Candidate surname | Other names

Centre Number | Candidate Number

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

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 ▶

P68676A

Pearson

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

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

An alternative formula for standard deviation is

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

Spearman's rank correlation coefficient

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

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

P68676A0220

Answer ALL questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

1 Tomasz found information about the weather on the Isle of Skye for the 31 days of May 2019 and the 31 days of July 2019

The table gives, for the Isle of Skye, some information about the number of days with rainfall during May.

Month	May	July
Absolute risk of rainfall on a day	0.452	

(Source: *www.isleofskyeweather.co.uk*)

(a) Using the information in the table, find the number of days with rainfall in May 2019
You must show your working.

.... days

(2)

The relative risk of rainfall for a day in July compared with a day in May is 1.29

(b) Find the number of days with rainfall in July 2019
You must show your working.

.... days

(2)

(Total for Question 1 is 4 marks)

2 Bill investigated the heights of females competing in different sports.

Using data from the internet, he recorded the height, in cm, of each female in a sample of 15 female rugby players and the height, in cm, of each female in a sample of 15 female basketball players.

He drew the box plot below for the recorded heights of the 15 female basketball players.

Rugby

Basketball

150 155 160 165 170 175 180 185 190 195
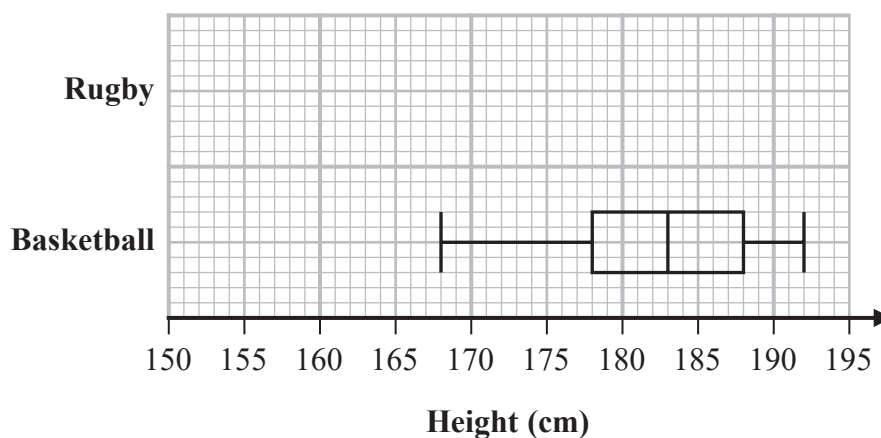
Height (cm)

(Source: *englandrugby.com, gbbasketball.com*)

The table gives information about the recorded heights of the 15 female rugby players.

Greatest height	182 cm
Median height	170 cm
Lower quartile	166 cm
Range	26 cm
Interquartile range	10 cm

(a) Using the information in the table, draw on the grid above a box plot for the recorded heights of the 15 female rugby players.

(3)

www.mrc-papers.com 55258711

P 6 8 6 7 6 A 0 4 2 0

(b) Compare the two distributions of heights.
Give **three** comparisons and interpret one of your comparisons.

....

....

....

....

....

....

....

(4)

(c) Considering the method that Bill used to collect his data, comment on the reliability
of your comparisons in part (b).

....

....

....

....

(2)

(Total for Question 2 is 9 marks)

3 Irina is investigating whether the percentage of the population of a country living in urban areas has an effect on the life expectancy in that country.

(a) Suggest a hypothesis Irina could use for her investigation.

....

....
(1)

Irina collected the following information about 10 countries.

The Urban population (%) refers to the percentage of the population of the country who live in urban areas.

Country	Urban population (%)	Life expectancy (years)
Australia	89	82
Bangladesh	34	73
China	56	76
Egypt	43	73
France	80	82
Macedonia	57	76
Norway	81	82
Panama	67	79
Thailand	50	75
USA	82	80

(Source: *CIA World Factbook*)

Irina used statistical software to draw a scatter diagram for the information in the table.

(b) Give a reason why a scatter diagram is an appropriate diagram to use.

....

....
(1)

(c) For this investigation, which variable is the explanatory variable?
Give a reason for your answer.

....

....

....
(2)

The scatter diagram from the statistical software is shown below.
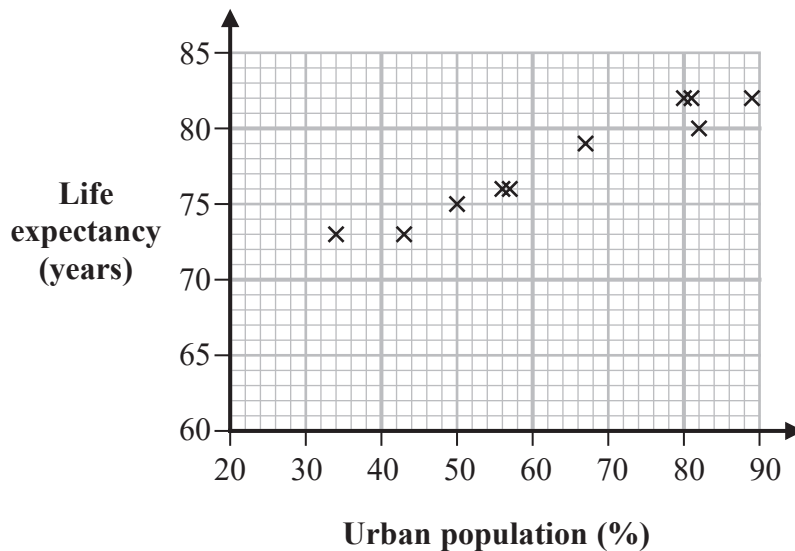
(d) Explain, giving a statistical reason, whether or not this scatter diagram supports your hypothesis in part (a).

....

....

(2)

For these 10 countries, the double mean point of the data is (63.9, 77.8).

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

(2)

Using statistical software, Irina finds that the gradient of the line of best fit should be 0.19

(f) Interpret the gradient of the line of best fit.

....

....

(1)

Irina now finds that South Africa has Urban population 65% and Life expectancy 63 years.

(g) Determine how this information for South Africa fits with the relationship shown in the scatter diagram for the other countries.

....

....

....

(2)

(Total for Question 3 is 11 marks)

4 A hospital department has three specialist consultants, **A**, **B** and **C**.

The department manager uses 2-digit random numbers to carry out a simulation of the number of consultations per week for each specialist.

According to hospital records, the average number of consultations per week for each specialist is shown in the following table.

Specialist	A	B	C
Average number of consultations	16	24	10

The department manager allocates the following 2-digit numbers to each specialist.

Specialist	A	B	C
2-digit numbers	00 to 31	32 to 79	80 to 99

(a) Explain why this is an appropriate way for the department manager to allocate the 2-digit numbers.

....

....

....

(1)

Also, according to hospital records, the expected total number of consultations in a week for all three specialists is 50

The table below shows the first 45 results from the simulation.

Specialist	A	B	C
Simulated number of consultations	18	19	8

The last five random numbers used by the department manager are

<div align="center">06   74   59   14   29</div>

(b) Using these last five numbers, complete the table below to show the final result of the simulation.

Specialist	A	B	C
Simulated number of consultations			

(1)

Each consultation is planned to last no longer than 20 minutes.

(c) According to the simulation, how much more time a week should be allocated to specialist **A** than to specialist **C** for all of their consultations?

.... minutes

(1)

The department manager decides that the results of her simulation are sufficient for her to plan how much time is needed per week for consultations with each specialist.

(d) Explain why the manager is **not** correct.

....

....

....

....

....

(2)

(Total for Question 4 is 5 marks)

5 An adventure travel company runs expeditions for groups of 20 people.

The company provides all the food and equipment for these expeditions.

The manager of the company wants to trial a new sports drink to see if it is better than the usual sports drink.

He decides on the following plan.

* Before the expedition each of the 20 people will complete a questionnaire about their fitness and energy levels.
* The 20 people will be divided into two groups of 10 people, a test group and a control group.
* The test group will be given the new sports drink during the expedition.
* The control group will be given the usual sports drink during the expedition.
* After the expedition each of the 20 people will again complete a questionnaire about their fitness and energy levels.
* The energy levels of the two groups after the expedition will be compared.

(a) Explain briefly why a control group should be used.

....

....

(1)

(b) Explain how the manager could use matched pairs in order to select the people in the two groups.

....

....

....

(1)

(c) Assess the reliability of any conclusions that the manager may reach as a result of using his plan.

Give **two** reasons for your answer.

....

....

....

....

....

....

(3)

(Total for Question 5 is 5 marks)

6 Paul uses the Petersen capture recapture method in order to find an estimate for the number of trout living in a lake.

To do this, Paul catches an initial sample of 50 trout from the lake and he tags each trout. He then releases the trout back into the lake.

For his second sample, Paul catches a sample of 80 trout from the lake. He finds that 16 of these trout have tags.

(a) Explain what needs to be true about his two samples for them to be valid for the capture recapture method.

....

....

(1)

Using his results, Paul is able to work out an estimate for the number of trout living in the lake.

(b) Find Paul's estimate for the number of trout and discuss the validity and the reliability of his estimate.

....

....

....

....

....

....

....

....

(5)

**(Total for Question 6 is 6 marks)**

P 6 8 6 7 6 A 0 1 1 2 0

7 Andy is investigating the lateness of the trains that arrive at Manchester Piccadilly.

(a) Explain why it is sensible for Andy to collect secondary data for his investigation.

....

....

(1)

Andy found out data on the lateness of each of 238 trains arriving at Manchester Piccadilly from London on Monday to Friday for one week.

The histogram below was drawn using Andy's data.
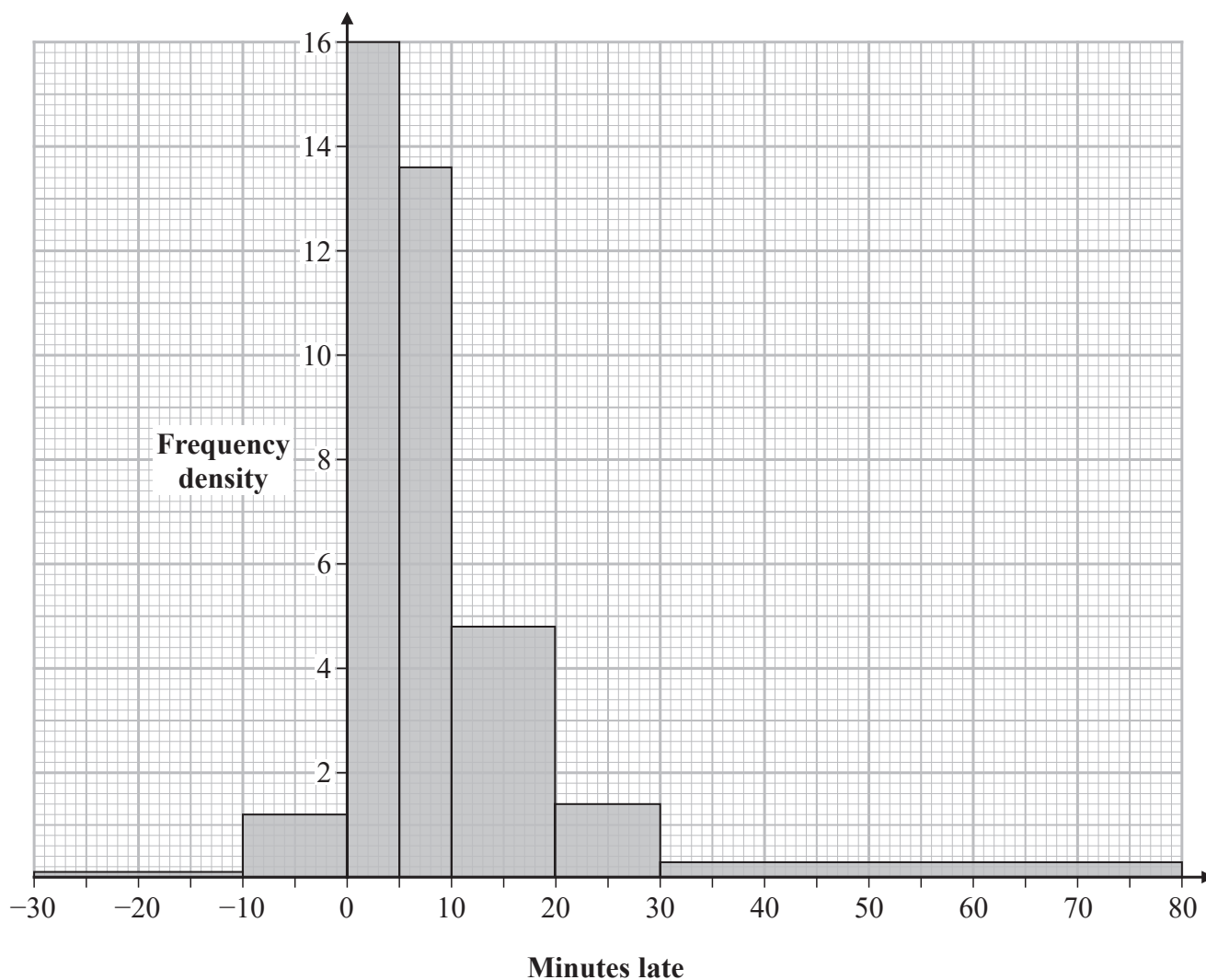
(Source: *www.recenttraintimes.co.uk*)

P 6 8 6 7 6 A 0 1 2 2 0

A negative time for a train means that the train arrived early at Manchester Piccadilly.

In the week, 14 trains arrived early.

(b) Find how many of the 238 trains did **not** arrive more than 10 minutes late at Manchester Piccadilly.

....

(3)

For a long distance journey, such as London to Manchester Piccadilly, a train is only considered to be late when it arrives more than 10 minutes late.

Andy says that more than 1 in 3 trains from London arrive late at Manchester Piccadilly.

(c) Determine whether or not Andy is correct.

You should comment on the reliability of your conclusion.

....

....

....

....

(4)

(Total for Question 7 is 8 marks)

8 An engineering production line is set up to produce machine parts with a diameter of 24.50 mm.

For quality control, random samples of size 4 are taken and the mean diameter of each sample is calculated.

The production line is set so that the sample means should be normally distributed with a mean of 24.50 mm and a standard deviation of 0.005 mm.
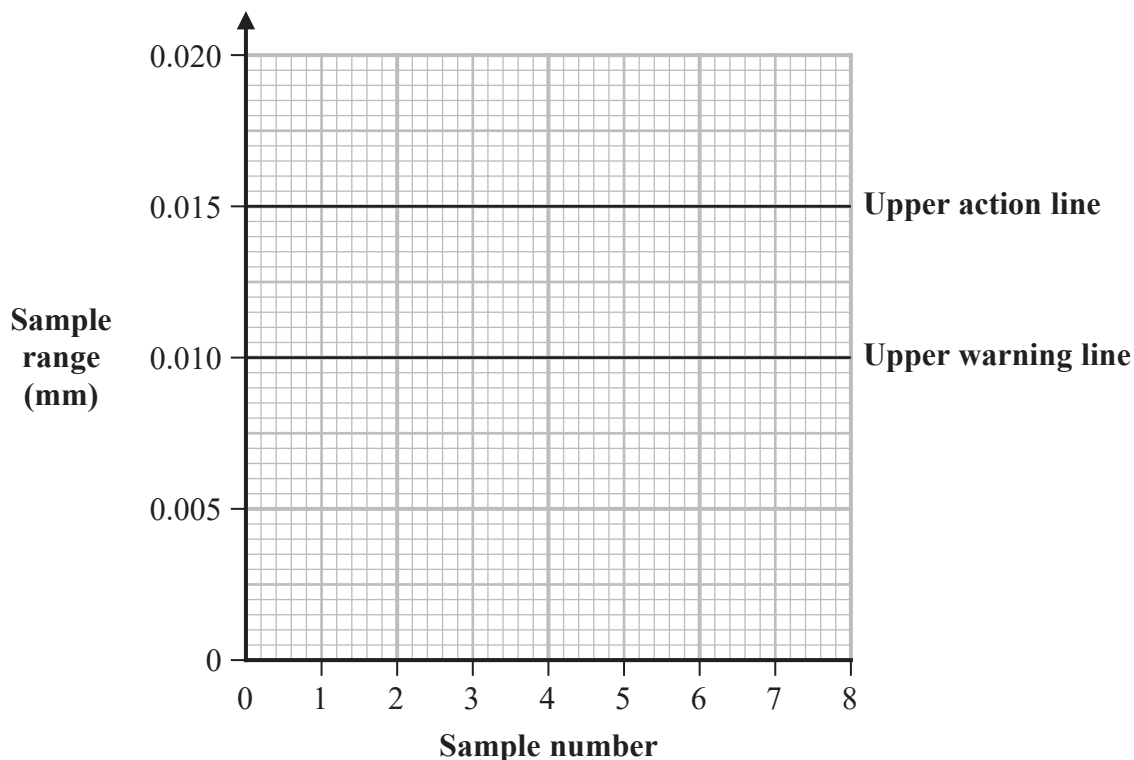
(a) Using this information, draw warning lines on the control chart for the sample means.

Action lines have already been drawn on the chart.
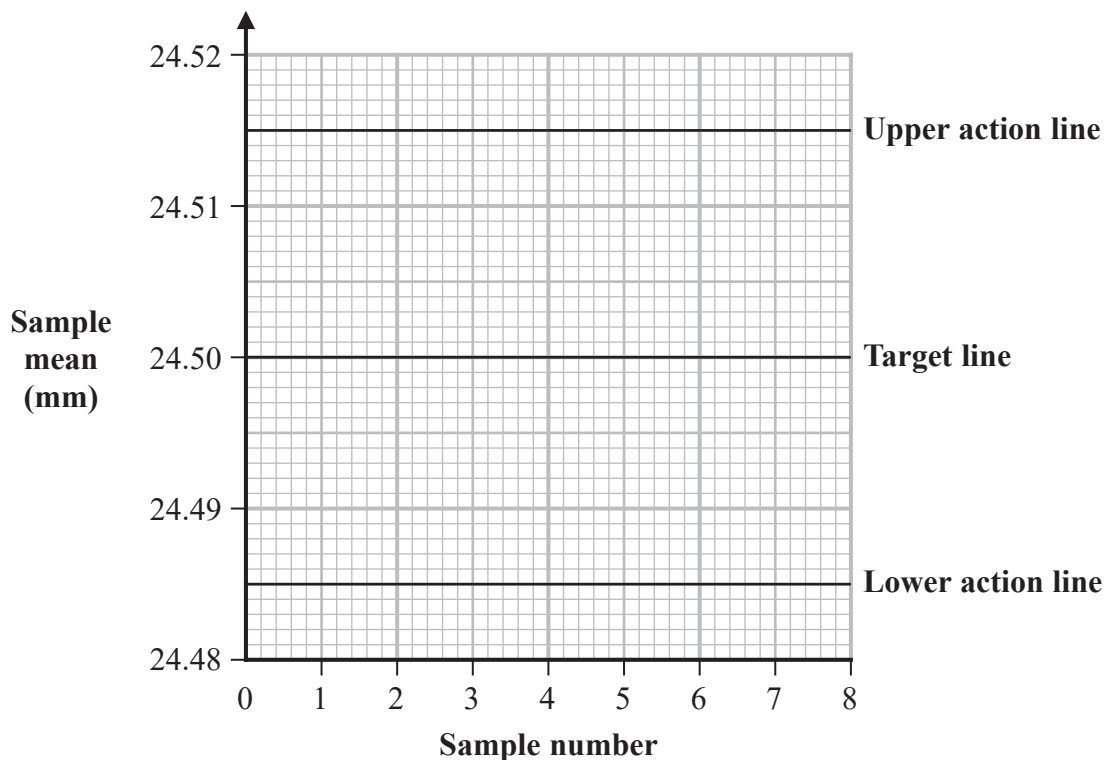
(2)

A control chart for the sample ranges is also used and is shown below.

www.mrc-papers.com
55258711

P 6 8 6 7 6 A 0 1 4 2 0

(b) Explain why it is not appropriate to have lower warning lines and lower action lines on the control chart for sample ranges.

....

....

(1)

The first two samples taken have the following summary statistics.

	Sample 1	Sample 2
Sample mean (mm)	24.502	24.496
Sample range (mm)	0.006	0.008

(c) Plot the summary statistics for these two samples on the control charts.

(2)

The machine parts in Sample 3 had the following diameters:

24.498 mm   24.492 mm   24.501 mm   24.509 mm

(d) Use these results to complete **both** control charts for Sample 3

(4)

(e) Describe what action needs to be taken after Sample 3

....

....

(1)

(Total for Question 8 is 10 marks)

Turn over ▶

9 Three types of sausage are sold in a farm shop, pork, beef and vegan.

The comparative pie charts show information about the number of packs of these sausages sold in the farm shop in May 2015 and in May 2019

The pie charts are drawn accurately and the angles for the vegan sectors are shown.

May 2019

May 2015

60°

90°

Key:
■ Pork   ■ Beef   □ Vegan
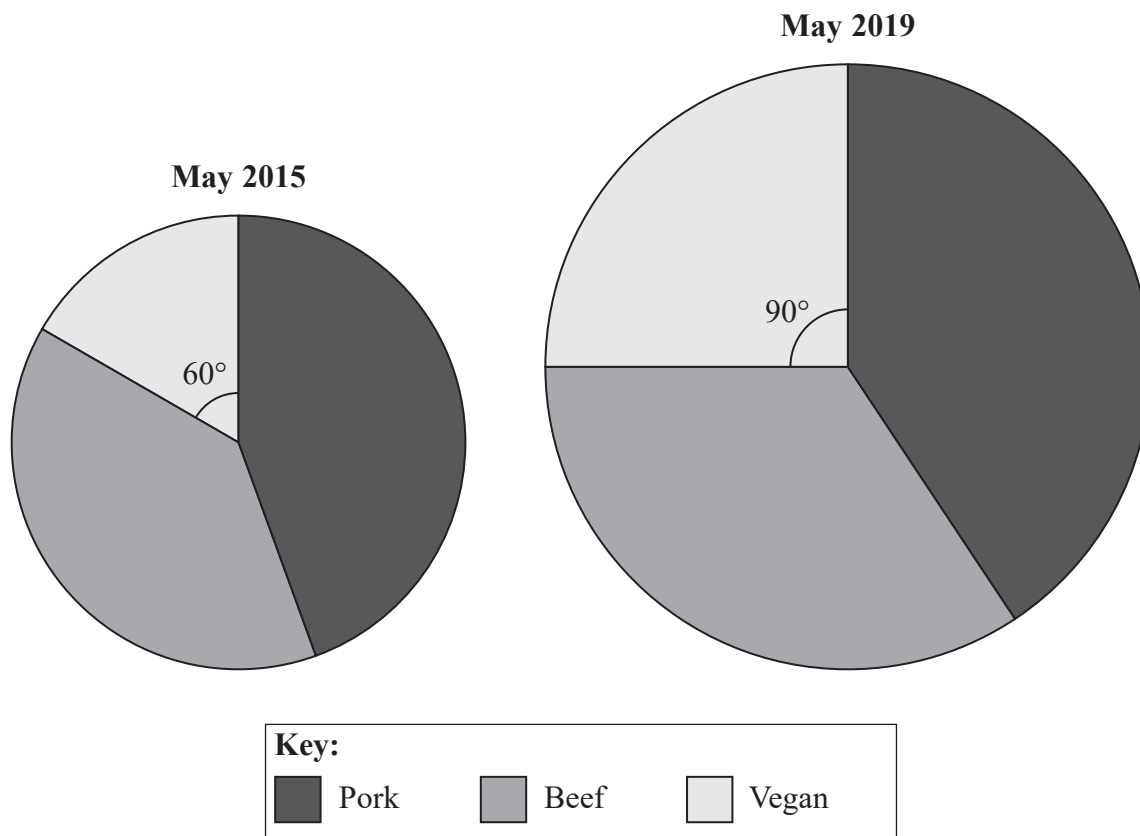
The radius of the pie chart for May 2019 is greater than the radius of the pie chart for May 2015

(a) Explain what can be deduced from this information.

....

(1)

30 packs of vegan sausages were sold in the farm shop in May 2015

(b) Work out the number of packs of vegan sausages that were sold in the farm shop in May 2019

....

(3)

(Total for Question 9 is 4 marks)

10 Changes in the cost of living in the United Kingdom are measured by the consumer price index (CPI).

The table shows the index numbers for June 2019, using 2015 as base year, of the items that contribute to the CPI.

The table also shows the weightings for how the spending by the Jones family is divided between these items.

Item	Index number	Jones family weighting	
Food, alcohol and tobacco	106.4	17	
Clothing and footwear	102.0	8	
Housing and services	107.5	32	
Transport	112.5	6	
Communication, recreation and leisure	108.9	25	
Health, education and miscellaneous	106.6	12	

(Source: *Office for National Statistics*)

Using 2015 as base year, the CPI for June 2019 was 107.9

Compare the overall change in the cost of living for the Jones family between 2015 and June 2019 with the change in the CPI between 2015 and June 2019

Explain whether or not it is possible to conclude from your calculations that the Jones family is financially better off or worse off in June 2019 as compared with 2015
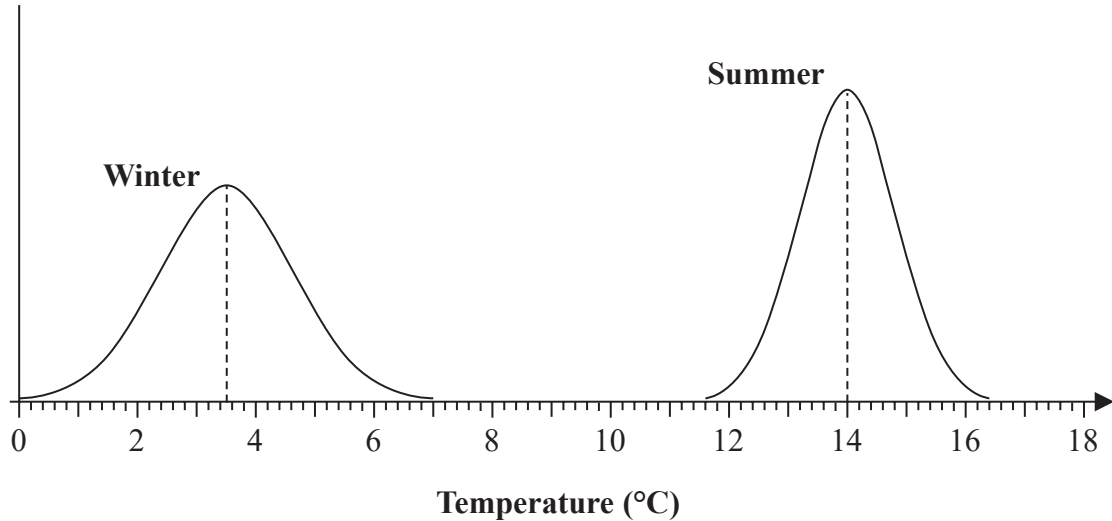
...

...

...

...

...

(Total for Question 10 is 5 marks)

11 The mean of the daily winter temperatures and the mean of the daily summer temperatures in the UK have been recorded for each year from 1910 to 2018

This information has been used to model the two distributions of the mean temperatures.

The graphs below give information about these models.

Temperature (°C)

(Source: *www.metoffice.gov.uk*)

(a) Write down the name of the distribution that is suggested by each of these graphs.

....

(1)

(b) Comment on the difference between the means of these two distributions.

....

....

....

(2)

The standard deviation for the distribution of the summer temperatures is 0.8 °C, correct to one decimal place.

(c) Using the graph for the winter temperatures, calculate an estimate for the standard deviation of the distribution of the winter temperatures, correct to one decimal place.

.... °C

(2)

The maximum height of the graph above the temperature axis for the winter temperatures is less than the maximum height of the graph above the temperature axis for the summer temperatures.

(d) Explain why.

....

....

<div align="right">(1)</div>

Carol says that the standard deviation of **daily** temperatures in summer for the years from 1910 to 2018 will also be 0.8 °C.

(e) Explain whether or not Carol is correct.

....

....

....

<div align="right">(2)</div>

Greta assumes that the means of the daily temperatures each year are independent.

She works out the probability that the means of the daily summer temperatures are greater than 15.6 °C for two consecutive years.

She concludes that this probability is less than 0.001

(f) Using the model for the daily summer temperatures, assess Greta's conclusion.

<div align="right">(5)</div>

<div align="right">(Total for Question 11 is 13 marks)</div>

TOTAL FOR PAPER IS 80 MARKS

BLANK PAGE

www.mrc-papers.com

P 6 8 6 7 6 A 0 2 0 2 0