Monyetla Bursary Project

Gr 12

Maths Lit

Data Handling – Part 2

Interpreting and analysing data:

- After representing data visually, it is important to interpret and analyse the data, by taking the following into account:
    1. Using percentages in a table or graph is useful for comparing relationships in size, but does not give any information regarding the actual sample or population size
    2. Using actual sample or population values gives an indication of the size, but not of the relationship between data categories
    3. The choice of scale of the axes and the point at which the axes cross will affect the impression created by the graph
    4. Graphs show trends in data more clearly than data values in a table
- The following questions should be asked when analysing and interpreting data:
    1. What was the size of the sample?
    2. Was the sample randomly chosen and representative?
    3. What methods were used to collect the data and did the collector/recorder remain neutral and impartial?
    4. Was the data collected fact of opinion?
    5. How was the data organised and/or grouped?
    6. Which measures of central tendency and spread were used?
- Be aware of the fact that data can be used and manipulated to favour an argument or circumstance.
- Interpretation and analysis of the data should happen at every stage of the statistical cycle.

Measures of central tendency:

A value that provides an indication of the 'middle' or 'centre' of the data.

- It gives a single value against which other values in the data set can be compared
- It is representative of the majority of values in the data set

1. **Mean:**
    - 'average'
    - Mean = $\frac{\text{sum of all values in data set}}{\text{total number of values in data set}}$
    - Only for numerical data

2**. Median:**

- All data values must first be arranged in ascending order
- Middle value of the ordered data set
- Only for numerical data

3. **Mode:**

- Data value(s) that occur(s) most frequently in a set
- If two values occur most frequently – bimodal
- For both numerical and categorical data

When to use the Mean, Median or Mode:

- **Outlier –** a value that is far away from most of the other values

e.g. 6 : 51 : 56 : 57 : 63 : 65

- **Mean –** best used to describe the average of a set of data that does not have an outlier
- **Median –** best used to describe the 'middle' value of a set of data that does have an outlier
- **Mode –** is usually used when the data is categorical or when asked to choose the most popular item

Measures of spread:

A value that provides an indication of how 'spread out' the data is

1. Range

- indicates the distance between the highest and the lowest values
- Range = highest value – lowest value
- If the range is small, the data is clustered together, and if the range is larger, the data is more spread apart

2. Interquartile range:

IQR = $Q_3 - Q_1$

- Quartiles divide a data set into four equal parts
- The middle quartile (Q_2) is the median of the data set
- The first, or lower, quartile (Q_1) indicates the first quarter of the data set
- The third, or upper, quartile (Q_3) indicates the upper quarter of the data set