



SCATTER PLOT

ACTIVITY 1

- 2.2 A table of data, showing the price of crude oil at the end of each year in US Dollars (\$) (to the nearest dollar) per barrel, is given.

Year	2011	2012	2013	2014	2015	2016	2017	2019	2020	2021
Price	90	92	98	54	37	54	60	61	49	75

- 2.2.1 Determine the equation of the least squares regression line for the price of crude oil per year (3)
- 2.2.2 Calculate the value of the correlation coefficient. (1)
- 2.2.3 Use calculations to predict the price of crude oil (in US \$) per barrel at the end of 2018. Discuss the validity of the answer that you obtained by using your answer in 2.2.2. (2)

ACTIVITY 2

During the month of June patients visited a number of medical facilities for treatment.

The table shows the number of patients treated on certain dates during the month of June

Dates in the month of June	3	5	8	12	15	19	22	26
Number of patients treated	270	275	376	420	602	684	800	820

- 2.1 On **DIAGRAM SHEET 2**, draw a scatter plot of the given data. (3)
- 2.2 Determine the equation of the least squares regression line of patients treated (y) against date (x). (3)
- 2.3 Estimate how many patients have been treated on the 24th of June. (2)
- 2.4 Draw the least squares regression line on the grid on **DIAGRAM SHEET 2**. (3)
- 2.5 Calculate the correlation coefficient of the data. Comment on the strength of the relationship between the variables. (3)
- 2.6 Given that the mean for patients treated on certain dates is 528,63 calculate how many patients were within one deviation of the mean (3)





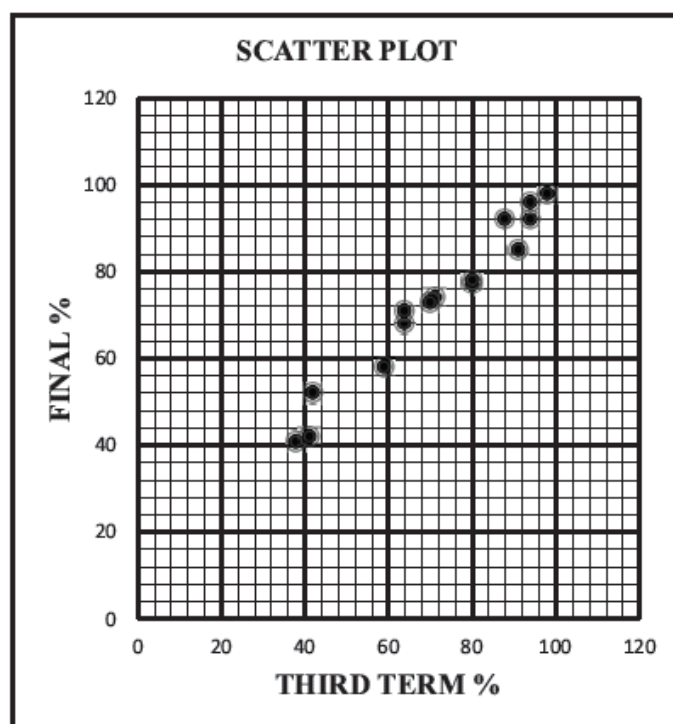
MATHEMATICS

STATISTICS

ACTIVITY 3

The table shows the percentages scored by a sample of 15 candidates in the third term and final examinations of 2022. The table and the scatter plot below represent these marks.

Third	71	80	59	38	41	98	80	88	91	94	64	94	70	42	64
Final	74	77	58	41	42	98	78	92	85	92	68	96	73	52	71



- 2.1 Determine the equation of the least squares regression line for the data, rounding off your answers to 3 decimal places. (3)
- 2.2 Write down the value of the correlation coefficient, r , between the 3rd term and final exam percentages. (1)
- 2.3 A candidate scored 48% in the third term.
- 2.3.1 Use the equation of the least squares regression line to predict his final percentage. Round your answer off to the nearest whole number. (2)
- 2.3.2 Give a reason why the prediction can be regarded as reliable. (1)
- 2.4 The least squares regression line is used to predict that the final percentage of a candidate who scored 50% in the third term is 80%.
- 2.4.1 Why would this prediction be unreliable? (1)
- 2.4.2 Would adding the point (20;10) to the original data set increase or decrease the gradient of the least squares regression line? (1)
- [9]





ACTIVITY 4

Mr Siphokazi supplements his pension by mowing lawns for customers. He measures the areas (x) (in m^2) of 12 of his customers' lawns and the time (y) in minutes, it takes him to mow these lawns. He works 8 hours a day. He recorded the data.

Area (x) (m^2)	360	120	845	602	1 190	530	245	486	350	1 005	320	250
Time (y) (minutes)	50	28	130	75	120	95	55	70	48	110	55	60

- 1.1 Determine the equation of the least squares regression line. (3)
- 1.2 Calculate the value of r , the correlation coefficient for the data. (2)
- 1.3 Given that Mr Siphokazi charges a flat call out fee of R150, as well as R50 per half hour (or part thereof), estimate the charge for mowing a customer's lawn that has an area of 560 m^2
(For example: 100 minutes would be taken as 2 hours). (3)
- 1.4 The local high school wants Mr Siphokazi to mow their rugby field which is rectangular, 100 meters long by 70 meters wide.
 - 1.4.1 Use the regression equation found in 1.1 to calculate the time it would take to mow this area. (1)
 - 1.4.2 Is it possible for him to complete this job in a day?
Give a reason for your answer. (1)

[10]





MATHEMATICS

STATISTICS

ACTIVITY 5

The following table shows a comparison of the distances (centimetres) jumped by 6 long jumpers and the hours spent practising their jumps in a week.

Long jumper	1	2	3	4	5	6
x : Hours practised	4,5	2	3,5	4	8	3
y : Distance jumped (cm)	650	420	580	490	780	525

- 2.1 Determine the equation for the least squares regression line for the data. (3)
 - 2.2 Predict the distance jumped by a long jumper who practiced for 5,4 hours. (2)
 - 2.3 Comment on the validity of your answer in QUESTION 2.2. Motivate your answer. (2)
 - 2.4 At the end of the event, they found that the measuring tape used was broken and all distances were decreased by 13 cm. How will this influence the:
 - 2.4.1 Mean jump of the event? (1)
 - 2.4.2 Range of the jumps during this event? (1)
 - 2.4.3 Standard deviation? (1)
- [10]

ACTIVITY 6

The table below shows the time spent teaching (in hours) and the average results achieved by the learners in their weekly tests as percentages (%).

TIME SPENT TEACHING (in hours)	2	8	4	6	12	10	11
AVERAGE RESULT OF 12A (as %)	42	62	48	52	64	63	67
AVERAGE RESULT OF 12B (as %)	9	63	45	47	61	64	62

- 1.1 Determine the equation of the least squares regression line of Grade 12A. (3)
 - 1.2 Write down the correlation coefficient of the Grade 12A results in respect of the time spent teaching. (1)
 - 1.3 Comment on the correlation between the time spent teaching and the average result of Grade 12A. (1)
 - 1.4 The equation of the least squares regression line of Grade 12B is $y = 15,74 + 4,54x$. Calculate the difference in the result achieved by each class had Mrs Molefe spent the average time to complete a particular section. (3)
 - 1.5 Identify an outlier in Grade 12B. (1)
 - 1.6 Indicate a valid reason for this outlier. (1)
- [10]



MATHEMATICS

STATISTICS

ACTIVITY 7

The goal-scorers in a netball game practice scoring at training during the week. In the tournament during the weekend, the number of goals they score from the total number of attempts they made, is recorded as a percentage. This statistic is referred to as the successful goal-shoot average.

The table below shows the number of goal shots practiced during the week and the successful goal-shoot average during the tournament for 8 goal-scorers.

Number of goal shots practiced	280	400	540	595	375	430	500	650
Successful goal shoot average (%)	73	75	83	89	80	76	82	91

- 2.1 Determine the equation of the least squares regression line. (3)
 - 2.2 Calculate the correlation coefficient for the data. (2)
 - 2.3 Comment on the correlation between the number of goal shots practiced and the successful goal-shoot average. (2)
 - 2.4 A player practiced 465 goal shots. What is her expected successful goal-shoot average for the next tournament? (2)
- [9]

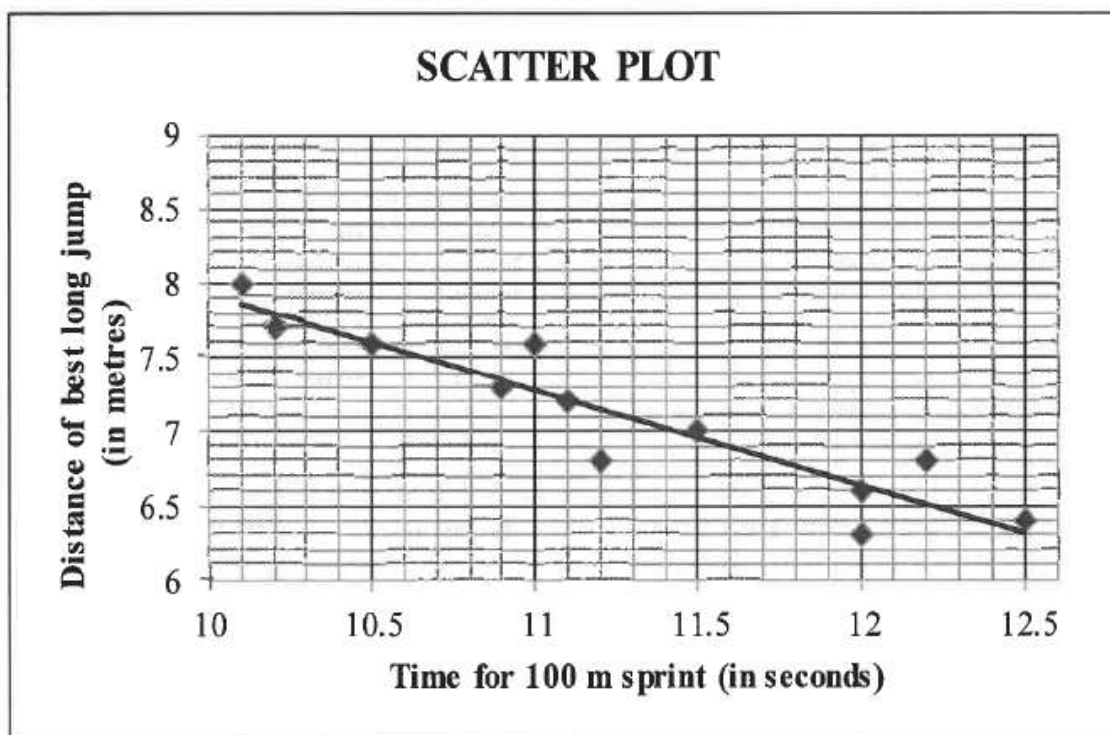


ACTIVITY 8

The table below shows the time (in seconds, rounded to ONE decimal place) taken by 12 athletes to run the 100 metre sprint and the distance (in metres, rounded to ONE decimal place) of their best long jump.

Time for 100 m sprint (in seconds)	10,1	10,2	10,5	10,9	11	11,1	11,2	11,5	12	12	12,2	12,5
Distance of best long jump (in metres)	8	7,7	7,6	7,3	7,6	7,2	6,8	7	6,6	6,3	6,8	6,4

The scatter plot representing the data above is given below.



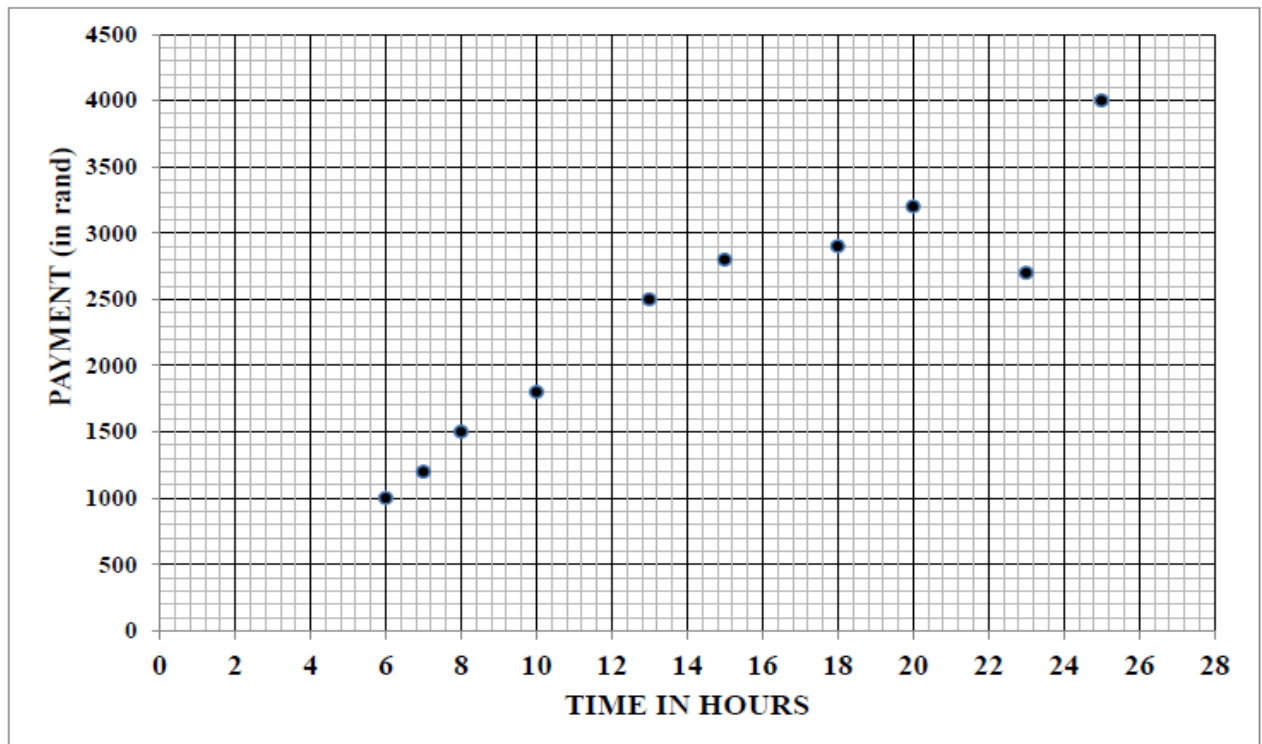
The equation of the least squares regression line is $\hat{y} = a + bx$.

- 1.1 Determine the values of a and b . (3)
 - 1.2 An athlete runs the 100 metre sprint in 11,7 seconds. Use $\hat{y} = a + bx$ to predict the distance of the best long jump of this athlete. (2)
 - 1.3 Another athlete completes the 100 metre sprint in 12,3 seconds and the distance of his best long jump is 7,6 metres. If this is included in the data, will the gradient of the least squares regression line increase or decrease? Motivate your answer without any further calculations. (2)
- [7]

ACTIVITY 9

A group of students did some part-time work for a company. The number of hours that the students worked and the payment (in rand) received for the work done is shown in the table below. The scatter plot is drawn for the data.

Number of hours worked	6	7	8	10	13	15	18	20	23	25
Payment (in rand)	1000	1200	1500	1800	2500	2800	2900	3200	2700	4000



- 2.1 Calculate the standard deviation of the number of hours worked. (1)
 - 2.2 Determine the number of hours that a student needed to work in order to receive a payment that was more than one standard deviation above the mean. (3)
 - 2.3 Determine the equation of the least squares regression line of the data. (3)
 - 2.4 Mapula who worked for 11,5 hours was omitted from the original data. Calculate the possible amount that the company has to pay Mapula. (2)
 - 2.5 Use the scatter plot to identify an outlier and give a possible reason for this point to be an outlier. (2)
- [11]



MATHEMATICS

STATISTICS

ACTIVITY 10

Ten students trained to take part in the 100m sprint event for the Mayoral Athletics Competition. The following data table shows the number of days each student trained and the time he/she ran.

Number of days training	10	70	50	60	20	60	100	30	90	30
Time (in seconds)	17	13	13	11	17	19	10	17	12	14

- 2.1 Calculate the MEAN and STANDARD DEVIATION of the times of the students. (4)
 - 2.2 Calculate an equation for the least squares regression line for the data. (3)
 - 2.3 Calculate the correlation coefficient. (1)
 - 2.4 If Simon trained for 80 days, estimate the time it might take him to finish. (2)
- [10]**

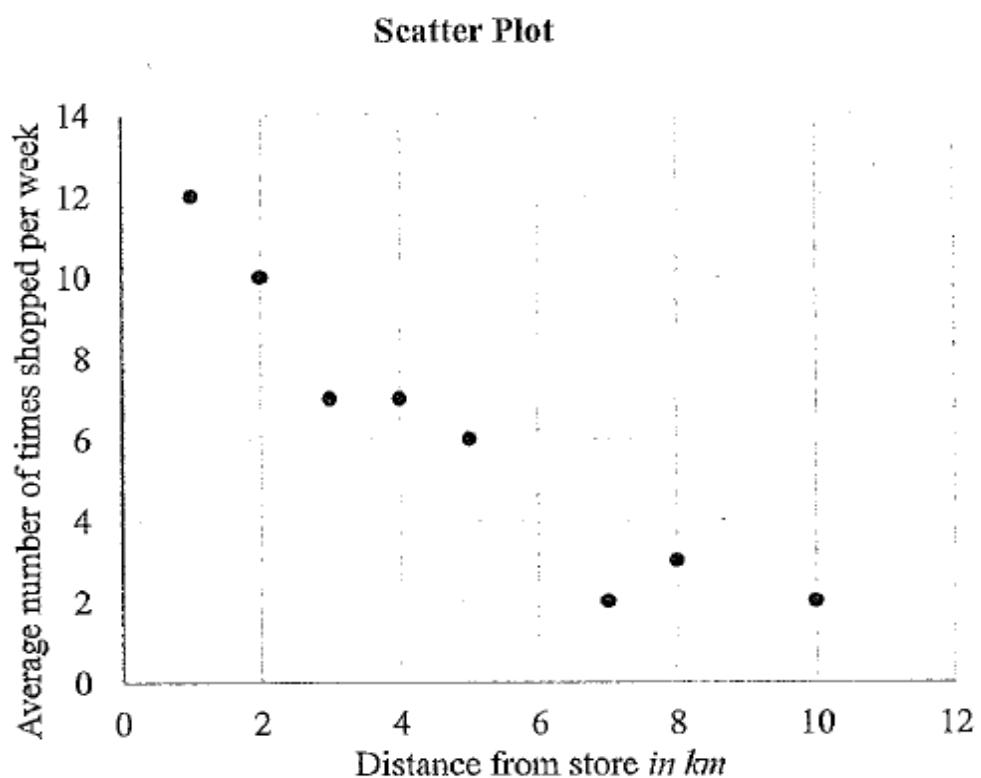


ACTIVITY 11

A survey was conducted at a local supermarket to establish the relationship between the distance (in kilometres) that shoppers stay from the store and the average number of times that they shop at the store in a week. The results are shown in a table below.

Distance from store in km	1	2	3	4	5	7	8	10
Average number of times shopped per week	12	10	7	7	6	2	3	2

The above data is represented in a scatter plot below.



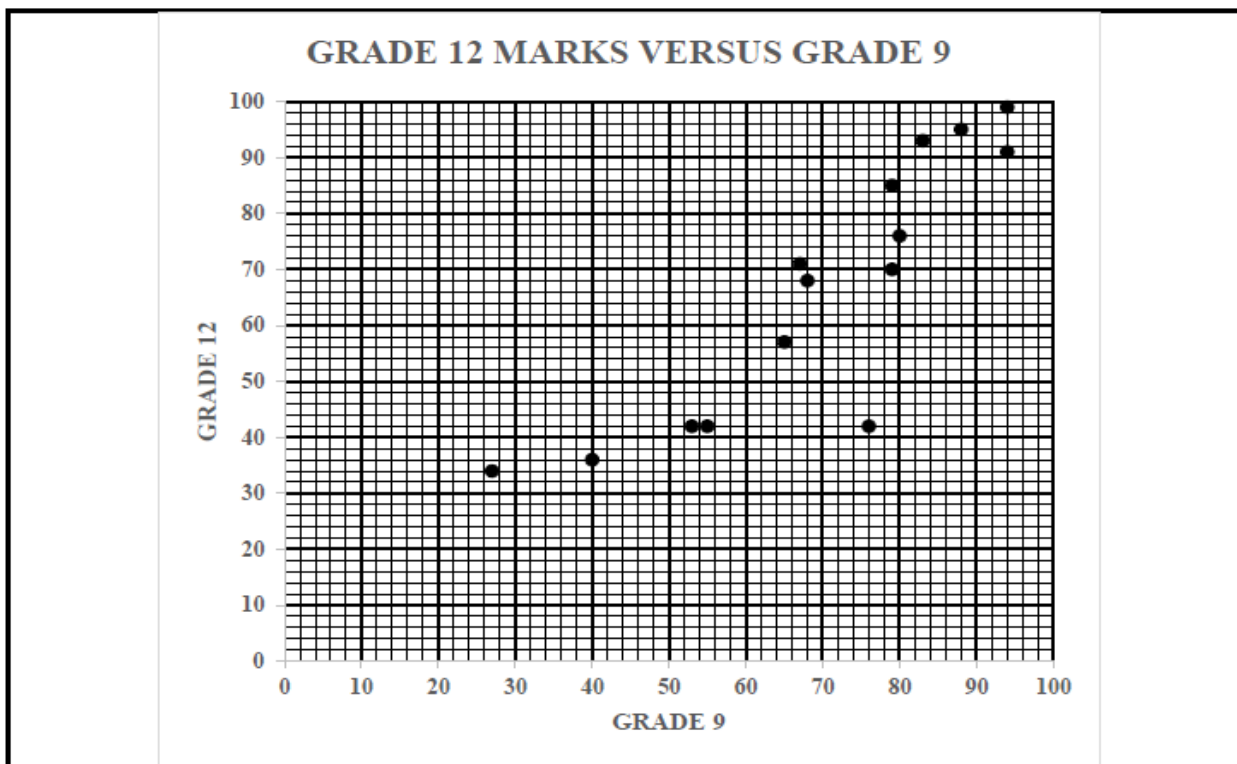
- 2.1 Calculate the correlation coefficient of the data. (2)
- 2.2 Comment on the strength of the relationship between the distance a shopper lived from the store and the average number of times she/he shopped at the store in a week. (1)
- 2.3 Calculate the equation of the least squares regression line of the data. (3)
- 2.4 Use your answer in QUESTION 2.3 to predict the average number of times that a shopper living 6 km from the supermarket will visit the store in a week. (2)

[8]

ACTIVITY 12

Every year the Grade 9s have to choose whether to take Mathematics or Mathematical Literacy. 15 learners were selected at random and their marks as a percentage for their Grade 9 report and their Grade 12 final exam are recorded in a table.

Grade 9	27	40	53	55	65	67	68	76	79	79	80	83	88	94	94
Grade 12	34	36	42	42	57	71	68	42	70	85	76	93	95	99	91



- 1.1 Determine the equation of the least squares regression line. (3)
 - 1.2 If a learner in Grade 9 final exams obtains 41%, estimate what mark in percentage did he get in Grade 12. [Round off your answer to the nearest integer] (2)
 - 1.3 Draw the least squares regression line on the scatter plot (2)
 - 1.4 Calculate the correlation coefficient. (2)
 - 1.5 Is the Grade 9 report mark a good predictor of the final matric mark? Motivate your answer. (2)
- [11]**



MATHEMATICS

STATISTICS

ACTIVITY 13

The data below shows the marks obtained by ten Grade 12 learners from two different Mathematics classes.

Class A	16	36	20	38	40	30	35	22	40	24
Class B	45	70	44	56	60	48	75	60	63	38

- 2.1 Make use of the grid provided in the ANSWER BOOK to draw a scatter plot for the data. (2)
- 2.2 Calculate the equation for the least squares regression line for this data. (3)
- 2.3 Draw the least squares regression line for the data on the scatter plot drawn in QUESTION 2.1 (2)
- 2.4 Describe the relationship between the classes. (2)
- 2.5 Calculate the mean and standard deviation for Class B (4)
- 2.6 How many scores in Class B fall within one standard deviation of the mean? (2)

[15]



ACTIVITY 14

A training company wants to know if there is a link between the hours spent in training by a particular category of employee and their productivity (units produced per day). The following data was extracted from files of 10 employees.

Employee	1	2	3	4	5	6	7	8	9	10
Hours in training	16	36	20	38	40	30	35	22	40	24
Productivity (units produced per day)	45	70	44	56	60	48	75	60	63	38

- 1.1 Draw a scatter plot of the data on the grid provided in the ANSWER BOOK. (3)
- 1.2 Determine the equation of the least squares regression line. (3)
- 1.3 Draw the least squares regression line on the scatter plot drawn in QUESTION 1.1. (2)
- 1.4 Use the least squares regression line to predict the productivity (units produced per day), if a particular category of employee spent 25 hours in training. (2)
- 1.5 Determine the correlation coefficient of the data. (1)
- 1.6 Comment on the strength of relationship between the hours spent in training and the number of units produced per day. (1)

[12]





MATHEMATICS

STATISTICS

ACTIVITY 15

A Mathematics teacher wants to create a model by which she can predict a learner's final marks. She decided to use her 2015 results to create the model.

Preparatory exam (x)	55	35	67	85	91	48	78	72	15	75	69	37
Final exam (y)	57	50	74	80	92	50	80	81	23	80	75	42

- 1.1 Determine the equation of the least squares regression line in the form $y = a + bx$. (3)
 - 1.2 Use the equation of the regression line to predict the final exam mark for a learner who attained 46% in the preparatory examination. (2)
 - 1.3 Could you use this equation to estimate the preparatory exam mark for a learner who attained 73% in the final exam? Give a reason for your answer. (2)
 - 1.4 Show that the point $(\bar{x}; \bar{y})$ lies on the regression line. (4)
 - 1.5 Determine the correlation coefficient of the data. (1)
 - 1.6 Describe the correlation between preparatory and final exam results. (2)
- [14]**





MATHEMATICS

STATISTICS

ACTIVITY 16

A company called Math π rates sent out a number of advertising pamphlets. The table below shows the number of pamphlets sent out to a certain place. The table also shows the orders that were received after the distribution of the pamphlets.

Number of Pamphlets	Number of Orders
600	350
1000	550
500	300
700	300
600	330
100	200
500	350
800	450
300	250
900	500

- 2.1 Represent the above information in a scatter plot on the grid provided in the ANSWER BOOK. (2)
- 2.2 Determine, using your calculator, the equation of the least squares regression line. (2)
- 2.3 Draw the least squares regression line on the scatter plot drawn for QUESTION 2.1. (2)
- 2.4 Calculate the possible number of orders if 200 pamphlets were sent out. (2)

[8]





MATHEMATICS

STATISTICS

ACTIVITY 17

The term latitude refers to how far a location is from the equator. Latitudes in the Northern Hemisphere range from 0° at the equator to 90° N at the North Pole.

Below are the latitudes of several cities in the Northern Hemisphere along with the average maximum temperature for April in degrees Celsius.

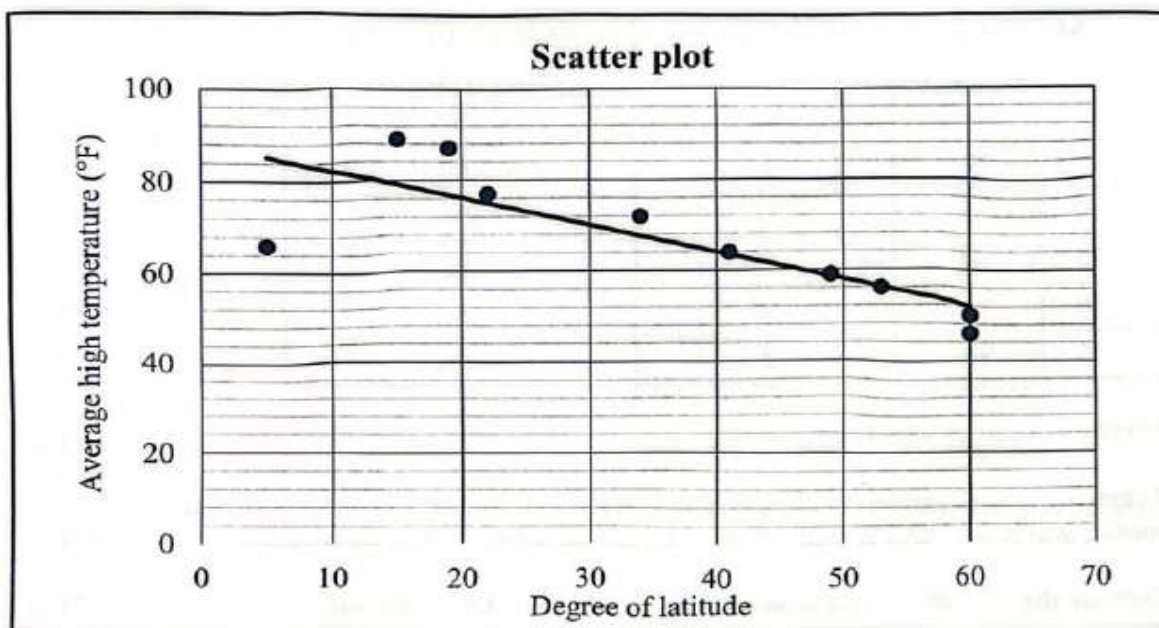
CITY	LATITUDE (degrees)	AVERAGE MAXIMUM TEMPERATURES FOR APRIL ($^\circ\text{C}$)
Lagos, Nigeria	6	32
London, England	52	13
Calcutta, India	23	36
Rome, Italië	42	20
Moscow, Russia	56	8
Cairo, Egipt	30	28
San Juan, Puerto Rico	18	29

- 2.1 Draw a scatter plot for the information above on the diagram provided in the answer book. (2)
 - 2.2 Determine the equation of the least squares regression line. (3)
 - 2.3 The city of Madrid has a latitude of 40 degrees. Determine the estimated average maximum temperature for April for this city. (2)
 - 2.4 Calculate the correlation coefficient of the data and describe the correlation between latitude and the average maximum temperature for April. (3)
- [10]



ACTIVITY 18

The scatter plot below shows the degrees of latitude of 10 cities in the Northern Hemisphere and their average high temperatures (in °F). The least squares regression line is also shown.



- 1.1 Use the scatter plot to describe the strength of the correlation between the latitude of a city and its average high temperature. Motivate your answer. (2)

The table below shows the data that was used to draw the scatter plot above.

Latitude in degrees	5	19	34	53	22	41	60	15	60	49
Average high temperature (in °F)	66	87	72	56	77	64	46	89	50	59

- 1.2 Determine the equation of the least squares regression line for the data. (3)
- 1.3 Predict the average high temperature of a city with a latitude of 28 degrees. (2)
- 1.4 Write down the standard deviation of the average high temperature of the 10 cities. (1)
- 1.5 Determine the number of cities of which the average high temperature is greater than one standard deviation above the mean. (3)
- [11]**

ACTIVITY 19

Eight couples entered a dance competition. Two judges gave marks for their performances. The marks (out of 20) are shown in the table below.

COUPLE	1	2	3	4	5	6	7	8
JUDGE 1	18	4	6	8	5	12	10	14
JUDGE 2	15	6	3	5	5	14	8	15

- 2.1 Determine the equation of the least squares regression line of the marks given by the two judges. (3)
 - 2.2 A ninth couple had a late entry in the competition and received a mark of 15 by JUDGE 1. Calculate to the nearest integer, the mark they possibly would have received from JUDGE 2. (2)
 - 2.3 Are the judges consistent when allocating marks?
Motivate your answer with appropriate calculations. (3)
- [8]**

ACTIVITY 20

10 girls compete to see who can throw a ball the furthest. They want to see whether there is a relationship between their height and the distance which they can throw. The following table shows the height of each girl (x cm) to the nearest cm and the distance (y m) she can throw the ball.

	A	B	C	D	E	F	G	H	I	J
x	122	124	133	138	144	156	158	161	164	168
y	41	38	56	56	29	54	59	61	68	67

- 2.1 Determine the equation for the least squares regression line. (3)
 - 2.2 Estimate the distance (to the nearest metre) which a ball can be thrown by a girl who is 150 cm tall. (2)
 - 2.3 Describe the trend of the data above. (1)
 - 2.4 Do you think this trend can continue indefinitely? Give a valid reason for your answer. (2)
- [8]**



MATHEMATICS

STATISTICS

ACTIVITY 21

A Mathematics teacher wants to create a model by which she can predict a learner's final marks. She decided to use her 2015 results to create the model.

Preparatory exam(x)	55	35	67	85	91	48	78	72	15	75	69	37
Final exam(y)	57	50	74	80	92	50	80	81	23	80	75	42

- 1.1 Determine the equation of the least squares regression line in the form $y = a + bx$. (3)
- 1.2 Draw a scatter plot and show the regression line (3)
- 1.3 Predict the final mark for a learner who attained 46% in the preparatory examination. (2)
- 1.4 Determine the correlation coefficient of the data. (1)
- 1.5 Describe the relationship between preparatory and final exam results. (1)
- 1.6 Could you use this equation to estimate the preparatory exam mark for a learner who attained 73% in the final exam? Give a reason for your answer. (2)

[12]



ACTIVITY 22

The leg strength and the number of leg presses done per day by a random sample of ten, eighteen-year-old boys were recorded as shown in the table below.

Number of leg presses done per day (x)	36	136	51	126	90	43	77	68	103	124
Strength of upper leg (y)	0.2	0.85	0.35	0.91	0.73	0.34	0.61	0.59	0.78	0.90

- 2.1 Determine the equation of the regression line and write your answer in the form. $y = \dots$ (3)
- 2.2 Use the regression line to predict the leg strength of the eighteen-year-old boy if he does 110 leg-presses per day. (2)
- 2.3 An eighteen-year-old boy does 250 leg presses per day. Can your regression line formula predict the strength of his legs? (Explain your answer). (2)
- 2.4 Calculate the correlation coefficient of the above data set. (2)
- 2.5 Use your answer in 2.4 to describe the relationship between the number of leg presses and the leg strength of an eighteen-year-old boy. (2)

[11]

