Please check the examination details below before entering your candidate information

Candidate surname	Other names

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

Centre Number	Candidate Number

Tuesday 16 June 2020

Morning (Time: 1 hour 30 minutes)	Paper Reference **1ST0/2H**

Statistics

Paper 2
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 ▶

R62714A

1/1/1/1/

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 = $\frac{\text{number of births} \times 1000}{\text{total population}}$)

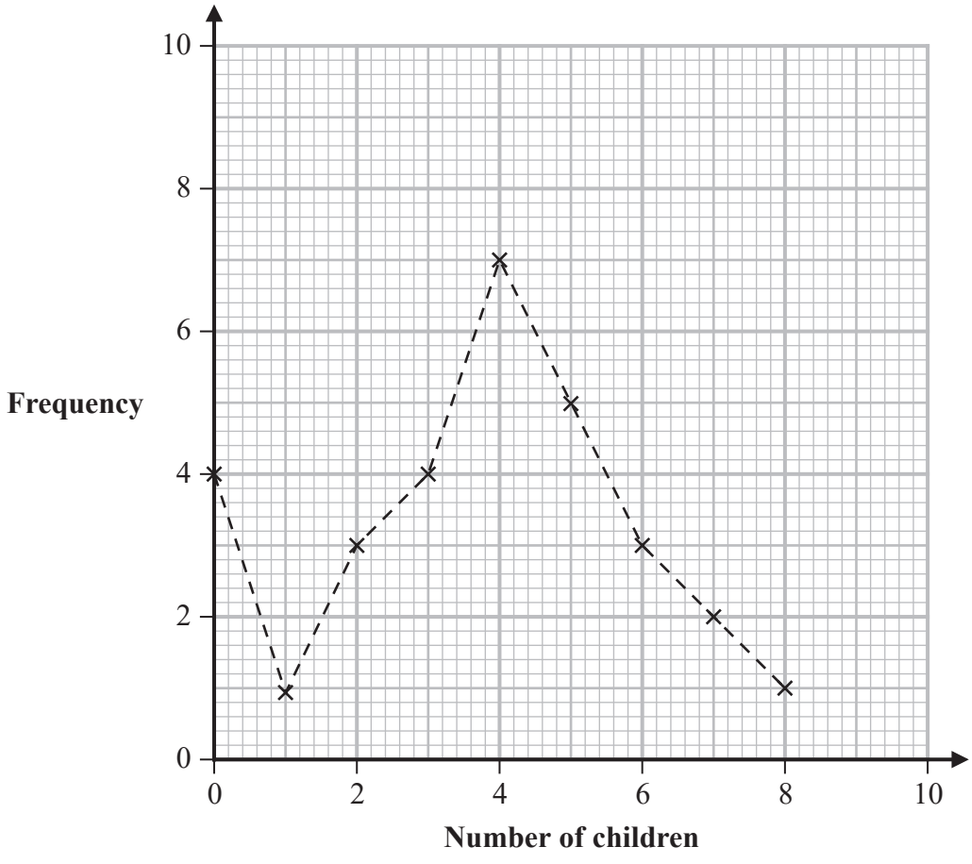
Answer ALL questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

1 The number of children living in each of 30 houses in a village was recorded in 1918 and in 2018

The frequency polygon has been drawn for the number of children recorded in 1918



The table shows information about the number of children living in each of the 30 houses in 2018

Number of children	0	1	2	3	4	5
Frequency	6	9	8	4	2	1

(a) On the grid, draw a frequency polygon for the information for 2018

(2)

(b) Compare the distributions of the number of children in these 30 houses in 1918 and in 2018

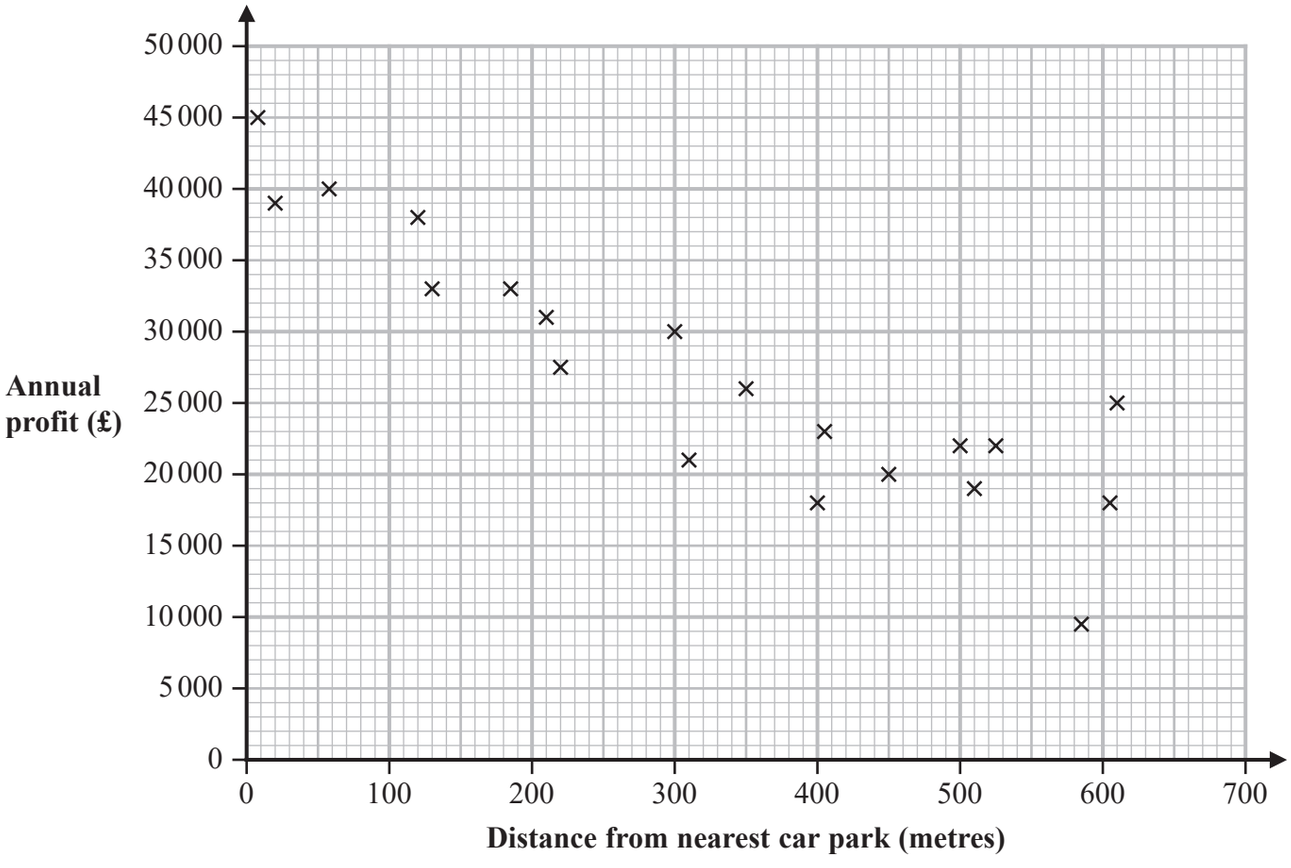
(2)

(Total for Question 1 is 4 marks)

2 Mike is investigating whether or not there is a relationship between the distance from the nearest car park to a restaurant and the annual profit made by the restaurant.

He finds out the distance, in metres, from the nearest car park to a restaurant and the annual profit, in £, made by the restaurant. He does this for twenty restaurants in a city.

Mike then draws a scatter diagram for this information.



(a) Explain why annual profit is the response variable for this scatter diagram.

(1)

Mike's hypothesis is that, for these restaurants, the further the restaurant is from its nearest car park the less is its annual profit.

(b) Explain, giving a statistical reason, whether or not the scatter diagram supports Mike's hypothesis.

(2)

Mike wants to draw a line of best fit on the scatter diagram. Using statistical software he obtains the following information about these restaurants.

Mean distance from the nearest car park	325 m
Mean annual profit	£27 000
Intercept of the line of best fit on the Annual profit axis	40 000

(c) (i) Using this information, draw a line of best fit on the scatter diagram.

(2)

(ii) Interpret the value of the intercept of the line of best fit on the Annual profit axis.

(1)

Restaurant A and restaurant B are two other restaurants in the city.

Restaurant A is 250 m from its nearest car park.
Restaurant B is 700 m from its nearest car park.

Mike uses the scatter diagram to find an estimate for the annual profit of each of these restaurants.

(d) Explain which of these two estimates will be the more reliable estimate.

(2)

Mike finds a positive correlation between the number of tables at a restaurant and its annual profit.

He concludes that as the number of tables increases this causes the annual profit to increase.

(e) Explain whether or not this conclusion is valid.

(1)

Mike reads an article in a newspaper that says that restaurant profits for the top 100 restaurants had fallen from £345 million to £125 million in the past year.

(Source: *telegraph.co.uk*)

(f) Using the data in the newspaper article, calculate the percentage decrease in restaurant profits for the top 100 restaurants in the past year.

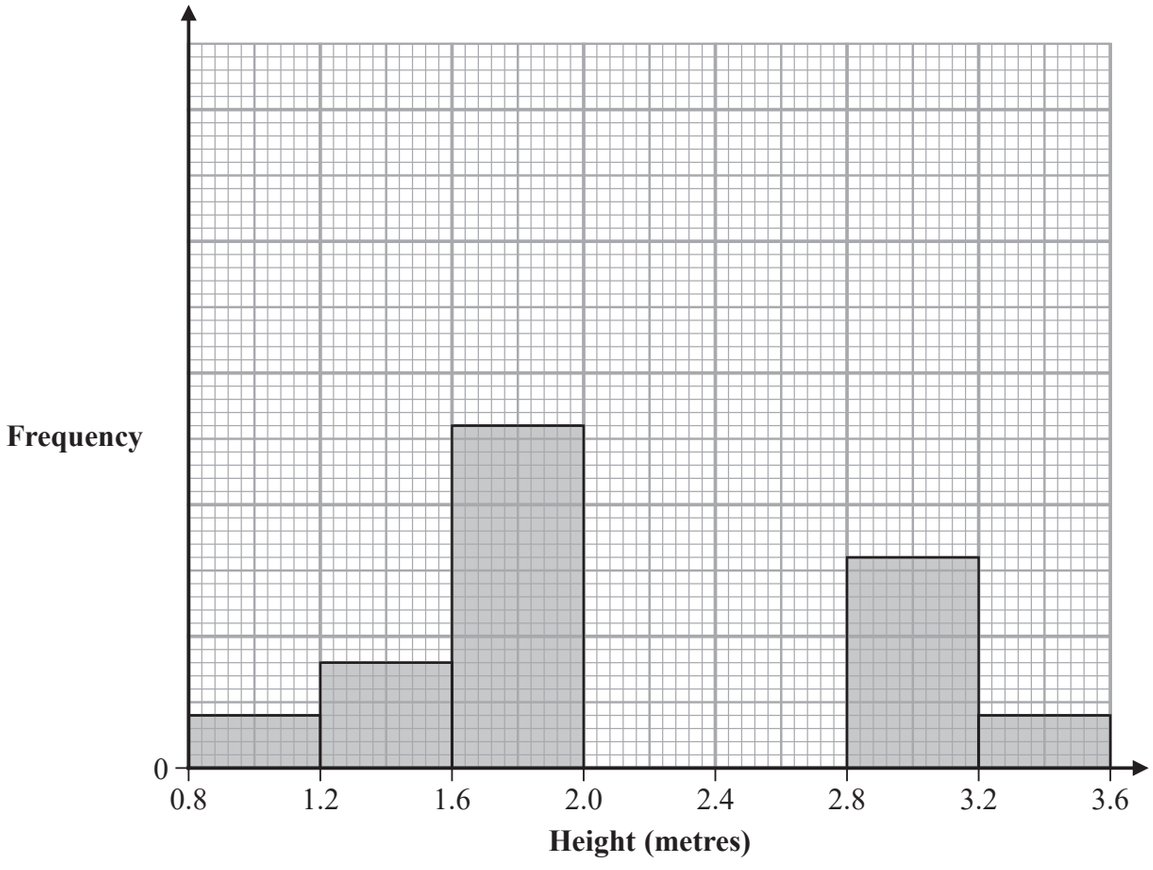
.....%

(1)

(Total for Question 2 is 10 marks)

3 The Forestry Commission planted Field Maple trees and Silverleaf Maple trees in region A. They measured the heights of the trees after 14 years.

The incomplete histogram and incomplete grouped frequency table give information about the heights, in metres, of the Field Maple trees in region A.



Height (h metres)	Frequency
$0.8 < h \leqslant 1.2$	2
$1.2 < h \leqslant 1.6$	4
$1.6 < h \leqslant 2.0$	13
$2.0 < h \leqslant 2.4$	25
$2.4 < h \leqslant 2.8$	10
$2.8 < h \leqslant 3.2$	
$3.2 < h \leqslant 3.6$	

(Source: *data.gov.uk*)

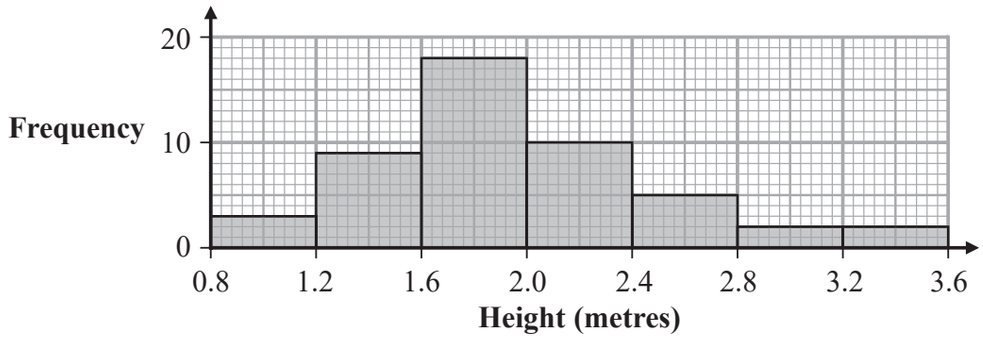
(a) Use the information in the histogram to complete the table.

(2)

(b) Use the information in the table to complete the histogram.

(2)

The histogram below gives information about the heights of the Silverleaf Maple trees after 14 years.



(c) Identify and interpret the type of skew shown in the histogram for Silverleaf Maple trees.

(2)

The Forestry Commission also planted Field Maple trees and Silverleaf Maple trees in region B.
The grouped frequency table below gives information about the heights of these trees.

	Frequency	
Height (h centimetres)	**Field Maple**	**Silverleaf Maple**
$0 < h \leqslant 80$	1	1
$80 < h \leqslant 160$	14	4
$160 < h \leqslant 240$	27	32
$240 < h \leqslant 320$	21	13
Total	**63**	**50**

The estimate of the mean for Field Maple trees is calculated to be 206.3 cm to 1 decimal place.

David thinks that the estimate of the means for Field Maple trees and for Silverleaf Maple trees suggests that Field Maple trees are taller than Silverleaf Maple trees.

(d) Is David correct?
You must show your working.
Give one limitation of your conclusion.

(4)

(Total for Question 3 is 10 marks)

4 Matthew is collecting information about road safety in his town.

He wants to collect information from drivers who live in his town.

Matthew plans to ask drivers at his local car park about their views on road safety.
He plans to collect a quota sample of 20 drivers from each of three age groups.

Matthew's three age groups of drivers are

39 years old or younger 40 years old to 59 years old 60 years old or older

(a) Comment on whether Matthew's plans are appropriate.

(2)

Matthew wants to know how many drivers, who live in his town, drive faster than the speed limit on the motorway.
Matthew knows that he will have to ask a sensitive question so he plans to use the random response technique to find out this information.

(b) Design a random response question that Matthew could use on a questionnaire in order to collect this information.

(3)

(Total for Question 4 is 5 marks)

5 A supermarket manager is planning the design of a new petrol station.

He wants to use the results of spinning a coin to simulate on which side of a car the fuel filler cap is.
He assumes that the fuel filler cap is equally likely to be on the left of the car as it is to be on the right of the car.
He wants to use the simulation in order to predict how many cars will arrive at the petrol station before there are four consecutive cars with filler caps on the same side.

(a) Explain how he could use a coin to carry out the simulation.

(2)

A new cafe is also opening at the supermarket.
Three different types of hot drink – tea, coffee and hot chocolate – will be sold in the cafe.

The table gives, for each type of hot drink, the fraction of the total number of hot drinks sold in other cafes of this supermarket chain.

Tea	Coffee	Hot chocolate
$\frac{2}{8}$	$\frac{5}{8}$	$\frac{1}{8}$

The manager wants to predict how many of each type of hot drink the cafe will sell in an hour the first Saturday the cafe is open.

The manager expects that there will be 40 customers in the cafe that hour.

He assigns the numbers shown in the table below to each type of hot drink.

Type of hot drink	Tea	Coffee	Hot chocolate
Numbers	0, 1, 2	3, 4, 5	6, 7, 8

He will then use a calculator to generate 40 random numbers between 0 and 9 to simulate 40 customers.

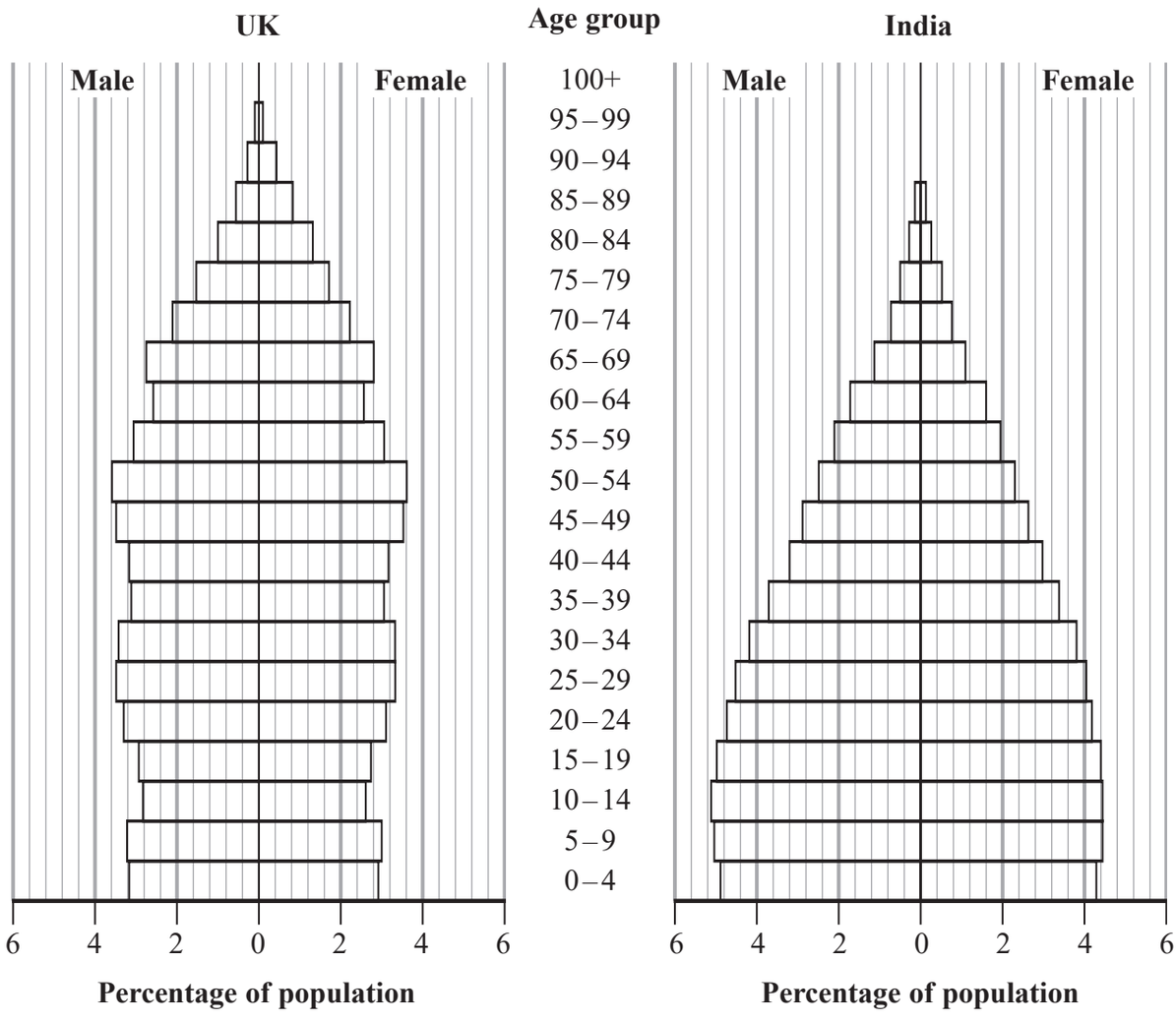
(b) Comment on the suitability of the manager's plan.

(3)

(Total for Question 5 is 5 marks)

6 The two population pyramids show the percentage of males and the percentage of females in each age group in the UK and in India for 2016

Each pyramid is based on the total population of that country for 2016



(Source: *www.populationpyramid.net*)

(a) Compare the percentages of males and females in the UK who are 24 years old or younger with the percentages of males and females in India who are 24 years old or younger.
You should give two comparisons.

(2)

The population of the UK in 2016 was 65 648 000 and the number of live births to women aged 15 to 49 was 775 300

(Source: *ons.gov.uk*)

(b) Calculate the general fertility rate for the UK in 2016

You may use

$$\text{general fertility rate} = \frac{\text{total number of live births in the year to women aged 15 to 49}}{\text{number of women aged 15 to 49}} \times 1000$$

.....

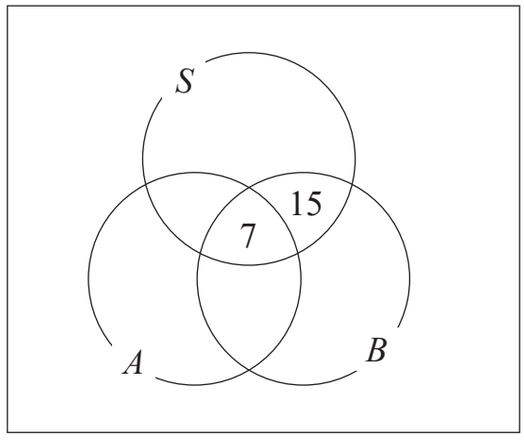
(4)

(Total for Question 6 is 6 marks)

7 The manager of a garage records optional extra features sold with each new car sold last month. The optional extra features are air conditioning (A), bluetooth (B) and sat nav (S).

Last month,
80 cars were sold
7 cars were sold with sat nav and bluetooth and air conditioning
15 cars were sold with sat nav and bluetooth only
21 cars were sold with sat nav and air conditioning only
10 cars were sold with bluetooth and air conditioning only
8 cars were sold with sat nav only
33 cars in total were sold with bluetooth
44 cars in total were sold with air conditioning.

(a) Complete the Venn diagram using this information.



(3)

Inge picks at random a car that was sold last month.

She thinks that the probability that the car has bluetooth given that it has sat nav is greater than the probability that the car has bluetooth given that it has not got sat nav.

(b) Is Inge correct?
You must show your working.

(5)

Second-hand cars are also sold at the garage.
The manager records whether the second-hand cars have a full service history or not.
She also records whether each second-hand car has a breakdown or not in the first year after being sold.

The table gives information about the second-hand cars sold at the garage in 2018

	Number of cars sold	Number of cars breaking down in the first year
With a full service history	440	22
Without a full service history	710	71

Paul says that the relative risk of a second-hand car without a full service history having a breakdown in the first year after being sold, compared with a second-hand car with a full service history having a breakdown in the first year after being sold, is 2

(c) (i) Show that Paul is correct.

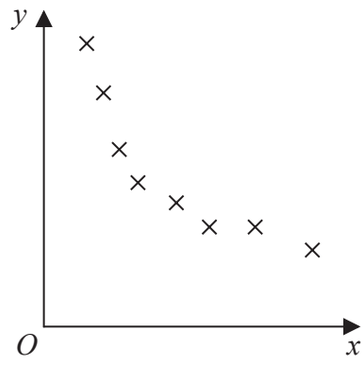
(2)

(ii) Interpret a relative risk of 2 in this context.

(1)

(Total for Question 7 is 11 marks)

8 Here is a scatter diagram.



The Spearman's rank correlation coefficient and the Pearson's product moment correlation coefficient are each going to be calculated for the data shown in the scatter diagram.

(a) How would you expect the values of these two correlation coefficients to compare?

(2)

The Australian Government collected data on crop yields from fields. The data collected were the yield of wheat grown in a field, in tonnes per hectare (t/ha), and the yield of other crops grown in the same field (t/ha) in a different year.

The table gives the Pearson's product moment correlation coefficient for the wheat yield and the barley yield and for the wheat yield and the oats yield. The table also gives the equation of the regression line in each case.

Explanatory variable (x)	**Response variable (y)**	**Pearson's product moment correlation coefficient**	**Regression equation**
Wheat yield (t/ha)	Barley yield (t/ha)	0.79	$y = 1.24x - 0.30$
Wheat yield (t/ha)	Oats yield (t/ha)	0.51	$y = 1.52x - 1.05$

(Source: *grdc.com.au*)

(b) Compare the relationship between wheat yield and barley yield with the relationship between wheat yield and oats yield.

You should refer to both correlation coefficients and to the equations of both regression lines in your comparison.

(6)

This year Louise has a field planted with wheat. She wants to use the information in the table to decide whether to plant barley or to plant oats in the field next year. She plans to plant the crop which is predicted to give the greatest yield.

Louise solves the equation $1.24x - 0.30 = 1.52x - 1.05$

She gets the answer $\frac{75}{28}$

(c) (i) Show that the answer Louise gets is correct.

(2)

(ii) Explain how Louise could use her answer to decide which of barley or oats to plant next year.

(2)

(iii) Give a limitation of the data that could affect the result of Louise's decision.

(1)

(Total for Question 8 is 13 marks)

9 A company manufactures tablets of medication X.

The tablets have a target mass of medication X of 200 mg.

The company uses quality assurance to monitor the mass of medication X in each tablet.

Samples of the tablets are taken from the production line at regular intervals and the mean mass of medication X in the tablets in each sample is found.

The sample means should be normally distributed with a mean of 200 mg and a standard deviation of 2.5 mg.

(a) Find the upper action limit for the sample means for medication X.

..... mg

(2)

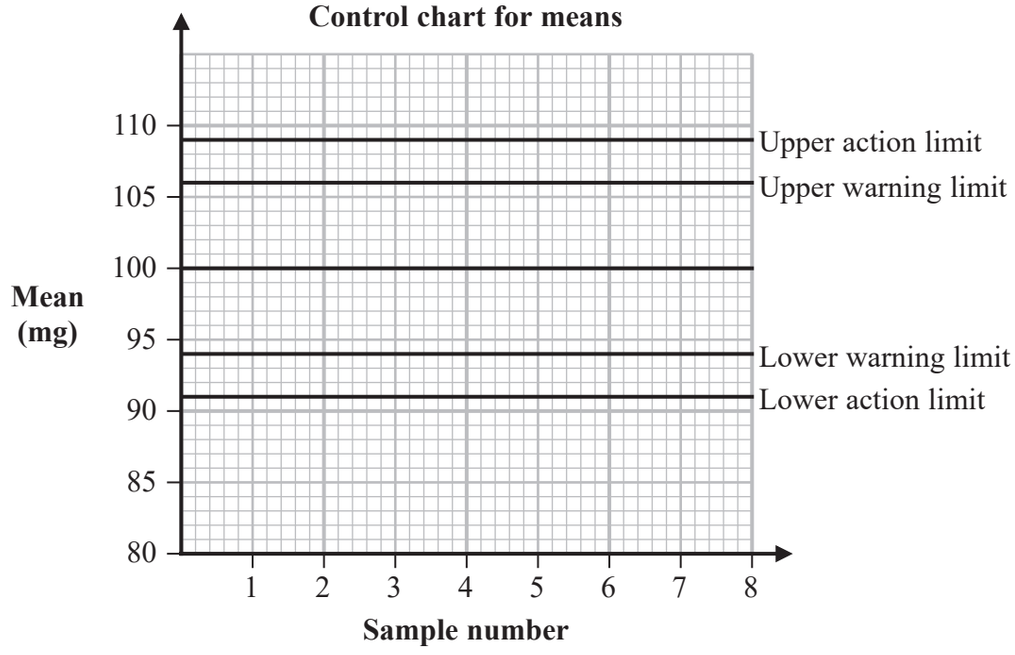
The pharmacist in charge of monitoring the production of the tablets of medication X wants to set the upper action limit closer to the target mass of 200 mg.

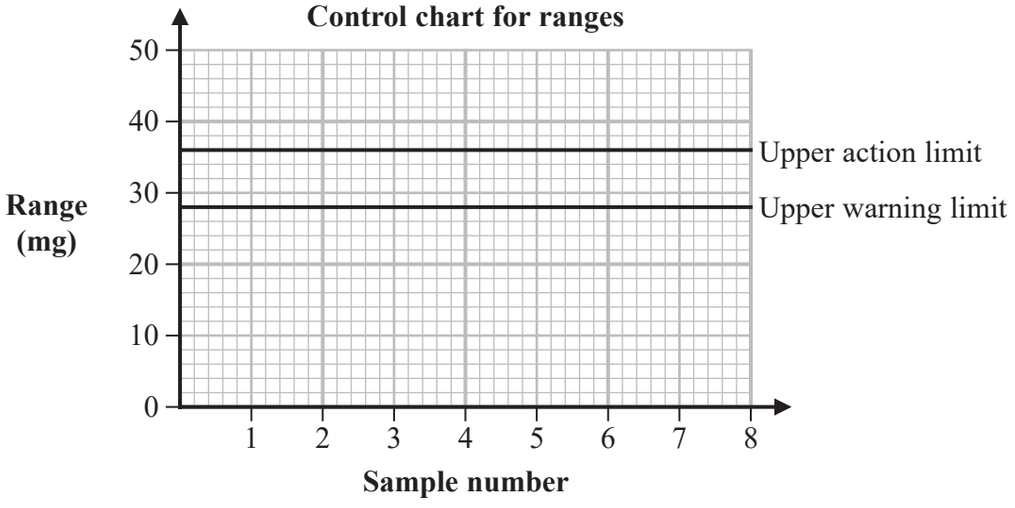
(b) What effect would you expect this to have on the number of times the production process may need to be stopped?

(1)

The company also manufactures tablets of medication Y and the company uses quality assurance to monitor the mass of medication Y in the tablets.

Here are the control charts for the sample means and for the sample ranges of the mass of medication Y in the tablets.





A sample of tablets of medication Y is taken.

The mean of the sample is 107 mg.
The range of the sample is 39 mg.

(c) Determine what action, if any, should be taken.
Justify your answer by referring to both the sample mean and the sample range.

(2)

(Total for Question 9 is 5 marks)

10 Kate is investigating the lengths of earthworms.
She wants to find out if the lengths of common earthworms can be modelled by a normal distribution.
She has found out that the greatest length of the common earthworm is 36 cm.

(Source: *www.nationalgeographic.com*)

Kate plans to group the data that she will collect in the table shown below.

Length (l cm)	Frequency
$0 < l \leqslant 5$	
$5 < l \leqslant 10$	
$10 < l \leqslant 20$	
$20 < l \leqslant 30$	
$30 < l \leqslant 35$	
$35 < l \leqslant 40$	

(a) Comment on whether Kate's choice of class intervals in the grouped frequency table is appropriate.

(2)

Bien is researching the lengths of time, in days, lived by flies.
He wants to know if the lengths of time lived by flies can be modelled by a normal distribution.

He has collected data for 80 flies.

The table gives information about the length of time lived by each fly.

Length of time lived (days)	Frequency
0–5	5
6–10	8
11–15	12
16–17	13
18–19	18
20–21	15
22–25	9

Here is Bien's plan to test whether the lengths of time lived by flies can be modelled by a normal distribution.

> Using the data collected I will draw a histogram. To do this I will need to calculate the frequency densities. Here is an example of a frequency density calculation for my data:
>
> Class interval 6–10
> Frequency 8
>
> $$\text{Frequency density} = \frac{8}{10 - 6} = 2$$
>
> I will then use the mean and standard deviation of the data to work out the amount of data within 1 standard deviation of the mean and the amount of data within 2 standard deviations of the mean.

(b) Comment on whether Bien's plan is appropriate.

(4)

(Total for Question 10 is 6 marks)

11 Bella collected the finish times, x minutes, of the men's elite wheelchair race at the 2018 London Marathon.

She uses statistical software to calculate the following summary statistics.

Number of athletes = 39 Median = 97 minutes

$\sum x = 4171$ $\sum x^2 = 469\,657$

(Source: *http://results-2018.virginmoneylondonmarathon.com*)

Calculate the skew for the distribution of the finish times and interpret this value in the context of Bella's data.

(Total for Question 11 is 5 marks)

TOTAL FOR PAPER IS 80 MARKS