

Student Name

Mathematics Stage 5 Core

DATA ANALYSIS B

Book 2 Represent datasets involving
2 numerical variables

Version: 260207

Report mistakes & suggest improvements:
<https://MrDingMaths.com>

CONTENTS

Scatterplots	2
Line of Best Fit	7

OVERVIEW

SYLLABUS CONTENT

MA5-DAT-C-02 displays and interprets datasets involving bivariate data

Represent datasets involving 2 numerical variables, using a scatter plot and a line of best fit, by eye

- Gather data on a topic of interest involving 2 numerical variables
- Represent the data using a scatter plot
- Create a line of best fit, by eye, on an existing scatter plot

SCATTERPLOTS

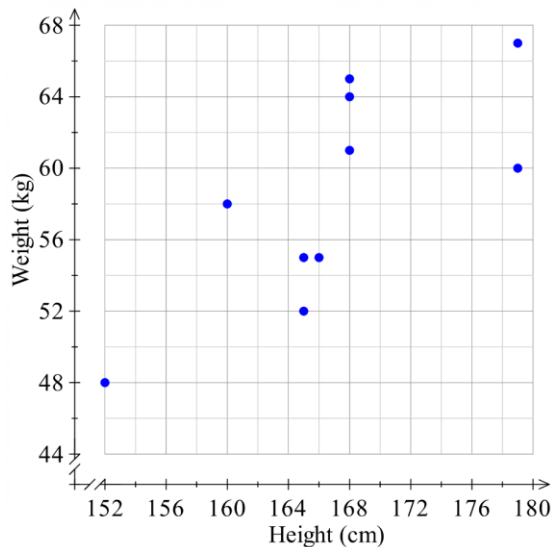
- A **scatterplot** shows two variables on a Cartesian Plane.
- A **bivariate data point** (x, y) has two parts:
 - x is the independent variable.
 - y is the dependent variable.

Example Plot data on a scatterplot.

The heights and weights of 10 students were measured and the results shown in the table.

Height (cm)	179	165	160	179	152	168	168	165	168	166
Weight (kg)	60	55	58	67	48	64	61	52	65	55

Plot this data as a scatterplot.



Comment on any relationship you can see.

Taller people tend to be heavier.

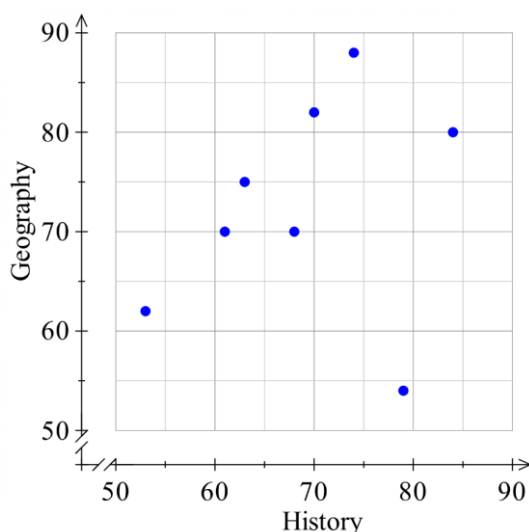
Are taller people *always* heavier than a shorter person?

Why did we use the language “taller people *tend to be* heavier”, not “taller people *are* heavier”

The test results of 8 students are shown.

History test (H)	84	53	63	74	68	79	70	61
Geography test (G)	80	62	75	88	70	54	82	70

Plot this data as a scatterplot.



Comment on any relationship you see.

People who are good at history also tend to be good at geography.

Is there likely to be a causal relationship between the two variables? i.e. does being good at history cause you to be good at geography?

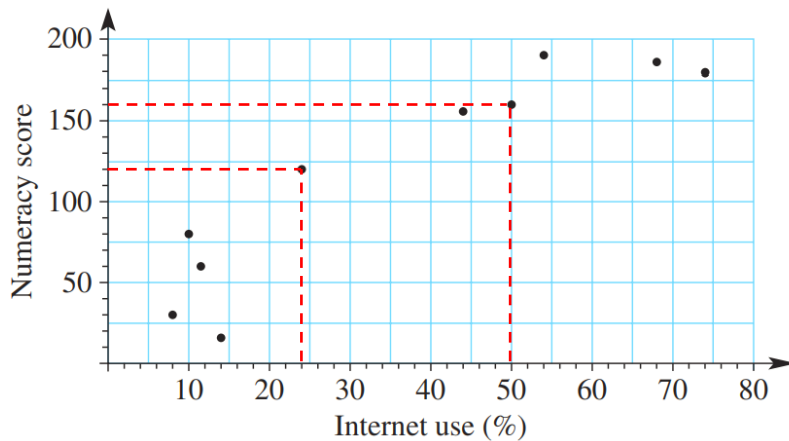
No, the two subjects are unrelated. It is more likely that people who have well-developed study habits are good at both.

INTERPRETING A SCATTERPLOT

- Interpreting a scatterplot is similar to interpreting a linear graph.
- Draw horizontal and vertical lines from the data point to the axes.

Example Interpret a scatterplot.

The average numeracy score for year six students and their general rate of internet use (%) for 10 countries are displayed in the scatterplot below.



What is the average numeracy score for the country which has an internet use rate of 24%?

120

What is the internet use (%) for the country which has an average numeracy score 160?

50%

How many countries have internet use of more than 50%? 3

How many countries have a numeracy score greater than 100? 5

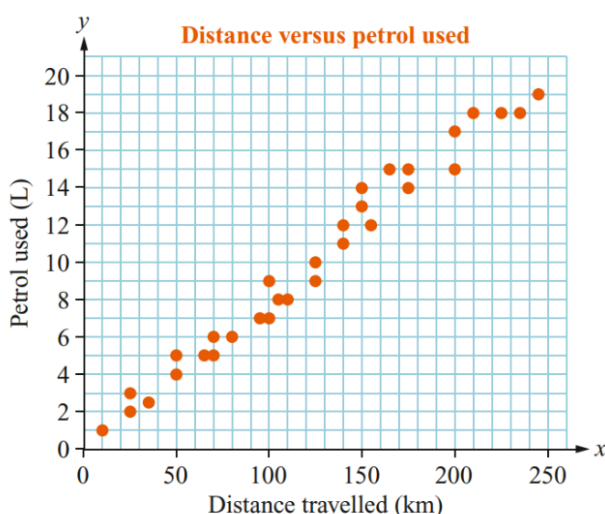
Comment on any relationship you can see.

It is likely that more internet use means more numeracy score.

Why can we **not** conclude that more internet use leads to better numeracy achievement?

.....

The distance travelled is graphed against petrol used



How much petrol was used to travel 210 km?

18 L

How far was travelled using 19 L?

245 km

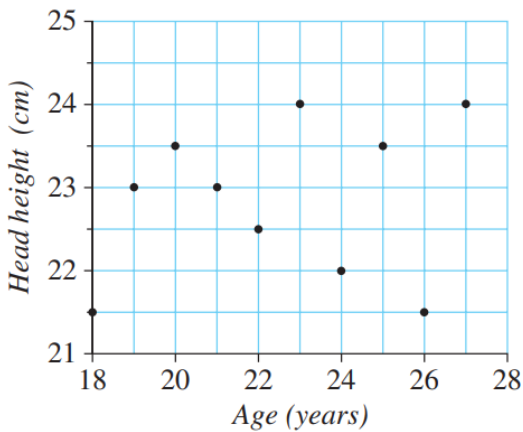
Comment on any relationship you see.

The greater the distance travelled, the more petrol used.

Can you conclude that greater distance travelled leads to more petrol used?

Yes, more petrol is needed to travel a further distance.

1 The scatterplot shows the head height to age for 10 people.

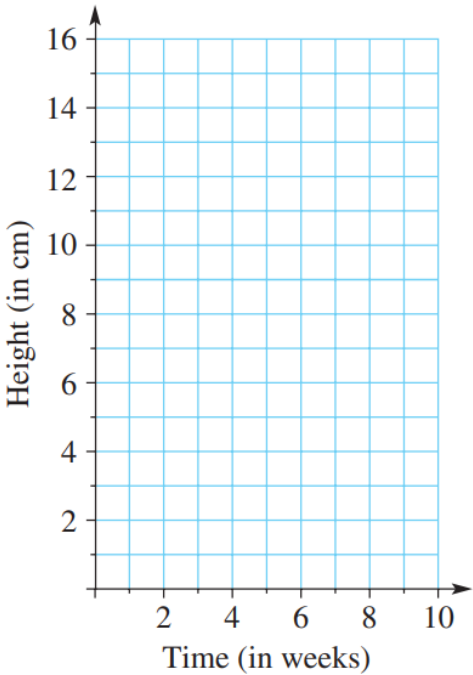


- a What is the head height for the person who is 21 years old?
23 cm
- b What is the age of the person who has a head height of 22 cm?
24 years old
- c What is the greatest head height?
24 cm
- d What is the age of the youngest person?
18 years old
- e How many people have a head height greater than 23 cm?
4 people
- f Is there a clear relationship between these two variables?
no

2 The table shows the height in cm of a eucalypt tree seedling as it grows.

Time (in weeks)	0	1	2	3	4	5	6	7	8	9	10
Height (in cm)	0	6.6	8.8	9.0	10.5	12.0	13.5	15.2	15.4	15.8	15.9

- a Plot the data as a scatterplot.
- b What is the increase in the height of the seedling between week 1 and 2?
2.2 cm
- c What is the increase in the height of the seedling during the last week?
0.1 cm
- d How many weeks does it take for the seedling to increase in height from 9 cm to 12 cm?
2 weeks
- e Estimate the height of the seedling after 4.5 weeks.
≈ 11 cm
- f Estimate the time taken for the seedling to grow to a height of 14 cm.
≈ 6.5 weeks



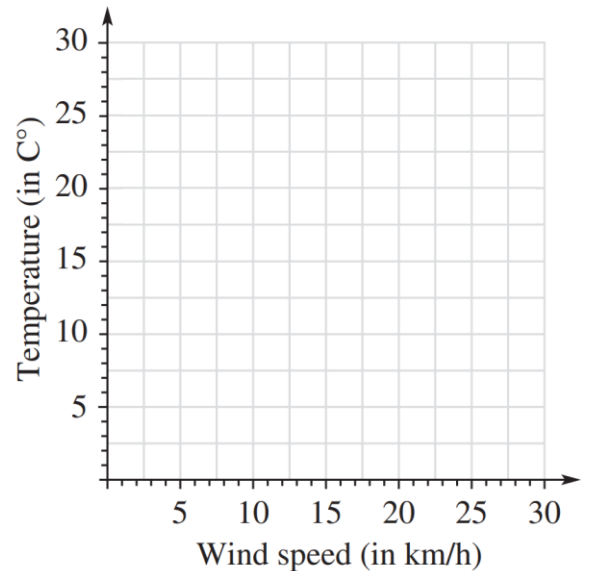
- 3 The maximum wind speed and maximum temperature in a location were recorded for 2 weeks. The data is displayed in the table below.

a Construct a scatterplot of the data given in the table.

Wind speed (in km/h)	2	6	12	15	19	20	22	25	17	14	5	11	24	13
Temperature (in °C)	28	26	23	22	21	22	19	16	20	24	25	24	19	26

- b Are there any conclusions to be drawn from the scatterplot?

As wind speed increases, temperature decreases.



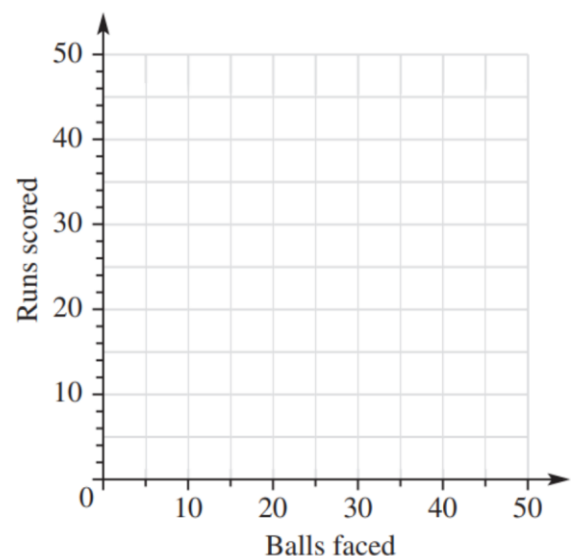
- 4 The table shows the number of runs scored and the number of balls faced by batsmen in a one-day cricket match.

a Construct a scatterplot of the data given in the table.

Balls faced	49	29	26	16	19	13	16	10	28	40	6
Runs scored	47	27	10	8	21	3	13	6	15	30	2

- b Are there any conclusions to be drawn from the scatterplot?

As the balls faced increased the number of runs increased.

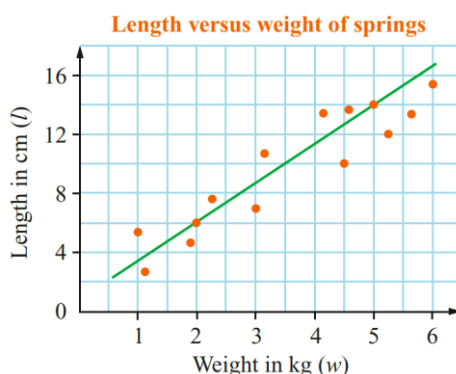


LINE OF BEST FIT

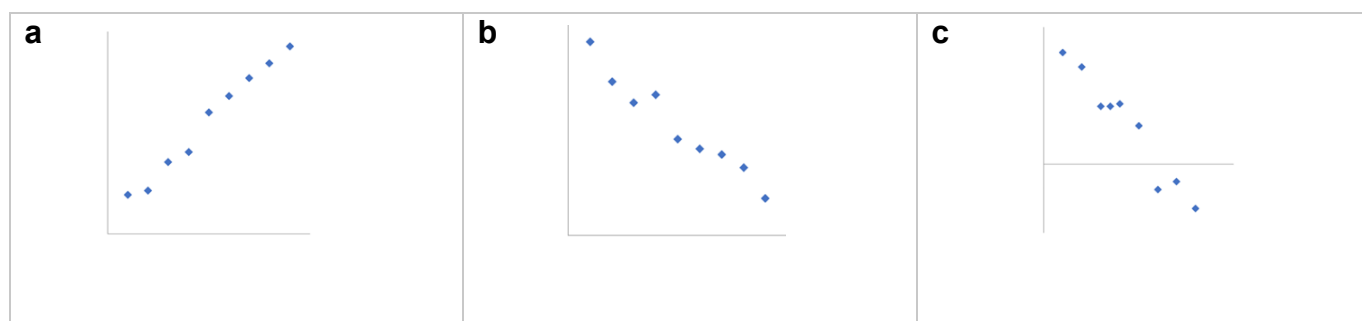
- A line of best fit:
 - Represents most of the points as closely as possible.
 - Goes through as many points as possible.
 - Has roughly the same number of points above and below it.
 - Is drawn so that the distances of points from the line are as small as possible.
- Do **NOT** extend the line past existing data points, as the model may not make sense outside of this range.
- A line of best fit may not “fit” the data nicely if there is an outlier.

Example Fit a line of best fit by eye.

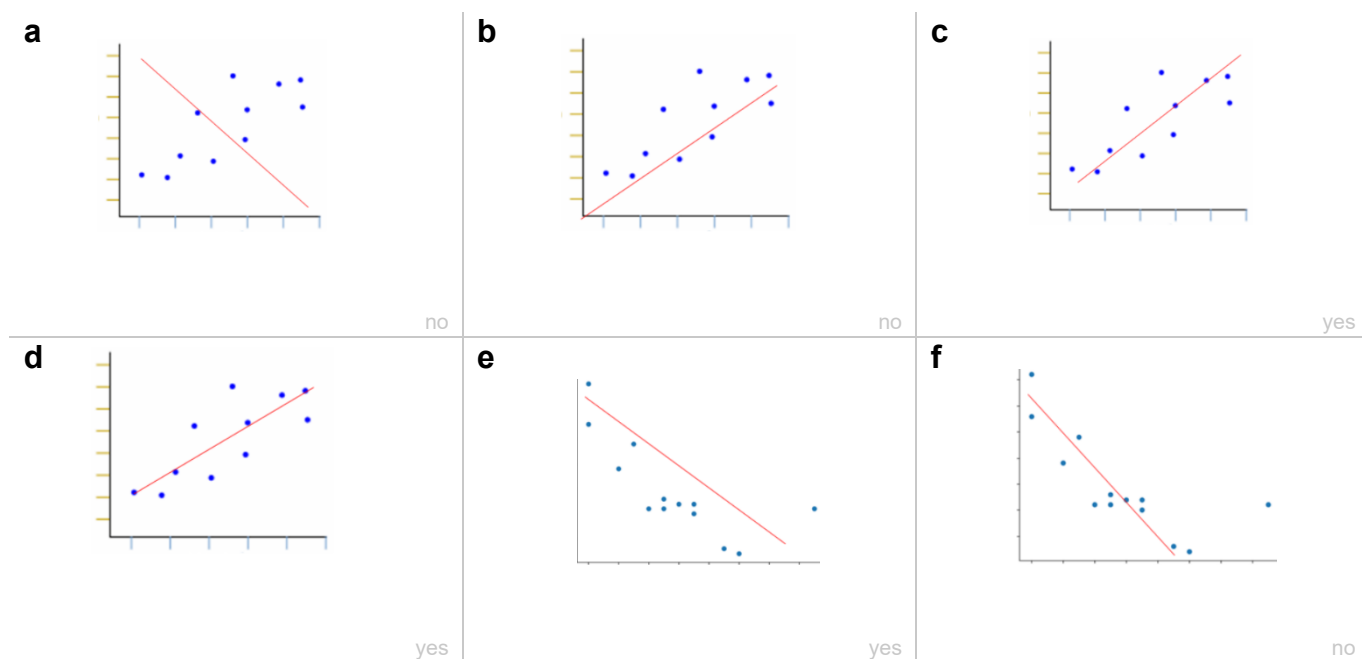
Draw a line of best fit through the data.



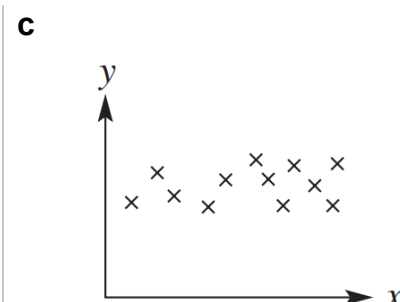
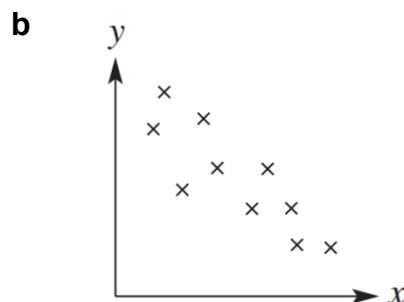
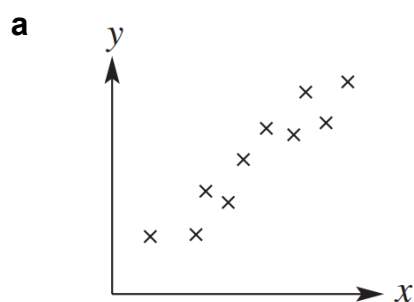
Draw lines of best fit for the following datasets.



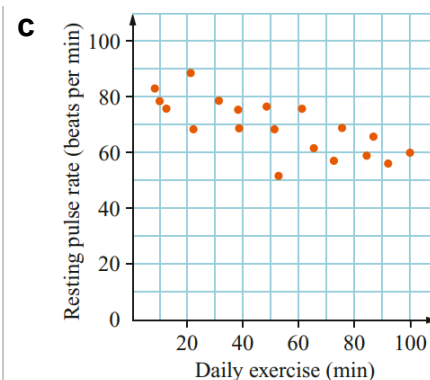
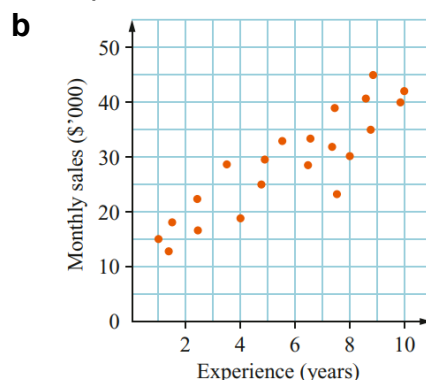
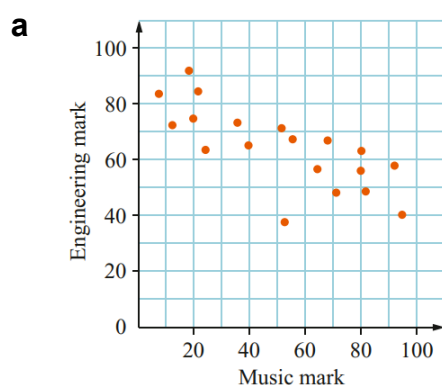
Decide whether these lines of best fit are appropriate.



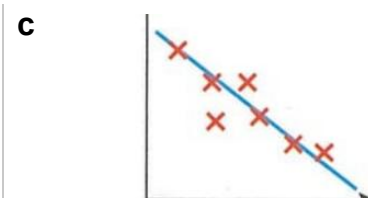
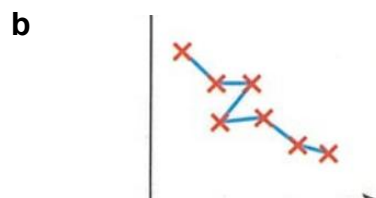
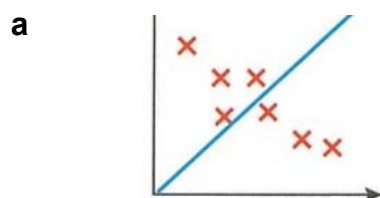
1 Sketch lines of best fit to these scatterplots.



2 Draw a line of best fit on these scatterplots.



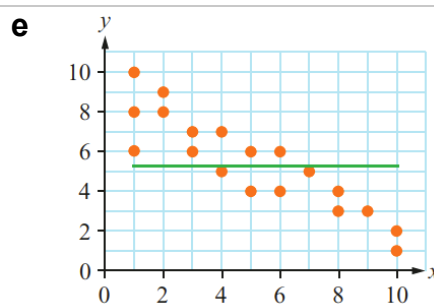
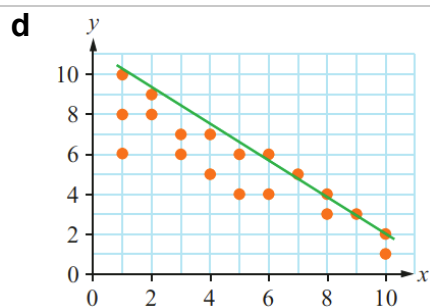
3 Explain why these are not lines of best fit.



Does not fit the data

The line of best fit is a single straight line

The line extends past the data.

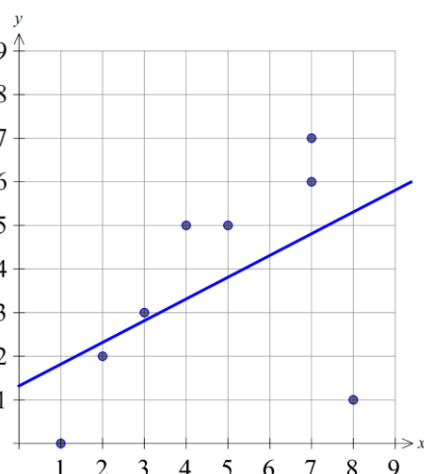
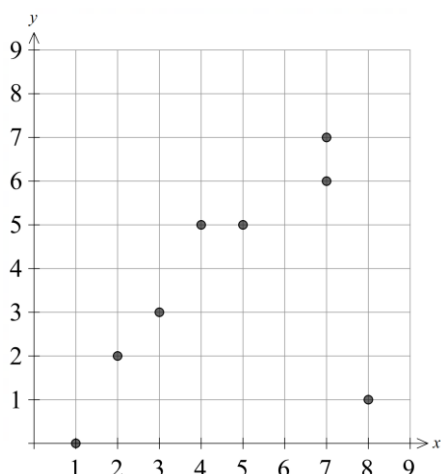


Too many datapoints below line of best fit.

Doesn't fit the data.

4 Outliers can have a large impact on the line of best fit.

Consider the dataset below. A computer was used to generate the line of best fit.



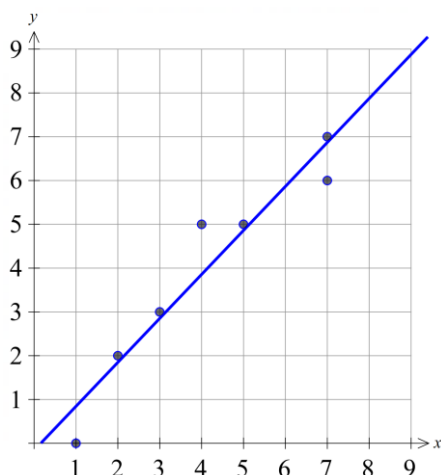
Complete the sentences:

The line of best fit doesn't seem to fit the data nicely. This is because of the outlier: (.....,).

(8, 1)

The outlier is causing the gradient of the line of best fit to be steep than expected.

less

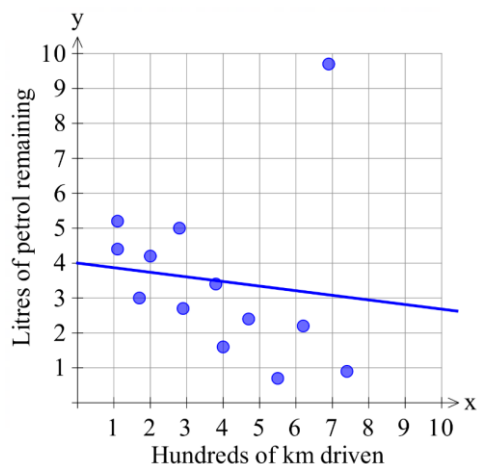


If we remove the outlier, the gradient of the line of best fit is, and fits the data better.

steeper

5 A survey collected data on cars with a 50 L petrol tank.

It compared distance driven and litres of petrol remaining in the tank.



Identify the outlier. Circle the data point.

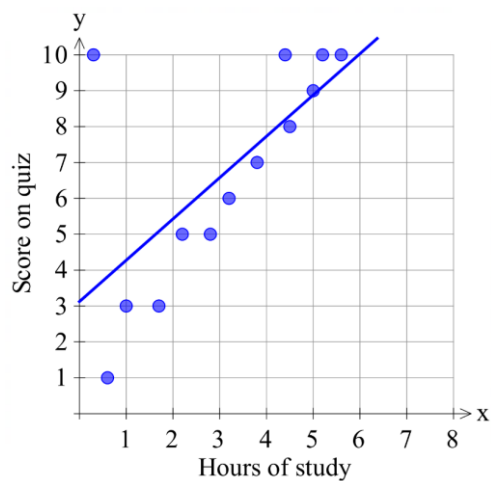
(6.9, 7.8)

If the outlier were removed, what would happen to the gradient of the line of best fit? Would it be steeper or less steep?

The gradient of the line of best fit would become more negative and steeper.

6 A survey collected data on hours studying and score on quiz.

Identify the outlier. Circle the data point.



(0.3, 10)

If the outlier were removed, what would happen to the gradient of the line of best fit? Would it be steeper or less steep?

The gradient of the line of best fit would become more positive, and steeper.