

Monyetla Bursary Project

Gr 12

Maths Lit

Data Handling – Part 1

Developing questions:

- **Survey** – a statistical study that collects data in order to see trends or to form some type of conclusion. It is often done in the form of a questionnaire.
- **Data** – pieces of information that have been observed and/or recorded.
- **Bias** – to favour one or more responses unfairly through the wording of a question or the design of a survey.

Tips for questions:

- Questions must be short, simple and easy to understand.
- Answers must be one word or a choice between two or more possible answers. Tick boxes must be provided where there is a choice between answers.
- Do not ask for sensitive information.
- Only ask relevant questions.
- Do not ask vague questions.
- Do not ask negative questions such as “Don’t you like cats?”
- The questionnaire should not take more than 10 minutes to complete.

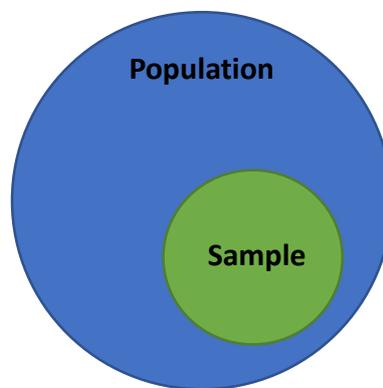
Collecting data:

- **Observation** – carefully watching and recording behaviour or some characteristic. Useful if you are studying animals or trying to collect information about something visual that does not need any more information.
- **Interviews** – a meeting in which someone is asked questions in a systematic way. Useful when you have to collect a lot of information and/or the survey participants have trouble understanding the questions.

- **Questionnaires** – a list of questions, usually printed or online, designed to collect specific information from the respondents. Useful when gathering sensitive information or if you need a large number of people to respond.

Population vs Sample:

- **Unbiased** – fair, unprejudiced and neutral. Choose participants randomly.
- **Representative sample:**
 1. The size of the group needs to be large enough to ensure for reliable data.
 2. The sample must include people from different genders, different age groups and different socio-economic groups



Classifying and organising data:

| Categorical data |
|--|
| <ul style="list-style-type: none"> • Divided into groups according to the properties of the data • 'Things' instead of numbers • e.g. gender groups, flavours of chips, makes of cars, colour of hair |

| Numerical data |
|--|
| <ul style="list-style-type: none"> • Consist of numbers – amounts and measurements • Discrete data – obtained from <u>counting</u> things and can only take on a finite amount of values (whole numbers). E.g. pets in a household, number of fish, number of people • Numerical data – obtained from <u>measuring</u> things and can take on an infinite number of values (any decimal number). E.g. temperature, distance, mass |

Single, multiple and stacked bar graphs:

- Shows the frequency of each data value, by means of bars
- Used for discrete categorical data
- Single bar graphs – one data value per category
- Compound bar graphs – multiple and stacked bar graphs
- Multiple bar graphs – two or more data values per category are compared and represented by bars next to each other
- Stacked bar graphs – two or more data values per category are compared and represented by bars being stacked on top of each other

Histograms:

- Shows the frequency of each data value by means of bars
- Used for continuous data
- Data is usually grouped in class intervals
- No spaces between the bars – indicates the continuous nature of the data
- Class intervals on horizontal axis where each bar represents one class or interval

Line and broken line graphs:

- A line graph shows the trend between plotted points of continuous data – points are joined to show continuous nature of data
- A broken line graph shows the trend between plotted points of discrete data – points are not joined to show the discrete nature of the data
- These graphs are also effective in showing the relationship between two variables and multiple sets of data; and how these data sets change in relation to each other

Scatter plots:

- A scatter plot is a graph whereby one variable is plotted against another variable; in order to show the relationship between them
- May form a pattern:
 - Increasing straight line 'pattern' – positive correlation
 - Decreasing straight line 'pattern' – negative correlation
 - Scattered randomly without any 'pattern' – no correlation

Which representation to use:

- In general, use:
 - **Pie charts** when you are trying to compare **parts of a whole**.
 - **Bar graphs** to compare the **frequency of discrete data**.
 - **Histograms** to compare the **frequency of continuous data**.
 - **Line and broken line graphs** to **track trends/changes** over time.
 - **Scatter plots** to show whether there is any **correlation between 2 variables**.
 - **Box-and-whisker plots** to show the **spread of data**.