + kx_3 + . . . + kx_n}{n} = \frac{k(x_1 + x_2 + x_3 + . . . + x_n)}{n} = k\bar{x}$$

EXAMPLE 5 THE MEAN OF 8, 9, 6, 8, 4, IS 7. IF YOU MULTIPLY EACH OF VALUE BY 5, YO
WILL OBTAIN 40, 45, 30, 40, 20. THEN THE NEW MEAN IS $5 \times 7 = 35$

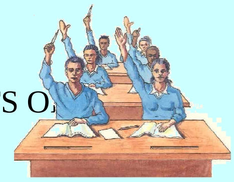
Note:

1. THE MEAN IS UNIQUE.
2. THE MEAN IS AFFECTED BY EXTREME VALUES.

2 The median

THE FOLLOWING ACTIVITY WILL HELP YOU TO REVISE WHAT YOU LEARNED IN PREVIOUS GRADE

ACTIVITY 6.5



1. FIND THE MEDIAN FOR EACH OF THE FOLLOWING SETS O
   - **A** 5, 2, 9, 7, 3 **B** 12 , 8, 10, 14, 13, 9
2. WHAT DID YOU OBSERVE ABOUT THE MIDDLE TERM WHEN THE NUMBER OF OBSERVATIONS IS ODD OR EVEN?

A SECOND MEASURE OF LOCATION OF QUANTITATIVE DATA IS THE median

Definition 6.7

The median is the value that lies in the middle of the data when it is arranged in ascending or descending order. So, half the data is below the median and half the data is above the median.

EXAMPLE 6 FIND THE MEDIAN OF EACH OF THE FOLLOWING:

A 6, 7, 9, 7, 11, 13, 15 **B** 27, 23, 36, 38, 27, 40, 45, 39

SOLUTION:

A FIRST ARRANGE THE DATA IN ASCENDING ORDER AS 6, 7, 7, 9, 11, 13, 15

THERE ARE SEVEN VALUES (AN ODD NUMBER OF VALUES) AND THE MIDDLE
THE 4TH ELEMENT OF THE LIST WHICH IS 9.

THEREFORE 9 IS THE MEDIAN OF THE DATA.

B FIRST, ARRANGE THE DATA IN ASCENDING ORDER AS 23, 27, 27, 36, 38, 39, 40, 45

THERE ARE EIGHT VALUES (AN EVEN NUMBER). THE TWO MIDDLE VALUES AR
AND 5TH ELEMENTS OF THE LIST WHICH ARE 36 AND 38. THE MEDIAN IS HALF TH
OF 36 AND 38. SO, THE MEDIAN IS $\frac{36+38}{2} = 37$.

EXAMPLE 7 FIND THE MEDIAN OF THE FOLLOWING DISTRIBUTION

V	1	2	3	4	5
f	2	3	2	4	2

SOLUTION: THERE ARE 13 DATA VALUES. SO, THE MEDIAN IS THE 7TH PIECE OF DATA, WHICH IS 3.

NOTE THAT THE MEDIAN OF A SET OF DATA WITH VALUES ARRANGED IN ASCENDING ORDER IS:

- **I** THE MIDDLE VALUE OF THE LIST IF THERE IS AN ODD NUMBER OF VALUES.
- **II** HALF OF THE SUM OF THE TWO MIDDLE VALUES IF THERE IS AN EVEN NUMBER OF VALUES.

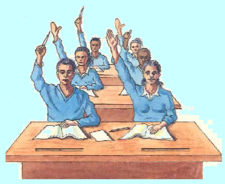
Properties of the median

1. THE MEDIAN CAN BE OBTAINED EVEN WHEN SOME OF THE DATA VALUES ARE NOT
2. IT IS NOT AFFECTED BY EXTREME VALUES.
3. IT IS UNIQUE FOR A GIVEN DATA SET.

3 The mode

THE FOLLOWING ACTIVITY SHOULD HELP YOU TO RECALL WHAT YOU HAVE LEARNT ABOUT MODE PREVIOUSLY.

ACTIVITY 6.6



1. FIND THE MODE(S) OF THE FOLLOWING DATA
   - **A** 5, 7, 8, 7, 9, 11
   - **B** M, F, M, F, F
2. CAN YOU FIND THE MEAN AND MEDIAN FOR THE ABOVE DATA?
3. DISCUSS YOUR OBSERVATION.

A THIRD MEASURE OF LOCATION IS THE **mode**. THE MODE CAN BE FOUND FOR BOTH QUANTITATIVE AND QUALITATIVE DATA.

Definition 6.8

The value of the variable which occurs most frequently in a data set is called the mode.

EXAMPLE 8 FIND THE MODE OF EACH OF THE FOLLOWING DATA SETS:

- **A** 4, 6, 12, 10, 7
- **B** 12, 10, 11, 13, 10, 14, 12, 18, 17
- **C** 9, 8, 7, 10, 6, 8

SOLUTION:

A IT HAS NO MODE BECAUSE EACH VALUE OCCURS ONLY ONCE.

B THE VALUES 10 AND 12 BOTH OCCUR TWICE, WHILE THE OTHERS OCCUR ONLY IT HAS TWO MODES AND THE DATA IS A BIMODAL.

C 8 IS THE MODE BECAUSE IT OCCURRED TWICE (MOST FREQUENTLY).

EXAMPLE 9 FIND THE MEAN, MEDIAN AND MODE OF THE FOLLOWING DISTRIBUTI TEMPERATURES IN A CERTAIN TOWN FOR ONE MONTH.

Temperature in °C(V)	20	21	23	24	26	28
Number of days(f)	2	4	5	9	3	7

SOLUTION: MEAN $\bar{x} = \dfrac{(20\times 2)+(21\times 4)+(23\times 5)+(24\times 9)+(26\times 3)+(28\times 7)}{2+4+5+9+3+7}$

$$= \frac{40+84+115+216+78+196}{30} = \frac{729}{30} = 24.3$$

THEREFORE, THE MEAN IS 24.3ºC.

THE NUMBER OF OBSERVATIONS IS AN EVEN NUMBER WHICH IS 30. SO, THE MEDIA THE SUM OF THE 15TH AND 16TH VALUES.

I.E., MEDIAN $= \dfrac{15^{\text{TH}}\text{ VALUE} + 16^{\text{TH}}\text{ VALUE}}{2} = \dfrac{24+24}{2} = 24$

THEREFORE, THE MEDIAN IS 24ºC.

THE VALUE WITH HIGHEST FREQUENCY IS THE NUMBER 24. THEREFORE, THE MODE

NOTE THAT A SET OF DATA CAN HAVE NO MODE, ONE MODE (unimodal), TWO MODES (bimodal) OR MORE THAN TWO MODES (multimodal). IF THERE IS NO OBSERVATION THAT OCCURS WITH T HIGHEST FREQUENCY, WE SAY THE DATA HAS no mode

Properties of the Mode

1. THE MODE IS NOT ALWAYS UNIQUE.
2. IT IS NOT AFFECTED BY EXTREME VALUES.
3. THE MODE CAN ALSO BE USED FOR QUALITATIVE DATA.

Exercise 6.3

1. **A** FIND THE MEAN, MODE AND MEDIAN OF THE FOLLOWING DATA.

   11, 9,14, 3, 11, 4, 10, 21, 8, 15, 350

   B WHICH MEASURE OF LOCATION IS PREFERABLE FOR THIS DATA?

2 GIVEN BELOW IS A FREQUENCY DISTRIBUTION OF VALUES V.

A FIND THE MEAN, MODE AND MEDIAN OF THE FOLLOWING DISTRIBUTION.

B HOW MANY OF THE VALUES ARE NON-NEGATIVE?

V	–2	–3	0	1	2	3
f	3	2	3	6	5	1

3 GIVEN THE NUMBERS 5, 6, 7, 10, 12, WHICH NUMBER MUST BE REMOVED IN ORDER T MAKE THE MEAN OF THE RESULTING VALUES 7.5?

4 GIVEN THE NUMBERS 10, 12, 19, 15, 8, WHAT NUMBER COULD BE INCLUDED SO THAT THE MEDIAN IS 11? (EXPLAIN)

5 GIVEN 3, 4, X, 5, Y, 12. FIND THE VALUES OF x AND y, IF THE MODE OF THE DATA IS 3 AND THE MEAN IS 6.

6 IF THE MEAN OF a, b, c, d IS k, THEN WHAT IS THE MEAN OF

A $a + b,\ 2b,\ c + b,\ d + b$? **B** $ab,\ b^2,\ cb,\ db$?

7 CALCULATE THE MEAN, MEDIAN AND MODE OF THE FOLLOWING DATA;

Value	10	15	20	25	30	35	40
Frequency	15	10	50	4	10	8	3

8 IN A SURVEY OF THE NUMBER OF OCCUPANTS OF CARS, THE FOLLOWING DATA RES

Number of occupants	1	2	3	4
Number of cars	7	11	7	x

A IF THE MEAN NUMBER OF OCCUPANTS IS $2\frac{1}{3}$, FIND x.

B IF THE MODE IS 2, FIND THE LARGEST POSSIBLE VALUE OF x.

C IF THE MEDIAN IS 2, FIND THE LARGEST POSSIBLE VALUE OF x

9 A RESEARCHER TABULATED THE NUMBER OF CASES HEARD BY 8 JUDGES ON A GIVI COURT AND FOUND THE FOLLOWING DATA:

Judges	1	2	3	4	5	6	7	8
Count of cases	6	3	1	2	0	5	5	4

A FIND THE MEAN, MEDIAN AND MODE.

B THE RESEARCHER REPORTED THAT OVER HALF OF THE JUDGES HEARD ABOVE WHAT DOES THE RESEARCHER MEAN BY THE "AVERAGE"?

10 THE FOLLOWING RAW DATA REPRESENTS THE NUMBER OF HIV/AIDS PATIENTS WAI
COUNSELLING AT 8:00 AM ON 40 CONSECUTIVE SATURDAYS AT A CERTAIN HOSPITA

11	6	5	8	11	6	3	7	4	6
5	4	13	14	9	11	13	8	10	9
10	9	6	5	10	7	8	7	8	3
8	7	8	9	6	10	11	8	8	4

A DRAW A FREQUENCY DISTRIBUTION TABLE.

B CALCULATE THE MEAN, MEDIAN AND MODAL NUMBER OF HIV/AIDS PATIENTS

C DRAW A HISTOGRAM.

11 IN A MATHEMATICS TEST THE SCORES FOR BOYS WERE 6, 7, 8, 7, 5 AND THE SCORES FOR GIRLS WERE 6, 3, 9, 8, 2, 2, 5, 7,3

A FIND THE MEAN SCORE FOR THE BOYS.

B FIND THE MEAN SCORE FOR THE GIRLS.

C FIND THE MEAN SCORE FOR BOTH THE BOYS AND GIRLS.

D WHAT DO YOU CONCLUDE?

12 THE MODE OF SOME DATA IS 20. IF EACH VALUE IN THE DATA IS INCREASED BY 2, W
BE THE MODE OF THE NEW DATA?

13 FIND THE MEAN, MEDIAN AND MODE OF THE DATA REPRESENTED BY THE HISTOGR

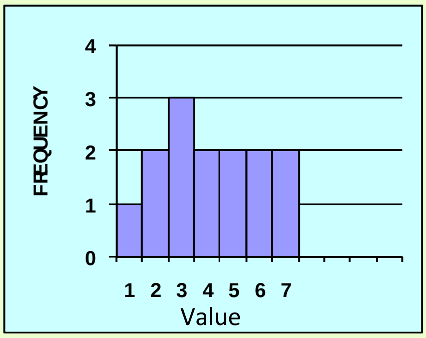


Figure 6.3

14 AN AGRICULTURAL DEVELOPMENT STATION SELLS SEEDLINGS OF PLANT THROUGH
CLAIMS THAT THE AVERAGE HEIGHT OF THE PLANTS AFTER ONE YEAR'S GROWTH
A SAMPLE OF 24 OF THE PLANTS WERE MEASURED AFTER ONE YEAR WITH THE FOL
RESULTS (IN CM).

6	7	7	9	34	56	85	89	89	90	90	91
91	92	93	93	94	95	95	96	97	97	99	93

A FIND THE MEAN AND THE MEDIAN HEIGHT OF THE SAMPLE.

B IS THE STATION'S CLAIM ABOUT AVERAGE HEIGHT JUSTIFIED?

15 IN ORDER TO RECEIVE A GRADE OF A IN HER MATHEMATICS EXAM, ABEBA NEEDS A MEAN
SCORE OF 90 AND ABOVE ON 4 TESTS. SO FAR ABEBA HAD SCORED 80, 91 AND 93 ON
WHAT IS THE LOWEST SCORE THAT SHE MUST GET IN HER LAST TEST IN ORDER
GRADE OF A?

6.1.4 Measures of Dispersion for Ungrouped Data

WHEN COMPARING SETS OF DATA, IT IS USEFUL TO HAVE A WAY OF MEASURING TH
SPREAD OF THE DATA.

Group Work 6.2

CONSIDER THE FOLLOWING THREE SETS OF DATA.

Group	Values							Total	Mean	Mode	Median
A	7	7	7	7	7	7	7				
B	4	5	6	7	7	9	1				
C	1	7	12	7	2	19	1				

A COMPLETE THE TABLE BY FINDING THE SUM OF EACH GROUP AND THE MEAN, AND MODE.

B ARE THE MEANS EQUAL? ARE THE MODES EQUAL? ARE THE MEDIANS THE SAM

C COMPARE THE VARIATION OF EACH GROUP?

- I WHICH GROUP SHOWS MOST VARIATION?
- II WHICH GROUP SHOWS NO VARIATION?
- III WHICH GROUP SHOWS SLIGHT VARIATION?

D COMPARE THE DIFFERENCE BETWEEN THE MEAN AND EACH OBSERVED VALU GROUP A, B AND C.

- I IN WHICH GROUP IS THE MEAN CLOSEST TO EACH VALUE?
- II IN WHICH GROUP IS THE DIFFERENCE BETWEEN THE MEAN AND EACH DA VALUE THE LARGEST?

E CALCULATE THE RANGE FOR EACH GROUP.

Dispersion OR Variation IS THE SCATTER (OR SPREAD) OF DATA VALUES FROM A MEASU CENTRAL TENDENCY.

THERE ARE SEVERAL MEASURES OF DISPERSION THAT CAN BE CALCULATED FOR A SET SECTION, WE WILL CONSIDER ONLY THREE OF THEM, NAMELY, range, variance, AND THE standard deviation.

1 Range

THE SIMPLEST AND THE MOST CRUDE MEASURE OF DISPERSION OF QUANTITATIVE DATA

Definition 6.9

The range R of a set of numerical data is the difference between the highest and the lowest values. i.e.,

Range = Highest value – Lowest value

EXAMPLE 1 THE AGES OF SIX STUDENTS ARE 24, 20, 18, 13, 16, 15 YEARS, RESPECTIVELY. WHAT IS THE RANGE?

SOLUTION: RANGE = HIGHEST VALUE – LOWEST VALUE = 24 – 13 = 11 YEARS.

EXAMPLE 2 FIND THE RANGE OF THE DISTRIBUTION GIVEN IN THE TABLE BELOW.

V	2	8	9	13	15	18
f	3	4	2	1	5	4

SOLUTION: THE MAXIMUM VALUE IS 18 AND THE MINIMUM VALUE IS 2.

RANGE = MAXIMUM VALUE – MINIMUM VALUE = 18 – 2 = 16

2 Variance (^{2})

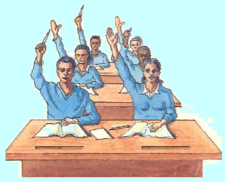
Definition 6.10

Variance, denoted by (^{2}), is defined as the mean of the squared deviations of each value from the arithmetic mean.

3 Standard deviation ( )

THE FOLLOWING ACTIVITY WILL HELP YOU TO LEARN THE STEPS USED TO FIND VARIANCE AND DEVIATION.

ACTIVITY 6.7



CONSIDER THE FOLLOWING DATA SET:

2, 3, 10, 6, 9

A FIND THE MEAN $\bar{x}$

B FIND THE DEVIATION OF EACH DATA VALUE FROM THE MEAN ($x - \bar{x}$).

C SQUARE EACH OF THE DEVIATIONS ($(x - \bar{x})^2$)

D FIND THE MEAN OF THESE SQUARED DEVIATIONS AND ITS PRINCIPAL SQUARE ROO

THE standard deviation IS THE MOST VALUABLE AND WIDELY USED MEASURE OF DISPERSI

Definition 6.11

Standard deviation, denoted by , is defined as the positive square root of the mean of the squared deviations of each value from the arithmetic mean.

The actual method of calculating σ can be summarized in the following steps:

Step 1 FIND THE ARITHMETIC MEAN $\bar{x}$ OF THE DISTRIBUTION.

Step 2 FIND THE DEVIATION OF EACH DATA VALUE FROM THE MEAN $(x-\bar{x})$.

Step 3 SQUARE EACH OF THESE DEVIATIONS, $(x-\bar{x})^2$.

Step 4 FIND THE MEAN OF THESE SQUARED DEVIATIONS. THIS VALUE IS CALLED variance AND IS DENOTED BY σ^2.

Step 5 TAKE THE PRINCIPAL SQUARE ROOT OF σ^2, I.E.

STANDARD DEVIATION = $\sqrt{\text{VARIANCE}}$

EXAMPLE 3 FIND THE VARIANCE σ^2 AND THE STANDARD DEVIATION σ FOR THE FOLLOWING DATA:

3, 5, 8, 11, 13

SOLUTION:

x	$(x-\bar{x})$	$(x-\bar{x})^2$.
3	−5	25
5	−3	9
8	0	0
11	3	9
13	5	25
TOTAL 40		68

VARIANCE $(\sigma^2) = \frac{68}{5} = 13.6$

Standard deviation (σ) = $\sqrt{\sigma^2} = \sqrt{13.6} \approx 3.7$

EXAMPLE 4 FIND THE VARIANCE AND STANDARD DEVIATION OF THE POPULATION WHOSE DISTRIBUTION IS GIVEN IN THE FOLLOWING TABLE.

V	2	3	5	6	8
f	3	4	4	5	4

SOLUTION: FIRST, THE MEAN HAS TO BE CALCULATED.

$$\bar{x} = \frac{3\times2+4\times3+4\times5+5\times6+4\times8}{3+4+4+5+4} = \frac{100}{20} = 5$$

x	f	xf	$x-\overline{x}$	$(x-\overline{x})^2$	$f(x-\overline{x})^2$
2	3	6	-3	9	27
3	4	12	-2	4	16
5	4	20	0	0	0
6	5	30	1	1	5
8	4	32	3	9	36
Total	20	100	0		84

VARIANCE $(\sigma^2) = \frac{84}{20} = 4.2$

STANDARD DEVIATION $(\sigma) = \sqrt{\quad} \approx 4$

THEREFORE, THE POPULATION VARIANCE AND STANDARD DEVIATION ARE 4.2 AND 2.05

Properties of variance and standard deviation

Group Work 6.3

CONSIDER THE FOLLOWING DATA WHICH SHOWS THE SUGAR IN KILOGRAMS SOLD BY A SMALL SHOP FOR FIVE DAYS.

6, 4, 8, 9, 3

I FIND THE MEAN.

II FIND THE VARIANCE AND STANDARD DEVIATION.

III IN THE NEXT FIVE DAYS, IF THE DAILY SALES IS EXACTLY TWICE.

- A FIND THE MEAN OF SALES FOR THE NEXT FIVE DAYS.
- B FIND THE VARIANCE AND STANDARD DEVIATION FOR THE NEXT FIVE DAYS.
- C COMPARE YOUR ANSWERS ABOVE WITH THOSE OBTAINED IN A AND B ABOVE.
- D DISCUSS THE COMPARISON YOU DID ABOVE.

IV IF THE DAILY SALES GIVEN FOR THE FIRST FIVE DAYS ARE DOUBLED I.E. IF THE DAILY SALES WERE 12, 8, 16, 18 AND 6,

- A FIND THE MEAN, VARIANCE AND STANDARD DEVIATION.
- B COMPARE THE ABOVE RESULT WITH THOSE OBTAINED IN I AND II AND DISCUSS THE RESULTS.

THE ABOVE GROUP WORK WILL HELP YOU TO OBSERVE THE FOLLOWING PROPERTIES.

1 IF A CONSTANT c IS ADDED TO EACH VALUE OF A POPULATION, THEN THE NEW VARIANCE IS THE SAME AS THE OLD VARIANCE. THE NEW STANDARD DEVIATION IS THE SAME AS THE OLD STANDARD DEVIATION.

Proof:-

LET $x_1, x_2, x_3, \ldots, x_n$ BE n OBSERVATIONS WITH MEAN μ AND VARIANCE σ^2.

ADDING c: $x_1 + c,\ x_2 + c,\ x_3 + c,\ \ldots,\ x_n + c$. THEN THE NEW MEAN IS $\bar{x} + c$

$$\text{NEW VARIANCE} = \frac{\left[(x_1 + c) - (\bar{x} + c)\right]^2 + \left[(x_2 + c) - (\bar{x} + c)\right]^2 + \ldots + \left[(x_n + c) - (\bar{x} + c)\right]^2}{n}$$

$$= \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n} = \sigma^2. \text{ (The original variance)}$$

AND THE NEW STANDARD DEVIATION IS $\sqrt{\text{VARIANCE}} = \sqrt{\sigma^2} = \sigma$ (*the original standard deviation*)

EXAMPLE 5 GIVEN 1, 2, 6, 3

A FIND THE VARIANCE **B** FIND THE STANDARD DEVIATION.

C ADD 2 TO EACH VALUE AND FIND THE VARIANCE AND STANDARD DEVIATION OF THE RESULTING NUMBERS

SOLUTION: $\bar{x} = \dfrac{1+2+6+3}{4} = 3$

$x - \bar{x}: -2, -1, 3, 0$ AND $(x - \bar{x})^2: 4, 1, 9, 0$

A $\sigma^2 = \dfrac{4+1+9+0}{4} = \dfrac{14}{4} = 3.5$

B $\sigma = \sqrt{3.5} \approx 1.87$

C ADDING 2: 3, 4, 8, 5

NEW MEAN: $\bar{x} = \dfrac{3+4+8+5}{4} = \dfrac{20}{4} = 5 = 3 + 2$

$x - \bar{x}: -2, -1, 3, 0$ AND $(x - \bar{x})^2: 4, 1, 9, 0$

NEW: $\sigma^2 = \dfrac{4+1+9+0}{4} = \dfrac{14}{4} = 3.5$

NEW $\sigma = \sqrt{3.5} \approx 1.87$

THEREFORE, THE OLD VARIANCE = THE NEW VARIANCE

THE OLD STANDARD DEVIATION = THE NEW STANDARD DEVIATION.

2 IF EACH VALUE OF A POPULATION FUNCTION IS MULTIPLIED BY A CONSTANT, THEN

I THE NEW VARIANCE IS k^2 TIMES THE OLD VARIANCE

II THE NEW STANDARD DEVIATION IS $|k|$ TIMES THE OLD STANDARD DEVIATION.

Proof:-

CONSIDER $x_1, x_2, \ldots, x_n$ WHOSE MEAN IS $\bar{x}$ AND VARIANCE IS s^2.

MULTIPLYING EACH DATA VALUE BY c GIVES US A NEW MEAN OF $c\bar{x}$.

THEN, NEW VARIANCE = $\dfrac{(cx_1 - c\bar{x})^2 + (cx_2 - c\bar{x})^2 + (cx_3 - c\bar{x})^2 + \ldots + (cx_n - c\bar{x})^2}{n}$

$$= \frac{c^2[(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + (x_3-\bar{x})^2 + \ldots + (x_n-\bar{x})^2]}{n}$$

$= c^2 \times$ THE OLD VARIANCE $= c^2 s^2$

THEREFORE, NEW STANDARD DEVIATION = $\sqrt{c^2 s^2} = |c|s$

Exercise 6.4

1 FIND THE RANGE, VARIANCE AND STANDARD DEVIATION OF THE FOLLOWING DATA.

4, 2, 3, 3, 2, 1, 4,3, 2, 6

2 FIND THE RANGE, VARIANCE AND STANDARD DEVIATION OF THE DISTRIBUTION IN THE TABLE BELOW.

V	–1	–2	0	1	2
f	2	1	3	3	1

3 FIND THE RANGE, VARIANCE AND STANDARD DEVIATION OF THE DATA FROM THE HISTOGRAM IN THE FIGURE BELOW.

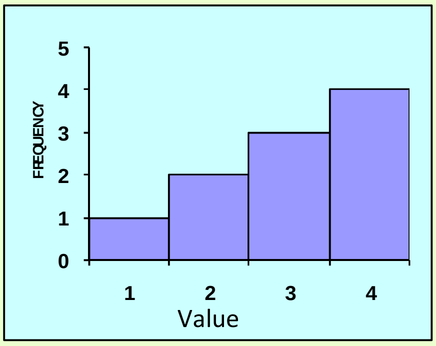


Figure 6.4

4 WHAT IS THE VALUE OF y, IF THE STANDARD DEVIATION OF 8, 8, 8, 8, y, 8 IS 0?

5 IF THE VARIANCE OF a, b, c, d IS k, THEN WHAT IS

A THE VARIANCE OF $a + c, b + c, 2c, d + c$?

B THE STANDARD DEVIATION OF $a + c, b + c, 2c, d + c$?

C THE VARIANCE OF ac, bc, c^2, dc?

D THE STANDARD DEVIATION OF ac, bc, c^2, dc?

6 IF A POPULATION FUNCTION x HAS MEAN $M(x) = 2$ AND $M(x^2) = 8$, FIND ITS STANDARD DEVIATION.

6.2 PROBABILITY

"The true logic of this world is the calculus of probabilities". James Clerk Maxwell

HISTORICAL NOTE:

The first inquiry into the science of Probability was made by Girolamo Cardano (1501-1576), an Italian physician and mathematician. Cardano predicted the date of his own death. Since he was healthy at the end of the day, he poisoned himself to make his prediction come true!

IN YOUR GRADE 8 LESSONS, YOU HAVE DISCUSSED THE WORD PROBABILITY. YOU OFTEN USE IT. "*The probability of winning a game is low*", OR "*there is a high probability that it will rain today*", ETC. IN THESE TWO SENTENCES, THE WORD PROBABILITY DESCRIBES ESTIMATES OF THE POSSIBILITIES.

PROBABILITY IS A NUMERICAL VALUE THAT DESCRIBES THE LIKELIHOOD OF THE OC EVENT IN AN EXPERIMENT.

THE FOLLOWING GROUP WORK WILL HELP YOU RECALL WHAT YOU HAVE LEARNED ON GRADE 8.

Group Work 6.4

ABEL THROWS A FAIR DIE ONCE. BASED ON THIS EXPERIMENT, DISCUSS THE FOLLOWING:

1. IS IT POSSIBLE TO PREDICT THE NUMBER THAT SH... UPPER FACE OF THE DIE? WHY?
2. LIST THE SET OF ALL POSSIBLE OUTCOMES.
3. WRITE AN EXAMPLE OF AN EVENT FROM THE EXPERIMENT.
4. WHAT CAN YOU SAY ABOUT THE FOLLOWING EVENTS?
   - I THE NUMBER ON THE UPPER FACE OF THE DIE IS SEVEN
   - II THE NUMBER ON THE UPPER FACE OF THE DIE IS AN INTEGER.
     - A WHICH OF THE ABOVE EVENTS IS CERTAIN?
     - B WHICH OF THE ABOVE EVENTS IS IMPOSSIBLE?
5. DETERMINE THE PROBABILITIES OF THE FOLLOWING EVENTS.
   - A THE NUMBER ON THE UPPER FACE OF THE DIE IS 2.
   - B THE NUMBER ON THE UPPER FACE OF THE DIE IS 7.
   - C THE NUMBER ON THE UPPER FACE OF THE DIE IS LESS THAN 7
6. DISCUSS THE FOLLOWING TERMS.
   - A EXPERIMENT
   - B POSSIBILITY SET
   - C EVENT
   - D IMPOSSIBLE EVENT
   - E CERTAIN EVENT

Definition 6.12

An experiment is a trial by which an observation is obtained but whose outcome cannot be predicted in advance.

Experimental probability

PROBABILITY DETERMINED USING DATA COLLECTED FROM EXPERIMENTS IS CALLED EXPERIMENTAL PROBABILITY.

EXAMPLE 1 THE NUMBERS 1 TO 20 ARE EACH WRITTEN ON INDIVIDUAL CARDS. ONE CARD IS CHOSEN AT RANDOM.

A LIST THE SET OF ALL POSSIBLE OUTCOMES.

B LIST THE ELEMENTS OF THE FOLLOWING EVENTS:

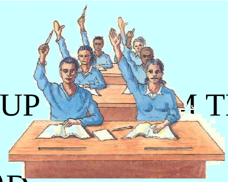
- **I** THE NUMBER IS LESS THAN 5.
- **II** THE NUMBER IS GREATER THAN 15.
- **III** THE NUMBER IS GREATER THAN 21.
- **IV** THE NUMBER IS DIVISIBLE BY 5.
- **V** THE NUMBER IS A PRIME.

SOLUTION:

A $S = \{1, 2, 3, \ldots, 19, 20\}$

B
- **I** $\{1, 2, 3, 4\}$
- **II** $\{16, 17, 18, 19, 20\}$
- **III** $\{\ \}$ OR $\varnothing$ SINCE NO CARD HAS A NUMBER GREATER THAN 20.
- **IV** $\{5, 10, 15, 20\}$
- **V** $\{2, 3, 5, 7, 11, 13, 17, 19\}$

ACTIVITY 6.8



ARRANGE YOURSELVES INTO GROUPS OF 5. LET EACH GROUP [illegible] THE FOLLOWING ACTIVITIES.

1 TAKE A COIN, TOSS IT 5 TIMES, 10 TIMES AND 15 TIMES, RECORD YOUR OBSERVATIONS IN THE FOLLOWING TABLE.

	Number of tosses			Total
Number of times a coin is tossed	5	10	15	
Number of times the coin shows up Heads				
Number of times the coin shows Tails				

WHAT PROPORTION OF THE NUMBER OF TOSSES SHOWS HEADS? A TAILS? WHAT IS THE PROBABILITY THAT THE OUTCOME IS HEAD?

2 THROW A DIE 20 TIMES. RECORD THE OBSERVATION AND COMPLETE THE FOLLOWING TABLE.

Number on the upper face of the die	1	2	3	4	5	6
Number of times it shows up						

A FIND THE NUMBER OF TIMES 3 IS ON THE UPPER FACE OF THE DIE.

B FIND THE NUMBER OF TIMES 6 IS ON THE UPPER FACE OF THE DIE.

C FIND THE NUMBER OF TIMES 7 IS ON THE UPPER FACE OF THE DIE.

D WRITE THE PROPORTION OF EACH NUMBER.

E WHAT IS THE PROBABILITY THAT THE NUMBER THAT SHOWS ON THE UPPER FACE OF THE DIE IS 4?

SUPPOSE WE TOSS A COIN 100 TIMES AND GET A HEAD 45 TIMES, AND A TAIL 55 TIMES. T

WOULD SAY THAT IN A SINGLE TOSS OF A COIN, THE PROBABILITY OF GETTING A HEA $\frac{45}{100}=\frac{9}{20}$

AGAIN SUPPOSE WE TOSS A COIN 500 TIMES AND GET A HEAD 260 TIMES, AND A TAIL 24 THEN WE SAY THAT IN A SINGLE TOSS OF A COIN, THE PROBABILITY OF GETTING

$\frac{260}{500}=\frac{13}{25}$. SO FROM VARIOUS EXPERIMENTS, WE MIGHT OBTAIN DIFFERENT PROBABILITIES FOR THE SAME EVENT. HOWEVER, IF AN EXPERIMENT IS REPEATED A SUFFICIENTLY LARGE NUM THE RELATIVE FREQUENCY OF AN OUTCOME WILL TEND TO BE CLOSE TO THE THEORE OF THAT OUTCOME.

Definition 6.13

The possibility set (or sample space) for an experiment is the set of all possible outcomes of the experiment.

EXAMPLE 2

A GIVE THE SAMPLE SPACE FOR TOSSING A COIN.

B WHAT IS THE SAMPLE SPACE FOR THROWING A DIE?

SOLUTION:

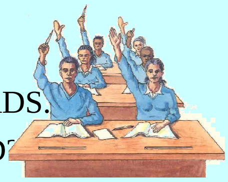
A WHEN WE TOSS A COIN THERE ARE ONLY TWO POSSIBLE OUTCOMES: HEADS (H) OR TAILS (T). SO S = {H,T}.

B WHEN WE THROW A DIE THE SCORE CAN BE ANY OF THE NUMBERS 1, 2, 3, 4, 5, 6, SO S = {1,2,3,4,5,6}.

Definition 6.14

An event is a subset of the possibility set (sample space).

ACTIVITY 6.9



SUPPOSE WE TOSS A COIN 1000 TIMES AND OBTAIN 495 HEADS.

A HOW MANY TIMES WAS THE EXPERIMENT PERFORMED?

B IF OUR EVENT IS HEADS, HOW MANY TIMES DID THIS EVENT OCCUR?

C WHAT IS THE PROBABILITY OF HEADS BASED ON THE RESULT OF THIS EXPERIMENT?

Definition 6.15

If an experiment has n equally likely outcomes and if m of these represent a particular event, then the probability of this event occurring is $\frac{m}{n}$.

EXAMPLE 3 IN AN EXPERIMENT OF SELECTING STUDENTS AT RANDOM, A TEACHER FOUND THE FOLLOWING RESULT AFTER 50 TRIALS.

Student	BOY	GIRL	TOTAL
Number	20	30	50

WHAT IS THE PROBABILITY THAT A RANDOMLY SELECTED STUDENT IS A GI

SOLUTION THE PROBABILITY THAT A RANDOMLY SELECTED STUDENT WILL BE A GIRL IS THE RATIO OF THE NUMBER OF GIRLS TO THE TOTAL NUMBER OF TRIALS.

$$P(\text{A GIRL WILL BE SELECTED}) = \frac{30}{50} = \frac{3}{5}$$

IN DECIMAL FORM THE PROBABILITY IS 0.6.

A TREE DIAGRAM IS ONE WAY OF SHOWING THE POSSIBLE OUTCOMES OF A EXPERIMENT.

EXAMPLE 4 IN AN EXPERIMENT OF TOSSING TWO COINS,

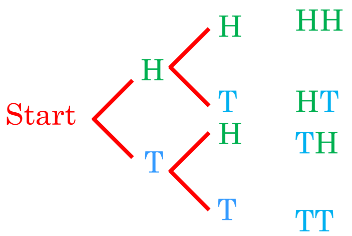
A WHAT ARE THE POSSIBLE OUTCOMES?

B HOW MANY DIFFERENT POSSIBLE OUTCOMES ARE THERE?

C WHAT IS THE PROBABILITY OF THE COINS LANDING WITH

I TWO HEADS? **II** TWO TAILS? **III** ONE HEAD?

SOLUTION: USING A TREE DIAGRAM, WE GET



A THE SET OF POSSIBLE OUTCOMES IS S = {HH, HT, TH, TT}.

B THERE ARE 4 POSSIBLE OUTCOMES.

C **I** THE EVENT TWO HEADS = {HH} HAS ONE MEMBER, SO

$$\text{P (TWO HEADS)} = \frac{1}{4}$$

II $\text{P (TWO TAILS)} = \frac{1}{4}$

III THE EVENT ONE HEAD = {HT, TH} HAS TWO MEMBERS, SO

$$\text{P (ONE HEAD)} = \frac{2}{4} = \frac{1}{2}$$

IN REAL SITUATIONS, IT MIGHT NOT ALWAYS BE POSSIBLE TO PERFORM AN EXPE CALCULATE PROBABILITY. IN SUCH SITUATIONS, WE NEED TO DEVELOP ANOTHER AP THE PROBABILITY OF AN EVENT.

IN THE NEXT SECTION, YOU WILL DISCUSS A THEORETICAL APPROACH OF FINDING PROB

Theoretical probability of an event

Definition 6.16

THE THEORETICAL PROBABILITY OF AN EVENT E, WRITTEN AS P(E), IS DEFINED AS FOLLOWS:

$$P(E) = \frac{\text{NUMBER OF OUTCOMES FAVOURABLE TO THE EVENT } E}{\text{TOTAL NUMBER OF POSSIBLE OUTCOMES } S}$$

YOU CAN WRITE THE PROBABILITY OF AN EVENT AS A FRACTION, A DECIMAL, OR A PER

EXAMPLE 5 A FAIR COIN IS TOSSED ONCE. WHAT IS THE PROBABILITY OF GETTING A HEAD?

SOLUTION:

S = { H, T}

E = {H}

$$\text{P (HEAD)} = \frac{n(E)}{n(S)} = \frac{1}{2} = 0.5$$

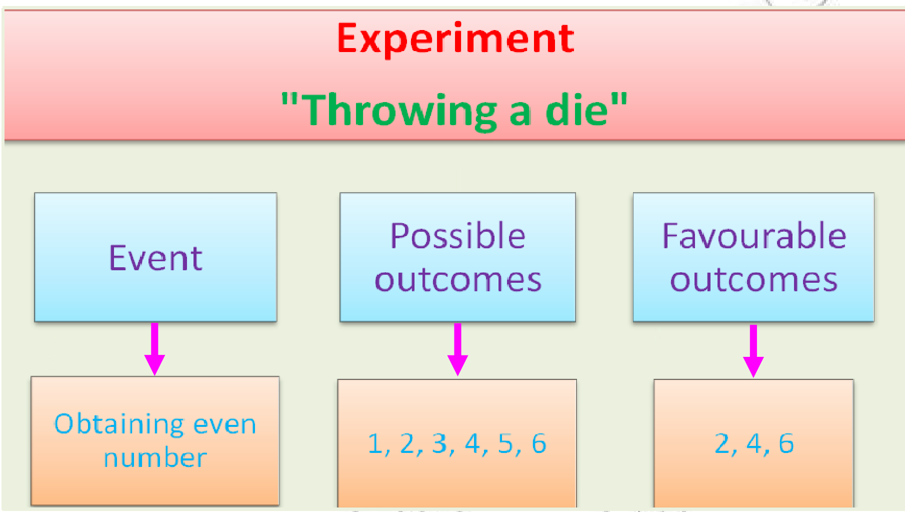
EXAMPLE 6 IF WE THROW A DIE ONCE, WHAT IS THE PROBABILITY THAT AN EVEN NUMBER WILL SHOW ON THE UPPER FACE OF THE DIE?

SOLUTION:

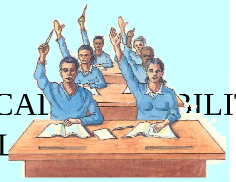
S = {1, 2, 3,4, 5, 6}

E = {2, 4, 6}

$$P(\text{EVEN}) = \frac{n(E)}{n(S)} = \frac{3}{6} = \frac{1}{2}.$$



ACTIVITY 6.10



WE ARE GOING TO INVESTIGATE WHETHER THE THEORETICAL PROBABILITY OF A COIN LANDING ON HEADS IS BACKED UP BY EXPERIMENTAL

A TOSS A COIN 10 TIMES, 20 TIMES, 30 TIMES . . . .

B KEEP A RECORD OF YOUR RESULTS,

Number of throws	Number of heads
10	
20	
30	

C FOR EACH ROW IN THE TABLE, WHAT PROPORTION OF THE THROWS LANDED AS HEADS?

HOW DO YOUR ANSWERS COMPARE WITH $P(\text{HEADS}) = \frac{1}{2}$ (THE THEORETICAL PROBABILITY)

Definition 6.17

Let S be the possibility set of an experiment and each element of S be equally likely to occur. Then the probability of the event E occurring, denoted by $P(E)$, is defined as:

$$P(E) = \frac{\text{NUMBER OF ELEMENTS IN } E}{\text{NUMBER OF ELEMENTS IN } S} = \frac{n(E)}{n(S)}$$

EXAMPLE 7 A DIE IS THROWN ONCE. WHAT IS THE PROBABILITY THAT THE NUMBER APPEARING WILL BE

A 3? **B** A NUMBER LESS THAN 5?

SOLUTION: THERE ARE SIX POSSIBLE OUTCOMES: {1, 2, 3, 4, 5, 6}. HENCE, $n(S) = 6$.

A ONLY ONE OF THESE OUTCOMES IS 3. HENCE THE PROBABILITY THAT 3 WILL BE ON THE UPPER FACE OF THE DIE IS $\frac{1}{6}$.

B {1, 2, 3, 4} IS THE REQUIRED SET, WHICH HAS FOUR ELEMENTS. HENCE THE PROBABILITY IS $\frac{4}{6} = \frac{2}{3}$.

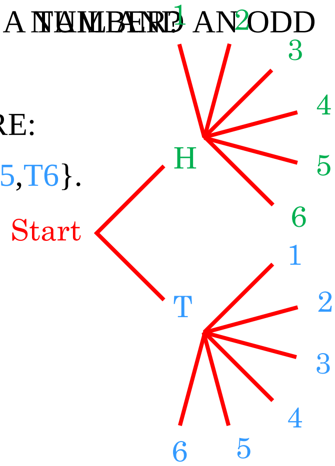
EXAMPLE 8 A DIE AND A COIN ARE TOSSED TOGETHER.

A SKETCH A TREE DIAGRAM SHOWING THE OUTCOMES OF THIS EXPERIMENT.

B WHAT IS THE PROBABILITY OF GETTING A HEAD AND AN EVEN NUMBER?

C WHAT IS THE PROBABILITY OF GETTING A TAIL AND AN ODD NUMBER?

SOLUTION:



A THE OUTCOMES OF THIS EXPERIMENT ARE:

S = {H1, H2, H3, H4, H5, H6, T1, T2, T3, T4, T5, T6}.

SO, $n(S) = 12$

B E_1 = {H2, H4, H6}. HENCE $P(E_1) = \frac{3}{12} = \frac{1}{4}$.

C E_2 = {T1, T3, T5}. HENCE $P(E_2) = \frac{3}{12} = \frac{1}{4}$.

EXAMPLE 9 USE A TREE DIAGRAM TO LIST THE SAMPLE SPACE (POSSIBILITY SET) SHOWING THE POSSIBLE ARRANGEMENT OF BOYS AND GIRLS IN A FAMILY WITH EXACTLY THREE CHILDREN.

A WHAT IS THE PROBABILITY THAT ALL THREE CHILDREN ARE BOYS?

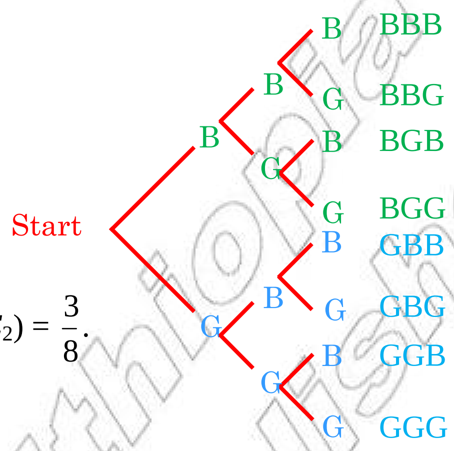
B WHAT IS THE PROBABILITY THAT TWO CHILDREN ARE BOYS AND ONE IS A GIRL?

c WHAT IS THE PROBABILITY THAT NONE OF THE CHILDREN IS A BOY?

d WHAT IS THE PROBABILITY THAT AT LEAST ONE OF THE CHILDREN IS A GIRL?

e WHAT IS THE PROBABILITY THAT ALL THREE CHILDREN ARE OF THE SAME SEX?

SOLUTION:



S = {BBB, BBG, BGB, BGG, GBB, GBG, GGB, GGG}.

THUS, $n(S) = 8$.

a E_1 = {BBB}. HENCE $P(E_1) = \frac{1}{8}$.

b E_2 = {BBG, BGB, GBB}. HENCE $P(E_2) = \frac{3}{8}$.

c E_3 = {GGG}. HENCE $P(E_3) = \frac{1}{8}$.

d E_4 = {BBG, BGB, BGG, GBB, GBG, GGB, GGG}.

HENCE $P(E_4) = \frac{7}{8}$.

(ALTERNATIVELY, HAVING AT LEAST ONE GIRL IS ALL OUTCOMES EXCEPT BBB

I.E., 8 – 1 = 7 OUTCOMES, GIVING THE SAME RESULT, $P(E_4) = \frac{7}{8}$.)

e E_5 = {BBB, GGG}. HENCE, $P(E_5) = \frac{2}{8} = \frac{1}{4}$.

Note: FOR ANY EVENT E,

✓ $0 \leq P(E) \leq 1$.

✓ IF $P(E) = 1$ THEN THE EVENT IS certain

✓ IF $P(E) = 0$, THEN THE EVENT IS impossible

FOR EXAMPLE, IF A BALL IS TAKEN FROM A BOX CONTAINING ONLY RED BALLS, THEN

P (BALL IS RED) = 1 AND P (BALL IS BLACK) = 0.

Exercise 6.5

1 TWO DICE ARE SIMULTANEOUSLY THROWN ONCE. WHICH OF THE FOLLOWING EVENTS CAN OCCUR.

a A = THE SAME NUMBER IS SHOWN ON EACH DIE.

b B = THE SUM OF THE NUMBERS IS 13.

c C = THE PRODUCT OF THE TWO NUMBERS IS 1.

d D = THE QUOTIENT OF THE TWO NUMBERS IS 7.

2 THREE COINS ARE TOSSED AT THE SAME TIME. DRAW A TREE DIAGRAM FOR THE OUTCOMES OF THIS EXPERIMENT. WHAT IS THE POSSIBILITY SET?

3 A BAG CONTAINS FOUR RED BALLS AND THREE BLACK BALLS. WHAT IS THE POSSIBILITY SET FOR COLOUR, IF 2 BALLS ARE CHOSEN AT RANDOM?

4 TOSS A COIN AND KEEP A RECORD OF WHETHER IT LANDS HEADS OR TAILS. DO THIS AT LEAST 20 TIMES FOR EACH EXPERIMENT AND PERFORM AT LEAST FIVE EXPERIME YOUR RESULTS IN A TABLE LIKE THE FOLLOWING.

Experiment	Number of coin tosses	Number of heads obtained
1		
2		
3		
4		
5		
TOTAL		

A DO YOU FEEL THAT THE TWO OUTCOMES "HEAD" AND "TAIL" ARE EQUALLY LIKELY?

B DO YOUR EXPERIMENTAL RESULTS SUPPORT THIS FEELING?

C WHAT IS THE RATIO OF THE NUMBER OF HEADS TO THE NUMBER OF TOSSES IN EACH EXPERIMENT?

D WHAT RATIO DO YOU HAVE FOR THE TOTAL NUMBER OF HEADS TO THE TOTAL NUMBER OF TOSSES?

5 A FAIR DIE IS ROLLED ONCE. CALCULATE THE PROBABILITY OF GETTING:

A AN ODD NUMBER **B** A SCORE OF 5

C A PRIME NUMBER **D** A SCORE OF 0

6 A NUMBER IS SELECTED AT RANDOM FROM THE WHOLE NUMBERS 1 TO 20, BOTH INCLUSIVE. FIND THE PROBABILITY THAT THE NUMBER SELECTED IS:

A EVEN **C** A MULTIPLE OF 3 **E** THE SQUARE OF 2

B A MULTIPLE OF 2 AND 3 **D** EVEN OR ODD **F** THE SQUARE OF 6

7 A BAG CONTAINS FIVE RED BALLS, THREE BLACK BALLS AND WHITE BALLS. A BALL IS DRAWN OUT OF THE BAG AT RANDOM. WHAT IS THE PROBABILITY THAT THE BALL

A WHITE? **B** RED? **C** BLACK?

8 A BAG CONTAINS 100 IDENTICAL CARDS ON WHICH THE NUMBERS 1 TO 100 ARE MARKED. A CARD IS DRAWN RANDOMLY. FIND THE PROBABILITY THAT THE NUMBER ON THE C

A AN EVEN NUMBER **B** AN ODD NUMBER **C** A MULTIPLE OF 7

D A MULTIPLE OF 5 **E** A MULTIPLE OF 3 **F** LESS THAN 76

G GREATER THAN 32 **H** A FACTOR OF 24

Key Terms

- analysis
- arithmetic mean
- average
- classification
- collection
- descriptive statistics
- equally likely
- event
- frequency
- frequency distribution
- histogram
- interpretation
- measure of central tendency
- measure of dispersion
- measure of location
- median
- mode
- outcomes
- presentation
- population
- population function
- possibility set
- primary data
- probability
- range
- raw data
- sample
- sample space
- secondary data
- standard deviation
- statistical data
- statistics
- tabulation
- variable (or variate)
- variance

Summary

1 STATISTICS IS THE SCIENCE OF COLLECTING, ORGANIZING, PRESENTING, ANALYSING AND INTERPRETING DATA IN ORDER TO DRAW CONCLUSIONS.

2 A POPULATION IS THE COMPLETE COLLECTION OF INDIVIDUALS, OBJECTS OR MEASUREMENTS THAT HAVE A CHARACTERISTIC IN COMMON.

3 A SMALL PART (OR A SUBSET) OF A POPULATION IS CALLED A SAMPLE.

4 IF THE CATEGORIES OF A CLASSIFICATION ARE BASED ON SOME ATTRIBUTE OR CHARACTERISTICS WHOSE VALUES ARE NOT NUMBERS, THEN IT IS CALLED QUALITATIVE CLASSIFICAT

5 IF THE CHARACTERISTIC OF INTEREST IS NUMERICAL, THEN IT IS CALLED QUANTITATIVE CLASSIFICATION.

6 DESCRIPTIVE STATISTICS IS A BRANCH OF STATISTICS CONCERNED WITH SUMMARIZING AND DESCRIBING A LARGE AMOUNT OF DATA.

7 DATA IS SAID TO BE primary, IF IT IS OBTAINED FIRST-HAND FOR THE PARTICULAR PURPOSE ON WHICH ONE IS CURRENTLY WORKING.

8 DATA THAT HAS BEEN PREVIOUSLY COLLECTED FOR A DIFFERENT PURPOSE IS CALLED SECONDARY DATA.

9 A STATISTICAL TABLE IS A SYSTEMATIC PRESENTATION OF DATA IN COLUMNS AND ROWS.

10 THE QUANTITY THAT WE MEASURE FROM OBSERVATIONS IS CALLED A VARIATE (OR VARIABLE).

11 THE DISTRIBUTION OF A POPULATION FUNCTION IS THE FUNCTION THAT ASSOCIATES WITH EACH VARIATE OF THE POPULATION FUNCTION THE CORRESPONDING FREQUENCY.

12 A FREQUENCY DISTRIBUTION IS A DISTRIBUTION SHOWING THE NUMBER OF OBSERVATIONS ASSOCIATED WITH EACH DATA VALUE.

13 A HISTOGRAM IS A PICTORIAL REPRESENTATION OF A FREQUENCY DISTRIBUTION IN WHICH THE VARIABLE(S) ARE PLOTTED ON THE x-AXIS AND THE FREQUENCY OF OCCURRENCE IS PLOTTED ON THE y-AXIS.

14 IF $x_1, x_2, x_3, \ldots x_n$ ARE THE n OBSERVATIONS OF A VARIABLE x, THEN THE MEAN ($\bar{x}$) IS GIVEN BY

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

15 THE MEDIAN OF A VARIABLE IS THE VALUE THAT FALLS IN THE MIDDLE OF THE DATA WHEN ARRANGED IN ASCENDING OR DESCENDING ORDER.

16 THE MODE OF A VARIABLE IS THE MOST FREQUENT OBSERVATION OF THE VARIABLE THAT OCCURS IN THE DATA SET.

17 THE RANGE R OF A SET OF NUMERICAL DATA IS THE DIFFERENCE BETWEEN THE MAXIMUM AND MINIMUM VALUES.

RANGE = MAXIMUM VALUE − MINIMUM VALUE

18 STANDARD DEVIATION IS THE SQUARE ROOT OF THE MEAN OF THE SQUARED DEVIATION OF EACH VALUE FROM THE ARITHMETIC MEAN.

19 THE OUTCOMES OF AN EXPERIMENT ARE SAID TO BE EQUALLY LIKELY, WHEN THE EXPERIMENT IS REPEATED A LARGE NUMBER OF TIMES, EACH OUTCOME OCCURS EQUALLY OFTEN.

20 THE POSSIBILITY SET FOR AN EXPERIMENT IS THE SET OF ALL POSSIBLE OUTCOMES OF THE EXPERIMENT. IT IS ALSO KNOWN AS THE SAMPLE SPACE.

21 AN EVENT IS A SUBSET OF THE POSSIBILITY SET.

22 IF S IS THE POSSIBILITY SET OF AN EXPERIMENT AND EACH ELEMENT OF S IS EQUALLY LIKELY, THEN THE PROBABILITY OF AN EVENT E OCCURRING, DENOTED BY $P(E)$, IS DEFINED AS:

$$P(E) = \frac{\text{NUMBER OF ELEMENTS IN } E}{\text{NUMBER OF ELEMENTS IN } S} = \frac{n(E)}{n(S)}$$

Review Exercises on Unit 6

1 WHAT IS MEANT BY SUMMARIZING AND DESCRIBING DATA?

2 THE MARKS OF 30 STUDENTS IN A MATHEMATICS TEST ARE GIVEN BELOW.

3	5	4	6	8	12	14	5	6	5
8	5	9	10	9	10	12	10	12	10
12	13	10	15	14	15	14	15	14	14

A CONSTRUCT A FREQUENCY DISTRIBUTION TABLE.

B DRAW A HISTOGRAM TO REPRESENT THE DATA.

C WHAT PERCENT OF THE STUDENTS HAVE SCORED LESS THAN 15?

3 REFER TO THE FOLLOWING HISTOGRAM TO ANSWER THE QUESTIONS THAT FOLLOW.

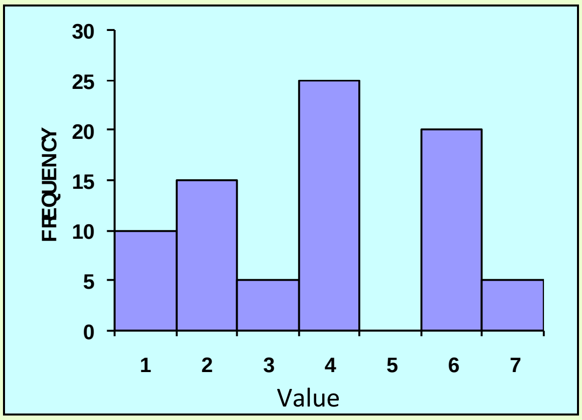


Figure 6.5

A PREPARE A FREQUENCY DISTRIBUTION TABLE.

B WHAT IS THE HIGHEST VARIABLE?

C WHAT IS THE HIGHEST FREQUENCY?

D HOW MANY VARIATES OCCUR 5 TIMES?

E WHICH VARIATES HAVE THE MINIMUM FREQUENCY?

4 FIND THE MEAN, MEDIAN, MODE, RANGE, VARIANCE AND STANDARD DEVIATION OF THE POPULATION FUNCTION WHOSE DISTRIBUTION IS GIVEN IN THE TABLE BELOW.

V	2	3	4	5	6
f	2	4	1	2	3

5 FIND THE MEAN, MEDIAN, MODE, RANGE AND VARIANCE OF THE HISTOGRAM GIVEN BELOW.

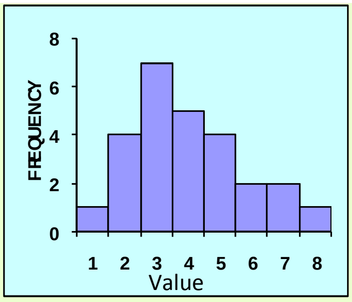


Figure 6.6

6 WHY CAN THE PROBABILITY OF AN EVENT NOT BE $2\sqrt{2}-$, $\frac{\ }{2}$ OR $\frac{3}{2}$?

7 AN INTEGER n, $1 \le n \le 144$, IS PICKED AT RANDOM. WHAT IS THE PROBABILITY THAT n IS THE SQUARE OF AN INTEGER?

8 GIVEN THE FOLLOWING VALUES OF A POPULATION FUNCTION:

5, 4, 7, 3, 6, 5, 3, 1, 5, 7, 5, 9.

FIND THE PROBABILITY THAT A RANDOMLY CHOSEN VALUE FROM THE DATA IS

A A MODAL VALUE;

B BELOW THE MEAN VALUE;

C ANY OF THE NUMBERS 1, 4, 6 OR 9;

D AN ODD NUMBER GREATER THAN THE MEAN VALUE.

9 TWO FAIR DICE ARE ROLLED ONCE. WHAT IS THE PROBABILITY THAT THE DIFFERENCE OF THE TWO NUMBERS SHOWN IS 1?

10 GIVEN BELOW IS THE FREQUENCY DISTRIBUTION OF A POPULATION V.

V	−10	−5	0	5	10
f	5	10	5	20	10

IF AN ELEMENT IS DRAWN RANDOMLY FROM THE POPULATION FIND THE PROBABILITY THAT IT IS

A NON-NEGATIVE; **B** NON-ZERO;

C LESS THAN OR EQUAL TO −5; **D** POSITIVE;

11 THE MEDIAN OF $x - 4$, x, $2x$ AND $2x + 12$ IS 9, WHERE x IS A POSITIVE INTEGER. FIND THE VALUE OF x.

12 THE TABLE BELOW SHOWS THE NUMBER OF STUDENTS WHO SCORED MARKS 3, 4 OR 5 IN A MATHS TEST.

Mark	3	4	5
Number of Students	3	x	4

IF THE MEAN MARK IS 4.1, HOW MANY STUDENTS SCORED 4?

13 IN A CLASS OF BOYS AND GIRLS, THE MEAN WEIGHT OF 8 BOYS IS 55 KG AND THE MEAN WEIGHT OF A GROUP OF GIRLS IS 48 KG. THE MEAN WEIGHT OF ALL THE CHILDREN HOW MANY GIRLS ARE THERE?

14 THERE ARE 24 RIGHT-HANDED STUDENTS IN A CLASS OF 30. WHAT IS THE PROBABILITY THAT A STUDENT CHOSEN AT RANDOM WILL BE LEFT-HANDED?

15 SUPPOSE YOU WRITE THE DAYS OF THE WEEK ON 7 IDENTICAL PIECES OF PAPER. YOU MIX THEM IN A BOWL AND SELECT ONE AT A TIME. WHAT IS THE PROBABILITY THAT T SELECT WILL HAVE THE LETTER R IN IT?

16 A PAIR OF DICE ARE ROLLED. FIND THE PROBABILITY THAT THE SUM OF THE NUMBERS ON THE UPPER FACES IS:

A 9; **B** GREATER THAN 9; **C** EVEN; **D** NOT GREATER THAN 9;

E GREATER THAN 9 AND EVEN; **F** GREATER THAN 9 OR EVEN.

17 FROM THE MEMBERS OF A FARMERS' ASSOCIATION, 50 FARMERS CULTIVATED WHEAT. AN AGRICULTURAL EXPERT WANTS TO STUDY THE FARMERS' YIELD IN TERMS OF (HARVESTED PER HECTARE AND FOUND THE FOLLOWING

50	45	45	50	46	48	55	48	52	54
51	52	45	55	46	50	55	54	49	51
48	46	51	52	47	45	49	54	46	48
53	52	48	46	55	47	51	47	50	53
47	53	48	45	54	48	50	46	52	54

A PREPARE A FREQUENCY DISTRIBUTION TABLE THAT REPRESENTS THE DATA.

B DRAW A HISTOGRAM.

C FIND THE MODE OF THE DATA.

D IF THE WEREDA AGRICULTURE OFFICE WANTS TO PRAISE FARMERS WHO PRODUCED MORE THAN 52 QUINTALS PER HECTARE, HOW MANY FARMERS WILL THEY PRAISE?

18 WHICH OF THE FOLLOWING IS TRUE?

A THE MEAN, MODE AND MEDIAN OF A POPULATION FUNCTION CANNOT BE EQUAL.

B THE RANGE AND THE STANDARD DEVIATION OF A POPULATION FUNCTION ARE INVERSELY RELATED.

C THE RANGE OF A POPULATION FUNCTION CAN BE A NON-POSITIVE NUMBER.

D THE SUM OF THE DEVIATIONS OF EACH VALUE OF A POPULATION FROM THE MEAN WILL ALWAYS BE ZERO.