Census@school

# Data Handling & Probability
## Grades 10, 11 and 12

Statistics South Africa

your leading partner in quality statistics

# 1

# Grade 10 Data Handling

In this chapter you will:
- Revise the language of data handling
- Determine measures of central tendency of lists of data, of data in frequency tables and data in grouped frequency tables
- Determine quartiles and the five number summary
- Determine percentiles
- Determine measures of dispersion (range and inter-quartile range)
- Illustrate the five number summary with a box and whisker diagram

**WHAT YOU LEARNED ABOUT DATA HANDLING IN GRADE 9**

*In Grade 9 you covered the following data handling concepts:*
- ***Collecting data :*** *including distinguishing between samples and populations*
- ***Organising and summarising data :*** *using tallies, tables and stem-and-leaf displays; determining measures of central tendency (mean, median, mode); determining measures of dispersion (range, extremes, outliers)*
- ***Representing data :*** *drawing and interpreting bar graphs, double bar graphs, histograms, pie charts, broken-line graphs, scatter plots.*
- ***Interpreting data :*** *critically reading and interpreting two sets of data represented in a variety of graphs.*
- ***Analysing data :*** *critically analysing data by answering questions related to data collection methods, summaries of data, sources of error and bias in the data*
- ***Reporting data*** *by drawing conclusions about the data; making predictions based on the data; making comparisons between two sets of data; identifying sources of error and bias in the data; choosing appropriate summary statistics (mean, median, mode, range) for the data and discussing the role of extremes and outliers in the data*

# THE LANGUAGE OF DATA HANDLING

✓ The word ***data*** is the plural of the word ***datum*** which means "a piece of information". So data are pieces of information.

## a) Organising Data

✓ In order to make sense of the data, we need to ***organise*** the data.

✓ Different sets of data can be sorted in different ways:

- You can write the data items in either ***alphabetical*** or ***numerical*** order.

  ***For example:***
  - The words elephant; lion; frog and crocodile can be ordered in ***alphabetical order*** as follows: crocodile; elephant; frog; lion
  - The numbers 32,1; 32,001; 32,0001 and 32,01 can be ordered in ***ascending numerical order*** as follows: 32,0001; 32,001; 32,01 and 32,1

- You can sort data items using a ***tally table***.

  A ***tally*** is a way of collecting information by making an appropriate mark for each item.
  A line is drawn for each item counted : ||||
  Every ***fifth*** line is drawn across the other four : 𝍸. This makes it easy to add up the number of items checked.

  ***For example***:
  The following ***tally table*** shows the favourite fruit of sixteen Grade 10 learners.

| **Favourite Fruit** | **Number of learners** |
|---|---|
| Apple | 𝍸 𝍷𝍷 |
| Banana | 𝍸 𝍷𝍷𝍷𝍷 |

✓ When collecting data, the number of times a particular item occurs is called its ***frequency***.

***For example***:
The following ***frequency tables*** show the same information about the favourite fruit of the sixteen Grade 10 learners.

| **Favourite Fruit** | **Frequency** |
|---|---|
| Apple | 7 |
| Banana | 9 |
| **TOTAL** | 16 |

| **Favourite Fruit** | Apple | Banana | **TOTAL** |
|---|---|---|---|
| **Frequency** | 7 | 9 | 16 |

✓ We can use a ***stem-and-leaf diagram*** to organise data.

With a stem-and-leaf diagram, we organise the data by using place value:

- The digits in the ***largest place*** are referred to as the ***stem.***
- The digits in the ***smallest place*** are referred to as the ***leaf*** (or leaves).
- The leaves are displayed to the right of the stem.

This means that for the number 45, the digit 4 is the stem and the 5 is the leaf.

**EXAMPLE 1**
Organise the following set of 25 data items using a stem-and-leaf-diagram:

6; 9; 12; 12; 14; 15; 16; 18; 18; 18; 19; 20; 20;
21; 21; 21; 22; 23; 28; 28; 29; 32; 33; 33; 37

**SOLUTION:**

| Stem | Leaves |
|---|---|
| 0 | 6; 9 |
| 1 | 2; 2; 4; 5; 6; 8; 8; 8; 9 |
| 2 | 0; 0; 1; 1; 1; 2; 3; 8; 8; 9 |
| 3 | 2; 3; 3; 7 |
| **KEY:** 1/4 = 14 | |

## b) Populations and Samples

✓ We can carry out a ***survey*** to find out information. We find out the information by asking questions.

✓ The word ***population*** is used in statistics *for the set of data being investigated.*

So, if we want to find out information about all the learners in a school, we could ask every single learner in the school. This group is called the ***population.***

✓ A ***sample*** is a subset of a population.
This means that a sample is much smaller than a population.

A ***subset*** is a set that is part of a larger set.

So, to find out information about all the learners in a school, we could ask selected learners in each grade instead of every learner in the school. These selected learners would be a ***sample*** and, if the sample is selected correctly, the results could be used to reach conclusions about the whole school.

# MEASURES OF CENTRAL TENDENCY OF LISTS OF DATA

- ✓ An ***average*** or ***measure of central tendency*** is a single number which is used to represent a collection of numerical data. The commonly used averages are the ***mean, median*** and ***mode***.

- ✓ When we calculate the mean, median and mode we are finding the value of a ***typical item*** in a data set.

## a) Mean

- ✓ The ***mean*** is the sum of all the values divided by the total number of values.

- ✓ The mean is the ***equal shares*** average. To find the mean you find the total of all the data items and share the total out equally.

- ✓ The mean is usually written as $\bar{x}$ (often read as "x bar").

  $\bar{x}$ is worked out using the following formula:

$$\bar{x} = \frac{sum\ of\ all\ the\ values}{total\ number\ of\ values}$$

$$= \frac{\Sigma x}{n}$$

- The symbol Σ (sigma) tells us to add all the values in the data set.
- The symbol $n$ is the total number of items in the data set.

**EXAMPLE 2**

Fourteen of the learners in a Grade 10 class were asked to work out how many kilometres they lived from school. The following list of data shows the distances in km:

4; 7; 1; 9; 4; 8; 11; 10; 19; 2; 5; 7; 19; 3

a) Calculate the mean distance these fourteen learners live from school.
b) What does the mean tell us about the distances travelled?
c) Use your scientific calculator to determine the mean distance travelled.

**SOLUTION:**

a) $\Sigma x = 4 + 7 + 1 + 9 + 4 + 8 + 11 + 10 + 19 + 2 + 5 + 7 + 19 + 3 = 109$ km

There are 14 terms in the data set, so $n = 14$

$$\text{Mean} = \bar{x} = \frac{\Sigma x}{n}$$

$$\bar{x} = \frac{109}{14}$$

$$\bar{x} = 7{,}7857...$$

$$\bar{x} \approx 7{,}79$$

The mean distance that the 14 learners live from their school is 7,79 km.

**<u>EXAMPLE 2</u> (continued)**

An ***outlier*** is a value far from most others in a set of data

b) The mean tells us that if all the distances are added together and shared out equally, each learner would travel 7,79 km.

Two of these learners live 19 km away from the school. These two ***outliers*** (of 19 km) affect the value of the mean and make it larger than it should be if only the distances that the other twelve learners live from the school were considered.

c) The key sequences for the Casio *fx-82ZA PLUS* and the Sharp EL-W535HT are as follows:

<u>CASIO</u>:
[MODE] [2 : STAT] [1: 1 – VAR]
4 [=] 7 [=] 1 [=] 9 [=] 4 [=] 8 [=] 11 [=] 10 [=] 19 [=] 2[=] 5 [=] 7 [=] 19 [=] 3 [=] [AC]
[SHIFT] [1 : STAT] [4 : VAR] [2 : $\bar{x}$] [=]

<u>SHARP</u>
[MODE] [1 : STAT] [0 : SD] [2ndF] [CA]
4 [DATA] 7 [DATA] 1 [DATA] 9 [DATA] 4 [DATA] 8 [DATA] 11 [DATA] 10 [DATA] 19 [DATA] 2 [DATA] 5 [DATA] 7 [DATA] 19 [DATA] 3 [DATA]
[ALPHA] [=] [$\bar{x}$]

Both calculators give the value $\bar{x}$ = 7,7857... ≈ 7,79 km

**EXAMPLE 3**

A representative sample of Secondary Schools in South Africa took part in the 2009 Census@School. The mean number of schools per province for 8 of the 9 provinces (Free State is excluded) was 78.

a) What is the total number of schools in the 8 provinces that took part in 2009 Census@School?
b) The number of schools in the Free State (54) is now added to the total in a).
   i) What is the total number of schools in the sample now?
   ii) What is the mean number of schools per province now?
c) What does this mean tell us about the number of schools in the sample?

**<u>SOLUTION:</u>**

a) Mean = $\frac{\text{total number of schools}}{\text{number of provinces}}$

$78 = \frac{\text{total number of schools}}{8}$

Total number of schools = 8 × 78 = 624

b) Number of provinces with the inclusion of the Free State = 8 + 1 = 9
   i) New total number of schools = 624 + 54 = 678
   ii) New mean = $\frac{\text{total number of schools}}{\text{number of provinces}} = \frac{678}{9}$ = 75,333... ≈ 75

c) The mean tells us that if the 678 schools were shared out equally amongst the 9 provinces, each province would get approximately 75 schools.

## b) Median

- ✓ The median is the middle value when all values are placed in ascending or descending order.
- ✓ There are as many values above the median as below it.
  - If there is an *odd number* of data items, the median is one of the data items.
  - If there is an *even number* of data items, the median is found by adding the two middle data items and dividing it by two.

**EXAMPLE 4**
Find the median of the following two sets of data:
a) 4; 7; 1; 9; 4; 9; 11; 10; 19; 2; 5; 8; 19
b) 4; 6; 1; 9; 4; 8; 11; 10; 19; 2; 5; 7; 19; 3

**SOLUTION:**

a) First arrange the data in ascending order: 1; 2; 4; 4; 5; 7; 8; 9; 9; 10; 11; 19; 19
There are 13 data items, and 13 is an odd number.
The middle item is the 7$^{th}$ one: 1; 2; 4; 4; 5; 7; **8**; 9; 9; 10; 11; 19; 19
***The median = 8***
Note that there are six data items to the left of 8 and six data items to the right of 8.

b) First arrange the data in ascending order: 1; 2; 3; 4; 4; 5; 6; 7; 8; 9; 10; 11;19; 19
There are 14 data items, and 14 is an even number.
The 7$^{th}$ and 8$^{th}$ terms are the two middle data items:
1; 2; 3; 4; 4; 5; **6; 7**; 8; 9; 10; 11;19; 19
The median is ***halfway*** between 6 and 7, so the ***median*** $= \frac{6+7}{2} = \frac{13}{2} = 6{,}5$
Note that 50% of the data items are less than 6,5 and 50% of the data items are more than 6,5.

## c) Mode

- ✓ The mode is the data item that occurs most frequently.
  - If there are two modes, then the data set is said to be ***bimodal***.
  - If there are more than two modes, then the data set is said to be ***multimodal***.
  - All the data items in a set may be different. In this case it has ***no mode***.
- ✓ The associated adjective is ***modal*** so we are sometimes asked to find the ***modal value.***

**EXAMPLE 5**
Find the mode of the following sets of data:
a) 3; 8; 9; 12; 17; 11; 9; 1; 10; 18
b) 1; 2; 3; 4; 4; 5; 7; 7; 8; 9; 10; 11; 19; 19
c) 1; 7; 8; 10; 51; 18; 2; 19; 11; 45

**SOLUTION:**

a) First arrange the data in ascending order: 1; 3; 8; 9; 9; 10; 11; 12; 17; 18
Look for the value that occurs most frequently: 1; 3; 8; **9; 9**; 10; 11; 12; 17; 18
***Mode = 9***

b) The data is already arranged in ascending order:
1; 2; 3; 4; 4; 5; 7; 7; 8; 9; 10; 11; 19; 19
Look for the value that occurs most often:
1; 2; 3; **4; 4**; 5; **7; 7**; 8; 9; 10; 11; **19; 19**
There are three modes, so the data set is multimodal.
***Modes = 4; 7 and 19***

c) First arrange the data in ascending order: 1; 2; 7; 8; 10; 11; 18; 19; 45; 51
Look for the value that occurs most frequently: 1; 2; 7; 8; 10; 11; 18; 19; 45; 51
None of the values are repeated.
***So there is no mode***

## EXERCISE 1.1

Round all decimal answers to 2 decimal places.

1) For each of the following sets of data find:
   i) The mean
   ii) The median
   iii) The mode

   a) 2; 5; 8; 4; 3; 4; 7; 6; 2; 4; 4
   b) R2,50; R3,00; R6,50; R1,25; R6,50; R2,50; R6,50
   c) 12 cm; 15 cm; 7 cm; 6 cm; 11 cm; 7 cm; 13 cm; 12 cm
   d) 120 kg; 112 kg; 118 kg; 111 kg; 113 kg; 114 kg; 119 kg; 125 kg; 109 kg; 130 kg

2) The mean height of a group of 10 learners is 166,8 cm.
   The tallest person in the group is 169,9 cm.
   Calculate the average height of the remaining 9 members in the group.

3) There are 4 children in a family. The two oldest children are twins. The mean of the 4 children's ages is 14,25 years, the median is 15,5 years and the mode is 16 years. Use this information to work out the ages of the 4 children.

4) Mr Molefe was reading the section in the 2009 Census@School that shows *Favourite Subject by Gender, Grade 8 to 12*. He was astonished at how few students in Secondary Schools in the sample chose Mathematics as their favourite subject.

   In an attempt to find out if the learners in his school shared the same feelings about Mathematics, he asked a representative sample of 100 boys and 100 girls in each grade what their favourite subject was. The following table shows the results of Mr Molefe's survey:

| **Percentage of the learners in the sample whose favourite subject is Mathematics** | | |
|---|---|---|
| **Grades** | **Boys** | **Girls** |
| 8 | 12,2% | 10,3% |
| 9 | 9,8% | 11,2% |
| 10 | 11,1% | 7,8% |
| 11 | 8,3% | 6,9% |
| 12 | | 12,5% |

   a) What is the mean percentage of girls in the table?
   b) Mr Molefe is still waiting for the results for the Grade 12 boys. He would like the mean for the boys in the whole sample to be 10%. What must the minimum percentage be from the Grade 12 boys in order for Mr Molefe to get the results that he wants?

5) A representative sample of 1 000 learners in each grade from Grades 3 to 12 was surveyed to determine the number of boys in each grade. The mean percentage of the boys in each of the grades, from Grade 3 to Grade 11 (9 grades), is 50%. The mean percentage of the boys in each of the grades, from Grades 3 to 12 (10 grades), is 49,68%. What percentage of the learners in the Grade 12 sample are boys?

# MEASURES OF CENTRAL TENDENCY OF DATA IN A FREQUENCY TABLE

✓ We can find the mode, median and mean of data in a ***frequency table.***

## a) Mode

✓ The mode is the value in the table that occurs most often.

✓ It is the value with the ***greatest frequency***.

✓ ***Remember:*** The mode is the ***value***, not the frequency.

**EXAMPLE 6**

Zanele did a survey of 10 of her friends.
She asked them how many siblings they had.
The frequency table shows the results of her survey:

Your ***sibling*** is your brother or sister.

| **Number of siblings** | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| **Frequency** | 2 | 3 | 4 | 1 |

Find the mode of the number of siblings.

**<u>SOLUTION</u>:**
The greatest frequency in the table is 4.
This means that four of her friends had 2 siblings.
So the mode = 2 siblings.

## b) Median

✓ One way to find the median is to list all the values in the frequency table in order of size.

✓ Another way is to work directly with the table.

**EXAMPLE 7**

Look again at the frequency table showing the results of Zanele's survey.

| **Number of siblings** | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| **Frequency** | 2 | 3 | 4 | 1 |

Find the median of the data in the frequency table.

**<u>SOLUTION</u>:**

<u>METHOD 1</u>: List all the values in order of size:

0; 0; 1; 1; **<u>1; 2</u>**; 2; 2; 2; 3

***The median*** $= \frac{1+2}{2} = \frac{3}{2} =$ ***1,5 siblings***

<u>METHOD 2</u>: Find the median directly from the table.

- The values in the table are already in order of size.
- Count along the frequencies to find where the middle value or values lie.
  For Zanele's data, adding the frequencies gives 2 + 3 + 4 + 1 = 10
  There are 10 values in the data. Half of 10 is 5.
  So the median is half-way between the 5$^{th}$ and 6$^{th}$ values.
  Count along the frequencies to find these values.

| **Number of siblings** | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| **Frequency** | 2 | 3 | 4 | 1 |
| | 2 values to here | 2+3=5 values to here | 6$^{th}$ value must be in here | |

So the 5$^{th}$ value must be 1 and the 6$^{th}$ value must be 2.

***The median*** $= \frac{1+2}{2} = \frac{3}{2} =$ ***1,5 siblings***

***NOTE:***

- *When there is an even number of data items, there is a possibility that the median is **not** one of the data items, and that it is a decimal value. As a result we can often end up with an answer like **1,5 siblings**.*
- *We **don't** need round off an answer like this because our interpretation of the situation is that "50% of Zanele's friends have less than 1,5 siblings (in other words 0 or 1), and 50% of Zanele's friends have more than 1,5 siblings (in other words 2 or 3)".*

## c) Mean

- ✓ One way to find the mean is to list all the values in the frequency table, add up these values, and then divide the answer by the number of values
- ✓ Another way is to work directly with the table using the formula $\bar{x} = \frac{\sum f.x}{n}$ where $f$ is the frequency, $x$ is the data item, and $n$ is the number of data items in the set of data.

**EXAMPLE 8**

Look again at the frequency table showing the results of Zanele's survey.

| **Number of siblings** | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| **Frequency** | 2 | 3 | 4 | 1 |

Find the mean of the data in the frequency table.

**<u>SOLUTION</u>:**

<u>METHOD 1</u>: List all the values in the table: 0; 0; 1; 1; 1; 2; 2; 2; 2; 3

***The mean*** $= \frac{\textit{total number of siblings}}{\textit{frequency}} = \frac{0+0+1+1+1+2+2+2+2+3}{10} = \frac{14}{10} =$ ***1,4 siblings***

<u>METHOD 2</u>: Find the mean directly from the table.

- To find the *total number of siblings* you must take the frequency of each value into account. We multiply each *number of siblings* by *its frequency*, and then add the numbers.
- Change the table to a vertical one, and add in another column:

| **VALUE**<br>**Number of siblings**<br>$x$ | **FREQUENCY**<br>**Number of Zanele's friends**<br>$f$ | **FREQUENCY × VALUE**<br>$f \times x$ |
|---|---|---|
| 0 | 2 | $2 \times 0 = 0$ |
| 1 | 3 | $3 \times 1 = 3$ |
| 2 | 4 | $4 \times 2 = 8$ |
| 3 | 1 | $1 \times 3 = 3$ |
| | ***n = 10*** | $\sum f.x = 14$ |

***The mean*** $= \bar{x} = \frac{\sum f.x}{n} = \frac{14}{10} = 1{,}4$ *siblings*

> ***NOTE:***
> - *The mean is **not necessarily** one of the data items. This means that it is possible to end up with an answer of **1,4 siblings**.*
> - *Again, we do not round off an answer like this because our interpretation of the situation is that "if the total number of siblings were shared out equally, each friend would get 1,4 siblings."*

**EXAMPLE 9**

A certain school provides buses to transport the learners to and from a nearby village. A record is kept of the number of learners on each bus for 26 school days.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 27 | 25 | 27 | 29 | 31 | 24 | 25 | 27 | 28 | 29 | 24 | 26 | 30 |
| 28 | 31 | 25 | 25 | 27 | 28 | 28 | 28 | 26 | 28 | 31 | 24 | 30 |

a) Organise the data in a frequency table
b) Use the table to calculate the total number of learners that were transported to school by bus.
c) Calculate the mean number of learners per trip, correct to one decimal place.
d) Explain what the mean represents.
e) Find the mode.
f) Find the median and explain what the median represents.

**SOLUTION:**

a)

| **Number of learners per trip ($x$)** | **Frequency ($f$)** | $f \times x$ |
|---|---|---|
| 24 | 3 | $3 \times 24 = 72$ |
| 25 | 4 | $4 \times 25 = 100$ |
| 26 | 2 | $2 \times 26 = 52$ |
| 27 | 4 | $4 \times 27 = 108$ |
| 28 | 6 | $6 \times 28 = 168$ |
| 29 | 2 | $2 \times 29 = 58$ |
| 30 | 2 | $2 \times 30 = 60$ |
| 31 | 3 | $3 \times 31 = 93$ |
| | $n = 26$ | $\sum f.x = 711$ |

b) Total number of learners that were transported to school by bus = 711

c) Mean number of learners on each bus $= \dfrac{\text{total number of learners on all the bus trips}}{\text{total number of bus trips}}$

$$= \frac{\sum f.x}{n}$$
$$= \frac{711}{26}$$
$$= 27{,}3461\ldots$$
$$\approx 27{,}3$$

d) The mean tells us that if each bus had exactly the same number of people each time, there would be approximately 27 learners on each bus.

e) The largest value in the frequency column is 6 and it goes with 28 learners. This means that the mode = 28 learners.

**EXAMPLE 9 (continued)**

f) There are 26 bus trips.
Half of 26 = 13, so the median lies between the 13$^{th}$ and the 14$^{th}$ values on the table.

| Number of learners per trip | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 |
|---|---|---|---|---|---|---|---|---|
| Frequency | 3 | 4 | 2 | 4 | 6 | 2 | 2 | 3 |
| | 3 values to here | 3 + 4 = 7 values to here | 7 + 2 = 9 values to here | 9 + 4 = 13 values to here | 13 + 6 = 19 values to here | | | |

So the 13$^{th}$ value is 27, and the 14$^{th}$ value is 28.
Median = $\frac{27+28}{2} = \frac{55}{2}$ = 27,5 learners.

The median tells us that for half of the bus trips there were less than 27,5 learners (which means 27 and less) on the bus and for half of the bus trips there were more than 27,5 learners (which means 28 or more) on the bus.

## d) Using a Scientific Calculator to Find the Mean of Data

✓ A scientific calculator makes it quicker and easier to find the mean of data in a frequency table.

✓ The key sequences for the CASIO *fx-82ZA* PLUS and the SHARP EL-W535HT that can be used to find the mean in Example 9 are as follows:

| CASIO | SHARP |
|---|---|
| • First add in a frequency column:<br>[SHIFT] [SETUP] [▼] [3:STAT] [1:ON]<br>• Then enter the data<br>[SETUP] [2:STAT] [1:1-VAR]<br>24 [=] 25 [=] 26 [=] 27 [=] 28 [=] 29 [=] 30 [=] 31 [=]<br>[▼] [▶]<br>3 [=] 4 [=] 2 [=] 4 [=] 6 [=] 2 [=] 2 [=] 3 [=]<br>[AC]<br>[SHIFT] [STAT] [1] [4:VAR] [2 :$\bar{x}$] | [MODE] [1 : STAT] [0 : SD]<br>[2ndF] [MODE] [CA]<br>24 [(x ; y)] 3 [DATA]<br>25 [(x ; y)] 4 [DATA]<br>26 [(x ; y)] 2 [DATA]<br>27 [(x ; y)] 4 [DATA]<br>28 [(x ; y)] 6 [DATA]<br>29 [(x ; y)] 2 [DATA]<br>30 [(x ; y)] 2 [DATA]<br>31 [(x ; y)] 3 [DATA]<br>[ALPHA] [4] [$\bar{x}$] |

## EXERCISE 1.2

1) In the 2009 Census@School survey, learners were asked to measure the length of their right foot. A group of Grade 10 learners measured the lengths of each other's feet and recorded the lengths obtained correct to the nearest 0,5 cm. The results are summarised in the table below:

| Length of foot (in cm) | 22,5 | 23 | 23,5 | 24 | 24,5 | 25 | 25,5 | 26 |
|---|---|---|---|---|---|---|---|---|
| Number of learners | 2 | 4 | 6 | 8 | 10 | 5 | 3 | 1 |

   a) Use the table to determine the following:
      i) The value of the mode
      ii) The value of the median
      iii) The mean foot length.

   b) What does each average tell you about the foot lengths for the group of Grade 10 learners?

2) Grade 11B did an Investigation Task to determine the effect of the Consumer Price Index (CPI) on inflation. The marks they obtained for their Investigations are as follows:

| Investigation mark (%) | 46 | 68 | 72 | 75 | 78 | 82 | 85 | 90 | 91 |
|---|---|---|---|---|---|---|---|---|---|
| Frequency (number of learners) | 1 | 3 | 5 | 6 | 4 | 2 | 2 | 1 | 1 |

   a) Determine the marks that are the mean, the median and the mode.
   b) Explain what each measure of central tendency tells you about the results of the investigation.

3) The table below is adapted from Census 2011 and shows the ***unemployment rate*** of 25 municipalities in the Western Cape. (The unemployment rate is the percentage of the total labour force that is **unemployed** but actively seeking employment and willing to work.) The rate has been rounded off to the nearest whole number.

| Unemployment rate (%) | 7 | 11 | 13 | 14 | 15 | 17 | 18 | 19 | 21 | 23 | 24 | 25 | 26 | 30 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequency (number of municipalities) | 1 | 3 | 1 | 4 | 2 | 1 | 2 | 1 | 1 | 3 | 1 | 2 | 1 | 1 |

   a) Determine the mean, median and mode of the unemployment rate for the 25 municipalities.
   b) Explain what each measure of central tendency tells you about the results in the table.

# MEASURES OF CENTRAL TENDENCY IN A GROUPED FREQUENCY TABLE

## a) Discrete Data, Continuous Data and Categorical Data

✓ There are three common types of data: ***discrete data***, ***continuous data*** and ***categorical data***.

- ***Discrete data*** consists of numerical values that are found ***by counting***. They are often whole numbers.
  Examples of discrete data are:
  - Number of children in a family
  - Number of rooms in a house
  - Marks scored in a maths test
- ***Continuous data*** consists of numerical values that are found ***by measuring***. They are given to a certain *degree of accuracy*. They cannot be given *exactly*. They are often decimals.
  Examples of continuous data are:
  - Foot lengths
  - Time taken to travel to school
  - Mass of books in a school bag
  - Amount of water drunk in a day.
- ***Categorical data*** consists of descriptions using names.
  Examples of categorical data are:
  - Heads or Tails
  - Boy or Girl
  - House, shack, flat or informal dwelling.

## b) Grouped Frequency Tables

✓ We can group discrete data and continuous data in ***grouped frequency tables***. In a grouped frequency table the values are grouped in ***class intervals***. The ***frequencies*** show the number of values in each class interval.

✓ We can find the mode, median and mean of data in a ***grouped frequency table.***

✓ If you do not have the raw data, you do not know the actual values in each class interval. So you cannot find the actual mode, median and mean of the data. However, you can make ***reasonable estimates*** of these answers.

## c) Mode

- ✓ The ***modal interval*** is the interval in the table that occurs most often. It is the group of values with the ***greatest frequency***.
- ✓ Note: ***Mode*** refers to a single value that occurs most often; ***modal interval*** refers to the group of values that occurs most often.
- ✓ ***Remember:*** The mode is the ***value***, not the frequency.

**EXAMPLE 10**

The choir teacher kept a record of the number of learners who attended the 40 choir practices during the year.

This frequency table gives a summary of the attendance.

| **Number of learners at choir practice ($x$)** | **Frequency ($f$)** |
|---|---|
| $0 < x \leq 10$ | 1 |
| $10 < x \leq 20$ | 2 |
| $20 < x \leq 30$ | 11 |
| $30 < x \leq 40$ | 9 |
| $40 < x \leq 50$ | 14 |
| $50 < x \leq 60$ | 3 |
| | $\boldsymbol{n = 40}$ |

Find the modal interval.

**<u>SOLUTION</u>:**

The interval with the greatest frequency is $40 < x \leq 50$.

This means that there were more times when there were from 40 up to and including 50 learners at the choir practices than any other interval.

So, the ***modal interval*** is $40 \leq x \leq 50$.

## d) Median

- ✓ You cannot find the median directly from a grouped frequency table. You can find the class interval that contains the median, and then find the approximate value of the median by finding the midpoint of the interval.

**EXAMPLE 11**
Look again at the choir teacher's summary of attendance. This time find the median number of learners at choir practice.

| Number of learners at choir practice ($x$) | Frequency ($f$) |
|---|---|
| $0 < x \leq 10$ | 1 |
| $10 < x \leq 20$ | 2 |
| $20 < x \leq 30$ | 11 |
| $30 < x \leq 40$ | 9 |
| $40 < x \leq 50$ | 14 |
| $50 < x \leq 60$ | 3 |
| | $n = 40$ |

**<u>SOLUTION</u>:**
There were 40 choir practices.
Half of 40 is 20, so the median is half-way between the 20th and 21st values.

*Count down the frequencies to find the position of the median*

| Number of learners at choir practice ($x$) | Frequency ($f$) | |
|---|---|---|
| $0 < x \leq 10$ | 1 | ... 1 value to here |
| $10 < x \leq 20$ | 2 | ... 1 + 2 = 3 values to here |
| $20 < x \leq 30$ | 11 | ... 3 + 11 = 14 values to here |
| $30 < x \leq 40$ | 9 | ... 14 + 9 = 23 values to here |
| $40 < x \leq 50$ | 14 | |
| $50 < x \leq 60$ | 3 | |
| TOTAL | 40 | |

The 15th to 23rd values are in the class interval $30 < x \leq 40$
So the 20th and 21st values are in this class.
So the ***median class is*** $\boldsymbol{30 < x \leq 40}$

To find ***an estimate of the median***, find the midpoint of the interval.
The midpoint of the interval $30 < x \leq 40$ is $\frac{30+40}{2} = \frac{70}{2} = 35$
So the ***median ≈ 35 learners***

The median tells us that 50% of the choir practices were attended by less than 35 learners and 50% of the choir practices were attended by more than 35 learners.

## e) Mean

✓ To find *an estimate of the mean:*

i) Find the midpoint of each interval to represent each class (usually written as *X*). Then you assume that each item in the interval has that value.

ii) Multiply the midpoint by the frequency in order to work out an estimate of the total of the values in the class.

iii) Add these values together to get an estimate of the total of all the values.

iv) Substitute the values directly into the formula $\bar{X} = \frac{\sum f.X}{n}$ where $\bar{X}$ is the approximate value of the mean, *X* is the value of the midpoint of each interval, and *f* is the frequency of that interval.

**EXAMPLE 12**

Look again at the choir teacher's summary of attendance.

| Number of learners at choir practice ($x$) | Frequency ($f$) |
|---|---|
| $0 < x \leq 10$ | 1 |
| $10 < x \leq 20$ | 2 |
| $20 < x \leq 30$ | 11 |
| $30 < x \leq 40$ | 9 |
| $40 < x \leq 50$ | 14 |
| $50 < x \leq 60$ | 3 |
| | $n = 40$ |

Find the approximate value of the mean number of learners who attended choir practice.

**SOLUTION:**

First add in another column and work out the ***midpoint of each interval***. Then, add another column and calculate ***frequency*** × ***mid-point value***.

| Number of learners at choir practice ($x$) | Frequency $f$ | Midpoint of the interval $X$ | $f \times X$ |
|---|---|---|---|
| $0 < x \leq 10$ | 1 | $\frac{0+10}{2} = 5$ | $1 \times 5 = 5$ |
| $10 < x \leq 20$ | 2 | $\frac{10+20}{2} = 15$ | $2 \times 15 = 30$ |
| $20 < x \leq 30$ | 11 | $\frac{20+30}{2} = 25$ | $11 \times 25 = 275$ |
| $30 < x \leq 40$ | 9 | $\frac{30+40}{2} = 35$ | $9 \times 35 = 315$ |
| $40 < x \leq 50$ | 14 | $\frac{40+50}{2} = 45$ | $14 \times 45 = 630$ |
| $50 < x \leq 60$ | 3 | $\frac{50+60}{2} = 55$ | $3 \times 55 = 165$ |
| | $n = 40$ | | $\sum f.X = 1\,420$ |

**<u>EXAMPLE 12</u> (continued)**

Mean = $\bar{X} = \frac{\sum f.X}{n} \approx \frac{1\,420 \text{ learners}}{40}$ = 35,5 learners

The mean tells us that if the total number of learners attending the 40 choir practices were shared out equally, then approximately 35,5 learners would have attended each choir practice.

> ***NOTE:***
> - *Once again, we don't round this amount off.*
> - *We use the value of the mean to **interpret** the situation, so don't have to end up with a value that is a whole number.*

## f) Using a Scientific Calculator to Find the Mean of Data

✓ A scientific calculator makes it quicker and easier to find the mean of grouped data.

✓ The key sequences for finding the mean in Example 12 using the CASIO *fx-82ZA* PLUS and the SHARP EL-W535HT are as follows:

| CASIO | SHARP |
|---|---|
| [SETUP] [2:STAT] [1:1-VAR]<br>***First enter the midpoints***<br>5 [=] 15 [=] 25 [=] 35 [=] 45 [=] 55 [=]<br>[▼] [▶]<br>***Then enter the frequencies***<br>1 [=] 2 [=] 11 [=] 9 [=] 14 [=] 3 [=]<br>[AC]<br>[SHIFT] [STAT] [1] [4:VAR] [2 : $\bar{x}$] | [MODE] [1 : STAT] [0 : SD]<br>[2ndF] [MODE] [CA]<br>***Enter the midpoints and frequencies together***<br>5 [(x ; y)] 1 [DATA]<br>15 [(x ; y)] 2 [DATA]<br>25 [(x ; y)] 11 [DATA]<br>35 [(x ; y)] 9 [DATA]<br>45 [(x ; y)] 14 [DATA]<br>55 [(x ; y)] 3 [DATA]<br>[ALPHA] [4] [$\bar{x}$] |

**EXAMPLE 13**

In a particular primary school in Pietermaritzburg, it was found that ninety of their Foundation Phase learners (Grades 1, 2 and 3) were accompanied to school by someone. The ages of the person accompanying the child were recorded, as shown in the table below.

| **Age (in years) ($x$)** | **Frequency ($f$)** |
|---|---|
| $0 < x \leq 10$ | 12 |
| $10 < x \leq 20$ | 30 |
| $20 < x \leq 30$ | 18 |
| $30 < x \leq 40$ | 12 |
| $40 < x \leq 50$ | 9 |
| $50 < x \leq 60$ | 6 |
| $60 < x \leq 70$ | 3 |

Use the information given in the table to

a) Determine the modal interval.
b) Estimate the mean age of the person accompanying a learner from the Foundation Phase.
c) Estimate the median age of the person accompanying a learner from the Foundation Phase.

**SOLUTION:**

a) The modal interval is $10 < x \leq 20$.
This means that more people in this age group accompanied the learners to school than any other age group.

b) To find the mean we have to take the midpoint of each class interval and then calculate *frequency* $\times$ *midpoint* for each class interval.

| **Age (in years) $x$** | **Midpoint $X$** | **Frequency $f$** | **$f.X$** |
|---|---|---|---|
| $0 < x \leq 10$ | $\frac{0+10}{2} = 5$ | 12 | $12 \times 5 = 60$ |
| $10 < x \leq 20$ | $\frac{10+20}{2} = 15$ | 30 | $30 \times 15 = 450$ |
| $20 < x \leq 30$ | $\frac{20+30}{2} = 25$ | 18 | $18 \times 25 = 450$ |
| $30 < x \leq 40$ | $\frac{30+40}{2} = 35$ | 12 | $12 \times 35 = 420$ |
| $40 < x \leq 50$ | $\frac{40+50}{2} = 45$ | 9 | $9 \times 45 = 405$ |
| $50 < x \leq 60$ | $\frac{50+60}{2} = 55$ | 6 | $6 \times 55 = 330$ |
| $60 < x \leq 70$ | $\frac{60+70}{2} = 65$ | 3 | $3 \times 65 = 195$ |
| | | $n = 90$ | $\sum f.X = 2\ 310$ |

$$\text{Mean} = \bar{X} = \frac{\sum f.X}{n} \approx \frac{2\ 310 \text{ years}}{90} = 25{,}7 \text{ years old}$$

*The mean tells us that if all the ages were added together, and then shared out equally amongst the 90 people, then each one would be 25,7 years old.*

**EXAMPLE 13 (continued)**

c) 90 people accompanied the learners to school.
$90 \div 2 = 45$. So the median lies half-way between the 45[th] and 46[th] person.

| Age (years) | Frequency | |
|---|---|---|
| $0 < x \leq 10$ | 12 | … 12 people to here |
| $10 < x \leq 20$ | 30 | … 12 + 30 = 42 people to here |
| $20 < x \leq 30$ | 18 | … 42 + 18 = 60 people to here. |
| $30 < x \leq 40$ | 12 | |
| $40 < x \leq 50$ | 9 | |
| $50 < x \leq 60$ | 6 | |
| $60 < x \leq 70$ | 3 | |

Both the 45[th] and 46[th] people lie in the interval $20 < x \leq 30$.
So the median interval is 20 years $< x \leq$ 30 years, and the approximate value of the median is 25 years.

The value of the median tells us that 50% of the people accompanying the Grade 1, 2 and 3 learners to school are less than or equal to 25 years old, and 50% are more than or equal to 25 years old.

**EXERCISE 1.3**

1) The table below represents the ages of the 90 people accompanying Foundation Phase learners to another primary school in Pietermaritzburg:

| **Age (in years) ($x$)** | **Frequency ($f$)** |
|---|---|
| $0 \leq x < 10$ | 4 |
| $10 \leq x < 20$ | 12 |
| $20 \leq x < 30$ | 25 |
| $30 \leq x < 40$ | 14 |
| $40 \leq x < 50$ | 10 |
| $50 \leq x < 60$ | 20 |
| $60 \leq x < 70$ | 5 |
| | $n = 90$ |

a) Use the information given in the table to
   i) Determine the modal interval.
   ii) Estimate the mean age of the people accompanying a learner from the Foundation Phase.
   iii) Estimate the median age of the people accompanying a learner from the Foundation Phase.

b) What do the modal class, the mean and the median tell you about the ages of the people who accompany the children to school?

2) The table below has been adapted from Census 2011. It lists the property values of 262 properties in a part of the Ikwezi municipality.

| **Property Value in Rand ($x$)** | **Frequency (number of households) ($f$)** |
|---|---|
| $0 \leq x < 50\,000$ | 172 |
| $50\,000 \leq x < 100\,000$ | 51 |
| $100\,000 \leq x < 150\,000$ | 18 |
| $150\,000 \leq x < 200\,000$ | 12 |
| $200\,000 \leq x < 250\,000$ | 9 |
| | $n = 262$ |

a) Use the information given in the table to
   i) Determine the modal interval.
   ii) Estimate the mean property value of the properties.
   iii) Estimate the median property value of the properties.

b) What does the modal class, the mean and the median tell you about these property values?

# QUARTILES AND THE FIVE NUMBER SUMMARY

## a) Quartiles

✓ ***Quartiles*** are the three values $Q_1$, $Q_2$ and $Q_3$ that divide a data set into ***four*** approximately equal parts. Each part consists of approximately 25% of the elements of the data set.

✓ $\boldsymbol{Q_1}$ is the ***lower quartile***; $\boldsymbol{Q_2}$ is the middle quartile or ***median*** and $\boldsymbol{Q_3}$ is the ***upper quartile***.

  ✓ The ***median*** divides an ordered data set into *two halves*.

  ✓ The ***quartiles*** divide an ordered data set into *four quarters*.

  ✓ The ***median*** is also the 2nd quartile (***M*** or $\boldsymbol{Q_2}$).

| *25% of the data* | *25% of the data* | *25% of the data* | *25% of the data* |
|---|---|---|---|

**Lower quartile $Q_1$** | **Median M or $Q_2$** | **Upper quartile $Q_3$**

✓ From the above diagram, one can see that:

- Approximately one quarter or 25% of the data is less than $Q_1$.
- Approximately three quarters or 75% of the data is more than $Q_1$.
- Approximately one half or 50% of the data is less than $Q_2$ and one half or 50% is more than $Q_2$.
- Approximately three quarters or 75% of the data is less than $Q_3$.
- About one quarter or 25% of the data is more than $Q_3$.
- Approximately one half or 50% of the data lies between $Q_1$ and $Q_3$.

**HOW TO FIND THE QUARTILES:**

i) Put the data items in order and find the median.

ii) Find the midpoint of the data items to the left of the median. This is the ***lower quartile*** $(\boldsymbol{Q_1})$.

iii) Find the midpoint of the data items to the right of the median. This is the ***upper quartile*** $(\boldsymbol{Q_3})$.

**EXAMPLE 14**

For each of the following sets of data

a) 23 65 33 101 23 21 102 18 26 9

b) 65 33 101 23 21 102 18 26 9

i) Find the median (M)

ii) Find the lower quartile ($Q_1$) and the upper quartile ($Q_3$)

**SOLUTION:**

a) First arrange the data in ascending or decending order.

9 18 21 23 23 26 33 65 101 102

i) There are ten data items (an even number of data items). To find the median, we have to find the mean of the middle two numbers.

$10 \div 2 = 5$, so the median lies between the 5th and the 6th terms.

9 18 21 23 **23 26** 33 65 101 102

So the median = M = $\frac{23+26}{2} = \frac{49}{2} = 24{,}5$

ii) To find the lower quartile ($Q_1$), take the data before the median and find the median of that.

There is an odd number of data items below the median. Take the middle one.

9 18 **21** 23 23 | 26 33 65 101 102

So, $Q_1 = 21$.

To find the upper quartile ($Q_3$), take the data after the median and find the median of that:

There is an odd number of data items above the median; take the middle one.

9 18 **21** 23 23 | 26 33 **65** 101 102

So, $Q_3 = 65$.

So, the three quartiles are 21; 24,5 and 65.

b) First arrange the data in ascending or decending order.

9 18 21 23 26 33 65 101 102

i) There are nine data items (an odd number of data items).

$9 \div 2 = 4{,}5$ so the median is the 5th term.

9 18 21 23 **26** 33 65 101 102

So the median = 26

ii) To find the lower quartile ($Q_1$), take the data before the median (26) and find the median of that.

There is an even number of data items below the median.

$4 \div 2 = 2$, so the lower quartile, $Q_1$, lies between the 2nd and 3rd terms.

9 18 | 21 23 **26** 33 65 101 102

So the $Q_1 = \frac{18+21}{2} = \frac{39}{2} = 19{,}5$

**EXAMPLE 14 (continued)**

To find the upper quartile ($Q_3$), take the data after the median and find the median of that.
There is an even number of data items above the median.
$4 \div 2 = 2$, and $5 + 2 = 7$, so $Q_3$ lies between the 7th and the 8th terms.

9 18 | 21 23 **26** 33 65 | 101 102

So the $Q_3 = \frac{65+101}{2} = \frac{166}{2} = 83$

So, the three quartiles are 19,5; 26 and 83.

## b) The Five Number Summary

- ✓ The ***five number summary*** consists of 5 items
  1) The minimum value in the data set;
  2) $Q_1$, the lower quartile;
  3) M, the median;
  4) $Q_3$, the upper quartile;
  5) The maximum value in the data set.

- ✓ Use the following method to find a five number summary:

| STEPS | EXAMPLE |
|---|---|
| 1) Put your numbers in the data set in order. | Find the five-number summary of the following data set: 1, 3, 5, 6, 12, 15, 23, 28, 31 |
| 2) Find the minimum and maximum values. | Minimum value = 1<br>Maximum value = 31 |
| 3) Find the median | There are 9 data items.<br>$9 \div 2 = 4{,}5$ so the median is the 5th data item<br>1 3 5 6 **12** 15 23 28 31<br>So the median = 12 |
| 4) Find $Q_1$ and $Q_3$. | There are 4 items below the median.<br>$4 \div 2 = 2$, so $Q_1$ lies between the 2nd and 3rd terms.<br>1 3 \| 5 6 **12** 15 23 28 31<br>So $Q_1 = \frac{3+5}{2} = 4$<br><br>There are 4 items above the median.<br>$5 + 2 = 7$, so $Q_3$ lies between the 7th and 8th terms.<br>1 3 \| 5 6 **12** 15 23 \| 28 31<br>So $Q_2 = \frac{23+28}{2} = 25{,}5$ |
| 5) Write down the five number summary. | **Minimum = 1**<br>**$Q_1$ = 4**<br>**Median = 12**<br>**$Q_3$ = 25,5**<br>**Maximum = 31** |

**EXERCISE 1.4**

1) Find the five number summary for each of the following data sets:
   a) 1 6 6 9 15 17 23 24 33 33 38 38 38 45 46 51
   b) 9 14 19 21 24 29 29 32 33 35 36 40 46 49
   c) 45 15 43 19 26 25 15 36 27 32 41 25 48
   d) 4 46 6 44 10 17 34 35 31 22 10 16

2) Johan is asked to find the five number summary for the following set of numbers:
   2 23 24 12 11 23 34 12 34 12 33 19 48 25 37 38 59

   Johan's answer is as follows:

   > 2 11 12 12 12 19 23 23 24 25 33 34 34 37 38 48 59
   >
   > Min = 2
   > $Q_1$ = 12
   > Median = 24
   > $Q_3$ = 34 or 37
   > Max = 59

   a) Give the correct solution to the question.
   b) Identify the mistake that Johan has made. Describe the misconception and explain to John why he is incorrect.

# PERCENTILES

✓ When you write a test and get a mark of 75%, it tells you how many questions you got right. ***But***, it doesn't tell you how well you did compared to the other people who wrote the test. ***Percentiles*** are values from 0 to 99 that tell you the ***percentage of the marks that are less than a particular mark.***

If the ***percentile of your test mark is 75***, it tells that

- ***75% of the marks are LESS than yours***.
- ***100% – 75% = 25% of the marks are MORE than yours***

✓ Percentiles can be used to compare values in any set of ***ordered*** data. You can calculate percentiles for income, mass, etc. Percentiles are often used in education and health-related fields to indicate how one person compares with others in a group.

> ***NOTE:***
> *There is a difference between a mark of **60%** (**a percentage** telling you that you got 60 out of 100) and a mark at the **60th percentile** (which tells us that approximately 60% of the marks are less than yours).*

✓ The following are some special percentiles:

- The ***median*** is at the 50th percentile
- The ***lower quartile*** is at the 25th percentile
- The ***upper quartile*** is at the 75th percentile.

✓ Scores that are in the ***95th percentile*** and above are unusually high while those in the ***5th percentile*** and below are unusually low.

***Example:***

a) The ***89th percentile*** is a number that 89% of the data items are below. We then also know that 100% – 89% = 11% of the data items are above that number.

b) If the ***20th percentile*** is 12, then 20% of the data items are less than 12, and 80% of the data items are more than 12.

c) If the ***90th percentile*** is 17, then 90% of the data items are less than 17, and 10% of the data items are more than 17.

✓ We usually find percentiles of a ***large*** number of items.

**EXAMPLE 15**

For the following set of 19 data items

72 71 65 60 62 58 67 57 70 73

50 61 51 55 64 68 69 59 63

a) At what percentile is 70?

b) Find the data item that is at the 20th percentile.

c) What is the median score (the score at the 50th percentile)?

**SOLUTION:**

A stem and leaf diagram can be used to get the data into ascending order:

| Stem | Leaf |
|---|---|
| 5 | 0 1 5 7 8 9 |
| 6 | 0 1 2 3 4 5 7 8 9 |
| 7 | 0 1 2 3 |
| **KEY:** 6/2 = 62 | |

a) ***To calculate at which percentile a data item lies:***

$$\text{Percentile} = \frac{\textit{number of data items that are less than or equal to } 70}{\textit{total number of data items}} \times 100\%$$

$$= \frac{16}{19} \times 100\%$$

$$= 84{,}210\ldots\%$$

$$\approx 84\%$$

***So 70 is at the 84th percentile.***

a) ***To calculate the data item which is at a given percentile:***

i) *Find out which term corresponds to the 20th percentile:*

Data item corresponding to the 20th percentile

$$= \frac{percentile}{100} \times number\ of\ data\ items$$

$$= \frac{20}{100} \times 19 \text{ data items}$$

$$= 3{,}8$$

$$\approx 4$$

ii) *Count along until you get to the 4th data item* – It is 57

***So 57 is the 20th percentile***

b) ***To calculate the data item which is at a given percentile:***

The data item corresponding to the 50th percentile

$$= \frac{percentile}{100} \times number\ of\ data\ items$$

$$= \frac{50}{100} \times 19$$

$$= 9{,}5$$

$$\approx 10$$

Count along to find the 10th data item – It is 63.

***So 63 is the 50th percentile or the median.***

## EXERCISE 1.5

1) Given the following set of data:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 24 | 34 | 35 | 37 | 39 | 40 | 41 | 45 | 46 | 48 |
| 50 | 52 | 54 | 55 | 56 | 56 | 59 | 59 | 60 | 61 |
| 63 | 66 | 67 | 68 | 69 | 70 | 73 | 75 | 77 | 77 |
| 78 | 79 | 79 | 80 | 84 | 84 | 86 | 86 | 86 | 89 |
| 94 | 95 | 96 | 97 | 98 | 98 | 100 | 101 | 102 | 103 |

a) At which percentile is:
   - i) 102?
   - ii) 34?
   - iii) 96?
   - iv) 70?

b) Find the item that corresponds to
   - i) The $30^{th}$ percentile
   - ii) The $50^{th}$ percentile
   - iii) The $10^{th}$ percentile
   - iv) The $65^{th}$ percentile

2) For the following stem and leaf diagram

| STEM | LEAF |
|---|---|
| 10 | 0 0 2 3 3 3 4 5 7 7 7 7 8 8 9 |
| 11 | 1 1 2 3 5 5 5 6 6 6 8 9 |
| 12 | 0 0 0 1 3 4 6 7 7 9 9 9 9 9 |
| 13 | 1 2 2 2 5 6 7 7 7 8 8 8 9 |
| 14 | 0 1 1 1 1 2 3 5 5 5 5 6 8 8 8 9 |
| KEY: 10/4 = 104 | |

a) What is the score that corresponds to the $11^{th}$ percentile?
b) Find the data item that corresponds to the $44^{th}$ percentile
c) At which percentile is a score of 135?

# MEASURES OF DISPERSION

- ✓ A ***measure of central tendency*** such as the mean, median and mode gives you a single measurement to stand for a set of data. A ***measure of dispersion*** or ***measure of spread*** tells you *how spread out the data is.*

- ✓ The data can either:
  - Be grouped closely together around the measure of central tendency, or
  - Be spread widely apart around the measure of central tendency.

## a) Range

- ✓ The range is the simplest measure of spread. It is the difference between the largest and smallest items of data.

***Range = Largest Value – Smallest Value***

- ✓ There are some limitations to using range:
  - It does not take into account anything about the distribution of any other piece of data except the smallest and largest value.
  - When data is given in a grouped frequency table, the range cannot be used.

## b) Interquartile Range

- ✓ The interquartile range (or ***IQR***) is the difference between the upper quartile and the lower quartile.

***Interquartile Range*** $= Q_3 - Q_1$

- ✓ The interquartile range is a better measure of dispersion than the range. It is not affected by any extreme values (very small or very large values). It is based on the ***middle half*** of the data. It is the *range between the upper and lower quartiles.*

- ✓ The ***semi-interquartile range*** is sometimes used. It is half of the interquartile range.

***Semi-interquartile range*** $= \frac{Q_3 - Q_1}{2}$

**EXAMPLE 16**

a) For the following set of data: 22 17 28 19 23 18 25 29 19 29
Find
   i) the range
   ii) the interquartile range
   iii) the semi-interquartile range

b) Approximately what percentage of the data items lie within the interquartile range?

**SOLUTION:**

a)

i) Arrange the data in order: 17 18 19 19 22 23 25 28 29 29

***Range = Largest Value – Smallest Value= 29 – 17 = 15***

ii) First find the median:
There are 10 terms.
$10 \div 2 = 5$ which means that the median lies between the 5th and 6th terms
Median = $\frac{22+23}{2} = \frac{45}{2} = 22{,}5$

The find the two quartiles

17 18 **19** 19 22 | 23 25 **28** 29 29

There are five terms to the left of the median and five terms to the right of the median.
So $Q_1 = 19$ and $Q_3 = 29$

***Interquartile range =*** $Q_3 - Q_1 = 28 - 19 = 9$

iii) ***Semi-interquartile range =*** $\frac{Q_3 - Q_1}{2} = \frac{28-19}{2} = \frac{9}{2} = 4{,}5$

b) Approximately 50% of the data items lie within the interquartile range.

**EXERCISE 1.6**

1) Find the median, lower quartile, upper quartile and the interquartile range for each of the following sets of data:
   a) 9 11 12 12 13 13 14 16 21 22 24
   b) 4 5 5 6 7 9 10 11 12 12 13 14 14 14
   c) 12 7 1 3 2 12 2 9 14 5 6 5 4 8 11 14

2) A group of 21 learners attending extra mathematics classes were required to write a test which was out of 50.
   Their results were:

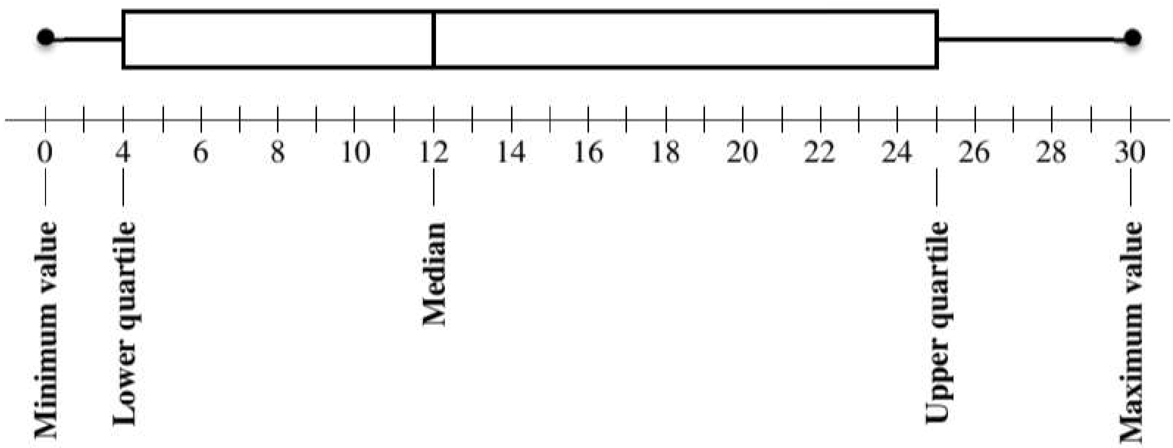
| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 17 | 8 | 19 | 9 | 12 | 28 | 11 | 16 | 20 | 14 | 29 |
| 23 | 37 | 23 | 26 | 4 | 35 | 26 | 18 | 45 | 7 | |

   a) Find the range.
   b) Find the lower and upper quartiles.
   c) Calculate the interquartile range.
   d) What do the lower and upper quartiles indicate about the results of the test?

# BOX AND WHISKER DIAGRAMS

✓ A graphical representation of the five number is known as a ***box and whisker diagram***, also sometimes called a ***box plot.***

- Vertical lines mark the two quartiles and the median. These are joined to make a ***box*** containing the middle half of the data. The box illustrates the ***interquartile range***.
- From the quartiles, horizontal lines are drawn to the minimum and maximum values. These lines are the ***whiskers***.




| HOW TO DRAW A BOX AND WHISKER DIAGRAM | |
|---|---|
| STEP 1: | Make sure that the data is arranged in ascending order |
| STEP 2: | Find the five number summary |
| STEP 3: | Draw a number line long enough to fit the minimum and maximum values. Make sure that the units are plotted correctly on the number line. |
| STEP 3: | Draw vertical lines at $Q_1$, M and $Q_3$ and then draw two horizontal lines to make the box. |
| STEP 4: | From the middle of the box, first draw a horizontal line to the minimum value and then draw a horizontal line to the maximum value. |

**EXAMPLE 17**

Draw a box-and-whisker diagram of the following set of data:

506 503 507 504 510 511 526 513
517 508 515 513 508 509 516

**SOLUTION:**

**STEP 1:** Arrange the data in ascending order

503 504 506 507 508 508 509 510 511 513 513 515 516 517 526

**STEP 2:** Find the five number summary.

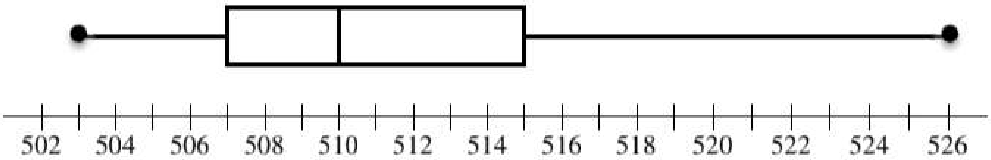
503 504 506 **507** 508 508 509 **510** 511 513 513 **515** 516 517 526

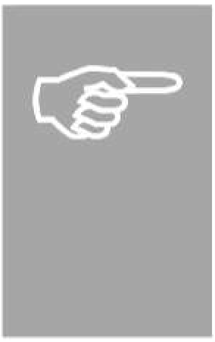
Minimum value = 503
$Q_1 = 507$
M = 510
$Q_3 = 515$
Maximum value = 526

**STEP 3:** Draw a number line long enough to go from 503 to 526.

**STEP 4:** Draw vertical lines at $Q_1$, M and $Q_3$ and form the box

**STEP 5:** Join the box to the minimum and maximum values to form the whiskers.





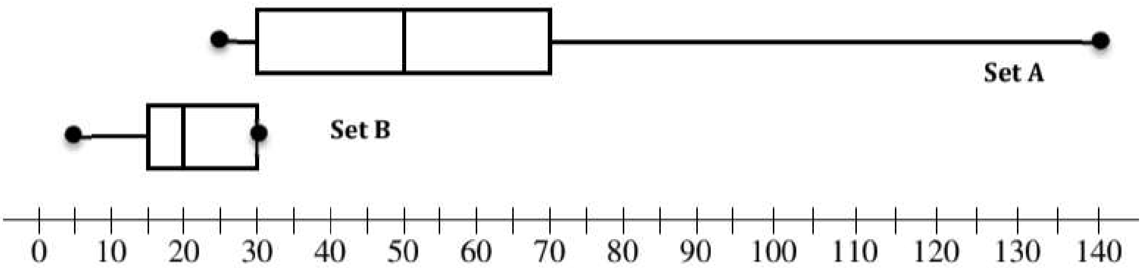
**EXAMPLE 18**

Two box and whisker diagrams are drawn on the same number line

a) List the five number summary of
   i) Set A
   ii) Set B

b) Find the range and interquartile range of
   i) Set A
   ii) Set B

c) Explain why the dot is on the edge of the box in Set B




**SOLUTION:**

a)

i) SET A:
Minimum value = 25
$Q_1 = 30$
M = 50
$Q_3 = 70$
Maximum value = 140

ii) SET B
Minimum value = 5
$Q_1 = 15$
M = 20
$Q_3 = 30$
Maximum value = 30

b)

iii) Range of Set A = maximum value – minimum value = 140 – 25 = 115

Interquartile range of Set A = $Q_3 - Q_1 = 70 - 30 = 40$

iv) Range of Set B = maximum value – minimum value = 30 – 5 = 25

Interquartile range of Set B = $Q_3 - Q_1 = 30 - 15 = 15$

c) The dot is on the edge of the box in Set B because the upper quartile and the maximum value in Set B are identical (they are both 30).

**EXERCISE 1.7**

1) The percentages achieved by a learner for a series of mathematics tests that he wrote throughout his Grade 9 year are as follows:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 35 | 45 | 50 | 28 | 39 | 49 | 55 | 35 | 56 | 49 |
| 43 | 37 | 28 | 53 | 55 | 38 | 47 | 51 | 30 | 58 |

   a) Calculate the five number summary
   b) Draw a box-and-whisker diagram to illustrate the five number summary.
   c) Find
      i) the range of the set of data
      ii) the interquartile range.

2) As part of the 2009 Census@School, the 26 Grade 10A learners measured their heights.

   The girls' heights (in centimetres) were:

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 150 | 150 | 153 | 155 | 156 | 158 | 160 | 161 | 164 | 164 | 166 | 170 | 170 |

   The boys' heights in centimetres were:

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 140 | 142 | 151 | 157 | 158 | 159 | 160 | 162 | 165 | 180 | 180 | 180 | 180 |

   a) Find the five number summary and the interquartile range for the girls and for the boys
   b) On the same number line draw two box and whisker diagrams to illustrate the girls' heights and the boys' heights.
   c) Use the five number summaries, the interquartile ranges and the box and whisker diagrams to write down two conclusions you can make about the heights of the girls and the boys.

# REFERENCES

Pike M. et al. (2011). *Classroom Mathematics Grade 10*. Heinemann.
Graham D. and Graham C. (1996). *Mainstream Mathematics for GCSE*. Macmillan
Speed B. (1997). *Higher Mathematics for GCSE*. Collins Educational.
Statistics South Africa (2010) *Census At School Results (2009).*