

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)

Monday 2 June 2025

Morning (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 ►

P76385A

©2025 Pearson Education Ltd.
Y: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

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

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

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



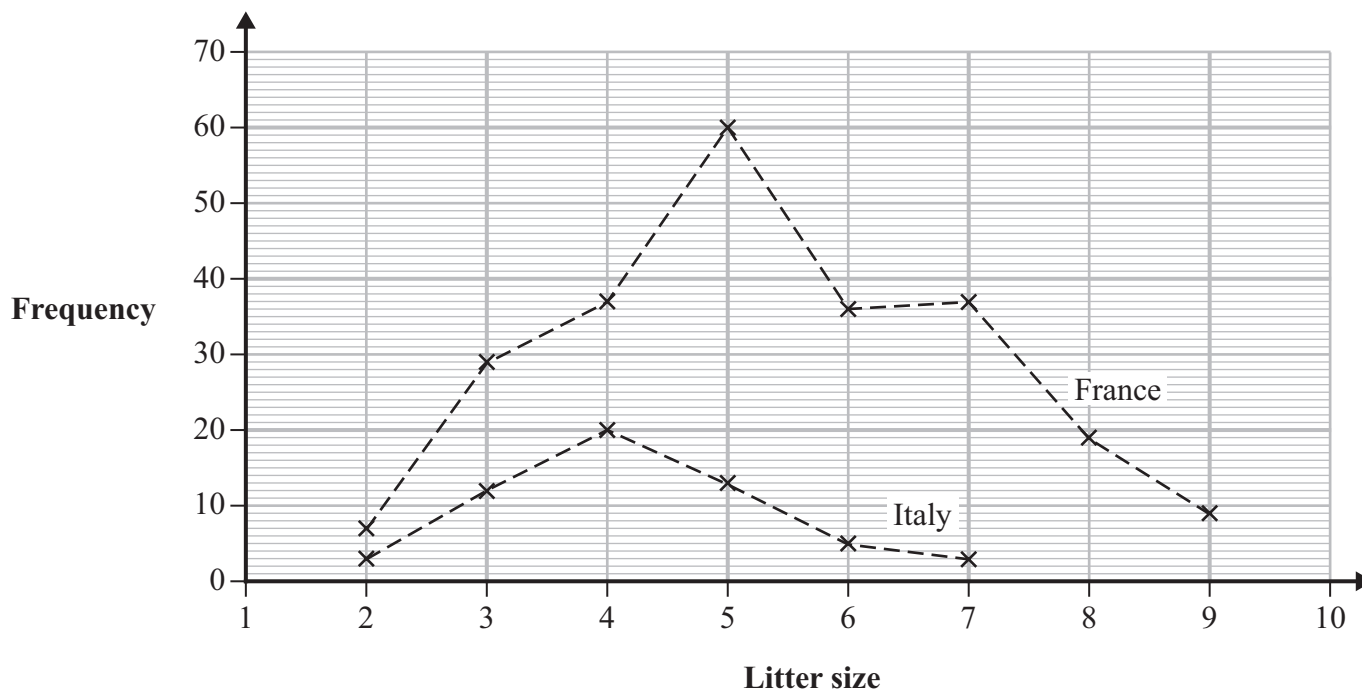
Answer ALL the questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

- 1 Rebecca collected data from the internet about the litter size of wild boar in France and in Italy.

The frequency polygons give information about the litter sizes.



(Source: www.academic.oup.com and www.researchgate.net)

Rebecca concludes that the average wild boar litter size in France is bigger than in Italy.

- (a) Do the frequency polygons support Rebecca's conclusion?

Give a reason for your answer.

(1)

- (b) Give another comparison of the distributions of the wild boar litter sizes in France and in Italy.

(1)

- (c) Discuss how the data collection could reduce the reliability of these conclusions.

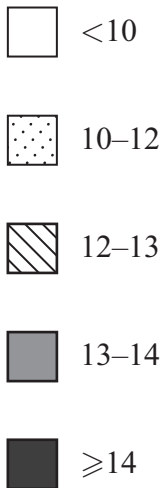
(1)

(Total for Question 1 is 3 marks)

- 2 The choropleth map gives information about the percentage change in the mileage driven in Great Britain for each region from 2020 to 2021

Percentage change in mileage driven in 2020 to 2021

Key:



(Source: <https://roadtraffic.dft.gov.uk>)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



The table shows the number of miles driven in London in 2020 and in 2021

2020	2021
16.6 billion miles	18.5 billion miles

(a) Show that London has been shaded correctly on the choropleth map.

(2)

(b) Which region had the lowest percentage change in the mileage driven between 2020 and 2021?

(1)

Amy concludes that the percentage change in the mileage driven between 2020 and 2021 is the same for Wales and South West England.

(c) Explain why Amy may **not** be correct.

(1)

Tom concludes that the number of motor vehicles using the roads has increased the most in Scotland between 2020 and 2021

(d) Explain whether or not the choropleth map supports this conclusion.

(2)

(Total for Question 2 is 6 marks)

DO NOT WRITE IN THIS AREA



3 Tsena is researching the ages of Oscar winners.

She has collected data from the internet about the ages of the winners of the Best Actor and Best Actress categories from 2000 to 2022

Data from the internet is an example of secondary data.

(a) Explain what is meant by secondary data.

(1)

Here are the ages of the winners of the Best Actor category from 2000 to 2022

40 36 47 29 43 37 38 45 50 48 60 50
39 55 44 32 41 41 59 37 83 53 54

(Source: www.openintro.org and www.oscars.org)

The stem and leaf diagram gives the ages of the winners of the Best Actress category from 2000 to 2022

(b) Complete the stem and leaf diagram with the ages of the winners of the Best Actor category.

You should also complete the key.

	Best Actress		Best Actor
2	2 5 6 8 8 9 9	2	
3	0 2 3 3 5 5	3	
4	4 4 5 5	4	
5	4	5	
6	0 0 1 2 3	6	
7		7	
8		8	

Key:

(Source: www.openintro.org and www.oscars.org)

(3)



- (c) Work out the median age for the winners of the Best Actress category from 2000 to 2022

.....
(1)

Michelle thinks it is more appropriate to use the interquartile range than the range as a measure of spread for these sets of data.

- (d) Give a reason why Michelle may think this.

.....
(1)

- (e) Work out the interquartile range for the winners of the Best Actress category from 2000 to 2022

.....
(2)

The table gives information about the ages of the winners of the Best Actor and Best Actress categories before 2000
It also gives information about the ages of the winners of the Best Actor category from 2000 to 2022

	Best Actor category before 2000	Best Actress category before 2000	Best Actor category from 2000 to 2022	Best Actress category from 2000 to 2022
Median	42	33	44	
Interquartile range	10.75	13.25	15	

- (f) Compare the distribution of ages for the winners of the Best Actor category with the distribution of ages for the winners of the Best Actress category.
You may use the blank column in the table to record your answers to parts (c) and (e) to help you.
You should interpret your comparisons in context.

(6)

(Total for Question 3 is 14 marks)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE

Turn over for next question



- 4 Agatha is investigating the times taken in the 2022 Olympic Women's Skiathlon race.

There are two parts to the Skiathlon race. There is a 7.5 km classic stage and a 7.5 km free stage.

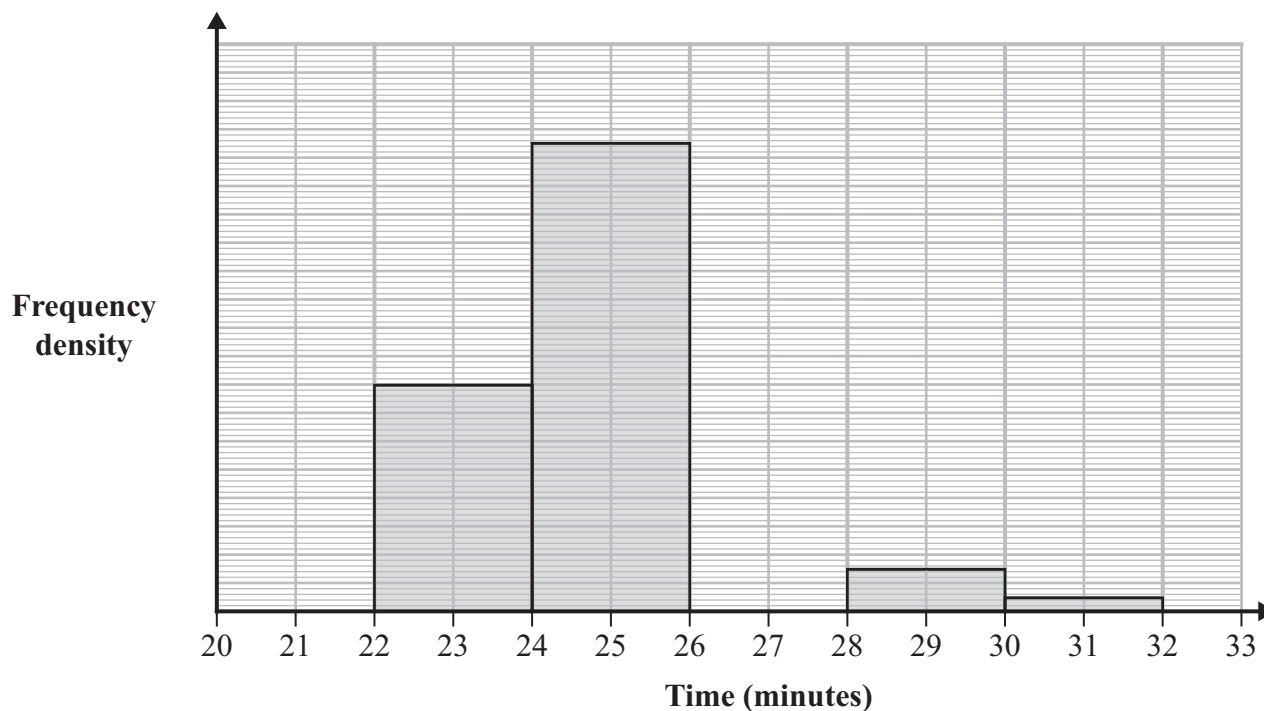
Agatha thinks there may be a relationship between the time taken by each skier to complete the classic stage and the time taken by each skier to complete the free stage. She suggests calculating a correlation coefficient to check.

- (a) Discuss the appropriateness of this suggestion.
You should consider the type of data in your answer.

(2)

Agatha found the times taken for the event by all the skiers.

The incomplete histogram and incomplete grouped frequency table give information about the times taken, in minutes, to complete the 7.5 km classic stage of the race.



(Source: *data.fis-ski.com*)

Time (t minutes)	Frequency
$22 < t \leq 24$	16
$24 < t \leq 26$	
$26 < t \leq 28$	11
$28 < t \leq 30$	3
$30 < t \leq 32$	1

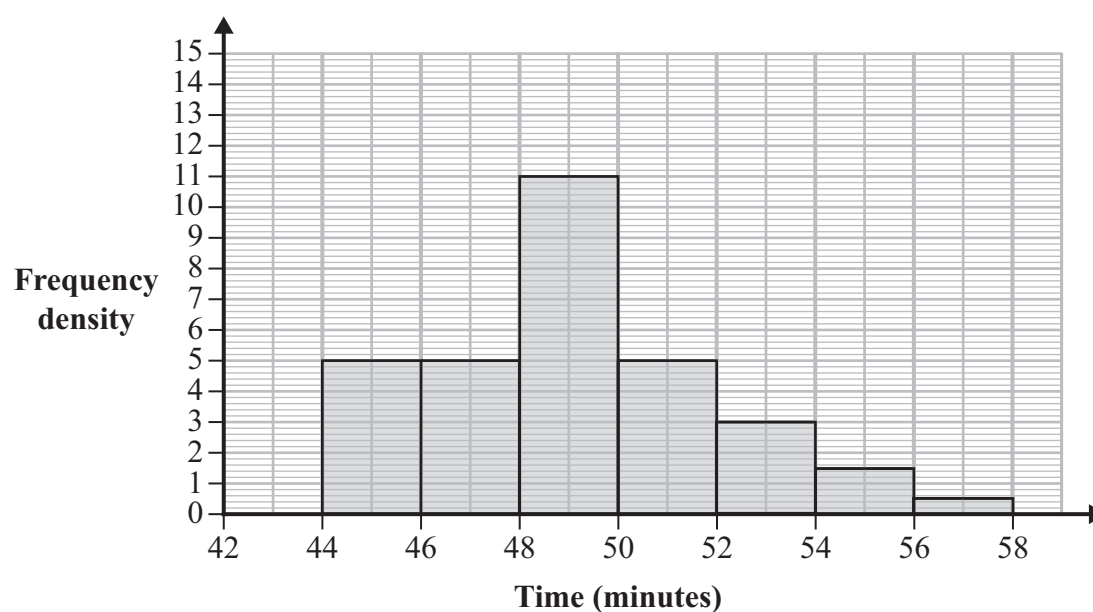
(b) Use the information in the histogram to complete the table.

(2)

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

(1)

Agatha also drew this histogram for the overall times for the race.



Agatha concludes that the median is in the $50 < t \leq 52$ minute time interval.

(d) Does the histogram support Agatha's conclusion?
Show calculations to support your answer.

(2)

A ski manufacturer produces wax used on skis.

It has made a new type of wax and wants to find out if using this wax improves the skiers' times.

Agatha thinks a matched pairs approach should be used to find out if the wax improves the time taken to complete the race.

Agatha suggests matching skiers based on the type of ski that they use and their times in the race. One skier in each pair would use their original wax and the other skier would use the new wax.

- (e) Comment on the appropriateness of this suggestion.
Give a reason for your answer.

(2)

(Total for Question 4 is 9 marks)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE

Turn over for next question



- 5 A council wants to encourage people to use the council leisure centres.
The council has put advertisements in the town centres to tell people about the opening times and the available facilities in the leisure centres.

Nico has been asked to find out if the advertising has been successful.
He plans to use cluster sampling where the clusters are customers at each leisure centre.
For the selected clusters he will ask the customers to complete a questionnaire about the advertisements.

- (a) Comment on the suitability of Nico's plan for sampling in this context.

(2)

Keti suggests that they should do a pilot study before they give out the questionnaire.

- (b) Give a reason why it is appropriate to do a pilot study.

(1)



The council also gave a voucher for one free admission to every child in the county.

The council thinks that some people have collected and used more than one of the free admission vouchers per child.

Nico is going to include a question about this on the questionnaire.

He thinks that he should use the random response technique.

(c) Explain why it is appropriate to use the random response technique in this context.

(1)

(d) Design a random response question that Nico could use on the questionnaire in order to collect this information.

(4)



Nico wants to find out if the number of people using the leisure centres increased after the advertisements and vouchers started a month ago.

He would also like to predict the number of people likely to use the leisure centres in the future.

Here is the plan for his investigation.

Data collection

- Collect data on the total number of people using each leisure centre on each day in the last 6 weeks.
- Each leisure centre will provide the total number of people for each of these dates.

Processing and presenting data

- Produce a time series graph for the total number of people at all of the leisure centres.
- Calculate 4-point moving averages for the total number of people and plot these to give a trend line.
- Identify any seasonality in the time series.
- If there is a trend then make predictions for the future total number of people by finding the mean total number of people for that particular day of the week and adding it to the trend line value.

- (e) Discuss whether Nico's plans for collecting, processing and presenting data are appropriate.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(6)

(Total for Question 5 is 14 marks)



6 Researchers wanted to know what kind of reduced-fat cheese that customers liked best.

They asked customers to rate 10 different cheeses and calculated an average score for 'Overall liking'.

The table gives information about some of their data.

Cheese	Customer overall liking	Yellowness of cheese	Customer overall liking rank	Yellowness of cheese rank		
A	6.69	5.2	1			
B	6.61	3.1	2			
C	6.56	4.3	3			
D	6.51	3.9	4			
E	6.48	4.8	5			
F	6.26	5.0	6			
G	6.22	6.9	7			
H	6.21	6.5	8			
I	6.04	5.8	9			
J	5.67	5.6	10			

(Source: www.sciencedirect.com)

- (a) Calculate Spearman's rank correlation coefficient between the 'customer overall liking' scores and the 'yellowness of cheese' scores.

(4)



The researchers calculated Spearman’s rank correlation coefficients between ‘customer overall liking’ scores and other characteristics of the cheeses.

In each case rank 1 represented the highest level of that quality.

The table gives some of their results.

Quality	Spearman’s rank correlation coefficient
Hole size	0.620
Toughness	−0.734
Stickiness	0.709
Fat content	−0.733
Salt content	0.787

(Source: www.sciencedirect.com)

- (b) Interpret in context the correlation for the characteristic that is most closely related to ‘customer overall liking’ scores.

(2)

The rankings for the characteristics ‘pleasantness of mouth feel’ and ‘customer overall liking’ were in perfect agreement.

- (c) Write down the Spearman’s rank correlation coefficient for the correlation between ‘pleasantness of mouth feel’ and ‘customer overall liking’.

(1)

(Total for Question 6 is 7 marks)



- 7 The Petersen capture recapture method was used to estimate the number of black bears in Pennsylvania in 2018

A group of researchers took into account the number of bears that died between initial tagging and the recapture.

A bear is considered to have a current tag if it was tagged in the previous year.

Taking into account the bears that died, the number of black bears with current tags available to be captured in 2018 was 690

During the year 2018, the number of black bears captured was 2017 of which 73 had a current tag.

(Source: *www.pgc.pa.gov*)

- (a) Calculate an estimate for the number of black bears in Pennsylvania in 2018

(2)

The researchers only considered a black bear to have a current tag during the second capture if it was tagged in the previous year.

- (b) Explain the effect on your estimate in part (a) if they did not do this and instead included all tags.

You should give a reason for your answer.

(2)

The approach used by the researchers considered the issues around tag loss.

- (c) Give one other assumption of the Petersen capture recapture method that they should have considered in their research.

(1)

(Total for Question 7 is 5 marks)



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

BLANK PAGE

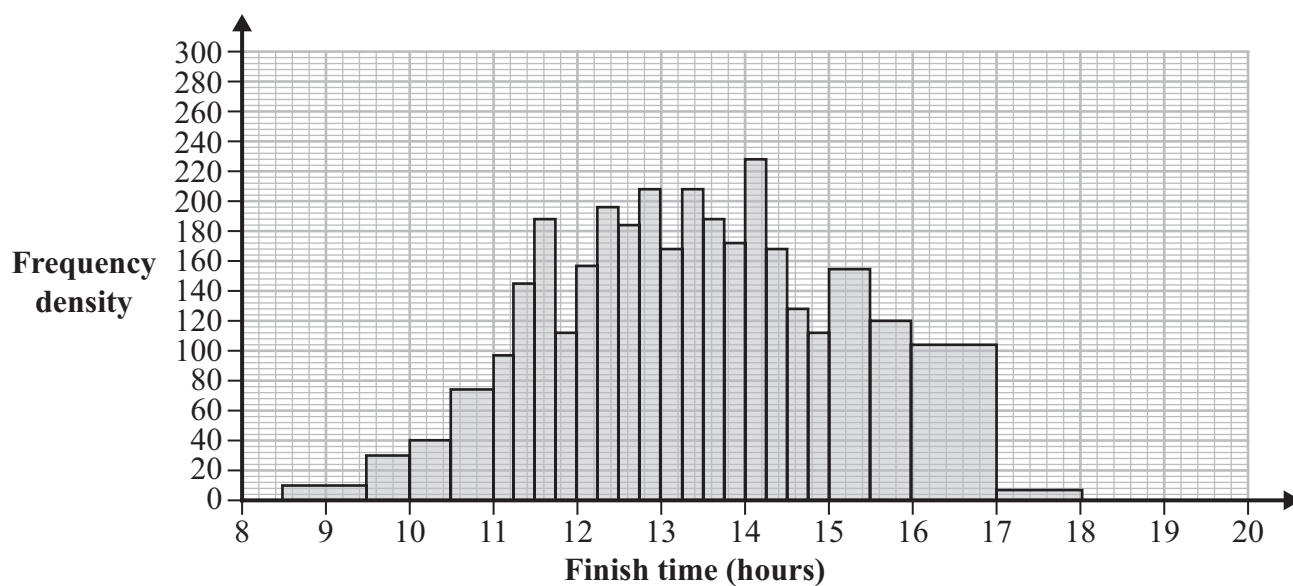
Turn over for next question



- 8 David collected data on the times, in hours, taken for men to complete the Ironman Canada 2022 race and the Ironman UK 2022 race.

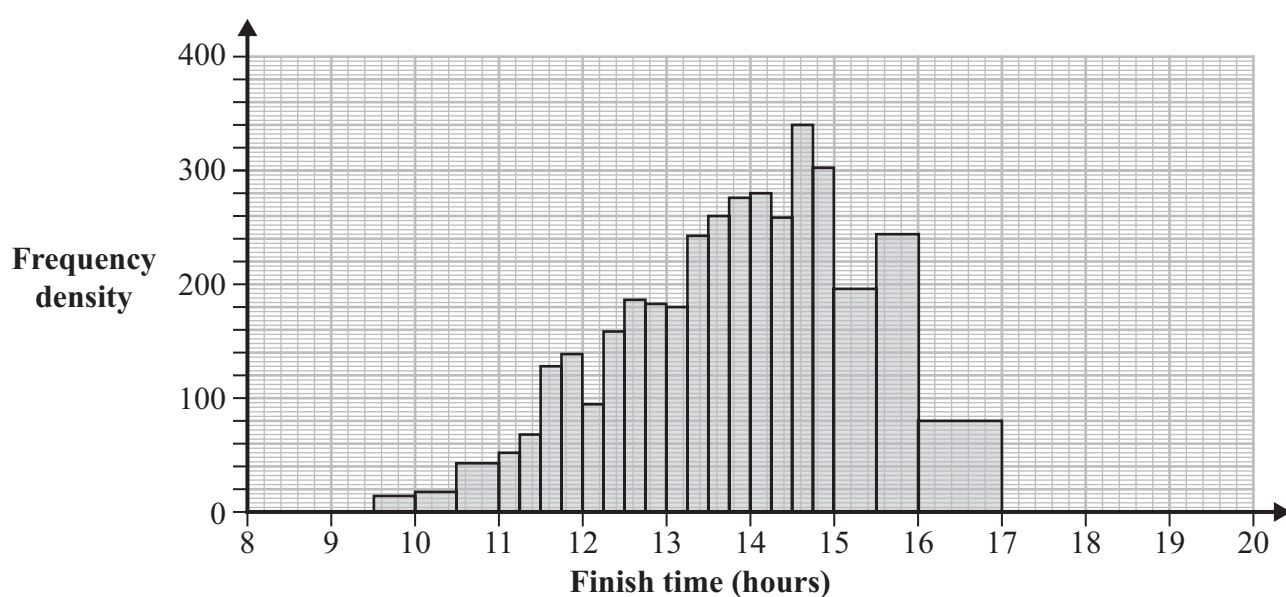
These graphs and summary statistics were produced for this data using statistical software.

Ironman Canada 2022 men's results



Mean	13.53	Median	13.49	Number of male competitors	992
Standard deviation	1.839	Skewness	0.0653		

Ironman UK 2022 men's results



Mean	13.98	Median	14.12	Number of male competitors	1128
Standard deviation	1.489	Skewness	-0.282		

(Source: www.coachcox.co.uk)



- (a) Show that there are no outliers in the times for the Ironman Canada 2022 men's results.

(3)

- (b) Compare, in context, the distribution of the Ironman Canada 2022 men's results with the distribution of the Ironman UK 2022 men's results.

(6)



Dan competed in the Ironman Canada 2022 race.

Here is information about the performance of all male competitors in the swim, bike and run stages of the race, and some of Dan's standardised scores.

Stage of race	Swim	Bike	Run
Mean (hours)	1.32	6.77	5.18
Standard deviation (hours)	0.21	0.85	0.95
Dan's standardised score	−1	−0.94	

Dan took 5.64 hours to complete the run.

- (c) Use standardised scores to compare Dan's performances relative to the other competitors in each stage of the race.
Give reasons for your comparisons.

(4)

(Total for Question 8 is 13 marks)



9 A company sells furniture.

The company has a telephone number that customers can use if they want to complain. At the end of the call customers are asked if they are satisfied or not.

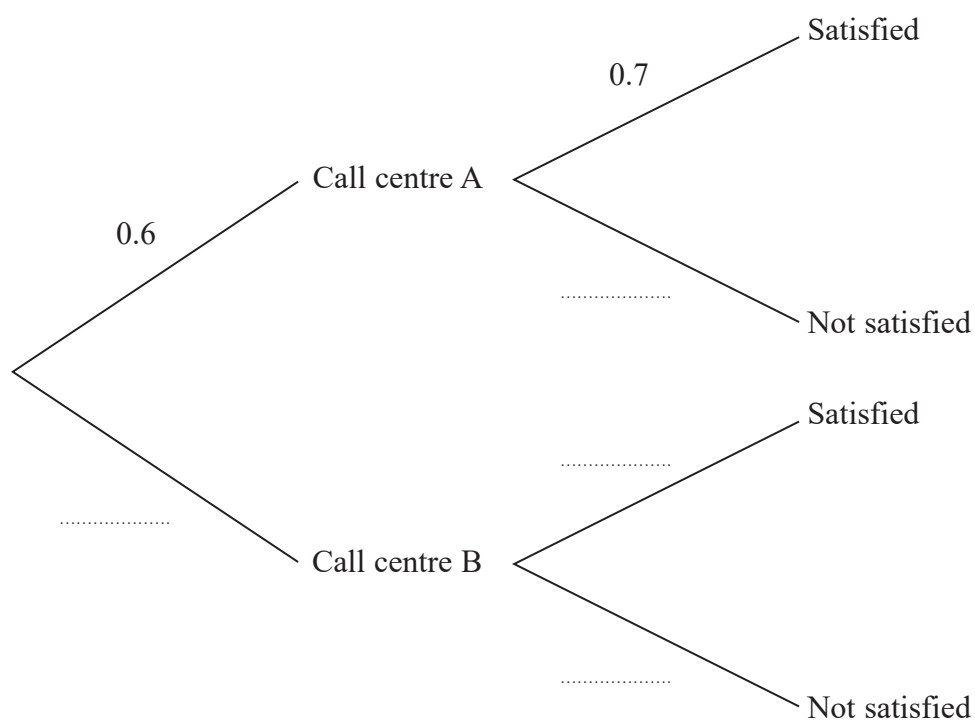
Originally the company had one call centre, Call centre A.
The company now has another call centre, Call centre B.

Since Call centre B opened,

- when a call is answered the probability that it is answered by Call centre A is 0.6
- of those calls answered by Call centre A, 70% result in a customer being 'satisfied'
- the probability of a call answered by Call centre B resulting in a customer being 'satisfied' is p

A call is picked at random.

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



(1)

Given that a customer is 'satisfied' the probability that the call was answered by Call centre B is 0.3

- (b) Find the probability of a call answered by Call centre B leading to a customer being satisfied.

(2)

The company sells tables that customers have to put together.

A table needs 4 connectors to hold it together.

A bag of 4 connectors is supplied with each table.

The company previously bought the connectors from QuickC.

The probability of a QuickC connector failing is 0.05

Connectors fail independently of each other.

A customer buys a table.

- (c) Work out the probability that the customer will need one or more replacement connectors because of the failure of QuickC connectors.

(2)



The company now buys the connectors from KlickT.
The probability of a KlickT connector failing is 0.16

Dina thinks that putting two more connectors in the pack would reduce the probability that a customer will need one or more replacement connectors to less than it was when QuickC connectors were used.

- (d) Is Dina correct?
You must show your working.

(4)

(Total for Question 9 is 9 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

