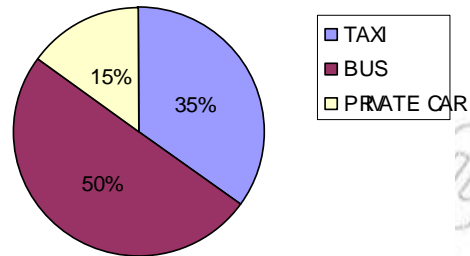


Unit 8



FURTHER ON STATISTICS

Unit Outcomes:

After completing this unit, you should be able to:

- *know basic concepts about sampling techniques.*
- *construct and interpret statistical graphs.*
- *know specific facts about measurement in statistical data.*

Main Contents

8.1 SAMPLING TECHNIQUES

8.2 REPRESENTATION OF DATA

8.3 CONSTRUCTION OF GRAPHS AND INTERPRETATION

8.4 MEASURES OF CENTRAL TENDENCY AND MEASURES OF VARIABILITY

8.5 ANALYSIS OF FREQUENCY DISTRIBUTIONS

8.6 USE OF CUMULATIVE FREQUENCY CURVES

Key terms

Summary

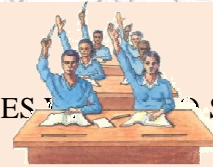
Review Exercises

INTRODUCTION

IN GRADE 9 AND GRADE 11, YOU DID SOME WORK IN STATISTICS, INCLUDING COLLECTING AND TABULATING OF STATISTICAL DATA, FREQUENCY DISTRIBUTIONS AND HISTOGRAMS, MEASURES OF LOCATION (MEAN, MEDIAN AND MODE(S), QUARTILES, DECILES AND PERCENTILES), MEASURES OF DISPERSION FOR BOTH UNGROUPED AND GROUPED DATA, AND SOME IDEAS OF PROBABILITY. IN THIS UNIT, YOU WILL STUDY DESCRIPTIVE STATISTICS.

8.1 SAMPLING TECHNIQUES

ACTIVITY 8.1



THE MINISTRY OF AGRICULTURE AND NATURAL RESOURCES WANTS TO STUDY THE PRODUCTIVITY BENEFITS OF USING IRRIGATION FARMING.

IF YOU WERE ASKED TO STUDY THIS, OBVIOUSLY YOU WOULD START BY COLLECTING DATA. ASK THE FOLLOWING QUESTIONS.

- 1 WHY DO YOU NEED TO COLLECT DATA? HOW WOULD YOU COLLECT THE DATA?
- 2 HOW MUCH DATA DO YOU NEED TO COLLECT?
- 3 FROM WHERE WOULD YOU COLLECT THE DATA?

STATISTICS AS A SCIENCE DEALS WITH THE PROPER COLLECTION, ORGANIZATION, PRESENTATION, ANALYSIS AND INTERPRETATION OF NUMERICAL DATA. SINCE STATISTICS IS USEFUL FOR MAKING DECISIONS OR FORECASTING FUTURE EVENTS, IT IS APPLICABLE IN ALMOST ALL SCIENCES. IT IS ALSO USEFUL IN SOCIAL, ECONOMIC AND POLITICAL ACTIVITIES. IT IS ALSO USEFUL IN SCIENTIFIC INVESTIGATIONS. SOME EXAMPLES OF APPLICATIONS OF STATISTICS ARE GIVEN BELOW.

1 Statistics in business

STATISTICS IS WIDELY USED IN BUSINESS TO MAKE BUSINESS A SUCCESSFUL ONE. A BUSINESS MUST KEEP A PROPER RECORD OF INFORMATION IN ORDER TO PREDICT THE COURSE OF THE BUSINESS, AND SHOULD BE ACCURATE IN STATISTICAL AND BUSINESS FORECASTING. STATISTICS CAN ALSO BE USED TO HELP IN FORMULATING ECONOMIC POLICIES AND EVALUATING THEIR EFFECT.

2 Statistics in meteorology

METEOROLOGISTS FORECAST WEATHER FOR FUTURE PERIODS. THE INFORMATION THEY OBTAIN FROM DIFFERENT SOURCES. HENCE THEIR FORECASTS ARE BASED ON STATISTICS THAT ARE COLLECTED.

3 Statistics in schools

IN SCHOOLS, TEACHERS RANK THEIR STUDENTS AT THE END OF A SEMESTER BASED ON INFORMATION COLLECTED THROUGH DIFFERENT METHODS (EXAMS, TESTS, QUIZZES, ETC.) WHICH GIVES AN INDICATION OF THE STUDENTS' PERFORMANCE.

Note:

- 1 COLLECTION OF DATA IS THE BASIS FOR ANALYSIS. STATISTICAL CARE MUST BE TAKEN AT THIS STAGE TO GET ACCURATE DATA. INACCURATE AND INADEQUATE DATA MAY LEAD TO WRONG OR MISLEADING CONCLUSIONS AND CAUSE POOR DECISIONS TO BE MADE.
- 2 RECALL THAT **population** IN STATISTICS MEANS THE COMPLETE COLLECTION OF ITEMS (INDIVIDUALS) UNDER CONSIDERATION.

IT IS OFTEN IMPRACTICAL AND TOO COSTLY TO COLLECT DATA FROM THE WHOLE POPULATION. A CENSUS SURVEY. CONSEQUENTLY, IT IS FREQUENTLY NECESSARY TO USE THE PROCESS OF SAMPLING FROM WHICH CONCLUSIONS ARE DRAWN ABOUT A WHOLE POPULATION. THIS LEADS YOU TO AN ESSENTIAL STATISTICAL CONCEPT WHICH IS IMPORTANT FOR PRACTICAL PURPOSES.

A **sample** IS A LIMITED NUMBER OF ITEMS TAKEN FROM WHICH ONLY A PART IS STUDIED/ INVESTIGATED.

A SAMPLE NEEDS TO BE TAKEN IN SUCH A WAY THAT IT IS A TRUE REPRESENTATION OF THE POPULATION. IT SHOULD NOT BE BIASED SO AS TO CAUSE A WRONG CONCLUSION. AVOIDING BIAS REQUIRES THE USE OF PROPER SAMPLING TECHNIQUES. BEFORE EXAMINING SAMPLING TECHNIQUES YOU NEED TO NOTE THE FOLLOWING.

DURING SAMPLING, THE FOLLOWING POINTS MUST BE CONSIDERED.

- 1 **Size of a sample:** THERE IS NO SINGLE RULE FOR DETERMINING THE SIZE OF A GIVEN POPULATION. HOWEVER, THE SIZE SHOULD BE ADEQUATE IN ORDER TO REPRESENT THE POPULATION.
 - I **Homogeneity or heterogeneity of the population:** IF THE POPULATION HAS A HOMOGENEOUS NATURE, A SMALLER SIZE SAMPLE IS SUFFICIENT. (FOR EXAMPLE, A SINGLE DROP OF BLOOD IS SUFFICIENT TO TAKE A BLOOD TEST FROM SOMEONE).
 - II **Availability of resources:** IF SUFFICIENT RESOURCES ARE AVAILABLE, IT IS ADVISABLE TO INCREASE THE SIZE OF THE SAMPLE.
- 2 **Independence:** EACH ITEM OR INDIVIDUAL IN THE POPULATION SHOULD HAVE AN EQUAL CHANCE OF BEING SELECTED AS A MEMBER OF THE SAMPLE.

Techniques of sampling

ACTIVITY 8.2



THE ATHLETICS FEDERATION DECIDES TO CONSTRUCT AN ATHLETICS ACADEMY IN SOME PART OF ETHIOPIA. FOR THIS PURPOSE, IT NEEDS TO STUDY THE POTENTIAL SOURCE OF ATHLETES SO AS TO DECIDE WHERE TO BUILD THE FACILITY.

- 1 IS IT POSSIBLE TO STUDY THE WHOLE POPULATION? WHY?

2 HOW WOULD THE FEDERATION COLLECT A SAMPLE POPULATION?

3 WHAT CHARACTERISTICS MUST BE FULFILLED BY THE SAMPLE?

THERE ARE VARIOUS TECHNIQUES OF SAMPLING, BUT THEY CAN BE BROADLY GROUPED INTO TWO MAIN CATEGORIES:

- A** RANDOM OR PROBABILITY SAMPLING.
- B** NON RANDOM OR NON PROBABILITY SAMPLING.

YOU WILL CONSIDER ONLY RANDOM (PROBABILITY) SAMPLING.

Random Sampling

IN THIS METHOD, EVERY MEMBER OF THE POPULATION HAS AN EQUAL CHANCE OF BEING SELECTED FOR THE SAMPLE. ONLY CHANCE DETERMINES WHICH ITEM IS TO BE SELECTED. THREE OF THE COMMONLY USED METHODS WHICH WILL BE DISCUSSED ARE **simple random sampling**, **systematic sampling** AND **stratified sampling**.

I Simple random sampling (SRS)

SIMPLE RANDOM SAMPLING IS CHARACTERIZED BY TOTAL RANDOMNESS. TO APPLY THIS METHOD, YOU MAY EITHER USE THE LOTTERY METHOD OR A TABLE OF RANDOM NUMBERS (AT THE END OF THE TEXTBOOK).

The Lottery method

IN THIS METHOD AN INVESTIGATOR

- ✓ PREPARES SLIPS OF PAPER WHICH ARE IDENTICAL IN SIZE AND COLOUR.
- ✓ WRITES NAMES OR CODE NUMBERS FOR EACH MEMBER OF THE POPULATION ON THE SLIP.
- ✓ FOLDS THE SLIPS AND PUTS THEM IN A CONTAINER, AND MIXES THEM WELL.
- ✓ A BLINDFOLD SELECTION IS THEN MADE UNTIL THE REQUIRED SAMPLE SIZE IS OBTAINED.

Example 1 A MATHEMATICS TEACHER IN A SCHOOL WANTS TO DETERMINE THE AVERAGE WEIGHT OF GRADE 12 STUDENTS. THERE ARE 6 SECTIONS OF GRADE 12 IN THE SCHOOL. ASSUMING THAT THERE ARE 45 STUDENTS IN EACH CLASS AND REQUIRING A SAMPLE SIZE OF 30 (5 FROM EACH SECTION), HOW CAN SHE USE THE LOTTERY METHOD TO SELECT HER SAMPLE?

Solution **A** PREPARE 45 CARDS OF SAME SIZE AND COLOUR, WITH NUMBER 0 WRITTEN ON 40 OF THEM AND THE NUMBER 1 WRITTEN ON 5 OF THEM.

- B** PUT THE CARDS ON A TABLE WITH THE NUMBERS FACING DOWN.
- C** INVITE THE STUDENTS (ONE AT A TIME) TO COME AND PICK A CARD.
- D** THOSE WHO PICK CARDS WITH THE NUMBER 0 ARE MEMBERS OF THE SAMPLE.
- E** REPEAT THE SAME PROCESS FOR EACH SECTION.

Note:

MAXIMUM CARE HAS TO BE TAKEN AT THIS STAGE TO GET ACCURATE DATA. INACCURATE OR INADEQUATE DATA MAY LEAD TO WRONG CONCLUSIONS. THUS,

- A** CARE SHOULD BE TAKEN SO THAT EACH STUDENT TAKES JUST ONE CARD.
- B** THE CARDS SHOULD BE WELL SHUFFLED BEFORE BEING TAKEN.
- C** THE SAME SET OF CARDS SHOULD BE USED THROUGHOUT ALL THE SECTIONS.

Using a table of random numbers

For this method, you need to use a table of random numbers, and you need to take the following steps.

- ✓ EACH MEMBER OF THE POPULATION IS GIVEN A UNIQUE NUMBER.
- ✓ SELECT ARBITRARILY ONE RANDOM NUMBER FROM RANDOM NUMBERS.
- ✓ STARTING WITH THE SELECTED RANDOM NUMBER, READ THE LIST OF RANDOM NUMBERS AND MATCH THESE WITH THE MEMBERS OF THE POPULATION IN THEIR CONSECUTIVE NUMBER ORDER.
- ✓ SORT THE SELECTED RANDOM NUMBERS INTO EITHER ASCENDING OR DESCENDING ORDER.
- ✓ IF YOU NEED A SAMPLE OF SIZE n , THEN SELECT THE SAMPLE THAT CORRESPONDS WITH THE FIRST n RANDOM NUMBERS.

Example 2 FOR THE PROBLEM IN EXAMPLE 1 ABOVE, USE THE TABLE OF RANDOM NUMBERS ATTACHED AT THE END OF THE TEXTBOOK TO SELECT A SAMPLE OF 30 STUDENTS (5 FROM EACH SECTION).

Solution

- A** GIVE EACH STUDENT A ROLE NUMBER FROM 1 TO 45 IN ORDER.
- B** SELECT ARBITRARILY ONE RANDOM NUMBER FROM RANDOM NUMBERS.
- C** FROM THE SELECTED RANDOM NUMBER, READ 45 CONSECUTIVE NUMBERS AND ATTACH EACH TO THE CONSECUTIVE NUMBERS GIVEN TO EACH MEMBER OF THE POPULATION.
- D** SORT THE SELECTED RANDOM NUMBERS (TO NUMBERS WITHIN THE 45) INTO ASCENDING OR DESCENDING ORDER.
- E** TAKE THE FIRST 5 RANDOM NUMBERS AND THE CORRESPONDING STUDENTS. THE STUDENTS WHOSE ROLE NUMBERS ARE SELECTED WILL BE PART OF THE SAMPLE.

II Systematic sampling

SYSTEMATIC SAMPLING IS ANOTHER RANDOM SAMPLING TECHNIQUE USED FOR SELECTING A SAMPLE FROM A POPULATION. IN ORDER TO APPLY THIS METHOD, YOU TAKE THE FOLLOWING STEPS:

IF N = SIZE OF THE POPULATION AND n = SIZE OF THE SAMPLE, THEN USE SAMPLING INTERVAL. AFTER THIS, YOU ARBITRARILY SELECT ONE NUMBER BETWEEN L AND $L + k - 1$. THEN EVERY NEXT SAMPLE MEMBER IS SELECTED BY CONSIDERING THE SELECTION OF THE SELECTED ONE.

Example 3 IN A CLASS, THERE ARE 80 STUDENTS WITH ROLL NUMBERS FROM 1-80. YOU NEED TO SELECT A SAMPLE OF 10 STUDENTS. HOW CAN YOU APPLY THE SYSTEMATIC SAMPLING TECHNIQUE?

Solution YOU APPLY THE SYSTEMATIC SAMPLING TECHNIQUE AS FOLLOWS

$$N = 80 \quad n = 10 \quad k = \frac{80}{10} = 8$$

FIRST, SORT THE LIST IN ASCENDING ORDER AND CHOOSE ONE NUMBER AT RANDOM FROM THE FIRST 8 NUMBERS. IF THE SELECTED NUMBER IS 5, THEN THE SAMPLE NUMBERS THAT YOU OBTAIN BY TAKING EVERY EIGHTH NUMBER UNTIL YOU GET THE TENTH SAMPLE NUMBER ARE

5, 13, 21, 29, 37, 45, 53, 61, 69, 77

WHAT DO YOU THINK WILL THE SAMPLE BE, IF THE FIRST RANDOMLY SELECTED NUMBER IS 3?

Note:

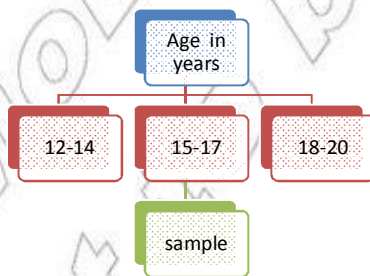
IN SYSTEMATIC SAMPLING, YOU USE $(n - 1)k$ WHERE k IS THE FIRST RANDOMLY SELECTED SAMPLE MEMBER OF A SAMPLE AND k IS THE SAMPLING INTERVAL.

III Stratified sampling

STRATIFIED SAMPLING IS USEFUL WHENEVER THE POPULATION UNDER CONSIDERATION HAS SOME IDENTIFIABLE STRATUM OR CATEGORICAL DIFFERENCE WHERE, IN EACH STRATUM, THE DATA ITEMS ARE SUPPOSED TO BE HOMOGENEOUS. IN THIS METHOD, THE POPULATION IS DIVIDED INTO HOMOGENEOUS GROUPS OR CLASSES CALLED STRATA AND A SAMPLE IS DRAWN FROM EACH STRATUM. ONCE YOU IDENTIFY THE STRATA, YOU SELECT A SAMPLE FROM EACH STRATUM EITHER BY RANDOM SAMPLING OR SYSTEMATIC SAMPLING.

CONSIDER THE FOLLOWING EXAMPLE.

Example 4 IF YOU CONSIDER STUDENTS IN A SECTION OF A CLASS OF AGE AS STRATA. IN SUCH A CASE, YOU COULD TAKE THE AGE GROUPS 12 – 14, 15 – 17 AND 18 – 20 AS STRATIFICATION OF THE STUDENTS.



Samples are taken from each stratum proportionally. In this case, strata are age groups from 12-14, 15-17 and 18-20.

Figure 8.1

SO FAR, THE THREE DIFFERENT SAMPLING TECHNIQUES ARE DISCUSSED. HOWEVER, NO TECHNIQUE IS BETTER THAN THE OTHERS. EACH HAS ITS OWN ADVANTAGES AND LIMITATIONS. ADVANTAGES AND LIMITATIONS OF RANDOM SAMPLING ARE MENTIONED BELOW.

Advantages of random sampling

- ✓ IT IS FREE FROM ANY PERSONAL BIAS OF THE INVESTIGATOR.
- ✓ THE SAMPLE IS A BETTER REPRESENTATIVE.

Limitations of random sampling

- ✓ IT NEEDS SKILL AND EXPERIENCE.
- ✓ IT REQUIRES TIME TO PLAN AND CARRY OUT.

Exercise 8.1

- 1 DEFINE THE TERM STATISTICS.
- 2 DESCRIBE THE DIFFERENCE BETWEEN THE STATISTICAL INFERENCE AND THE STATISTICAL INFERENTIAL STATISTICS.
- 3 EXPLAIN AND DESCRIBE THREE SAMPLING TECHNIQUES.
- 4 BY USING FURTHER READING, EXPLAIN OTHER TYPES OF PROBABILITY AND NON-PROBABILITY SAMPLING).
- 5 FROM A POPULATION OF SIZE 100 LISTED IN THE TABLE, SELECT A SAMPLE OF SIZE 20 AND THE FIRST RANDOMLY SELECTED NUMBER IS 4, DETERMINE:
 - A THE SAMPLING INTERVAL;
 - B ALL MEMBERS OF THE SAMPLE.
- 6 DISCUSS THE ADVANTAGES AND LIMITATIONS OF EACH TECHNIQUE.

8.2 REPRESENTATION OF DATA

ACTIVITY 8.3



THE FOLLOWING DATA OF STUDENTS' WEIGHTS (IN KG) IS COLLECTED:

65	48	52	55	62	58	47	53	65	71	54
50	62	51	49	54	60	68	53	57	62	59

PREPARE A GROUPED FREQUENCY TABLE FOR THE DATA USING A CLASS WIDTH OF 5.

AS YOU WELL KNOW, RAW DATA WHICH HAS BEEN COLLECTED AND EDITED, WILL NOT IMMEDIATELY GIVE REQUIRED INFORMATION. IT USUALLY NEEDS TO BE PUT INTO A FORM THAT MAKES IT EASIER TO UNDERSTAND AND INTERPRET, SUCH AS TABLES, GRAPHS OR DIAGRAMS. YOU STUDIED HOW TO REPRESENT DATA IN TABLES AND CHARTS IN GRADE 9. HERE, YOU WILL CONSIDER SOME OTHER DATA REPRESENTATIONS DISCUSSING THEIR IMPORTANCE, WITH THEIR STRENGTHS AND WEAKNESSES.

- A COMPUTATIONAL ANALYSIS AND DECISION MAKING,
- B PROVIDING INFORMATION FOR PUBLIC AWARENESS AND OTHER PURPOSES.

Tabular methods of data presentation

ONE OF THE COMMON WAYS OF REPRESENTING DATA IS THE OFTEN, YOU USE FREQUENCY DISTRIBUTION TABLES. A FREQUENCY DISTRIBUTION TABLE IS A TABLE WHICH LIST OF ALL DATA VALUES OBTAINED, WITH THEIR RESPECTIVE FREQUENCIES.

Example 1 THE FOLLOWING REPRESENTS THE AGES OF 20 WOMEN TAKEN WHEN THEY GAVE BIRTH TO THEIR FIRST CHILD.
 24, 25, 27, 26, 22, 28, 24, 25, 23, 24, 27, 26, 25, 24, 25, 25, 24, 25, 24, 26
 REPRESENT THE DATA USING A DISCRETE FREQUENCY DISTRIBUTION TABLE.

Solution YOU CAN REPRESENT THE ABOVE DATA USING DISCRETE FREQUENCY DISTRIBUTION AS FOLLOWS:

Age (in years) (x)	Tally marks	Number of women (f)
22		1
23		1
24	HHH	6
25	HHH	6
26		3
27		2
28		1

FROM THIS FREQUENCY DISTRIBUTION TABLE, YOU CAN DRAW SOME CONCLUSIONS ABOUT THE WOMEN. YOU CAN IDENTIFY THAT THE MAJORITY OF THE WOMEN FIRST GAVE BIRTH AT AGES OF 24 AND 25. THE ABOVE DATA CAN BE FURTHER SUMMARIZED USING A GROUPED FREQUENCY DISTRIBUTION AS FOLLOWS:

Age (in years)	Tally	Number of women
22 – 24	HHH	8
25 – 27	HHH HHH	11
28 – 30		1

THIS PROVIDES MORE CONCISE INFORMATION. FROM THIS, YOU CAN, FOR EXAMPLE, SAY THAT THE MAJORITY OF THE WOMEN FIRST GAVE BIRTH BEFORE THE AGE OF 28.

Graphical methods of data presentation

THE OTHER WAY IN WHICH YOU REPRESENT DATA IS THROUGH GRAPHICAL REPRESENTATIONS THAT YOU ARE GOING TO DISCUSS IN THE FOLLOWING SECTION INCLUDE BAR CHARTS, PIE CHARTS AND FREQUENCY GRAPHS.

Example 2 THE FOLLOWING BAR CHART REPRESENTS THE ENROLMENT IN PREPARATORY PROGRAMS IN ETHIOPIA FROM 1996 E.C TO 2000 E.C. CAN YOU USE THE BAR CHART TO ANSWER THE FOLLOWING?

- A** IS THE ENROLMENT INCREASING OR DECREASING OVER THE YEARS?
- B** BETWEEN WHICH TWO YEARS DOES FEMALE ENROLMENT INCREASE MOST SIGNIFICANTLY?

**ENROLLMENT IN PREPARATORY PROGRAM (11-12)
BY GENDER**

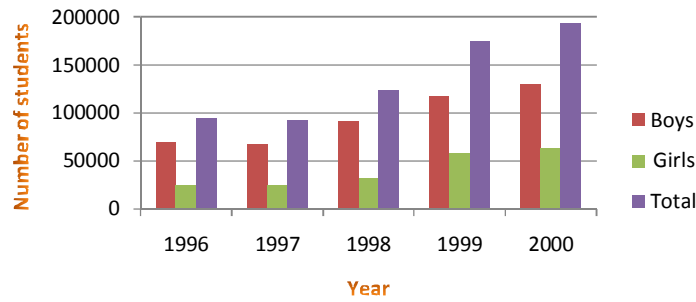


Figure 8.2

Solution

A THE ENROLMENT IS INCREASING STARTING FROM 1998.

B THERE SEEMS TO BE NO CHANGE IN THE NUMBER OF ENROLMENT TWO YEARS. BUT FROM 1999 ONWARDS, THERE IS A CONSIDERABLE INCREASE IN ENROLMENT OF

FROM THE ABOVE EXAMPLES, YOU SEE THAT DATA REPRESENTATION CAN BE A USEFUL PRESENT INFORMATION, FROM WHICH CONCLUSION COULD BE DRAWN.

Example 3 THE FOLLOWING BAR CHART REPRESENTS PERCENTAGE OF SW FEMALE GENITAL MUTILATION, BY EDUCATIONAL LEVEL.

A DETERMINE THE FOUR COUNTRIES THAT HAVE A GREATER PREVALENCE BETWEEN OLDER WOMEN (AGES 35 TO 39) AND YOUNGER WOMEN (AGES 15 TO 19).

B WHAT SIGNIFICANCE DOES THIS HAVE FOR POLICY MAKERS TRYING TO STOP FEMALE GENITAL MUTILATION?

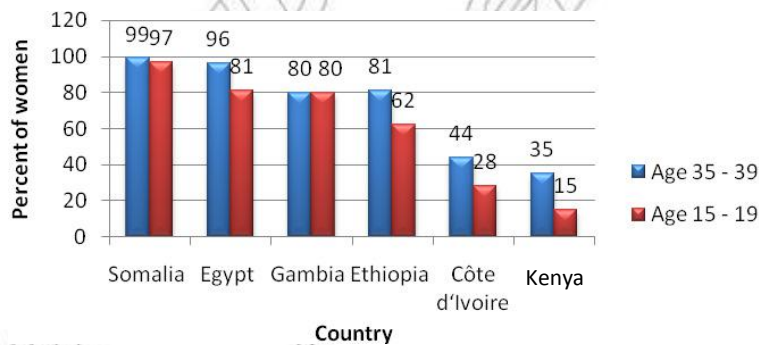


Figure 8.3: Source: Population Reference Bureau, Female Genital Mutilation/Cutting: Data and Trends update 2010

Solution

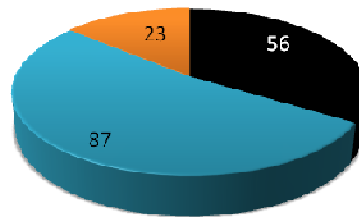
A THE FOUR COUNTRIES IN THE SURVEY THAT HAVE A GREATER PREVALENCE BETWEEN OLDER WOMEN (AGES 35 TO 39) AND YOUNGER WOMEN (AGES 15 TO 19) ARE KENYA, ETHIOPIA, CÔTE D'IVOIRE, AND EGYPT

B FOR POLICY MAKERS, IT MIGHT SUGGEST THAT THE LARGER DIFFERENCE IN PREVALENCE BETWEEN OLDER WOMEN (AGES 35 TO 39) AND YOUNGER WOMEN (AGES 15 TO 19) ARE DOING BETTER. THIS MAY BE A SIGN THAT THE PRACTICE IS BEING ABANDONED.

Example 4 AN AGRICULTURAL FIRM, WHICH PLANTS COFFEE, TEA AND OTHER HERBAL DRINKS, HAS CONDUCTED A SURVEY ON THE USE OF COFFEE, TEA AND OTHER HERBAL DRINKS IN A COMMUNITY, IN ORDER TO ASSESS THE MARKET POTENTIAL FOR ITS PRODUCTS. CAN THE CHART BELOW HELP IT IN MAKING DECISIONS?

Solution COFFEE SEEMS TO HAVE MORE OF A MARKET POTENTIAL. THE FIRM MIGHT NEED TO LAUNCH AN AWARENESS RAISING PROGRAM ABOUT THE HEALTH BENEFITS OF DRINKING HERBAL DRINKS.

Drinking habits of a community



■ Tea ■ Coffee ■ Other herbal drinks

Figure 8.4

Advantages of Graphical Presentation of Data

- 1 THEY ARE ATTRACTIVE TO THE EYE. SINCE GRAPHICAL PRESENTATION HAS VISUALIZING POWER, THEY CAN CONVEY MESSAGES EASILY.
E.G. WHILE READING BOOKS OR NEWSPAPERS, YOU FIRST GO TO THE PICTURES.
- 2 THEY ARE HELPFUL FOR MEMORIZING FACTS. PRESENTATION CREATED BY DIAGRAMS AND GRAPHS CAN BE RETAINED IN YOUR MIND FOR A LONG PERIOD OF TIME.
- 3 THEY FACILITATE COMPARISON. THEY HELP ONE TO MAKE QUICK AND ACCURATE COMPARISONS OF DATA. THEY BRING OUT HIDDEN FACTS AND RELATIONSHIPS. INFORMATION PRESENTED CAN BE EASILY UNDERSTOOD AT A GLANCE.

IN THE FOLLOWING SUB-UNIT, YOU WILL LEARN ABOUT THE INTERPRETATION OF GRAPHS.

Exercise 8.2

THE FOLLOWING IS THE WEIGHT IN KILOGRAMS OF 30 STUDENTS IN A CLASS.

52, 48, 55, 56, 57, 59, 60, 60, 52, 58, 55, 49, 50, 51, 52, 51, 57, 51, 54, 53, 55, 51, 53, 50, 60, 54, 50, 52, 48, 57

- 1 CONSTRUCT BOTH DISCRETE AND CONTINUOUS FREQUENCY DISTRIBUTIONS FOR THE ABOVE DATA.
- 2 ANSWER THE FOLLOWING QUESTIONS:
 - I WHAT IS THE NUMBER OF STUDENTS WHOSE WEIGHT IS
 - A BETWEEN 50 AND 55 KG
 - B LESS THAN 53 KG
 - C MORE THAN 54 KG
 - D BETWEEN 55 AND 60 KG
 - II IN WHICH WEIGHT GROUP DO THE WEIGHTS OF THE MAJORITY OF THE STUDENTS LIE?

8.3 CONSTRUCTION AND INTERPRETATION OF GRAPHS

IN THE ABOVE SUB UNIT, YOU DISCUSSED THE ADVANTAGES OF REPRESENTING DATA USING FORMS SUCH AS TABLES, GRAPHS OR DIAGRAMS. HERE, YOU WILL SEE WAYS TO ORGANISE AND PRESENT DATA SUCH AS HISTOGRAMS, FREQUENCY POLYGONS AND FREQUENCY CURVES, LINE GRAPHS AND PIE CHARTS OF FREQUENCY DISTRIBUTIONS.

ACTIVITY 8.4



THE FOLLOWING IS UNGROUPED DATA OF THE WEIGHT OF STUDENTS IN A CLASS (IN KILOGRAMS)

52, 48, 55, 56, 57, 59, 60, 60, 52, 58, 55, 49, 50, 51, 51
52, 57, 51, 54, 53, 55, 51, 53, 50, 60, 54, 50, 52, 48, 57

DRAW A HISTOGRAM.

8.3.1 Graphical Representation of Grouped Data

A FREQUENCY DISTRIBUTION CAN BE GRAPHICALLY PRESENTED IN THE FOLLOWING WAYS:

- I Histograms
- II Frequency polygons
- III Frequency curves

ACTIVITY 8.5



CONSIDER THE FOLLOWING GROUPED FREQUENCY DISTRIBUTION WHICH REPRESENTS THE WEEKLY WAGES OF 100 WORKERS.

Weekly wages in Birr (class limits)	Class boundaries	Class mid point	Number of workers
140 - 159	139.50 - 159.50	149.50	7
160 - 179	159.50 - 179.50	169.50	20
180 - 199	179.50 - 199.50	189.50	33
200 - 219	199.50 - 219.50	209.50	25
220 - 239	219.50 - 239.50	229.50	11
240 - 259	239.50 - 259.50	249.50	4
TOTAL			100

- 1 LOCATE THE CLASS BOUNDARIES ALONG HORIZONTAL AXIS).
- 2 ASSIGN A RECTANGULAR BAR FOR EACH CLASS BETWEEN ITS LOWER BOUNDARY AND UPPER CLASS BOUNDARY.
- 3 FIX THE HEIGHT OF EACH BAR AS THE FREQUENCY OF ITS CLASS.

i Histograms

HISTOGRAMS ARE USED TO ILLUSTRATE GROUPED OR CONTINUOUS DATA. AS YOU MAY RECALL, THERE IS AN IMPORTANT DIFFERENCE BETWEEN A BAR CHART AND A HISTOGRAM. A BAR CHART SHOWS QUANTITATIVE DISCRETE DATA AND HENCE THE VARIABLE AXIS IS JUST DIVIDED INTO SPACES. ON THE OTHER HAND, A HISTOGRAM ILLUSTRATES GROUPED OR CONTINUOUS DATA AND THEREFORE THE VARIABLE AXIS IS A CONTINUOUS NUMBER LINE. TO DRAW A HISTOGRAM, YOU NEED TO TAKE NOTE OF THE FOLLOWING:

Note:

- I CONSTRUCT A GROUPED FREQUENCY DISTRIBUTION.
- II LOCATE CLASS BOUNDARIES ALONG HORIZONTAL AXIS).
- III THE WIDTH OF THE BAR INDICATES THE CLASS INTERVAL.
- IV THE HEIGHT OF THE BARS INDICATE THE FREQUENCY OF EACH CLASS.

Example 1 THE HISTOGRAM OF THE DATA GIVEN ABOVE WILL LOOK LIKE THE FOLLOWING:

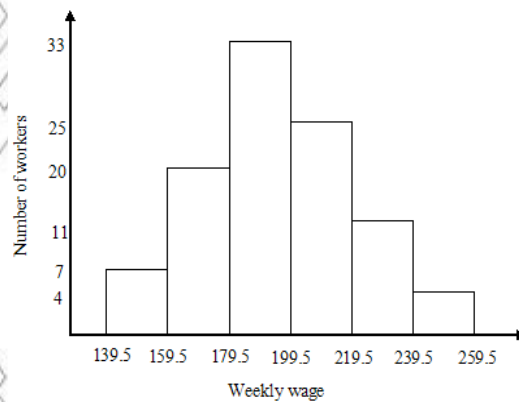


Figure 8.5

Example 2 THE SOIL LABORATORY SECTION OF AN AGRICULTURAL COLLEGE COLLECTED THE FOLLOWING DATA ABOUT THE LENGTH OF A KIND OF EARTHWORM, WHICH PLAGUES THE SURROUNDING FARMS.

LENGTH (CM)	0.5-1.5	1.5-2.5	2.5-3.5	3.5-4.5	4.5-5.5	5.5-6.5	6.5-7.5	7.5-8.5	8.5-9.5
FREQUENCY	4	7	14	20	19	17	10	7	2

Solution THE HISTOGRAM IS GIVEN BELOW. BECAUSE A HISTOGRAM REPRESENTS A CONTINUOUS VARIABLE, IT IS CONTINUOUS.

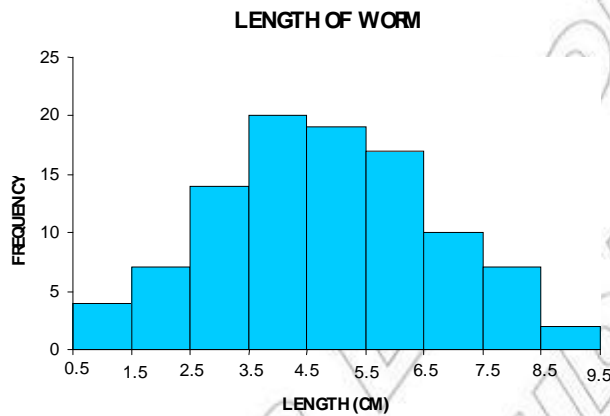


Figure 8.6

II Frequency polygons

THIS IS ANOTHER TYPE OF GRAPH USED TO REPRESENT DATA. IN DRAWING A FREQUENCY POLYGON, YOU PLOT THE MID POINTS (CLASS-MARKS) OF THE CLASS INTERVALS ON THE HORIZONTAL AXIS AND THE CORRESPONDING FREQUENCIES ON THE VERTICAL AXIS. AFTER PLOTTING THESE POINTS, YOU JOIN THEM BY CONSECUTIVE LINE SEGMENTS. THE RESULTING GRAPH IS A FREQUENCY POLYGON.

Example 3 THE FOLLOWING TABLE REPRESENTS MARKS OBTAINED BY STUDENTS IN A TEST. CONSTRUCT A FREQUENCY POLYGON.

Marks	Mid point	Number of students
15 - 20	17.5	3
20 - 25	22.5	17
25 - 30	27.5	10
30 - 35	32.5	5

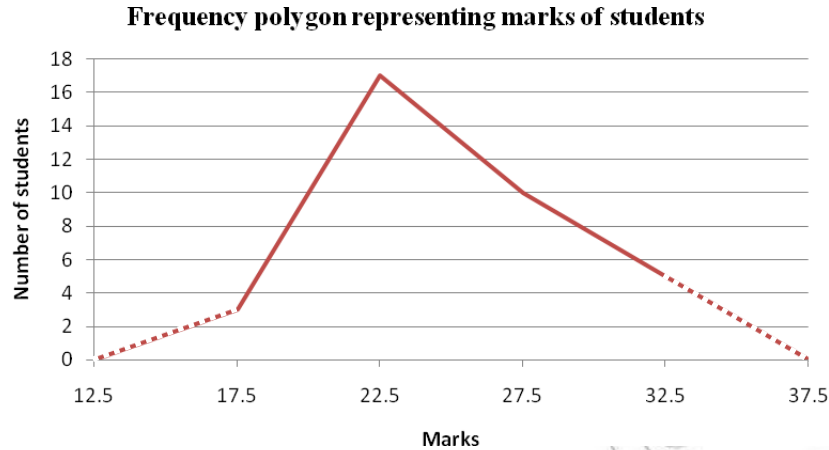


Figure 8.7

Note:

- 1 THE DOTTED LINES AT BOTH ENDS SHOW THAT THE POLYGON IS CLOSED. THEY ARE NEEDED IN ORDER FOR THE POLYGON TO BE CLOSED.
- 2 A FREQUENCY POLYGON WILL BE MORE MEANINGFUL IF IT IS SUPERIMPOSED ON THE SAME GRAPH. THAT IS, IF THE ABOVE MARKS ARE OF SECTION A STUDENTS, WE CAN ADD THE MARKS OF SECTION B STUDENTS AND HENCE COMPARE THEIR PERFORMANCE.
- 3 IT IS POSSIBLE TO DRAW A FREQUENCY POLYGON SIMPLY BY JOINING THE MID POINTS OF EACH BAR. FOR EXAMPLE, THE FREQUENCY POLYGON OF THE GROUPS OF WAGES OF WORKERS IS SHOWN BELOW.

Example 4 IN THE FIGURE BELOW, THE FREQUENCY POLYGON IS DRAWN FROM THE HISTOGRAM IN EXAMPLE 1 ABOVE.

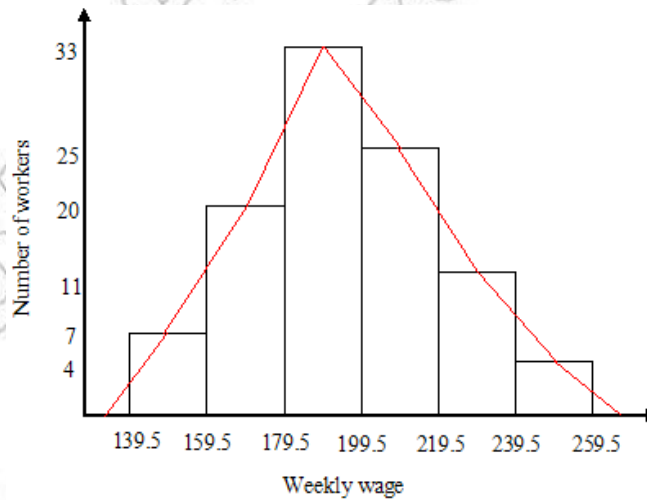


Figure 8.8

III Cumulative frequency curve (Ogive)

TO DRAW A CUMULATIVE FREQUENCY CURVE (OGIVE), YOU CUMULATIVE FREQUENCY TABLE. AN EXAMPLE IS GIVEN BELOW:

Example 5 DRAW A CUMULATIVE FREQUENCY CURVE (OGIVE) FOR THE D ABOVE.

Solution

Marks	Mid point	Number of students (f)	Cumulative Frequency
15 - 20	17.5	3	3
20 - 25	22.5	17	20
25 - 30	27.5	10	30
30 - 35	32.5	5	35

THE CUMULATIVE FREQUENCY ABOVE SHOWS THE NUMBER OF STUDENTS WHO SCORED LE EQUAL TO THE UPPER CLASS BOUNDARY OF THE CORRESPONDING CLASS. FOR INSTANCE, 20 THE NUMBER OF STUDENTS WHOSE SCORE IS LESS THAN OR EQUAL TO 25.

TO DRAW THE OGIVE, YOU PLOT EACH CUMULATIVE FREQUENCY AGAINST ITS UPPER BOUNDARY.

OGIVE CURVE

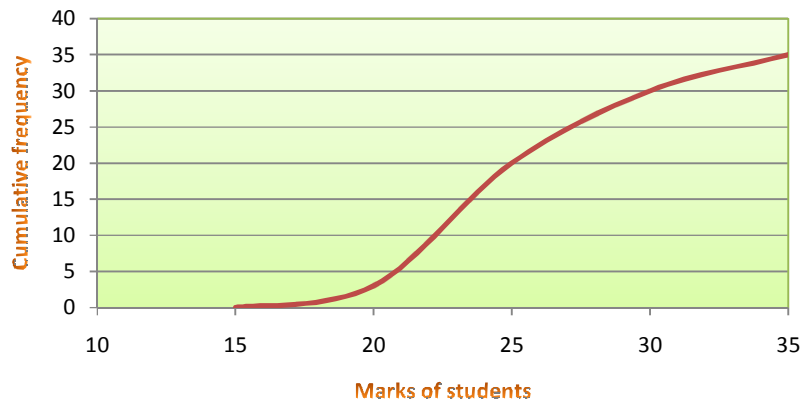


Figure 8.9

HISTOGRAMS OF GROUPED FREQUENCY DISTRIBUTIONS OFTEN DISPLAY A LOW FREQUENCY RISE STEADILY UP TO A PEAK AND THEN DROP DOWN TO A LOW FREQUENCY AGAIN ON THE THE PEAK IS IN THE CENTRE AND THE SLOPES ON EITHER SIDE ARE VIRTUALLY EQUAL TO THEN THE DISTRIBUTION IS SAID TO BE OTHERWISE, THE DISTRIBUTION IS SKWNESS IS LACK OF SYMMETRY IN THE DATA.

FOR A SKEWED DISTRIBUTION, IF THE PEAK LIES TO THE LEFT OF THE CENTRE, THEN THE DIS positively - skewed, AND IF THE PEAK OF THE DISTRIBUTION IS TO THE RIGHT, THE DISTRIBUTION IS SAID TO BE - skewed.

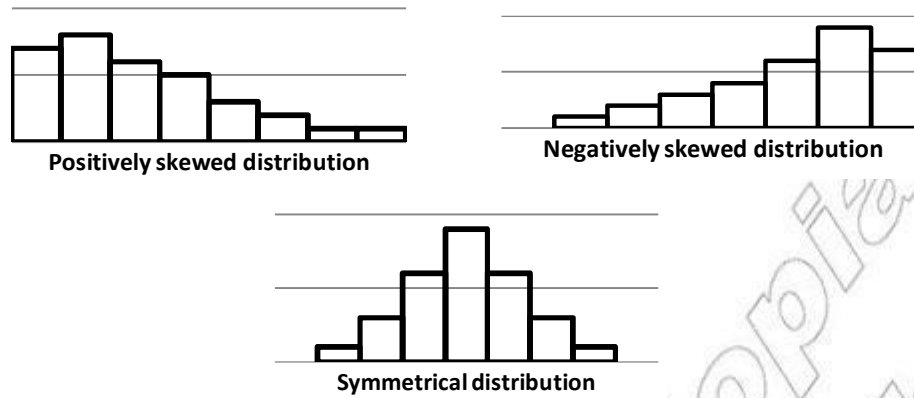


Figure 8.10

IV Frequency curves

FREQUENCY CURVES ARE SIMPLY SMOOTHED HISTOGRAMS.

Example 6 CONSTRUCT A FREQUENCY CURVE FOR THE DISTRIBUTION OF WEEKLY WAGES OF WORKERS GIVEN IN TABLE 8.5

Solution

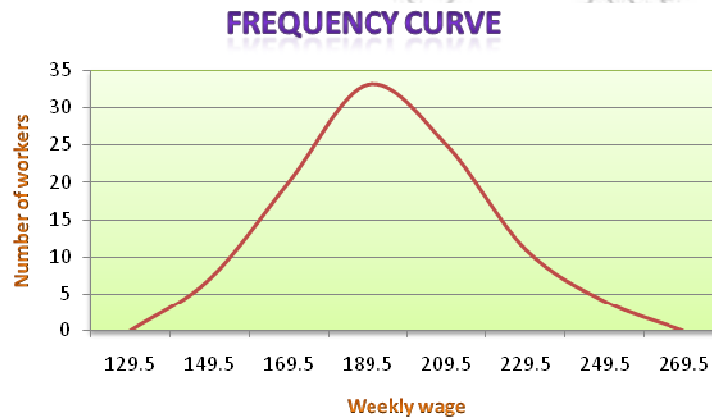


Figure 8.11

FREQUENCY CURVES CAN ALSO BE USED TO SHOW SKEWNESS. FOR THE GROUPED FREQUENCY DISTRIBUTION IN FIGURE 8.10 THE CORRESPONDING FREQUENCY CURVES ARE GIVEN BELOW

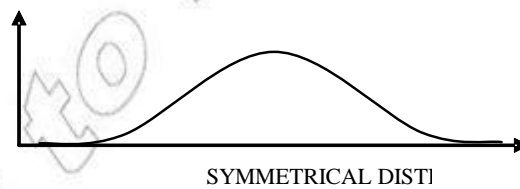


Figure 8.12

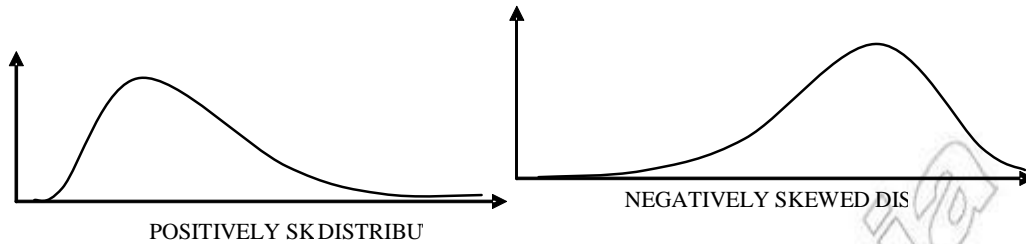


Figure 8.13

Example 7 DRAW THE FREQUENCY CURVE FOR THE DATA BELOW: EXAMPLE 2

Solution THE FREQUENCY CURVE IS GIVEN BELOW:

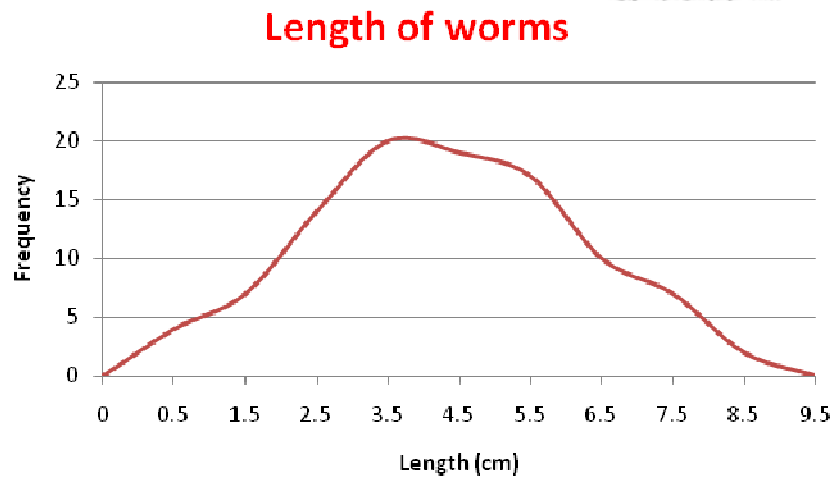


Figure 8.14

Representation of data using diagrams (Charts)

SO FAR, WE DISCUSSED REPRESENTATION OF DATA USING FREQUENCY POLYGONS AND FREQUENCY CURVES. THERE ARE ALSO OTHER FORMS OF DATA REPRESENTATION. HERE, WE REPRESENTATION OF DATA USING BAR CHARTS, LINE GRAPHS AND PIE CHARTS. FIRST OF FOLLOWING:

V Bar charts

ACTIVITY 8.6



- 1 WHAT IS BAR CHART?
- 2 CONSIDERING THE FOLLOWING CHART, EXPLAIN THE DIFFERENCE BETWEEN THIS CHART AND THE HISTOGRAM.

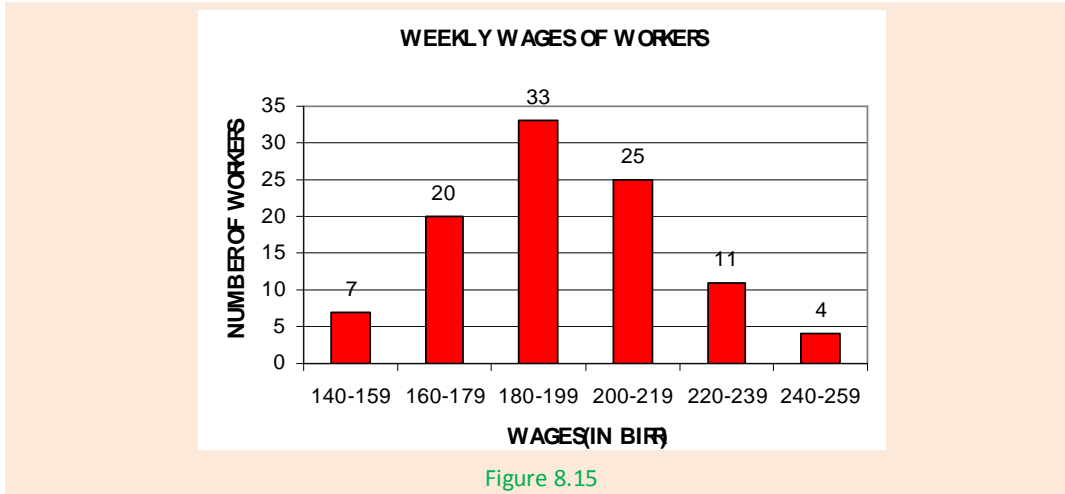


Figure 8.15

BAR CHARTS ARE LIKE HISTOGRAMS IN THAT FREQUENCIES ARE REPRESENTED WITH RECTANGLES BUT, WITH A SPACE BETWEEN EACH BAR. BAR CHARTS ARE ONE OF THE MOST COMMONLY USED REPRESENTATIONS FOUND IN NEWSPAPERS, MAGAZINES AND REPORT PAPERS.

THERE ARE DIFFERENT TYPES OF BAR CHARTS

- ✓ simple bar charts
- ✓ component (subdivided) bar charts
- ✓ grouped (multiple) bar charts
- A Simple bar charts

A SIMPLE BAR CHART IS A TYPE OF BAR CHART REPRESENTS THE FREQUENCIES OF SINGLE ITEMS WITHOUT CONSIDERING THE COMPONENT ITEMS.

Example 8 THE FOLLOWING TABLE DEPICTS TYPES AND QUANTITIES OF SHOES PRODUCED BY A CERTAIN FACTORY FOR FOUR CONSECUTIVE YEARS (IN THOUSANDS)

Year	Boots	Normal	Total
1990	3	7	10
1991	5	10	15
1992	4	6	10
1993	10	15	25

IF YOU CONSIDER THE TOTAL NUMBER OF PAIRS OF SHOES PRODUCED, ITS SIMPLE BAR CHART WILL LOOK LIKE THE FOLLOWING, WHICH RELATES ONLY YEAR AND TOTAL PAIRS OF SHOES PRODUCED YEAR. NOTICE THAT YOU ARE CONSIDERING A SINGLE ITEM WITHOUT CONSIDERING THE COMPONENTS.

Simple Bar chart showing number of pairs of shoes produced in a factory each year

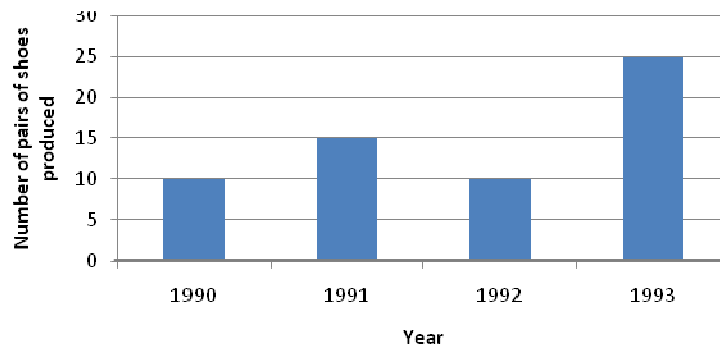


Figure 8.16

In order to draw a bar chart, take the following steps.

- 1 SET HORIZONTAL AND VERTICAL AXES.
- 2 LOCATE DATA VALUES/ CATEGORIES ON THE HORIZONTAL AXIS ON THE VERTICAL AXIS.
- 3 DRAW RECTANGULAR BARS.
- 4 NOTICE THAT THE SPACE BETWEEN EACH BAR MUST BE THE SAME.

YOU CAN ALSO USE MICROSOFT EXCEL OR ANY OTHER SOFTWARE TO DRAW SUCH CHARTS.

B Component bar charts

IN ADDITION TO THE FEATURES OF A SIMPLE BAR CHART, A COMPONENT BAR CHART TAKES INTO ACCOUNT THE RELATIVE CONTRIBUTION OF EACH PART OR COMPONENT TO THE TOTAL. SEE THE COMPONENT BAR CHART FOR THE DATA GIVEN IN THE PREVIOUS EXAMPLE.

Component Bar chart showing number of pairs of shoes produced in a factory each year

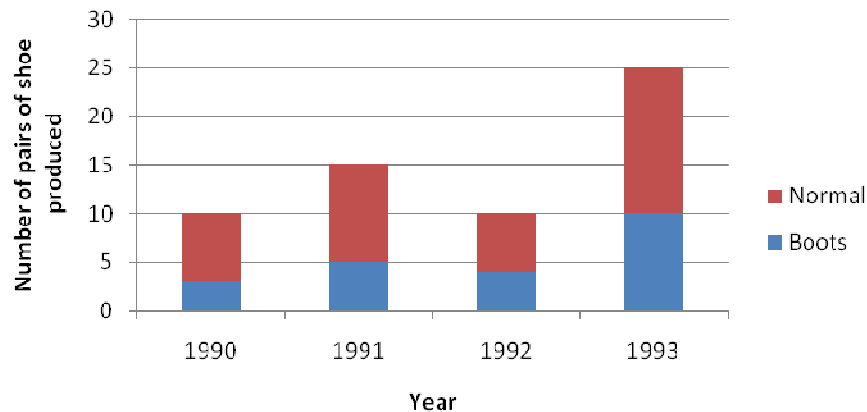


Figure 8.17

Note:
 ✓ THIS COMPONENT BAR CHART DESCRIBES NOT ONLY THE TOTAL NUMBER OF PAIRS OF SHOES BUT ALSO THE TYPES OF SHOES PRODUCED, ONE ON TOP OF THE OTHER SO THAT EACH COLUMN SUMS UP THE TOTAL.

C Multiple bar charts

THESE ARE BAR CHARTS THAT SHOW THE VARIOUS COMPONENTS BY SIDE. THEY HELP TO FACILITATE COMPARISON. THE MULTIPLE BAR CHART FOR THE ABOVE DATA IS GIVEN BELOW.

Multiple Bar chart showing pairs of shoes produced in a factory each year

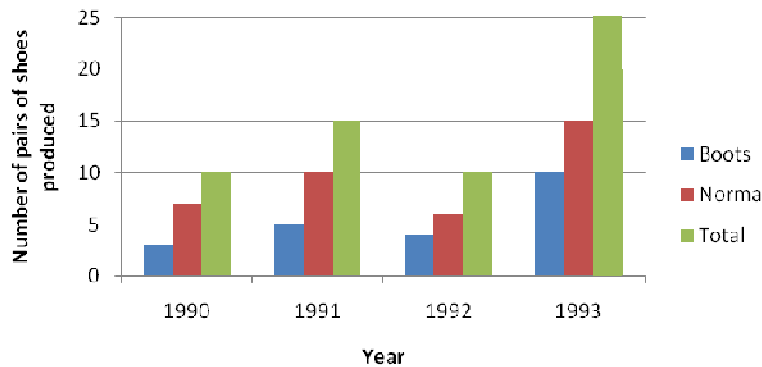


Figure 8.18

M Line graphs

A LINE GRAPH IS ANOTHER USEFUL WAY TO REPRESENT DATA. THE CATEGORIES REPRESENT TIME. SUCH GRAPHS PORTRAY CHANGES IN AMOUNT WITH RESPECT TO TIME BY USING LINE SEGMENTS. THESE GRAPHS ARE USEFUL FOR COMPARING SERIES OF DATA.

Example 9 THE FOLLOWING DATA REPRESENTS DAILY SALES FOR A SHOP.

Days	M	T	W	TH	F	SA
Sales in hundreds of Birr	4	3	6	8	2	5

Sales in hundreds of Birr

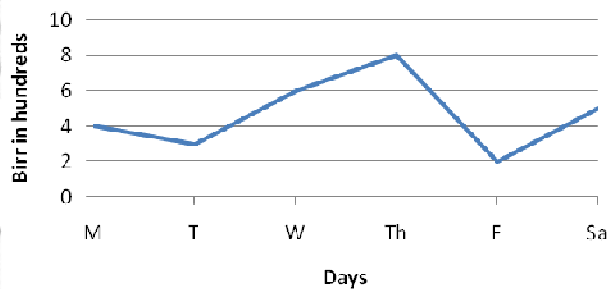


Figure 8.19

IN ORDER TO DRAW A LINE GRAPH, FIRST PLOT EACH QUANTITY AND THEN CONNECT EACH LINE SEGMENT.

VI Pie charts

A PIE CHART IS A PICTORIAL REPRESENTATION OF DATA SUBDIVISIONS IN A CIRCULAR REGION. THE VARIOUS COMPONENTS ARE CONVERTED INTO DEGREES BY TAKING PROPORTION OF 360°.

In order to draw pie chart

- I DRAW A CIRCLE WITH CONVENIENT RADIUS.
- II FIND THE RELATIVE FREQUENCY OF EACH ITEM.
- III CONVERT EACH RELATIVE FREQUENCY INTO AN ANGLE.
- IV DIVIDE THE CIRCLE ACCORDING TO THESE ANGLES.
- V DIFFERENT COMPONENTS APPEAR AS ADJACENT SECTORS OF THE CIRCLE.

Example 10 THE FOLLOWING DATA DEPICTS PREFERRED MODES OF TRANSPORT FOR 100 PEOPLE.

Type of transport	TAXI	BUS	PRIVATE
People who used	35	50	15

TO DRAW A PIE CHART THAT REPRESENTS THE GIVEN DATA, FIRST YOU NEED TO DETERMINE THE RELATIVE FREQUENCY (FROM THE NUMBER OF HOUSEHOLDS OF EACH TYPE OF TRANSPORT TO CALCULATE THE ANGLES:

Type of transport	No of people	Relative frequency	Angle
TAXI	35	$\frac{35}{100} \times 360^\circ =$	126°
BUS	50	$\frac{50}{100} \times 360^\circ =$	180°
PRIVATE CAR	15	$\frac{15}{100} \times 360^\circ =$	54°

TRANSPORT PREFERENCE

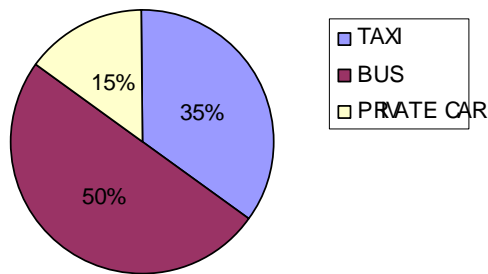


Figure 8.20

Note:

- 1 IN THE PIE CHART, NOTICE THAT THE AREAS OR DIMENSIONS OF THE COMPONENTS ARE PROPORTIONAL TO THE RELATIVE FREQUENCY
- 2 COMPONENTS REPRESENTING EQUAL PERCENTAGES ARE EQUAL IN AREA OF THE CIRCLE.
- 3 PIE CHARTS MAY NOT BE EFFECTIVE IF THERE ARE TOO MANY COMPONENTS

Exercise 8.3

- 1 THE FOLLOWING TABLE DEPICTS AGES OF PEOPLE IN A CERTAIN CITY. CONSTRUCT A HISTOGRAM.

Age (in years)	10 – 15	15 - 20	20 - 25	25 - 30	30 - 35
Number of people	17	23	15	14	12

- 2 DRAW A FREQUENCY POLYGON AND FREQUENCY CURVE FOR THE ABOVE DATA.

Age (in years)	20 - 26	26 - 32	32 - 38	38 - 44	44 - 50
Number of people	8	3	11	7	12

- 3 THE FOLLOWING TABLE REPRESENTS THE COSTS INCURRED IN BUILDING A HOUSE IN THREE MONTHS.

Month	Items			Total
	Cement	Steel	Labour	
MESKERE	70	90	50	210
TIKIMT	80	100	70	250
HIDAR	50	45	45	155

REPRESENT THE ABOVE DATA USING THE THREE TYPES OF BAR CHARTS.

- 4 REPRESENT THE FOLLOWING USING ANY TWO OF THE BAR CHARTS.

Year	Production in tonnes			Total
	Teff	Wheat	Maize	
1996	80	60	70	210
1997	100	150	180	430
1998	150	200	250	600

- 5 THE AGE DISTRIBUTION OF PEOPLE IN A VILLAGE IS AS FOLLOWS. FILL IN THE “DEGREE” COLUMN AND CONSTRUCT A PIE CHART.

Age	Number of people	Degree
UNDER 20	15	
20 – 40	60	
40 – 60	20	
OVER 60	5	

6 DRAW A PIE CHART FOR EACH OF THE FOLLOWING SETS OF

A

Crop	Production in tonnes
TEFF	500
WHEAT	700
MAIZE	800
BARLEY	500

B

Expenditure	Amount (in birr)
RENT	500
TRANSPORT	200
ELECTRICITY	1500
EDUCATION	100

8.4 MEASURES OF CENTRAL TENDENCY AND MEASURES OF VARIABILITY

8.4.1 Measures of Central Tendency

IN PREVIOUS GRADES, YOU STUDIED THE DIFFERENT MEASURES OF CENTRAL TENDENCY (MEAN AND MEDIAN) FOR UNGROUPED AND GROUPED DATA AND MEASURES OF VARIATION THAT IS RANGE, VARIANCE AND STANDARD DEVIATION. IN THIS SUB-UNIT, YOU WILL BRIEFLY REVISIT THESE CONCEPTS WITH THE HELP OF EXAMPLES AND PROCEED TO SEE OTHER MEASURES OF VARIATION SUCH AS INTER-QUARTILE RANGE AND MEAN DEVIATION.

ACTIVITY 8.7



CONSIDERING THE FOLLOWING UNGROUPED DATA: 50, 70, 45, 60, 77, 63, 62, 75, 54, 50, 55, 49, 53 OF THE WEIGHTS IN KG OF 15 STUDENTS, FIND:

- A THE MEAN
- B THE MEDIAN
- C THE RANGE
- D THE FIRST AND THIRD QUANTILES
- E THE MODE AND
- F THE STANDARD DEVIATION.

FROM THIS ACTIVITY, IT IS HOPED THAT YOU HAVE REVISED THE MEASURES OF CENTRAL TENDENCY AND MEASURES OF DISPERSION FOR UNGROUPED DATA. THE SAME APPROACH HOLDS FOR GROUPED DATA AS WELL. SOME EXAMPLES ARE GIVEN BELOW.

Example 1 CONSIDERING THE FOLLOWING GROUPED FREQUENCY DISTRIBUTION OF 150 STUDENTS:

Weight (in k.g) x	Class mid point (m)	Number of students (f)
48 - 49	48.5	5
50 - 51	50.5	23
52 - 53	52.5	15
54 - 55	54.5	25
56 - 57	56.5	8
58 - 59	58.5	7
		$\sum f = 83$

CALCULATE THE MEAN THE MEDIAN THE RANGE THE STANDARD DEVIATION

Solution **A** THE MEAN IS $\frac{\sum fm}{\sum f} = \frac{4415.5}{83} = 53.20$ KC

B THE MEDIAN IS

$$m_d = L + \left(\frac{\left(\frac{n}{2} - cf_b \right)}{f_c} \right) w$$

SO MEDIAN $51.5 + \left(\frac{\left(\frac{83}{2} - 28 \right)}{15} \right) \times 2 = 51.5 + \frac{(41.5 - 28)}{15} \times 2 = 51.5 + \frac{(13.5)}{15} \times 2$
 $= 51.5 + 1.8 = 53.3$ KC

C THE RANGE IS THE DIFFERENCE BETWEEN UPPER CLASS BOUNDARY OF THE HIGHEST CLASS (59.5) AND THE LOWER CLASS BOUNDARY OF THE LOWEST CLASS (47.5). THE RANGE IS $R = 59.5 - 47.5 = 12$

D THE STANDARD DEVIATION:

Weight (in k.g) x	Class mid point (m)	Number of students (f)	$x_i - 53.20$	$(x_i - 53.69)^2$	$f_i(x_i - 53.69)^2$
48 - 49	48.5	5	-4.70	22.09	110.45
50 - 51	50.5	23	-2.70	7.29	167.67
52 - 53	52.5	15	-0.70	0.49	7.35
54 - 55	54.5	25	1.30	1.69	42.25
56 - 57	56.5	8	3.30	10.89	87.12
58 - 59	58.5	7	5.30	28.09	196.63
$\sum f = 83$			$\sum_{i=1}^n f_i (x_i - \bar{x})^2 = 611.47$		

$$\text{THUS } s = \sqrt{\frac{\sum_{i=1}^n f_i (x_i - \bar{x})^2}{\sum_{i=1}^n f_i}} = \sqrt{\frac{611.47}{83}} = \sqrt{7.37} = 2.71 \text{ KC}$$

Example 2 CALCULATE THE Q₁ AND Q₃ OF THE FOLLOWING DATA.

x	f	cf
10 - 19	3	3
20 - 29	5	8
30 - 39	14	22
40 - 49	7	29

Solution

I Q_1 IS THE $\left(\frac{29}{4}\right)^{\text{TH}}$ ITEM $\left(7.25\right)^{\text{TH}}$ ITE IN THE 2ND CLASS

$$Q_1 = 19.5 + \left(\frac{7.25 - 3}{5}\right)10 = 19.5 + 8.5 = 28$$

II Q_2 IS THE $\left(\frac{2 \times 29}{4}\right)^{\text{TH}}$ ITEM $\left(14.5\right)^{\text{TH}}$ ITE IN THE 3RD CLASS

$$Q_2 = 29.5 + \left(\frac{14.5 - 8}{14}\right)10 = 29.5 + 4.64 = 34.14$$

III Q_3 IS THE $\left(\frac{3 \times 29}{4}\right)^{\text{TH}}$ ITEM $\left(21.75\right)^{\text{TH}}$ ITE IN THE 3RD CLASS

$$Q_3 = 29.5 + \left(\frac{21.75 - 8}{14}\right)10 = 29.5 + 9.82 = 39.32$$

Example 3 THE FOLLOWING IS THE AGE DISTRIBUTION OF STUDENTS IN ONE OF THE MODAL AGE.

age	Number of students
10 - 14	2
15 - 19	7
20 - 24	9
25 - 29	4
30 - 34	3

Solution THE MODAL CLASS IS 20-24TH CLASS BECAUSE ITS FREQUENCY IS THE HIGHEST. THE LOWER CLASS BOUNDARY OF THIS CLASS IS 19.5

$$\therefore L = 19.5$$

$$w = 24.5 - 19.5 = 5, \Delta_1 = 9 - 7 = 2, \Delta_2 = 9 - 4 = 5$$

$$\text{MODE} = L + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right)w = 19.5 + \left(\frac{2}{2+5}\right)5 = 19.5 + \frac{10}{7} = 19.5 + 1.43 = 20.93$$

Exercise 8.4

- 1 CALCULATE THE ARITHMETIC MEAN OF EACH DATA SET BELOW
A 76, 78, 69, 75, 84, 92, 11, 81, 10, 95 **B** 22, 22, 22, 22, 22, 22, 22, 22
- 2 IF THE MEAN OF 4, 7, 8, 6, 5 IS 6, THEN, FIND THE MEAN OF 8, 7 + 8, 8 + 8, 6 + 8, 5 + 8.
- 3 IF THE MEAN OF 5, 6, 10, 15, 19 IS 11 THEN, FIND THE MEAN OF $2 \times 5, 2 \times 6, 2 \times 10, 2 \times 15, 2 \times 19$.
- 4 IF THE MEAN OF A, B, C, D IS 5 THEN, FIND THE MEAN OF $3A + 3, 3B + 3, 3C + 3, 3D + 3$.
- 5 FIND THE MEAN OF EACH OF THE FOLLOWING DATA.

A

<i>x</i>	7	10	11	15	19
<i>f</i>	3	2	4	8	6

B

Marks	20	30	40	50	60	70
Number of students	8	12	20	10	6	4

C

Marks	0 - 9	10 - 19	20 - 29	30 - 39	40 - 49
<i>f</i>	5	10	8	13	4

- 6 FIND THE MEDIAN AND MODE OF EACH OF THE FOLLOWING DATA SETS

A 2, 7, 6, 8, 10, 1

B

<i>x</i>	10	12	15	16
<i>f</i>	4	6	8	3

C

Age	1 - 3	4 - 6	7 - 9	10 - 12	13 - 15
<i>f</i>	2	1	8	17	11

- 7 FIND Q_1, Q_2 AND Q_3 OF EACH OF THE FOLLOWING.

A 18, 11, 26, 20, 16, 8, 22, 23, 8, 12, 15, 13

B

<i>x</i>	5 - 9	10 - 14	15 - 19	20 - 24	25 - 29
<i>f</i>	3	4	6	7	3

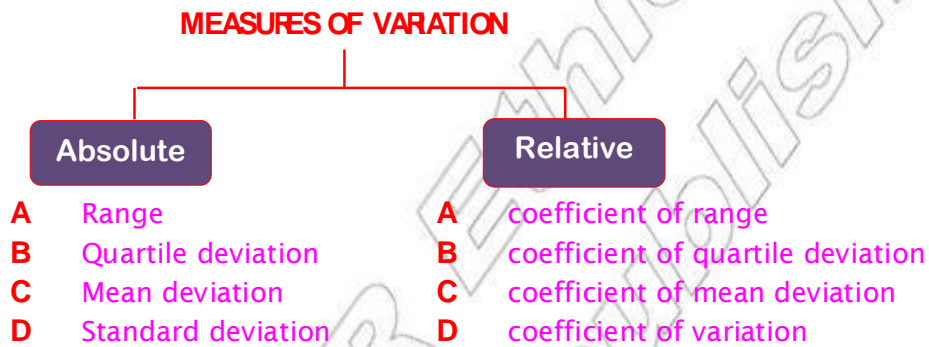
8.4.2 Further on Measures of Variation

IN PREVIOUS GRADES, YOU STUDIED THE DEFINITION OF VARIATION AND MEASURES OF SUCH AS RANGE, VARIANCE AND STANDARD DEVIATION. IN THIS SUB UNIT, YOU ARE GOING TO STUDY ADDITIONAL MEASURES OF VARIATION, NAMELY, MEAN DEVIATION AND SOME RELATIVE MEASURES OF VARIATION SUCH AS THE COEFFICIENT OF VARIATION.

RECALL THAT A MEASURE OF VARIATION CAN BE DEFINED IN EITHER OF THE FOLLOWING TWO WAYS:

- A** THE DEGREE TO WHICH NUMERICAL DATA TENDS TO SPREAD;
- B** THE SCATTER OR VARIATION OF VARIABLES ABOUT A CENTRAL VALUE.

VARIATION CAN BE MEASURED EITHER ABSOLUTELY OR RELATIVELY.



Note:

- ✓ ABSOLUTE MEASURES ARE EXPRESSED IN CONCRETE UNITS, WHICH THE DATA VALUE IS EXPRESSED, E.G. BIRR, KG, M, ETC.
- ✓ A RELATIVE MEASURE OF VARIATION IS THE RATIO OF ABSOLUTE VARIATION TO ITS CORRESPONDING AVERAGE. IT IS A PURE NUMBER THAT IS INDEPENDENT OF THE MEASUREMENT.

Mean deviation (MD)

WHEN WE CALCULATED STANDARD DEVIATION, YOU HAD TO TAKE THE SQUARE ROOT OF THE SUM OF THE SQUARES OF THE DEVIATIONS OF EACH OBSERVATION FROM THE MEAN. ANOTHER MEASURE OF DEVIATION THAT ALSO CONSIDERS ALL MEMBERS OF A DATA SET IS THE MEAN DEVIATION, WHICH YOU ARE GOING TO SEE NOW.

ACTIVITY 8.8



THE FOLLOWING DATA REPRESENTS MATHEMATICS EXAMINATION MARKS OUT OF 10 OF TEN STUDENTS.

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

- 1** FIND THE MEAN OF THE DATA.
- 2** FIND THE DEVIATION OF EACH VALUE FROM THE MEAN.

- 3 DETERMINE THE MEAN OF THESE DEVIATIONS.
- 4 OBSERVE THAT THIS MEAN IS 0. WHAT WILL THE MEAN BE IF YOU CONSIDER THE ABSOLUTE VALUES OF EACH DEVIATION?

Definition 8.1

MEAN DEVIATION IS THE SUM OF DEVIATIONS (IN ABSOLUTE VALUE) OF EACH ITEM FROM AVERAGE DIVIDED BY THE NUMBER OF ITEMS. IT CAN BE CONSIDERED AS THE MEAN OF ABSOLUTE VALUES OF DEVIATIONS OF EACH VALUE FROM A CENTRAL VALUE.

Note:

A DEVIATION MAY BE TAKEN FROM THE MEAN, MEDIAN OR MODE.

YOU WILL NOW SEE HOW TO CALCULATE MEAN DEVIATION ABOUT THE MEAN, THE MEDIAN OR MODE FOR UNGROUPED DATA, FOR DISCRETE FREQUENCY DISTRIBUTIONS AND FOR GROUPED DATA.

1 Mean deviation for ungrouped data

I Mean deviation from the mean MD(\bar{x})

TO CALCULATE THE MEAN DEVIATION FROM THE MEAN, FOLLOW THE FOLLOWING STEPS.

- Step 1:** FIND THE MEAN OF THE DATA SET.
- Step 2:** FIND THE DEVIATION OF EACH ITEM FROM THE MEAN (SINCE MEAN DEVIATION ASSUMES ABSOLUTE VALUE).
- Step 3:** FIND THE SUM OF THE DEVIATIONS.
- Step 4:** DIVIDE THE SUM BY THE TOTAL NUMBER OF ITEMS IN THE DATA SET.

$$MD(\bar{x}) = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + |x_3 - \bar{x}| + \dots + |x_n - \bar{x}|}{n} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

II Mean deviation about the median MD(m_d)

TO CALCULATE THE MEAN DEVIATION FROM THE MEDIAN, USE THE MEDIAN IN PLACE OF THE MEAN AND PROCEED IN THE SAME WAY, AS FOLLOWS.

- Step 1:** FIND THE MEDIAN OF THE DATA SET.
- Step 2:** FIND THE ABSOLUTE DEVIATION OF EACH ITEM FROM THE MEDIAN.
- Step 3:** FIND THE SUM OF THE DEVIATIONS.
- Step 4:** DIVIDE THE SUM BY THE TOTAL NUMBER OF ITEMS IN THE DATA SET.

$$MD(m_d) = \frac{|x_1 - m_d| + |x_2 - m_d| + |x_3 - m_d| + \dots + |x_n - m_d|}{n} = \frac{\sum_{i=1}^n |x_i - m_d|}{n}$$

III Mean deviation about the mode MD(m_0)

AGAIN PROCEED IN A SIMILAR WAY:

- Step 1:** FIND THE MODE OF THE DATA SET.
- Step 2:** FIND THE ABSOLUTE DEVIATION OF EACH ITEM FROM THE MODE.
- Step 3:** FIND THE SUM OF THE DEVIATIONS.
- Step 4:** DIVIDE THE SUM BY THE TOTAL NUMBER OF ITEMS IN THE DATA SET.

$$MD(m_0) = \frac{|x_1 - m_0| + |x_2 - m_0| + |x_3 - m_0| + \dots + |x_n - m_0|}{n} = \frac{\sum_{i=1}^n |x_i - m_0|}{n}$$

Example 4 FIND THE MEAN DEVIATION ABOUT THE MEAN OF THE DATA SET AND M FOLLOWING DATA.

5, 8, 8, 11, 11, 11, 14, 16

A MEAN DEVIATION ABOUT THE MEAN.

Step 1: CALCULATE THE MEAN OF THE DATA SET.

$$\bar{x} = \frac{5+8+8+11+11+11+14+16}{8} = \frac{84}{8} = 10.5$$

Step 2: FIND THE ABSOLUTE DEVIATION OF EACH ITEM FROM THE MEAN.

x	$ x - \bar{x} $
5	5.5
8	2.5
8	2.5
11	0.5
11	0.5
11	0.5
14	3.5
16	5.5
$\sum x - \bar{x} = 21$	

Step 3: FIND THE SUM OF THE DEVIATIONS, WHICH IS 21.

Step 4: DIVIDE THE SUM BY THE TOTAL NUMBER OF ITEMS IN THE DATA SET.

$$MD(\bar{x}) = \frac{\sum |x - \bar{x}|}{n} = \frac{21}{8} = 2.625$$

B MEAN DEVIATION ABOUT THE MEDIAN, MD(MD)

Step 1: CALCULATE THE MEDIAN OF THE DATA SET.

$$M_d = \frac{\left(\frac{8}{2}\right)^{\text{TH}} \text{ ITEM} + \left(\frac{8}{2} + 1\right)^{\text{TH}} \text{ ITEM}}{2} = \frac{4^{\text{TH}} \text{ ITEM} + 5^{\text{TH}} \text{ ITEM}}{2} = \frac{11 + 11}{2} = 11$$

Step 2: FIND THE ABSOLUTE DEVIATION OF EACH DATA ITEM FROM THE MEDIAN

x	$ x - m_d $
5	6
8	3
8	3
11	0
11	0
11	0
14	3
16	5
$\sum x - m_d = 20$	

Step 3: FIND THE SUM OF THE DEVIATIONS, WHICH IS 20.

Step 4: DIVIDE THE SUM BY THE TOTAL NUMBER OF ITEMS IN THE DATA SET

$$MD(m_d) = \frac{\sum |x - m_d|}{n} = \frac{20}{8} = 2.5$$

C MEAN DEVIATION ABOUT MODE, MD(MD)

Step 1: CALCULATE (IDENTIFY) THE MODE OF THE DATA SET MODE

Step 2: FIND THE ABSOLUTE DEVIATION OF EACH DATA ITEM FROM THE MODE

x	$ x - m_0 $
5	6
8	3
8	3
11	0
11	0
11	0
14	3
16	5
$\sum x - m_0 = 20$	

Step 3: FIND THE SUM OF THE DEVIATIONS, WHICH IS 20.

Step 4: DIVIDE THE SUM BY THE TOTAL NUMBER OF ITEMS IN THE DATA SET

$$MD(m_0) = \frac{\sum |x - m_0|}{n} = \frac{20}{8} = 2.5$$

2 Mean deviation for discrete frequency distributions

TO CALCULATE THE MEAN DEVIATION FOR A DISCRETE DISTRIBUTION ABOUT THE MEAN, THE MEDIAN AND THE MODE YOU TAKE SIMILAR STEPS AS IN THE PROCESS FOR DISCRETE DATA.

IF $x_1, x_2, x_3, \dots, x_n$ ARE VALUES WITH CORRESPONDING FREQUENCIES $f_1, f_2, f_3, \dots, f_n$ THE MEAN DEVIATION IS GIVEN AS FOLLOWS.

I Mean deviation about the mean $MD(\bar{x})$

Step 1: FIND THE MEAN OF THE DATA SET.

Step 2: FIND THE ABSOLUTE DEVIATION OF EACH ITEM FROM THE MEAN.

Step 3: MULTIPLY EACH DEVIATION BY ITS CORRESPONDING FREQUENCY.

Step 4: FIND THE SUM OF THESE DEVIATIONS MULTIPLIED BY THEIR FREQUENCIES.

Step 5: DIVIDE THE SUM BY THE SUM OF THE FREQUENCIES.

FOLLOWING THE STEPS OUTLINED ABOVE, YOU WILL GET THE MEAN DEVIATION ABOUT THE MEAN AS FOLLOWS.

$$MD(\bar{x}) = \frac{f_1|x_1 - \bar{x}| + f_2|x_2 - \bar{x}| + f_3|x_3 - \bar{x}| + \dots + f_n|x_n - \bar{x}|}{f_1 + f_2 + f_3 + \dots + f_n} = \frac{\sum_{i=1}^n f_i |x_i - \bar{x}|}{\sum_{i=1}^n f_i}$$

II Mean deviation about the median $MD(m_d)$

HERE, WE SIMPLY NEED TO REPLACE THE MEAN BY THE MEDIAN AND FOLLOW EACH STEP AS ABOVE. THIS WILL GIVE US THE MEAN DEVIATION ABOUT THE MEDIAN TO BE:

$$MD(m_d) = \frac{f_1|x_1 - m_d| + f_2|x_2 - m_d| + f_3|x_3 - m_d| + \dots + f_n|x_n - m_d|}{f_1 + f_2 + f_3 + \dots + f_n} = \frac{\sum_{i=1}^n f_i |x_i - m_d|}{\sum_{i=1}^n f_i}$$

III Mean deviation about the mode $MD(m_o)$

THE STEPS THAT WE NEED TO FOLLOW HERE ARE SIMILAR TO THE STEPS FOR THE MEAN AND MEDIAN. WE WILL USE THE MODE INSTEAD OF THE MEAN OR THE MEDIAN. FOLLOWING THE STEPS, WE WILL GET THE MEAN DEVIATION ABOUT THE MODE TO BE:

$$MD(m_o) = \frac{f_1|x_1 - m_o| + f_2|x_2 - m_o| + f_3|x_3 - m_o| + \dots + f_n|x_n - m_o|}{f_1 + f_2 + f_3 + \dots + f_n} = \frac{\sum_{i=1}^n f_i |x_i - m_o|}{\sum_{i=1}^n f_i}$$

Example 5 FIND THE MD OF THE FOLLOWING DATA ABOUT THE MEAN, MEDIAN AND THE MODE

x	f	cf
9	3	3
15	5	8
21	10	18
27	12	30
33	7	37
39	3	40

1 CALCULATING THE MEAN, THE MEDIAN AND THE MODE WHICH SET

A THE MEAN $= \bar{x} = \frac{3 \times 9 + 5 \times 15 + 10 \times 21 + 12 \times 27 + 7 \times 33 + 3 \times 39}{3 + 5 + 10 + 12 + 7 + 3} = \frac{984}{40} = 24.6$

B THE MEDIAN $= m_d = \frac{\left(\frac{40}{2}\right)^{th} + \left(\left(\frac{40}{2}\right) + 1\right)^{th}}{2} = \frac{20^{th} + 21^{th}}{2} = \frac{27 + 27}{2} = 27$ AND

C THE MODE $m_0 = 27$

2 YOU CALCULATE THE DEVIATIONS FROM THE MEAN, THE MEDIAN AND THE MODE

x	f	DEVIATION ABOUT THE MEAN		DEVIATION ABOUT THE MEDIAN		DEVIATION ABOUT THE MODE	
		$ x - \bar{x} $	$f x - \bar{x} $	$ x - m_d $	$f x - m_d $	$ x - m_0 $	$f x - m_0 $
9	3	15.6	46.8	18	54	18	54
15	5	9.6	48	12	60	12	60
21	10	3.6	36	6	60	6	60
27	12	2.4	28.8	0	0	0	0
33	7	8.4	58.8	6	42	6	42
39	3	14.4	43.2	12	36	12	36
	40		261.6		252		252

3 FIND THE SUM OF THE DEVIATIONS AND DIVIDE BY THE TOTAL FREQUENCIES TO GET THE MEAN DEVIATIONS WHICH WILL BE:

I $MD(\bar{x}) = \frac{\sum f|x - \bar{x}|}{\sum f} = \frac{261.6}{40} = 6.54$

II $MD(m_d) = \frac{\sum f|x - m_d|}{\sum f} = \frac{252}{40} = 6.3$

III $MD(m_0) = \frac{\sum f|x - m_0|}{\sum f} = \frac{252}{40} = 6.3$

3 Mean deviation for grouped frequency distributions

FOR CONTINUOUS GROUPED FREQUENCY DISTRIBUTIONS, MEAN DEVIATION IS CALCULATED IN THE SAME WAY AS ABOVE EXCEPT THAT f_i IS SUBSTITUTED BY THE MIDPOINT OF EACH CLASS (m_i) AND

$$\therefore MD(\bar{x}) = \frac{\sum_{i=1}^n f_i |m_i - \bar{x}|}{\sum_{i=1}^n f_i}, \quad MD(m_d) = \frac{\sum_{i=1}^n f_i |m_i - m_d|}{\sum_{i=1}^n f_i}, \quad MD(m_0) = \frac{\sum_{i=1}^n f_i |m_i - m_0|}{\sum_{i=1}^n f_i}$$

Example 6 FIND THE MEAN DEVIATION ABOUT THE MEAN, THE MEAN DEVIATION FOR THE FOLLOWING

x	0 – 5	6 – 11	12 – 17	18 – 23	24 – 29
f	5	8	7	10	3

Solution

1 FIRST, YOU HAVE TO FIND THE MEAN, MODE AND MEDIAN OF THE DISTRIBUTION

x	f	m	fm	cf
0 – 5	5	2.5	12.5	5
6 – 11	8	8.5	68	13
12 – 17	7	14.5	101.5	20
18 – 23	10	20.5	205	30
24 – 29	3	26.5	79.5	33
$\sum f = 33$		$\sum fm = 466.5$		

A MEAN = $\frac{\sum fm}{\sum f} = \frac{466.5}{33} = 14.14$

B MEDIAN = $L + \left(\frac{\frac{n}{2} - cf_b}{f_c} \right) w = 11.5 + \left(\frac{(16.5 - 13)}{7} \right) 6 = 11.5 + 3 = 14.5$

C MODE = $L + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) w = 17.5 + \left(\frac{3}{3 + 7} \right) 6 = 17.5 + 1.8 = 19.3$

2 DETERMINE THE DEVIATIONS AND CALCULATE THE MEAN DEVIATIONS.

x	f	m	$ m - \bar{x} $	$f m - \bar{x} $	$ m - m_d $	$f m - m_d $	$ m - m_0 $	$f m - m_0 $
0 – 5	5	2.5	11.64	58.20	12	60	16.8	84
6 – 11	8	8.5	5.64	45.12	6	48	10.8	86.4
12 – 17	7	14.5	0.36	2.52	0	0	4.8	33.6
18 – 23	10	20.5	6.36	63.6	6	60	1.2	12
24 – 29	3	26.5	12.36	37.08	12	36	7.2	21.6
$\sum f = 33$		$\sum f m - \bar{x} = 206.52$		$\sum f m - m_d = 204$		$\sum f m - m_0 = 237.6$		

THE MEAN DEVIATION WILL THEN BE

A MEAN DEVIATION ABOUT THE MEAN

$$MD(\bar{x}) = \frac{\sum f |m - \bar{x}|}{\sum f} = \frac{206.52}{33} = 6.26$$

B MEAN DEVIATION ABOUT MEDIAN

$$MD(m_d) = \frac{\sum f |m - m_d|}{\sum f} = \frac{204}{33} = 6.18$$

C MEAN DEVIATION ABOUT THE MODE

$$MD(m_0) = \frac{\sum f |m - m_0|}{\sum f} = \frac{237.6}{33} = 7.2$$

MEAN DEVIATION CAN BE USEFUL FOR APPLICATIONS. FOR THE “ARITHMETIC MEAN” YOU TAKE THE DEVIATION ABOUT THE MEAN, IF OUR AVERAGE IS “MEDIAN” THEN YOU TAKE THE DEVIATION ABOUT THE MEDIAN, AND IF OUR AVERAGE IS THE “MODE”, YOU TAKE MEAN DEVIATION ABOUT THE MODE

TO DECIDE WHICH ONE OF THE MEAN DEVIATIONS TO USE IN A GIVEN SITUATION CONSIDER THE FOLLOWING POINTS: IF THE DEGREE OF VARIABILITY IN A SET OF DATA IS NOT VERY HIGH, USE OF THE MEAN DEVIATION ABOUT THE MEAN IS COMPARATIVELY THE BEST FOR INTERPRETATION. WHENEVER THERE IS AN EXTREME VALUE THAT CAN AFFECT THE MEAN, MEAN DEVIATION ABOUT THE MEDIAN IS PREFERABLE.

MEAN DEVIATION, THOUGH IT HAS SOME ADVANTAGES, IS NOT COMMONLY USED FOR INTERPRETATION. RATHER, IT IS THE STANDARD DEVIATION THAT IS COMMONLY USED AND WHICH TENDS TO BE THE BEST MEASURE OF VARIATION.

Advantages of mean deviation

COMPARED TO RANGE AND QUARTILE DEVIATIONS, MEAN DEVIATION HAS THE FOLLOWING ADVANTAGES: RANGE AND INTER-QUARTILE RANGES (DISCUSSED BELOW) CONSIDER ONLY TWO VALUES; MEAN DEVIATION TAKES EACH VALUE INTO CONSIDERATION.

Limitation

BY TAKING ABSOLUTE VALUE OF DEVIATION, IT IGNORES SIGNS OF DEVIATION WHICH VIOLATES THE RULES OF ALGEBRA.

Exercise 8.5

1 CALCULATE THE MEAN DEVIATION ABOUT THE MEAN, MEDIAN AND MODE OF EACH OF THE FOLLOWING DATA SETS:

A 19, 15, 12, 20, 15, 6, 10

B 5, 6, 7, 9, 10, 10, 11, 12, 13, 17

2 CALCULATE THE MEAN AND MEDIAN ABOUT THE MEAN AND MEDIAN FOR THESE DATA SETS:

A

x	1	2	3	4	5
f	3	1	4	2	1

B

x	12	13	14	15	16	17
f	4	11	3	8	5	4

C

x	0 - 4	5 - 9	10 - 14	15 - 19	20 - 24
f	3	5	7	4	3

D

x	10 - 14	15 - 19	20 - 24	25 - 29
f	5	25	10	4

Range and inter-quartile range (IQR)

ACTIVITY 8.9



THE FOLLOWING ARE THE AVERAGED DAILY TEMPERATURES IN WEEK FOR TWO CITIES, A AND B.

CITY A: 15, 16, 16, 10, 17, 20, 14

CITY B: 13, 16, 15, 15, 14, 16, 17

- 1 FIND THE FIRST AND THE THIRD QUARTILES FOR EACH CITY
- 2 DETERMINE $Q_3 - Q_1$ FOR EACH CITY.
- 3 COMPARE WHICH CITY HAS A HIGHER VARIATION IN TEMPERATURE

IN THE PREVIOUS SUB-UNIT, WE MENTIONED RANGE AS THE DIFFERENCE BETWEEN THE HIGHEST AND THE LOWEST VALUES IN A DATA SET. SOMETIMES, IT MAY NOT BE POSSIBLE TO GET THE RANGE, ESPECIALLY IN OPEN ENDED DATA, WHERE HIGHEST OR LOWEST VALUE MAY BE UNKNOWN. IT MAY SOMETIMES ALSO BE TRUE THAT THE RANGE IS HIGHLY AFFECTED BY EXTREME VALUES. UNDER SUCH CIRCUMSTANCES, IT MAY BE OF INTEREST TO MEASURE THE DIFFERENCE BETWEEN THE THIRD QUARTILE AND THE FIRST QUARTILE, WHICH IS CALLED THE INTER-QUARTILE RANGE. INTER-QUARTILE RANGE IS A MEASURE OF VARIATION WHICH OVERCOMES THE LIMITATIONS OF RANGE. IT IS DEFINED AS FOLLOWS:

$$IQR = Q_3 - Q_1 \quad (\text{DIFFERENCE BETWEEN UPPER AND LOWER QUARTILES})$$

Example 7 CONSIDER THE FOLLOWING TWO SETS OF DATA:

A: 2, 7, 7, 7, 7, 7, 7, 10

B: 2, 3, 5, 8, 9, 10

THE RANGE OF A AND B ARE $10 - 2 = 8$ AND $10 - 2 = 8$

FROM WHICH YOU SEE THAT THEY HAVE THE SAME RANGE HOWEVER, IF YOU OBSERVE THE TWO SETS OF DATA, YOU CAN SEE THAT DATA B IS MORE VARIABLE THAN DATA A.

Example 8 CALCULATE THE IQR OF A AND B.

Solution

For data A:

$$Q_1 = \left(\frac{8+1}{4}\right)^{\text{TH}} \text{ ITEM} = (2.25)^{\text{TH}} \text{ ITEM WHICH IS 7, AND}$$

$$Q_3 = \left(\frac{3(8+1)}{4}\right)^{\text{TH}} \text{ ITEM} = 6.75^{\text{TH}} \text{ ITEM WHICH IS 7.}$$

$$\therefore \text{INTER-QUARTILE RANGE (IQR)} = 7 - 0$$

For Data B:

$$Q_1 = \left(\frac{6+1}{4}\right)^{\text{TH}} \text{ ITEM} = (1.75)^{\text{TH}} \text{ ITEM WHICH IS 2.75, AND}$$

$$Q_3 = \left[\frac{3(6+1)}{4}\right]^{\text{TH}} \text{ ITEM} = (5.25)^{\text{TH}} \text{ ITEM WHICH IS 9.25.}$$

$$\therefore \text{INTER - QUARTILE RANGE (IQR)} = Q_3 - Q_1 = 9.25 - 2.75 = 6.5$$

FROM WHICH YOU SEE CLEARLY THAT DATA B POSSESSES HIGHER VARIABILITY THAN DATA A. THE GREATER THE MEASURE OF VARIATION, THE GREATER THE VARIABILITY (DISPERSION) OF THE DATA.

\therefore SINCE $IQR_B > IQR_A$ DATA B IS MORE VARIABLE

Limitation of Inter-Quartile Range

- 1 IT ONLY DEPENDS ON TWO VALUES AND IT DOESN'T CONSIDER THE VARIABILITY OF EACH ITEM IN THE DATA SET.
- 2 IT IGNORES 50% OF THE DATA (THE TOP 25% ABOVE THE BOTTOM 25% BELOW Q_2) IT ONLY CONSIDERS THE MIDDLE 50% OF VALUES BETWEEN Q_1 AND Q_3 .

Standard deviation

YOU HAVE ALREADY SEEN HOW TO CALCULATE THE STANDARD DEVIATION FOR UNGROUPED AND GROUPED FREQUENCY DISTRIBUTIONS. YOU HAVE ALSO SEEN OTHER MEASURES OF DISPERSION

ACTIVITY 8.10

A CERTAIN SHOP HAS REGISTERED THE FOLLOWING DATA ON DAILY (IN 100 BIRR) FOR TEN CONSECUTIVE DAYS.

30 45 54 60 25 35 42 80 70 40

- 1 CALCULATE THE DIFFERENT MEASURES OF DISPERSION (IQR, MEAN DEVIATION AND STANDARD DEVIATION).
- 2 DISCUSS SIMILARITIES AND DIFFERENCES BETWEEN THESE MEASURES OF DISPERSION



FROM THE PREVIOUS DISCUSSION, YOU KNOW THAT MEAN DEVIATION AND STANDARD DEVIATION CONSIDER ALL THE DATA VALUES. HOWEVER, MEAN DEVIATION ASSUMES ONLY THE ABSOLUTE DEVIATIONS OF EACH DATA VALUE FROM THE CENTRAL VALUE (MEAN, MEDIAN OR MODE). IT MISSES ALGEBRAIC CONSIDERATIONS. TO OVERCOME THE LIMITATION OF MEAN DEVIATION WE HAVE A BETTER MEASURE OF VARIATION WHICH IS KNOWN AS STANDARD DEVIATION. YOU RECALL THAT STANDARD DEVIATION IS GIVEN BY:

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} \quad \text{for sample ungrouped data and}$$

$$= \sqrt{\frac{\sum(x_i - \bar{x})^2}{N}} \quad \text{for population data}$$

WHEN CONSIDERING STANDARD DEVIATION ~~AND NOT THE MEAN~~ DEVIATIONS, YOU ALWAYS TAKE THE DEVIATIONS FROM THE ARITHMETIC MEAN IN STANDARD DEVIATION. SINCE THE DEVIATION IS SQUARED, THE SIGN BECOMES NON-NEGATIVE WITHOUT VIOLATING THE RULES OF ALGEBRA. THUS, STANDARD DEVIATION IS THE ONE THAT IS MOSTLY USED FOR ANALYSIS AND INTERPRETATION. IT IS ALSO USED IN CONJUNCTION WITH THE MEAN FOR COMPARING DEGREES OF VARIABILITY AND CONSISTENCY OF TWO OR MORE DIFFERENT DATA SETS.

Example 9 FIND THE STANDARD DEVIATION OF THE FOLLOWING DATA.

- a** 6, 6, 6, 6, 6, 6, 6 $\bar{x} = 6$

x	$x - \bar{x}$	$(x - \bar{x})^2$
6	0	0
6	0	0
6	0	0
6	0	0
6	0	0
6	0	0
6	0	0
6	0	0
		$\sum(x - \bar{x})^2 = 0$

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{0}{n-1}} = 0$$

SINCE $s = 0$, IT INDICATES THAT THERE IS NO VARIABILITY IN THE DATA SET.

Note: THE GREATER THE STANDARD DEVIATION, THE HIGHER THE VARIABILITY.

8.5 ANALYSIS OF FREQUENCY DISTRIBUTIONS

ACTIVITY 8.11



CONSIDER THESE TWO GROUPS OF SIMILAR DATA:

DATA A: 1, 2, 3, 4, 5, 6, 7, 8, 9 AND

DATA B: 5, 4, 5, 5, 5, 5, 6, 5, 5

COMPARE THE TWO DATA SETS. WHICH OF THESE TWO DATA SETS IS MORE CONSISTENT? WHY? THE TWO DATA SETS GIVEN ABOVE BOTH HAVE AN AVERAGE OF 5. HOW IS IT POSSIBLE TO COMPARE THESE DATA SETS? CAN YOU CONCLUDE THAT THEY ARE THE SAME? IT IS OBVIOUS THAT THE TWO DATA SETS ARE NOT THE SAME IN CONSISTENCY.

Example 1 THE DAILY INCOME OF THREE SMALL SHOPS IS RECORDED AS FOLLOWS. WHICH SHOP HAS CONSISTENT INCOME (IN BIRR)?

Shops	A	B	C
	28	35	29
	36	32	39
	42	41	23
	25	27	33
	29	25	36

$$\bar{x}_A = \bar{x}_B = \bar{x}_C = 32$$

FOR SHOP A			FOR SHOP B			FOR SHOP C		
x	$x - \bar{x}$	$(x - \bar{x})^2$	x	$x - \bar{x}$	$(x - \bar{x})^2$	x	$x - \bar{x}$	$(x - \bar{x})^2$
28	-4	16	35	3	9	29	-3	9
36	4	16	32	0	0	39	7	49
42	10	100	41	9	81	23	-9	81
25	-7	49	27	-5	25	33	1	1
29	-3	9	25	-7	49	36	4	16
$\sum (x - \bar{x})^2 = 190$			$\sum (x - \bar{x})^2 = 164$			$\sum (x - \bar{x})^2 = 156$		

BASED ON THE VALUES IN THE ABOVE TABLE, WE SEE THAT

$$S_A = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{190}{5-1}} = \sqrt{\frac{190}{4}} = 6.89;$$

$$S_B = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{164}{5-1}} = \sqrt{\frac{164}{4}} = 6.40, \text{ AND}$$

$$S_C = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{156}{5-1}} = \sqrt{\frac{156}{4}} = 6.24$$

THE COMPARISON SHOWS THAT $S_C < S_B < S_A$. SINCE $S_C < S_B < S_A$, THEN SHOP C HAS THE MOST CONSISTENT INCOME. THE INCOME OF SHOP A IS HIGHLY VARIABLE.

IN THE DISCUSSION GIVEN ABOVE, YOU USED STANDARD DEVIATIONS TO COMPARE CONSISTENCY WHERE THE DATA SETS CONSIDERED HAVE THE SAME MEAN AND THE SAME UNIT. BUT, YOU MAY ALSO FACE DATA SETS THAT DO NOT HAVE THE SAME MEAN. YOU MAY ALSO FACE DATA SETS THAT DO NOT HAVE THE SAME UNIT. IF THE UNITS ARE DIFFERENT, IT WILL BE DIFFICULT TO COMPARE CONSISTENCY. FOR EXAMPLE, FOR TWO SETS OF DATA A AND B, IF $S_A = 1.7$ CM AND $S_B = 1.7$ CM WHICH OF THE DATA SETS A OR B IS MORE VARIABLE?

YOU CANNOT COMPARE KG TO CM. HENCE, YOU NEED TO SEE A RELATIVE MEASURE OF VARIATION WHICH IS A PURE NUMBER. SUCH A PURE NUMBER, WHICH IS USED AS A RELATIVE MEASURE OF VARIATION, IS THE COEFFICIENT OF VARIATION GIVEN BY:

$$CV = \frac{s}{\bar{x}} \times 100.$$

Example 2 CONSIDER THE FOLLOWING DATA ON THE MEAN AND STANDARD DEVIATION OF THE GROSS INCOMES OF TWO SCHOOLS A AND B.

School	Mean income (in Birr)	Standard deviation (in Birr)
A	8000	120
B	8000	140

FROM THIS TABLE, YOU SEE THAT BOTH SCHOOLS HAVE THE SAME MEAN OF 8000 BIRR. THE EQUALITY OF THE MEANS INDICATE THAT THESE TWO SCHOOLS HAVE THE SAME VARIATION CONSISTENCY?

OBVIOUSLY, THE ANSWER IS NO, BECAUSE THE TWO DATA SETS DO NOT HAVE THE SAME STANDARD DEVIATION. FOR SUCH A CASE, WHEN YOU NEED TO COMPARE THE CONSISTENCY OF TWO DATA SETS, YOU CAN USE ANOTHER MEASURE CALLED THE COEFFICIENT OF VARIATION (CV).

The Coefficient of variation IS A UNIT-LESS RELATIVE MEASURE THAT WE USE TO MEASURE THE DEGREE OF CONSISTENCY GIVEN AS A RATIO OF THE STANDARD DEVIATION TO THE MEAN.

$$C.V = \frac{\text{standard deviation}}{\text{mean}} \times 100 = \frac{s}{\bar{x}} \times 100$$

FOR THE ABOVE EXAMPLE, $C.V_A = \frac{120}{8000} \times 100 = 1.5$ AND $C.V_B = \frac{140}{8000} \times 100 = 1.75$.

WHEN YOU COMPARE THESE TWO DATA SETS, YOU CAN CONCLUDE THAT DATA SET A IS MORE CONSISTENT THAN DATA SET B BECAUSE DATA SET A HAS A HIGHER DEGREE OF VARIABILITY.

YOU CAN ALSO SEE THE RATIO OF THE COEFFICIENTS OF VARIATION, GIVEN AS

$\frac{C.V_A}{C.V_B} = \frac{1.5}{1.75} = \frac{120}{140} = \frac{6}{7}$ FROM WHICH YOU CAN CONCLUDE THAT THE DATA SET WITH LESSER STANDARD DEVIATION IS MORE CONSISTENT THAN THE DATA SET WITH LARGER STANDARD DEVIATION.

Example 3 THE FOLLOWING ARE THE MEAN AND THE STANDARD DEVIATION OF WEIGHT OF A SAMPLE OF STUDENTS.

height	weight
$\bar{x} = 168\text{CM}$	$\bar{x} = 54\text{KC}$
$S = 2.3\text{ CM}$	$S = 1.6\text{ KC}$

WHICH OF THE MEASURED VALUES (HEIGHT OR WEIGHT) HAS THE HIGHER VARIABILITY?

$$C.V(\text{HEIGHT}) = \frac{s}{\bar{x}} \times 100 = \frac{2.3\text{CM}}{168\text{CM}} \times 100 = 1.369$$

$$C.V(\text{WEIGHT}) = \frac{s}{\bar{x}} \times 100 = \frac{1.6\text{KG}}{54\text{KG}} \times 100 = 2.963$$

SINCE $C.V(\text{WEIGHT}) > C.V(\text{HEIGHT})$, THE STUDENTS HAVE GREATER VARIABILITY IN WEIGHT.

Exercise 8.6

1 TWO BASKETBALL TEAMS SCORED THE FOLLOWING POINTS IN GAMES AS FOLLOWS:

TEAM A: 42 17 83 59 72 76 64 45 40 32

TEAM B: 28 70 31 0 59 108 82 14 3 95

- A** CALCULATE THE STANDARD DEVIATION OF EACH TEAM.
- B** WHICH TEAM SCORED MORE CONSISTENT POINTS?

2 THE MEAN AND STANDARD DEVIATION OF GROSS INCOME OF TWO COMPANIES ARE GIVEN BELOW:

Company	Mean	Standard deviation
A	6000	120
B	10000	220

- A** CALCULATE THE C.V OF EACH COMPANY.
- B** WHICH COMPANY HAS THE MORE VARIABLE INCOME?

8.6 USE OF CUMULATIVE FREQUENCY CURVES

IN SECTION 8.3 OF THIS UNIT, YOU SAW THREE TYPES OF FREQUENCY CURVES WHOSE SHAPE IS SYMMETRICAL, SKEWED TO THE LEFT OR SKEWED TO THE RIGHT. THE SHAPE OF A FREQUENCY CURVE DESCRIBES THE DISTRIBUTION OF A DATA SET. SUCH A DESCRIPTION WAS MADE POSSIBLE BY DRAWING THE FREQUENCY CURVE OF A FREQUENCY DISTRIBUTION.

IN THIS SUB-UNIT, YOU WILL SEE HOW THE MEASURES OF CENTRAL TENDENCY (MEAN, MEDIAN AND MODE) DETERMINE THE SKEWNESS OF A DISTRIBUTION.

8.6.1 Skewness Based on the Relationships Between Mean, Median and Mode

ACTIVITY 8.12



CONSIDER THE FOLLOWING DATA

DATA A : 2, 3, 4, 5, 5, 6, 5, 7, 8 DATA B: 2, 3, 1, 4, 8, 5, 8, 7

- 1 CALCULATE AND COMPARE THE MEAN, MEDIAN AND MODE FOR EACH DATA SET.
- 2 CONSTRUCT FREQUENCY CURVES FOR EACH DATA SET.

RELATIVE MEASURES OF VARIATION HELP TO STUDY THE CONSISTENCY OR VARIATION OF DATA IN A DISTRIBUTION. HOW DO MEASURES OF CENTRAL TENDENCY HELP IN STUDYING THE SKEWNESS OF A DISTRIBUTION? WHAT HAPPENS TO THE SKEWNESS IF MEAN = MEDIAN = MODE?

A MEASURE OF CENTRAL TENDENCY OR A MEASURE OF VARIATION ALONE DOES NOT TELL US WHETHER THE DISTRIBUTION IS SYMMETRICAL OR NOT. IT IS THE RELATIONSHIP BETWEEN THE MEAN, MEDIAN AND MODE THAT TELLS US WHETHER THE DISTRIBUTION IS SYMMETRICAL OR SKEWED.

Example 1 CONSIDER THE FOLLOWING FREQUENCY DISTRIBUTION OF STUDENTS IN A CLASS:

- A DRAW THE HISTOGRAM AND FREQUENCY CURVE.
- B CALCULATE MEAN, MEDIAN AND MODE.
- C DESCRIBE RELATIONSHIPS BETWEEN THE MEAN, MEDIAN AND MODE AND THE SKEWNESS OF THE DISTRIBUTION.

Age	Number of students
13 – 14	5
15 – 16	15
17 – 18	30
19 – 20	15
21 – 22	5
	70

Solution

A THE HISTOGRAM AND FREQUENCY CURVE OF DISTRIBUTION AS FOLLOWS.

Histogram and frequency curve of age of students

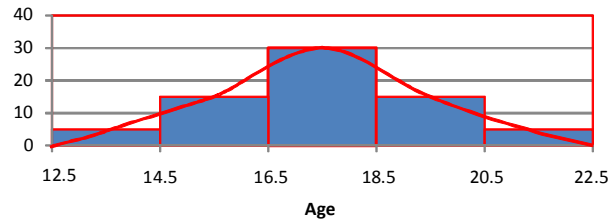


Figure 8.21

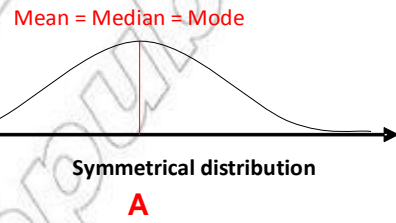
THIS APPEARS TO BE SYMMETRICAL.

B MEAN = MEDIAN = MODE = 17.5

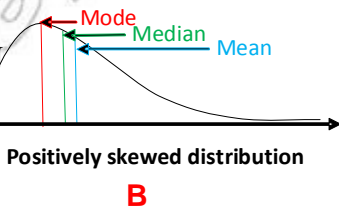
C FROM A AND B, YOU SEE THAT WHENEVER MEAN = MEDIAN = MODE, THE DISTRIBUTION IS SYMMETRICAL.

INVESTIGATE WHAT HAPPENS TO THE SKEWNESS OF A DISTRIBUTION, IF $MEAN > MEDIAN > MODE$. FROM THE DISCUSSIONS OUTLINED ABOVE, YOU CAN MAKE THE FOLLOWING GENERALIZATION:

I FOR A UNIMODAL DISTRIBUTION IN WHICH THE VALUES OF MEAN, MEDIAN AND MODE COINCIDE (I.E., $MEAN = MEDIAN = MODE$), THE DISTRIBUTION IS SAID TO BE **symmetrical**.



II IF THE MEAN IS THE LARGEST IN VALUE, AND THE MEDIAN IS LARGER THAN THE MODE BUT SMALLER THAN THE MEAN, THEN THE DISTRIBUTION IS POSITIVELY SKEWED. THAT IS, IF $MEAN > MEDIAN > MODE$, THEN THE DISTRIBUTION IS **positively skewed** (SKEWED TO THE RIGHT).



III IF THE MEAN IS SMALLEST IN VALUE, AND THE MEDIAN IS LARGER THAN THE MEAN BUT SMALLER THAN THE MODE, THEN THE DISTRIBUTION IS NEGATIVELY SKEWED. THAT IS, IF $MEAN < MEDIAN < MODE$ THEN THE DISTRIBUTION IS **negatively skewed** (SKEWED TO THE LEFT).

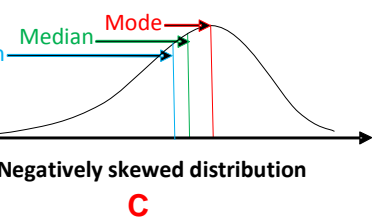


Figure 8.22

8.6.2 Skewness Based on Relationships Between Measures of Central Tendency and Measures of Variation

IN THE ABOVE DISCUSSION, WE USED THE RELATIONSHIPS BETWEEN MEASURES OF CENTRAL TENDENCY ONLY TO DETERMINE THE SKEWNESS OF A DISTRIBUTION. WITH THE HELP OF TENDENCIES AND STANDARD DEVIATION, IT IS ALSO POSSIBLE TO DETERMINE SKEWNESS OF A DISTRIBUTION. THIS IS SOMETIMES CALLED A MATHEMATICAL MEASURE OF SKEWNESS. MATHEMATICALLY, SKEWNESS CAN BE MEASURED IN ONE OF THE FOLLOWING WAYS BY CALCULATING A COEFFICIENT OF SKEWNESS.

- 1 Karl Pearson's coefficient of skewness
- 2 Bowley's coefficient of skewness

1 Karl Pearson's coefficient of skewness

KARL PEARSON'S COEFFICIENT OF SKEWNESS (PEARSON'S COEFFICIENT OF SKEWNESS) IS OBTAINED BY EXPRESSING THE DIFFERENCE BETWEEN THE MEAN AND THE MEDIAN RELATIVE TO THE STANDARD DEVIATION. IT IS USUALLY DENOTED BY

$$\text{COEFFICIENT OF SKEWNESS} = \frac{3(\text{MEAN} - \text{MED})}{\text{STANDARD DEV}}$$

THE INTERPRETATION OF SKEWNESS BY THIS APPROACH FOLLOWS OUR PRIOR KNOWLEDGE FROM THE PREVIOUS DISCUSSION, IF MEAN = MEDIAN, WE CAN SEE THAT THE DISTRIBUTION IS SYMMETRICAL. LOOKING AT PEARSON'S COEFFICIENT OF SKEWNESS, IF MEAN = MEDIAN, SO THE DISTRIBUTION IS SYMMETRICAL. FOLLOWING THE SAME APPROACH, WE CAN STATE THE INTERPRETATION ON SKEWNESS USING PEARSON'S COEFFICIENT OF SKEWNESS.

Interpretation

- 1 IF PEARSON'S COEFFICIENT OF SKEWNESS IS ZERO, THE DISTRIBUTION IS SYMMETRICAL.
- 2 IF PEARSON'S COEFFICIENT OF SKEWNESS IS POSITIVE, THE DISTRIBUTION IS SKEWED POSITIVELY (SKEWED TO THE RIGHT).
- 3 IF PEARSON'S COEFFICIENT OF SKEWNESS IS NEGATIVE, THE DISTRIBUTION IS SKEWED NEGATIVELY (SKEWED TO THE LEFT).

Example 2 CALCULATE KARL PEARSON'S COEFFICIENT OF SKEWNESS FROM THE DATA GIVEN BELOW AND DETERMINE THE SKEWNESS OF THE DISTRIBUTION.

x	11	12	13	14	15
f	3	9	6	4	3

Solution MEAN = 12.8, MEDIAN = 13, AND S = 1.2

$$\text{COEFFICIENT OF SKEWNESS} = \frac{3(\text{MEAN} - \text{MEDIAN})}{S} = \frac{3(12.8 - 13)}{1.2} = -0.5$$

$$\text{COEFFICIENT OF SKEWNESS} = -0.5 < 0$$

∴ THE DISTRIBUTION IS NEGATIVELY SKEWED.

2 Bowley's coefficient of skewness

PREVIOUSLY, YOU SAW HOW TO DETERMINE SKEWNESS USING THE RELATIONSHIPS BETWEEN MEAN, MEDIAN AND STANDARD DEVIATION. IT IS ALSO POSSIBLE TO DETERMINE SKEWNESS BY USING POSITIONAL MEASURES OF CENTRAL TENDENCY, THE QUANTILES. SUCH A COEFFICIENT OF SKEWNESS THAT USES QUANTILES, IS CALLED BOWLEY'S COEFFICIENT OF SKEWNESS.

BOWLEY'S COEFFICIENT OF SKEWNESS, WHICH IS USUALLY DENOTED BY

$$\text{BOWLEY'S COEFFICIENT OF SKEWNESS} = \frac{Q_3 + Q_1 - 2(\text{MEDIAN})}{Q_3 - Q_1}$$

The interpretation for skewness based on Bowley's coefficient of skewness is also the same as that of Pearson's that

- 1 IF BOWLEY'S COEFFICIENT OF SKEWNESS IS 0, THE DISTRIBUTION IS SYMMETRICAL.
- 2 IF BOWLEY'S COEFFICIENT OF SKEWNESS IS POSITIVE, THE DISTRIBUTION IS SKEWED POSITIVELY (SKEWED TO THE RIGHT).
- 3 IF BOWLEY'S COEFFICIENT OF SKEWNESS IS NEGATIVE, THE DISTRIBUTION IS NEGATIVELY SKEWED (SKEWED TO THE LEFT).

Example 3 FIND BOWLEY'S COEFFICIENT OF SKEWNESS FOR THE FOLLOWING DATA AND DETERMINE THE SKEWNESS OF THE DISTRIBUTION.

15, 18, 19, 3, 2, 7, 10, 6, 9, 8

Solution ARRANGING THE DATA IN ASCENDING ORDER, YOU GET:

2, 3, 6, 7, 8, 9, 10, 15, 18, 19.

FROM THIS ARRANGED DATA, YOU CAN DETERMINE THE QUANTILES AS

$$Q_1 = 4.5 \quad Q_2 = \text{MEDIAN} = 8.5 \quad Q_3 = 12.5$$

$$\begin{aligned} \text{BOWLEY'S COEFFICIENT OF SKEWNESS} &= \frac{Q_3 + Q_1 - 2(\text{MEDIAN})}{Q_3 - Q_1} \\ &= \frac{12.5 + 4.5 - 2(8.5)}{12.5 - 4.5} = \frac{17 - 17}{8} = 0 \end{aligned}$$

∴ THE DISTRIBUTION IS SYMMETRICAL.



Key Terms

bar chart	mean deviation
coefficient of variation	non random sampling technique
frequency curve	population
frequency polygon	sample random sampling technique
histogram	skewness
inter - quartile range	standard deviation
line graph	symmetrical distribution



Summary

- 1 **Statistics** REFERS TO METHODS THAT ARE USED FOR COLLECTING, ANALYZING AND PRESENTING NUMERICAL DATA.
- 2 STATISTICS IS HELPFUL IN BUSINESS RESEARCH, UNDERSTANDING OF ECONOMIC PROBLEMS AND THE FORMULATION OF ECONOMIC POLICY.
- 3 A **population** IS THE COMPLETE SET OF ITEMS WHICH ARE IN A PARTICULAR SITUATION.
- 4 IT IS NOT POSSIBLE TO COLLECT INFORMATION FROM THE WHOLE POPULATION BECAUSE IT IS COSTLY IN TERMS OF TIME, ENERGY AND RESOURCES. TO OVERCOME THESE PROBLEMS WE TAKE ONLY A CERTAIN PART OF THE POPULATION CALLED **sample**.
- 5 A SAMPLE SERVES AS REPRESENTATIVE OF THE POPULATION FROM WHICH WE CAN DRAW CONCLUSIONS ABOUT THE ENTIRE POPULATION BASED ON THE RESULTS OBTAINED FROM THE SAMPLE.
- 6 THERE ARE TWO METHODS OF SAMPLING
 - ✓ THE **Random (probability)** SAMPLING METHOD.
 - ✓ THE **Non-Random (non-probability)** SAMPLING METHOD.
- 7 IN **random sampling**, EVERY MEMBER OF THE POPULATION HAS AN EQUAL CHANCE OF BEING SELECTED.
- 8 **Raw data**, WHICH HAS BEEN COLLECTED, CAN BE PRESENTED IN TABULAR, GRAPHICAL AND PICTORIAL METHODS.
- 9 THE PURPOSE OF PRESENTING DATA IN FREQUENCY DISTRIBUTION IS TO
 - ✓ CONDENSE AND SUMMARIZE LARGE AMOUNT OF DATA.

- 10** THE PURPOSE OF PRESENTING DATA USING PICTORIAL METHOD
- ✓ FACILITATE COMPARISONS BETWEEN TWO OR MORE SETS OF DATA
 - ✓ CONVEY MESSAGES ABOUT THE NATURE OF DATA AT A GLANCE
- 11** Measures of variation HELP TO DECIDE THE DEGREE OF VARIABILITY.
- 12** THERE ARE TWO TYPES OF MEASURES OF VARIATION
- I absolute AND
 - II relative.
- 13** Range = $x_{max} - x_{min}$
- 14** Inter quartile range = $Q_3 - Q_1$
- 15** Mean deviation about the mean = $MD(\bar{x}) = \frac{\sum |x - \bar{x}|}{n}$
- 16** Mean deviation about the median = $MD(m_d) = \frac{\sum |x - m_d|}{n}$
- 17** Mean deviation about the mode = $MD(m_o) = \frac{\sum |x - m_o|}{n}$
- 18** FOR A GIVEN DISTRIBUTION:
- ✓ IF $\bar{x} > m_d > m_o$ THE DISTRIBUTION IS positively skewed
 - ✓ IF $\bar{x} < m_d < m_o$ THE DISTRIBUTION IS negatively skewed
 - ✓ IF $\bar{x} = m_d = m_o$ THE DISTRIBUTION IS symmetrical
- 19** Pearson's coefficient of skewness = $\frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$
- ✓ = 0 MEANS THE DISTRIBUTION IS symmetrical.
 - ✓ > 0 MEANS THE DISTRIBUTION IS positively skewed.
 - ✓ < 0 MEANS THE DISTRIBUTION IS negatively skewed.
- 20** Bowley's coefficient of skewness = $\beta = \frac{Q_3 + Q_1 - 2(\text{MEDIAN})}{Q_3 - Q_1}$
- ✓ = 0 MEANS THE DISTRIBUTION IS symmetrical.
 - ✓ > 0 MEANS THE DISTRIBUTION IS positively skewed.
 - ✓ < 0 MEANS THE DISTRIBUTION IS negatively skewed.



Review Exercises on Unit 8

- 1 DEFINE POPULATION AND SAMPLE.
- 2 WRITE DOWN ADVANTAGES OF SIMPLE RANDOM SAMPLING AND STRATIFIED SAMPLING TECHNIQUES.
- 3 EXPLAIN THE DIFFERENCE BETWEEN A FREQUENCY POLYGON AND A FREQUENCY CURVE.
- 4 WHAT BENEFITS DO STATISTICAL GRAPHS UNDERSTAND AND INTERPRET DATA?
- 5 WHAT DOES SKEWNESS MEAN ABOUT A DISTRIBUTION?
- 6 EXPLAIN SIMILARITIES AND DIFFERENCES BETWEEN EMPLOYED AND MULTIPLE BAR CHARTS.
- 7 DISCUSS MEAN DEVIATION FROM THE MEAN, MEDIAN AND MODE IN THE ADVANTAGES OF EACH.
- 8 WHAT LIMITATIONS DO RANGE AND INTERQUARTILE RANGE HAVE?
- 9 WHY IS IT USEFUL TO USE STANDARD DEVIATION AS A MEASURE OF DISPERSION?
- 10 THE AGES OF 50 PEOPLE ARE GIVEN BELOW

21	75	15	60	72	40	46	65	70	45
22	34	35	53	64	66	63	80	34	36
21	45	72	38	23	45	40	69	24	39
40	30	50	60	24	38	35	45	27	66
45	34	54	24	38	66	46	32	45	40

- A WHAT TYPE OF DATA IS THIS? (DISCRETE OR CONTINUOUS)
 - B SELECT SUITABLE CLASSES AND PREPARE A FREQUENCY DISTRIBUTION.
 - C DRAW A HISTOGRAM TO PRESENT THE DATA.
 - D DRAW A FREQUENCY POLYGON.
- 11 CONSIDER THE FOLLOWING TABLE

Year	Average production in tonnes		
	Wheat	Maize	Total
1960	440	250	690
1961	170	362	532
1962	620	657	1277

PRESENT THE ABOVE DATA USING

- A A SIMPLE BAR CHART
- B A COMPONENT BAR CHART
- C A MULTIPLE BAR CHART
- D A PIE CHART

12 A FIND THE MEAN, MEDIAN AND STMODE, ^{RP}QUANTILE AND QUARTILE OF THE FOLLOWING DATA:

Class	0 - 9	10 - 19	20 - 29	30 - 39
frequency	2	10	13	8

B USING THE ABOVE DATA CALCULATE

- I** THE MEAN DEVIATION ABOUT THE MEAN, MODE AND MEDIAN;
- II** THE RANGE AND INTER-QUARTILE RANGE;
- III** THE STANDARD DEVIATION;
- IV** THE COEFFICIENT OF VARIATION;
- V** PEARSON'S COEFFICIENT OF SKEWNESS AND ~~KNOWLEDGE OF THIS~~ ^{KNOWLEDGE OF THIS} DISTRIBUTION.

13 THE FOLLOWING DATA REPORTS THE PERFORMANCE OF ~~WORKERS~~ ^{WORKERS}.

Company	A	B
Average hours worked in a week	30	28
Standard deviation in performance	5	8

WORKERS OF WHICH COMPANY ARE MORE CONSISTENT IN THEIR PERFORMANCE?

14 THE FOLLOWING DATA REPRESENTS HOURS ~~IN A DAY WORKED~~ ^{IN A DAY WORKED} IN SOIL AND WATER CONSERVATION.

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

USING THE ABOVE DATA CALCULATE

- I** THE MEAN DEVIATION ABOUT THE MEAN, MODE AND MEDIAN.
- II** THE RANGE AND INTER-QUARTILE RANGE.
- III** THE STANDARD DEVIATION.
- IV** THE COEFFICIENT OF VARIATION.