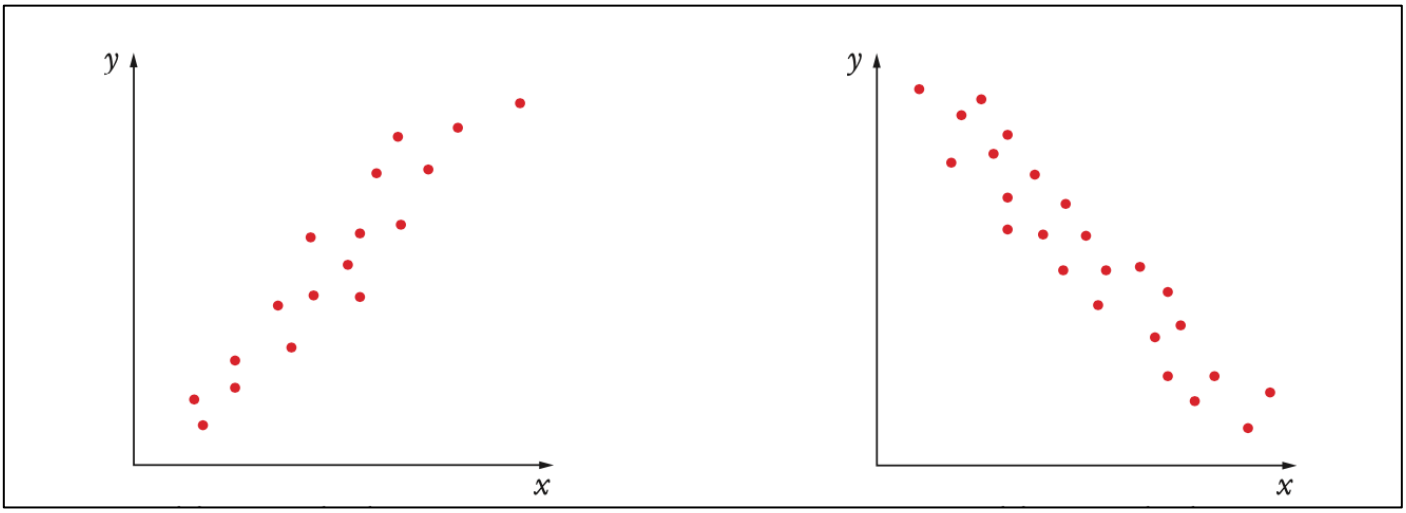
10MATH1
BIVARIATE DATA ANALYSIS
ONLINE LEARNING

LESSON 1- BIVARIATE DATA AND SCATTER PLOTS

Bivariate data is data that measures two different variables, for example, a person's height and their arm span, or the number of hours spent studying and marks on a Maths test. We can graph bivariate data on a **scatter plot**. A **scatter plot** is a graph of points on a number plane. The pattern produced often shows the data scattered around and does not always form a straight line. We can call the pattern linear or non-linear.

A scatter plot can also tell us about the relationship between the data. We call this relationship the **correlation** between the data. We use the words strength and direction when describing a scatter plot.



The graph on the left shows a **positive** relationship because both variables are increasing together. The graph on the right shows a **negative** relationship because as x increases, y decreases.

The **strength** of a relationship between two variables can be described as:

- **strong** if the points are close together
- **weak** if the points are more spread out.
- **perfect** if ALL points lie on a straight line.

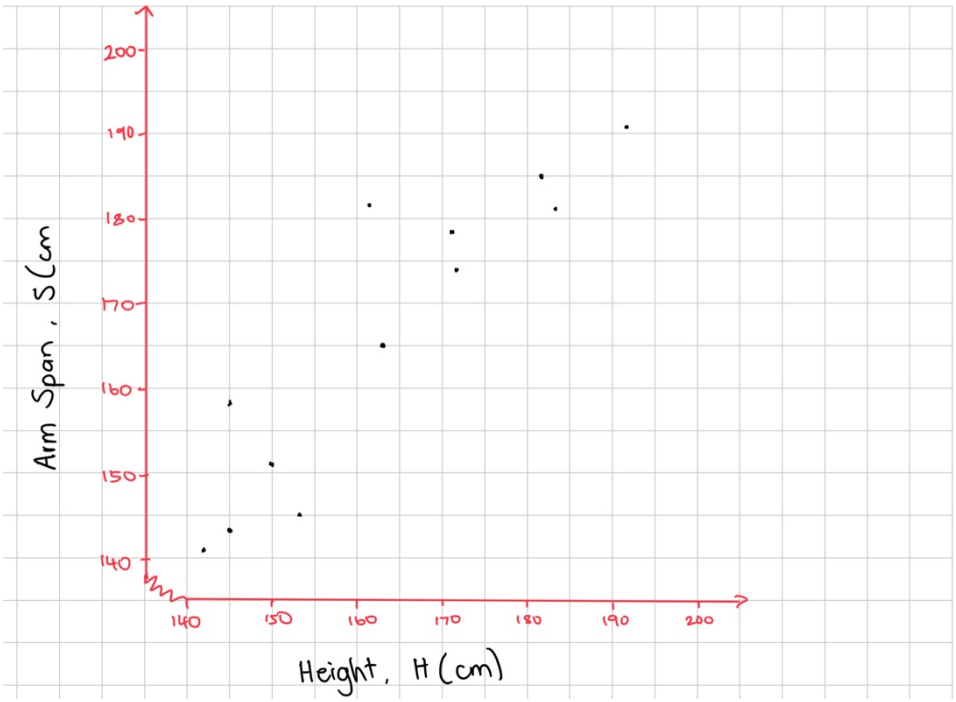
In scatter plots, y is the dependent variable and x is the independent variable. As y generally depends on x. For example, when comparing stride length and height, stride length would depend on the person's height, so height would be our independent variable, and stride length would be our dependent variable.

EXAMPLES:

1. The heights and arm spans of a group of students are shown in the table below:

Height H cm	162	182	153	145	172	163	150	142	183	145	192	171
Arm Span S cm	158	185	145	143	174	165	151	141	181	158	191	178

a) Plot the data on a scatter plot.



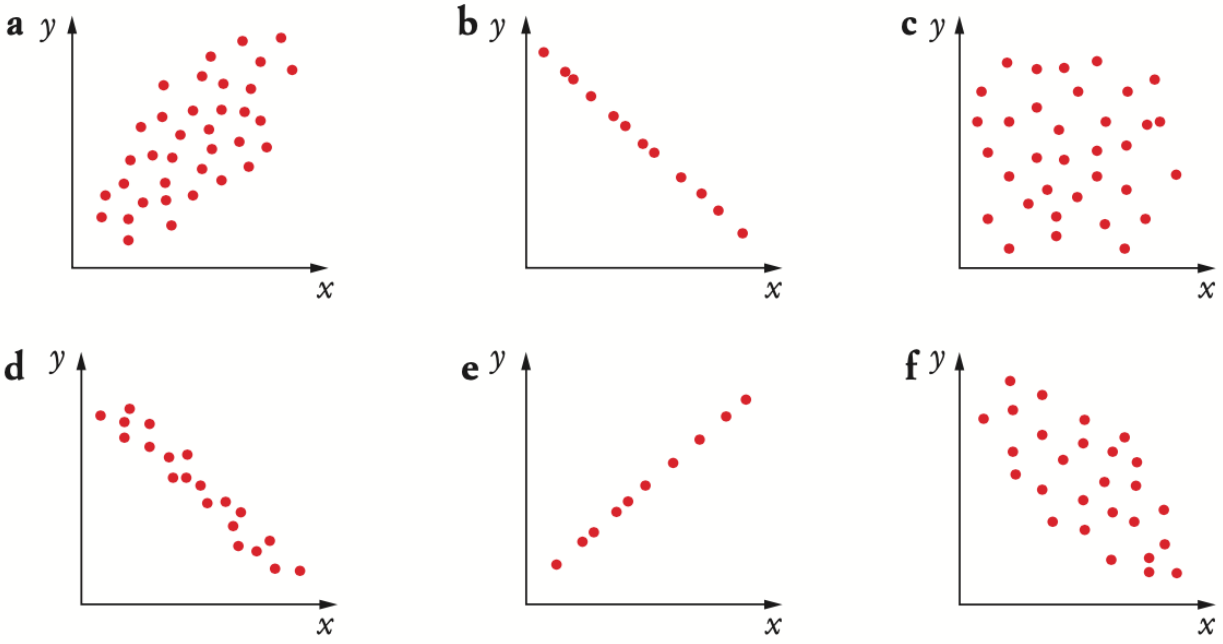
b) Describe the pattern of the plotted points

The points form a linear pattern

c) Describe the relationship between the students' heights and arm spans.

As the heights of the students increase, their arm spans tend to increase.

2. Describe the strength and direction of the relationship shown in each scatter plot below.



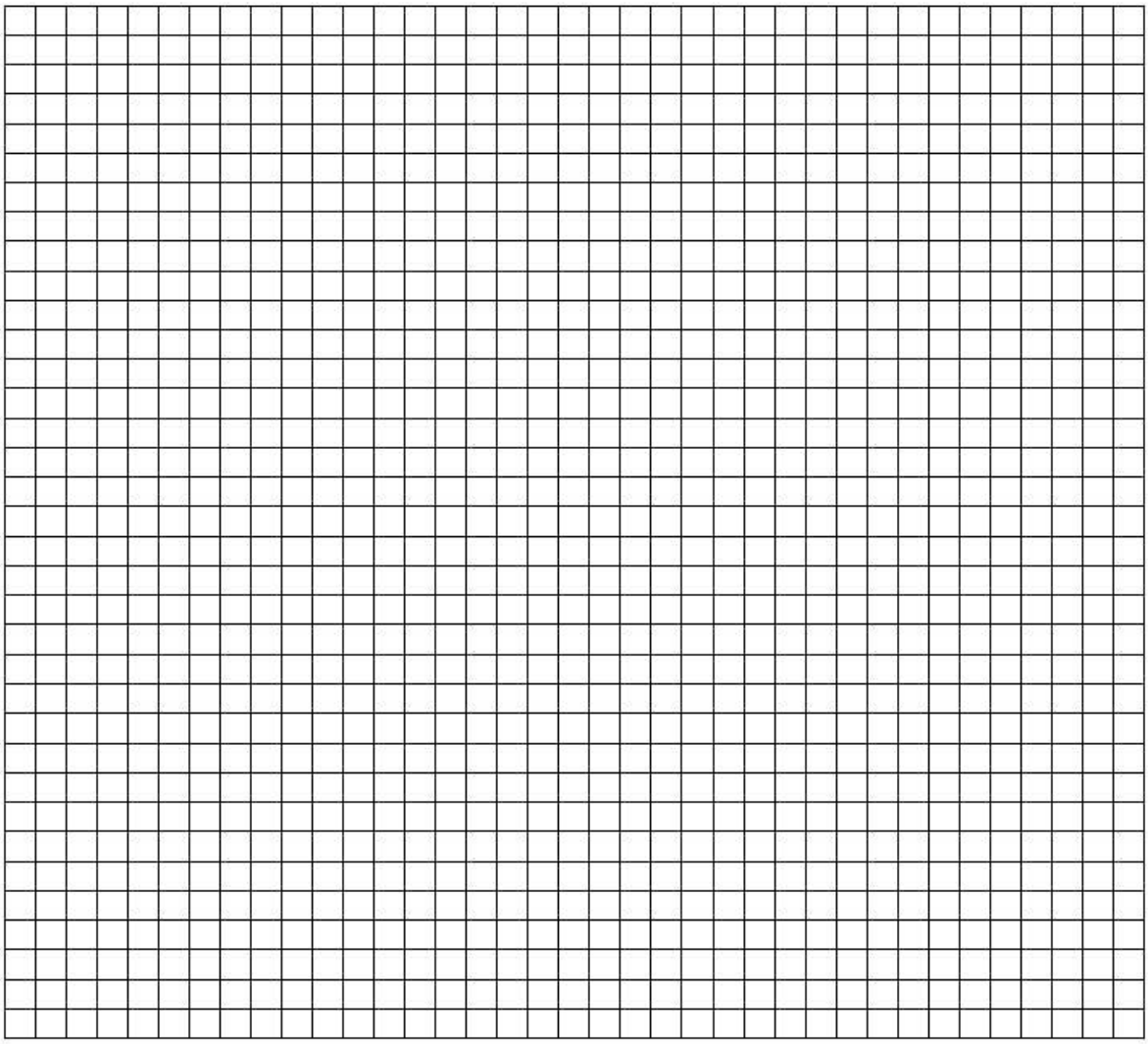
a) Weak positive relationship (the points are fairly spread out)
b) Perfect negative relationship (all points lie on a straight line)
c) No relationship (there is no pattern here)
d) Strong negative relationship
e) Perfect positive relationship
f) Weak negative relationship

YOUR TURN:

1. The height and handspans of a group of students are shown in the table below:

Height, H cm	168	175	175	156	160	173	171	180	185	175	182	180
Handspan, S cm	20.0	21.1	17.6	16.5	17.5	19.0	20.8	22.5	25.0	23.0	20.2	21.1

a) Plot the data in a scatter plot



b) Describe the pattern of the plotted points

_____

_____

c) Describe the relationship between the two variables

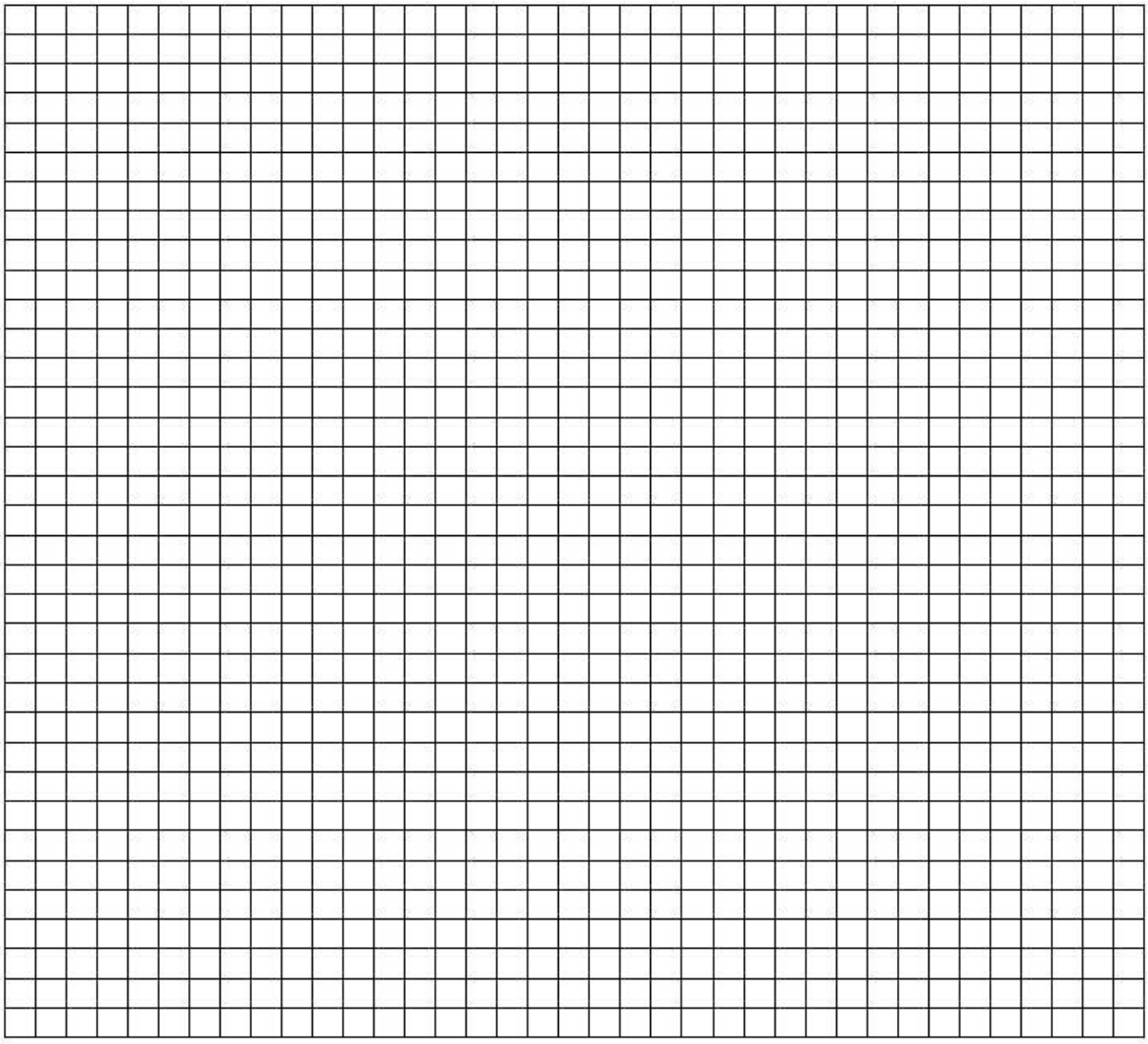
_____

_____

2. A survey was conducted to see whether there was a relationship between the age and height of students at a school. The results are in the table below.

Age, A years	14	16	15	13	11	14	17	15	12	11	14	16	13	18
Height, Hcm	162	174	182	162	132	173	187	160	154	145	165	171	151	181

a) Plot the data in a scatter plot



b) Which variable could be considered as the dependent variable? Give reasons.

_____

_____

_____

c) Describe the strength and direction of the relationship between the two variables.

_____

_____

LESSON 2- LINE OF BEST FIT

If two variables, x and y, show a strong linear relationship when graphed on a scatter plot, the linear relationship can be approximated by finding the **line of best fit** through the points and finding its equation, y=mx+b. We will learn to do this on paper, but it is best done through graphing technologies like Desmos. You may wish to use Desmos when completing the questions for this lesson.

A **line of best fit**:

- represents most or all the points as closely as possible.
- goes through as many points as possible.
- has roughly the same number of points above or below it.
- is drawn so that the distances of points from the line are as small as possible.

A line of best fit can be used to predict what might happen:

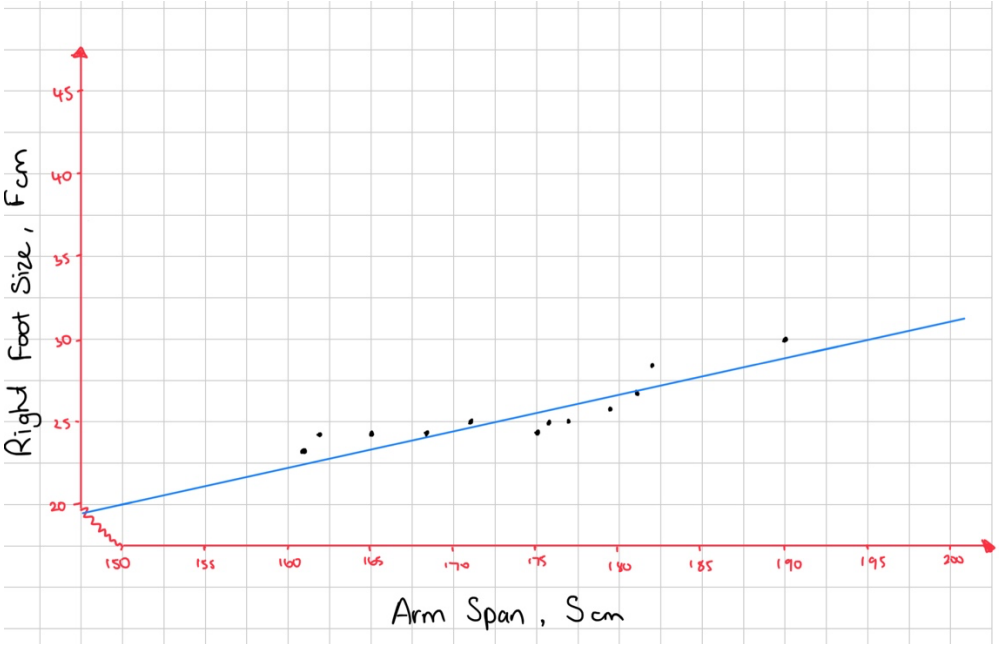
- between the points on a scatter plot, within the range of data (this is called **interpolation**)
- beyond the points on a scatter plot, outside the range of data (this is called **extrapolation**)

EXAMPLES:

The arm span and right foot size of 12 Year 10 students were measured and are shown below:

Arm Span, S cm	177	179	162	182	181	171	161	176	175	190	168	165
Right Foot Size, F cm	25	26	24	28	27	25	23	25	24	30	24	24

a) Graph the points on a scatter plot and construct a line of best fit.



b) Find the equation of the line of best fit.

We need to find the equation by using our point-gradient formula. Choose two points that fall on the line, I'm using (181, 27) and (150, 20). Using the gradient formula:

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$
$$m = \frac{20 - 27}{150 - 181}$$
$$m = 0.226\ (to\ 3dp)$$

Then using our point-gradient formula:

$$y - y_1 = m(x - x_1)$$
$$y - 27 = 0.226(x - 181)$$
$$y = 0.226x - 40.96 + 27$$
$$y = 0.226x - 13.906$$

c) Use the equation to estimate the foot size of a student with an arm span of 173cm.

y= 0.226(173)-13.906

y= 25.192

y= 25cm

d) Use the graph to interpolate the foot size of a student with an arm span of 185cm.

Reading up from 185cm, the foot size of the student would be about 28cm.

e) Use the graph to extrapolate the arm span of a student with a foot size of 31cm.

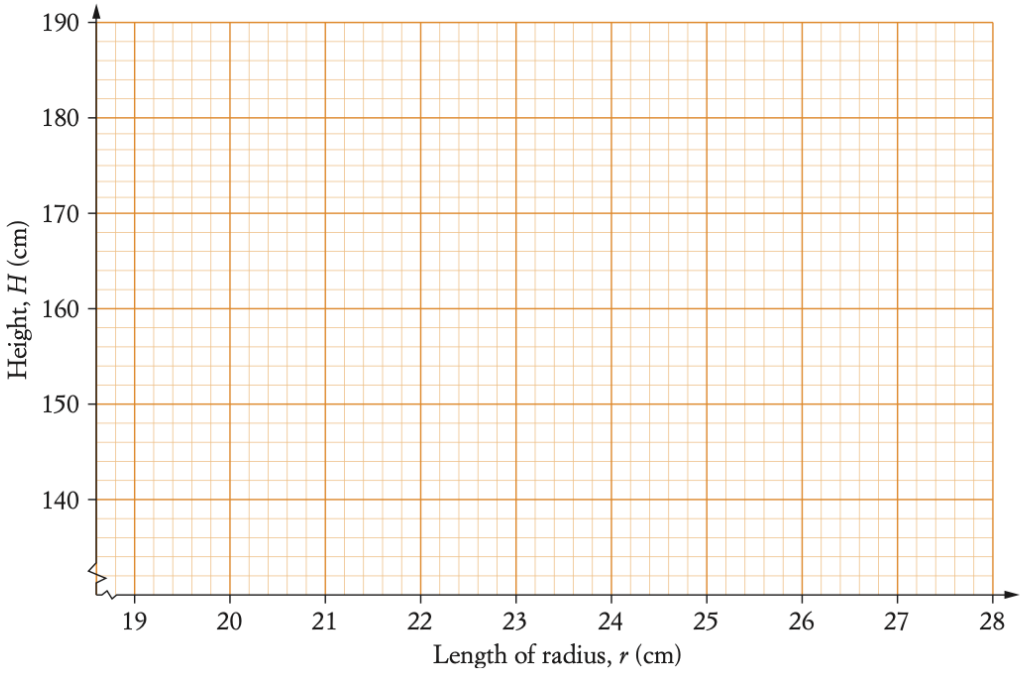
Reading across from 31cm, the armspan of the student would be about 198cm.

YOUR TURN:

1.

Forensic scientists can estimate people's heights from the lengths of their bones such as the tibia, femur, humerus and radius. The table below gives the heights of females and the length of their radius.

Length of radius, *r* (cm)	25.2	22	23	22.5	21.8	26.2	20.4	23.5	24.3	21.4
Height, *H* (cm)	173	158	165	161	158	179	152	167	169	156



a) Plot the points on the scatter plot above and draw the line of best fit.

b) Find the equation of the line of best fit

c) Use your equation to find the height of a female whose radius is 25cm long

d) If the radius is 27cm in length, use the line of best fit to predict the height of the female

2.

In this activity, we will use a spreadsheet to create a scatter plot and graph a line of best fit.
The heights of men and the lengths of their femur bone are recorded in the table below.

Length of femur, *f*(cm)	40	42.9	44.2	46.1	46.8	47	48.4	50.3	51.2	57.2
Height, *H* (cm)	162	165	164	173	174	178	179	182	186	200

1. Enter the data from the table into a spreadsheet. Type **Length of femur** in cell A1 and **Height** in B1.
2. To graph a scatter plot, select all the values in cells B1 to K2, and under the **Insert** menu, select **Scatter** and **Scatter with Straight Lines and Markers**.
3. To draw the line of best fit, select one of the points on the scatterplot and right-click. Select **Add Trendline**, **Linear** and **Display Equation on chart**, then **Close**.

LESSON 3- BIVARIATE DATE INVOLVING TIME

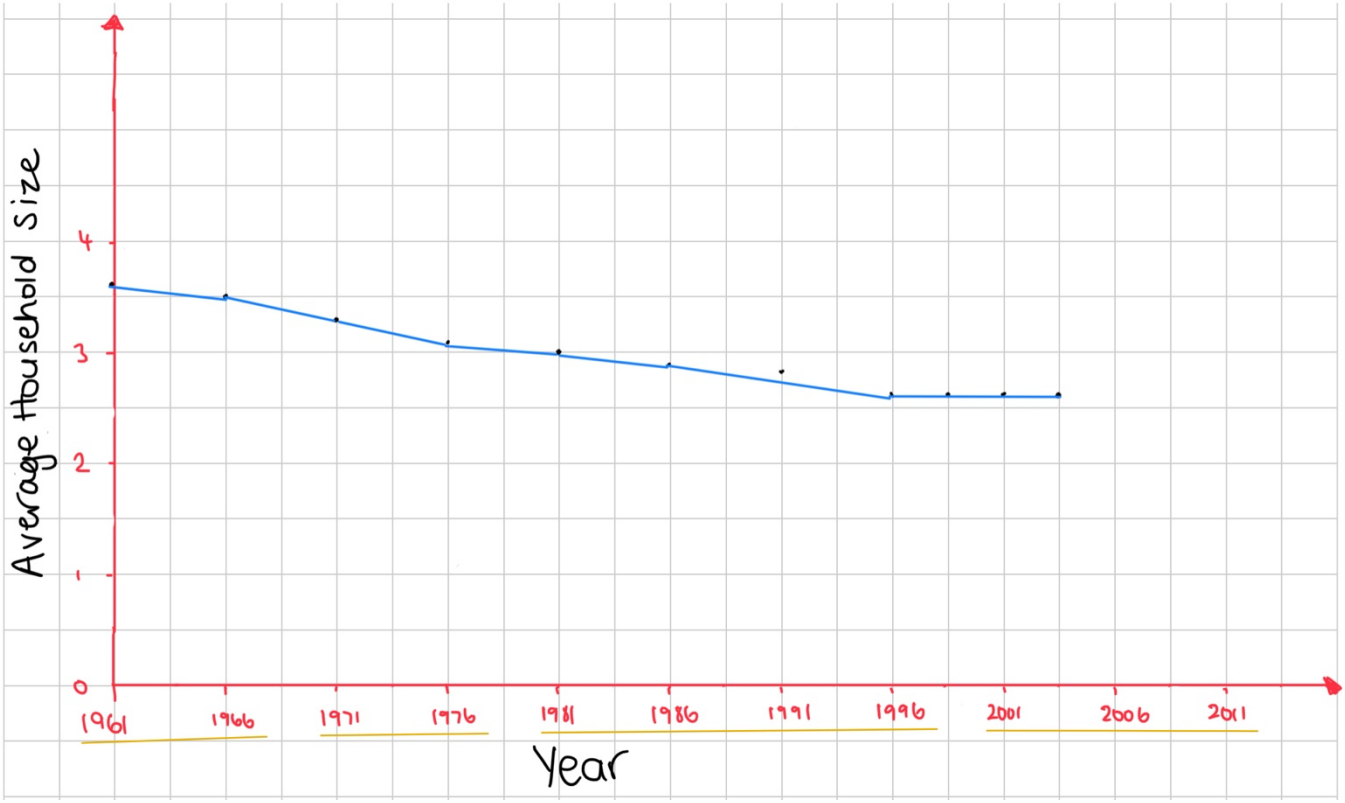
Bivariate data involving time, or sometimes referred to as time-series data, is two-variable data where the independent variable is time. Examples of time series data are population changes over time, weekly share prices, daily rainfall and patient's heart rates.

EXAMPLES:

The table below shows the average household size (number of people living in a house) between 1961 and 2011, according to the census:

Year	1961	1966	1971	1976	1981	1986	1991	1996	2001	2006	2011
Average Household size	3.6	3.5	3.3	3.1	3.0	2.9	2.8	2.6	2.6	2.6	2.6

a) Graph the data on a scatter plot and join the points



b) Use your graph to describe the change in average household size from 1961 to 2011.

The average household size decreases from 3.6 to 2.6 from 1961 to 1996 and from there remains little to no change.

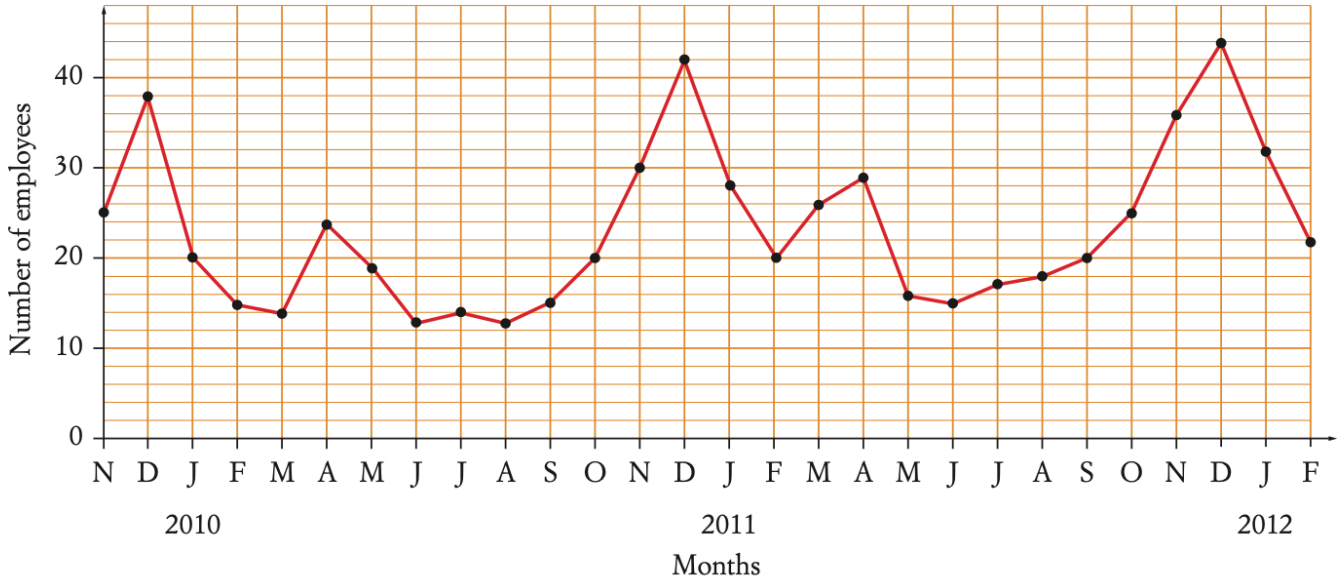
c) Based on your time series graph, estimate the household size for 2021.

2.4-2.6 people per household.

YOUR TURN:

1.

The number of people employed per month at SUPA SAVE SUPERMARKET from November 2009 to February 2012 is displayed in the time series graph below.



a) How many people were employed by the supermarket in:

i) November 2009 _____

ii) December 2010 _____

iii) June 2011 _____

b) In which month of the year were most people employed by the supermarket. Suggest a reason why.

_____

_____

c) In which month of the year were the least people employed. Suggest a reason why.

_____

_____

d) Describe how the number of people employed by the supermarket changes from November 2009 to February 2012.

_____

_____

_____

INVESTIGATION TIME:
The Australian Bureau of Statistic (ABS) is the official organisation in charge of collecting data for government departments. The data covers many areas – from population to employment, weekly earnings, weight and obesity in adults, to health of children in Australia.

Visit the ABS website www.abs.gov.au to answer the following questions.

1.

a) What is the current population of Australia?

_____

b) What is the predicted population for:

i) 2030?

_____

ii) 2040?

_____

c) What is Australia's rate of population increase?

_____

2. Go to 2016 Census Data by Location, and then to Data and analysis.

a) What was the population of NSW and its increase from 2011?

_____

b) Which state had the largest population increase?

_____

c) Which state had the smallest population increase?

_____

LESSON 4- STATISTICS IN THE MEDIA and INVESTIGATING STATISTICAL STUDIES

We live in a world of 24-hr news, whether it is from the newspaper, TV or the internet, which often quote results from surveys. When survey data is used in the media we need to consider:

- where the news comes from and what samples the statistics are based on
- who supplied the information
- the number of samples and what sample size was used
- the way in which the collected data has been presented.

EXAMPLES:
What concerns could be raised about the following claim?
"The Daily Sun newspaper reports that it has an average issue readership of 1.385 million and that its Travel liftout has a readership of 1.455 million"

The newspaper is reporting about its own readership so it could be reporting biased data. It also states that its issue readership is less than its Travel liftout. Where did they get these numbers from, as you can only get the Travel liftout by buying the actual newspaper.

YOUR TURN:

1. A TV network surveys 300 people in shopping centres between 9am and 11am to get feedback on its new game show.

   a) How may this survey be biased?

   _____

   b) Suggest a better method for obtaining feedback about its game show.

   _____

2. For the following pictures of statistics in the media, explain what is wrong with the graph, and why it is misleading.

a) What is wrong with this graph?

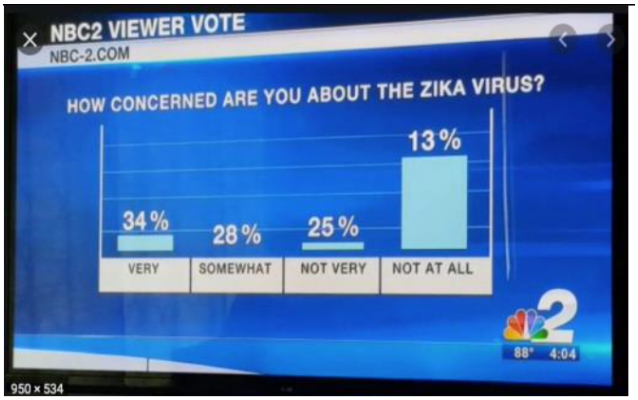
_____

_____

b) Why is it misleading?

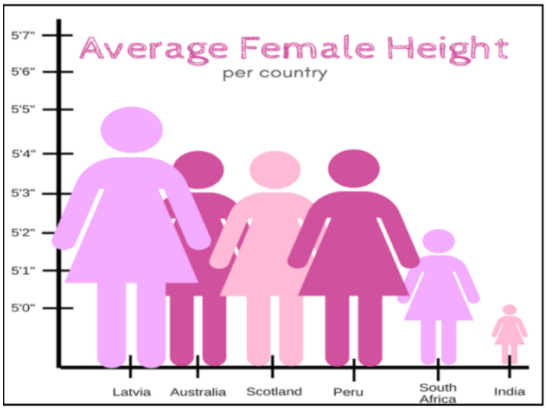
_____

_____



c) What is wrong with this graph?

d) Why is it misleading?



e) What is wrong with this graph?

f) Why is it misleading?

3. Is Australia becoming a warmer continent? Investigate this by looking at data from the Australian Bureau of Statistics and the Australian Bureau of Meteorology.

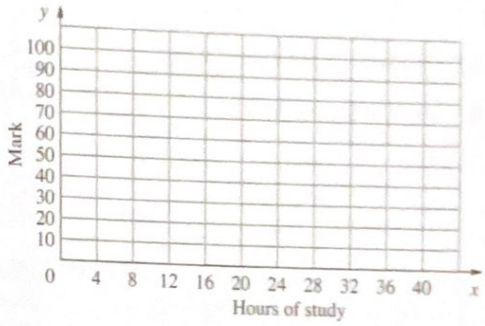
4. Investigate tobacco and alcohol use by teenagers in Australia. Include tables and graphs in your report. Refer to the National Drug Strategy Household Survey (www.nationaldrugstrategy.gov.au) and NSW Health (www.healthinsite.gov.au) and search 'alcohol and teenage statistics in Australia'. Write your report below.

UNIT 7: Scatter plots

Excel Advanced-level Mathematics Study Guide Years 9–10
Pages 187–228

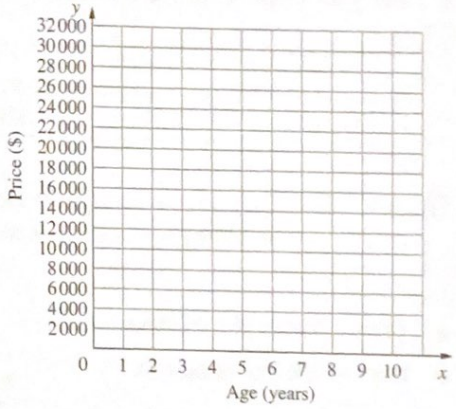
QUESTION 1 In a class, the number of hours each student spent studying for an examination and the mark each one was awarded were recorded as shown in the table below. Construct a scatter plot to show these data, and comment on any trends.

Student	Mark	Hours	Student	Mark	Hours
1	15	2	11	72	21
2	93	35	12	82	29
3	30	5	13	85	30
4	52	8	14	9	2
5	61	15	15	27	3
6	82	30	16	39	4
7	97	36	17	48	6
8	100	39	18	92	36
9	5	1	19	67	20
10	38	7	20	99	38



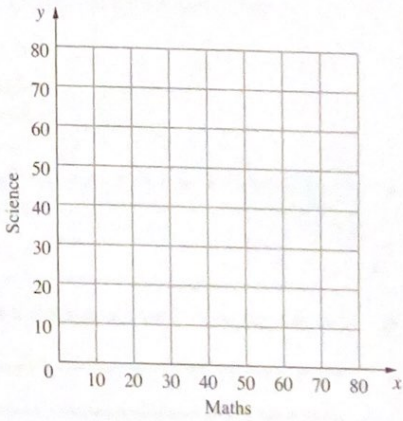
QUESTION 2 The following table shows the ages and advertised prices of a particular model car. Construct a scatter plot displaying the data and comment on any trends.

Age (years)	Price ($)	Age (years)	Price ($)
3	20 500	10	5 000
7	12 800	9	7 000
2	26 900	6	9 000
1	30 000	2	29 000
8	10 000	1	32 000
2	28 000	3	27 000
9	9 000	8	11 000
5	14 000	9	7 500
6	10 000	6	9 500



QUESTION 3 The assessment test results in Maths and Science of a class of 15 students are given in the table. Draw a scatter plot of these results and then draw in the line of best fit.

Student	Maths mark	Science mark	Student	Maths mark	Science mark
1	70	58	9	32	36
2	37	40	10	53	58
3	52	55	11	42	48
4	66	62	12	64	56
5	36	32	13	27	34
6	46	50	14	67	73
7	30	35	15	57	49
8	62	68			



Excel Advanced-level Mathematics Study Guide Years 9–10
Pages 187–228

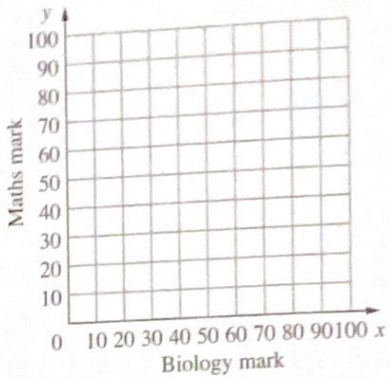
UNIT 8: Lines of best fit (1)

QUESTION **1** The data in the table to the right gives the marks out of 100 in a Maths and Biology test for a group of students.

Maths mark	Biology mark
84	74
88	82
51	47
62	55
100	90
68	59
79	73
74	64
95	90
48	42
40	34

a Plot this data on this graph.

b Draw a line of best fit.



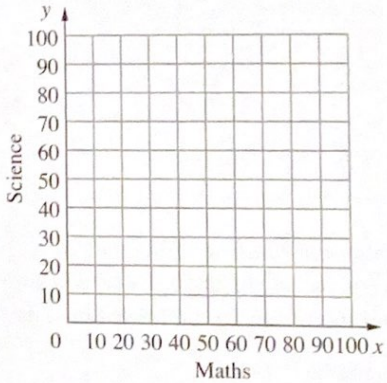
c Use your graph to find an equation for the line of best fit.

QUESTION **2** The table to the right shows the assessment results in Maths and Science of a class of 12 students.

Students	Maths	Science
1	75	63
2	42	47
3	57	62
4	71	69
5	41	39
6	51	57
7	35	42
8	67	75
9	37	43
10	58	65
11	47	55
12	69	63

a Draw a scatter plot of these results.

b Draw in the line of best fit.



c Use this line of best fit to predict Science marks for students with the following Maths results.

i 95

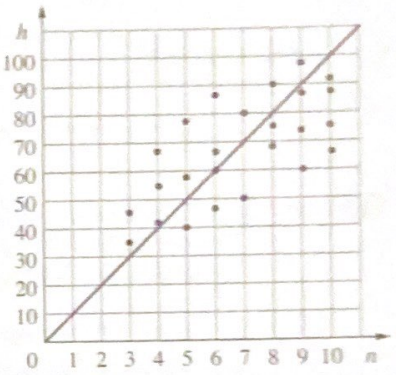
ii 20

d Use the graph to find the equation of the line of best fit.

Excel Advanced-level Mathematics Study Guide Years 9–10
Pages 187–228

UNIT 9: Lines of best fit (2)

QUESTION 1 A scatter plot shows the height (h cm) of some seedlings n weeks after they were planted. A line of best fit has been drawn on the scatter plot.



a What is the equation of the line of fit?

_____

b Using this equation, what would be the height of a seedling planted for:

i $8\frac{1}{2}$ weeks? _____

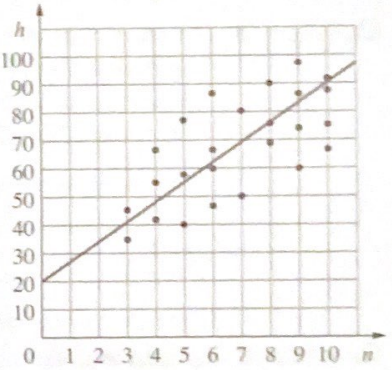
ii 13 weeks? _____

c Using this equation, how many weeks would you predict that a seedling had been planted if it had a height of:

i 55 cm? _____

ii 160 cm? _____

QUESTION 2 A different line of fit was drawn on the same scatter plot from question 1.



a What is the equation of the line of fit?

_____

b Using this equation, what would be the height of a seedling planted for:

i $8\frac{1}{2}$ weeks? _____

ii 13 weeks? _____

c Using this equation, how many weeks ago would you predict that a seedling had been planted if it had a height of:

i 55 cm? _____

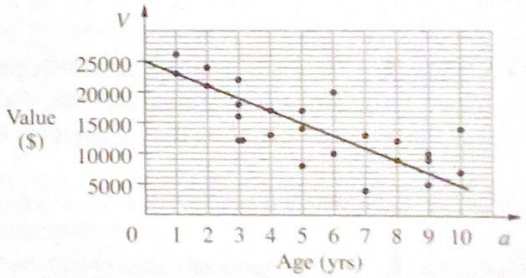
ii 160 cm? _____

d Briefly comment on the differences between predictions based on the different lines of fit.

_____

_____

QUESTION 3 A line of fit was drawn on this scatter plot that shows the age and value of some cars.



a What is the equation of the line of fit?

_____

b What is the predicted value of a car that is 5 years old?

_____

c What is the predicted age of a car that is valued at $5000?

_____

d After what age will the equation be no longer valid? Justify your answer.

_____

_____

Excel Advanced-level Mathematics Study Guide Year 10
Pages 187–228

UNIT 10: Investigating reports

QUESTION 1 A survey of some year 10 students found that 35% of the students couldn't swim the length of a 50 m pool. A politician, appalled by this statistic, suggested that all school students should have swimming lessons. Do you think that is a good response? Justify your answer.

QUESTION 2 A particular vitamin supplement, (brand X), is being advertised with a claim that 'studies show that brand X is the best'. What information about these studies would be helpful in deciding whether or not the claim is correct?

QUESTION 3 A phone survey just before an election asked whether voters would still vote for a particular politician if she changed her stance on an environmental issue. The survey was conducted for one of the politician's opponents. The politician claimed that she had no intention of changing her stance. Why do you think the question was being asked in the survey? Comment.

QUESTION 4 A current affairs show on television showed a program about a particular crime and the leniency of the sentence given to the perpetrator. At the end of the show the viewers were asked to vote whether or not they believed sentences for crimes were long enough. What problems might there be with the results of this survey?

QUESTION 5 An insurance company uses probabilities when determining the size of premiums that must be paid by their customers. Two neighbours, with exactly the same type of car, have to pay completely different amounts for insurance with the same company. Why might this be?

QUESTION 6 The results of a census in a local area showed that the area had a high percentage of young families. What implications might this have for government decision makers?