



SUBJECT and GRADE	Mathematics	Grade 12
TERM 3	Week 2	
TOPIC	Statistics	
AIMS OF LESSON	Revision of grade 10 and 11 statistical concepts.	
RESOURCES	Paper based resources	Digital resources
	Please go to your Statistics chapter in your textbook.	Click on the icon  in the lesson to see a video on how to use your calculator. Click on the icon  in the lesson to see a video of the concept.
INTRODUCTION: Dear learner in previous grades you have learnt about the different stages of statistical investigation. The different stages are: Collecting data, Sorting the data, Graphical representation of the data, Interpreting and Making conclusions. In this section we will investigate more techniques to manage, summaries and represent grouped - and ungrouped data.		
CONCEPTS AND SKILLS:		
Measures of central tendency (ungrouped data)		
Mean (\bar{x})	Median (Q_2)	Mode
<ul style="list-style-type: none"> The sum of all the values(x) of the data set divided by the number of values (n). $\bar{x} = \frac{\sum x}{n}$ → $\sum = \text{sum of}$ <p>https://youtu.be/HThpvMP </p>	<ul style="list-style-type: none"> The middle value of an arranged data set. Divides data in two equal sets. The position of $Q_2 = \frac{1}{2}(n + 1)$. <p>Arrange data in ascending order</p>	<ul style="list-style-type: none"> The value of the data set with the highest frequency/ most common value.

A single value that summarises the entire data set.



Example 1:

Consider the following set of marks (out of 50) for a mathematics test.

25	26	21	20	30	32	45	37	41	32
33	42	25	26	26	33	43	35	36	38

Determine the mean, median and mode.

Solution:

Mean

$$\bar{x} = \frac{\sum x}{n} = \frac{25 + 26 + 21 + 20 + 30 + 32 + 45 + 37 + 41 + 32 + 33 + 42 + 25 + 26 + 26 + 33 + 43 + 35 + 36 + 38}{20} = \frac{646}{20} = 32,3$$

Median

20, 21, 25, 25, 26, 26, 26, 30, 32, 32, 33, 33, 35, 36, 37, 38, 41, 42, 43, 45. Arrange data in ascending

Position of $Q_2 = \frac{1}{2}(n + 1) = \frac{1}{2}(20 + 1) = 10,5^{th}$ value.

$\therefore Q_2$ lies between the 10th and 11th value.

$\therefore Q_2 = \frac{32+33}{2} = 32,5$ (When the position of Q_2 falls between two values: find the sum of these values and divide it by 2)

Mode

26

CAN YOU?

Find the mean, median and mode of the data given below.

15	16	21	23	33	34	17
32	22	19	18	15	16	14

Answers:

Mean: 21,07

Median: 18,5

Mode: 15 and 16



Measures of central tendency (grouped data)

Example 2:

Find the estimated mean, median and mode of the data given below.

Class interval	Frequency	Midpoint of class interval	Frequency × Midpoint
$0 < x \leq 20$	6	10	60
$20 < x \leq 40$	8	30	240
$40 < x \leq 60$	10	50	500
$60 < x \leq 80$	3	70	210
$80 < x \leq 100$	2	90	180
	29		1190

Solution:

Mean

$$\bar{x} = \frac{\sum x}{n} = \frac{1190}{29} = 41,03$$

Median

Position of $Q_2 = \frac{1}{2}(n + 1) = \frac{1}{2}(29 + 1) = 15^{th}$ value.

The 15^{th} value falls in the $40 < x \leq 60$ class interval.

$\therefore Q_2 = 50$ (midpoint of class interval.)

Mode

$40 < x \leq 60$

Calculating the estimated mean.

1. Create additional columns **Midpoint of class interval** and **Frequency × Midpoint**.

2. Calculate the **Midpoint of class interval** and **Frequency × Midpoint**.

2.1 **Midpoint of class interval =**

$$\frac{\text{lower limit of class} + \text{upper limit of class}}{2}$$

2.2 Multiply the frequency column with the midpoint column for each class interval.

3. Calculate the sum of the **Frequency** column to find the “n” value.

4. Calculate the sum of the **Frequency × Midpoint** column to find the “ $\sum x$ ” value.

5. Substitute step 3 and 4 into $\bar{x} = \frac{\sum x}{n}$ to find the estimated mean.

<https://youtu.be/MgKljN7mpqk>



Median.

- Find the position of Q_2 .
- Determine in which class interval the position of Q_2 falls.
- Use the class midpoint as the value of Q_2 .

Mode

The class interval with the highest frequency.

<https://youtu.be/zjHfAhcU6kE>



CAN YOU?

Find the estimated mean, median and modal interval of the data given below.

Class interval	Frequency
$0 < x \leq 10$	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

Answers:

Mean 35.5

Median 35

Modal interval: $40 < x \leq 50$



Measures of dispersion

It helps us to determine how the data is spread around the mean..

Range	Quartiles	Percentiles
<p>Largest value – smallest value</p> <p>Interquartile range</p> $IQR = Q_3 - Q_1$ <p>Semi – interquartile range</p> $= \frac{Q_3 - Q_1}{2}$	<ul style="list-style-type: none"> Divides an arranged data set into quarters. Lower quartile (Q_1) Median (Q_2) Upper quartile (Q_3) The position of $Q_1 = \frac{1}{4}(n + 1)$. The position of $Q_3 = \frac{3}{4}(n + 1)$. 	<ul style="list-style-type: none"> Divides an arranged data set into 100 equal parts. The position of the k^{th} percentile = $\frac{k}{100}(n + 1)$.



<https://youtu.be/K3wsOqlq>

Example 3:

Determine the lower quartile, upper quartile, range, interquartile range, semi – interquartile range, and 80th percentile of the data below.

20, 21, 25, 25, 26, 26, 26, 30, 32, 32, 33, 33, 35, 36, 37, 38, 41, 42, 43, 45

Solution

<p>Lower quartile (Q_1)</p> <p>The position of $Q_1 = \frac{1}{4}(n + 1)$</p> $= \frac{1}{4}(20 + 1) = 5,25^{th} \text{ position.}$ $\therefore Q_1 = \frac{26 + 26}{2} = 26$	<p>Upper quartile (Q_3)</p> <p>The position of $Q_3 = \frac{3}{4}(n + 1)$</p> $= \frac{3}{4}(20 + 1) = 15,75^{th} \text{ position.}$ $\therefore Q_3 = \frac{37 + 38}{2} = 37,5$	<p>80th percentile</p> <p>The position of the 80th = $\frac{80}{100}(20 + 1) = 16,8^{th}$ position.</p> $\therefore 80^{th} \text{ percentile} = \frac{38+41}{2} = 39,5$
<p>Range = $45 - 20 = 25$</p>	<p>IQR = $Q_3 - Q_1 = 37,5 - 26 = 11,5$</p>	<p>Semi - IQR = $\frac{Q_3 - Q_1}{2} = \frac{37,5 - 26}{2} = 5,75$</p>

CAN YOU?

Determine the lower quartile, upper quartile, range, interquartile range, semi – interquartile range, and 60th percentile of the problem below. What do the lower quartiles indicate about the results of the test?

A group of 21 learners 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

Answers

Lower Quartile: 11.5 Upper Quartile: 27 Range: 41 IQR: 15.5 semi IQR : 7.75 60th percentile: 23 Q1 -25% of the results are less than 11,5



The five number summary

The five number summary consists of the following measures of dispersion:

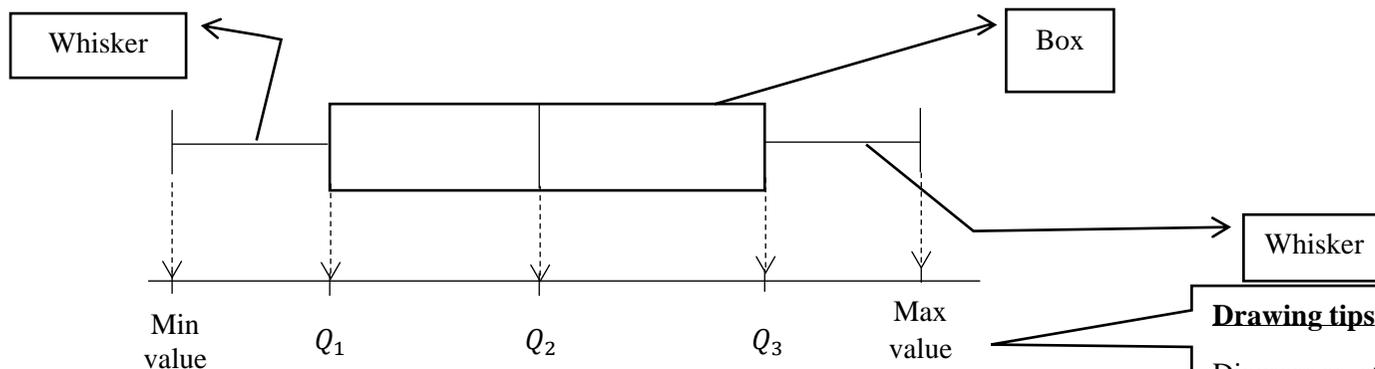
- ❖ The minimum value of the data set
- ❖ Q_1 25%
- ❖ Q_2 50% median
- ❖ Q_375%
- ❖ The maximum value of the data set

Outliers

A value that "lies outside" (is much smaller or larger than) most of the other values in a set of data.

Box and whisker diagram

The Box and Whisker diagram is a graphical representation of the five number summary.



Drawing tips
 Diagram must always have a box and whiskers.
 Clearly show five number summary on axis.

Important conclusions from box and whisker diagram.

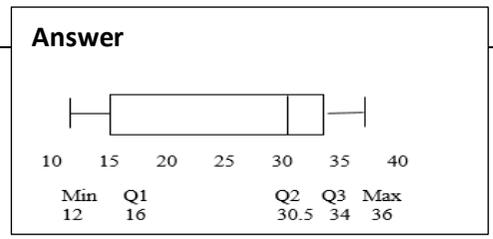
- 25% of data lies between the minimum value and Q_1 .
- 25% of data lies between Q_1 and Q_2 .
- 25% of data lies between Q_2 and Q_3 .
- 25% of data lies between Q_3 and the maximum value.

 <https://youtu.be/QI0oaZ-WTvU>

CAN YOU?

Draw a box and whisker diagram of the data below.

12, 12, 13, 14, 16, 22, 23, 24, 30, 31, 31, 31, 32, 34, 35, 35, 35, 36





Distribution of data

Positive skew/Skewed to the right

- A high volume of data around the lower values of the data set.
- The higher values of the data set are more spread out.
- Mean > Median.

Symmetrical distribution

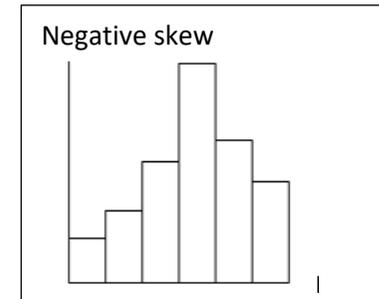
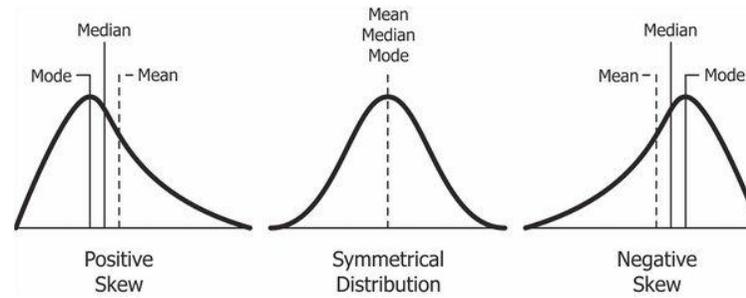
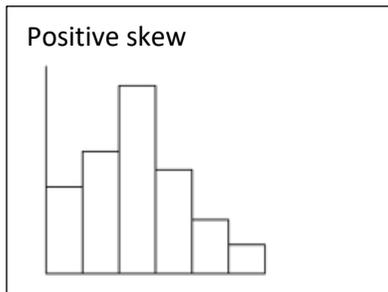
- A high volume of data situated around the mean.
- Data is symmetrical distributed around middle.
- Mean = Median.

Negative skew/Skewed to the left

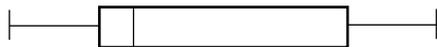
- A high volume of data around the higher values of the data set.
- The lower values of the data set are more spread out.
- Mean < Median.

Graphical representation of distribution

https://youtu.be/pfujiA5Mk_U



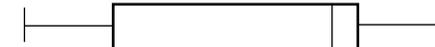
Positive skew



Symmetrical distribution



Negative skew

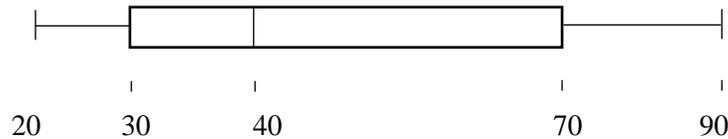


Skewness influences the mean. The more skew the data, the less the mean can be used for central tendency. If the data is skew to the left the mean is too low. If it is skew to the right the mean will be too high. The best measure of central tendency to use is the median, as this gives a better idea of what is happening to the central tendency of the data.



Example 4:

The box and whisker diagram below show the marks (out of 90) obtained by a class of nine learners.



Answer the following questions.

1. Comment on the skewness of the data.
2. Why will the mean not be a good measure of central tendency?
3. Do you think the test was too easy or too difficult? Give a reason for your answer.

Solution:

1. Skew to the right or positively skew.
2. The mean will be too high.
3. Too difficult 50% of learners had less than 40%.

CAN YOU?

The data below shows the number of people visiting a local clinic per day for 6 days.

12, 29, 13, 63, 12, 3

1. Draw a box and whisker diagram.
2. Use your calculator to determine the mean of the data
3. Calculate the mean – median
4. Comment on the skewness/ spread of the data

Answer

2. mean $\approx 14,1$
3. mean - median = $14,1 - 17,5 = -3,4$
4. The data is clustered more closely on the right and is slightly more spread out on the left. It is negatively skewed. The difference between the mean and the median is negative which also indicates that the data is negatively skewed

Cumulative frequency graph (ogive)

Definition:

The total of a frequency and all frequencies so far in a frequency distribution. It is the “running total” of frequencies.

What must I be able to do?

- Use a cumulative frequency table to draw an ogive.
- Determine values by using the ogive.
- Use a given ogive and complete a cumulative frequency table from it.
- Use the ogive to determine percentiles and quantiles.

What is important when you draw an ogive?

1. It must have a heading.
2. Mark and label axes.
3. Plot points. (Upper boundary of class ; cumulative frequency of class) **GROUND** ogive at (lower boundary of first interval ; 0)



<https://youtu.be/ETeQxOHX3xc>

When drawing the curve, you may **NOT** use a ruler. It is an S – shaped curve.



Example 5:

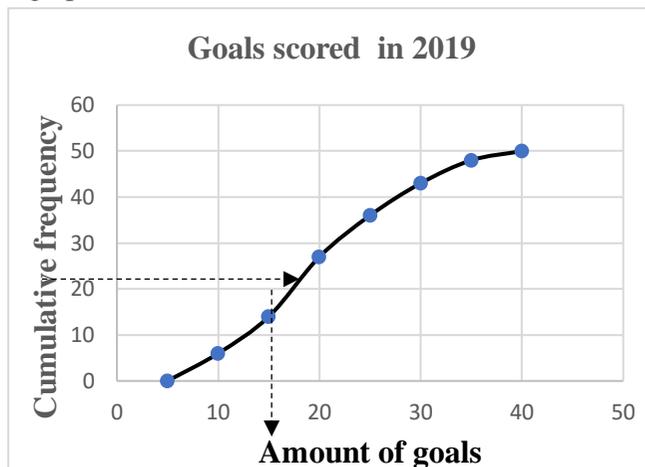
The following frequency table shows the amount of goals scored by the Protea netball team in the 50 matches they played in 2019.

Amount of goals	Frequency	Cumulative frequency	Coordinates
$5 < x \leq 10$	6	6	(10 ; 6)
$10 < x \leq 15$	8	14	(15 ; 14)
$15 < x \leq 20$	13	27	(20 ; 27)
$20 < x \leq 25$	9	36	(25 ; 36)
$25 < x \leq 30$	7	43	(30 ; 43)
$30 < x \leq 35$	5	45	(35 ; 45)
$35 < x \leq 40$	2	50	(40 ; 50)

1. Draw the cumulative frequency graph(ogive) of the given data.
2. Use the graph to determine the median.

Solution:

1.



2. Position of $Q_2 = \frac{1}{2}(n + 1) = \frac{1}{2}(50 + 1) = 25,5^{th}$ value. $\therefore Q_2 = 19$

Guidelines on drawing an ogive.

1. Create a **cumulative frequency** column.
2. Create a **coordinates** column.
3. Ground ogive at (lower boundary ; 0).
4. Plot points of coordinates column.
5. Connect points with smooth curve.

How to find quartiles and percentiles graphically.

1. Find the position. (on y-axis)
2. Find this position on cumulative frequency axis.
3. Draw a dashed line towards the ogive.
4. Draw a dashed line from the ogive towards the $x -$ axis.
5. Read the value off $x -$ axis.

CAN YOU?

A survey was done on 50 people to determine the distances they travel to work daily. The following table shows the results of the survey:

Distance in km	Frequency
$0 < d \leq 5$	2
$5 < d \leq 10$	7
$10 < d \leq 15$	4
$15 < d \leq 20$	13
$20 < d \leq 25$	16
$25 < d \leq 30$	8

1. Draw the cumulative frequency graph(ogive) of the given data.
2. Use the graph to determine the 60th percentile
3. Use your graph to estimate the median number of kilometers travelled per week.

Answers: (2) ≈ 54

(3) Median between 25th and 26th learner ≈ 50 km per week



Standard deviation and Variance

Standard deviation(σ) is a quantity expressing by how much the members of a group differ from the mean value for the group. The bigger the value of the standard deviation the further away the data lies from the mean(\bar{x}). This tells us that the mean is not the best measure of central tendency, and the median will be a better option. The smaller the value of the standard deviation, the closer the data is lying to the mean. Therefore, the mean is a reliable measurement of central tendency and can be used and trusted. The variance is the standard deviation squared(σ^2). Interval for data to be k standard deviations from the mean ($\bar{x} - k\sigma; \bar{x} + k\sigma$).

How to calculate standard deviation with your casio calculator.

1. Press mode and select STAT.
2. Select 1 – VAR.
3. Enter values. After entering each individual value press = before entering the next value.
4. After entering all the individual values press AC.
5. To find σ , press shift then STAT(at 1).
6. Then press Var to find σ .



<https://youtu.be/YYj8OMg LHSA>

Calculate standard deviation with a formula.



<https://youtu.be/laTFpp-uzp0>

Example 6:

The following sample of weekly wages of ten people was taken from a printing company.

R2 250; R2 250; R3 000; R3 300; R3 300
R3 600; R3 900; R4 350; R4 350; R5 250

1. Calculate the mean weekly wage.
2. Calculate the standard deviation of the weekly wage.
3. Determine the percentage of workers which lie within ONE standard deviation of the mean.

Solution:

1. $\bar{x} = \frac{\sum x}{n} = \frac{35550}{10} = 3555$
2. $\sigma = 900,12$
3. $(\bar{x} - k\sigma; \bar{x} + k\sigma)$
 $= (3555 - 900,12; 3555 + 900,12)$
 $= (2654,88; 4455,12)$
 $\therefore 7$
 $\therefore \frac{7}{10} \times 100 = 70\%$ of people earn within one standard deviation.

CAN YOU?

The following test results was recorded:

12; 19; 40; 33; 25; 24; 15; 38

1. Calculate the standard deviation of the test scores.
2. Determine how many test scores lies outside of one standard deviation of the mean.

Answers: 1. 9,74 2. 4

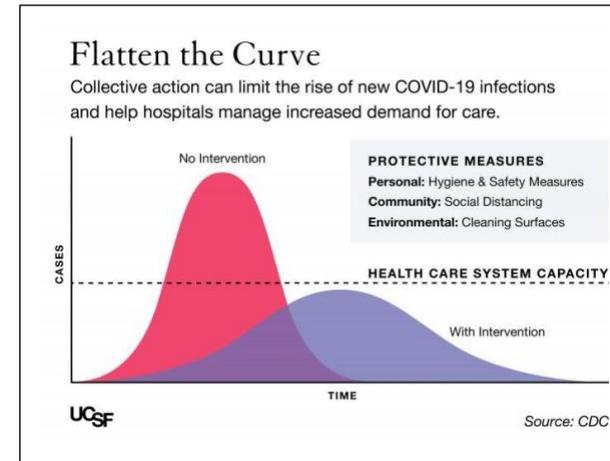


Consolidation

- Being able to draw graphs from given data and understanding statistics concepts is important but added to this...
- Analysing graphs and commenting on the distribution of data is an important skill that would assist in the interpretation of data.

Flatten the curve !!! What is the curve?

The "curve" researchers are talking about refers to the projected number of people who will contract COVID-19 over a period of time.
 (To be clear, this is not a hard prediction of how many people will definitely be infected, but a theoretical number that's used to model the virus' spread.)



ACTIVITY	Mind Action Series Revision Exercise pg. 288 – 291	Via Africa Mixed Exercise pg. 106	Classroom Mathematics Revision Exercise pg. 304 - 305	Siyavula Revision Exercise pg. 370 - 372	Platinum Revision Exercise pg. 242
-----------------	--	---	--	--	--