**LEARNER GUIDE**

Numeracy Level 2

**Unit standard 9009 Level 2 Credits 3**

# TABLE OF CONTENTS

[TABLE OF CONTENTS i](#_Toc80360422)

[PERSONAL INFORMATION 1](#_Toc80360423)

[INTRODUCTION 2](#_Toc80360424)

[Structure 3](#_Toc80360425)

[Programme methodology 3](#_Toc80360426)

[What Learning Material you should have 4](#_Toc80360427)

[Different types of activities you can expect 5](#_Toc80360428)

[Assessments 6](#_Toc80360429)

[Learner Support 7](#_Toc80360430)

[Learner Administration 7](#_Toc80360431)

[Learner Expectations 8](#_Toc80360432)

[UNIT STANDARD 9009 9](#_Toc80360433)

[Organise and Represent Data 12](#_Toc80360434)

[Use Statistics In Work Or Every Day Life 12](#_Toc80360435)

[Graphical Representation Of Data 13](#_Toc80360436)

[Bar graphs 13](#_Toc80360437)

[Line Graphs or Curves 14](#_Toc80360438)

[Pie Charts 14](#_Toc80360439)

[Stemplots 16](#_Toc80360440)

[Displaying Data From a Table 19](#_Toc80360441)

[Activity 1 (SO1, AC1-7) 25](#_Toc80360442)

[Frequency distribution and range 25](#_Toc80360443)

[Activity 2 (SO1, AC4-7) 27](#_Toc80360444)

[Summarising Data 27](#_Toc80360445)

[Mean 27](#_Toc80360446)

[Median 28](#_Toc80360447)

[Mode 28](#_Toc80360448)

[Using centres and averages 29](#_Toc80360449)

[Quartiles 30](#_Toc80360450)

[Example 1: Q1, 1st quartile 31](#_Toc80360451)

[Example 2: Median, 2nd quartile 32](#_Toc80360452)

[Example 3: 3rd quartile 32](#_Toc80360453)

[Activity 2 (SO1, AC4-7) 33](#_Toc80360454)

[Modelled Data 34](#_Toc80360455)

[Activity 2 (SO1, AC4-7) 34](#_Toc80360456)

[Annexure A 35](#_Toc80360457)

**PERSONAL INFORMATION**

|  |  |
| --- | --- |
| ***NAME*** |  |
| ***CONTACT ADDRESS*** |  |
|  |
| ***Code*** |  |
| ***Telephone (H)*** |  |
| ***Telephone (W)*** |  |
| ***Cellular*** |  |
| ***Learner Number*** |  |
| ***Identity Number*** |  |
| ***EMPLOYER*** |  |
| ***EMPLOYER CONTACT ADDRESS*** |  |
|  |
| ***Code*** |  |
| ***Supervisor Name*** |  |
| ***Supervisor Contact Address*** |  |
|  |
| ***Code*** |  |
| ***Telephone (H)*** |  |
| ***Telephone (W)*** |  |
| ***Cellular*** |  |

**INTRODUCTION**

***Welcome to the learning programme***

Follow along in the guide as the training practitioner takes you through the material. Make notes and sketches that will help you to understand and remember what you have learnt. Take notes and share information with your colleagues. Important and relevant information and skills are transferred by sharing!



This learning programme is divided into sections. Each section is preceded by a description of the required outcomes and assessment criteria as contained in the unit standards specified by the South African Qualifications Authority. These descriptions will define what you have to know and be able to do in order to be awarded the credits attached to this learning programme. These credits are regarded as building blocks towards achieving a National Qualification upon successful assessment and can never be taken away from you!

## Structure

### Programme methodology



The programme methodology includes facilitator presentations, readings, individual activities, group discussions and skill application exercises.

**Know what you want to get out of the programme from the beginning and start applying your new skills immediately. Participate as much as possible so that the learning will be interactive and stimulating.**

The following principles were applied in designing the course:

* Because the course is designed to maximise interactive learning, you are encouraged and required to participate fully during the group exercises
* As a learner you will be presented with numerous problems and will be required to fully apply your mind to finding solutions to problems before being presented with the course presenter’s solutions to the problems
* Through participation and interaction the learners can learn as much from each other as they do from the course presenter
* Although learners attending the course may have varied degrees of experience in the subject matter, the course is designed to ensure that all delegates complete the course with the same level of understanding
* Because reflection forms an important component of adult learning, some learning resources will be followed by a self-assessment which is designed so that the learner will reflect on the material just completed.

This approach to course construction will ensure that learners first apply their minds to finding solutions to problems before the answers are provided, which will then maximise the learning process which is further strengthened by reflecting on the material covered by means of the self-assessments.

***Different role players in delivery process***

* Learner
* Facilitator
* Assessor
* Moderator

### What Learning Material you should have

This learning material has also been designed to provide the learner with a comprehensive reference guide. It is important that you take responsibility for your own learning process; this includes taking care of your learner material. You should at all times have the following material with you:

|  |  |
| --- | --- |
| ***Learner Guide*** | ***This learner guide is your valuable possession:***This is your textbook and reference material, which provides you with all the information you will require to meet the exit level outcomes. During contact sessions, your facilitator will use this guide and will facilitate the learning process. During contact sessions a variety of activities will assist you to gain knowledge and skills. Follow along in the guide as the training practitioner takes you through the material. Make notes and sketches that will help you to understand and remember what you have learnt. Take and share information with your colleagues. Important and relevant information and skills are transferred by sharing!This learning programme is divided into sections. Each section is preceded by a description of the required outcomes and assessment criteria as contained in the unit standards specified by the South African Qualifications Authority. These descriptions will define what you have to know and be able to do in order to be awarded the credits attached to this learning programme. These credits are regarded as building blocks towards achieving a National Qualification upon successful assessment and can never be taken away from you! |
| ***Formative Assessment Workbook*** | The Formative Assessment Workbook supports the Learner Guide and assists you in applying what you have learnt. The formative assessment workbook contains classroom activities that you have to complete in the classroom, during contact sessions either in groups or individually.You are required to complete all activities in the Formative Assessment Workbook. The facilitator will assist, lead and coach you through the process. These activities ensure that you understand the content of the material and that you get an opportunity to test your understanding.  |

### Different types of activities you can expect

To accommodate your learning preferences, a variety of different types of activities are included in the formative and summative assessments. They will assist you to achieve the outcomes (correct results) and should guide you through the learning process, making learning a positive and pleasant experience.



The table below provides you with more information related to the types of activities.

| ***Types of Activities*** | ***Description*** | ***Purpose*** |
| --- | --- | --- |
| ***Knowledge Activities*** | You are required to complete these activities on your own.  | These activities normally test your understanding and ability to apply the information. |
| ***Skills Application Activities*** | You need to complete these activities in the workplace  | These activities require you to apply the knowledge and skills gained in the workplace |
| ***Natural Occurring Evidence*** | You need to collect information and samples of documents from the workplace. | These activities ensure you get the opportunity to learn from experts in the industry.Collecting examples demonstrates how to implement knowledge and skills in a practical way |

### Assessments

The only way to establish whether a learner is competent and has accomplished the specific outcomes is through the assessment process. Assessment involves collecting and interpreting evidence about the learners’ ability to perform a task.

**To qualify and receive credits towards your qualification, a registered Assessor will conduct an evaluation and assessment of your portfolio of evidence and competency.**

**This programme has been aligned to registered unit standards. You will be assessed against the outcomes as stipulated in the unit standard by completing assessments and by compiling a portfolio of evidence that provides proof of your ability to apply the learning to your work situation.**



***How will Assessments commence?***

***Formative Assessments***

The assessment process is easy to follow. You will be guided by the Facilitator. Your responsibility is to complete all the activities in the Formative Assessment Workbook and submit it to your facilitator.

***Summative Assessments***

You will be required to complete a series of summative assessments. The Summative Assessment Guide will assist you in identifying the evidence required for final assessment purposes. You will be required to complete these activities on your own time, using real life projects in your workplace or business environment in preparing evidence for your Portfolio of Evidence. Your Facilitator will provide more details in this regard.

**To qualify and receive credits towards your qualification, a registered Assessor will conduct an evaluation and assessment of your portfolio of evidence and competency.**

### Learner Support

**The responsibility of learning rests with you, so be proactive and ask questions and seek assistance and help from your facilitator, if required.**



Please remember that this Skills Programme is based on outcomes based education principles which implies the following:

* You are responsible for your own learning – make sure you manage your study, research and workplace time effectively.
* Learning activities are learner driven – make sure you use the Learner Guide and Formative Assessment Workbook in the manner intended, and are familiar with the workplace requirements.
* The Facilitator is there to reasonably assist you during contact, practical and workplace time for this programme – make sure that you have his/her contact details.
* You are responsible for the safekeeping of your completed Formative Assessment Workbook and Workplace Guide
* If you need assistance please contact your facilitator who will gladly assist you.
* If you have any special needs please inform the facilitator

## Learner Administration



***Attendance Register***

You are required to sign the Attendance Register every day you attend training sessions facilitated by a facilitator.

***Programme Evaluation Form***

On completion you will be supplied with a “Learning programme Evaluation Form”. You are required to evaluate your experience in attending the programme.

Please complete the form at the end of the programme, as this will assist us in improving our service and programme material. Your assistance is highly appreciated.

### Learner Expectations

Please prepare the following information. You will then be asked to introduce yourself to the instructor as well as your fellow learners



|  |
| --- |
| Your name:  |
|  |
|  |
| The organisation you represent:  |
|  |
|  |
| Your position in organisation:  |
|  |
|  |
| What do you hope to achieve by attending this course / what are your course expectations? |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |

# UNIT STANDARD 9009

#### Unit Standard Title

Apply basic knowledge of statistics and probability to influence the use of data and procedures in order to investigate life related problems

#### NQF Level

2

#### Credits

3

#### Purpose

This Unit Standard is designed to provide credits towards the mathematical literacy requirement of the NQF at Level 2. The essential purposes of the mathematical literacy requirement are that, as the learner progress with confidence through the levels, the learner will grow in:

* A confident, insightful use of mathematics in the management of the needs of everyday living to become a self-managing person
* An understanding of mathematical applications that provides insight into the learner’s present and future occupational experiences and so develop into a contributing worker
* The ability to voice a critical sensitivity to the role of mathematics in a democratic society and so become a participating citizen

#### Learning Assumptions

The credit value is based on the assumption that people starting to learn towards this unit standard are competent in Mathematics and Communications at NQF level 1

#### Range

This unit standard includes the requirement to:

* Identify issues suited to resolution by basic statistical methods.
* Work with existing data.
* Generate statistics through the use calculators and other available technology.
* Represent data in the form of tables, charts and graphs.
* Use statistics and representations of data to
* Summarise real-life and or work related issues within the experience of the learner.
* Give opinions on statistics and representations of data.
* More detailed range statements are provided for specific outcomes and assessment criteria as needed

#### Specific Outcomes and Assessment Criteria

**Specific outcome 1:** Apply various techniques to organise and represent data in order to model situations for specific purposes

Range: Techniques include:

* Using a variety of methods to represent statistics including pie charts, bar graphs, stem and leaf plots;
* Reading tables (e. g., the meaning of row and column headings and the relationship between age by gender by province);
* Extracting a suitable set of data from tables and databases (e. g., census data, tables in newspapers, HIV data; weather data);
* Recording and organising data into tables;
* Calculating measures of centre and spread such as mean, median, mode, and range; the use of quartiles in classifying data items ("Measures of centre and spread" should be handled via examples, which are directly related to the life or work experiences of each learner. For example workers` wages and learners` test scores).

**Assessment criteria**

* Questions about sets of data that can be dealt with through statistical methods are identified correctly.
* Existing tables are understood correctly through a proper application of row and column headings.
* Raw data or statistics in the body of tables are used correctly.
* Effective methods to record and organise data are used to solve problems.
* Calculations of statistics are correct.
* Appropriate statistics are used to answer questions.
* Scales used in graphical representations and tables are consistent with the data, are correct, clear and appropriate to the situation and target audience.

**Specific outcome 2:** Give opinions on the implications of the modelled data for the required purpose

Range: Purposes include:

* Determining trends in societal issues such as crime and health;
* Identifying relevant characteristics of target groups such as age range, gender, socio-economic group, cultural belief, and performance;
* Considering the attitudes or opinions of people on current issues relevant to the life experience of the learners;
* Determining weather patterns for a given region.

**Assessment criteria**

* Verbal (written or oral) explanation of findings is based on the representation of the data.
* Trends, group profiles and attitudes are justified.
* Appropriate information is extracted from representations in order to answer questions.

#### Essential embedded knowledge

The following essential embedded knowledge will be assessed through assessment of the specific outcomes in terms of the stipulated assessment criteria. Candidates are unlikely to achieve all the specific outcomes, to the standards described in the assessment criteria, without knowledge of the listed embedded knowledge. This means that the possession or lack of the knowledge can be inferred directly from the quality of the candidate `s performance against the standards.

* Methods for selecting, organising data and calculating statistics
* The meaning of concepts such as centre and spread
* Techniques for representing and drawing conclusions from statistics

#### Critical cross field outcomes

* Identify and solve problems using critical and creative thinking: Give opinions, based on data and statistics, on a variety of problems and issues
* Collect, analyse, organise and critically evaluate information: Select organise, and give opinions on statistics to make sense of situations related to the life or work of the learner
* Communicate effectively: Use everyday language and mathematical language to represent data, statistics and probabilities and to communicate conclusions
* Use mathematics: Use mathematics to describe and represent situations and to solve life related problems.

# Organise and Represent Data

#### Outcome

Apply various techniques to organise and represent data in order to model situations for specific purposes

#### Outcome Range

* Techniques include: Using a variety of methods to represent statistics including pie charts, bar graphs, stem and leaf plots;
* Reading tables (e. g., the meaning of row and column headings and the relationship between age by gender by province);
* Extracting a suitable set of data from tables and databases (e. g., census data, tables in newspapers, HIV data; weather data);
* Recording and organising data into tables;
* Calculating measures of centre and spread such as mean, median, mode, and range; the use of quartiles in classifying data items ("Measures of centre and spread" should be handled via examples, which are directly related to the life or work experiences of each learner. For example workers` wages and learners` test scores).

#### Assessment criteria

* Questions about sets of data that can be dealt with through statistical methods are identified correctly
* Existing tables are understood correctly through a proper application of row and column headings
* Raw data or statistics in the body of tables are used correctly.
* Effective methods to record and organise data are used to solve problems
* Calculations of statistics are correct
* Appropriate statistics are used to answer questions
* Scales used in graphical representations and tables are consistent with the data, are correct, clear and appropriate to the situation and target audience

## Use Statistics In Work Or Every Day Life

Statistics is the collection and analysis of numerical data in large quantities. This means that you gather information about a subject and then you analyse the information or data, so that you can distinguish trends. It is a very useful and easy way to “see” the story the numbers are telling.

Every time before an election, one of the organisations, Markinor, who collect and analyse data, will tell us before the election which political party will win the election and by how big a margin they will win the election. This is an example of gathering information and then analysing the information in order to find out what the trends are.

In the workplace, you can gather information about how many passengers you collect every day with your bus, how much fuel your bus uses, how many employees are off sick during winter, how much stationery is used by the administration department, etc. Once you have the information, you can analyse it to find out what the trend is.

## Graphical Representation Of Data

Once you have analysed the information, you will want to present it in such a way that everyone understands the information.

Charts and graphs are good ways of presenting information. They are visually appealing and make it easy for users to see comparisons, patterns, and trends in data. For instance, rather than having to analyze several columns of worksheet numbers, you can see at a glance whether sales are falling or rising over quarterly periods, or how the actual sales compare to the projected sales.

### Bar graphs

Bar graphs compare measurements at intervals, the bars run horizontally. Column charts compare measurements at intervals and provide a view of data at a specific time. The bars run vertically

The example below shows a column chart indicating how many ice creams were sold from January to May.

If you use a bar chart, the bars will run horizontally and not vertically as with a column chart.

*Number Of Ice Creams Sold*



* In a bar graph, the heights of the bars are important.
* When you draw a bar graph, state clearly what you are representing on the two axes. This means that you have to label the axes. Also insert it on the graph above.
* Draw the axes at right angles to each other
* Choose a scale for the vertical axis and write in the units.
* Use a ruler to help you read off the height of a bar.

### Line Graphs or Curves

Show the changes in data or trends over a given period of time. They are used to emphasize rather than compare. We use a dot to show the height of each bar. If we join the dots, we get a line graph.

*Example: Number Of Ice Creams Sold*

0

10

20

30

40

50

60

70

80

90

100

Jan

Feb

March

April

May

### Pie Charts

Show the breakdown of a total. A pie chart is a good way to show how a fixed number is divided. The whole circle (360ْ) represents the total number (or 100%) and we express each part as a fraction or percentage of the whole. A pie chart is constructed by converting the share of each component into a percentage of 360 degrees.

In the example on the next page, a total of 2000 cars were sold in 1996 and the pie chart shows the breakdown of the 2000 cars: which manufacturer sold how many cars.



A pie chart is useful to display the total value of your monthly expenses per your household budget.

Below is an example of a household budget in a table and then displayed as a pie chart:

|  |
| --- |
| **Expenses** |
| Mortgage |  R 3,200.00  |
| Car Loan |  R 2,200.00  |
| Water and lights |  R 430.00  |
| Groceries |  R 900.00  |
| Transport |  R 200.00  |
| Total |  R 6,930.00  |



From this you can see that this couple spend most of their money on paying for their house and car.

### Stemplots

Bar graphs can be a bit complicated to make although they are very easy to understand. If I have a small amount of data there is a quicker way to create a simple graph that has many of the same characteristics as a histogram. Let me define ‘a small amount of data’ as about 100 items or less. I’ve happily used this technique with several hundred items because it also has other hidden properties. So what we agree upon as small is just a matter of taste and convenience.

The result is called a ‘stemplot’ and I am going to demonstrate how to create it. It’s must easier to show you how to create a stemplot than it is to tell you how to do it.

You are arranging a birthday party for your grandfather and the list of guests amounts to 50. You have written down everyone’s age, to help you with the seating arrangements.

There are 50 numbers in total:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 53 | 97 | 66 | 99 | 30 |
| 81 | 19 | 9 | 31 | 67 |
| 61 | 4 | 5 | 73 | 54 |
| 42 | 27 | 49 | 29 | 30 |
| 28 | 13 | 60 | 13 | 34 |
| 28 | 59 | 87 | 29 | 62 |
| 38 | 40 | 40 | 78 | 98 |
| 69 | 39 | 62 | 56 | 90 |
| 9 | 34 | 12 | 69 | 93 |
| 38 | 58 | 83 | 28 | 5.  |

These numbers consist of two digits so they may range from 00 to 99.

There are three steps in making a stemplot:

1. Write down the stems. The stem consists of the first digit of each number remembering that ‘4’ is actually ‘04’: The stems are written vertically down the page with a vertical line to their right. In this case the stems range from ‘0’ to ‘9’.
2. Create the leaves. Write down the second digit to the right of the stem that contains its first digit. For example, ‘29’ uses ‘2’ as its stem with the ‘9’ written to the right of the vertical bar.
3. The last step is to sort the values of the leaves for each stem.

The figure below shows the 3 steps.



Figure 1 The three steps needed to create a stemplot

I’ll now take you through the steps just to make sure you understand them. The stems are easy to understand so we write down and put a vertical line to their right (Step 1).

The first number is ‘53’ so we write the ‘3’ to the right of the stem ‘5’. The next number is ‘97’ so we write the ‘7’ to the right of the stem ‘9’. The third number is ‘66’ so we write the ‘6’ to the right of the stem ‘6’. Continue in this way until all the numbers are finished and you should end up with the leaves (Step 2). You must write the right-hand digit next to the ones that are already associated with the stem (left-hand digit). If you used the normal tally method of putting strokes or lines down instead of the second digit, you end up with a stem and leaf plot.

Finally we sort the number in each stem with the lowest number at the left.

When we are finished, we have something like a histogram but on its side. Turn the manual sideways or turn you head sideways in order to see its shape.

We have something a histogram can’t give us: the values of the items! We also have another benefit that we often find very useful: the numbers are now sorted! Check the list: 4, 5, 5, 9, 9, 12, 13, 13, 19, and so forth until we get to 99.

Now you know how many guests aged between 0 and 9 years, 10 to 19 years, etc. will attend the party. You also know their individual ages. Now you can arrange seating for the youngest children near their mothers, you can place the teenagers away from the very old people, and so on.

#### Create a stemplot with decimals

You will need to make some adjustments for numbers with more than two digits or numbers with decimals. Let us say that the list of numbers on the next page represents the average maximum winter temperatures taken over a period of time at 10:00.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 13.6 | 18.9 | 11.0 | 8.1 | 14.8 |
| 14.6 | 17.4 | 11.0 | 17.4 | 9.3 |
| 19.1 | 18.1 | 9.2 | 22.4 | 15.5 |
| 12.9 | 10.8 | 6.2 | 10.3 | 13.7 |
| 13.8 | 16.6 | 19.3 | 7.9 | 15.8 |
| 17.3 | 3.6 | 16.9 | 20.6 | 6.2 |
| 11.6 | 11.0 | 16.8 | 22.1 | 12.4 |
| 6.7 | 14.9 | 14.7 | 18.9 | 20.0 |
| 15.2 | 10.5 | 14.3 | 16.5 | 6.1 |
| 8.3 | 12.5 | 15.8 | 14.3 | 15.8 |

**Create a stemplot**

Find the minimum and maximum numbers by scanning the rows and columns. The minimum is 3.6 and the maximum is 22.4. Use the whole numbers for the stem and the fraction for the leaf. If you do so you should end up with the stemplot in Figure 2.



Figure 2 Stemplot using decimals

I want to call your attention to several interesting items in this stemplot. (Interesting to one who wants to understand data that is.) The value of 3.6 is pretty far away from the other values. Values that are far away from their nearest neighbours may be what is called an outlier in statistics-talk. They lie outside the data. They may be really there or they may be there because of some error in the process of collecting or recording the information. I would check this value just to make sure it’s really 3.6 and not 6.3! You might also note that values 22.1 and 22.4 are also a bit far from their neighbours as well and may be outliers.

The values in stem ‘6’ also catch my attention. There appear to be too many values in this stem to go along with the rest of the distribution. These values would also need to be verified.

The overall shape of the distribution from 7.9 to 20.6 is fairly symmetrical even with the gaps that appear the values between 12.0 and 13.9 (the two stems of 12 and 13).

Now that we took a detour in looking at the previous distribution, let’s get back to the stemplot. You may also split the stems if it helps you. If you do then each stem appears twice. The leaves 0-4 are put with the upper stem while the leaves 5-9 are put with the lower stem. In the previous example, if I split the stems 14 and 15 I would obtain the grouping shown in Figure 3:



Figure 3 Stemplot with split stems

## Displaying Data From a Table

The following table was taken from the 2001 census and published by Stats SA.

Columns

|  |  |  |  |
| --- | --- | --- | --- |
| **Language** | Male | Female | Total |
| **Afrikaans** | 2,900,214 | 3,083,212 | 5,983,426 |
| **English** | 1,772,483 | 1,900,720 | 3,673,203 |
| **IsiNdebele** | 342,366 | 369,455 | 711,821 |
| **IsiXhosa** | 3,726,376 | 4,180,777 | 7,907,153 |
| **IsiZulu** | 5,045,450 | 5,631,855 | 10,677,305 |
| **Sepedi**Rows | 1,987,170 | 2,221,810 | 4,208,980 |
| **Sesotho** | 1,704,071 | 1,851,115 | 3,555,186 |
| **Setswana** | 1,774,785 | 1,902,231 | 3,677,016 |
| **SiSwati** | 571,429 | 623,002 | 1,194,431 |
| **Tshivenda** | 482,134 | 539,623 | 1,021,757 |
| **Xitsonga** | 1,001,446 | 990,761 | 1,992,207 |
| **Other** | 126,117 | 91,175 | 217,292 |
| **Total** | **21,434,041** | **23,385,736** | **44,819,777** |

This table lists the languages along with the number of people, grouped by gender, who use this language as their home language. Look at the table. It contains a few very important details.

Firstly, a table is made up of rows and columns. The rows extend horizontally across the page and the columns extend vertically down the page. Secondly, the table has a caption that describes the contents of the table. In this case the caption is below the table while in other cases the caption may be above the table. This table has a total column as its last row that may or may not appear in other tables. Thirdly, each column is labelled (Language, Male, Female and Total) so that you know what is in each column of data.

A table is pretty simple. It lists data in rows and columns and you can find the information you are looking for by going to either the row or columns of interest and looking either across or down. The intersection of the row and column (where they cross) is the data you require.

Once data has been collected it will be displayed in a table. From the table, you can make certain deductions. However, the best way to display data is to convert the table into a graph.

In this example, if I want to find out how many people reported that they speak a home language that is not one of the official nine languages, I would look in the row called ‘Other’ and read the number to its right in the ‘Total’ column: 217,292.

The purpose of a graph is to provide a visual summary of data. Graphs are the most effective way to communicate data and a good graph shows facts that would be very difficult or impossible to see from a table.

The visual impact of a graph is much stronger than looking at rows and rows of data in a table. I can’t get excited over a table of numbers but a good graph can tell me plenty. There is one problem with a graph and you must be aware of it. Graphs are so easy to use and so powerful that some people look at them and forget to think. A graph might look pretty, but it is very easy to deceive the person looking at it. But I’ll show you what to do to create a good graph and at the same time, show you what to look for when someone is trying to ‘sell’ you bad data.

Graphs, like tables, should be clearly labelled to show the variables that are being presented and the units being used. There are three things to remember when putting data in a graph:

* Make your data stand out
* Avoid clutter on the graph
* Use visual perception to get the facts to others.

The table on the next page shows the actual values of the age of all South Africans grouped into five years age groups. The table also displays the percentage of each age group. Note that the column totals do not sum to 100%. I only used three decimal figures to calculate the results in order to demonstrate the potential problems with round-off errors.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Age** | Male | % | Female | % | Total | % |
| **0-4** | 2,223,731 | 10.3% | 2,226,085 | 9.5% | 4,449,816 | 9.9% |
| **5-9** | 2,425,804 | 11.3% | 2,427,751 | 10.3% | 4,853,555 | 10.8% |
| **10-14** | 2,518,956 | 11.7% | 2,542,961 | 10.8% | 5,061,917 | 11.2% |
| **15-19** | 2,453,079 | 11.4% | 2,528,642 | 10.8% | 4,981,721 | 11.1% |
| **20-24** | 2,099,293 | 9.7% | 2,195,230 | 9.3% | 4,294,523 | 9.5% |
| **25-29** | 1,899,124 | 8.8% | 2,035,814 | 8.7% | 3,934,938 | 8.7% |
| **30-34** | 1,594,488 | 7.4% | 1,746,412 | 7.4% | 3,340,900 | 7.4% |
| **35-39** | 1,441,507 | 6.7% | 1,630,264 | 6.9% | 3,071,771 | 6.8% |
| **40-44** | 1,233,632 | 5.7% | 1,385,832 | 5.9% | 2,619,464 | 5.8% |
| **45-49** | 967,604 | 4.5% | 1,119,776 | 4.7% | 2,087,380 | 4.6% |
| **50-54** | 769,499 | 3.5% | 868,521 | 3.7% | 1,638,020 | 3.6% |
| **55-59** | 552,323 | 2.5% | 652,943 | 2.7% | 1,205,266 | 2.6% |
| **60-64** | 444,510 | 2.0% | 620,784 | 2.6% | 1,065,294 | 2.3% |
| **65-69** | 304,763 | 1.4% | 483,164 | 2.0% | 787,927 | 1.7% |
| **70-74** | 232,547 | 1.0% | 398,922 | 1.7% | 631,469 | 1.4% |
| **75-79** | 136,436 | 0.6% | 231,101 | 0.9% | 367,537 | 0.8% |
| **80-84** | 90,835 | 0.4% | 180,111 | 0.7% | 270,946 | 0.6% |
| **85+** | 45,907 | 0.2% | 111,425 | 0.4% | 157,332 | 0.3% |
| **Total** | **21,434,038** | **99.1%** | **23,385,738** | **99.0%** | **44,819,777** | **99.1%** |

The bar chart on the next page shows the details of the above table visually. It makes more sense to look at it this way, does it not? Something else you can note about this graph is that is consists of two bar charts displayed back to back.



The table below shows the number of units that were sold per item. To visually represent this data, a column chart or bar graph would be best:

|  |
| --- |
| **Sales** |
|  | Jan | Feb | Mar |
| **Item 1** | 1000 | 700 | 900 |
| **Item 2** | 400 | 600 | 700 |
| **Item 3** | 100 | 500 | 200 |
| **Item 4** | 5000 | 7000 | 6000 |
| **Totals** | 6500 | 8800 | 7800 |



This table shows school subjects:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1st quarter | 2nd quarter | 3rd quarter | 4th quarter |
| subject |  |  |  |  |
| English | 60 | 62 | 66 | 71 |
| Geography | 73 | 69 | 61 | 76 |
| History | 51 | 49 | 55 | 53 |
| Mathematics | 43 | 41 | 39 | 45 |
| Science | 46 | 43 | 47 | 53 |
| Second language | 58 | 61 | 53 | 57 |

And the column chart looks like this:



## Activity 1 (SO1, AC1-7)

## Frequency distribution and range

When you are drawing a chart or graph, you will need the frequency distribution and range of the information in order for the graph or chart to make sense.

* **Frequency distribution**: where you arrange (distribute) data in some kind of order. A frequency distribution tells you how often certain numbers or values occur.
* **Population**: The objects we are busy investigating (in the example in the activity, the learners in this class)
* **Range**: The difference between the lowest and highest items in a set of data is called the range of the data set.

Frequency or relative frequency distributions are most commonly displayed in histograms. A histogram looks very similar to a bar chart.

The histogram and the stemplot display the overall shape of a distribution of values as well as deviations from the centre of the distribution. We can see distributions that are symmetrical (Figure 1), skewed to the right (Figure 2) and skewed to the left (Figure 3). We have also seen that data appear to have a centre and also a spread of values around this centre. Some distributions have a single peak and others have more than one peak.

Figure 1 shows test results.



Figure 2 shows the histogram of the length of words occurring in War and Peace by Leo Tolstoy, 1869. This book consists of more than 550,000 words.



Figure 3 is a histogram of a word list. I am using the term ‘word list’ instead of dictionary because this list does not contain definitions. It’s an open source word list that is used for checking the spelling of words. **Error! Reference source not found.** shows the distribution of the number of letters in each word.

This histogram, in ‘statistics-talk’, is skewed to the left because it has a long tail that appears on the left. It’s completely different, perhaps, from what you would have expected. There are no one-letter words and there are no twelve-letter words.



#### Range

The range is defined as the difference between the largest and smallest values in the data set.

range = largest - smallest

The range is one measure of the spread of a set of data. If the range is very large we may expect the values in the data set to be spread widely.

## Activity 2 (SO1, AC4-7)

## Summarising Data

There are often situations in which it is useful to summarise a whole set of data by describing it with a single number. The sum of a set of numbers can be used to summarise data (e.g. the total mass of the pack of rugby forwards is an indication of the possible power of a rugby team).

The total is not always a useful summarizing number, however. Suppose we wish to compare the heights of 10 year old and 9 year old children in a group. The totals cannot be used for the data given because there are different numbers of children in the group.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Height of 10 year olds(m) | 1.76 | 1.77 | 1.8 | 1.66 | 1.6 | 1.79 | 1.8 |  |  |
| Height of 9 year olds(m) | 1.69 | 1.7 | 1.5 | 1.42 | 1.42 | 1.75 | 1.67 | 1.62 | 1.6 |

Although the total for the 9 year olds (14.37 m) is higher than the total for the 10 year olds (11.12 m), when we consider individual heights it seems that 10 year olds are typically taller than 9 year olds.

What we actually need is a single number which is typical or representative of the heights of individual 9 year and 10 year olds. One such a number is the arithmetic mean or average.

### Mean

**Arithmetic mean** = (sum of cases)/total number of cases

1. Mean height of 10 year olds = 12.18/7= 1.74
2. Mean height of 9 year olds = 14.37/9 = 1.6

The "mean" is the "average" you're used to, where you add up all the numbers and then divide by the number of numbers.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1st quarter | 2nd quarter | 3rd quarter | 4th quarter |
| subject |  |  |  |  |
| English | 60 | 62 | 66 | 71 |
| Geography | 73 | 69 | 61 | 76 |
| History | 51 | 49 | 55 | 53 |
| Mathematics | 43 | 41 | 39 | 45 |
| Science | 46 | 43 | 47 | 53 |
| Second language | 58 | 61 | 53 | 57 |

If we look at the table above, to calculate the average mark per year for English:

Add the marks for each quarter: 259

Divide by the number of terms: 4

**Average = 64.75**

### Median

**Median** is the middle value in a spread of values arranged in order from the lowest to the highest.

The median of an even number of items is the mean of the two middle items when the items are arranged according to size.

The median of an odd number of items is the middle item when the items are arranged according to size.

When do we use it? Suppose, for example, that at a party there are ten people aged 14, 15, 16, 14, 15, 16, 16, 15, 60, and 65 respectively. The mean of these ages is 24.6 which are not at all typical of the people at the party. A better statistic would be a median.:

14 14 15 15 **16** 16 16 60 65 **= the mean is 16 – there are 9 ages which is an odd number, so the median is the age right in the middle, which is 16.**

If we leave out one of the ages, say 65, there is an even number of items, namely 8.

14 14 15 **15 16** 16 16 60: **now we calculate the mean of the two middle ages 15 and 16, and our median becomes 15.50**

### Mode

When numbers occur frequently in a set of data, the number occurring most frequently is the mode. This is used if you have the same number occurring so frequently in a set of data that it can be regarded as the typical item. Suppose, for example that members of a group are asked to contribute to a gift for another person, and the contributions are

R1, R1, R2, R100, R100, R0.50, R2, R1, R1, R100, R100, R100.

This set of data can be described well by saying that the majority of people each contributed R100.

About the only hard part of finding the mean, median, and mode is keeping straight which "average" is which. Just remember the following:

**mean: regular meaning**

**median: middle**

**mode: most often**

### Using centres and averages

For each set of data (information) that you have collected, you will have to decide which of the three statistics (mean, mode, and median) will give you the best description of the data.

The **mode** is useful when dealing with nominal data (grouped data) like eye colour or ordinal data (ordered data) like shoe sizes

We now know how to calculate the mean, mode and average but we must still learn how and when to use them. Although the mode is very important in many real life situations it is not used that often in statistical calculations. In many situations the mode is not useful at all because there is no mode. The median is used more frequently than the mode because it is able to describe the data set with more flexibility. The median is also easily understood. The mean, or average, is the most commonly used measure of the centre of a set of data because it is backed by statistical theory. Let’s now see these different ways of measuring the centre of data in action.

Assume that I am the owner of a sporting body and I employ twelve sportsmen (or sportswomen). The annual salaries of each sportsman (sportswoman) are as follows:

* six receive R200,000
* four receive R400,000
* one receives R800,000
* the superstar receives R2,400,000

What is their average salary?

If I used the mode to calculate the average salary it would be R200,000 because this value occurs with the highest frequency (6 times).

If I use the median I must first order the salaries as shown in the table below and then determine the middle value:

|  |  |
| --- | --- |
| Order | Salary |
| 1 | R200,000  |
| 2 | R200,000  |
| 3 | R200,000  |
| 4 | R200,000  |
| 5 | R200,000  |
| 6 | R200,000  |
| 7 | R400,000  |
| 8 | R400,000  |
| 9 | R400,000  |
| 10 | R400,000  |
| 11 | R800,000  |
| 12 | R2,400,000. |

There are an even number of entries (12) so I must obtain the average of the middle two values (R200,000 and R400,000) and indicate that the median is R300,000.

If I use the average (mean) I must sum all the salaries and divide by the number of players which is R500,000 (R60,000,000 divided into 12 players).

Which is the best number to use for the average salary?

The mode indicates that the highest frequency and most players receive R200,000. This number just doesn’t seem right to me even though half the players receive this salary and the other half receive more.

The average produces R500,000. No player receives this amount but then it is an average. However, 10 players receive less that R500,000 and only two receive more. Again, this just doesn’t look right to me.

The median also produces a value that no player receives but at least half the players receive less than this amount and half receive more than this amount. This is the middle of the road average so I would agree with it. As a matter of fact, if the superstar was paid R24,000,000, this measurement of the centre (R300,000) would not change and I would still have half of the players being less and half being paid more. However, if I used the average (mean) I would find that the average salary would be R2,300,000!

The median is a good choice to use for the central value when the distributions are skewed to the right or to the left. Salaries are almost always skewed to the right (very few people obtain very high salaries while most of the workers receive salaries that are on the lower end.) Note that the median value always has half the values on one side and half the values on the other side. There is always a middle value that exists or is the average of the two centre values. The values don’t matter, only their positions when sorted matter.

When the distribution of values is more or less symmetrical and there are no outliers, then the average or mean is the best value to use for the centre of the data set.

Values that skew the distribution of values of the data affect the mean or average, sometimes dramatically. If I did pay my superstar R24,000,000 and advertised that my average salary bill is R2,300,000, the other 11 players would be at my door asking why their salary is so low!

## Quartiles

‘What! More things to learn?’ No, you already know them but you don’t know you know them!

Three of the numbers are already familiar to you: minimum value, maximum value and the median. The minimum and maximum values are used to calculate the range, so you know them. The median is the middle-most number in the data set and divides the data set into two separate but equal sizes of data: 50% of the values are above the median and 50% of the values are below the median.

Now take each half and divide it in half again (find the median of each half). What do we have? With the lower half of the data, we have the values that fall below (and above) 25% of the data and with the upper half of the data, we have the values that fall above (and below) 75% of the data.

When we calculate the median of the lower and upper half of the data set we calculate the ‘lower quartile’ and ‘upper quartile’ of the data. We also have the last two numbers of our five-number summary.

* The ‘lower quartile’ (lower quarter point) is also called the ‘25th percentile’ or the ‘1st quartile’ or just Q1: 25% of the data fall below this value.
* The ‘upper quartile’ (upper quarter point) is also called the ‘75th percentile’ or the ‘3rd quartile’ or just Q3: 75% of the data fall below this value.
* The median is also called the ‘50th percentile’, the ‘2nd quartile’ or just Q2: 50% of the data fall below this value.
* The minimum and maximum values are often called Q0 and Q4, respectively

The five-number summary consists of Q0, Q1, Q2, Q3 and Q4 (minimum, 1st quartile, median, 3rd quartile and maximum values). I told you that you knew five-number summary, didn’t I?

Quartiles divide your data set into 4 equal sized groups of data.

|  |  |
| --- | --- |
| Number | Value |
| 1 | 15 |
| 2 | 17 |
| 3 | 17 |
| 4 | 18 |
| 5 | 18 |
| 6 | 18 |
| 7 | 19 |
| 8 | 19 |
| 9 | 20 |
| 10 | 20 |
| 11 | 20 |
| 12 | 21 |
| 13 | 22 |
| 14 | 23 |
| 15 | 24 |
| 16 | 29 |
| 17 | 32 |
| 18 | 33 |
| 19 | 33 |
| 20 | 41 |

### Example 1: Q1, 1st quartile

In order to calculate the first quartile (25th percentile) we just use the equation and insert the correct number to obtain the correct element to choose.



The value returned is not a whole number so split it into two pieces 5 and 0.25.

The 5th value in the table is 18. The value 0.25 means that Q1 is 0.25 times the distance between the 5th and the 6th entry. So we have to calculate this value:

5.25th item = 5th item + 0.25×(6th item – 5th item) or

Q1 = 18 + 0.25×(18 – 18) = 18

Therefore, the 1st quartile or Q1 is 18. This also means that 25% of the values are 18 or lower.

### Example 2: Median, 2nd quartile

In order to calculate the median or the 2nd quartile we just use the same formula but put in different values.



The value returned is not a whole number so split it into two pieces 10 and 0.5.

The 10th item in the table is 20. The value 0.5 means that the median is 0.5 times the distance between the 10th and 11th entry. So we have to calculate this value: 10.5th item = 10th item + 0.5×(11th item – 10th item).

Therefore the median = Q2 = 20 + 0.5×(20 – 20) = 20.

Therefore, the median is 20. This also means that 50% of the values are lower than 20.

### Example 3: 3rd quartile

In order to calculate the median or the 3rd quartile we just use the same formula but put in different values.



The value returned is not a whole number so split it into two pieces 15 and 0.75.

The 15th item in the table is 24. The value 0.75 means that the median is 0.75 times the distance between the 15th and 16th entry. So we have to calculate this value: 15.75th item = 15th item + 0.75×(16th item – 15th item).

Therefore the median = Q3 = 24 + 0.75×(29 – 24) = 27.75.

Therefore, the 3rd quartile or Q3 is 27.75. This also means that 75% of the values are lower than 27.75.

## Activity 2 (SO1, AC4-7)

# Modelled Data

#### Outcome

Give opinions on the implications of the modelled data for the required purpose

#### Outcome Range

Purposes include:

* Determining trends in societal issues such as crime and health;
* Identifying relevant characteristics of target groups such as age range, gender, socio-economic group, cultural belief, and performance;
* Considering the attitudes or opinions of people on current issues relevant to the life experience of the learners;
* Determining weather patterns for a given region.

#### Assessment criteria

* Verbal (written or oral) explanation of findings is based on the representation of the data.
* Trends, group profiles and attitudes are justified
* Appropriate information is extracted from representations in order to answer questions

## Activity 2 (SO1, AC4-7)

## Annexure A



