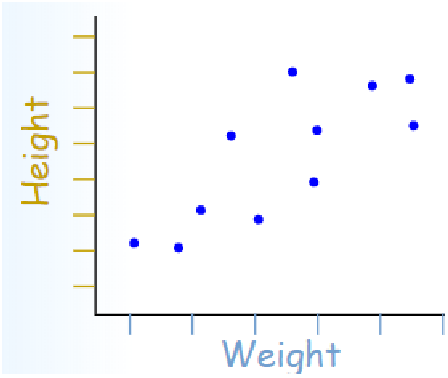
Week 15/16
Term 2
2022

HAWKER COLLEGE
Engage | Inspire | Achieve

Learning Brief

EM3

Goals



This week and next week we are going to:

- Apply further graphical concepts
- Describe the patterns and features of bivariate data
- Describe the association between the two variables, identifying the dependent (*y*) and independent variable (*x*)
- Find the line of best fit
- Use the line of best fit to make predictions by interpolation or extrapolation

HAVE YOU GOT A RULER AND A CALCULATOR? YOU WILL NEED ONE FOR THE TEST.

Theoretical Components

Resources:
PDF file: Week 15 and 16 Notes and Exercises

Knowledge Checklist:

- Bivariate data
- Scatterplots
- Independent and dependent variables
- Correlation/causation
- Line of best fit

Order:

1. Work through the Week 16 and 17 notes and exercises
2. Complete the Portfolio task
3. Complete the reflection at the end of the booklet
4. Show your teacher your completed booklet.

Practical Components

Work through the exercises and show the complete tasks to your teacher.

Be sure to ask for help as you need for the successful completion of all tasks.

You will need some graph paper. There are some attached to the end of this booklet. If you need more, see your teacher.

Portfolio Task

See the last page of the booklet

Other

Have you organised your folders yet? Don't forget you are allowed to take your classwork to the test as reference.

ESSENTIAL MATHEMATICS 3

WEEK 16 AND 17 – SCATTERPLOTS

Scatterplots

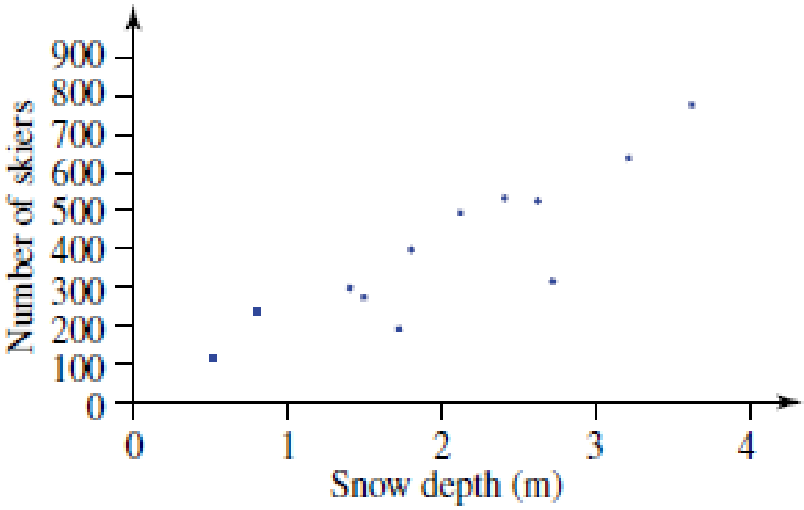
The manager of a small ski resort has a problem. He wishes to be able to predict the number of skiers using his resort each weekend in advance so that he can organise additional resort staffing and catering if needed. He knows that good deep snow will attract skiers in big numbers by scant covering is unlikely to attract a crowd. To investigation the situation further he collects the following data over twelve consecutive weekends at his resort.

Depth of snow (m)	Number of skiers
0.5	120
0.8	250
2.1	500
3.6	780
1.4	300
1.5	280
1.8	410
2.7	320
3.2	640
2.4	540
2.6	530
1.7	200

The data in this example are know as **bivariate** data. For each item (weekend), two variables are considered (depth of snow and number of skiers). When analysing bivariate data, we are interested in examining the relationship between the two variables. In the case of the ski resort data, we might be interested in finding out:

- Are visitor numbers related to depth of snow?
- If there is a relationship, then is it always true or is it just a guide? In other words, how strong is the relationship?
- Is it possible to predict the likely number of skiers if the depth of snow is known?
- How much confidence could be placed in the prediction?

To help answer these questions the data can be graphed on a **scatterplot**. Each of the data points is represented by a single visible point on the graph.



When drawing a scatterplot, it is important to choose the correct variable to assign to each of the axis. In this situation, we are assuming that the number of skiers depends on the amount of snow. It is called the **dependent variable**. The snow depth affects the number of skiers (rather than the other way around) and it is called the **independent variable**.

The independent variable is always placed on the x-axis.

The dependent variable is the variable that responds to changes in the independent variable.

Notice how the scatterplot for the ski resort data shows a general upward trend. It is not a perfectly straight line, but it is still clear that a general trend or relationship has formed: as the depth of snow increases, so does the number of skiers also increases.

When one variable increases with another it is said that there is a *positive correlation* between the variables.

Relationships can show either positive or negative correlation.

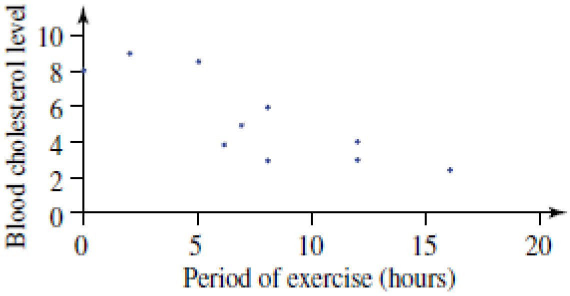
Example 1

Consider the following table which 10 Year 11 students were surveyed to find out the amount of time that they spend doing exercise each week. This is compared with their blood cholesterol level.

Period of exercise (h)	6	8	12	16	2	0	5	8	7	12
Blood cholesterol level	4	3	3	3	9	8	9	6	5	4

Construct a scatterplot and determine the relationship between the two variables.

Solution

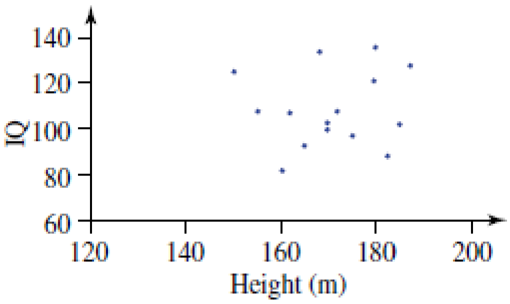


Notice how the trend shows a general downward trend. So, as the amount of exercise increases, the blood cholesterol decreases.

When one variable decrease while the other increase it is said that a *negative correlation* exists between the variables.

Notice in this case the points are not as closely aligned as in the ski resort data. We could say that the relationship (or correlation) between the variables is only a weak relationship. In general terms, the closer the points are to forming a straight line, the stronger the relationship is between the variables.

Sometimes we find that there is no relationship between the variables. Consider the example in which a researcher was looking for a link between people's heights and their IQs (intelligence quotients). The points appear to be randomly dispersed across the scatterplot. In cases like this, it can be concluded that there is no clear relationship between the variables.



Exercise 1

1. For each of the following pairs of variables, identify the independent variable and the dependent variable. If it is not possible to identify this, then write 'not appropriate'.

   a. The age of an AFL footballer and his annual salary

      I:

      D:

   b. The growth of a plant and the amount of fertiliser it receives

      I:

      D:

c. The number of books read in a week and the eye colour of the readers

I:

D:

d. The voting intentions of a person and their weekly consumption of red meat

I:

D:

e. The number of members in a household and the size of the house

I:

D:

f. The month of the year and the electricity bill for that month

I:

D:

g. The mark obtained for a maths test and the number of hours spent preparing for the test

I:

D:

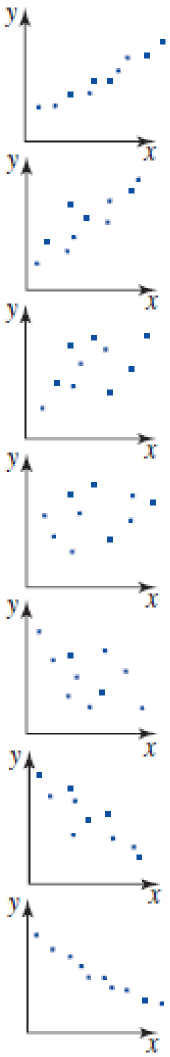
h. The cost of grapes (in dollars per kilogram) and the season of the year

I:

D:

Drawing conclusions/causation

When data is graphed, we can often estimate by eye (rather than measure) the type of correlation involved. Our ability to make these qualitative judgements can be seen from the following examples, which summaries the different types of correlation that might appear in a scatterplot.



Strong positive correlation
Conclusion: The greater the level of x, the greater the level of y.

Moderate positive correlation
Conclusion: There is evidence to show that the greater the level of x, the greater the level of y.

Weak positive correlation
Conclusion: There is a little evidence to show that the greater the level of x, the greater the level of y.

No correlation
Conclusion: There appears to be no relationship between the variables x and y.

Weak negative correlation
Conclusion: There is a little evidence to show that the greater the level of x, the smaller the level of y.

Moderate negative correlation
Conclusion: There is evidence to show that the greater the level of x, the smaller the level of y.

Strong negative correlation
Conclusion: The greater the level of x, the smaller the level of y.

Notice how the conclusion drawn for each scatterplot is slightly different. If the correlation is a strong one, then the resulting conclusion can be more definite than if it were weak.

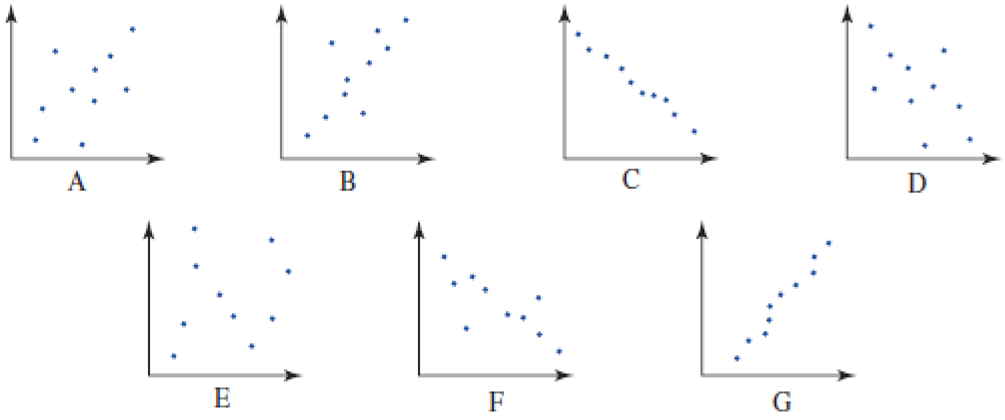
When drawing conclusions from a scatterplot or in summarising its trend it is important to avoid using statements like ‘x causes y’. Just because there is a strong relationship between two variables, it does not mean that one variable causes the other.

In fact, a strong correlation might have resulted because variable y is causing changes in x, or it could be that there is some third factor that is causing changes in both variables x and y.

To illustrate this point, a Dutch researcher compared the human birth rates (births per 1000 population) in different areas with the stork population in those areas. He found that there was a strong positive correlation between the stork population in the different areas and the human birth rate in those areas. What could he conclude? That storks cause babies? Absolutely not! His conclusion could only be along the lines of ‘the greater the stork population, the greater the human birth rate’. In this case, there was a third factor that was causing the apparent relationship. Storks prefer nesting in rural areas and for social demographic reasons, rural dwellers tend to have larger families than their cosmopolitan counterparts. In other words, the land usage of the areas was causing changes in both the stork population and the human birth rate.

Exercise 2

1. Match the following scatterplots with the correlation it shows



a Strong positive correlation
b Moderate positive correlation
c Weak positive correlation
d No correlation
e Weak negative correlation
f Moderate negative correlation
g Strong negative correlation

2. A pie seller at a football match notices that there seems to be a relationship between the number of pies that he sells and the temperature of the day. He records the following data.

Daily temperature (°C)	12	22	26	11	8	18	14	16	15	16
Number of pies sold	620	315	295	632	660	487	512	530	546	492

a. Draw a scatterplot of the data.

b. State the type of correlation the scatterplot shows and draw a suitable conclusion from it.

c. Suggest why the plot is not perfectly linear.

3. A researcher is investigating the effect of living in air-conditioned buildings upon general health. She records the following data.

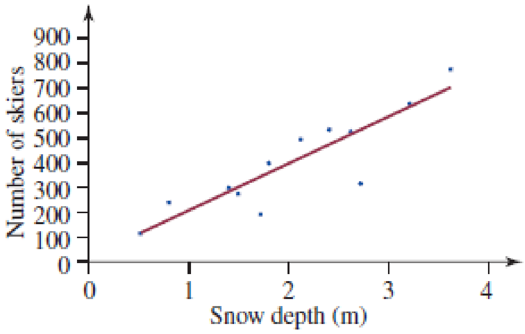
Hours spend each week in airconditioned buildings	2	13	6	48	40	0	10	0	2	5	18	10	12	26	30
Number of days sick due to flu and colds	3	6	2	15	13	8	14	1	16	9	9	6	7	14	15

a. Draw a scatterplot of the data.

b. State the type of correlation the scatterplot shows and draw a suitable conclusion from it.

c. The researcher finishes her experimental report by concluding that air-conditioning is the cause of poor health. Is she correct to say '… is the cause …'? What other factors could have influenced the relationship shown by the scatterplot?

Drawing a straight line through the set of points



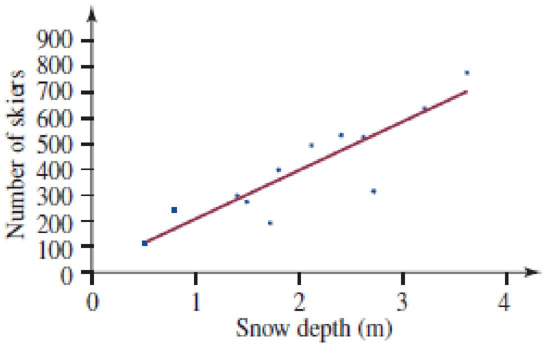
To help analyse a scatterplot, we need to fit a straight line through the whole set of points. The process of fitting a line to a set of points is often referred to as regression. It is usually not possible to rule a single straight line through all the points. We are looking for the straight line which most closely fits the data; that I, the **line of best fit**. The positioning of this line by eye will clearly rely upon some careful judgement.

Exercise 3

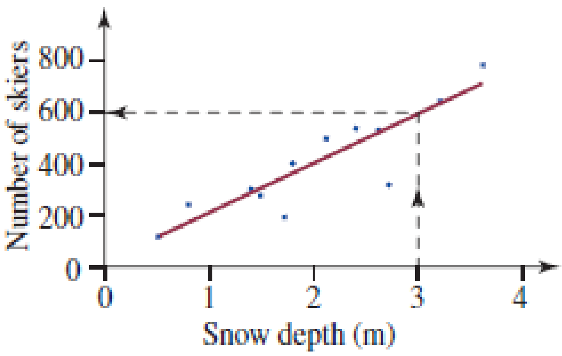
1. Return to Exercise 2 question 2 and 3 and fit a line of best fit to the scatterplots you have drawn.

Making predictions

We can use our line of best fit to make predictions about our data. Below is the scatterplot (with the line of best fit) for the depth of snow and the number of skiers.

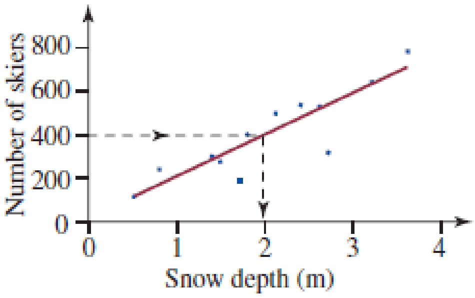


We can use the line of best fit to find the number of skiers when the snow depth was 3 m.

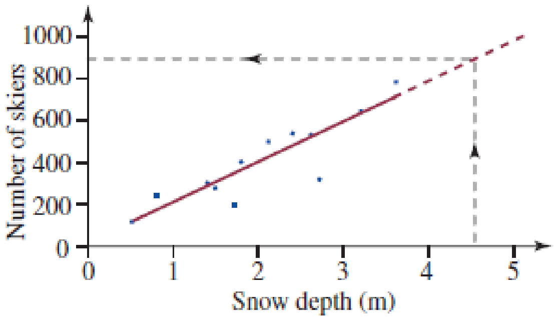


By using a ruler to draw a vertical and horizontal line, we can see that there would be 600 skiers.

Likewise, we can predict the depth of snow needed to draw a crowd of 400 skiers.



The snow needs to be 2 m deep to draw a crowd of 400 skiers.
Using the data in this way is called **interpolation**.
We can also extend our line of best fit to make predictions beyond the data collected. This is called **extrapolation**.



By extrapolating the line of best fit beyond 4 m, we can predict that 4.5 m would draw approximately 900 skiers.

You must be very careful making predictions this way. In the above example, we are assuming that the relation between snow depth and number of skiers continues in a smooth (linear) upward trend. This many not be the case.

Exercise 4

1. A manufacturer who is interested in minimising the cost of training gives 15 of his machine operators different amounts of training. He then measures the number of machine errors made by each of the operators. The results are shown in the table below.

Hours spent training	2	5	7	3	8	9	16	18	10	4	6	12	10	13	14
Number of errors during first week	43	16	25	31	24	18	6	8	24	34	27	15	22	6	8

a. Draw a scatterplot of this data. Here the **hours spent training** is the independent variable and goes on the *x* axis. The **number of errors** is the dependent variable and goes on the *y* axis. For the scale, start both axis at zero and go up by 5s.

b. Draw a line of best fit and extrapolate in both directions.

c. Use this line to answer the following questions.

i. If an operator had **zero** hours of training, how many errors would you expect?

ii. How many hours of training would you need before you would be confident of making no errors?

2. The height (in centimetres) of 21 Aussie Rules football players were recorded against the number of marks they took in a game of football. The data are shown in the following table.

Height (cm)	Number of marks taken
184	6
194	11
185	3
175	2
186	7
183	5
174	4
200	10
188	9
184	7
188	6

Height (cm)	Number of marks taken
182	7
185	5
183	9
191	9
177	3
184	8
178	4
190	10
193	12
204	14

a. Draw a scatterplot of this data. Let height be the independent variable – it goes on the *x* axis. Start at 0 and go up by 20. Number of marks is the dependent variable, and it goes on the *y* axis, start at 0 and go up by 2.

b. Draw a line of best fit and extrapolate in both directions.

c. Use this line to answer the following questions.

   i. How tall do you need to be to mark 18 marks?

   ii. If a player made no marks predict their heights.

d. What is the problem with these results?

3. The following table shows the fare charged by a bus company for journeys of differing lengths.

Distance (km)	Fare
1.5	$2.10
0.5	$2.00
7.5	$4.50
6	$4.00
6	$4.50

Distance (km)	Fare
2.5	$2.60
0.5	$2.00
8	$4.50
4	$3.50
3	$3.00

a. Represent the data using a scatterplot. Draw the trend line (line of best fit) and extrapolate the line in both directions.

b. According to the trend line estimate the cost for travelling 10 km,

c. Use the trend line to find the cost of travelling 0 km. Comment on this result.

Week 15&16 Portfolio Task

Hypothesis: the length of a person’s foot is approximately 15% of their height.

You are going to investigate to see whether this is true.

1. Complete the table blow by collection data from 10 people from your school.

	Foot length cm	Height m
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

2. Draw a scatterplot using the date above. Add a line of best fit and label as ‘my school’

3. Complete this table to show what people's foot length would be if it is 15% of their height.

Foot length cm	Height m
	1.0
	1.2
	1.4
	1.6
	1.8

4. Plot those points on your scatterplot graph. Add a line that best fit and label it as 'hypothesis'.

5. Use your scatter graph to evaluate whether the hypothesis was true for your school. Write down your response base on the following:

   - How close was your line of best fit to the hypothesis line?
   - Did any one from your school fail on or very near the hypothesis line?
   - Was the hypothesis correct or can you improve it?

MARKING RUBRIC

CRITERIA	EXPECTATIONS	POSS	MULT	GIVEN	TOTAL
Practical	Student completes practical work, including exercises of the brief to an acceptable standard set by the teacher.	**2**	**3**		**/6**
Portfolio Task	Student completes the investigation task of the week to an acceptable standard set by the teacher.	**2**	**2**		**/4**
Reasoning and Communications	Student responses are accurate and appropriate in presentation of mathematical ideas, with clear and logical working out shown.	**4**	**-**		**/4**
Concepts and Techniques	Student submitted work selects and applies appropriate mathematical techniques to solve practical problems and demonstrates proficiency in the use of mathematical facts, techniques and formulae.	**4**	**-**		**/4**
	Submission Guidelines				
Timeliness	Student submits the exercises and investigation by the set deadline. See scoring guidelines for specific details.	**2**	**-**		**/2**
				FINAL	**/20**

Student Reflection:
How did you go with this week's work? What was interesting? What did you find easy? What do you need to work on?