

Please check the examination details below before entering your candidate information

Candidate surname

Other names

Centre Number

Candidate Number

Pearson Edexcel
Level 1/Level 2 GCSE (9–1)

--	--	--	--	--

--	--	--	--	--

Specimen Paper (Set 2)

(Time: 1 hour 30 minutes)

Paper Reference **1ST0/1H**

Statistics

Paper 1
Higher Tier

You must have:

Ruler graduated in centimetres and millimetres, protractor,
pair of compasses, pen, HB pencil, eraser, scientific calculator.

Total Marks

Instructions

- Use **black** ink or ball-point pen.
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions.
- Answer the questions in the spaces provided – *there may be more space than you need.*
- Scientific calculators may be used.
- You must **show all your working out** with **your answer clearly identified** at the **end of your solution**.



Information

- The total mark for this paper is 80.
- The marks for **each** question are shown in brackets – *use this as a guide as to how much time to spend on each question.*

Advice

- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.

Turn over ►

S61445A

©2018 Pearson Education Ltd.

1/1/1/




Pearson

Higher Tier Formulae

You must not write on this page.

Anything you write on this page will gain NO credit.

$$\text{Skew} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

$$\text{Standard deviation} = \sqrt{\frac{1}{n} \sum (x - \bar{x})^2}$$

An alternative formula for standard deviation is

$$\text{standard deviation} = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

Spearman's rank correlation coefficient

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

Rates of change (e.g. Crude birth rate = $\frac{\text{number of births} \times 1000}{\text{total population}}$)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

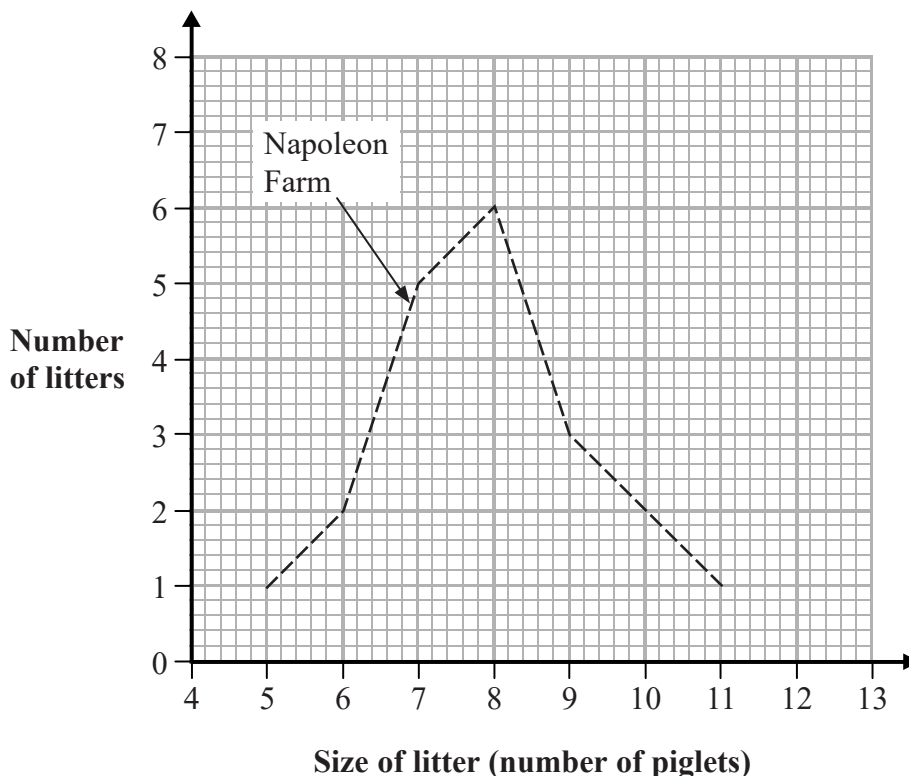
DO NOT WRITE IN THIS AREA

Answer ALL questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

- 1 The frequency polygon shows information about the size of litters (number of piglets) born to pigs on Napoleon Farm one year.



The table shows information about the size of litters born to pigs on Wilbur Farm for the same year.

Size of litter (number of piglets)	8	9	10	11	12
Number of litters	2	4	7	5	2

- (a) On the grid, draw a frequency polygon for the information about Wilbur Farm. (2)
- (b) Compare the distributions of the sizes of the litters for the two farms.

.....

.....

.....

.....

.....

(2)

(Total for Question 1 is 4 marks)



S 6 1 4 4 5 A 0 3 2 4

2 Samira wrote a questionnaire to investigate mobile phone use by the students at her school. She decided to select a simple random sample from the 850 students at her school.

(a) Explain what is meant by ‘random’ in this case.

(1)

(b) State the population for Samira’s investigation.

(1)

Samira obtained a list of all the students at her school, numbered 001 to 850, to use as a sampling frame.

Samira decided to select her sample using random numbers generated by her calculator.

Here are her first 16 random numbers.

317	310	940	753	323	471	236	046
053	310	729	593	812	398	797	243

Terri says these 16 random numbers will **not** give Samira 16 students for her sample.

(c) Explain why Terri is correct.

Give **two** reasons.

(2)

Terri suggests that Samira should select her sample by writing all the students’ names on pieces of paper, then picking them from a box without looking.

(d) Comment on whether Terri’s method is appropriate.

(2)

(Total for Question 2 is 6 marks)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

3 The table gives the average UK annual car insurance price for male drivers for each of the years 2016 and 2017

Year	2016	2017
Average car insurance price (£)	781	890

(Source: *www.confused.com*)

- (a) Using 2016 as the base year, calculate the index number for the average car insurance price for male drivers in 2017
Give your answer correct to the nearest whole number.

.....
(2)

The index number for the average UK annual car insurance price for female drivers for 2017 using 2016 as the base year is 113

- (b) Compare how the average UK annual car insurance prices changed from 2016 to 2017 for male drivers and for female drivers.

.....
(2)

(Total for Question 3 is 4 marks)



S 6 1 4 4 5 A 0 5 2 4

4 Claire buys packs of sports cards.

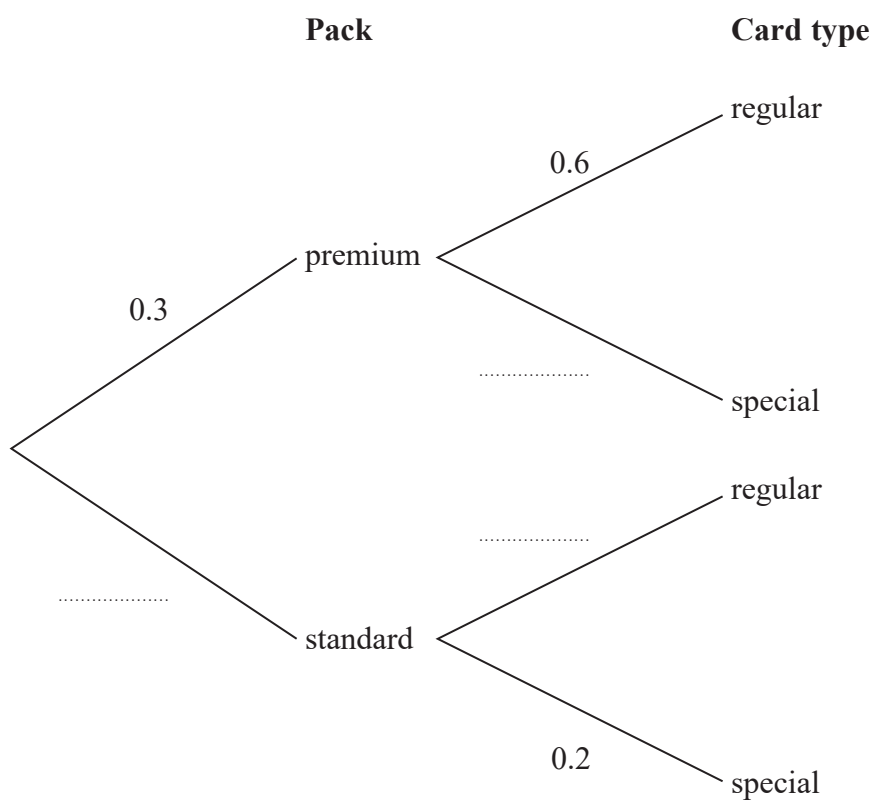
The cards can be bought in premium packs and in standard packs.

Of the packs that Claire buys, 30% are premium packs and 70% are standard packs.

In each premium pack there are 6 regular cards and 4 special cards.

In each standard pack there are 4 regular cards and 1 special card.

(a) Complete the probability tree diagram for this information.



(1)

Claire picks at random one of the packs she has bought, opens the pack and takes at random one card from the pack.

(b) Work out the probability that the card is a regular card.

.....

(3)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Given that the card is a regular card,

(c) work out the probability that it came from a premium pack.

.....
(2)

(Total for Question 4 is 6 marks)



- 5 The table and time series graph give information about the numbers of houses with a value of greater than £40 000 that were sold in Wales in the years 2014 to 2016

Year	Quarter	Number of houses sold	4-point moving average
2014	1	19 010	
	2	25 390	
	3	25 530	23 607.5
	4	24 500	23 662.5
2015			23 882.5
	1	19 230	24 427.5
	2	26 270	25 080
	3	27 710	26 130
	4	27 110	25 282.5
2016	1	23 430	25 212.5
	2	22 880	
	3	27 430
	4	25 710	

(Source: *www.gov.uk*)

The final 4-point moving average is missing from the table and from the time series graph.

- (a) (i) Calculate this missing 4-point moving average.
Write your answer in the table.

(2)

- (ii) Plot this 4-point moving average on the time series graph.

(1)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

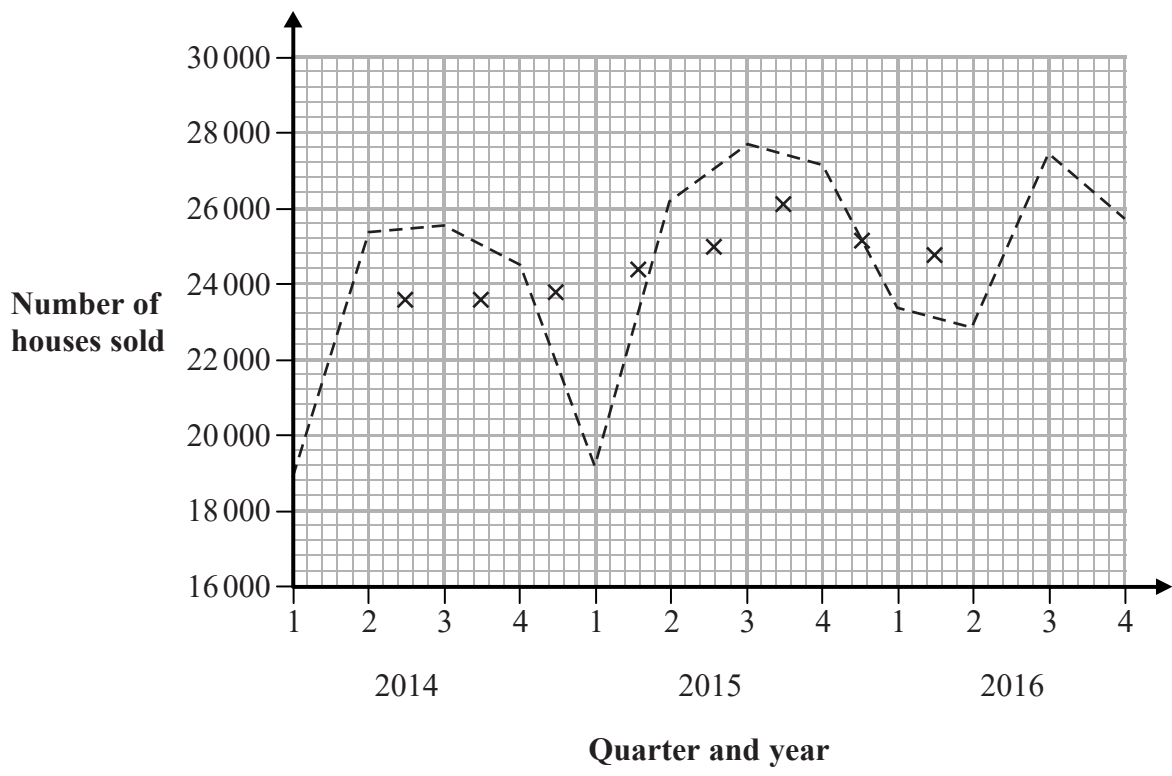
DO NOT WRITE IN THIS AREA



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



A newspaper wants to include this graph with an article on house sales in Wales.

(b) Why might this graph be misleading?

(1)

(c) For which quarter each year was the number of houses sold in Wales the greatest?

(1)

(d) Describe and interpret the trend shown by the time series graph.

(2)



Matthew draws a trend line on the time series graph.

He calculates the gradient of his trend line correctly.

Here is his calculation

$$\text{gradient} = 2600 \div 8 = 325$$

(e) Interpret this gradient in the context of the graph.

(1)

Matthew wants to use his trend line to predict the numbers of houses sold in Wales in 2017 by extrapolation.

(f) Give a danger associated with using extrapolation when making predictions.

(1)

(Total for Question 5 is 9 marks)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

6 David wants to compare the association between the shoulder height and the weight of male lions with the association between the shoulder height and the weight of female lions.

He has collected data from the internet for 30 male lions and for 30 female lions.

(a) Give a reason why it is appropriate to collect secondary data from the internet rather than to collect primary data in this case.

(1)

David plans to compare the associations by drawing comparative box plots.

(b) Explain why this is not appropriate for his investigation.

(1)

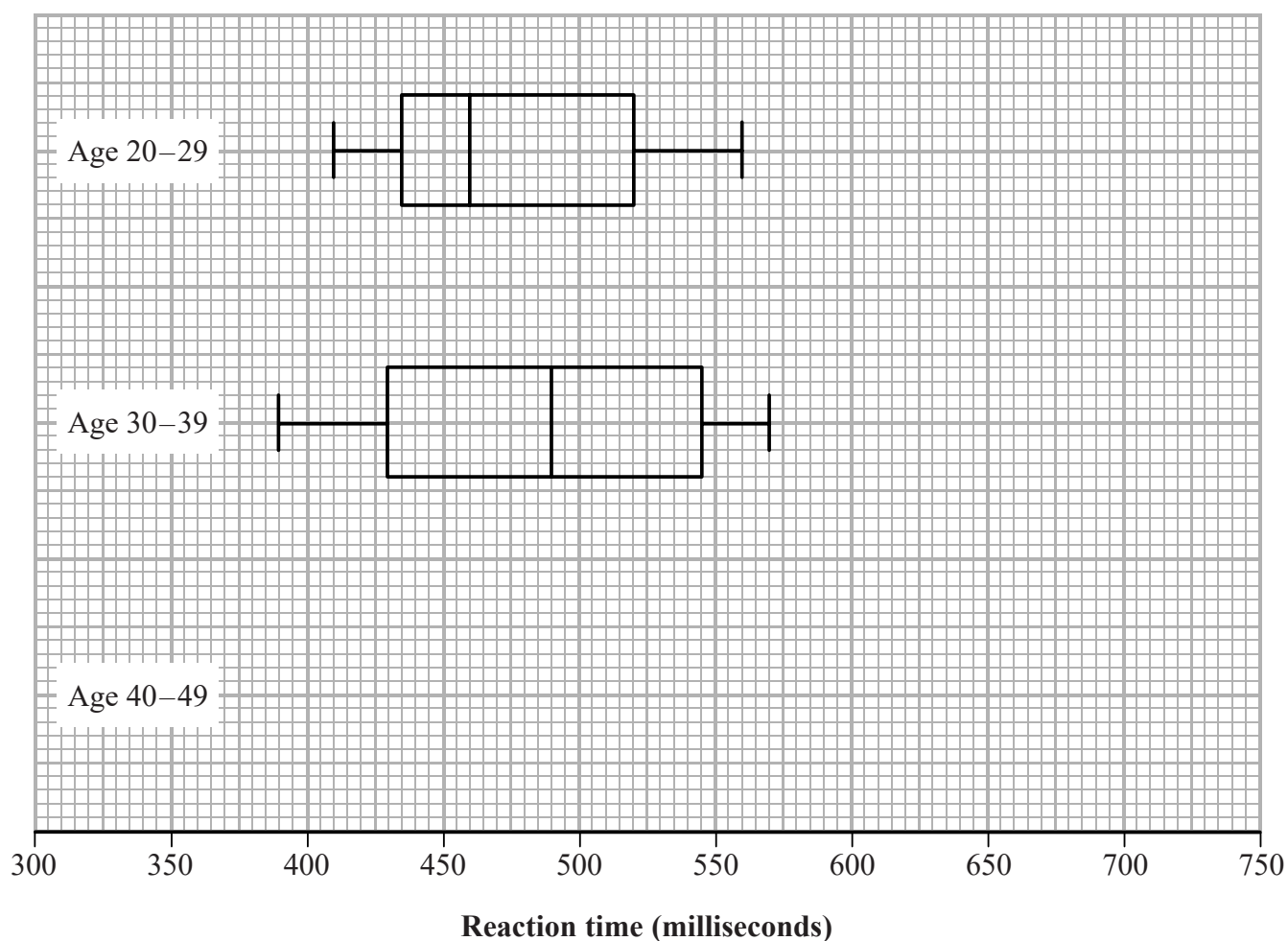
(c) Identify which diagrams and calculations David should plan to include in his report. Explain how these diagrams and calculations could be interpreted.

(6)

(Total for Question 6 is 8 marks)



- 7 The box plots show information about the reaction times, in milliseconds, for 20–29 year old men and for 30–39 year old men.



(Source: *journals.plos.org*)

The table below gives information about the reaction times, in milliseconds, for 40–49 year old men.

Least time	Lower quartile	Median	Upper quartile	Greatest time
430	481	533	579	725

- (a) On the grid above, draw a box plot for the information in the table.

(2)



(b) Compare the distributions of the reaction times.
Interpret one of your comparisons.

.....

.....

.....

.....

.....

.....

.....

.....

.....

(4)

Talia wants to investigate whether the amount of sleep has an effect on a reaction time.

She is going to compare reaction times following 2 hours sleep with reaction times following 8 hours sleep.

She plans to use matched pairs in order to collect her data.

(c) Comment on whether or not it is appropriate to use matched pairs.

.....

.....

.....

(2)

Talia wants to investigate further whether the amount of sleep has an effect on a reaction time.

Her variables will be age, amount of sleep and reaction time.

Here is a list of statistical words.

- rank quantitative categorical qualitative multivariate

(d) Circle the **two** words that describe the data that Talia will be collecting.

(1)

(Total for Question 7 is 9 marks)



- 8 Remi wants to find out about the number of people that avoid paying tax. He used a random response question to collect his data.

This is a suitable technique to use.

(a) Explain why.

(1)

Here is the random response question that Remi used on a questionnaire.

Roll a fair dice.

If you get **1, 2 or 3**, tick box A.
If you get **4, 5 or 6**, answer this question.

Have you ever avoided paying tax?

If **yes**, tick box A. If **no**, tick box B.

A B

Remi sent the questionnaire to a sample of people.

For this question,

426 people ticked box A

354 people ticked box B

A group of researchers also collected data on the number of people that avoid paying tax by directly asking people.

They found that 5.9% of the people they asked said that they had avoided paying tax.

(Source: www.emeraldinsight.com)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(b) Compare the results of Remi's questionnaire with the results obtained by the researchers.

You should consider whether the use of the random response technique was effective and you should give a limitation of your conclusion.

.....

.....

.....

.....

(5)

(Total for Question 8 is 6 marks)



9 Margaret and Paul are collecting data on turtles.

Margaret wants to estimate the number of turtles in a lake.

She catches a sample of 100 turtles from the lake.
She tags each turtle and then puts them back into the lake.

Three days later Margaret catches 60 turtles from the lake.
She finds that 12 of them have been tagged.

(a) Work out an estimate of the total number of turtles in the lake.

.....

(2)

Margaret wants to use this estimate in a report about the turtles in this lake.

(b) How reliable is this estimate?
Give reasons for your answer.

.....
.....
.....
.....
.....

(2)

Paul has collected data on 24 male turtles.

He has used statistical software to calculate the following summary statistics

$$\bar{x} = 102.58 \quad \sum x^2 = 256\,510$$

where the values of x are the breadths, in millimetres, of the shells of the turtles.

(Source: *www.public.iastate.edu*)

(c) Calculate, using Paul's summary statistics, the standard deviation of the breadths of the shells of the male turtles.

..... mm

(2)



Information about the breadth of the shells of female turtles was also recorded.

Mean	88.3 mm
Standard deviation	7.07 mm
Least breadth	74.0 mm
Greatest breadth	106 mm

(Source: *www.public.iastate.edu*)

Libby says that assuming that the breadths are normally distributed then there are no outliers for the breadths of the shells of female turtles.

(d) Is Libby correct?

You must show your working.

(4)

(Total for Question 9 is 10 marks)



10 Tina has collected data on the deadlifts, in kilograms, for some female powerlifters.

She has summarised her results in the table.

Age (years)	Mean deadlift (kilograms)	Frequency
20–29	134	30
30–39	146	15
40–49	125	25

(Source: www.openpowerlifting.org)

Tina wants to calculate the mean deadlift for all 70 women.

She thinks that she has three possible methods.

Method A Find the mean of 134, 146 and 125

Method B Work out the total of (class midpoint of age \times mean deadlift) for each class and divide by 70

Method C Work out the total of (mean deadlift \times frequency) for each class and divide by 70

Determine which of these calculations Tina should use.

Explain why the calculation is appropriate.

.....

.....

.....

.....

(Total for Question 10 is 2 marks)



- 11 The table gives the time that Lena took to complete successfully each of two skills tests, test A and test B.

The mean and the standard deviation for each of these tests for all the students who completed the tests are also given.

		All students	
	Lena	Mean	Standard deviation
Test A (minutes)	16.3	14.4	1.5
Test B (minutes)	21.6	19.8	2.4

- (a) Using standardised scores, compare Lena's performance in Test A relative to the other students with her performance in Test B relative to the other students.
Explain how you reach your conclusion.

(5)

The times taken to complete successfully each test are normally distributed.

- (b) Work out an estimate for the percentage of the times taken by the students to complete test B which are between 19.8 minutes and 22.2 minutes.

(2)

(Total for Question 11 is 7 marks)

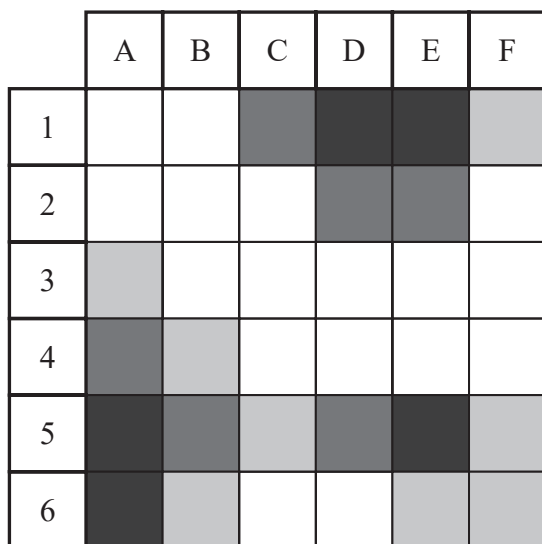


12 Hilary and Mika are investigating the distribution of bees in a nature reserve.

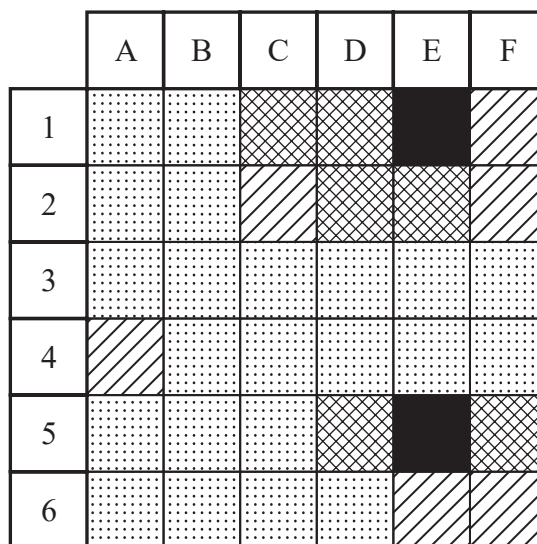
They have divided the nature reserve into 36 squares of equal size. The number of bees observed in each square in a fixed time period was recorded. The number of wildflowers observed in each square was also recorded.

The choropleth maps below are drawn using this information.

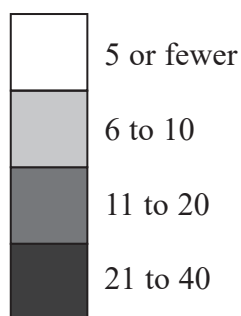
Bees



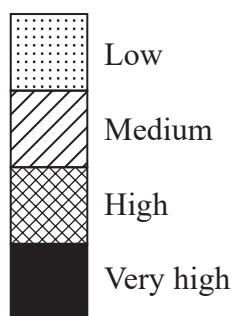
Wildflowers



Key: Number of bees



Key: Number of wildflowers



Hilary says that there are more bees in those regions of the nature reserve with greater numbers of wildflowers.

- (a) Does the information in the choropleth maps support her conclusion?
Give reasons for your answer.

.....

.....

.....

.....

(2)



Hilary and Mika are producing a report about the nature reserve.

They want to include information about the types of small mammals recorded in the nature reserve in one week in 2008 and in one week in 2018

Hilary thinks that they should use pie charts.

Mika thinks that they should use comparative pie charts.

(b) What advice would you give to Hilary and to Mika on their choice of diagram?

.....

.....

.....

.....

(2)

Mika is drawing his comparative pie charts.

He has the following information about the total number of small mammals recorded in one week in 2008 and in one week in 2018 in the nature reserve.

	2008	2018
Total number of small mammals	236	349

Mika is going to use a circle of radius of 5 cm for his pie chart for 2008

(c) Calculate the radius of the circle that Mika should use for his comparative pie chart for 2018
Give your answer correct to 1 decimal place.

..... cm

(2)

(Total for Question 12 is 6 marks)



13 Sarah has collected bivariate data. Her data has positive correlation for which Spearman's rank correlation coefficient is greater than Pearson's product moment correlation coefficient.

(a) Using the unscaled axes, sketch a possible scatter diagram for Sarah's data.



(1)

Nick has collected data on the amount of money spent on ski equipment and the time taken to complete a particular ski course.

He finds that for his data Pearson's product moment correlation coefficient is -0.65

(b) Which of the following conclusions is appropriate for Nick to make?

Tick (✓) the appropriate conclusion.

Spending more money on ski equipment will definitely reduce the time taken to complete the ski course

As the amount of money spent on ski equipment increases, the time taken to complete the ski course decreases

Both of the above

None of the above

(1)

(c) Give a reason for your answer to part (b).

.....

.....

.....

(1)

(Total for Question 13 is 3 marks)

TOTAL FOR PAPER IS 80 MARKS



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE

