Module 6

Junior Secondary Mathematics

Data Handling

THE COMMONWEALTH of LEARNING

Science, Technology and Mathematics Modules

for Upper Primary and Junior Secondary School Teachers of Science, Technology and Mathematics by Distance in the Southern African Development Community (SADC)

Developed by
The Southern African Development Community (SADC)

Ministries of Education in:

- **Botswana**
- **Malawi**
- **Mozambique**
- **Namibia**
- **South Africa**
- **Tanzania**
- **Zambia**
- **Zimbabwe**

In partnership with The Commonwealth of Learning

Published jointly by The Commonwealth of Learning and the SADC Ministries of Education.

Residents of the eight countries listed above may obtain modules from their respective Ministries of Education. The Commonwealth of Learning will consider requests for modules from residents of other countries.

ISBN 1-895369-65-7

SCIENCE, TECHNOLOGY AND MATHEMATICS MODULES

This module is one of a series prepared under the auspices of the participating Southern African Development Community (SADC) and The Commonwealth of Learning as part of the Training of Upper Primary and Junior Secondary Science, Technology and Mathematics Teachers in Africa by Distance. These modules enable teachers to enhance their professional skills through distance and open learning. Many individuals and groups have been involved in writing and producing these modules. We trust that they will benefit not only the teachers who use them, but also, ultimately, their students and the communities and nations in which they live.

The twenty-eight Science, Technology and Mathematics modules are as follows:

Upper Primary Science
Module 1: *My Built Environment*
Module 2: *Materials in my Environment*
Module 3: *My Health*
Module 4: *My Natural Environment*

Junior Secondary Science
Module 1: *Energy and Energy Transfer*
Module 2: *Energy Use in Electronic Communication*
Module 3: *Living Organisms' Environment and Resources*
Module 4: *Scientific Processes*

Upper Primary Technology
Module 1: *Teaching Technology in the Primary School*
Module 2: *Making Things Move*
Module 3: *Structures*
Module 4: *Materials*
Module 5: *Processing*

Junior Secondary Technology
Module 1: *Introduction to Teaching Technology*
Module 2: *Systems and Controls*
Module 3: *Tools and Materials*
Module 4: *Structures*

Upper Primary Mathematics
Module 1: *Number and Numeration*
Module 2: *Fractions*
Module 3: *Measures*
Module 4: *Social Arithmetic*
Module 5: *Geometry*

Junior Secondary Mathematics
Module 1: *Number Systems*
Module 2: *Number Operations*
Module 3: *Shapes and Sizes*
Module 4: *Algebraic Processes*
Module 5: *Solving Equations*
Module 6: *Data Handling*

A MESSAGE FROM THE COMMONWEALTH OF LEARNING

The Commonwealth of Learning is grateful for the generous contribution of the participating Ministries of Education. The Permanent Secretaries for Education played an important role in facilitating the implementation of the 1998-2000 project work plan by releasing officers to take part in workshops and meetings and by funding some aspects of in-country and regional workshops. The Commonwealth of Learning is also grateful for the support that it received from the British Council (Botswana and Zambia offices), the Open University (UK), Northern College (Scotland), CfBT Education Services (UK), the Commonwealth Secretariat (London), the South Africa College for Teacher Education (South Africa), the Netherlands Government (Zimbabwe office), the British Department for International Development (DFID) (Zimbabwe office) and Grant MacEwan College (Canada).

The Commonwealth of Learning would like to acknowledge the excellent technical advice and management of the project provided by the strategic contact persons, the broad curriculum team leaders, the writing team leaders, the workshop development team leaders and the regional monitoring team members. The materials development would not have been possible without the commitment and dedication of all the course writers, the in-country reviewers and the secretaries who provided the support services for the in-country and regional workshops.

Finally, The Commonwealth of Learning is grateful for the instructional design and review carried out by teams and individual consultants as follows:

- Grant MacEwan College (Alberta, Canada):
  General Education Courses
- Open Learning Agency (British Columbia, Canada):
  Science, Technology and Mathematics
- Technology for Allcc. (Durban, South Africa):
  Upper Primary Technology
- Hands-on Management Services (British Columbia, Canada):
  Junior Secondary Technology

Dato' Professor Gajaraj Dhanarajan
President and Chief Executive Officer

ACKNOWLEDGEMENTS

The Mathematics Modules for Upper Primary and Junior Secondary Teachers in the Southern Africa Development Community (SADC) were written and reviewed by teams from the participating SADC Ministries of Education with the assistance of The Commonwealth of Learning.

CONTACTS FOR THE PROGRAMME

The Commonwealth of Learning
1285 West Broadway, Suite 600
Vancouver, BC V6H 3X8
Canada

Ministry of Education
Private Bag 005
Gaborone
Botswana

Ministry of Education
Private Bag 328
Capital City
Lilongwe 3
Malawi

Ministério da Eduação
Avenida 24 de Julho No 167, 8
Caixa Postal 34
Maputo
Mozambique

Ministry of Basic Education,
Sports and Culture
Private Bag 13186
Windhoek
Namibia

National Ministry of Education
Private Bag X603
Pretoria 0001
South Africa

Ministry of Education and Culture
P.O. Box 9121
Dar es Salaam
Tanzania

Ministry of Education
P.O. Box 50093
Lusaka
Zambia

Ministry of Education, Sport and
Culture
P.O. Box CY 121
Causeway
Harare
Zimbabwe

COURSE WRITERS FOR JUNIOR SECONDARY MATHEMATICS

Ms. Sesutho Koketso Kesianye: *Writing Team Leader*
Head of Mathematics Department
Tonota College of Education
Botswana

Mr. Jan Durwaarder: Lecturer (Mathematics)
Tonota College of Education
Botswana

Mr. Kutengwa Thomas Sichinga: Teacher (Mathematics)
Moshupa Secondary School
Botswana

FACILITATORS/RESOURCE PERSONS

Mr. Bosele Radipotsane: Principal Education Officer (Mathematics)
Ministry of Education
Botswana

Ms. Felicity M Leburu-Sianga: Chief Education Officer
Ministry of Education
Botswana

PROJECT MANAGEMENT & DESIGN

Ms. Kgomotso Motlotle: Education Specialist, Teacher Training
The Commonwealth of Learning (COL)
Vancouver, BC, Canada

Mr. David Rogers: *Post-production Editor*
Open Learning Agency
Victoria, BC, Canada

Ms. Sandy Reber: *Graphics & desktop publishing*
Reber Creative
Victoria, BC, Canada

TEACHING JUNIOR SECONDARY MATHEMATICS

Introduction

Welcome to *Data Handling*, Module 6 of Teaching Junior Secondary Mathematics! This series of six modules is designed to help you to strengthen your knowledge of mathematics topics and to acquire more instructional strategies for teaching mathematics in the classroom.

The guiding principles of these modules are to help make the connection between theoretical maths and the use of the maths; to apply instructional theory to practice in the classroom situation; and to support you, as you in turn help your students to apply mathematics theory to practical classroom work.

Programme Goals

This programme is designed to help you to:

- strengthen your understanding of mathematics topics
- expand the range of instructional strategies that you can use in the mathematics classroom

Programme Objectives

By the time you have completed this programme, you should be able to:

- develop and present lessons on the nature of the mathematics process, with an emphasis on where each type of mathematics is used outside of the classroom
- guide students as they work in teams on practical projects in mathematics, and help them to work effectively as a member of a group
- use questioning and explanation strategies to help students learn new concepts and to support students in their problem solving activities
- guide students in the use of investigative strategies on particular projects, and thus to show them how mathematical tools are used
- guide students as they prepare their portfolios about their project activities

How to work on this programme

Congratulations on reaching Module 6!

Data Handling, or Descriptive Statistics, is a relatively new concept in Maths teaching, basically a response to the need for Maths literacy in an information-rich society. As we have said in earlier modules, do the Exercises and Assignments yourself, transfer some concepts into your own classroom, and interact with your colleagues about this newer aspect of school Maths.

ICONS

Throughout each module, you will find the following icons or symbols that alert you to a change in activity within the module.

Read the following explanations to discover what each icon prompts you to do.

	**Introduction**	Rationale or overview for this part of the course.
	**Learning Objectives**	What you should be able to do after completing this module or unit.
	**Text or Reading Material**	Course content for you to study.
	**Important—Take Note!**	Something to study carefully.
	**Self-Marking Exercise**	An exercise to demonstrate your own grasp of the content.
	**Individual Activity**	An exercise or project for you to try by yourself and demonstrate your own grasp of the content.
	**Classroom Activity**	An exercise or project for you to do with or assign to your students.
	**Reflection**	A question or project for yourself—for deeper understanding of this concept, or of your use of it when teaching.
	**Summary**	
	**Unit or Module Assignment**	Exercise to assess your understanding of all the unit or module topics.
	**Suggested Answers to Activities**	
	**Time**	Suggested hours to allow for completing a unit or any learning task.
	**Glossary**	Definitions of terms used in this module.

CONTENTS

Module 6: Data Handling

Module 6 – Overview 3

Unit 1: What is data handling? 5
- Section A Data handling or statistics? 6
- Section B Reasons for including data handling in the curriculum 7
- Section C Misconceptions as to what data handling / statistics implies 8
- Section D How do people make decisions? 9
- Section E Stages in statistical investigation 11
- Section F Data and information 11
- Section G Type of data 12
- Section G1 Qualitative versus quantitative data (variables) 12
- Section G2 Ordinal versus nominal qualitative data (variables) 13
- Section G3 Discrete versus continuous data (variables) 13
- Answers to self mark exercises Unit 1 16

Unit 2: Methods of data collection 17
- Section A Data collection with a purpose 19
- Section B Population, sample and random sampling 20
- Section C Misconceptions in inferential statistics 23
- Section D Methods of data collection 24
- Section D1 Surveys 24
- Section D1.1 Types of surveys 25
- Section D2 Experiments 37
- Section D3 Simulations 39
- Section D3.1 Pupils' simulation activity: Dice and disease in the classroom 41
- Section D3.2 Analytical model for dice and disease 44
- Section E Choice of data collection method 46
- Answers to self mark exercises Unit 2 48

Unit 3: Data representation 57
- Section A Represent or model data 58
- Section B Tables 58
- Section C Nature and format of data 58
- Section D Graphical representations 60
- Section D1 Bar charts 60
- Section D2 Line-stick graphs 62
- Section D3 Histograms 62
- Section D4 Pie charts 67
- Section D5 Pictograms 67
- Section D6 Line graphs/charts 69
- Section D7 Frequency polygons 71
- Section D8 Stem-leaf diagrams 72
- Section D9 Scatter diagrams 75
- Section E Representing data for understanding 79
- Section F Misconceptions of pupils in descriptive statistics. 86

Section G Making nonsense of statistics 88
Section H Interpreting data. 92
Answers to self mark exercises Unit 3 96

Unit 4: Measures of central tendency.....108
Section A Averages: mean, mode, median 110
Section B The concept of the mean.....111
Section B1 The average or mean has three meanings 112
Section B2 Misconceptions and pupils' errors 112
Section C Mean, median and mode for ungrouped data.....113
Section D Which is the best average to use?.....117
Section E Mean, mode and median: classroom lessons 119
Section F Mean, median and mode for grouped discrete data.....136
Section G Estimation of the median, quartiles and percentiles.....143
Section G1 Estimation of the median, quartiles and percentiles from cumulative frequency curves 143
Section G2 Estimation of median, quartiles and percentiles by linear interpolation.....148
Section G3 Estimation of the median from a histogram.....151
Section H Estimation of the mode from grouped data 153
Section I Boxplots or box and whiskers diagrams.....155
Answers to self mark exercises Unit 4 159

Unit 5: Measures of dispersion 170
Section A Interquartile range for ungrouped data 171
Section B Interquartile range for grouped data.....172
Section C Standard deviation of ungrouped data.....174
Section D Standard deviation of grouped data.....175
Answers to self mark exercises Unit 5 180

References 182

Glossary 183

Module 6
Data Handling

Introduction to the module

Over the past decade terms such as ‘data handling’, ‘exploratory data analysis’ and ‘data visualisation’ have replaced the use of the term ‘statistics’ in the secondary curricula. The current meaning of data handling emphasises collecting, organising, representing, analysing and interpreting data closely connected to the world of the pupils. The visual representation of data is of major importance at the secondary level. Data handling is seen as going beyond statistics (the science of data collection, representation, analysis and interpretation for decision making). Statistics can be said to be the content, data handling is the whole learning environment in which data are explored.

Aim of the module

The module aims at providing you with:

(a) ideas and materials to support the learning of pupils in data handling

(b) materials to extend and consolidate your knowledge on the topic of data handling

(c) knowledge on common misconceptions related to the topic of data handling

(d) ideas for pupil centred approach to the coverage of the concepts

(e) questions to make you reflect on your present practice as a teacher when covering the topic data handling and how to relate your present practice to the ideas presented in this module.

Structure of the module

This model first looks into the reasons to include data handling in the secondary curriculum and at misconceptions as to what data handling is. Unit 1 also looks at the basic concept of ‘data’, the various types of data and its distinction from information. Unit 2 covers the three basic methods for data collection, i.e., (a) surveys (b) experiments and (c) simulation. Attention is paid to questionnaires and interviews as means to collect data in surveys. Unit 3 covers the multiple ways in which data can be represented graphically. The major emphasis is on the question of which representation is most appropriate in the given context and less emphasis is on the techniques for the producing the various graphical representations. Unit 4 covers the measures for central tendency. The emphasis is again on ‘what is the most appropriate measure to use in the given context’ rather than on the techniques of computation. Unit 5 covers the measures for spread or dispersion and uses box plots, introduced in Unit 4, as a visual representation of spread.

Objectives of the module

When you have completed this module you should be able to create with confidence, due to enhanced background knowledge, a learning environment for your pupils in which they can:

(i) apply various techniques to collect, represent and analyse data

(ii) demonstrate a critical approach to data represented in the media

(iii) justify choices for representing and analysing collected data in a specific format

Prerequisite module

No special modules need to be covered prior to this module. The knowledge on data handling you have and experience as a teacher is all that is needed.

Unit 1: What is data handling?

Introduction to Unit 1

In this unit you will learn about the basic concept involved in data handling, i.e., What is data? You will also reflect on why this topic is included in the junior secondary school syllabus. Pupils generally do not have problems in drawing charts, calculating means and other techniques but fail to give meaning to these activities. The misconceptions as to what data handling is are discussed. Data handling serves a purpose: making of decisions. This unit looks at how pupils (people) make decisions.

Purpose of Unit 1

The main aim of this unit is to look at the basic questions: What is data handling? Why is it included in the junior secondary syllabus? What is data? How can types of data be distinguished? What erroneous ideas do exist with respect to data handling? What is the most appropriate method to use in the classroom to facilitate the learning of data handling?

Objectives

When you have complete this unit you should be able to:

- state reasons for including data handling in the curriculum
- defend the use of the term data handling versus the use of the word statistics
- state the purpose of data handling
- discuss the way people make decisions
- differentiate between data and information
- differentiate between qualitative and quantitative data
- differentiate between discrete and continuous quantitative data
- differentiate between nominal and ordinal qualitative data
- differentiate between descriptive and inferential statistics
- state reasons for the need of data collection
- sets activities to pupils to acquire knowledge on type of data
- set activities to make pupils aware of the base on which they make decisions
- list common misconceptions related to the question “what is statistics?”
- state the four stages of a statistical investigation

Time

To study this unit will take you about five hours.

Unit 1: What is data handling?

Section A: Data handling or statistics?

The term data handling has replaced the more traditional term statistics in the primary and secondary school curricula. The shift from statistics to data handling is more than a mere play on words. It signals a different methodology to be used in the classroom. One can identify two different approaches to the teaching / learning of data handling

(1) Statistical ideas and techniques are to be covered and the task of the teacher is to find suitable examples and activities to illustrate these.

(2) There are in the world around us a lot of questions and situations we like to understand, to describe, to explore and to access, and a range of statistical techniques will be an appropriate tool to do this.

Write down your view on why data handling should be taught at the junior secondary school level.

What method do you presently use to cover data handling?

Do you consider data handling to be part of mathematics or do you feel it should stand on its own?

Compare what you wrote with the information below and see how it agrees or disagrees with what you wrote.

It is the second view which is implicitly or explicitly expressed in the syllabus for the Junior Secondary School level in many countries. It is the view that data handling is **most effectively learned by pupils in the context of projects** which explore issues seen as relevant by the pupils rather than a set of skills and processes which have familiar illustrations. Projects such as "Smoking and Health", "Beer cans and pollution", "Teenage pregnancies" all start with a situation or an issue, and the need for relevant data handling techniques should develop when the need arises. In this way pupils are motivated to learn certain data handling concepts because they need them to analyse, describe and represent their collected data to answer the question they set at the start of their project.

This is also expressed in the Cockcroft Report (Cockcroft 1982) that states:

Statistics is essentially a practical subject and its study should be based on the collection of data, wherever possible by pupils themselves. It should consider the kinds of data which is appropriate to collect, the reasons for collecting the data and the problems in doing so, the ways in which the data may legitimately be manipulated and the kind of interference which may be drawn. (p 234)

The challenge is to produce a range of realistic activities and small projects that will allow the development of all the data handling work considered to be appropriate and useful at lower secondary level. If for a certain technique or concept no suitable activity can be found, it most likely does not deserve a place in the syllabus. Pupils need data handling tools to (i) understand and

critically look at data presented in the media and (ii) present and analyse data collected by themselves to answer questions related to their own world.

There is an ongoing debate whether or not data handling / statistics should be included in mathematics or whether it should stand on its own. The nature of data handling is rather different from pure mathematics. Mathematics deals with developing a logical system based on axioms through theorems. It leads to 'true' knowledge within the system. Statistics is concerned with making sense of data, representing and summarising the data and making decisions based on this data. However the outcomes are 'probable' not 'sure' as in the case with mathematical knowledge. In order to make this distinction clear to pupils, some educators plead for a separation of mathematics and data handling as two rather distinct disciplines. Others feel that at secondary school level both mathematics and data handling are attempts to describe, explore and understand patterns in numbers and shapes in the world around us and hence both should be under the umbrella of mathematics.

Data handling at the secondary level mainly covers the part of statistics called **descriptive statistics**. This part of statistics deals with the collection, representation (in tables, charts and diagrams) and analysis (calculation mean, measures of spread, etc.) of data. The part of statistics dealing with drawing conclusions, testing of hypothesis is called inferential statistics. This part is not covered rigorously in Module 6.

Section B: Reasons for including data handling in the curriculum

One of the main objectives, universally accepted, for secondary education is to prepare the learners for the world of work, economy, politics they are to participate in after completing their education. The reliance of the society on data analysis is obvious: decisions for the future, and for further development in the social, political, economic and other realms are taken based on statistics. In the learning and teaching of data handling the purpose should not be lost sight of. The aim is: **decision making**. Statistics (data handling) is all about exploring data in order to answer questions, and is the study of the variation of that data. The basic idea is: looking for patterns in the variation and trying to understand them. Some authors state that mathematics is about rigour, abstraction and certainty, while statistics is about uncertainty, investigation and estimation within a context. Others will not make such a distinction and feel that both mathematics and statistics are trying to describe the world around us by using models.

Data handling has decision making as its aim. For example:

A shoe factory will be interested in the most common shoe sizes in order to make a **decision** on the production process.

The Ministry of Education will be interested in the trend in the number of pupils starting each level of education in order to make **decisions** related to building of schools, training of teachers, etc.

Nearly all sciences, e.g., economics, business studies, political science, geography, biological sciences (including medicine), psychology, etc., all make use of statistics in order to make decisions based on the available data.

Section C: Misconceptions as to what data handling / statistics implies

That data handling is aimed at **decision making** is frequently lost sight of and some of the misconceptions among learners about data handling are:

1. Statistics is surveys

   Statistics is seen as restricted to questionnaire based data collection, representation and analysis.

   Data collection through **experiments** (scores on tossing a dice 500 times, length of maize plants grown using three different types of fertiliser, comparing two different teaching methods using an experimental and a control group) and through **simulation** (using random number devices to model a real life situation that is not available for experimentation or too dangerous for experiments or too costly, e.g., spread of a disease) form two other major methods which should be included in the school curriculum.

2. Statistics is art

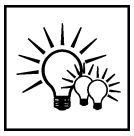
   A common misconception among younger pupils. They will produce pages of colourful neatly drawn charts with no interpretation.

3. Statistics is sums

   A common idea among older pupils, who have learned how to compute various values from raw data. However there is no rationale (why would you calculate a mode, median or mean in this particular case?) nor interpretation (what is this calculated value now telling us?).

4. Statistics is data collection

   Data are collected for the sake of it, without real purpose.



Practice task 1

The objective of this task is to find out pupils' ideas about data handling / statistics.

Ask pupils to write down (in small groups) what they understand by "data handling" (or if they are more familiar with the term statistics use statistics).

Categorise the responses. Are some responses reflecting the misconceptions listed above?

Using guided discussion discuss with pupils what data handling entails and what its purpose is.

Write an evaluative report on this class activity. Questions to consider are: What are pupils' ideas about data handling? Did the discussion go well? Did all participate? What ideas were brought forward? What ideas for questions or projects did come up? Why do you think it is important for a teacher to be aware of the ideas of the pupils?

Section D: How do people make decisions?

There is clearly a need to include the **purpose** in the learning of data handling. Textbooks frequently concentrate on the mechanical computation of statistics (number crunching) and production of charts. However data handling is at all times related to answering a question, to make a decision. Each and every person makes decisions—but frequently in an unscientific way. Decisions of people are often based on personal, informal data collection in cases where a more scientific approach would be more appropriate.

Write down how you make your decisions in various situations. For example when buying clothes, shoes or household items or when deciding how to spend a free weekend, which friends to invite to a party, etc.

Compare what you wrote with the information below and see in how far it agrees or disagrees with what you wrote.

For example:

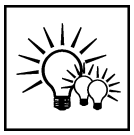
Lesego is going to buy a new pair of takkies.

She has read a report in a consumers magazine that compared takkies from three different makes on good fit, how long they last, how healthy they are for the feet, price, etc. From the three brands A, B and C, C is recommended as the best buy.

Lesego also asked some friends—they all are wearing brand A (the type considered to be 'cool', C is considered to be 'square'—the 'out' thing).

Which brand will she buy?

Research indicated that Lesego will go for brand A! People base their choices and decisions on the personal, informal data they collect—not on scientific (statistical) data analysis. People are influenced by friends, peers, advertisement (on radio and television, in the news papers) and make decisions based on this very selective and frequently one sided (biased) information.



Practice task 2

The objective of this activity is to make pupils aware of how they make decisions.

1. Ask pupils to discuss and write down (in small groups) how they would make decisions in the following situations, what facts they would take into account and how they would collect these 'facts' (add some more ideas appropriate for your class, or ask pupils to add their own situation requiring decision taking).

   (a) The type of drinks to buy for a class party.

   (b) The make of sport shoes to buy.

   (c) The type of clothes to wear to a party.

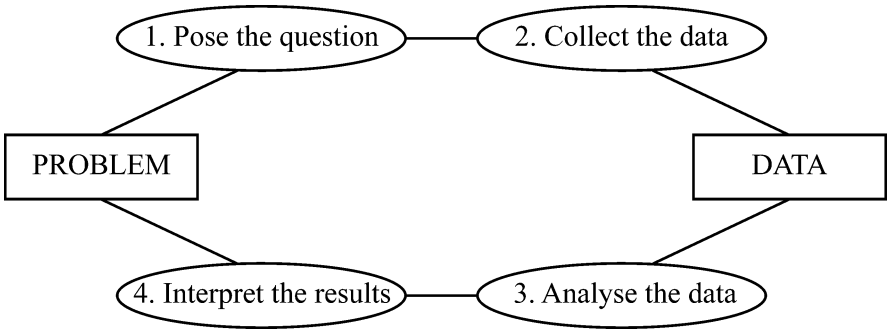
   (d) The make of school bag to buy.

   (e) The boy/girl to go out with to a disco.

2. Categorise the responses and analyse whether the pupils based their decisions on personal and informal data or whether they used a scientific method to collect and analyse data.
3. Use guided discussion to discuss with pupils what the most reliable method is to obtain data on which to base decisions in a specific situation.
4. Write an evaluative report on this class activity. Questions to consider are: What data do pupils use to make decisions? Did the discussion go well? Did all participate? What ideas were brought forward?

Section E: Stages in statistical investigation

A scientific data analysis goes through four distinct stages. The starting point is a question / problem one wants to find an answer to, or wants to make a well founded decision based on the available facts.



Posing the 'correct' question is important. If you do not pose a clear, well-defined question you might not know what data to collect or the question might be such that you cannot find data to answer it. Ill posed questions often lead to data that are impossible or very difficult to analyse and hence will lead to unreliable results.

Section F: Data and information

Asking pupils in your school their favourite colour gives you a list of the colours favoured by pupils. This is (raw) data. Measuring the height of each pupil gives you a collection of measures—again data. Facts, numbers, measures are collected on a certain target group of objects or people. The target group under study is called a population. A population is defined as the entire collection of objects having at least one similar characteristic. The pupils in your class can form a population that share the characteristics of being in the same class. The teachers in the district could form a target population, as do the maize plants growing in the school farm, the bolts produced by a machine in a factory, etc.

Once a set of data has been collected one can try to summarise, combine, compare the data—this leads to information. The height of all the pupils in your school is data, stating that the mean height is 1.62 m is information obtained from the data. This information can become data again if the mean height of pupils in each of the 24 classes in the school are collected.

A misconception is that data is 'number'. 5 is a number (actually a numeral representing the number), but "5 pupils in your class failed the mathematics tests" is data. Data is 'numbers' with a context. The word 'numbers' is in inverted commas as data can also be qualitative, i.e., not expressed in number form.

As teachers we have to encourage pupils to ask the following questions (and they will only do so if we do it ourselves).

a) How was this data collected?

b) What was the source of the data?

c) Is it accurate data or is it estimates?

d) Are they qualities, quantities (counts, measurements, rates, proportions, averages)?

Data is frequently presented in such a way as to favour a certain view or idea. For example, a factory wants to make their product look better than similar products from another factory. A government might want to play down the unemployment figures, a school might want to show how well their students are performing in an examination, a factory might want to show that their wages are high by selecting a specific type of average. By using selective data and representing it in a specific way the general public might be given the wrong impression. It is therefore very important to be critical towards data and the presentation of data in the media.

Self mark exercise 1

1. Practising statisticians claim that statistics is not mathematics. Why do you suppose they think so?
2. Data handling / statistics is included in all syllabi for the lower secondary school. Do you think pupils need to study data handling / statistics? Justify your stand.

Suggested answers are at the end of this unit.

Section G: Type of data

Data can be categorised in different ways. The type of data is very important as it determines the way the data can be represented and analysed. The different types of data are discussed in the following sections.

Section G1: Qualitative versus quantitative data (variables)

Qualitative data: Data that cannot be described by a number is referred to as qualitative data. For example: sex, region of the country, examination grades A, B, ..., preferred drink, political alliance, make of car, a person's blood group (O, A, B, or AB), etc.

In qualitative data the 'values' are words to identify defining categories. Qualitative data is also referred to as categorical data and leads most often to frequency counts for the categories.

Qualitative data are frequently given a numerical code, but any arithmetic done with these codes is meaningless. For example if gender is coded as 1 for male and 2 for female, then averaging to get 1.5 is meaningless. The numbers are only used to ease tabulating of results, but the numbers themselves have no numerical meaning.

Quantitative data is expressed by numbers: age, height, shoe size, income are examples of quantitative data.

As data is made up of a collection of **variables** you can also speak of qualitative variables and quantitative variables.

A **variable** is a factor which describes or characterises some aspect of the population and has different values (numerical or words/categories).

If you want to find out who is helping the pupils with their assignments, the variable is those people helping the pupil. The 'values' this variable can take are, for example: friend, mother, father, sister, uncle, ...

If you want to find the arm span of the pupils in a class, the variable is 'armspan'—a length taking all possible values within a certain range.

Section G2: Ordinal versus nominal qualitative data (variables)

Qualitative data can be ordinal or nominal. Data is **ordinal** when there is an underlying continuity between the words. For example, the response to "How often do you use games in the teaching of mathematics" might be rated on a scale as very often, often, sometimes, never. These words are clearly graded in order but their position on the continuum scale is incompletely defined. The 'distances' between these labels are not clear.

Similarly pupils might be asked to respond to the question "School uniforms should be abolished" by ticking one of the options: strongly agree, agree, indifferent, disagree, strongly disagree. These replies are the values taken by the ordinal variable.

Nominal data is data that names the categories, for example, favourite type of food of pupils in your class, number of different types of animals found in a wildlife park, a person's blood group, brands of rice.

Section G3: Discrete versus continuous data (variables)

Discrete data that can take only specific values or fall in a specific category. Discrete data is frequently the result of counting (quantitative data) or classifying (categorical data). Examples: shoe size, number of pupils in each form, number of songs on a CD, salary of the teacher, score on a dice, favourite drink.

It needs to be emphasised that both quantitative data and qualitative data can be discrete. It is a common misconception to equate qualitative and discrete data.

Discrete data can be displayed in bar charts (most appropriate for categorical data), bar-line diagrams (most appropriate for discrete qualitative data) or pie charts (discrete data with limited number of categories or values).

Continuous data. Data that can take any value within a certain range is called continuous data. Continuous data results in most cases from measuring, for example: pupils' mass, height, arm span, time spend on homework.

Continuous data is best displayed in histograms. In a histogram the frequencies are proportional to the area of the bar. When cases with bars of

the same width are considered the histogram becomes a bar graph with the bars touching each other. Details will be discussed in Unit 3, especially the problem of where the boundaries between two bars are to be exactly located.

Note that it is rather common to display certain discrete data (for example scores on a test, number of children in a family, i.e., numerical data that can be ordered) in a bar graph with the bars touching each other. This strictly speaking is not correct; to show discrete data properly, the bars should be separated.

Self mark exercise 2

State whether the following data is (a) qualitative or quantitative (b) discrete or continuous

1. The number of pips in oranges
2. The mass of oranges
3. The taste of oranges
4. The colour of cars in a car park
5. The number of desks in each classroom
6. The number of goals scored by the football teams that played on Saturday
7. The brands of toothpaste on sale in supermarket
8. The size of the dresses of the girls in school
9. The length of the sentences in this module (i) in words (ii) in centimetres
10. The attitude of pupils towards mathematics
11. The names of the teachers in the school
12. The age of the persons in your family
13. Number of voters that voted in the last elections
14. Salaries earned by teachers
15. Most popular music track during a school disco
16. The geographic districts (provinces) in the country
17. The audibility of teachers when speaking in class
18. Flavours of ice-cream sold in a super market
19. The grades (A, B, C, ...) by candidates obtained in an examination
20. The smell of different bathing soaps

Suggested answers are at the end of this unit.

You have come to the end of this unit. As a last assignment you are to bring what you have learned into the classroom.

Practice task 3

The objective of this activity is to make pupils aware of different types of data.

1. Write a lesson plan with these objectives (a) to introduce pupils to the different types of data (b) to practice and consolidate classifying of different types of given data (c) to illustrate different types of data with pupil generated examples.
2. Write an evaluative report on the lesson. Questions to consider are: Did pupils find difficulties in classifying data as discrete-continuous, qualitative-quantitative? Were the pupils able to generate their own examples? Were pupils well motivated to work on the activity? Were the objectives achieved? Did you meet some specific difficulties?

Present the lesson plan and report to your supervisor.

Summary

This unit introduced the subject of statistics by stressing **project-based learning**, **interpretation** of results, and **taking informed decisions**. These themes will be given more attention in subsequent units. Watch for an ongoing theme or motif about the true purpose of learning (and doing) statistics.

Unit 1: Answers to self mark exercises

Self mark exercise 1

1. In the view of some statisticians, mathematics deals with rigour and certainty. Statistics on the other hand deals with stochastic processes, i.e., processes related to chance. Hence the statistics lead to conclusions that are valid with a (stated) degree of probability.
2. Making decisions in today's world frequently requires the ability to analyse and interpret data, e.g., advertising claims, weather reports, statements of politicians, environmental impact reports, reports on population growth and spread of diseases. As a result data handling skills are increasingly fundamental to individuals in order to participate as well informed citizens in the society.

Self mark exercise 2

N.B. For some qualitative data, you can argue whether it is discrete or continuous. For example like or dislike for a subject: you can see these as discrete, but you might also look at them as the ends of a continuous scale with extreme ends: very great liking and very great dislike with all possible shades of like to dislike in between. In the table these arguable cases are indicated with √?

Question	Qualitative	Quantitative	Discrete	Continuous
1		√	√	
2		√		√
3	√		√?	√?
4	√		√	
5		√	√	
6		√	√	
7	√		√	
8		√	√	
9		√	(i) √	(ii) √
10	√		√?	√?
11	√		√	
12		√	√ (in years)	√ (exact time)
13		√	√	
14		√	√	
15	√		√	
16	√ (if named)	√ (if number)	√	
17	√		√?	√?
18	√		√	
19	√		√	
20	√		√?	√?

Unit 2: Methods of data collection

Introduction to Unit 2

In this unit you will look at methods of collecting data. There are three basically different ways to collect data: (a) using a survey (b) carrying out an experiment (c) carrying out a simulation. Data collection in the form of a survey frequently uses questionnaires and / or an interview. The unit will look at the characteristics of a well structured questionnaire. When collecting data, the question of the number of items needed for analysis in order to make well founded decisions will arise. This relates to concepts such as population, randomness of the sample and how well the sample represents the population. These will be looked at in this unit.

Purpose of Unit 2

The aim of this unit is to look at methods of data collection and questionnaire construction. The issue of sampling is introduced without going into details of the various methods of sampling.

Objectives

At the end of this unit you should be able to:

- distinguish between population and sample
- state reasons for sampling
- explain what is meant by a random sample
- explain common misconceptions in descriptive statistics
- select the most appropriate method of data collection (survey, experiment and simulation) for a given problem
- list four different types of surveys: a) questionnaire survey, b) interview, c) administering tests and d) structured observations
- explain and illustrate each of the above four types of surveys
- select the most appropriate type of survey given a problem requiring a survey method
- list and illustrate five characteristics of a well designed questionnaire
- explain the importance of a well designed questionnaire
- design a questionnaire for a given questionnaire survey
- illustrate and explain advantages and disadvantages of open or closed questions in a questionnaire
- list four points requiring special attention in a structured interview
- design a data collection sheet for a specified survey
- explain what a frequency distribution and a grouped frequency distribution is

- state reasons for using grouped frequency distributions
- explain disadvantages of using grouped frequency distributions
- design appropriate pupil centred activities for pupils to collect data

Time

To study this unit will take you about eight hours.

Unit 2: Methods of data collection

Section A: Data collection with a purpose

This unit is going to look at data collection. Although the emphasis in this unit is on methods of data collection, it should not be forgotten that prior to the collection of the data, objectives for the collection are to be set. The question: What is the purpose of collecting this data? is to be kept clearly in mind and answered before starting to collect data. The data that needs to be collected in order to answer the question is to be clearly identified. It is not uncommon for pupils to collect data NOT needed to answer the questions set or NOT to collect data that is needed. Pupils have the tendency at times to start to collect data in an ill organised way. This leads to situations where they might end up with data that is impossible or difficult to analyse or the omission of data needed to answer the question they set initially.

For example:

Suppose you want to find out what the favourite sport is of boys and girls in your school. Your objective is to make a list of the rank order of sports favoured by boys and a similar list for the girls in your school.

Pupil A made a questionnaire with one question:

Please write down what your favourite sport is.
My favourite sport is _____.

This pupil cannot answer the set question as the gender of the respondent is not included in the questionnaire. Data needed to answer the question was NOT collected.

Pupil B made a questionnaire as follows:

Please complete this questionnaire.

Are you male or female? (give a tick) Male ____ Female ____

In which Form are you? I am in Form _____

What is your favourite sport? My favourite sport is _____

This pupil collected data not needed to answer the question set in the objective, by including a question asking which Form the respondent is in.

It is very important that the data collected corresponds to the objective(s) set.

Self mark exercise 1

1. You want to find out how much time pupils in your class, on average, spend on preparing for a science test and whether there is a difference between boys and girls in this respect.
   a) What data would you collect?
   b) Give two examples of data that would be inappropriate to collect to answer the question.
2. You want to find out what is the favourite colour of pupils in your school.
   a) What data would you collect?
   b) Give two examples of data that would be inappropriate to collect to answer the question.
3. You want to find out to what occupation the pupils in your class aspire to.
   a) What data would you collect?
   b) Give two examples of data that would be inappropriate to collect to answer the question.

Suggested answers are at the end of this unit.

Section B: Population, sample and random sampling

For any question to be answered, a target group under study has to be identified before definition of variables, i.e., the data to be collected. The target group under study is called a population. A population is defined as the entire collection of objects with at least one similar characteristic. For example, if you want to find the proportion of the pupils at the school who regularly drink Coca Cola, you can ask each pupil in the school a question such as: "Do you drink Coca Cola at least once a week?" and record the answers. But talking with every pupil in the school would be time consuming and probably quite difficult to arrange. In this case, you might then try the question on just a portion of the pupils and based on the responses of that portion of the pupils make generalisation about all the pupils in the school on whether they regularly drink Coca Cola or not. This portion of the pupils is called the "sample." A sample is defined as a portion of the entire collection of objects of similar characteristics (the population). In the above example, the pupils in the school form the population and the portion of the pupils that were asked to respond to the question is called the sample.

The 'real' statistician will not see the physical objects as population and sample but the **collection of observations** obtained or **measurements** taken on the physical objects.

Need for Sampling

In order to get accurate data, on which decisions can be based, one would have to consider the whole population under study. However, in many cases it is not possible to obtain information about all members of a population for the following reasons:

1. The collection of information may destroy the sample, e.g., testing the life span of electric bulbs or electric fuses.
2. The population may be infinite, as in the results of throwing two dice (the experiment can be continued indefinitely).
3. It may be impracticable to make a measurement for every member of the population, e.g., finding the shoe size of all people in a country or the length of rice grains in a 50 kg bag. Even if a measurement could be made for each member of a population, considerations of time and expense usually dictate otherwise.

Due to these constraints it becomes inevitable in most cases to consider samples in obtaining information about a population. But how large should the sample size be in order for the sample to be representative of the population? The underlying idea to this question is how many measurements or observations should be taken so that the sample displays the true characteristics of the parent population? There is no definite answer that can be given to this question. It seems obvious that "the larger the sample size, the more accurate are the results," since the large sample is likely more closely an approximation to the population. If the sample size is small, there is a high chance that the results may not give a true picture of the parent population, because the sample size is too small to be representative of the population. However it is a misconception to think that the sample size is to be proportional to the population. A relatively small sample (say 150) can give reliable information on the whole population provided the sample is **representative**.

Purpose of Sampling

Since it is not possible at times to consider the whole population, a sample which is representative of the population is drawn from the population. That means the characteristics of the sample are the same as that of the population. The sample is linked to its population through **estimation theory**.

This theory states that if **statistics** such as mean, median, variance, standard deviation, etc., are calculated from the sample, it is possible to use these sample statistics to estimate **parameters** such as mean, median, variance, standard deviation etc., for the entire population from where the sample was drawn. The theory tries to generalise the characteristic of the sample to that of the parent population. This is a topic for advanced statistics and will not be discussed here.

Note the use of the word statistics in the above: a sample is described by measures such as mean, mode, median, variance, range—these measures are called statistics.

The word parameter refers to the same measures for the population. The mean of a sample is a statistic, the mean of a population is a parameter.

Estimation theory explains how statistics can be used to obtain parameters.

Random Sampling

In order for a sample to be representative of the whole population, each member of the population must have an equal chance of being chosen. A sample chosen in this way is called a random sample. To take a sample of 30 from all the pupils in your school you could place the names of all pupils on separate pieces of paper and put these in a box. You then pick from the box 30 slips. The names on these slips form the pupils that are going to be in your sample.

If your question to be answered requires pupils from each form to be in the sample and there are 18 forms in the school, you might make a sample by randomly picking two pupils from each form (giving a sample of size 36).

Anything that distorts data so that it does not fairly represent the population is called **bias**. The way in which data is collected and samples are formed must avoid bias so that any results produced are reliable. There is a famous example of an opinion poll giving the wrong prediction due to bias, when in 1948 an extensive telephone poll predicted that the new president of the USA was to be Dewey. However it turned out that Truman won the election. The incorrect prediction was due to bias: the telephone sample was not representative for the voting population as—at that time—many people did not have phones (apart from higher-paid voters who tended to vote for Dewey).

Pupils should understand that results obtained on a sample **may** apply to the whole population provided the sample was representative for the population.

Self mark exercise 2

What is wrong with the way that each of these samples is selected?

1. The head of the school wants to choose a new school uniform that pupils will be happy to wear. She asks 5 pupils from form 1 for their ideas. The new uniform is to be based on the responses of these pupils.
2. Batteries are produced by a machine. Every 50th battery is tested to ensure production standards are maintained.
3. A take away sells pies. Pies are made throughout the day. A sample from the first batch of pies produced is tested every day.
4. To decide what drinks to buy for the end of year party for all forms (200 pupils) you ask the first 10 pupils of your class entering the classroom for their choice. Based on their responses you buy the drinks.
5. A pupil is doing a survey into pupils' opinions about the school tuck shop. She picks pupils at random from those coming to the tuck shop to buy something.
6. You want to find out pupils' attitude in your school towards the learning of mathematics. You ask the 15 members of the mathematics club for their opinion.

Suggested answers are at the end of this unit.

Section C: Misconceptions in inferential statistics

The part of data handling / statistics dealing with the collection, representation and analysis (calculation of mean, etc.) is called **descriptive statistics**. The part dealing with the interpretation, testing of hypotheses and the drawing of conclusions is called **inferential statistics**. There are a number of well documented misconceptions in inferential statistics. It is important for a teacher to be aware of these so as to plan lessons such that the misconceptions are avoided—or when noted are addressed.

1) Concepts "population" and "sample."

   In data handling population refers to the complete group of persons, objects or measurements in which we are interested.

   A sample is part of that population, implicitly meant to be a representation of that population.

   These meanings differ from what the lay person understands by these words. A population is seen as a group of people in a certain region whether it be as large as a continent or as small as a village. The word sample has a wider general interpretation met in a context such as 'a free sample'.

2) There is no variability in the 'real world'.

   There is the idea that a sample accurately represents the population, i.e., that there is no variability (sampling errors). Pupils erroneously believe that, for instance, sample and population must have the same mean.

   For example: 1000 pupils were measured and the average height was found to be 168 cm. A sample of 10 pupils is taken from the 1000 randomly. The first pupil selected has a height of 162 cm. What will be the expected average height of the 10 pupils in the sample? The common error response is: 168 cm, i.e., the sample is thought to be identical to the population, so the given 162 cm is completely ignored.

3) Unwarranted confidence in small samples.

   Findings based on small samples are thought to be representative for the population, ignoring the fact that small samples, due to variability, might differ substantially from the population.

4) Insufficient respect for small differences in large random samples.

   Small differences in large random samples can be highly significant.

5) The size of a sample should be directly related to the population size.

   It is not so much the size as the fact whether or not the sample is representative that matters. If you are to test batteries on their life span, a well selected sample of 100 batteries will yield results reliable enough to make statements on the whole population of millions of batteries of that type. A sample of 200 would not yield more reliable results.

For students' misconceptions that are common in **descriptive statistics**, see Unit 3 Section F.

Section D: Methods of data collection

Data can be collected through

- Surveys (Section D1)
- Experiments (Section D2)
- Simulations (Section D3)

Surveys form a good starting point for pupils for gathering their own data and representing and analysing the data gathered. Experiments require more data gathering than surveys do, since they use both test and control subjects, or before-and-after testing. Simulations are similar to experiments but still more involved since they use random number devices—dice, spinners, random number tables or Random-key on calculator—to model real-world situations. In the following sections each method of data collection is discussed in some more detail.

Section D1: Surveys

What is a survey?

A survey is a method of collecting existing data to find the answer to a question. You collect the data from people (their mass, their favourite holiday activity, their view on the quality of the postal service) or objects (the content or mass of packaged cereals, different brands of cooking oil available in different supermarkets). You might want to know what activity pupils are involved in outside their school work, a survey, i.e., collection of data of pupils' responses to questions, might help you to answer the question. A survey based on the entire population is called a **census**. Surveys always have to start with a well defined **purpose**. The purpose of the survey is frequently formulated in the form of a **hypothesis**: a statement that might be true or false. To test the hypothesis a survey can be carried out.

Here are some examples of hypotheses:

1) Girls and boys perform equally well in mathematics in our school.
2) Pupils in the school are of the opinion that the amount of homework is about right.
3) Pupils feel that everybody in the school—including teachers—should participate in at least one sport.

Several factors have to be considered when setting a hypothesis:

a) Can you test it?

b) Can enough data be collected to give a valid outcome?

c) Can the type of data to be collected be analysed?

Self mark exercise 3

Are the following correctly formulated hypotheses, i.e., can they be answered? Justify your answer. If not restate the hypothesis in a correct form.

1. Group work in mathematics is better than whole class teaching.
2. There is life after death.
3. There will be an outbreak of malaria in Botswana every five years.
4. Pupils often participate in sporting activities.
5. Pupils are given too much homework.

Suggested answers are at the end of this unit.

Section D 1.1: Types of surveys

There are various types of surveys. We will look at the following:

- Questionnaire survey
- Interview
- Tests
- Structured observations
- Secondary data

Questionnaire survey

Questionnaires are collections of printed questions to be answered in writing by persons from the target group. Questionnaires are frequently used to collect data on opinions of people. A business might seek information about how the buyers view their products, a political party might want to know the opinion of voters in a region on their wishes for educational facilities in the region. In both cases a questionnaire presented to a well selected group of the population might be used to collect the required data.

Questionnaire design

The design of a questionnaire is most important. The way questions are phrased, the way answers are expected to be given and the overall layout of the questionnaire have an impact of the validity of the data collected.

Be very specific about what type of data you are looking for and why, to avoid questions not needed for the survey to be included in the questionnaire or questions needed for the survey to be (erroneously) omitted. The data collected in the questionnaire should be relevant to the research question(s) / objectives and it should be possible to analyse the data collected.

If a questionnaire is used to collect data:

a) Make it as short as possible. Long questionnaires lead to a poor response rate and/or unreliable responses (e.g., people quickly tick randomly, skip questions).

b) Questions should be phrased clearly and unambiguously. For example don't use two-in-one questions.

Examples: How often do you go to a sport activity and how do you go there?

This questions asks for two things and should be split. In addition the question is not clear: How often - do you mean per week, per month? Responses to such a question may be: not so often, from time to time, etc., which would be difficult to analyse.

c) Language and vocabulary must be at the level of the respondents i.e., as simple as possible so that everybody can understand what the question wants.

d) Do not include questions bearing no relevance to the research question, i.e., focus on questions directly relevant to the research objectives only. Each question should be directly related to one of the set objectives. If a question does not relate to an objective it is clearly out of place and should not be included.

For example if sex, school, number of years in teaching are NOT used as variables in the research there is no need to collect this data.

e) Avoid leading questions, i.e., be neutral. For example:

(i) Why do you use teaching aids? - implying that one IS using teaching aids, but perhaps the respondent is NOT using teach aids.

(ii) How would you improve your handball skills? - implying that they need improvement, the respondent might feel that his/her skills are inferior and do need further improvement.

(iii) Would you not agree that there is too much violence among pupils? - implying that the person setting the questionnaire thinks so and seeks support for his/her view. Better would be to ask:

Do you think that pupils are too violent with each other?
YES ☐ NO ☐

f) Ensure a neat, well typed and well spaced layout of the questionnaire with sufficient space for answering the questions.

g) Be aware that responses do not necessarily (i) reflect the real view of the respondent (ii) reflect the actions of the respondent especially when 'sensitive' questions (income or sex habits of person, for example) are asked.

h) Ask (in general) short questions which can be answered precisely.

i) Ask questions in a logical sequence.

j) In a closed questionnaire provide tick boxes to allow respondents to tick their choice. This makes it easier to complete and to analyse.

Questions that look rather straightforward can still cause problems. For example:

Are you a regular church goer? YES ☐ NO ☐

Are you in favour of trading on Sundays?
YES ☐ NO ☐ UNDECIDED ☐

A person might wonder whether a mosque or synagogue is considered as a church. A person might be against trade on Sundays but still wants to be able to buy 'take-away' or petrol on Sundays. These problems are generally overcome when an interview is used and the interviewer can clarify questions on the spot.

Administering the questionnaire

There are various ways in which the questionnaire can be administered.

a) Given personally to (random selected) individuals from the target population and to be filled in in your presence.

b) Given personally to (random selected) individuals from the target population and to be collected later. The individual can complete the questionnaire in his/her own time.

c) Given personally to a (random selected) group from the target population and to be filled in in your presence.

   A class of pupils can be asked to fill in a questionnaire related to menu and organisation of meals in the dining hall in your presence.

d) Posting questionnaires to the randomly selected individuals forming the sample.

Always when administering a questionnaire, in whatever form, explain—in an introduction or orally—the purpose of the questionnaire, and inform the respondents of what you are going to do with the data collected.

Self mark exercise 4

1. The following questions were included in a questionnaire on radio listening habits of pupils in the school. Explain why each question is unsuitable and rewrite the question so that it could be included in the questionnaire.

   a) When do you listen to the radio?

   b) What do you like about radio programmes?

   c) Don't you agree that the radio gives the best news reports?

   d) Shouldn't there be more educational programmes on the radio?

   e) How could the news readers improve their presentation of the news?

   f) How big are you?

   g) What do you think of the new improved menu in the dining hall?

2. The following questions were included in a questionnaire on the use of the school library. Explain why each question is unsuitable and rewrite the question so that it could be included in the questionnaire.

   a) Do you use the library frequently?

   b) Are you in the final year of school?

   c) Which books do you read?

   d) Don't you think that more books on science are needed in the library?

e) The journal section in the library is not up to date. Do you agree?

f) How could the library be improved?

3. Above were listed four ways to administer questionnaires. List advantages and disadvantages of each method mentioned.

4. a) Which of the following questions do you think are biased?

   b) Write down what makes them biased.

   c) Write a better question to replace the biased ones.

   (i) Most people think that the bus services are rather poor. Do you agree?

   (ii) Do you think that the bus services are better than they used to be?

   (iii) Do you think that smoking should be banned on buses?

   (iv) Do you agree that the buses should run more frequently?

   (v) Secondary school students often behave badly on buses. Do you think special buses should be provided for them?

   (vi) Don't you agree that beauty contests are degrading?

   (vii) Don't you think that the use of make-up is unnatural and bad for the skin?

   (viii) Do you agree that capital punishment should be brought back for murder and rape?

Suggested answers are at the end of this unit.

Format of the questions

Questions can be presented in two formats: closed or open. We look at them in more detail below.

I. Closed-ended: the respondents are to select responses from pre-specified answers

Advantages:

- A large sample can be used.
- Anonymity of respondents (ticking answers, handwriting will not 'detect' respondent).
- Responses can be easily (computer) analysed.
- Responses from different respondents are comparable.

Disadvantages:

- No guarantee that questions will be interpreted in the same way by all respondents.
- Questions might not be understood by respondents.
- Bias of researcher might come in: researcher chooses the questions and the responses, the respondents might not find 'their' questions and 'their' responses and just choose from what is presented. For example, only a

NO /YES option is offered but respondent feels "it depends on ..." or a scale does not offer a neutral position (strongly agree - agree - disagree - strongly disagree; the person might be neutral and neither agree nor disagree with the statement).

- Low return rate might affect reliability and validity of the research as the non respondents might differ in opinion from the respondents.
- Some respondents might be 'questionnaire tired' and random tick some answers.
- Generally respondents do not answer ALL questions—there are some blanks in the majority of cases. This might affect the outcome of the study.

To make a valid reliable closed-ended questionnaire a **preliminary pilot** in open format should be presented to a representative group. The responses given will indicate what questions (and how phrased) to include and what to include as possible responses in the closed-ended questionnaire. The closed questionnaire needs also to be piloted to reveal ambiguities, poorly worded questions or questions that are ill understood by respondents.

One should check a questionnaire on the following:

- Responses to questions should be consistent with the wording of the question.

  Question: Do you prefer hot or cold drinks? YES ☐ NO ☐

  The responses do not 'fit' the question. HOT ☐ COLD ☐ could have been given as responses.

- Responses should cover all possible answers to a question.

  Question: Do you discuss your grades with your friends?

  Responses ALWAYS ☐ SOMETIMES ☐ NEVER ☐

  If you did discuss your marks with most friends except for one or two, it will be difficult to answer the question. A box with USUALLY ☐ would improve the question.

- Responses offered should be balanced: the number of positive and negative responses should be the same.

  Example:
  The responses OUTSTANDING VERY GOOD GOOD POOR are unbalanced since there are three responses at the 'positive' side and only one at the 'negative' side. A more balanced scale would be

  VERY GOOD GOOD FAIR POOR VERY POOR

II. Open-ended: Each question represents a topic on which the respondent can freely comment.

Advantage

- Respondents can really bring out their point of view and present what they feel is most relevant.
- Good information for decision-taking often arises from thoughtful comments.

Disadvantages:

- Data is difficult to analyse.
- Responses of different respondents are difficult to compare.
- Bias of researcher can come into the categorisation of the responses.
- Answers might be scanty and lacking the data the researcher is looking for.

The main source of bias is in the analysis of the open-ended data. The researcher will have to use judgement as to how to categorise various responses. To safeguard against this bias the data should be analysed by at least TWO independent individuals.

Self mark exercise 5

1. Suggest 8 questions to be included in a questionnaire to be included in a survey of

   a) homework habits of the pupils in the school

   b) reading habits of the pupils in the school

   c) eating habits of the pupils in the school

   d) pupils' views on school rules

2. Design a questionnaire to test each hypothesis:

   a) Pupils in school are less superstitious than their parents / guardians.

   b) Girls, given a free choice, would hardly ever choose to wear a dress in preference to something else.

   c) Most cars these days use unleaded petrol.

   d) Pupils in form 3 are better in estimating length than pupils in form 1.

   e) Left-handed pupils are better in mathematics than right-handed pupils.

   f) The more time you use to prepare for a test the higher your score.

3. Which of the following questions included in various questionnaires are likely to get honest responses?

   a) I help enough at home. YES ☐ NO ☐

   b) Are you a kind person? YES ☐ NO ☐

   c) My family is poor. YES ☐ NO ☐

4. The following questions are likely to get responses that would not be very useful. Rewrite these statements so the responses would give you more useful information.

   a) I get 9 hours of sleep each night.
   ALWAYS ☐ SOMETIMES ☐ NEVER ☐

   b) I do my maths assignments.
   ALWAYS ☐ SOMETIMES ☐ NEVER ☐

   c) I go out during the weekend.
   ALWAYS ☐ SOMETIMES ☐ NEVER ☐

Suggested answers are at the end of this unit.

Interview

An interview is an oral, in-person administration of a standard set of questions that is prepared in advance. A structured interview (the questions have to be prepared as for a questionnaire and can be structured—respondent to choose from alternatives provided by the interviewer—or semi-structured) is generally a more reliable source of data collection than a questionnaire. All responses are coded, tabulated, and summarised numerically. The feedback received can be improved by probing and follow up questions. The quantity of data that can be collected is less than when using a questionnaire. Combining questionnaire design with interview is an option worthwhile to consider.

Advantages of using an interview to collect data

1. high response rate
2. more detailed information can be collected (follow up questions / clarifying question if misunderstood)

Disadvantages of using an interview

1. expensive and time consuming
2. possible bias in the way questions are asked and responses recorded

Administering tests

The term **"tests"** refers to the use of test scores as data. This technique involves subject response to either written or oral questions to measure knowledge, ability, aptitude, or some other traits that describe a characteristic of the subject.

For example:

(i) You want to find out pupils' skills in problem solving. You set a test with a variety of problems each testing a particular problem solving strategy. The scores on the test are your data.

(ii) You want to find out the skills of pupils in estimating the length of line segments in the range 0 cm to 50 cm. You prepare a test with various line segments drawn on it and pupils are asked to write down the estimated length. This will give you the raw data.

(iii) You want to find out pupils' skills in making models of 3D objects. The test set to pupils is to make a prescribed 3D object.

Tests require careful construction to ensure they will test what you want them to tests (i.e., the test is to be valid). Each test item must relate clearly to an objective to be tested. As tests require scoring a scheme is to be prepared as to how scores will be awarded to the responses (particularly how to allocate scores to partial correct responses).

Structured observations

Data can be collected by observation, visually and / or auditory, and the observations systematically recorded. For example, if you want to find out whether a pedestrian crossing should be made at a particular place you could use an observation sheet to tally the number of people crossing at or near that

point. If you want to find out whether boys or girls ask more questions in a classroom you can sit in the classroom and tally whenever a question is asked.

Data collection sheet are frequently used when collecting data by counting. When we want to find the favourite colour of pupils in the class the responses of each pupil could be recorded in a simple record sheet.

Data collection sheet

Colour	Tally	Frequency
red	| | |	3
blue	𝍸 | |	7
black	..	..
green	..	..
white	..	..
silver	..	..
orange	..	..
.....	..	..

The colour mentioned by each pupil is recorded in the **tally** column by a single stroke. To make counting easier groups of 5 are recorded as 𝍸.

The total number of times each colour is mentioned is called the **frequency**.

A table for discrete data with the totals included is called a **frequency distribution**.

For large amounts of discrete or continuous data the data is organised into **groups** or **classes**. Data collected in groups with totals included is called a **grouped frequency distribution**.

Grouped frequency distribution table: mass of boys in a class

Mass m kg	Tally	Frequency
$50 \leq m < 55$	|	1
$55 \leq m < 60$	𝍸 |	6
$60 \leq m < 65$	𝍸 𝍸 | |	12
		..
		..
		..
		..
.....	..	..

The masses of boys in the class are grouped in **class intervals** of equal width (5 kg). $50 \leq m < 55$ means 50 kg or more, but less than 55 kg.
The boundaries between classes are at 55 kg, 60 kg, etc. This should be made clear to pupils by asking questions such as the following.

Questions to ask are:

1. Elliot has a mass of 54.9 kg; in which class interval is his mass recorded?
2. Tumisang has a mass of 55.0 kg; in which class interval is his mass recorded?

In presenting an exercise to pupils on making distribution tables ensure you cover all the cases:

(a) frequency distribution of discrete data

(b) grouped frequency distribution of discrete data

(c) grouped frequency distribution of continuous data

An example of each is given in the following self mark exercise.

Self mark exercise 6

1. The colour of 40 cars in a car park are listed.

red	blue	red	silver	red
red	red	white	blue	silver
red	black	green	red	blue
green	blue	black	white	yellow
white	white	red	blue	silver
green	blue	green	yellow	black
white	white	green	blue	red
red	white	black	silver	green

   a) Make a frequency distribution for the data.

   b) Which colour of car is most popular? (This colour is called the mode.)

2. The ages of the 40 teachers in a school are listed.

27	35	26	33	24
43	34	20	47	56
42	49	57	34	54
29	39	50	21	37
58	30	28	26	20
34	33	27	41	59
47	62	52	29	30
25	37	29	44	22

   a) Make a grouped frequency distribution using classes 20 - 29, 30 - 39, 40 - 49, 50 - 59 and 60 - 69.

   b) In which class interval are most of the teachers? (This is called the modal class.)

3. The height of 36 girls in a class was recorded to the nearest cm.

148	155	156	175	160	165
159	179	161	173	167	154
163	158	160	172	147	170
158	155	165	178	168	157
166	171	157	159	172	162
155	172	165	168	157	173

   a) Make a grouped frequency table taking the class intervals for the height h as $145 \leq h < 150$, $150 \leq h < 155$, etc.

   b) How many girls are less than 160 cm?

   c) How many girls are 155 cm or taller?

4. a) Make a data collection sheet to record the month in which the pupils in the class were born.

   b) Make a frequency table for the data.

   c) In which month did most births occur?

Suggested answers are at the end of this unit.

Secondary data

The above listed methods to collect data through surveys are called **primary data** collection. The required data is directly obtained by the person collecting the data. The data is the result of direct observation, tests, questionnaires and interviews. At times one might make use of published data, for example, government statistics on number of pupils attending the various levels of the education system, the number of teachers teaching each subject, the number of children born in each district, etc. This type of data is called **secondary data** as it was not directly collected by the researcher. Secondary data need to be approached with great caution. Are the secondary data up-to-date (i.e., is the data accurate)? How was the data collected (i.e., is the data reliable)?

In your assignment you are required to facilitate the learning of pupils on one (or more) of the following issues related to data collection:

a. What data is to be collected in a given situation
   (See self mark exercise 1)

b. How is a sample to be taken to ensure the outcomes are reliable
   (See self mark exercise 2)

c. How to formulate correct hypotheses (See self mark exercise 3)

d. How to formulate correct questions to be included in a questionnaire
   (See self mark exercise 4)

e. Designing a questionnaire to collect data to answer a specific question or hypothesis (See self mark exercise 5)

f. How to tabulate various types of data (Self mark exercise 6)

You might decide to cover several of the above by setting a small project to groups (4 per group is recommended). However prior to pupils embarking on data collection, preparing a questionnaire, setting hypotheses, recording the data in tables, they have to be made aware of and have to discuss the issues covered in the self mark exercises 1 to 6.

Practice task 1

1. Choose one (or more) of the topic listed above (a to f).
2. Write a lesson plan with clearly stated objectives. Prepare worksheets for the pupils to work in groups.
3. a) Write an evaluative report on the lesson. Questions to consider are: Did pupils meet difficulties? Were pupils well motivated to work on the activity? Were the objectives achieved? Did you meet some specific difficulties in preparing the lesson or during the lesson?

   b) Present the lesson plan and report to your supervisor.

Two-way tables to tabulate data

Data can at time be collected and represented in **two-way tables**. These tables can be used when two different variables are involved. Reading and interpretation of these tables will be one of the objectives.

Self mark exercise 7

1. The table displays data on pupils with spectacles in a school.

Pupils wearing glasses in the school

	Wear glasses	
	YES	NO
Girls	82	273
Boys	86	295

a. How many girls wear glasses?

b. How many pupils do not wear glasses?

c. How many pupils were in the survey?

d. Do the results prove or disprove the hypothesis "More boys than girls wear glasses"? Justify your answer.

2. The two-way table displays the results of a survey to test the hypothesis "More girls are left-handed than boys."

Pupils being left-handed in the school

	Left-handed	
	YES	NO
Girls	10	60
Boys	15	90

Do the results prove or disprove the hypothesis? Justify your answer.

Suggested answers are at the end of this unit.

Guidelines for designing and using an observation sheet

In the following assignment you are to guide pupils in the designing of a data collection / data observation sheet. Next pupils in groups collect the data and present the data in tabulated form to the class, explaining their design and pointing out any valid conclusions that can be drawn from the data.

The following information is to be given and explained to pupils:

Draw up your observation sheet after you have decided on the categories you are going to use.

Make some initial decisions on how you hope to organise and analyse the data.

Decide when and where to collect the data.

Practice task 2

1. Set to groups of pupils the task to design an observation / data collection sheet and to collect data on topics of their choice. Suggestions could be given if needed, for example (see below) or add some other ideas from pupils or yourself.

   a. pupils coming late for the first class in the morning

   b. pupils leaving the classroom (with permission) during lessons

   c. how frequently pupils use toilets

   d. places pupils group during morning break

   e. clubs in which pupils are participating

   f. favourite magazine

   g. favourite type of music

   h. commonest shoe size

For all tabulations collect data for boys and girls separately.

Pupils present their results to the class and explain the ways they collected their data and the ways they tabulated the data. They are also to mention valid conclusions they draw from their tabulated data.

2. a. Write an evaluative report on the activity. Questions to consider are: Did pupils meet difficulties? Were pupils well motivated to work on the activity? Were the objectives achieved? Did you meet some specific difficulties in co-ordinating the activity?

   b. Present the report to your supervisor.

Section D2: Experiments

The scientific method is used to collect data resulting from an experiment. Unlike in a survey where existing data is collected in an experiment, a situation is experimentally manipulated and the data resulting from the experimentation is collected. This frequently takes the form of collecting data on an experimental group and a control group. To find out whether or not pupils learn expansion of algebraic expression better using concrete manipulatives or using a multiplication table model, the researcher will use two comparable groups of pupils (a pre-test establishes that existing groups are comparable on algebra knowledge at the start of the project, or two comparable groups are formed). Next the two groups cover the topic using one of the two methods. Data from a post-test is used to determine whether or not one of the two methods is superior.

These types of situations are very common: Is medicine X more effective than medicine Y to cure a disease?

Do batteries of factory A last longer than batteries produced by factory B?

At other times the purpose of the experiment might not be comparing but testing, for example a product: Under what pressure do gas bottles explode?

Will cables marked that they safely can be used up to a load of 500 kg indeed not break under a lighter load?

The outcomes of the data analysis of these experiments is for decision taking. Which batteries to buy? Which medicine to bring on the market? How to set the machines making the gas bottles in a factory? What cables to use for a lift so people can use it without fear that the cables might break?

In experiments, with the exception of the variable being changed all, other conditions should remain the same throughout the experiment. For example, if you are looking at the effect of fertiliser on the growth of sorghum plants the seedlings should be of the same height and strength at the beginning of the experiment. The trays with seedlings should be in the same environment (same amount of light, same temperature condition) and only the amount (or type) of fertiliser should vary—ensuring that one tray with seedlings has no fertiliser applied at all (the 'control' tray).

Practice task 3

The objective of this activity is for pupils to design an experiment, and to collect and tabulate data from the experiment.

1. Give different experiments to groups of pupils to enhance discussion when the outcomes are presented to the class.

   Some suggestions are:

   a) What is the reaction time of boys / girls to catch a ruler between thumb and finger when the ruler is released? Is there difference between using left hand or right hand?

   b) How many 'nonsense' words can pupils recall after studying them for x minutes?

   c) How accurately do pupils measure the height of a desk?

   d) How many times does a drawing pin land head up?

   e) How strong a pillar can you make with an A4 sheet of paper?

   f) How high does a table tennis ball rebound from different surfaces?

   g) How fast do pupils react (pressing a buzzer) after seeing a red light (a green bulb and a red bulb are lighted randomly at various intervals)

2. a) Write an evaluative report on the activity. Questions to consider are: Did pupils meet difficulties? Were pupils well motivated to work on the activity? Were the objectives achieved? Did pupils control factors that might lead to unreliable results? Did you meet some specific difficulties in facilitating this activity? What did you learn yourself from the activity?

   b) Present the lesson plan and report to your supervisor.

Section D3: Simulations

There are questions people like to be answered for decision making but the data cannot be obtained by surveys or experiments. For example: the spread of a disease among people, animals, trees or plants; the number of the impalas in a wild life park in relation to the number of lions; the (average) length of a queue or the (average waiting) time at a public phonebooth; the waiting time of cars to find a parking place in the car park near a shopping centre.

Data to investigate these questions is difficult or impossible to obtain. Data can be generated in a simulation (using random numbers from calculator or table) as all data involves aspects of randomness.

A note on random numbers.

Random numbers can be found in tables or on calculators having a RANDOM key. Random numbers consist of lists of digits 0, 1, 2, 3, ... 9 which are such that each digit has an equal probability of appearing (p = 0.1 for each digit). The numbers might be listed individually or grouped in some way (for example in groups of three).

The RANDOM key on a calculator produces random numbers. Pressing the key produces a number between 0 and 1 to 3 decimal places (end zeros are not displayed!). The decimal part is used to make a list of random numbers (do not forget the 0 if this is to be the third digit).

Here is a list as generated by the calculator.
817 806 553 734 790 376 668 379 735 916

Suppose you have a group of 70 pupils and you want to select 8. The people can be given the numbers 01 02 03 ... to 70

You change the random number list so the digits are grouped in pairs:
81 78 **06** **55** **37** **34** 79 **03** ..

The person with the bolded numbers will be in the sample, because these pairs of digits lie between 1 and 70.

Another way to select the persons in your sample is multiplying the random number by 70 and rounding to the nearest whole number. The whole number identifies the person to be included. Continue this process until you have the size of sample required.

For example:

A car park has 50 parking lots. It opens at 8.30 am at the same time as the shops near the parking place open. What will be the average time to fill the car park?

Random digits are now used to represent the number of cars arriving each minute.

Digit	Number of cars arriving each minute
0 or 1	0
2 or 3	1
4 or 5	2
6 or 7	3
8 or 9	4

Using the random numbers listed on the previous page:
817 806 553 734 790 376 ...

The first random digit is 8 and this corresponds with 4 cars arriving.

The second random digit is 1, this corresponds with 0 car arriving.

Next digit is 7, corresponding with 3 cars arriving, etc.

So the number of cars arriving are as follows:
4, 0, 3, 4, 0, 3, 2, 2, 1, 3, 1, 2, 3, 4, 0, 1, 3, 3, ...

At the same time a dice is used to determine the number of cars leaving the car park but in the first 15 minutes no cars depart.

Number on dice	Number of cars leaving
1 or 2	0
3 or 4	1
5 or 6	2

A table can now be used to record how many cars are parked each minute until there are 50 cars and the park is full. For example, the table for the first 20 minutes could look as illustrated.

Time (min after opening)	Arrival	Departure	Total cars in park
0	0	0	0
1	4		4
2	0		4
3	3		7
4	4		11
5	0		11
6	3		14
7	2		16
8	2		18
9	1		19
10	3		22
11	1		23
12	2		25
13	3		28
14	4		32
15	0	1	31
16	1	2	30
17	3	2	31
18	3	0	34
19	3	1	36
20	3	1	38
...			

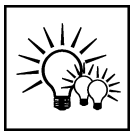
Continuing the table until 50 cars are in the car park gives the time needed to fill the park.

The simulation is repeated several times, or results of pairs of pupils (ensure they start with a different random number) are pooled together.

Averaging the times obtained gives the mean time needed to fill the car park. The model is clearly based on assumption: number of cars arriving per

minute in the range 0 to 4, number of cars leaving is 0, 1 or 2. Early in the morning this might be realistic, towards closing time of the shops obviously hardly any cars will arrive in the park and more will be leaving—arrival and departure will have to be simulated in other forms.

Similar methods as described above can be used to determine the average waiting time in a queue. In each time period (3 or 5 minutes for example) take the last person to join the queue and determine how long the person has to wait. The numbers of people attended to per time period can be simulated by numbers on a dice, the number of people arriving in the queue by random numbers. If the queue exceeds say 30 people, the 31st (or more) will leave since they don't want to wait for so long.



Practice task 4

Objective of this activity is for pupils to collect and tabulate data from a simulation.

1. Give the same simulation to different groups of pupils.You can choose from the suggestions in the above section or design your own. Ensure each group uses different start digits for the random numbers they use. Outcomes of the groups are to be presented and discussed with the whole class.
2. a) Write an evaluative report on the activity. Questions to consider are: Did pupils meet difficulties? Were pupils well motivated to work on the activity? Were the objectives achieved? Did pupils bring up ideas to improve on the simulation to make it more realistic? Did you meet some specific difficulties in facilitating this activity? What did you learn yourself from the activity?

   b) Present the lesson plan and report to your supervisor.

Section D3.1: Pupils simulation activity: Dice and disease in the classroom

Below is an outline for a classroom activity with pupils to simulate the spread of a disease. What follows are the lesson outline and notes for the teacher. Work through them before trying out the activity with your pupils in assignment 5.

Assumptions:

- Pupils are aware that germs spread disease.
- You can avoid catching some diseases by avoiding risky encounters.
- The definition of a risky encounter varies with illness. For example, such illnesses as a common cold may be spread by an activity as common as shaking hands, whereas AIDS is frequently spread by sexual contact. Cures do not exist for either illnesses.

Goal of activity:

To model the exponential growth of the common cold, AIDS, or any other communicable disease. The activity underscores the effect that a friend's or partner's previous behaviour may have on a current relationship and on a society at large. The activity works best with thirty or more pupils.

A risky encounter

If the sum is less than or equal to the **cut-off number** 5, the encounter has been risky.

A **risky encounter** means that if either of the pupils is a carrier of the disease, it is passed to the other pupil.

	1	**2**	**3**	**4**	**5**	**6**
1	**2**	**3**	**4**	**5**	6	7
2	**3**	**4**	**5**	6	7	8
3	**4**	**5**	6	7	8	9
4	**5**	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

The probability p that an encounter will be risky (cut-off number less than or equal to 5) is

$$p = \frac{10}{36} \approx 0.28$$

If, say, ten infected individuals are in the population at a particular time, those ten individuals will interact with ten other individuals and a fraction, p, of those encounters, approximately three encounters, will be risky and result in new infections.

It is possible to model the effect of an **education campaign**. Suppose that for three stages the cut-off number is set to five, as described previously, and in subsequent stages the cut-off number is reduced to three. This change lowers the infection rate, which could happen if the population were educated about the dangers of risky encounters.

The class activity

The activity is introduced by the teacher reminding the pupils of the points listed under "Assumptions."

The pupils represent a group of people interacting socially. A possibility exists that members of the group are infected with a communicable disease. During interaction between individuals the disease is passed along if an individual has contact with an infected person and has a risky encounter. The aim is to track the spread of the disease and to observe how many people become infected.

- Pupils are NOT informed when encounters will be risky and when not.
- Pupils are given an ID number, an encounter sheet and a pair of dice.

- Pupils walk around the room to ‘encounter’ friends.
- On the sheet, pupils enter the ID number of the person they ‘encounter’ and the score on the pair of dice thrown during this encounter.
- Both pupils involved in the encounter write the information on their data sheet.
- The stage should be timed (approximately 2 minutes) such that pupils can make two to four encounters.
- Teacher announces end of stage 1 and start of stage 2. In stage 2 the same procedure as stage 1 is followed. This is once more repeated in stages 3 to 5. (More stages can be included.)

My ID number

Stage 1

ID number				
Dice score				

Stage 2

ID number				
Dice score				

Stage 3

ID number				
Dice score				

Stage 4

ID number				
Dice score				

Stage 5

ID number				
Dice score				

After the three stages data is going to be collected and analysed. All the ID numbers are on the board.

(i) The teacher announces (and explains the reason for) the cut-off point for risky encounters.

(ii) Pupils identify (circle) on their encounter sheet whether or not their encounter was risky.

(iii) Randomly one ID number is chosen: this is the person carrying the disease (say ID xxx).

(iv) Systematically the teacher guides pupils through the 'encounters' looking at column 1, stage 1.

a) Who met with ID xxx?

b) Was the encounter risky or not?

NO: go to column 2 and repeat a. Who met with ID xxx? and look at whether the encounter was risky or not.

YES: write down (on the board) the ID yyy of the person also infected and go to column 2. Who met with ID xxx or ID yyy? Where the encounters risky or not?

Step by step the first stage is analysed and the total number of pupils infected by the end of stage 1 is noted.

Next analyse stage 2 and 3 in a similar manner. Present the data graphically.

The data can be analysed again (or the simulation can be repeated) with different scenarios. For example, increase or decrease probability of a risky encounter; simulate the effect of an information campaign by lowering the cut-off point for encounters in stage 3 and higher.

Practice task 5

Objective of this activity is for pupils to collect and tabulate data from the simulation.

1. Carry out the simulation dice and disease activity with your class.
2. a) Write an evaluative report on the activity. Questions to consider are: Did pupils meet difficulties? Were pupils well motivated to work on the activity? Were the objectives achieved? Did pupils bring up ideas to improve on the simulation to make it more realistic? Did you meet some specific difficulties in facilitating this activity? What did you learn yourself from the activity?

   b) Present the lesson plan and report to your supervisor.

OPTIONAL

Section D3.2: Analytical model for dice and disease (for background knowledge of teacher)

Analytical, or pencil-and-paper, models can also be devised to describe the activity.

Assumption:

The disease spreads through the population in stages and each individual interacts with exactly one other individual in each stage.

Our goal is to produce a sequence of numbers that we denote

$$\{\mathbf{I_0}, \mathbf{I_1}, \mathbf{I_2}, \mathbf{I_3}, \mathbf{I_4}, \mathbf{I_5}, \ldots\ldots\}$$

in which I_n stands for the number of infected individuals at the n-th stage.

The initial population of infected individuals can always be taken as $\mathbf{I_0} = 1$.

We look for a rule that has the form

$\mathbf{I_{n+1}}$	=	$\mathbf{I_n}$	+	$\mathbf{N_n}$
Number of infected at stage n		Number of infected at stage n		Number of new infected at stage n

If I_n infected individuals are involved in stage n, they will interact with I_n healthy individuals, producing

$$p \times \mathbf{I_n}$$

newly infected individuals.

Therefore

$\mathbf{I_{n+1}}$	=	$\mathbf{I_n}$	+	$p\mathbf{I_n}$
Number of infected at stage n+1		Number of infected at stage n		Number of new infected at stage n
	=	$(1+p)\mathbf{I_n}$		

$\mathbf{I_0} = 1$

$\mathbf{I_1} = (1+p)\mathbf{I_0} = (1+p)$

$\mathbf{I_2} = (1+p)\mathbf{I_1} = (1+p)(1+p) = (1+p)^2$

$\mathbf{I_3} = (1+p)\mathbf{I_2} = (1+p)(1+p)^2 = (1+p)^3$

$\mathbf{I_n} = (1+p)\mathbf{I_{n-1}} = (1+p)(1+p)^{n-1} = (1+p)^n$ for n=0, 1, 2, ..

Refining the model

Remove the assumption that infected individuals interact only with healthy individuals.

The number of newly infected people we called N_n is really proportional to the number of interactions between healthy individuals and infected individuals.

Let **T** be the total population.

Assuming no births, deaths or immigration, it follows that the number of healthy individuals at stage n is $\mathbf{T} - \mathbf{I_n}$.

The number of interactions between infected individuals and healthy individuals is proportional to the product

$$\mathbf{I_n} \times (\mathbf{T} - \mathbf{I_n})$$

So we can write

$\mathbf{I_{n+1}}$	=	$\mathbf{I_n}$	+	$\mathbf{N_n}$
Number of infected at stage n+1		Number of infected at stage n		Number of new infected at stage n
	=	$\mathbf{I_n}$	+	$a\mathbf{I_n} \times (\mathbf{T} - \mathbf{I_n})$

Note that the number of newly infected individuals is zero for $\mathbf{I_n} = 0$ or $\mathbf{I_n} = \mathbf{T}$

Section E: Choice of data collection method

The method of data collection to be used depends on the nature of the question one wants to answer, i.e., the hypothesis to be tested.

For example:

a) You want to test the hypothesis that there is no difference in the leisure time activities of boys and girls whatever their age.

You will in this case collect data, using a questionnaire or interview, of a sample of boys and girls of different ages and ask them what their leisure time activities are. Hence you collect data using a survey.

b) You think that the height that pupils can jump depends on the length of their legs. You will collect data by using an experiment: measuring length of legs and the height of the jump of a random sample of pupils (separating boys and girls).

c) Waiting time in queue during lunch time.

Although you could obtain data by observation, this would take a long time to gather sufficient data. A simulation might be the most appropriate way to collect data.

Self mark exercise 8

Which form of data collection would you use in the following cases:

1. Germination rates of seeds.
2. Pupils are better at estimating mass than volume.
3. Litter: causes and possible solutions.
4. Teenage pregnancies: causes and possible solutions.
5. Blind-folded short people can walk in a straight line for a greater distance than blind-folded tall people.
6. Girls have better memories than boys.
7. What is the most popular magazine among pupils in your school?
8. Reaction time with right hand versus left hand.
9. Preferred type of video watched.
10. People are of the opinion that animals should not be used for drug testing.

Suggested answers are at the end of this unit.

Practice task 6

Objective of this activity is for pupils to carry out a small project by setting a question or hypothesis to be researched, collecting and tabulating the relevant data, and stating any valid conclusion from the tabulated data. Allow pupils to choose their own research question.

1. a) Write an evaluative report on the activity. Questions to consider are: Did pupils meet difficulties? Were pupils well motivated to work on the activity? Were the objectives achieved? Did pupils bring up ideas to improve on the validity of the data to be collected? Where the research questions realistic? Did you meet some specific difficulties in facilitating this activity? What did you learn yourself from the activity?

   b) Present the lesson plan and report to your supervisor.

Summary

This unit has ranged widely over the topic of how to gather "good" data from a sample. Your teaching outcome should be (eventually) a few classroom projects which engage your students in collecting real data. Don't expect that the projects which your students can tackle will convey all, or even most, of the concepts taught in this unit. And conversely, don't try to teach the unlearned concepts by lecturing. Your students will better grasp the reality of data handling from doing the projects and learning from them.

A good practice is to assign a different data-gathering project to each grouping of two or three students. Each group can add new conceptual layers to its project as the class progresses through the stages of data gathering, representation, and so on.

Unit 2: Answers to self mark exercises

Self mark exercise 1

1. a) Gender and time spend on preparation.

   b) Any other from 1a. such as age, favourite leisure time, etc.

2. a) Using a sample of pupils in which pupils of all classes are represented and about an equal number of boys and girls collect their favourite colour.

   b) Anything different from 2a.

3. a) As the occupation aspired to might depend on the gender, data to collect is: gender and occupation he/she wants to take up.

   b) Anything other than in 3a, e.g., form of the pupil – as all pupils are in your form.

Self mark exercise 2

1. Biased sample as pupils from all classes are not involved. Sample too small.
2. Machines frequently make systematic errors, i.e., every tenth or hundredth item might have a defect. Hence a sample of products produced by a machine should not take equally spaced items.
3. Biased sample as the first production is not representative for the product produced throughout the day. The first batch might be very outstanding as workers are just starting, the ingredients have just been freshly made, etc. A sample should include pies produced throughout the day, i.e., from each batch.
4. Biased sample as the pupils are not representative for the all the pupils in the school.
5. Biased sample. Those not going to the tuckshop are not included!
6. Biased sample as the pupils participating in the mathematics club are likely to have positive attitudes towards mathematics. A sample representative for the whole school should be taken.

Self mark exercise 3

1. The statement is not specific enough as it is not clear what is meant by "better." Leading to better results? Or "better" in the opinion of pupils?

   Hence correct forms could be:

   Pupils prefer group work in mathematics over whole class teaching.

   Pupils' achievement in mathematics is significantly better when they work in groups than when a whole class teaching method is used.

2. No data can be collected to prove or disprove the statement.

   What could be investigated, for example, is pupils' opinion in junior secondary on whether or not they believe that there is life after death.

   The hypothesis could then be that pupils in junior secondary school believe that there is life after death.

3. Not enough data can be collected to give a valid outcome. Data over a very long period 50 – 100 years would be needed to see whether or not a five-year cycle can be discovered.

4. Not specific enough as the word 'often' is open to multiple interpretation.

   Pupils participate in sporting activities more than three times a week.

5. Not specific enough as 'too much' is very relative and needs to be specified.

   Pupils are given homework that will take them daily 2.5 hours to complete according to the teachers.

   Pupils spend between 2 – 4 hours daily to complete their homework.

Self mark exercise 4

1. a) Question is not specific enough. Responses could be "At the weekend" "At 6 o'clock" "When I am working on my homework."

   b) The responses will be so varied that no proper analysis will be possible as the question is not specific enough.

   c & d) Are leading questions

   e) The open question will lead to a wide variety of responses that are difficult to analyse and is biased: it suggests that the news reader need to improve the presentation but people might feel that the news is read all right.

   f & g) Irrelevant questions in relation to the topic: radio listening habits.

   Suggestions for questions and their format to include in the questionnaire are below (need to be worked out in more detail):

   a) Indicate the times and days you generally listen to the radio by completing the table below.

DAY	TIMES (for example 6.00 am to 6.30 am, 5.00 pm to 8.00 pm)
Monday	
Tuesday	
Wednesday	
Thursday	
Friday	
Saturday	
Sunday	

b) What radio programmes do you listen to? Complete the table.

Programme	Listen ALWAYS	MOST OF THE TIME	SOME-TIMES	NEVER
News				
Request programmes				
Top 10				
Educational programmes				
Church services				
… list here the programmes offered on the local radio service ..				

c) How do you rate (tick) the level of the news supplied by the following media?

MEDIA RATING

	VERY GOOD	GOOD	FAIR	POOR	VERY POOR
Daily Times					
Guardian					
Sun					
Radio					
TV					
…. (others)					

d) Indicate whether you feel that more or less time should be given to the following programmes or whether it is OK.

PROGRAMME Time allocated

	SHOULD INCREASE	IS OK	SHOULD REDUCE
News			
Sport			
Request programmes			
Popular music			
Classical music			
Educational programmes			
Religious programmes			
…			

e) How do you rate the reading of the news? (tick your choice)

(i) NEEDS IMPROVEMENT _____

(ii) IS FAIR _____

(iii) IS VERY GOOD _____

If you ticked (i) suggest what improvement is needed.

_____

2. a) Not specific enough to lead to useful data.

How often do you use the library during a week?

Never _____

1 or 2 times _____

3 or 4 times _____

5 or more times _____

b) Why only final year?

In what Form are you? I am in Form _____

c) The question is too general to give data that can be analysed meaningfully.

Better would be to ask:

Which of the following books do you read or borrow from the library? Complete the table.

	NEVER	SOMETIMES	FREQUENTLY
Novels			
Science fiction			
Study books			
Picture books			
Dictionaries			
Encyclopaedia			

d) Biased question suggesting that more science books are needed.

Better to formulate as:

How do you rate the following book sections in the library? Complete the table.

BOOKS	MORE NEEDED	SUFFICIENT	TOO MANY
Novels			
Science fiction			
Picture books			
Dictionaries			
Encyclopaedia			

Study books in

Science			
Mathematics			
History			
Geography			
Art			
Languages			

e) Biased question, suggesting that the journals are outdated.

Rephrase in unbiased format:

The Journals available in the journal section are

Current (latest issues are available) _____

Slightly out of date (latest issue is not available, the issues available are less than a year old) ___

Greatly outdated (issues available are more than a year old) _____

f) Question is too general and generates a wide range of responses, is difficult to analyse and/or does not give the data one is looking for.

Rephrase to make more specific.

How do you rate the following aspects of the library?

Complete the table. If you feel improvement is needed give a suggestion as to how this might be done.

	SATISFACTORY	SHOULD IMPROVE	BY
Sitting space			
Opening hours			
Catalogue			
Service at borrowing desk			
Shelving of books			
Access to books			

3. a) Given personally to (randomly selected) individuals from the target population and to be filled in in your presence.

Advantages

- High return rate.
- Questions not understood can be explained as the person filling the questionnaire can ask for clarification.

Disadvantages

- Time consuming. Hence likely only a relatively small sample can be used.
- Those participating might not feel free to honestly respond as the person is present and knows who filled in the questionnaire (lack of confidentiality).
- Person might quickly fill in spaces (tick randomly responses, skip open questions), as the person is waiting to receive the filled in questionnaire. This might lead to unreliable responses and incomplete questionnaires.

b) Given personally to (random selected) individuals from the target population and to be collected later. The individual can complete the questionnaire at his/her own time.

Advantages

- High return rate.
- Individuals have more time to complete questionnaire (do not feel under pressure) hence higher reliability of the responses.
- If collection not completed method a. can be used.

Disadvantages

- Time consuming. Hence likely only a relatively small sample can be used.
- Confidentiality not guaranteed.

c) Given personally to a (randomly selected) group from the target population and to be filled in in your presence.

Advantages

- High return rate.
- Economical use of time, high number of respondents can be reached at once.
- Confidentiality can be ensured.
- Questions on questionnaire can be explained if respondents ask for clarification.

Disadvantages

- Presence of researcher might reduce reliability (people tend at times to answer the way they think the person asking the question is expecting them to; people might feel under pressure to fill form in quickly).

d) Posting questionnaires to the randomly selected individuals forming the sample.

Advantages

- Economical use of time.
- Confidentiality can be ensured.

Disadvantages

- Can be expensive.

- Often a low return rate.
- Those responding are the ones interested in the topic, those not interested might not respond. This makes the sample biased.
- Questions might be misunderstood by respondents and there is no way to correct this.

4. Biased are

   (i) (ii) (iii) (iv) (vi) (vii) (viii) as an opinion is suggested
   (v) expresses an opinion not all people will agree with

   (i) How do you rate the bus services in the country?

   Very good ___ Good ___ Fair ___ Poor ___ Very poor ___

   (ii) Compare the present bus services with those 5 years ago.
   Do you think that they have
   improved ___ are still the same ___ deteriorated ___

   (iii) What is your opinion on smoking on buses? Tick your opinion.

   Should be allowed _____

   Should be restricted to special areas in the bus _____

   Should be banned completely _____

   (iv) What do you think about the frequency of the bus services?

   Should increase ___ Is sufficient ___ Could be cut back ___

   (v) Should secondary school students travel on special buses?
   YES ☐ NO ☐

   If YES, please explain _____

   (vi) What is your opinion of beauty contests? Tick your choice and give reasons.

   _____ Should be held, because _____

   _____ Should be banned, because _____

   (vii) What is your opinion on the use of face make-up? Tick your choice.

   _____ Should be used, because _____

   _____ Should not be used, because _____

   (viii) What is your opinion on capital punishment?

   _____ Capital punishment should be completely banned

   _____ Capital punishment should be used for the following types of crimes_____

Self mark exercise 5

1/2. A wide variety of responses is possible. Keep the points mentioned on good questionnaire design in mind. Avoid bias and leading questions. The Self mark exercise 4 has given you examples of properly structured questions.

3. None of them, as most people will feel (in their own opinion) that they do enough at home, that they are (reasonably) kind and that their family should not be described with the word 'poor'.

4. a) In general I sleep each night (tick the most appropriate)

   Less than 4 hours _____ 4 – 5 hours _____ 6 – 7 hours _____

   7- 8 hours _____ more than 8 hours _____

   b & c) add a box for USUALLY

Self mark exercise 6

1.

Colour	Tally	Frequency
Red	𝍸 𝍸	10
Blue	𝍸 ||	7
Green	𝍸 |	6
Silver	||||	4
White	𝍸 ||	7
Black	||||	4
Yellow	||	2

Mode: RED

2.

Age	Tally	Frequency
20 –29	𝍸 𝍸 ||||	14
30 – 39	𝍸 𝍸 |	11
40 – 49	𝍸 ||	7
50 – 59	𝍸 ||	7
60 - 69	|	1

Modal class 20 –29

3.

Height	Tally	Frequency
$140 \le h < 150$	||	2
$150 \le h < 155$	|	1
$155 \le h < 160$	𝍸 𝍸 |	11
$160 \le h < 165$	𝍸	5
$165 \le h < 170$	𝍸 ||	7
$170 \le h < 175$	𝍸 ||	7
$175 \le h < 180$	|||	3

b. 14 c. 33

Self mark exercise 7

1. a) 82 b) 568 c) 736

   d) No, 23.1% (1 dp) of the girls and 22.6% (1 dp) of the boys wear glasses.

   To the nearest percent both 23%. There seems to be no significant difference at all.

2. No. For girls and for boys 1 out of 7 is left handed.

Self mark exercise 8

Collect data by experiment for 1, 2, 5, 6 and 8

Use a survey to collect data in cases 3, 4, 7, 9 and 10