Monyetla Bursary Project

Grade 11

Lesson 4: Data Handling

Measure of central tendency:

Refer to three types of averages that we can use to compare the other data values with.

1. Mean

The mean is calculated by adding all the values in a data set before you divide by the number of values in the data set.

$$\text{mean} = \frac{\text{sum of all data values in the set}}{\text{number of data values in the set}}$$

2. Median

The median is the value that lies in the middle of the data set when the data set is arranged in ascending order.

*   Uneven set:

Step 1: Arrange in ascending order:
1;   2 ;   2;   3;   3;   4;   4;   4;   5;   5;   6,   6
7;   7;   9

Step 2: Find the value in the middle:
1;   2;   2;   3;   3;   4;   4;   (4;)   5;   5;   6,   6
7;   7;   9

The median = 4 minutes.

- Even set:

Arrange in ascending order:

6,59m 7,02m ; 7,05m; 7,08m;   7,10m;   7, 11m   two values in the middle

Find the two middle values:

6,59 m; 7,02 m; 7,05 m; 7,08 m; 7,10 m; 7,11 m

Add the two middle values and divide the answer by 2:

$$\text{Median} = \frac{\text{Sum of two values in the middle}}{2}$$

$$= \frac{7,05 + 7,08}{2}$$

$$= \frac{14,13}{2}$$

$$\approx 7,07 \text{ m (rounded to 2 decimals)}$$

3. Mode

The mode is the value or item that appears most in a set of data.

The mode is normally intended for categorical data but it is possible to use it for numerical data sets. With small sets of numerical data, it is not really a 'central' value.

It is possible to have more than one mode. Make sure you list all modal values.

It is also possible to have no mode.

Measures of spread:

The range is the difference between the highest and the lowest data value. It shows how spread out the values are on the number line.

Range = highest – lowest

Which measure of central tendency to use:

The purpose of determining an average is that it should give a type of summary of the data set in one single figure. This average must convey the idea of what can be expected in the middle.

MODE:

The mode is the only 'average' that can be used for categorical data – when data cannot be ranked in any way, the mode is the only 'average' to be used.

For very large sets of numerical data it is a suitable numerical 'average'.

MEAN:

When numerical values are used, the mean is normally the better choice because all data values are included in the calculation.

MEDIAN:

When there are extreme values or outliers present in the data set, the median is usually the better choice.