

# **Grade 12 – Book C** **(First edition – CAPS)**

## **TEACHER’S GUIDELINES**

### **CONTENTS:**

	<u>Page:</u>
C1. Trigonometry	3
C2. Data handling	135

This book was compiled and processed by E.J. Du Toit in 2023.

Website: [www.abcbooks.co.za](http://www.abcbooks.co.za)

Copyright © 2023. All copyrights reserved. No part of this publication may be reproduced in any form, unless written consent was obtained.

WITH SPECIAL THANKS AND ACKNOWLEDGMENT TO THE DEPARTMENT OF EDUCATION FOR THE USE OF EXTRACTS FROM OLD PAPERS.

ISBN 978-1-928336-70-9

Also visit [www.abcmathsandscience.co.za](http://www.abcmathsandscience.co.za) for extra exercise, tests and exam papers.

## Chapter C2

### Data handling

#### C2.1 Revision grade 11:

##### C2.1.1 Measurement of central tendencies:

- \* **Mode:** This is the count that occurs the **most**. With grouped data, the modal class is the class with the highest frequency.
- \* **Median:** This is the count that occurs precisely in the **middle** of the arranged data.
- \* **Mean:** The total of all the values divided by the number of entries for a specific set of data. The formula  $\bar{x} = \frac{\sum x}{n}$  with  $\bar{x}$  as the mean,  $\sum$  means the sum of,  $x$  is all the value in the data set and  $n$  is the number of entries.  
For grouped data the formula  $\bar{x} = \frac{\sum fx}{n}$  is used with  $fx$  the product of the midpoint and the frequency of each class.

##### C2.1.2 Distribution (dispersion) of data:

- \* **Range:** (Highest value – lowest value).
- \* **Median:** Divide the arranged data in two halves.
- \* **Quartiles:** Divide the arranged data in 4 quarters. Each two quarters are separated by a quartile:
  - The lower quartile ( $Q_1$ ) is between the first and second quarters.
  - The middle quartile ( $Q_2$ ) better known as the median (M) is between the second and third quarters and divide the data in two equally large sets of data.
  - The upper quartile ( $Q_3$ ) is between the third and fourth quarters.
- \* **Deciles:** Divide the arranged data in ten equal groups.  
The median is the 5<sup>th</sup> decile.
- \* **Percentiles:** Divide the arranged data in 100 equal groups.  
The median is the 50<sup>th</sup> percentile.

If there are outliers in the data, the range will not be used as a measure of dispersion of the data; then we would rather use the interquartile range (IQR). The IQR is based on the distribution of the middle half of the data around the median.

$$\therefore \text{IQR} = Q_3 - Q_1$$

### C2.1.3 Cumulative frequency:

Ex. 1 (a) Complete the following frequency table according to the pulse (per minute) of the players of the two rugby teams, after a match.  
Summarize both teams on one frequency table.

Rugby team A:	87	98	99	99	102	105	105	107	108	108	111	111
Rugby team B:	101	102	108	110	112	113	114	118	119	121	124	124
	126	126	128	129	129	130	131	133	135	135	141	

Frequency of the pulse of the rugby players:

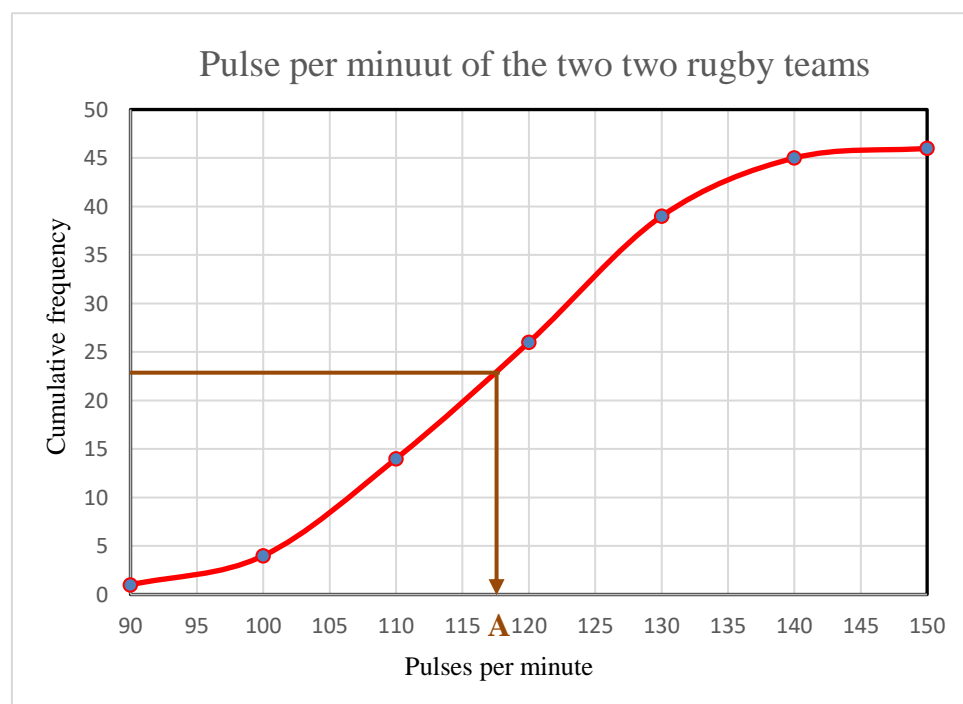
Pulse	Frequency	Cumulative frequency	
81 - 90	1	1	Start by writing down F(1)
91 - 100	3	4	Add F(3) to the 1
101 - 110	10	14 *	Add F(10) to the 4
111 - 120	11	26	Add F(10) to the 14
121 - 130	14	39	Add F(13) to the 24
131 - 140	6	45	Add F(6) to the 37
141 - 150	1	46	Add F(1) to the 43

(b) How many of the players have a pulse of less than 111 per minute?  
Use the data in (a).

14 read from \*

### C2.1.4 Ogives:

Ex. 2 (a) Use the cumulative frequency table in ex. 1 and draw an ogive of the data.  
(b) Use the ogive to determine the median. Show on the ogive where the reading were done.



To draw the ogive:

- \* The upper boundary of each class interval is indicated on the x-axis.
- \* The y-axis is divided in units to cover the cumulative frequency.
- \* Each point dotted on the graph thus consists of the upper boundary of the class interval as x-coordinate and the cumulative frequency as y-coordinate.  
Eg. (90 ; 1) and (100 ; 4).
- \* Connect the points with a free hand.
- \* Remember to label the axis!

(b) There are 46 data entries. The median will therefore lie between entries 23 and 24. Therefore, go to the cumulative entry of 23, 5 on the y-axis and from there read the median from the x-axis.

∴ Median = 117 at A

## C2.1.5 Variance and standard deviation:

### C2.1.5.1 Without calculator:

Ex.3 The number of cars sold (per month) in one calendar year by the sales consultants of a certain car dealer is as follow:

23 32 11 34 16 14 27 12 17 18 29 19

(a) Determine the monthly mean number of cars sold by the car dealer for this year.

$$\bar{x} = \frac{11 + 12 + 14 + 16 + 17 + 18 + 19 + 23 + 27 + 29 + 32 + 34}{12} = \frac{252}{12} = 21$$

(b) Complete the table below:

Number of cars sold ( $x$ )	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
11	$11 - 21 = -10$	100
12	$12 - 21 = -9$	81
14	$14 - 21 = -7$	49
16	$16 - 21 = -5$	25
17	$17 - 21 = -4$	16
18	$18 - 21 = -3$	9
19	$19 - 21 = -2$	4
23	$23 - 21 = 2$	4
27	$27 - 21 = 6$	36
29	$29 - 21 = 8$	64
32	$32 - 21 = 11$	121
34	$34 - 21 = 13$	169
	$\sum_{i=1}^n (x_i - \bar{x})^2 =$	678

(c) Determine the variance of the number of cars sold.

$$\text{Variance} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{12} = \frac{678}{12} = 56,5$$

(d) Determine the standard deviation of the number of cars sold.

$$\text{Standard deviation}(\sigma) = \sqrt{\text{Variance}} = \sqrt{56,6} \approx 7,52$$

(e) By using the standard deviation, calculated in (d), make a statement regarding the sales tactics of the consultants.

**7 of the monthly sales figures of the consultants lie within the standard deviation of 7,52 of the mean. The sales figures of the consultants lie close to the mean. The team of consultants had consistent sales for the year.**

### C2.1.5.2 With calculator:

Steps if using the CASIO fx-82 ZA Plus II scientific calculator:

\* Press Mode 2(STAT)

\* Press 1(VAR)

\* Screen:  $x$

1

2

3

→ Key in the values followed by =

\* Press AC to clear the screen

\* Press Shift 1 – STAT option appears on the screen

\* Press 4(VAR) – VAR option appears on the screen

\* Press 2( $\bar{x}$ ) = for the mean

\* Press Shift 1 – STAT option appears on the screen

\* Press 4(VAR) – VAR option appears on the screen

\* Press 3( $\sigma x$ ) = Standard deviation

Example:

21 ; 24 ; 27 ; 32

21 = 24 = 27 = 32 =

26

4,06201 ...  $\approx$  4,06

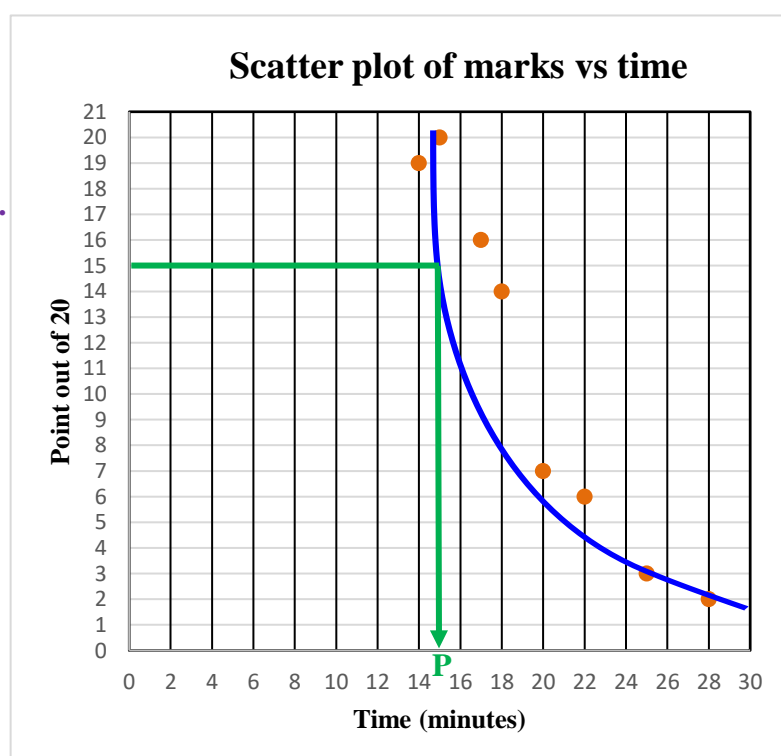
Use the data in example 3 to calculate the standard deviation ( $\sigma$ ) by using the calculator.

### C2.1.6 Scatter plots:

Ex. 4 A group of students write a test at the end of a topic. The test consists out of 20 multichoice questions. It appears as if the students who studied harder for the test, answers the test quicker. The table below shows the time and marks for each student:

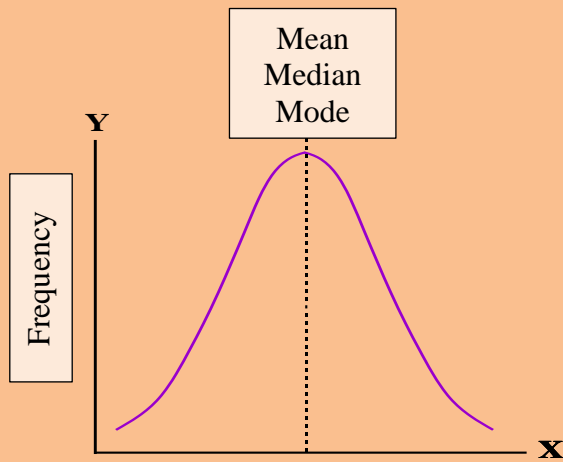
Student:	A	B	C	D	E	F	G	H
Time (in minutes):	25	20	18	22	15	17	28	14
Mark out of 20:	3	7	14	6	20	16	2	19

- (a) Draw a scatter plot to represent the information in the table.
- (b) Draw the line of best fit on the graph.
- (c) Describe the correlation between the marks and the time according to the line of best fit.
- (d) Predict the time that a student scoring 15 out of 20 will take to complete the test.
- (b) See graph in blue
- (c) Exponential correlation
- (d) At P: it can take about 15 minutes to score a of 15 out of 20.



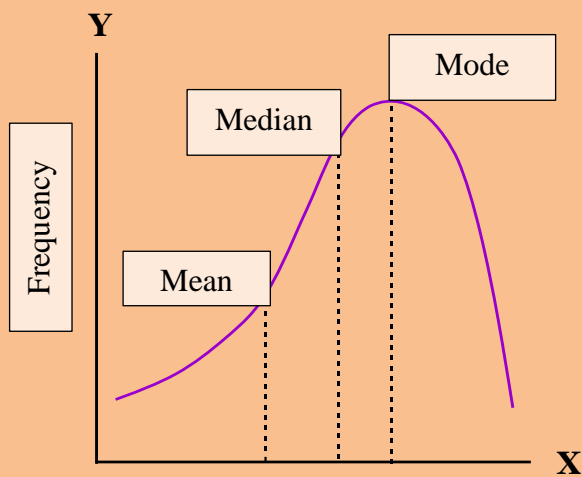
## C2.1.7 Analysis of distribution:

### Symmetrical distribution:



→  $\text{Mean} - \text{Median} = 0$

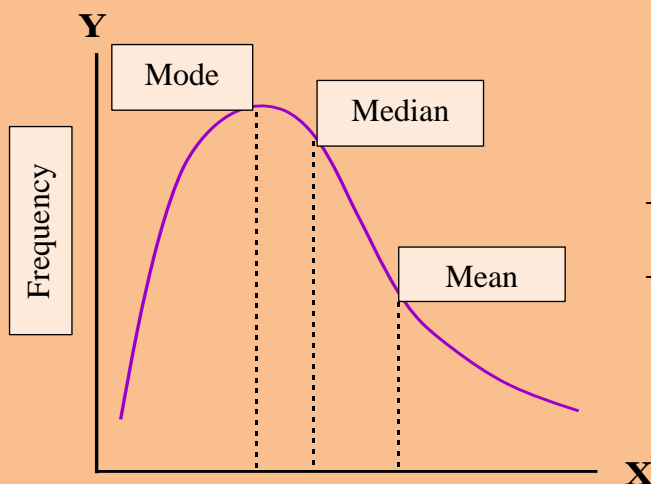
### Distribution negative skew:



→  $\text{Mean} - \text{Median} < 0$

→ **Negative skew to the left**

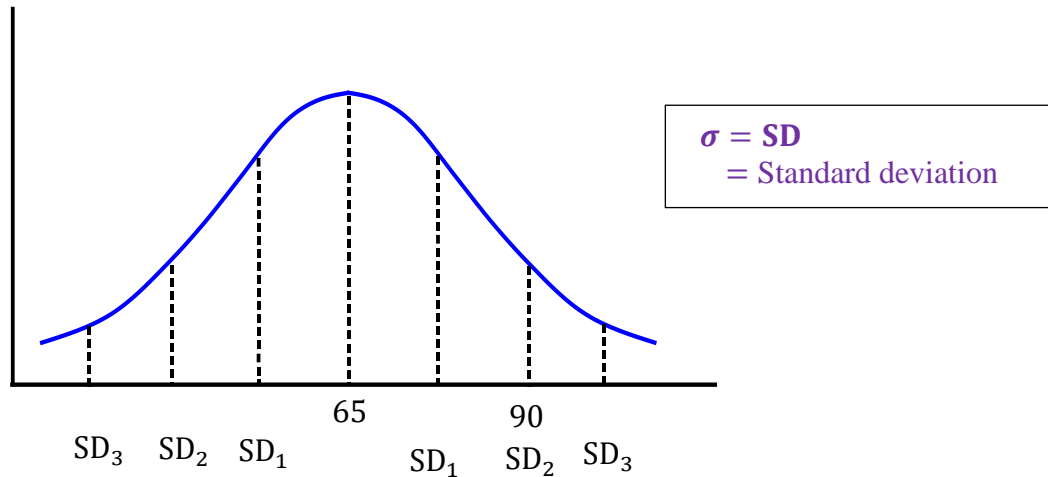
### Distribution positive skew:



→  $\text{Mean} - \text{Median} > 0$

→ **Positive skew to the right**

**Ex. 5** Two universities require a Mathematics-entrance examination for prospective first year students. Renata took the entrance examination at both universities. The normal distribution graph below represents the results for University A.



\* University B announced the following:

$$\text{Mean } (\bar{x}) = 49$$

$$\text{Standard deviation } (\sigma) = 5$$

(a) Calculate the standard deviation for the entrance examination of University A.

$$\sigma = \frac{90 - 65}{2} = 12,5$$

(b) Renata's results were as follow:

University A: 78

University B: 60

In which entrance examination did she achieve better marks in comparison to the other students who also wrote the test? Motivate your answer with relevant calculations.

University A:  $78 - 65 = 13$  - See graph!

Her results lie somewhat more than one standard deviation from the mean.

University B: See at \*!

$$\bar{x} + \sigma = 49 + 5 = 54$$

$$\bar{x} + 2\sigma = 49 + 10 = 59$$

Her results lie somewhat more than two standard deviations from the mean.

**$\therefore$  Her percentage was better at University B in comparison to the other students.**

## Exercise 1:

Date: \_\_\_\_\_

(1) Calculate the mean ( $\bar{x}$ ) and the standard deviation ( $\sigma$ ) of each of the following sets of data by using the calculator: [Give your answer correct to one decimal.]

(a) 18 ; 24 ; 36 ; 11 ; 24

$$\bar{x} = 22,6$$

$$\sigma = 8,2365 \dots \approx 8,2$$

(b) 312 ; 203 ; 254 ; 215 ; 456 ; 300

$$\bar{x} = 290$$

$$\sigma = 84,2911 \dots \approx 84,3$$

(c) 3 ; 2 ; 8 ; 10 ; 4 ; 3 ; 3 ; 4 ; 5 ; 3

$$\bar{x} = 4,5$$

$$\sigma = 2,4186 \dots \approx 2,4$$

(d) 3 380 ; 3 456 ; 3 654 ; 3 020

$$\bar{x} = 3 377,5$$

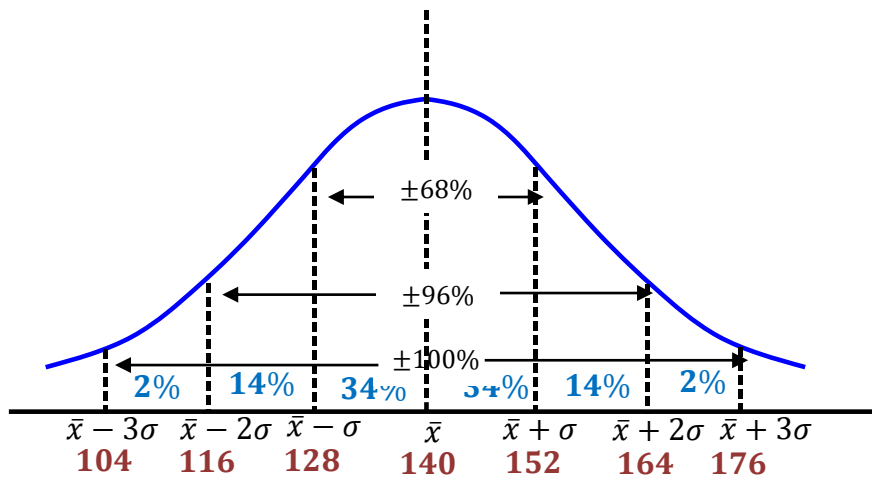
$$\sigma = 229,3616 \dots \approx 229,4$$

(e) 44 ; 32 ; 56 ; 37 ; 65 ; 33 ; 40

$$\bar{x} = 43,8571 \dots \approx 43,9$$

$$\sigma = 11,4321 \dots \approx 11,4$$

(2) The number of SMS messages sent by a group of teenagers was recorded over a period of a week. The data was found to be normally distributed with a mean of 140 messages and a standard deviation of 12 messages.



Answer the following questions with reference to the information provided in the graph:

(a) What **percentage** of teenagers sent **less than** 128 messages?  $140 - 12 = 128$

$$\therefore \text{less than 1 standard deviation left of the mean} \rightarrow 50\% - 34\% = 16\%$$

(b) What percentage of teenagers sent between 116 and 152 messages?

$$116 (140 - 2 \times 12) \text{ minutes are 2 standard deviation left of the mean} \rightarrow \therefore 48\%$$

$$152 (140 + 12) \text{ minutes are 1 standard deviation right of the mean} \rightarrow \therefore 34\%$$

$$\therefore 48\% + 34\% = 82\%$$





(3) During a cycle rally the following times (in seconds) were recorded:

86 96 71 65 86 78 69 63 94 100 72 85 86 89

(a) Determine the mode, median and mean of the data.

63 65 69 71 72 78 85 86 86 86 89 94 96 100

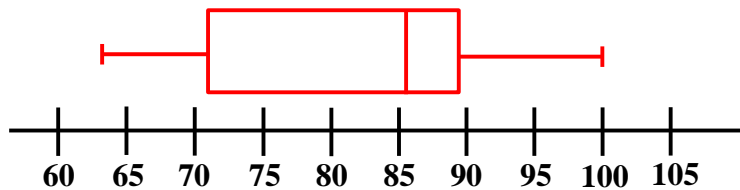
$$\text{Mode} = 86 \qquad \text{Median} = \frac{85 + 86}{2} = \frac{171}{2} = 85,5$$

$$\text{Mean} = \frac{63+65+69+71+72+78+85+86+86+86+89+94+96+100}{14} = \frac{1140}{14} \approx 81,4$$

(b) Write down the five-number summary of the data.

14 (63 ; 71 ; 85,5 ; 89 ; 100)

(c) Draw a box and whisker plot of the data.



(d) Describe the distribution of the data as in die box and whisker plot.

Distribution is skew to the left.

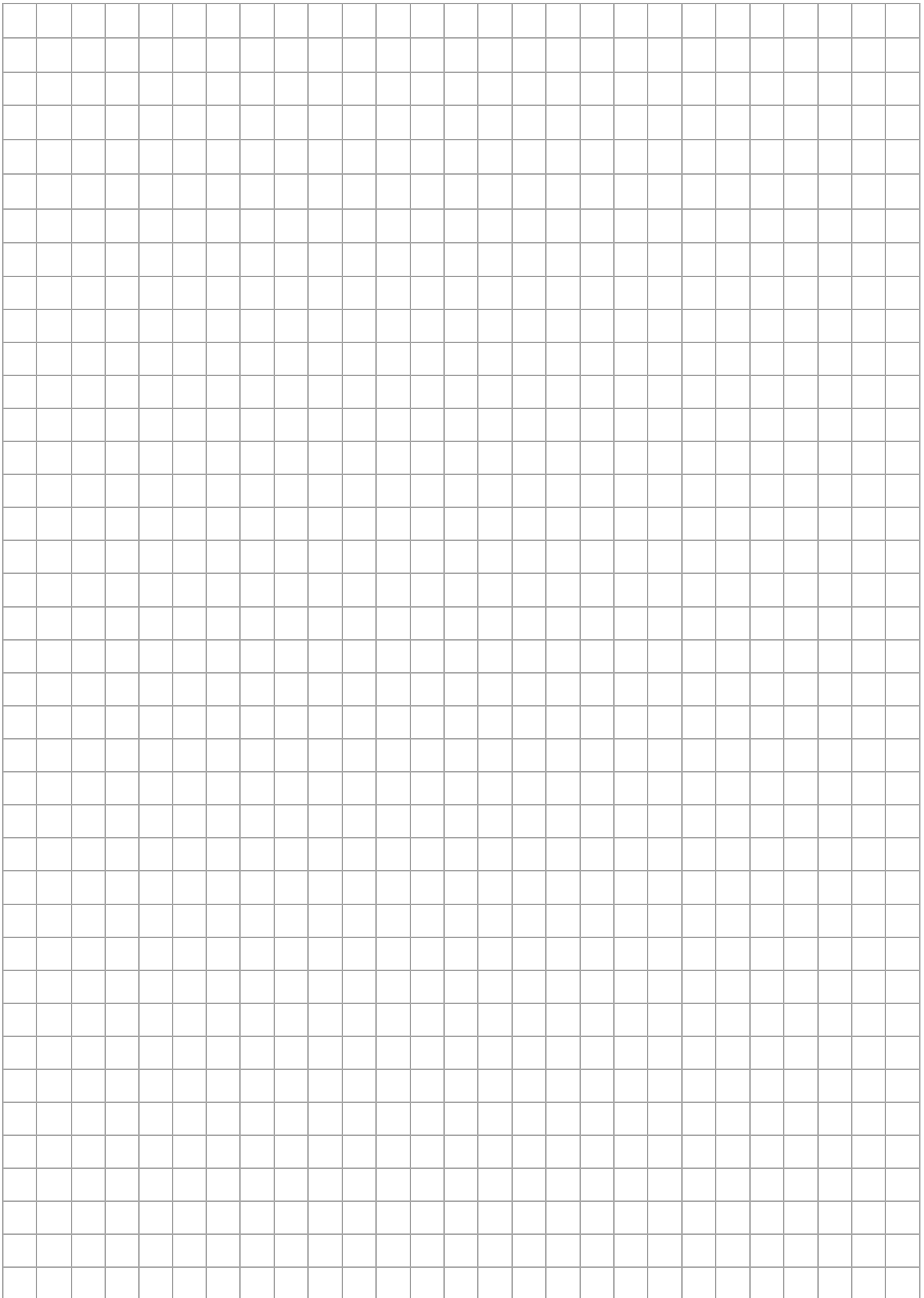
(e) Calculate the standard deviation of the time in which the rally is completed by completing the table:

Time in seconds ( $x$ )	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
63	$(63 - 81,4) = -18,4$	338,56
65	$(65 - 81,4) = -16,4$	268,96
69	$(69 - 81,4) = -12,4$	153,76
71	$(71 - 81,4) = -10,4$	108,16
72	$(72 - 81,4) = -9,4$	88,36
78	$(78 - 81,4) = -3,4$	11,56
85	$(85 - 81,4) = 3,6$	12,96
86	$(86 - 81,4) = 4,6$	21,16
86	$(86 - 81,4) = 4,6$	21,16
86	$(86 - 81,4) = 4,6$	21,16
89	$(89 - 81,4) = 7,6$	57,76
94	$(94 - 81,4) = 12,6$	158,76
96	$(96 - 81,4) = 14,6$	213,16
100	$(100 - 81,4) = 18,6$	345,96
	$\sum_{i=1}^n (x_i - \bar{x})^2 =$	1 821,44

$$\sigma = \sqrt{\frac{1821,44}{14}} = 11,4026 \dots \approx 11,4$$

(f) Which measurement describes the central tendency best?

Although the mode and median are close, the mean will best describe the central tendency.



- (4) A farmer makes use of piece workers to harvest vegetables. He aims to harvest two hectares daily. The table below shows the number of workers versus the total number of hours it takes the workers to harvest the two hectares of vegetables daily:

Day	1	2	3	4	5	6	7	8	9
Number of hours	4	12	8	6	9	10	12	4	5
Number of workers	36	13	16	24	15	14	36	40	28

- (a) Determine the median for the number of hours worked.

4 ; 4 ; 5 ; 6 ; 8 ; 9 ; 10 ; 12 ; 12

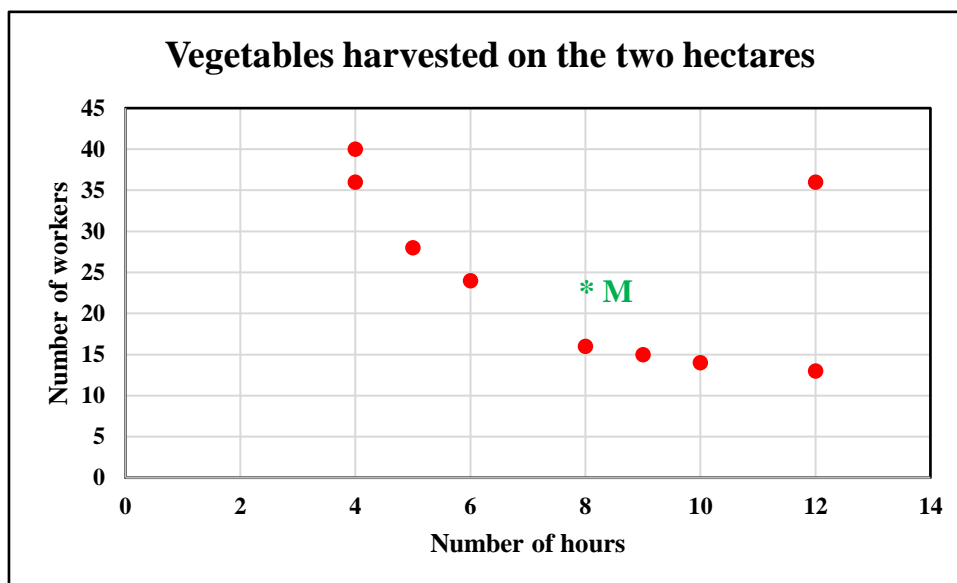
∴ Median = 8

- (b) Determine the median for the number of workers.

13 ; 14 ; 15 ; 16 ; 24 ; 28 ; 36 ; 36 ; 40

∴ Median = 24

- (c) Present the data in the table as a scatter plot. Show the median clearly on the graph. Mark the median with an M. **M(8 ; 24)**



- (d) Would the line of best fit be best described as a linear, quadratic or exponential curve for above information?

**Exponential curve will be the line of best fit – with one outlier.**

- (e) Calculate the mean number of hours per day worked by the workers for this specific period.

$$\text{Mean number of hours} = \frac{4+12+8+6+9+10+12+4+5}{9} = \frac{70}{9} = 7,7777 \dots \approx 7,8$$

- (f) Use a calculator and calculate the variance and standard deviation of the number of hours worked per day by the workers.

$$\text{Variance} = (\sigma)^2 = 9,0617 \dots \approx 9,1$$

$$\text{Standard deviation } (\sigma) = 3,0102 \dots \approx 3,0 \rightarrow \text{First do this one on the calculator and then}$$



- (5) The following table shows the number of cars, according to the make, offered by a secondhand dealer:

Make:	Number sedans:	Number bakkies:	Frequency:
Fiat	12	4	16
Ford	8	7	15
Mazda	8	10	18
Nissan	5	7	12
Opel	16	2	18
Tata	18	8	26
Toyota	3	3	6
Volkswagen	7	2	9
TOTAL	77	43	120

- (a) Complete the table.
- (b) Is this data discrete or continuous? Motivate your answer.

**Discrete → the cars are counted.**

- (c) Present the table as a pie chart. Show all calculations.

$$\text{Fiat: } \frac{16}{120} \times 360 = 48^\circ$$

$$\text{Opel: } \frac{18}{120} \times 360 = 54^\circ$$

$$\text{Ford: } \frac{15}{120} \times 360 = 45^\circ$$

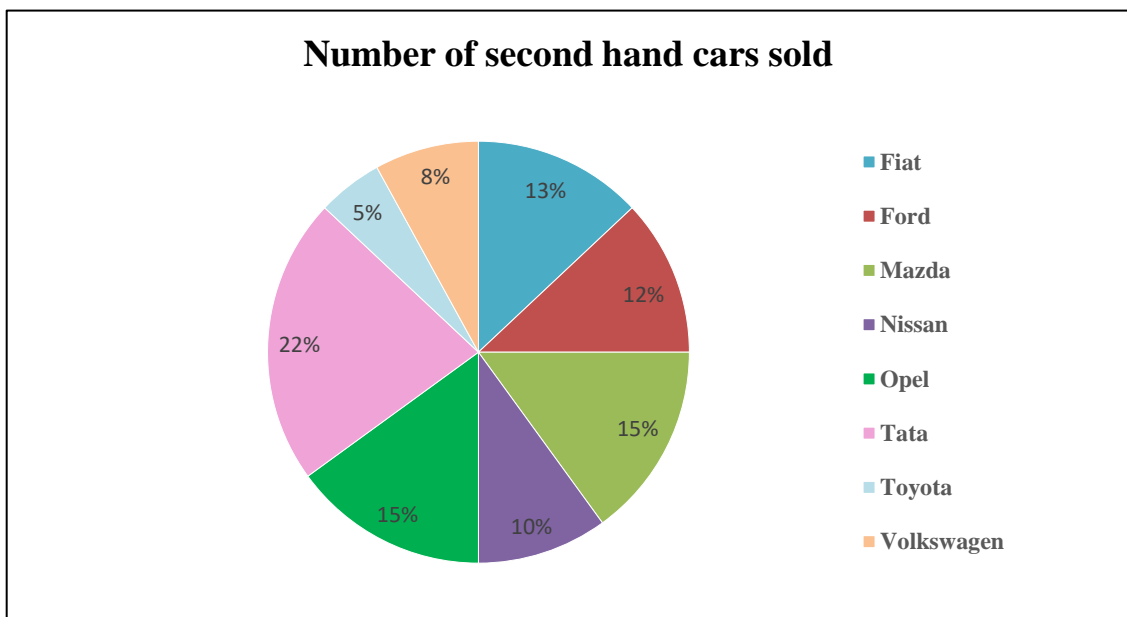
$$\text{Tata: } \frac{26}{120} \times 360 = 78^\circ$$

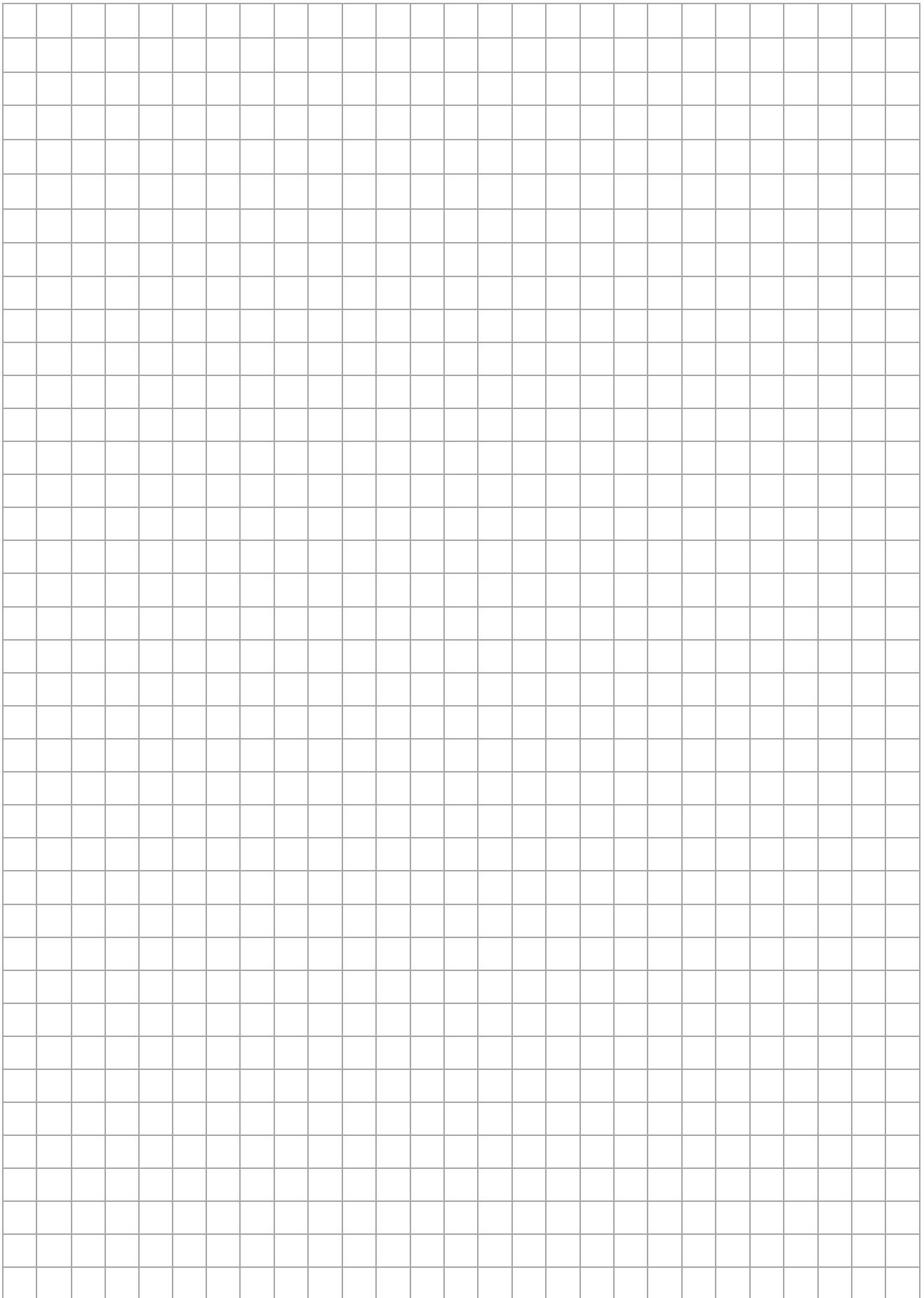
$$\text{Mazda: } \frac{18}{120} \times 360 = 54^\circ$$

$$\text{Toyota: } \frac{6}{120} \times 360 = 18^\circ$$

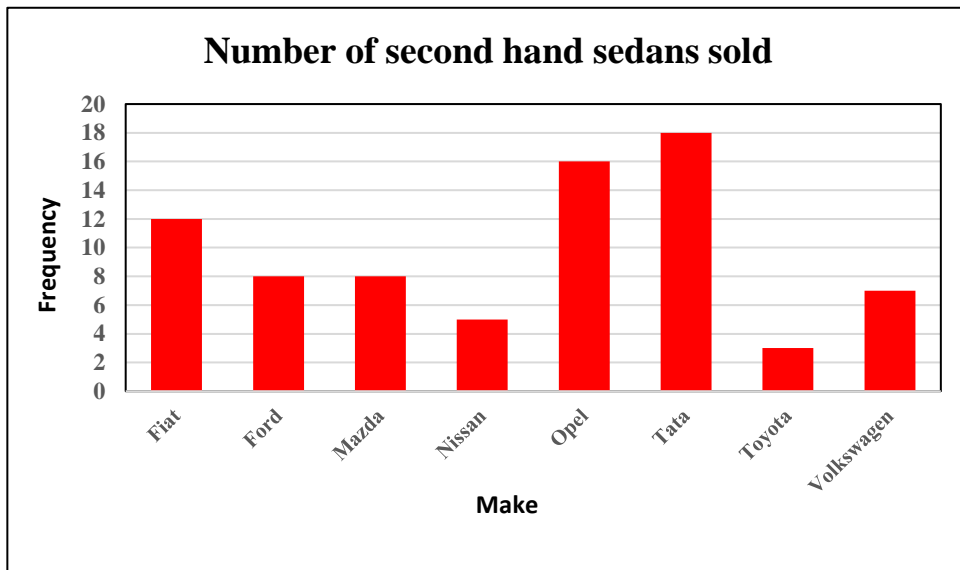
$$\text{Nissan: } \frac{12}{120} \times 360 = 36^\circ$$

$$\text{Volkswagen: } \frac{9}{120} \times 360 = 27^\circ$$

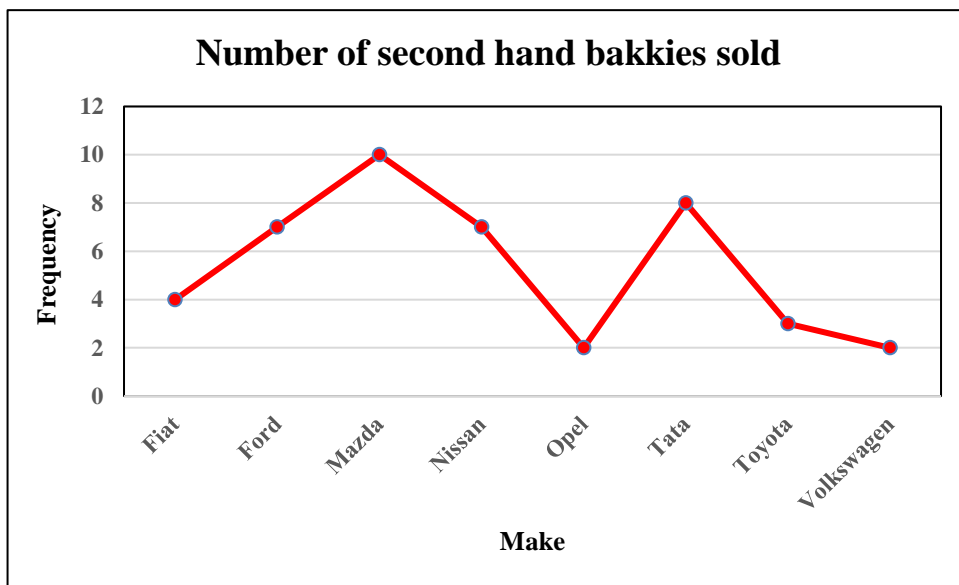




- (d) Present the the number of sedans available in the different models as a bar chart.



- (e) Present the the number of bakkies available in the different models as a line graph.



- (6) The following is the heights, to the nearest cm, of all the girls in grade 8:

141 145 176 165 156 137 169 176 155 187 156 153 161 163 156 155 161  
 150 159 177 148 155 160 145 168 149 167 169 157 171 166 156 161 169  
 170 169 163 172 188 139 169 167 155 153 149 166 178 155 168 181 177

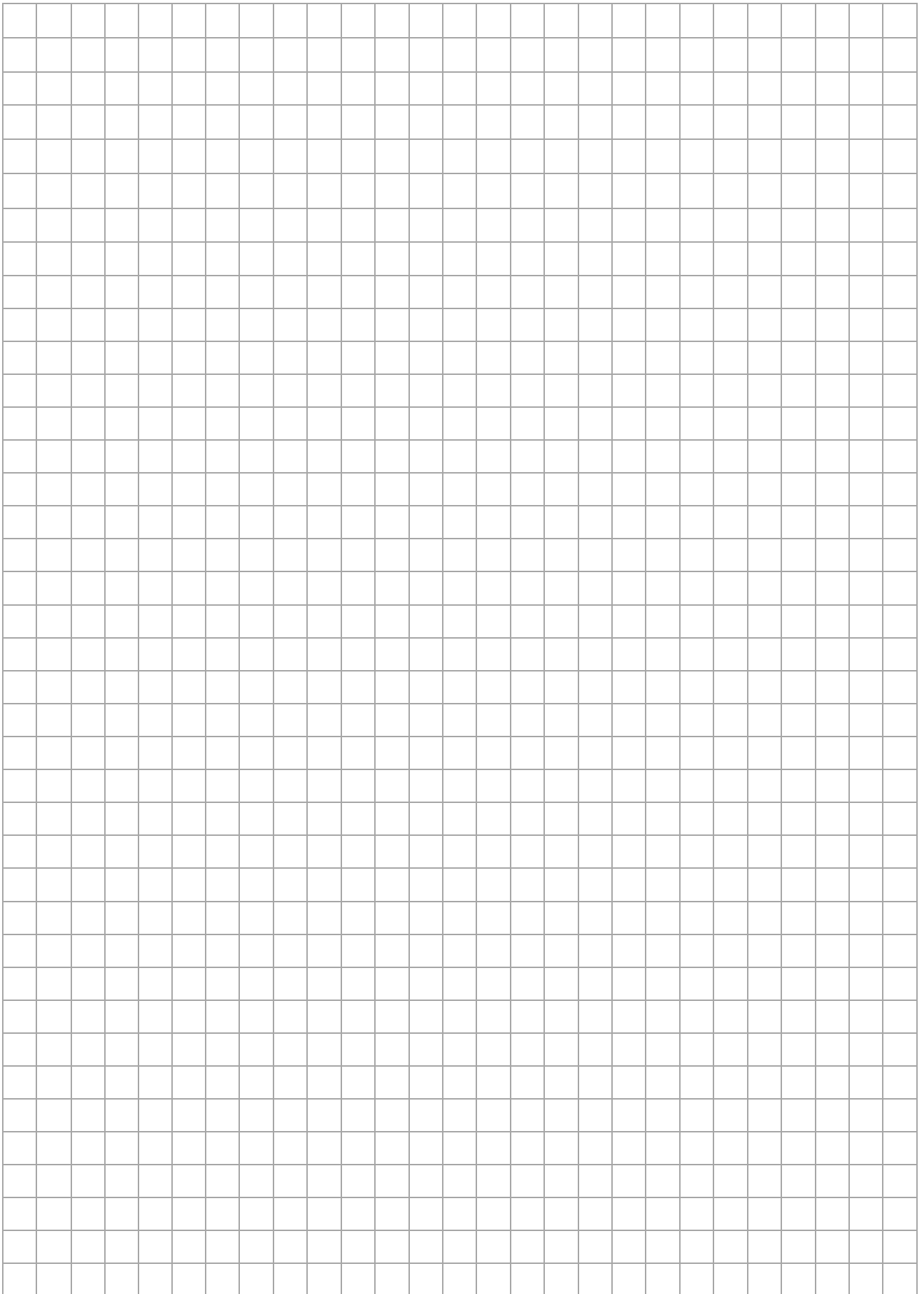
- (a) Is this data discrete or continuous? Motivate your answer.

**Continuous** → Height can be measured to a certain degree of accuracy.

- (b) Is this data quantitative or qualitative? Motivate your answer.

**Quantitative** → Data is number based, counted or measured.



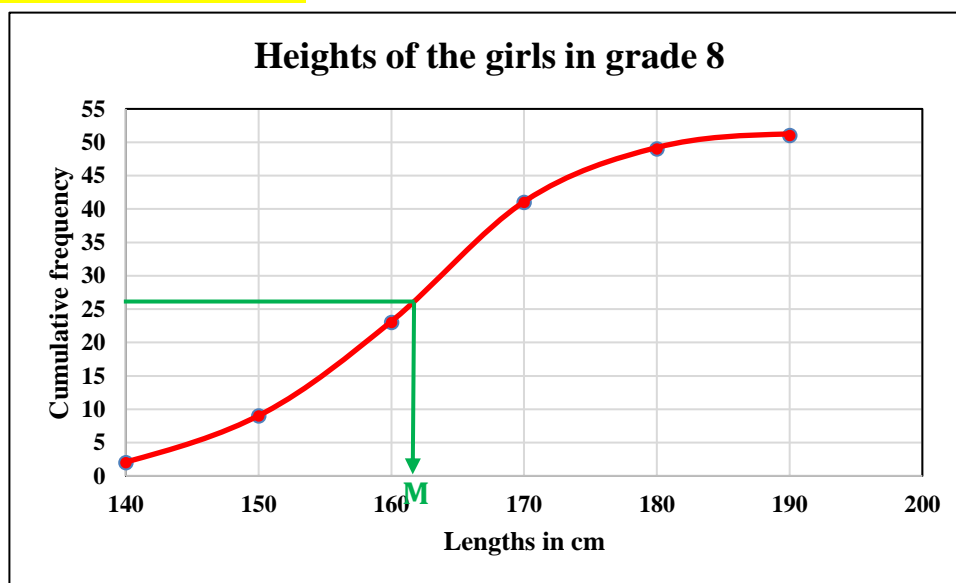


- (c) Use the data and complete the table below and draw an ogive of the data in the table:

Heights in cm:	Tally:	Frequency (F):	Cumulative F:	Points to plot:
131 – 140	//	2	2	(140 ; 2)
141 – 150	### //	7	9	(150 ; 9)
151 – 160	### ### ////	14	23	(160 ; 23)
161 – 170	### ### ### ///	18	41	(170 ; 41)
171 – 180	### //	7	49	(180 ; 49)
181 – 190	///	3	51	(190 ; 51)

- (d) Indicate on the ogive where the reading for the median will be done. What is the median?

At M → Median = 162



- (e) Determine the modal class. **Modal class → (161 – 170)**

- (f) Present the data as a stem-and-leave diagram.

<b>Heights (in cm):</b>		
<b>Stem:</b>	<b>Leaf:</b>	<b>Frequency (F):</b>
13	7 9	2
14	1 5 5 8 9 9	6
15	0 3 3 5 5 5 5 5 6 6 6 6 7 9	14
16	0 1 1 1 3 3 5 6 6 7 7 8 8 9 9 9 9 9	18
17	0 1 2 6 6 7 7 8	8
18	1 7 8	3
<b>Key: 13 – 7 represents 137</b>		<b>51</b>

- (g) Use the the stem-and-leave diagram in (f) and determine the median. Compare this median with the median determined in (d). **See in purple on the stem-and-leaf diagram.**

**Mode → 155 and 169. The one mode (169) lies within the modal class.**

- (h) Use the the stem-and-leave diagram in (f) and determine the median. Compare this median with the median determined in (d). **See in green on the stem-and-leaf diagram.**

**Median → 161. It is very close to the median as indicated on the ogive.**

## C2.2 Regression line:

To examine the relationship between two variables we use a scatter plot. The independent variable (is not influenced by other variables), is shown on the  $x$ -axis. The dependent variable (it is this variable's outcome that is examined), is shown on the  $y$ -axis. The line of best fit (as studied in previous grades) is also known as the linear regression line. If the points all lie in a perfect straight line, the line is defined by  $y = bx + a$ . If the points are spread out but have the general tendency of a straight line, the line is known as a regression line with equation  $\hat{y} = a + bx$ . To accurately calculate the equation of the regression line (thus not by inspection), we make use of the least squares method.

### C2.2.1 Regression line – without a calculator:

To calculate the equation of the regression line ( $\hat{y} = a + bx$ ), without a calculator, we proceed as follow:

- \* Calculate the mean of all the  $x$ -values ( $\bar{x}$ ) as well as the mean of all the  $y$ -values ( $\bar{y}$ ).
- \* Set up a table as in example 6 below and complete the table.
- \* Calculate the sum of  $(x - \bar{x})(y - \bar{y})$  and the sum of  $(x - \bar{x})^2$ .
- \* Substitute the values of  $\sum(x - \bar{x})(y - \bar{y})$  and  $\sum(x - \bar{x})^2$  in the gradient formula to calculate  $b$ .
- \* Use the equation,  $\hat{y} = a + bx$  to calculate the value of  $a$ .
- \* Write the equation in standard form.

**Ex. 6 Determine the equation of the regression line of the following set of data:**

$x$	<b>23</b>	<b>28</b>	<b>29</b>	<b>36</b>	<b>37</b>	<b>45</b>
$y$	<b>56</b>	<b>73</b>	<b>81</b>	<b>103</b>	<b>111</b>	<b>119</b>

$x$	$y$	$(x - \bar{x})$	$(y - \bar{y})$	$(x - \bar{x})(y - \bar{y})$	$(x - \bar{x})^2$
<b>23</b>	<b>56</b>	<b>-10</b>	<b>-34,5</b>	<b>345</b>	<b>100</b>
<b>28</b>	<b>73</b>	<b>-5</b>	<b>-17,5</b>	<b>87,5</b>	<b>25</b>
<b>29</b>	<b>81</b>	<b>-4</b>	<b>-9,5</b>	<b>38</b>	<b>16</b>
<b>36</b>	<b>103</b>	<b>3</b>	<b>12,5</b>	<b>37,5</b>	<b>9</b>
<b>37</b>	<b>111</b>	<b>4</b>	<b>20,5</b>	<b>82</b>	<b>16</b>
<b>45</b>	<b>119</b>	<b>12</b>	<b>28,5</b>	<b>342</b>	<b>144</b>
<b><math>\sum x = 198</math></b>	<b><math>\sum y = 543</math></b>	<b><math>\sum = 0</math></b>	<b><math>\sum = 0</math></b>	<b><math>\sum = 932</math></b>	<b><math>\sum = 310</math></b>
<b><math>\bar{x} = 33</math></b>	<b><math>\bar{y} = 90,5</math></b>				

$$\therefore b = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{932}{310} = 3,00645 \dots \approx 3 \text{ correct to two decimals.}$$

With  $\hat{y} = a + bx$

$$\therefore 90,5 = a + 3 \times 33 \rightarrow \text{Substitute } \bar{x} = 33 ; \bar{y} = 90,5 \text{ and } b = 3,00645 \dots \text{ in.}$$

$$\therefore 90,5 = a + 3,00645 \dots \times 33$$

$$\therefore a = -8,712 \dots \approx -8,71$$

$$\therefore \hat{y} = -8,71 + 3x$$

### **C2.2.2 Regression line – with a calculator:**

To calculate the equation of the regression line ( $\hat{y} = a + bx$ ), with a calculator (CASIO fx-82 ZA PLUS II), we proceed as follow:

- \* Press Mode → 2: STAT
- \* Press 2: A + BX
- \* Key in the  $x$ -values and press = after each  $x$ -value.
- \* Move the marker to the right and then to the top and key in the  $y$ -values. Remember the =.
- \* Press AC to clear the screen.
- \* Press Shift 1 and then 5:Reg
- \* Press 1:A = to determine  $a$ .
- \* Press AC Shift 1 then 5 and then 2:B = to determine  $b$ .
- \* Write the equation in standard form.

**Verify example 6 by using your calculator!**

### **C2.2.3 Correlation coefficient – with a calculator:**

The correlation coefficient ( $r$ ) is a measure of the strength of a linear relationship between two variables presented on a scatter plot. The correlation coefficient can vary between  $-1$  and  $1$ . The closer the correlation coefficient is to  $1$ , the stronger the correlation between the variables will be to perfectly positive and the closer the correlation coefficient is to  $-1$ , the stronger the correlation between the variables will be to perfectly negative. If the correlation coefficient is  $0$ , then there will be no correlation between the two variables.

To calculate the correlation coefficient ( $r$ ) by using a calculator (CASIO fx-82 ZA PLUS II) we proceed as follow:

[All the steps as to calculate the equation of the regression line.]

[Therefore, after calculating the regression line, directly going on to calculate the correlation coefficient – last step.]

- \* Press Mode → 2: STAT
- \* Press 2: A + BX
- \* Key in the  $x$ -values and press = after each  $x$ -value.
- \* Move the marker to the right and then to the top and key in the  $y$ -values. Remember the =.
- \* Press AC to clear the screen.
- \* Press Shift 1 and then 5:Reg
- \* Press 1:A = to determine  $a$ .
- \* Press AC Shift 1 then 5 and then 2:B = to determine  $b$ .
- \* **Press AC Shift 1 and then 5 and then 3:r = to determine  $r$ .**

**Ex. 7 Determine the correlation coefficient by using a calculator and interpret the strength of the correlation between  $x$  and  $y$  in example 6.**

**$r \approx 0,97 \rightarrow$  There is a very strong correlation between  $x$  and  $y$  in example 6.**