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)

Wednesday 5 June 2024

Afternoon (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 ▶

P75437A

F:1/1/1/1/1/1/

P 7 5 4 3 7 A 0 1 2 0

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

Answer ALL questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

1 The table gives the total labour force and the unemployment rate for the UK in 2017 and in 2018

Year	Total labour force (million)	Unemployment rate (%)
2017	33.80	4.38
2018	32.41	4.00
2019	34.04	.....

(Source: *www.economicsonline.co.uk* and *www.ons.gov.uk*)

In 2019 the total number of unemployed people was 1.29 million.

$$\text{Unemployment rate} = \frac{\text{Number of unemployed people}}{\text{Total labour force}} \times 100$$

(a) (i) Using the formula above, work out the unemployment rate for the UK in 2019
Give your answer correct to 2 decimal places and write your answer in the table.

(2)

(ii) Using your answer to part (a)(i), what conclusion can be drawn about the unemployment rate in the UK between 2017 and 2019?

.....

.....

(1)

Bob says, without doing any calculations, that the total number of people unemployed decreased from 2017 to 2018

(b) Using the data in the table, assess Bob's claim.

.....

.....

.....

(2)

(Total for Question 1 is 5 marks)

2 A fjord is a deep and narrow part of a sea with steep land on three sides.

Emily is investigating the length of fjords in Norway.
She collects some data from the internet and puts the data into a grouped frequency table.

The grouped frequency table below shows information about the results she collected.

Length of fjord (l km)	Frequency
$0 \leqslant l < 50$	199
$50 \leqslant l < 100$	17
$100 \leqslant l < 150$	12
$150 \leqslant l < 200$	3
$200 \leqslant l < 250$	1

(Source: *https://en.wikipedia.org/wiki/List_of_Norwegian_fjords*)

(a) Work out the number of fjords that have a length of at least 100 km.

.....

(2)

(b) (i) Calculate an estimate of the mean length of these fjords.
Give your answer to 1 decimal place.

..... km

(3)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(ii) Explain why your answer to part (b)(i) is only an estimate.

.....

.....

(1)

(iii) How could Emily have improved the accuracy of her answer to part (b)(i)?

.....

.....

(1)

Emily plans to use a frequency polygon to represent the lengths of the fjords.

(c) Discuss whether or not a frequency polygon would be an appropriate diagram to use.

.....

.....

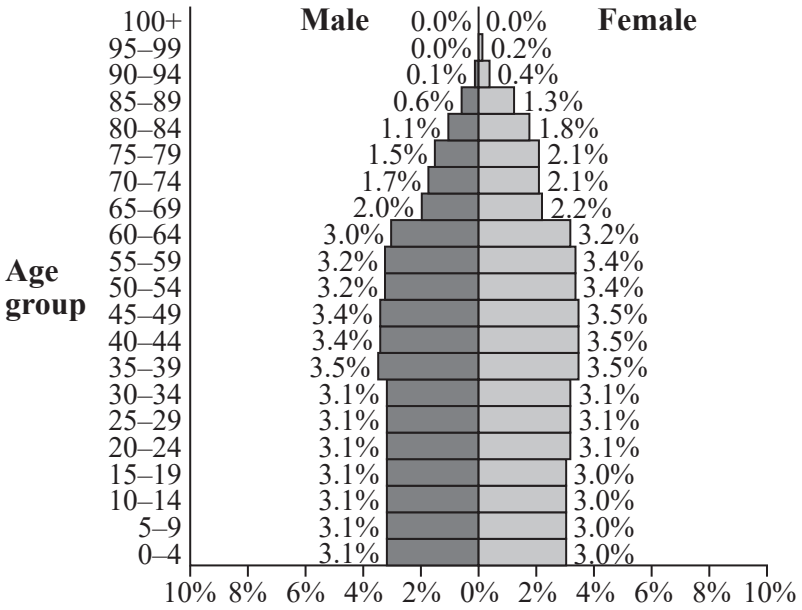
.....

(2)

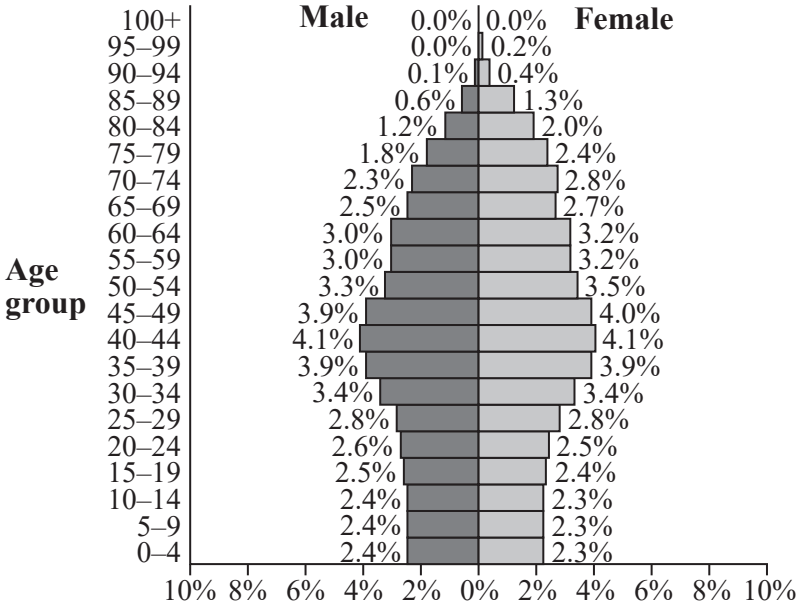
(Total for Question 2 is 9 marks)

3 The two population pyramids give information about the percentage of the population who are male and who are female for each age group in France and in Italy in 2010

2010 France **Population 62 444 566**



2010 Italy **Population 59 822 450**



(Source: *www.populationpyramid.net*)

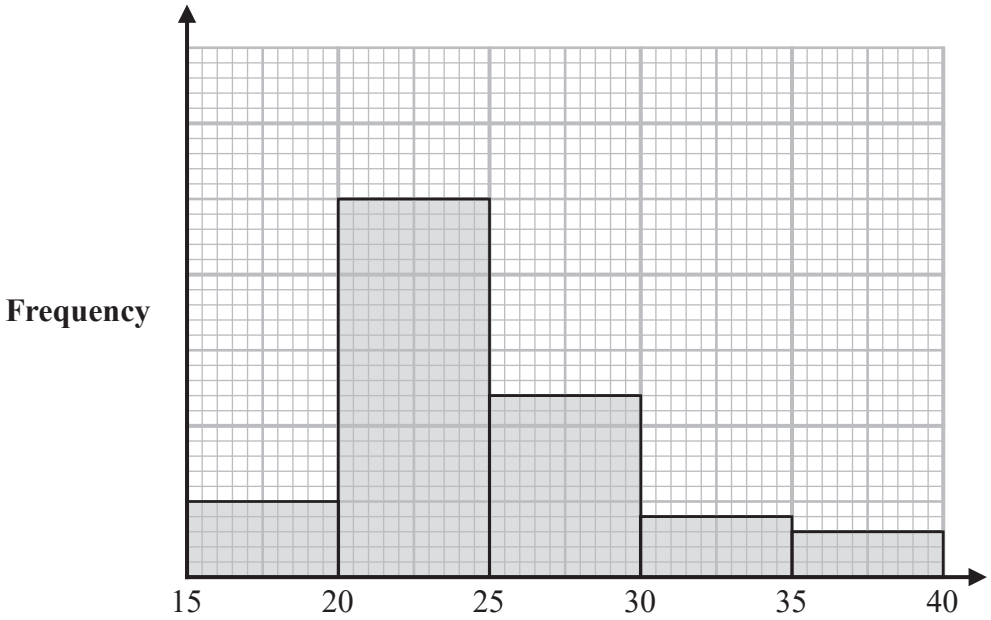
Tommy is investigating how the populations of Italy and France differ in 2010
He uses the two population pyramids to reach the following two conclusions.

- The percentage of people aged 50–54 was lower in France than the percentage of people aged 50–54 in Italy.
- The number of males aged 40–44 in France was greater than the number of males aged 40–44 in Italy.

Assess Tommy's two conclusions.
You should show clearly the value of any statistics you use in your answer.

(Total for Question 3 is 5 marks)

4 The time taken, in minutes, for some runners to complete a 5 km run was recorded. The incomplete histogram and incomplete grouped frequency table give information about the times taken, in minutes, for these runners to complete the 5 km run.



Time taken to run 5 km (t minutes)	Frequency
$15 < t \leqslant 20$	5
.....	25
$25 < t \leqslant 30$	.....
$30 < t \leqslant 35$	4
$35 < t \leqslant 40$	3

(Source: *www.parkrun.org.uk*)

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

(2)

(b) Estimate the number of runners that took less than or equal to 23 minutes to complete the race.

.....

(2)

(c) Identify and interpret the skew shown on the histogram.

.....

.....

.....

(2)

(Total for Question 4 is 6 marks)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

5 The management of a factory is considering changing the working hours of their employees.
Muhammad and Rose want to get the views of the employees in the factory.

Employees in the factory work on the production line or in the warehouse or in the office.

20 employees work on the production line.
15 employees work in the warehouse.
25 employees work in the office.

Muhammad plans to use a questionnaire.
He plans to take a sample of the employees and ask them the questions on his questionnaire.
For his sample, he decides to ask all of the employees who work on the production line.

(a) (i) Name this sampling technique.

.....

(1)

(ii) Give two reasons why using this sampling technique may not be appropriate.

1 .....

.....

2 .....

.....

(2)

Muhammad wants to find out how many extra hours each employee would be willing to work each week.

(b) Design a closed question that Muhammad could use in his questionnaire.

(2)

Muhammad collects the completed questionnaires.
He finds that some of the employees on the production line have not responded.

(c) Suggest **two** ways in which Muhammad could have reduced the number of non-responses.

.....

.....

.....

.....

(2)

Rose decides to take a 10% systematic sample of all the 60 employees in the factory.

(d) Describe in detail how this sample could be selected.

.....

.....

.....

.....

.....

.....

.....

.....

(3)

Rose plans to use a face-to-face interview.

(e) How would using a face-to-face interview rather than a questionnaire improve the quality of the responses?

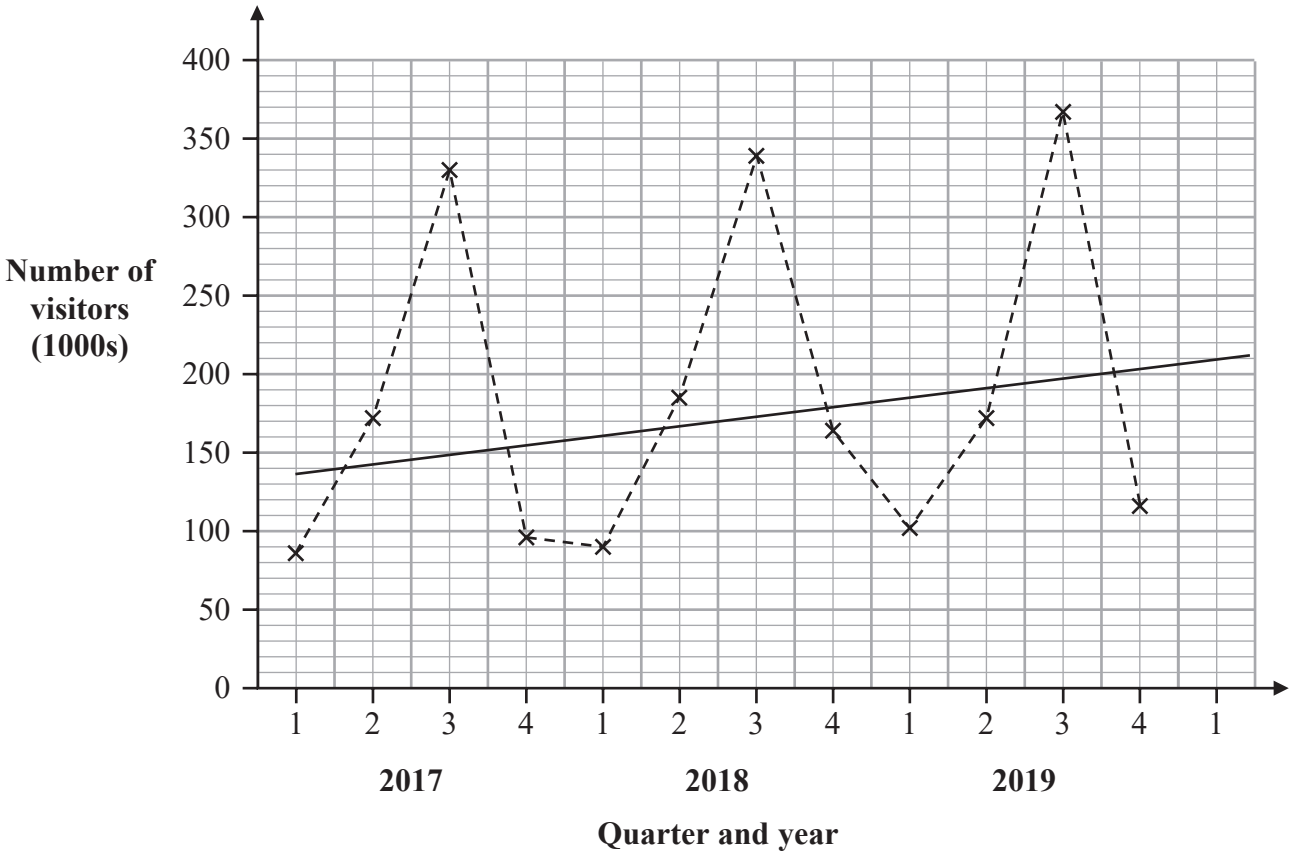
.....

.....

(1)

(Total for Question 5 is 11 marks)

6 The time series graph gives information about the numbers, in thousands, of visitors to Canada from the UK for each quarter for the years 2017 to 2019



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

A trend line has been drawn on the time series graph.

(a) Describe and interpret the trend shown by the graph.

(2)

(b) Discuss any seasonal variation shown by the graph and interpret one in context.
Do **not** do any calculations.

(3)

(c) (i) Work out the mean seasonal variation for Quarter 1
Give your answer correct to 1 decimal place.

..... thousand

(2)

(ii) Interpret your answer to part (c)(i) in context.

.....

.....

(1)

(d) (i) Use your answer to part (c)(i) and the time series graph to predict the number of visitors to Canada from the UK in Quarter 1 of 2020
You must show your working.

..... thousand

(2)

(ii) Discuss the reliability of using your answer to part (c)(i) to predict the number of visitors to Canada from the UK in Quarter 1 of 2025
Give a reason for your answer.

.....

.....

(2)

(Total for Question 6 is 12 marks)

7 Roberta is investigating how the ages of brides getting married in the UK has changed from 2003 to 2013
She collects official data from the internet using the website 'Office for National Statistics'.

(a) Explain why this website will give reliable data.

.....

.....

(1)

Roberta wrote the following hypothesis before she collected her data,

Has the age of brides increased between 2003 and 2013?

(b) Explain why it is not appropriate to use this as a hypothesis.

.....

.....

(1)

The table gives information about the data that Roberta collected.

Age of bride	**2003**	**2013**
Under 25	53 837	32 197
25 to 34	130 138	126 377
35 to 44	54 103	44 422
45 to 54	21 858	24 952
55 and over	10 173	12 874
Total	**270 109**	**240 822**

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

DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA DO NOT WRITE IN THIS AREA

Roberta wants to compare the proportion of brides in each age group and the total number of brides in each age group by presenting the results in charts for 2003 and 2013
She discusses how to present the results with Andria.

Andria thinks that they should use pie charts.

Roberta thinks that they should use comparative pie charts.

(c) What advice would you give to Andria and to Roberta on their choice of charts?

.....

.....

.....

.....

.....

.....

(3)

Roberta used a circle with a radius of 4 cm for her comparative pie chart for 2003

(d) Calculate the radius of the circle for her comparative pie chart for 2013
Give your answer correct to 2 decimal places.

..... cm

(2)

Roberta is going to present her comparative pie charts to non-statisticians.

(e) What extra information could she include on her pie charts to help the non-statisticians make comparisons?

.....

.....

(1)

(Total for Question 7 is 8 marks)

8 Faiza owns a factory that makes footballs.
The production line is set up to make footballs with a diameter of 219 mm.
As a quality control check, random samples are taken and the mean diameter of each sample is calculated.

The sample means should be normally distributed with a mean of 219 mm and a standard deviation of 0.05 mm.

(a) Would the sample means or the population values be more consistent?
Give a reason for your answer.

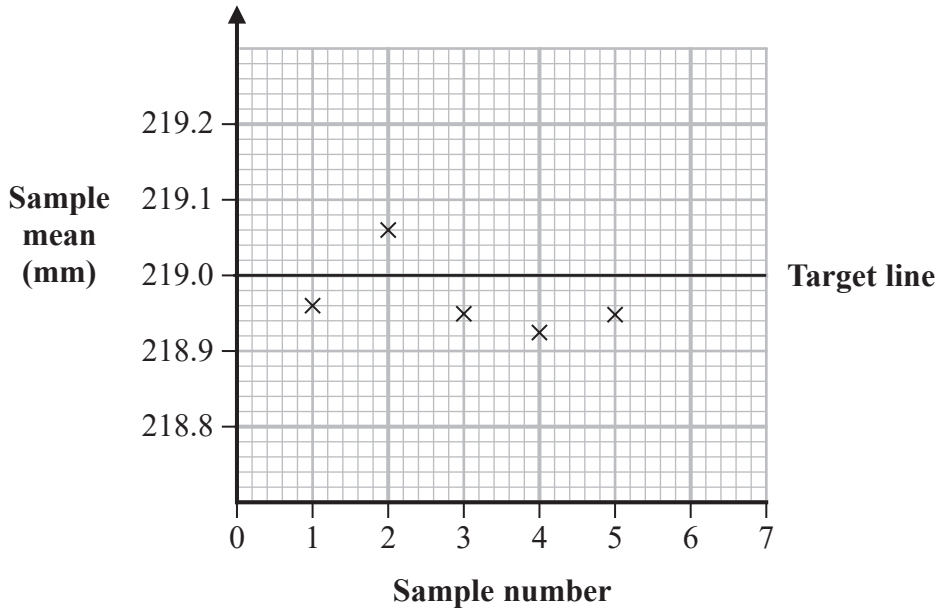
.....

.....

.....

(2)

Faiza draws a control chart showing the first 5 samples.



Sample 6 and sample 7 have the following sample means.

	Sample 6	Sample 7
Sample mean (mm)	219.12	219.05

(b) Plot these two sample means on the quality control chart.

(2)

(c) By completing the control chart, determine what actions, if any, Faiza should have taken based on the information given after each of these last two samples.
You must justify your answer.

(5)

(d) Faiza decides to complete more quality control checks and draw control charts for these.
From the four options below, circle the two other measures that Faiza could measure and use in these control charts.

range **mode** **median** **skew**

(1)

(Total for Question 8 is 10 marks)

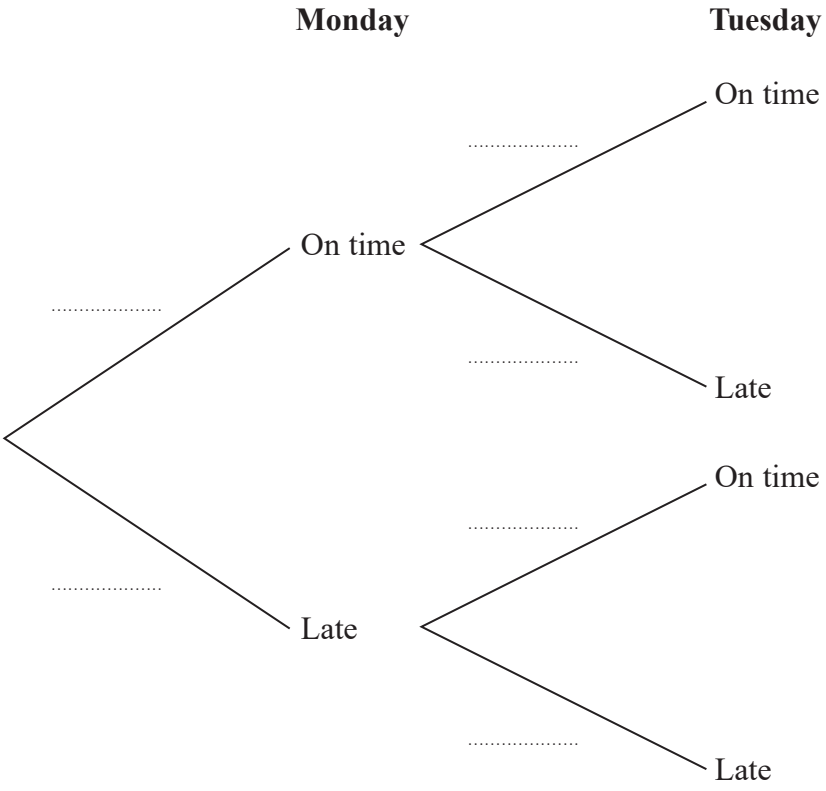
9 Peter drops his children off at a breakfast club every Monday morning and every Tuesday morning.
He either drops them off on time or late.

The probability that Peter drops his children off on time on a Monday morning is three times the probability that he drops them off late on a Monday morning.

The probability that Peter drops them off on time on a Monday morning **and** on a Tuesday morning is 0.6

The probability that Peter drops them off late on a Tuesday morning given that he has dropped them off late on a Monday morning is 0.3

(a) Complete the tree diagram below.



(4)

Peter drops his children off at the breakfast club every Monday morning and Tuesday morning over a 4-week period.

Assume that the probability of dropping them off on time on both a Monday morning and a Tuesday morning remains constant and that each week is independent.

(b) Work out the probability that, over a 4-week period, he drops his children off on time on a Monday morning and on time on a Tuesday morning on 3 or more occasions.

.....

(4)

(Total for Question 9 is 8 marks)

10 During a sports day competition students take part in two races.
Each student runs a 100 m race and a 400 m race.
The table shows the mean and the standard deviation of the times taken by the students in the 100 m race and in the 400 m race.

	Mean (seconds)	**Standard deviation (seconds)**
100 m race	14.5	2.3
400 m race	57.2	5.4

Dominic and Kai take part in the two races.
The incomplete table below gives their times and standardised scores.

	Dominic		**Kai**	
	Time (seconds)	**Standardised score**	**Time (seconds)**	**Standardised score**
100 m race	13.35	.....	13.58	–0.4
400 m race	56.12	–0.2	.....	0.3

Dominic and Kai make the following conclusions.

- Dominic concludes that he performed better in the 400 m race compared to the 100 m race.
- Kai concludes that he finished over one second slower than Dominic in the 400 m race.

Complete the table and assess Dominic's and Kai's conclusions.
Give a reason for each of your decisions.

.....

.....

.....

.....

(Total for Question 10 is 6 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