



GuideMyGrades

— YOUR GOALS. OUR EXPERTISE. —

EDEXCEL A LEVEL MATHEMATICS: QUESTIONS BY TOPIC

Data Presentation and Interpretation (12)

A Level Mathematics Topic 12

24 questions · 224 marks · Mark schemes included

Pearson Edexcel A Level Mathematics (9MA0), June 2018 - June 2025

 guidemygrades.com

 @Guidemygrades

 @Guidemygrades

 @Guidemygrades

 info@guidemygrades.com

 Piccadilly Business Centre, Unit C, Aldow Enterprise Park
Blackett Street, Manchester, M12 6AE

HOW TO USE THIS BOOKLET

Questions are grouped by specification sub-topic and run from the lowest mark tariff to the highest within each one.

Click any row below to jump straight to that sub-topic or question.

On every question page, click ANSWER in the top-right corner to jump to its mark scheme.

Diagrams, graphs, tables and data are printed exactly as they appeared in the original papers.

QUESTION	SOURCE	MARKS	PAGE
12.1 DIAGRAMS AND HISTOGRAMS		2 Qs · 24 marks	3
1. Q3	Paper 3: Statistics and Mechanics October2020	10	4
2. Q6 <i>also Topic 11 & 14</i>	Paper 3: Statistics and Mechanics June2023	14	8
12.2 SCATTER DIAGRAMS, CORRELATION AND REGRESSION		7 Qs · 53 marks	13
3. Q2 <i>also Topic 15</i>	Paper 3: Statistics and Mechanics June2024	6	14
4. Q2 <i>also Topic 11 & 15</i>	Paper 3: Statistics and Mechanics November2021	6	18
5. Q2 <i>also Topic 15</i>	Paper 3: Statistics and Mechanics June2018	7	22
6. Q2 <i>also Topic 11 & 15</i>	Paper 3: Statistics and Mechanics October2020	7	24
7. Q3 <i>also Topic 11 & 15</i>	Paper 3: Statistics and Mechanics June2019	9	28
8. Q6 <i>also Topic 11 & 15</i>	Paper 3: Statistics and Mechanics June2022	9	32
9. Q4 <i>also Topic 11</i>	Paper 3: Statistics and Mechanics June2025	9	37
12.3 MEASURES OF LOCATION AND SPREAD		13 Qs · 130 marks	41
10. Q4 <i>also Topic 11</i>	Paper 3: Statistics and Mechanics June2023	6	42
11. Q3	Paper 3: Statistics and Mechanics June2022	7	44
12. Q3	Paper 3: Statistics and Mechanics June2023	7	46
13. Q3	Paper 3: Statistics and Mechanics November2021	8	48
14. Q4 <i>also Topic 11 & 13 & 14</i>	Paper 3: Statistics and Mechanics June2019	9	50
15. Q5	Paper 3: Statistics and Mechanics June2024	10	54
16. Q5 <i>also Topic 11 & 13</i>	Paper 3: Statistics and Mechanics November2021	11	58
17. Q2 <i>also Topic 11 & 13</i>	Paper 3: Statistics and Mechanics June2022	12	62
18. Q4 <i>also Topic 11</i>	Paper 3: Statistics and Mechanics June2018	13	66
19. Q5 <i>also Topic 11 & 13 & 14</i>	Paper 3: Statistics and Mechanics June2019	13	70
20. Q5 <i>also Topic 11 & 13 & 14</i>	Paper 3: Statistics and Mechanics June2018	14	73
21. Q5 <i>also Topic 11 & 13 & 14 & 15</i>	Paper 3: Statistics and Mechanics October2020	15	77
22. Q2	Paper 3: Statistics and Mechanics June2025	5	82
12.4 OUTLIERS AND CLEANING DATA		2 Qs · 17 marks	84
23. Q3	Paper 3: Statistics and Mechanics June2024	6	85
24. Q2 <i>also Topic 14</i>	Paper 3: Statistics and Mechanics June2019	11	87



QUESTIONS

24 questions · 224 marks · 4 sub-topics



SUB-TOPIC 12.1

DIAGRAMS AND HISTOGRAMS

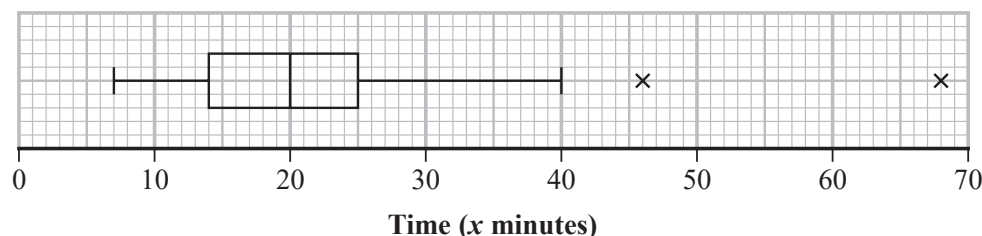
Sub-topic 12.1 · 2 questions · 24 marks · lowest tariff first



3. Each member of a group of 27 people was timed when completing a puzzle.

The time taken, x minutes, for each member of the group was recorded.

These times are summarised in the following box and whisker plot.



- (a) Find the range of the times. (1)

- (b) Find the interquartile range of the times. (1)

For these 27 people $\sum x = 607.5$ and $\sum x^2 = 17\,623.25$

- (c) calculate the mean time taken to complete the puzzle, (1)

- (d) calculate the standard deviation of the times taken to complete the puzzle. (2)

Taruni defines an outlier as a value more than 3 standard deviations above the mean.

- (e) State how many outliers Taruni would say there are in these data, giving a reason for your answer. (1)

Adam and Beth also completed the puzzle in a minutes and b minutes respectively, where $a > b$.

When their times are included with the data of the other 27 people

- the median time increases
 - the mean time does not change
- (f) Suggest a possible value for a and a possible value for b , explaining how your values satisfy the above conditions. (3)
- (g) Without carrying out any further calculations, explain why the standard deviation of all 29 times will be lower than your answer to part (d). (1)





(Total for Question 3 is 10 marks)



6. A medical researcher is studying the number of hours, T , a patient stays in hospital following a particular operation.

The histogram on the page opposite summarises the results for a random sample of 90 patients.

- (a) Use the histogram to estimate $P(10 < T < 30)$

(2)

For these 90 patients the time spent in hospital following the operation had

- a mean of 14.9 hours
- a standard deviation of 9.3 hours

Tomas suggests that T can be modelled by $N(14.9, 9.3^2)$

- (b) With reference to the histogram, state, giving a reason, whether or not Tomas' model could be suitable.

(1)

Xiang suggests that the frequency polygon based on this histogram could be modelled by a curve with equation

$$y = kxe^{-x} \quad 0 \leq x \leq 4$$

where

- x is measured in **tens of hours**
- k is a constant

- (c) Use algebraic integration to show that

$$\int_0^n xe^{-x} dx = 1 - (n+1)e^{-n} \quad (4)$$

- (d) Show that, for Xiang's model, $k = 99$ to the nearest integer.

(3)

- (e) Estimate $P(10 < T < 30)$ using

- (i) Tomas' model of $T \sim N(14.9, 9.3^2)$

(1)

- (ii) Xiang's curve with equation $y = 99xe^{-x}$ and the answer to part (c)

(2)

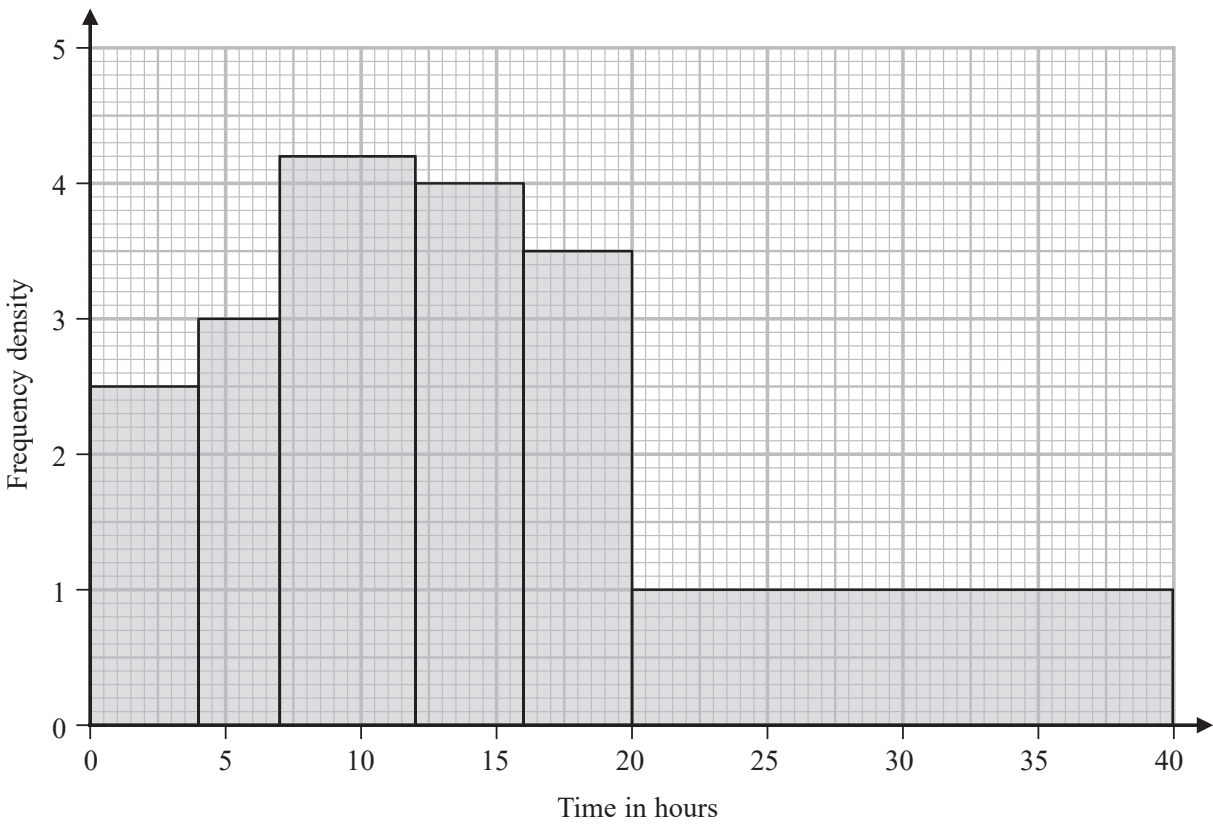
The researcher decides to use Xiang's curve to model $P(a < T < b)$

- (f) State one limitation of Xiang's model.

(1)



Question 6 continued









SUB-TOPIC 12.2

SCATTER DIAGRAMS, CORRELATION AND REGRESSION

Sub-topic 12.2 · 7 questions · 53 marks · lowest tariff first



2. Amar is studying the flight of a bird from its nest.

He measures the bird's height above the ground, h metres, at time t seconds for 10 values of t

Amar finds the equation of the regression line for the data to be $h = 38.6 - 1.28t$

- (a) Interpret the gradient of this line.

(1)

The product moment correlation coefficient between h and t is -0.510

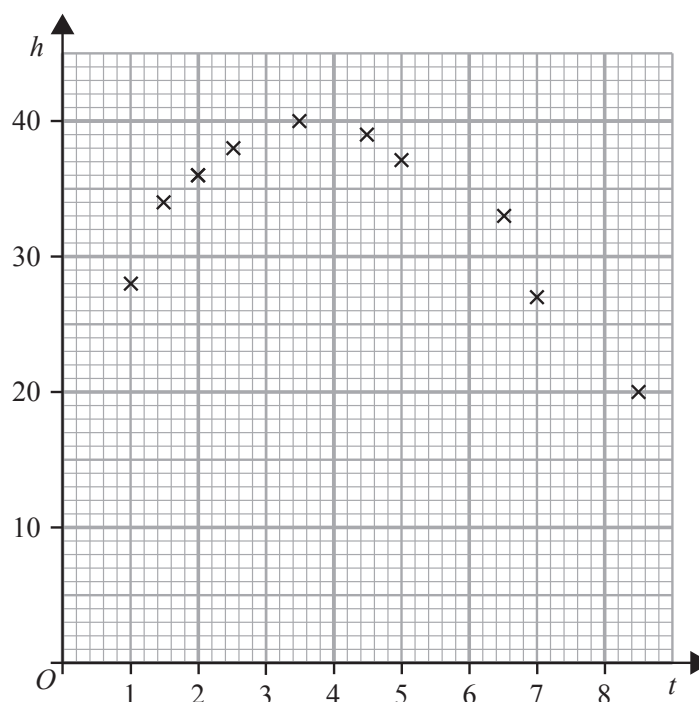
- (b) Test whether or not there is evidence of a negative correlation between the height above the ground and the time during the flight.

You should

- state your hypotheses clearly
- use a 5% level of significance
- state the critical value used

(3)

Jane draws the following scatter diagram for Amar's data.



- (c) With reference to the scatter diagram, state, giving a reason, whether or not the regression line $h = 38.6 - 1.28t$ is an appropriate model for these data.

(1)

Jane suggests an improved model using the variable $u = (t - k)^2$ where k is a constant.

She obtains the equation $h = 38.1 - 0.78u$

- (d) Choose a suitable value for k to write Jane's improved model for h in terms of t only.

(1)

Question 2 continued

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.



Question 2 continued

(Total for Question 2 is 6 marks)



2. Marc took a random sample of 16 students from a school and for each student recorded
 - the number of letters, x , in their last name
 - the number of letters, y , in their first name

His results are shown in the scatter diagram on the next page.

- (a) Describe the correlation between x and y .

(1)

Marc suggests that parents with long last names tend to give their children shorter first names.

- (b) Using the scatter diagram comment on Marc's suggestion, giving a reason for your answer.

(1)

The results from Marc's random sample of 16 observations are given in the table below.

x	3	6	8	7	5	3	11	3	4	5	4	9	7	10	6	6
y	7	7	4	4	6	8	5	5	8	4	7	4	5	5	6	3

- (c) Use your calculator to find the product moment correlation coefficient between x and y for these data.

(1)

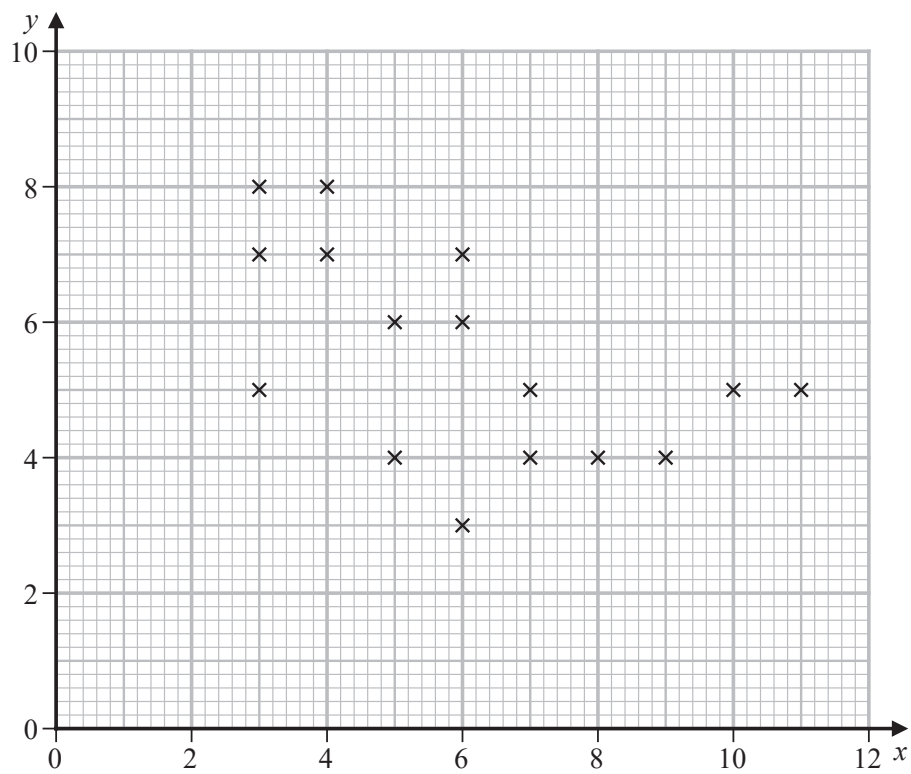
- (d) Test whether or not there is evidence of a negative correlation between the number of letters in the last name and the number of letters in the first name.

You should

- state your hypotheses clearly
- use a 5% level of significance

(3)

Question 2 continued.



Question 2 continued.



Question 2 continued.

(Total for Question 2 is 6 marks)



2. Tessa owns a small clothes shop in a seaside town. She records the weekly sales figures, £ w , and the average weekly temperature, t °C, for 8 weeks during the summer. The product moment correlation coefficient for these data is -0.915

(a) Stating your hypotheses clearly and using a 5% level of significance, test whether or not the correlation between sales figures and average weekly temperature is negative. (3)

(b) Suggest a possible reason for this correlation. (1)

Tessa suggests that a linear regression model could be used to model these data.

(c) State, giving a reason, whether or not the correlation coefficient is consistent with Tessa's suggestion. (1)

(d) State, giving a reason, which variable would be the explanatory variable. (1)

Tessa calculated the linear regression equation as $w = 10\,755 - 171t$

(e) Give an interpretation of the gradient of this regression equation. (1)

Question 2 continued

(Total for Question 2 is 7 marks)



2. A random sample of 15 days is taken from the large data set for Perth in June and July 1987. The scatter diagram in Figure 1 displays the values of two of the variables for these 15 days.

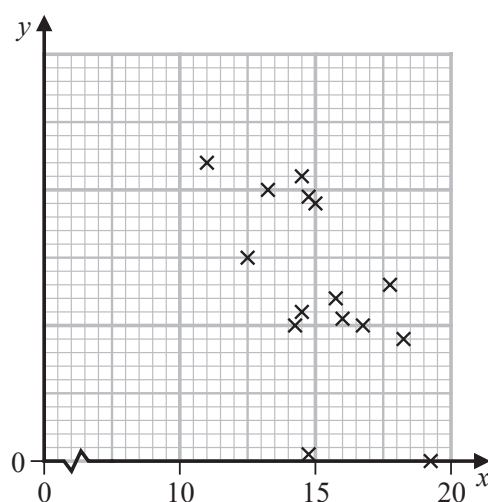


Figure 1

- (a) Describe the correlation.

(1)

The variable on the x -axis is Daily Mean Temperature measured in $^{\circ}\text{C}$.

- (b) Using your knowledge of the large data set,

- suggest which variable is on the y -axis,
- state the units that are used in the large data set for this variable.

(2)

Stav believes that there is a correlation between Daily Total Sunshine and Daily Maximum Relative Humidity at Heathrow.

He calculates the product moment correlation coefficient between these two variables for a random sample of 30 days and obtains $r = -0.377$

- (c) Carry out a suitable test to investigate Stav's belief at a 5% level of significance. State clearly

- your hypotheses
- your critical value

(3)

On a random day at Heathrow the Daily Maximum Relative Humidity was 97%

- (d) Comment on the number of hours of sunshine you would expect on that day, giving a reason for your answer.

(1)

Question 2 continued



Question 2 continued

(Total for Question 2 is 7 marks)



3. Barbara is investigating the relationship between average income (GDP per capita), x US dollars, and average annual carbon dioxide (CO_2) emissions, y tonnes, for different countries.

She takes a random sample of 24 countries and finds the product moment correlation coefficient between average annual CO_2 emissions and average income to be 0.446

- (a) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not the product moment correlation coefficient for all countries is greater than zero.

(3)

Barbara believes that a non-linear model would be a better fit to the data.

She codes the data using the coding $m = \log_{10} x$ and $c = \log_{10} y$ and obtains the model $c = -1.82 + 0.89m$

The product moment correlation coefficient between c and m is found to be 0.882

- (b) Explain how this value supports Barbara's belief.

(1)

- (c) Show that the relationship between y and x can be written in the form $y = ax^n$ where a and n are constants to be found.

Question 3 continued

[illegible]

Question 3 continued

[illegible]

Question 3 continued

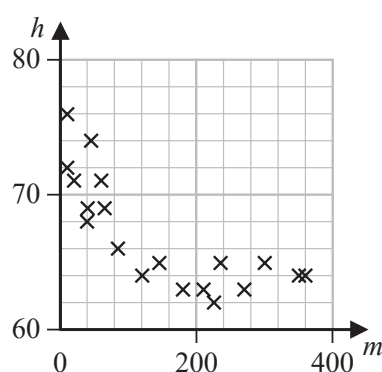
(Total for Question 3 is 9 marks)



6. Anna is investigating the relationship between exercise and resting heart rate. She takes a random sample of 19 people in her year at school and records for each person

- their resting heart rate, h beats per minute
- the number of minutes, m , spent exercising each week

Her results are shown on the scatter diagram.



- (a) Interpret the nature of the relationship between h and m

(1)

Anna codes the data using the formulae

$$x = \log_{10} m$$

$$y = \log_{10} h$$

The product moment correlation coefficient between x and y is -0.897

- (b) Test whether or not there is significant evidence of a negative correlation between x and y

You should

- state your hypotheses clearly
- use a 5% level of significance
- state the critical value used

(3)

The equation of the line of best fit of y on x is

$$y = -0.05x + 1.92$$

- (c) Use the equation of the line of best fit of y on x to find a model for h on m in the form

$$h = am^k$$

where a and k are constants to be found.

(5)



Question 6 continued



Question 6 continued



Question 6 continued



Question 6 continued

(Total for Question 6 is 9 marks)



4. Kay is studying the variables Daily Total Sunshine (x) and Daily Total Rainfall (y) from the large data set for Leeming in 2015

Kay starts with 5th May and then selects every 10th day thereafter.

- (a) State the name of the sampling technique Kay uses.

(1)

Kay wants to find the regression line of y on x for these data.

- (b) Using your knowledge of the large data set, explain how Kay might need to clean these data before finding the equation of the regression line.

(1)

The equation of the regression line Kay finds is $y = 0.741 + 0.199x$

- (c) Using your knowledge of the large data set,

- (i) state the units of the gradient of the regression line,
(ii) give an interpretation of the y -intercept of the regression line.

(2)

Kay's teacher claimed that the greater the amount of sunshine in a day the lower the amount of rain there should be.

- (d) State, giving a reason, whether the teacher's claim is true for Kay's data.

(1)

The teacher used all the data for these variables from the large data set for Leeming in 2015, as a sample.

The teacher calculated the product moment correlation coefficient for x and y to be -0.160

In a suitable test to determine whether there is evidence to support the teacher's claim, the p -value was 0.015

- (e) Using a 5% level of significance, state the hypotheses and conclusion for this test.

(2)

- (f) For the test in part (e) describe

- (i) the sample,
(ii) a possible population.

(2)

Question 4 continued

Question 4 continued



Question 4 continued

(Total for Question 4 is 9 marks)



SUB-TOPIC 12.3

MEASURES OF LOCATION AND SPREAD

Sub-topic 12.3 · 13 questions · 130 marks · lowest tariff first



4. A study was made of adult men from region A of a country. It was found that their heights were normally distributed with a mean of 175.4 cm and standard deviation 6.8 cm.

- (a) Find the proportion of these men that are taller than 180 cm.

(1)

A student claimed that the mean height of adult men from region B of this country was different from the mean height of adult men from region A .

A random sample of 52 adult men from region B had a mean height of 177.2 cm

The student assumed that the standard deviation of heights of adult men was 6.8 cm both for region *A* and region *B*.

- (b) Use a suitable test to assess the student's claim.

You should

- state your hypotheses clearly
- use a 5% level of significance

(4)

- (c) Find the p -value for the test in part (b)

(1)

Question 4 continued

(Total for Question 4 is 6 marks)



Question 3 continued

(Total for Question 3 is 7 marks)



Question 3 continued

(Total for Question 3 is 7 marks)



Question 3 continued.

(Total for Question 3 is 8 marks)



4. Magali is studying the mean total cloud cover, in oktas, for Leuchars in 1987 using data from the large data set. The daily mean total cloud cover for all 184 days from the large data set is summarised in the table below.

Daily mean total cloud cover (oktas)	0	1	2	3	4	5	6	7	8
Frequency (number of days)	0	1	4	7	10	30	52	52	28

One of the 184 days is selected at random.

- (a) Find the probability that it has a daily mean total cloud cover of 6 or greater.

(1)

Magali is investigating whether the daily mean total cloud cover can be modelled using a binomial distribution.

She uses the random variable X to denote the daily mean total cloud cover and believes that $X \sim B(8, 0.76)$

Using Magali's model,

- (b) (i) find $P(X \geq 6)$

(2)

- (ii) find, to 1 decimal place, the expected number of days in a sample of 184 days with a daily mean total cloud cover of 7

(2)

- (c) Explain whether or not your answers to part (b) support the use of Magali's model.

(1)

There were 28 days that had a daily mean total cloud cover of 8

For these 28 days the daily mean total cloud cover for the **following** day is shown in the table below.

Daily mean total cloud cover (oktas)	0	1	2	3	4	5	6	7	8
Frequency (number of days)	0	0	1	1	2	1	5	9	9

- (d) Find the proportion of these days when the daily mean total cloud cover was 6 or greater.

(1)

- (e) Comment on Magali's model in light of your answer to part (d).

(2)

Question 4 continued

[illegible]

Question 4 continued

[illegible]

(Total for Question 4 is 9 marks)

Question 5 continued

[illegible]

Question 5 continued

This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

Question 5 continued

(Total for Question 5 is 10 marks)



5. The heights of females from a country are normally distributed with
- a mean of 166.5 cm
 - a standard deviation of 6.1 cm

Given that 1% of females from this country are shorter than k cm,

- (a) find the value of k (2)

- (b) Find the proportion of females from this country with heights between 150 cm and 175 cm (1)

A female, from this country, is chosen at random from those with heights between 150 cm and 175 cm

- (c) Find the probability that her height is more than 160 cm (4)

The heights of females from a different country are normally distributed with a standard deviation of 7.4 cm

Mia believes that the mean height of females from this country is less than 166.5 cm.

Mia takes a random sample of 50 females from this country and finds the mean of her sample is 164.6 cm

- (d) Carry out a suitable test to assess Mia's belief.
You should
- state your hypotheses clearly
 - use a 5% level of significance
- (4)**

Question 5 continued.

Question 5 continued.



Question 5 continued.

(Total for Question 5 is 11 marks)



Question 2 continued

Question 2 continued



Question 2 continued

(Total for Question 2 is 12 marks)



4. Charlie is studying the time it takes members of his company to travel to the office. He stands by the door to the office from 08 40 to 08 50 one morning and asks workers, as they arrive, how long their journey was.

(a) State the sampling method Charlie used.

(1)

(b) State and briefly describe an alternative method of non-random sampling Charlie could have used to obtain a sample of 40 workers.

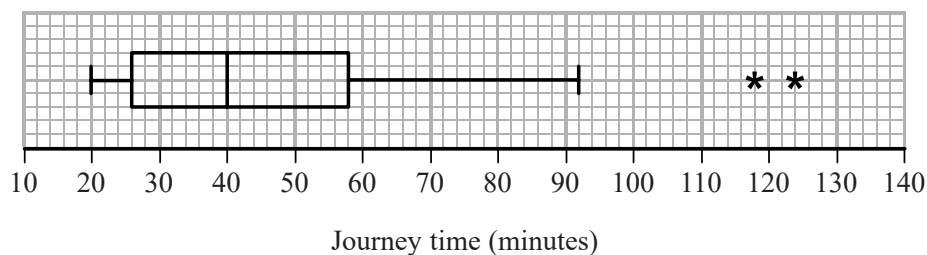
(2)

Taruni decided to ask every member of the company the time, x minutes, it takes them to travel to the office.

(c) State the data selection process Taruni used.

(1)

Taruni's results are summarised by the box plot and summary statistics below.



$$n = 95 \quad \sum x = 4133 \quad \sum x^2 = 202\,294$$

(d) Write down the interquartile range for these data.

(1)

(e) Calculate the mean and the standard deviation for these data.

(3)

(f) State, giving a reason, whether you would recommend using the mean and standard deviation or the median and interquartile range to describe these data.

(2)

Rana and David both work for the company and have both moved house since Taruni collected her data.

Rana's journey to work has changed from 75 minutes to 35 minutes and David's journey to work has changed from 60 minutes to 33 minutes.

Taruni drew her box plot again and only had to change two values.

(g) Explain which two values Taruni must have changed and whether each of these values has increased or decreased.

(3)

Question 4 continued



Question 4 continued



Question 4 continued

(Total for Question 4 is 13 marks)



5. A machine puts liquid into bottles of perfume. The amount of liquid put into each bottle, D ml, follows a normal distribution with mean 25 ml

Given that 15% of bottles contain less than 24.63 ml

- (a) find, to 2 decimal places, the value of k such that $P(24.63 < D < k) = 0.45$

(5)

A random sample of 200 bottles is taken.

- (b) Using a normal approximation, find the probability that fewer than half of these bottles contain between 24.63 ml and k ml

(3)

The machine is adjusted so that the standard deviation of the liquid put in the bottles is now 0.16 ml

Following the adjustments, Hannah believes that the mean amount of liquid put in each bottle is less than 25 ml

She takes a random sample of 20 bottles and finds the mean amount of liquid to be 24.94 ml

- (c) Test Hannah's belief at the 5% level of significance.
You should state your hypotheses clearly.

(6)

Question 5 continued

[illegible]

Question 5 continued

(Total for Question 5 is 13 marks)



Question 5 continued



Question 5 continued



Question 5 continued

(Total for Question 5 is 14 marks)



Question 5 continued

Question 5 continued



Question 5 continued



Question 5 continued

(Total for Question 5 is 15 marks)



(Total for Question 2 is 3 marks)



SUB-TOPIC 12.4

OUTLIERS AND CLEANING DATA

Sub-topic 12.4 · 2 questions · 17 marks · lowest tariff first



[illegible]

(Total for Question 3 is 6 marks)



2.

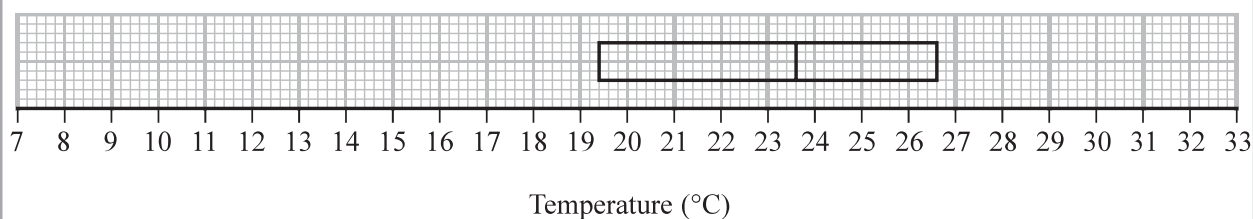


Figure 1

The partially completed box plot in Figure 1 shows the distribution of daily mean air temperatures using the data from the large data set for Beijing in 2015

An outlier is defined as a value
 more than $1.5 \times \text{IQR}$ below Q_1 or
 more than $1.5 \times \text{IQR}$ above Q_3

The three lowest air temperatures in the data set are 7.6°C , 8.1°C and 9.1°C
 The highest air temperature in the data set is 32.5°C

(a) Complete the box plot in Figure 1 showing clearly any outliers.

(4)

(b) Using your knowledge of the large data set, suggest from which month the two outliers are likely to have come.

(1)

Using the data from the large data set, Simon produced the following summary statistics for the daily mean air temperature, $x^\circ\text{C}$, for Beijing in 2015

$$n = 184 \quad \sum x = 4153.6 \quad S_{xx} = 4952.906$$

(c) Show that, to 3 significant figures, the standard deviation is 5.19°C

(1)

Simon decides to model the air temperatures with the random variable

$$T \sim N(22.6, 5.19^2)$$

(d) Using Simon's model, calculate the 10th to 90th interpercentile range.

(3)

Simon wants to model another variable from the large data set for Beijing using a normal distribution.

(e) State two variables from the large data set for Beijing that are **not** suitable to be modelled by a normal distribution. Give a reason for each answer.

(2)

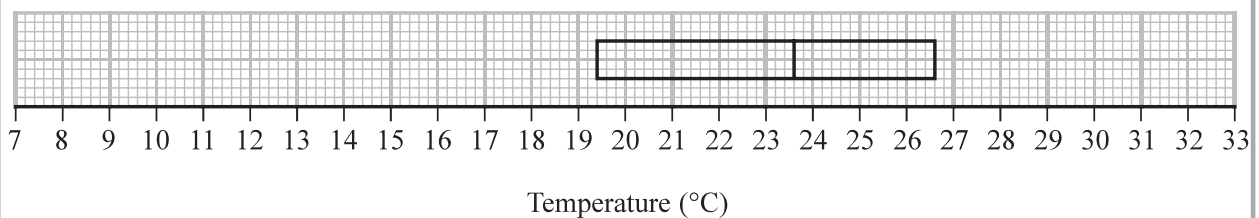
This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.



This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

Question 2 continued

Only use this grid if you need to redraw your box plot.



(Total for Question 2 is 11 marks)

MARK SCHEMES

official mark schemes



Qu 3	Scheme	Marks	AO
(a)	$[68 - 7 =]$ 61 (only)	B1 (1)	1.1b
(b)	$[25 - 14] =$ 11	B1 (1)	1.1b
(c)	$\left[\mu \text{ or } \bar{x} = \frac{607.5}{27} = \right] =$ 22.5	B1 (1)	1.1b
(d)	$\sigma = \sqrt{\frac{17\,623.25}{27} - "22.5"{}^2}$ <u>or</u> $\sqrt{146.4629\dots}$ $= 12.10218\dots$ awrt 12.1	M1 A1 (2)	1.1b 1.1b
(e)	$\mu + 3\sigma = "22.5" + 3 \times "12.1\dots" =$ awrt 59 so only one outlier	B1ft (1)	1.1b
(f)	Median increases implies that both values must be > 20 Mean is the same means that $a + b = 45$ So possible values are: e.g. $b = 21$ and $a = 24$ (o.e.)	M1 M1 A1 (3)	3.1b 1.1b 2.2b
(g)	Both values will be less than 1 standard deviation from the mean and so the standard deviation of all 29 values will be smaller	B1 (1)	2.4
		(10 marks)	
Notes			
(a)	B1 for correctly interpreting the box plot to find the range (more than 1 answer is B0)		
(b)	B1 for correct understanding of IQR and answer of 11		
(c)	B1 for 22.5 only (or exact equivalent such as $\frac{45}{2}$). Allow 22 mins and 30 secs.		
(d)	M1 for a correct expression including square root. Allow $\sqrt{146}$ or better. Ft their mean A1 for awrt 12.1 NB Allow use of $s = 12.3327\dots$ or awrt 12.3		
(e)	B1ft for a correct calculation or value based on their μ and σ and compatible conclusion		
(f)	1 st M1 Correct start to the problem and a correct statement about the values based on median Allow if their final two values are both > 20 2 nd M1 for a correct explanation leading to equation $a + b = 45$ (o.e. e.g. equidistant from mean) Allow if their final two values sum to 45 A1 for a correct pair of values (both > 20 with a sum of 45) and at least some attempt to explain how their values satisfy at least one of the conditions (both > 20 <u>or</u> $a + b = 45$). Ignore $a =$ or $b =$ labels NB The values for a and b do not need to be integers.		
(g)	B1 for a correct explanation. Must mention that both values are less than 1 sd (ft their answer to (d)) from the mean		

Qu 6	Scheme	Marks	AO
(a)	$2 \times 4.2, 4 \times 4, 4 \times 3.5, 10 \times 1$ ($= 8.4 + 16 + 14 + 10 = 48.4$) $[\text{So } P(10 < T < 30) =] \quad \left[\frac{48.4}{90} \right] = \frac{121}{225} = 0.53777... \quad \underline{\underline{0.53 \sim 0.54}}$ (2sf OK)	M1 A1	1.1b 1.1b
(b)	(Not suitable as) data is not symmetric <u>or</u> is skew (normal is symmetric) (“Even” distribution or a diagram <u>on its own</u> is not enough so B0)	(2) B1 (1)	2.4
(c)	$\int x e^{-x} (dx) = \int x d(-e^{-x})$ $= [-x e^{-x}] - \int (-e^{-x}) (dx) \quad (+c)$ $\int_0^n x e^{-x} (dx) = [-x e^{-x} - e^{-x}]_0^n = (-n e^{-n} - e^{-n}) - [-(0) - 1]$ $= 1 - (n+1)e^{-n} \quad (*)$	M1 A1 dM1 A1cso*	2.1 1.1b 1.1b 1.1b
(d)	Require area = 90 i.e. $k \int_{(0)}^{(n)} x e^{-x} dx = 90$ (ignore limits) Using the result in part (c) with $n = 4$ gives $k[1 - 5e^{-4}] = 90$ $(k =) \underline{\underline{99.0729...}} \quad (*)$	M1 M1 A1cso*	3.1a 2.1 1.1b
(e)(i)	$[P(10 < T < 30) =] \quad 0.64863... \quad \text{awrt } \underline{\underline{0.649}}$	B1	1.1b
(ii)	$[\text{No. of patients} =] \quad (99) \left[(1 - 4e^{-3}) - (1 - 2e^{-1}) \right] \quad (= 53.1...)$ $\text{Prob} = \frac{0.5366... \times 99}{90} = 0.59027... [\text{or } 0.5907...] = \text{awrt } \underline{\underline{0.590 \text{ or } 0.591}}$	M1 A1	3.4 3.2a
(f)	eg Patients might stay longer than 40 hours (Can ignore other comments unless clearly contradictory.)	B1	3.5b
(14 marks)			
Notes			
(a)	M1 for an attempt to find the number between 10 and 30 (2 correct products or 48 or 48.4 seen) A1 for 2sf answer in $[0.53 \sim 0.54]$ NB use of 48 gives 0.5333... [Correct ans implies 2/2]		
(b)	B1 for a comment suggesting not suitable based on (lack of) symmetry <u>or</u> “not bell shaped”		
(c)	1 st M1 for attempting integration by parts in right direction. Must have $u = x$ and $v = \pm e^{-x}$ 1 st A1 for a correct first step, correct first integration and expression for second integral 2 nd dM1 (dep on 1 st M1) for all integration attempted and some use of at least one limit * 2 nd A1 for cso with no incorrect working seen. Minimum is correct int and use of limits seen.		
(d)	1 st M1 for realising need area under the curve (implied by the integral) = 90 2 nd M1 for use of (c) with $n = 4$ and set = 90 May be implied by sight of 99.07... or better * A1cso for $k = 99$ or awrt 99.1 NB Allow use of $k = 99$ and show area = awrt 89.9 with a conclusion to score 3/3		
(e)(i)	B1 for awrt 0.649		
(ii)	M1 for use of (c) with $n = 1$ and $n = 3$ Don’t need the 99. Implied by sight of awrt 0.54 A1 for awrt 0.590 or awrt 0.591 Allow 0.59 from correct working seen.		
(f)	B1 eg for comment, in context, about the upper limit for <u>time</u> (<u>t or x</u>)(time/hour may be implied)		

Qu 1	Scheme	Marks	AO
	2nd M1 for correct interpretation of “at least 12” and writing or using $1 - P(D \leq 11)$ We are <u>not</u> attempting to ft their incorrect 0.155 on our calculators here. A1 for awrt 0.212 [Answer only 3/3]		
(c)	B1 for 100 but must be seen in part (c) i.e. between (b) and (d)		
(d)	1st M1 for attempting normal with mean = 100 <u>or</u> ft their answer to (c) May be implied by the correct mean and a correctly labelled s.d. (σ) or var (σ^2)		
	1st A1 for correctly labelled standard deviation allow $\sqrt{\frac{250}{3}} = \sqrt{83.3...} = 9.1(28....)$ <u>or</u> correctly labelled variance. Implied by $N(\mu, \frac{250}{3})$ <u>or</u> correct answer.		
	2nd M1 for attempt at continuity correction i.e. sight of 95 ± 0.5 2nd A1 for awrt 0.689 [Answer only 4/4]		
NB	If they don't state the model for 1st M1 but just give probabilities with probability statements (Y is any letter):		
$\sigma = \frac{250}{3}$	1 st M1 implied by: $P(Y > 94.5) = 0.52(63...)$, $P(Y > 95) = 0.52(39...)$, $P(Y > 95.5) = 0.52(15..)$		
No cc	1st M1 1st A1 implied by: $P(T > 95) = 0.70(805...)$ 1st M1 1st A1 2nd M1 implied by: $P(T > 94.5) = 0.72(657...)$		
	Exact binomial gives 0.68567... and will likely score 0/4		

Qu 2	Scheme	Marks	AO
(a)	e.g. The <u>height</u> (h) <u>decreases</u> by about <u>1.28 m</u> for <u>each second</u> of the flight	B1 (1)	3.4
(b)	$H_0: \rho = 0$ $H_1: \rho < 0$ [5% 1-tail cv =] (+) 0.5494 [$r = -0.510$ not sig] there is <u>insufficient</u> (o.e.) evidence of a negative <u>correlation</u> between <u>height</u> (or h) and <u>time</u> (or t)	B1 M1 A1 (3)	2.5 1.1a 2.2b
(c)	No – since points seem to follow a curve/quadratic (rather than a line) or since points are “non-linear” but regression line/ model is linear or e.g. between ($t = 5$ and 7) height drops by much more than 2.56 m or e.g. gradient is positive up to $t = 3.5$ (line gradient < 0) or e.g. gradient is positive initially (line gradient < 0) or e.g. gradient is positive and then negative	B1 (1)	2.4
(d)	[$h = 38.1 - 0.78 (t - k)^2$ with] a suitable k i.e. in the range 3~4.5	B1 (1)	3.3
(6 marks)			
Notes			
(a)	B1 for a suitable interpretation in context [value can be 1.3 or 1.28 or “just over 1”] per sec Must have underlined words (o.e.) and units “m” or metres and “s” or seconds NB “descends” implies “height decreases” Condone e.g. “decreases by – 1.28 m”		
(b)	B1 for both hypotheses correct in terms of ρ [accept a p or p but not r or r] Must be attached to H_0 and H_1 M1 for a critical value corresponding to their H_1 : 1-tail: awrt ± 0.549 or 2-tail (B0 scored for H_1) : awrt ± 0.632 (tables 0.6319) If hypotheses are in words and can deduce whether one or two-tail then use their words. If no hypotheses or their H_1 is not clearly one or two-tail assume one-tail A1 a correct conclusion in context mentioning <u>correlation</u> and <u>height</u> and <u>time</u> A comparison or statement such as “not sig” is not needed but if seen must be correct. Do NOT award this A mark if contradictory comments or working seen e.g. “reject H_0 ” or comparison of 0.510 with significance level of 0.05 or e.g. $-0.549 > -0.510$ NB Can award B0M1A1		
SC	B0(for 2-tail) M0(for cv = ± 0.549) scored: Allow 1 mark (score as B0M0A1) for conclusion such as: “ <u>insufficient</u> evidence of (negative) <u>correlation</u> between <u>height</u> and <u>time</u> of flight”		
(c)	B1 for saying no and giving a suitable supporting reason Don’t allow “correlation” on its own instead of “gradient” B0 for simply saying “points don’t lie close to a straight line” Need mention of curve or some other feature of scatter plot that <u>differs</u> from regression line. B0 for just “non-linear” without mention of the model being linear B0 for simply comparing 1 or 2 points – need a comment about general pattern		
(d)	B1 for a value of k in the range [3, 4.5] Do not need $k = \dots$ Accept a value embedded in Jane’s model. ISW any errors in multiplying out bracket.		



Qu 2	Scheme	Marks	AO
(a)	Negative	B1 (1)	1.2
(b)	Marc's suggestion <u>is compatible</u> because it's <u>negative correlation</u>	B1 (1)	2.4
(c)	$(r =) - 0.54458266...$ awrt <u>-0.545</u>	B1 (1)	1.1b
(d)	$H_0 : \rho = 0$ $H_1 : \rho < 0$ [5% 1-tail cv =] $(\pm) 0.4259$ (significant result / reject H_0) There <u>is</u> evidence of negative <u>correlation</u> between the <u>number of letters</u> in (or <u>length</u> of) a student's last <u>name</u> and their first <u>name</u>	B1 M1 A1 (3)	2.5 1.1a 2.2b
(6 marks)			
Notes			
(a)	B1 for "negative" Allow "slight" or "weak" etc Allow a description e.g. "as x increases y decreases" or in context e.g. "people with longer last names tend to have shorter first names" A comment of "negative skew" is B0 Need to see distinct or separate responses for (a) and (b)		
(b)	B1 for a comment that suggests data is compatible with the suggestion and a suitable reason such as "there is negative correlation" <u>or</u> a description in x and y or in context <u>or</u> the points lie close to a line with <u>negative gradient</u> <u>or</u> draw line $y = x$ and state that <u>more points below the line</u> so <u>supports (or is compatible with)</u> his suggestion A reason based on just a single point is B0 e.g. "11 letters in last name has only 5 in first name"		
(c)	B1 for awrt -0.545		
(d)	B1 for both hypotheses correct in terms of ρ M1 for a critical value compatible with their H_1 : 1-tail: awrt ± 0.426 (condone ± 0.425) or 2-tail (B0 scored for H_1) : awrt ± 0.497 If hypotheses are in words and can deduce whether one or two-tail then use their words. If no hypotheses or their H_1 is not clearly one or two tail assume one-tail A1 for compatible signs between cv and r and a correct conclusion in context mentioning <u>correlation</u> and <u>number of letters</u> or <u>length</u> and <u>name</u> (ft their value from (c)) Do NOT award this A mark if contradictory comments or working seen e.g. "accept H_0 " or comparison of 0.426 with significance level of 0.05 etc NB The M1A1 can be scored independently of the hypotheses		

Qu 1	Scheme	Marks	AO
(a)	1 st B1 for a correct set of values for c . Allow $\{\frac{1}{8}, \frac{2}{8}, \dots, \frac{8}{8}\}$ 2 nd B1ft for correct probs from their values for c , consistent with discrete uniform distrib'n Maybe as a prob. function. Allow $P(X = x) = \frac{1}{9}$ for $0 \leq x \leq 8$ provided $x = \{0, 1, 2, \dots, 8\}$ is clearly defined somewhere.		
(b)	B1 for using correct model to get $\frac{4}{9}$ (o.e.)		
SC	Sample space $\{1, \dots, 8\}$ If scored B0B1 in (a) for this allow $P(C < 4) = \frac{3}{8}$ to score B1 in (b)		
(c)	B1ft for comment that states that the model proposed is or is not a good one based on their model in part (a) and their probability in (b) $(b) - 0.315 > 0.05$ Allow e.g. “it is not suitable”; “it is not accurate” etc $(b) - 0.315 \leq 0.05$ Allow a comment that suggests it <u>is</u> suitable No prob in (b) Allow a comparison that mentions 50% or 0.5 and rejects the model No prob in (b) and no 50% or 0.5 or $(b) > 1$ scores B0 Ignore any comments about location or weather patterns.		
(d)	B1 for a sensible refinement considering variations in month or location Just saying “not uniform” is B0 Context & “non-uniform” Allow mention of different locations, months <u>and</u> non-uniform <u>or</u> use more locations to form a new distribution with probabilities based on frequencies Context & “binomial” Allow mention of different locations, months <u>and</u> binomial Just refined model Model must be outlined and discrete and non-uniform e.g. higher probabilities for more cloud cover <u>or</u> lower probabilities for less cloud cover Continuous model Any model that is based on a continuous distribution. e.g. normal is B0		

Qu 2	Scheme	Marks	AO
(a)	$H_0 : \rho = 0$ $H_1 : \rho < 0$ Critical value: -0.6215 (Allow any cv in range $0.5 < cv < 0.75$) $r < -0.6215$ so significant result and there is evidence of a negative correlation between w and t	B1 M1 A1 (3)	2.5 1.1a 2.2b
(b)	e.g. As temperature increases people spend more time on the beach and less time shopping (o.e.)	B1 (1)	2.4
(c)	Since r is close to -1 , it is consistent with the suggestion	B1 (1)	2.4
(d)	t will be the explanatory variable since sales are likely to depend on the temperature	B1 (1)	2.4
(e)	Every degree rise in temperature leads to a drop in weekly earnings of £171	B1 (1)	3.4
		(7 marks)	
Notes			
(a)	B1 for both hypotheses in terms of ρ M1 for the critical value: sight of ± 0.6215 or any cv such that $0.5 < cv < 0.75$ A1 must reject H_0 on basis of comparing -0.915 with -0.6215 (if $-0.915 < 0.6215$ is seen then A0 but may use $ r $ o.e. which is fine) <u>and</u> mention “negative”, “correlation/relationship” and at least “ w ” and “ t ”		
(b)	B1 for a suitable <u>reason to explain</u> negative correlation using the context given. e.g. “As temperature drops people are more likely to go shopping (than to the beach)” e.g. “As temperature increases people will be outside rather than in shops” A mere description in context of negative correlation is B0 SO e.g. “As temperature increases people don’t want to go shopping/buy clothes” is B0 e.g. “Less clothes needed as temp increases” is B0		
(c)	B1 for a suitable reason e.g. “strong”/”significant”/”near perfect” “correlation”, $ r $ close to 1 <u>and</u> saying it is consistent with the suggestion. Allow “yes” followed by the reason.		
(d)	B1 For identifying t <u>and</u> giving a suitable reason. Need idea that “ w <u>depends</u> on t ” or “ w <u>responds</u> to t ” or “ t <u>affects</u> w ” (o.e.) Allow t (temperature) <u>affects</u> the other variable etc Just saying “ t is the independent variable” or “ t <u>explains</u> change in w ” is B0 N. B. Suggesting causation is B0 e.g. “ t causes w to decrease”		
(e)	B1 for a description that conveys the idea of rate per degree Celsius. Must have 171, condone missing “£” sign.		

Qu 2	Scheme	Marks	AO
(a)	Negative	B1 (1)	1.2
(b)(i)	Rainfall	B1	2.2b
(ii)	mm <u>or</u> Pressure hPa or Pascals or hectopascals or mb or millibars	B1ft (2)	1.1b
(c)	$H_0: \rho = 0$ $H_1: \rho \neq 0$ Critical value: $-0.361(0)$ $r < -0.3610$ so significant result and there is evidence of a correlation between Daily Total <u>Sunshine</u> and Daily Maximum Relative <u>Humidity</u>	B1 M1 A1 (3)	2.5 1.1b 2.2b
(d)	Humidity is high and there is evidence of correlation and $r < 0$ So expect amount of sunshine to be <u>lower</u> than the <u>average</u> for Heathrow(oe)	B1 (1)	2.2b
(7 marks)			
Notes			
(a)	B1 for stating negative. “Negative skew” is B0 though		
(b)(i)	B1 for mentioning “rainfall” (allow “rain” <u>or</u> “precipitation”) <u>or</u> “pressure” (if more than 1 answer both must be correct) NB the other quantitative variable for Perth is: Daily Mean Wind Speed and scores B0 [Not allowed “wind speed” since $r = +0.15$ and in winter might expect wind to raise temp]		
(ii)	B1ft for giving the correct units. If Daily Mean Wind Speed (kn) or knots “Wind speed” and “knots” would score B0B1 but any other variable scores B0B0		
(c)	B1 for both hypotheses correct in terms of ρ M1 for the correct critical value compatible with their H_1 : allow $\pm 0.361(0)$ If the hypotheses are 1-tail then allow cv of ± 0.3061 e.g. Alternative hypothesis with $r < \pm 0.377$ implies a one-tail test <u>or</u> H_0 and H_1 in words saying “ H_0 : there is no correlation, H_1 : there is correlation” is two-tail If there are no hypotheses (or they are nonsensical) assume 2-tail so M1 for $\pm 0.361(0)$ A1 for a correct conclusion in context based on comparing -0.377 with their cv. Condone incorrect inequality e.g. $-0.3610 < -0.377$ as long as they reject H_0 Do not accept contradictory statements such as “accept H_0 so there is evidence of ...” Can say “support for Stav’s <u>belief</u> ”(o.e.e.g. “claim”) or “evidence of a correlation between <u>sunshine</u> and <u>humidity</u> ” condone “negative correlation” or comments such as “if humidity is high amount of sunshine will be low”		
(d)	B1 for stating <u>low</u> amount of sunshine (o. e.) and some reference to $r < 0$ or fog Check for the following 2 features: (i) low sunshine: allow ≤ 5 hrs (LDS mean for 2015 is 5.3, humidity 97% is 4.1, $\geq 97\%$ is 3.1) (ii) negative correlation may be described in words e.g. “high humidity gives low sunshine” <u>or</u> fog (LDS says $>95\%$ humidity is foggy) so less sunshine		



Question	Scheme		Marks	AOs
3(a)	$H_0 : \rho = 0 \qquad H_1 : \rho > 0$		B1	2.5
	Critical value 0.3438		M1	1.1a
	(0.446 > 0.3438) so there is evidence that the product moment correlation coefficient (pmcc) is greater than 0/there is positive correlation		A1	2.2b
			(3)	
(b)	The value is close(r) to 1 or there is strong(er) (positive) correlation		B1	2.4
			(1)	
(c)	$\log_{10} y = -1.82 + 0.89(\log_{10} x)$	$y = ax^n \rightarrow$ $\log_{10} y = \log_{10} (ax^n)$	M1	1.1b
	$y = 10^{-1.82+0.89(\log_{10} x)}$	$\log_{10} y = \log_{10} a + \log_{10} x^n$	M1	2.1
	$y = 10^{-1.82} \times 10^{0.89(\log_{10} x)}$ [$= 10^{-1.82} \times 10^{(\log_{10} x)^{0.89}}$]	$\log_{10} y = \log_{10} a + n \log_{10} x$ [$\log_{10} a = -1.82, n = 0.89$]	M1	1.1b
	$y = 0.015x^{0.89}$	$y = 0.015x^{0.89}$	A1A1	1.1b 1.1b
			(5)	
	(9 marks)			
Notes				
(a)	B1: for both hypotheses correct in terms of ρ M1: for the critical value: sight of 0.3438 or any cv such that $0.25 < cv < 0.45$ A1: a comment suggesting a significant result/ H_0 is rejected on the basis of <u>seeing</u> +0.3438 and which mentions “pmcc/correlation/relationship” and “greater than 0/positive” (not just $\rho > 0$) or an answer in context e.g. ‘as “income”(o.e.) increases, “CO ₂ /emissions”(o.e.) increases’ A contradictory statement scores A0 e.g. ‘Accept H_0 , therefore positive correlation’			
(b)	B1: for suitable reason e.g. r is close(r) to 1 or “strong(er)”/“near perfect” “correlation” Do not allow ‘association’			
(c)	For both methods, once an M0 is scored, no further marks can be awarded and condone missing base 10 throughout Method 1: (working to the model) M1: Correct substitution for both c and m (may be implied by 2 nd M1 mark) M1: Making y the subject to give an equation in the form $y = 10^{a+b(\log_{10} x)}$ (may be implied by 3 rd M1 mark) M1: Correct multiplication to give an equation in the form $y = 10^a \times 10^{b(\log_{10} x)}$ (this line implies M1M1M1 provided no previous incorrect working seen) Method 2: (working from the model) M1: Taking the log of both sides (may be implied by 2 nd M1 mark) M1: Correct use of addition rule (may be implied by 3 rd M1 mark) M1: Correct multiplication of power (this line implies M1M1M1 provided no previous incorrect working seen) A1: $n = 0.89$ or $a = \text{awrt } 0.015$ or $y = ax^{0.89}$ or $y = \text{awrt } 0.015x^n$ (dep on M3) A1: $n = 0.89$ and $a = \text{awrt } 0.015$ / $y = \text{awrt } 0.015x^{0.89}$ (dep on M3) do not award the final A1 if answer is given in an incorrect form e.g. $y = 0.015 + x^{0.89}$			

Question	Scheme		Marks	AOs
6(a)	eg As the number of minutes <u>exercise</u> (m) increases the resting <u>heart rate</u> (h) decreases or the gradient of the curve is becoming flatter with increasing m : diminishing effect of each <u>additional minute</u> of <u>exercise</u>		B1	2.4
			(1)	
(b)	$H_0 : \rho = 0$ $H_1 : \rho < 0$		B1	2.5
	Critical value – 0.3887 (Allow $\pm$)		M1	1.1b
	There is evidence that the product moment <u>correlation</u> is <u>less than 0/</u> <u>there is a negative correlation</u>		A1	2.2b
			(3)	
(c)	$\log_{10} h = -0.05 \log_{10} m + 1.92$	$h = am^k \rightarrow \log_{10} h = \log_{10} am^k$	M1	1.1b
	$\log_{10} h = -\log_{10} m^{0.05} + 1.92$ or $\log_{10} h = \log_{10} m^{-0.05} + 1.92$ or $h = 10^{1.92 - 0.05 \log_{10} m}$ oe	$\log_{10} h = \log_{10} a + \log_{10} m^k$ or $\log_{10} a = 1.92$	M1	2.1
	$\log_{10} hm^{0.05} = 1.92$ or $\log_{10} \left(\frac{h}{m^{-0.05}} \right) = 1.92$ or $h = 10^{1.92} \times 10^{-0.05 \log_{10} m}$ oe	$\log_{10} h = \log_{10} a + k \log_{10} m$	M1	1.1b
	$hm^{0.05} = 10^{1.92}$ or $\frac{h}{m^{-0.05}} = 10^{1.92}$ or $h = 10^{1.92} \times 10^{\log_{10} m^{-0.05}}$	$\log_{10} a = 1.92$ and $k = -0.05$	M1	1.1b
	$h = 10^{1.92} m^{-0.05}$ or $h = 83.17...m^{-0.05}$ or $a = \text{awrt } 83.17$ and $k = -0.05$		A1	1.1b
			(5)	
Notes: (9 marks)				
(a)	B1	eg Idea as one increases the other decreases (in context). Allow use of m and h eg As m increases h decreases. Do not allow negative correlation with no context or $\rho < 0$ Allow there is a negative correlation/association/relationship/exponential between minutes <u>exercise</u> (m) and resting <u>heart rate</u> (h) oe		
(b)	B1	Both hypotheses correct in terms of ρ (allow p)		
	M1	For the cv of -0.3887 or any cv such that $0.3 < \text{cv} < 0.5$		
	A1	Independent of hypotheses. Correct conclusion that implies reject H_0 on basis of seeing -0.3887 or if they give 0.3887 we must see the comparison $0.3887 < 0.897$ and which mentions “pmcc/correlation/relationship” and less than 0/ negative or $\rho < 0$ A contradictory statement scores A0 eg Accept H_0 therefore negative correlation		
(c)		In this part once M0 is scored no more marks can be scored. Condone no base		
	M1	May be implied by 2nd M1 mark Method 1: Correct substitution for both x and y Method 2 : Taking the log of both sides		
	M1	May be implied by 3rd M1 mark Method 1: Correct use of the power log rule or making h the subject Method 2 : Correct use of the addition/subtraction log rule		
	M1	This line implies M1M1M1 Method 1: Correct use of the addition/subtraction log rule or eq ⁿ in the form $h = 10^{1.92} \times 10^{-0.05 \log m}$ Method 2: A second correct step for correct use of the power log rule		
	M1	This line implies M1M1M1M1 Method 1: Correct removal of logs or $h = 10^{1.92} \times 10^{\log m^{-0.05}}$ Method 2: Log a (or a) and k correct		
	A1	Allow $h = \text{awrt } 83.2m^{-0.05}$ NB award 5/5 for $a = \text{awrt } 83.2$ and $k = -0.05$ or $h = \text{awrt } 83.2...m^{-0.05}$ or $h = 10^{1.92} m^{-0.05}$		



Question	Scheme		Marks	AOs
4(a)	Systematic (sampling)		B1	1.2
			(1)	
4(b)	The (Daily Total) <u>Rainfall</u> data may contain “ <u>tr</u> ” entries (these will need a suitable value substituted before calculations can take place.)		B1	2.4
			(1)	
4(c)(i)	mm/h (o.e.) e.g. mm h^{-1} or mm hrs^{-1} or $\frac{\text{mm}}{\text{hours}}$		B1	1.1b
(ii)	When there is <u>no sun</u> (shine) there is (on average) 0.7(41) (mm) <u>of rain</u> or the amount of <u>rain</u> when there is <u>no sun</u> (shine) (o.e.)		B1	2.4
			(2)	
4(d)	e.g. Not consistent (since); Kay’s line says <u>positive</u> correlation or gradient or regression coefficient or regression line or 0.199 is <u>positive</u>		B1	2.4
			(1)	
4(e)	$H_0 : \rho = 0$ and $H_1 : \rho < 0$		B1	2.5
	[p-value < 5% so significant result] there is evidence to <u>support</u> the <u>teacher’s claim</u>		B1	2.2b
			(2)	
4(f)(i)	The sample is <u>rainfall</u> (y) and <u>sunshine</u> (x) for <u>May~Oct</u> in (Leeming) in <u>2015</u>		B1	1.1b
(ii)	e.g. (rainfall and sunshine) for: (Leeming) for all of 2015, or Leeming anytime (not just May~Oct) or Leeming for May~October for other years too or May ~Oct in UK etc		B1	3.3
			(2)	
(9 marks)				
Notes:				
(a)	B1	for systematic (ignore other non-contradictory descriptions). Condone misspelling. If a clear choice is given e.g. stratified or systematic we take the final answer.		
(b)	B1	for mention of “ <u>tr</u> ” or “trace” entries in <u>rain</u> (fall) [or y] data. Ignore “n/a”. Ignore any comment about what to do with trace.		
(c)(i)	B1	for mm/h or equivalent in words.		
(ii)	B1	for the idea of amount of <u>rain</u> when <u>no sun</u> . “Minimum rain” is B0 Don’t need value or units but if given must be correct or consistent with (i)		
(d)	B1	for a suitable reason and saying <u>not</u> consistent (o.e.) e.g. “false” or “untrue” or “no” Reason only needs to be about the line and may be a description e.g. as x increases y increases or at least 2 values of x substituted and y values evaluated		
(e)	B1	for both hypotheses correct in terms of ρ (condone attempt at ρ that looks like p)		
	B1	for a correct conclusion in context with no contradiction. Using words “ <u>support</u> ” and “ <u>teacher’s claim</u> ” or “ <u>negative correlation</u> ” and “ <u>sun</u> or <u>x</u> ” and “ <u>rain</u> or <u>y</u> ” [If a comparison is given it must be $0.015 < 0.05$]		
(f)(i)	B1	for recognising that all the data/days for <u>these variables</u> in <u>LDS</u> for <u>2015</u> is the sample. Condone omitting “Leeming” B0 for “sunshine and rainfall for Leeming in 2015” It could be the population.		
(ii)	B1	for a suitable attempt to describe a population. The teacher’s sample must be a subset.		



Qu 4	Scheme	Marks	AO
(a)	[Let N = height from region A; $P(N > 180) =]$ 0.24937... awrt <u>0.249</u>	B1 (1)	1.1b
(b)	$H_0: \mu = 175.4$ $H_1: \mu \neq 175.4$ [S = height from region B] $\bar{S} \sim N\left(175.4, \frac{6.8^2}{52}\right)$ Allow $\sigma^2 =$ awrt 0.889 [$P(\bar{S} > 177.2)] = 0.02814...$ [0.028... > 0.025, Not sig, do not reject H_0] <u>Insufficient</u> evidence to <u>support</u> student's <u>claim</u>	B1 M1 A1 A1 (4)	2.5 3.3 3.4 2.2b
(c)	[p -value = $2 \times 0.02814... =$] 0.05628... in range <u>0.056~0.06</u> or <u>5.6(%)~6(%)</u>	B1ft (1)	1.2
(6 marks)			
Notes			
(a)	B1 for awrt 0.249		
(b)	B1 for both hypotheses correct in terms of μ (See below for one-tail test) M1 for selecting the correct model, may be implied by standardisation using correct values <u>or</u> may be implied by a correct <u>value</u> in 1 st A1 e.g.(Prob =) 0.028 or awrt 0.972, ($Z =$) 1.9(08..) (CV=) 177.25 Condone use of S (or any other letter) instead of $\bar{S}$ Condone use of $\bar{S} \sim N\left(177.2, \frac{6.8^2}{52}\right)$ but this will lose 2nd A mark		
ALT	1 st A1 for probability of awrt 0.028 (allow 0.03 if $P(\bar{S} > 177.2)$ is seen) Condone $1 - 0.02814... = 0.9718...$ (awrt 0.972) only if clearly compared with 0.975 Allow $Z = 1.9(088...)$ <u>and</u> comparison with 1.96 (or better: calc gives 1.95996...) <u>or</u> CR of $[\bar{S}]... 177.248...$ (awrt 177.25) Allow $[\bar{S}] > 177.248...$ (awrt 177.25) Implied by diagram or correct interpretation of inequality with their CV (Ignore any attempt at a lower CR for $\bar{S}$) 2 nd A1 (dep on 1 st A1 and use of correct model. Use of $N(177.2, ...)$ scores A0) for a conclusion using context: e.g. does <u>not support</u> student's <u>claim</u> <u>or</u> e.g. <u>insufficient</u> evidence of a <u>difference in heights</u> Do not allow 2 nd A mark for contradictory statements e.g. “significant” so “no support for claim”		
(c)	B1ft for answer in range 0.056~0.06 or 5.6%~6% (Ranges are inclusive, condone missing %) (can fit their probability, provided < 0.5, from part (b) but not 0.025 leading to 5%)		
NB	One-tail test [Max of 3/5 for (b) and (c)] In (b) B0 (hypotheses) M1(model as above) 1 st A1[for probability <u>or</u> Z compared with 1.6449 <u>or</u> CR $[\bar{S}]... \text{ or } > 176.95...$ (awrt 177)] 2 nd A1 for conclusion in context that <u>supports claim</u> or “ <u>heights of men from B is different from/greater than</u> from A” In (c) B0		

Question	Scheme	Marks	AOs
3(a)	tr	B1	1.2
		(1)	
(b)(i)	$\mu = \frac{174.9}{31} = 5.6419\dots$ awrt 5.64	B1	1.1b
(ii)	$\sigma_r = \sqrt{\frac{3523.283}{31} - \mu^2}$	M1	1.1b
	$= 9.04559\dots$ awrt 9.05	A1	1.1b
		(3)	
(c)	Leuchars is in the North and Camborne is in the South	M1	2.4
	The mean is smaller for Leuchars than Camborne therefore there is no evidence that Dian's belief is true	A1ft	2.2b
		(2)	
(d)	eg $p = 0.27$ is unlikely to be constant.	B1	2.4
		(1)	
(7 marks)			
Notes:			
(a)	B1	Allow Tr or trace or Trace	
(b)(i)	B1	For a correct mean awrt 5.64	
(ii)	M1	For a correct expression for sd including the $\sqrt{\quad}$ Ft their mean	
	A1	awrt 9.05 (Allow $s = 9.1932\dots$ awrt 9.19) NB awrt to 9.05 or 9.19 with no working is M1 A1	
(c)	M1	For stating Leuchars is North of Camborne oe eg Camborne is further south	
	A1ft	M1 must be awarded. A correct conclusion and correct comment about the means ft their mean in (b) Allow No	
	SC	for No and there are only 2 places used so there is insufficient data. Mark as M0A1 on open	
(d)	B1	A correct reason referring to - independence (needs context as to what is independent) eg consecutive 14 days unlikely to be independent. - probability [of rain] not being constant. - Allow a comment that conveys the idea that the proportion of days with no rain will be different over the year.	

Qu 3	Scheme	Marks	AO
(a)	Need to replace tr with a numerical value Value of tr is between 0 and 0.05 suggest using e.g 0.025 , 0 <u>or</u> value ,, 0.05	M1 A1 (2)	1.2 1.1b
(b)(i)	$\left[\bar{x} = \frac{389.3 \sim 390.8}{184} \right] = 2.119\dots$ awrt <u>2.12</u> allow $\frac{195}{92}$ or $2\frac{11}{92}$	B1	1.1b
(ii)	$[\sigma =] \sqrt{\frac{(\text{awrt})4336}{184} - \bar{x}^2}$ <u>or</u> allow $[\sigma^2 =] \frac{(\text{awrt})4336}{184} - \bar{x}^2$ <u>or</u> awrt 19.1 = 4.367... awrt <u>4.37</u>	M1 A1 (3)	1.1b 1.1b
(c)(i)	Only covers May~Oct (so not a suitable sample)	B1	1.1b
(ii)	e.g. Winter months are <u>missing</u> when we'd expect <u>more rain</u> so expect estimate in (b)(i) to be an <u>underestimate</u> (oe)	B1 (2)	2.4
		(7 marks)	
Notes			
(a)	M1 for recognising that tr must be replaced (oe) with a numerical value The following examples would score M0: The tr values are worth 0 so ignore (not replacing) <u>or</u> must remove outliers <u>or</u> fill gaps in table <u>or</u> make widths the same <u>or</u> need to find mid-points A1 for using a suitable value: e.g. 0.025 (or allow 0) i.e. any value in [0, 0.05] (these give $\sum x = 390$ (3sf), use of 0.05 gives 390.8, use of 0 gives 389.3 allow in (b)(i))		
(b)(i)	B1 for awrt 2.12 <u>or</u> allow simplified fraction or mixed number. B0 for $\frac{390}{184}$		
(ii)	M1 for a correct expression for standard deviation or variance. Allow $\sum x^2 =$ awrt 4336 Ignore their label σ or σ^2 Can ft their mean A1 for awrt 4.37 [Use of s gives 4.3791... so for correct use seen allow awrt 4.38]		
SC	Using n = 155 Allow M1 for expression $[\sigma =] \sqrt{\frac{(\text{awrt})4336}{155} - \bar{x}^2} = \sqrt{21.64\dots}$ or 4.65...		
Part (c) can effectively be marked together.			
(c)(i)	B1 for a comment mentioning that data is just from May~Oct (so not representative of the whole year). Just saying “only 184 days so not representative” is B0, must mention May ~ Oct		
(ii)	B1 for comment that <u>missing/winter</u> months expected to have more rain (oe) and “underestimate”(oe) We are looking for all 3 of these ideas here: 1. A statement or implication that missing data is from winter or different months. 2. A suggestion about the rainfall in these months (probably more rain). 3. A statement about the impact on the estimate in (b)(i) <u>equivalent</u> to saying it would be an underestimate or the (actual) mean will be higher.		
SC	If you see “Leeming or N or NE has less rain in winter months” – please send to review		



Qu 3	Scheme	Marks	AO
(a)	Hectopascal <u>or</u> hPa	B1	1.2
(b)	$\bar{x} = \bar{y} + 1010$ <u>or</u> $\frac{214}{30} + 1010$ $= 1017.1333... \text{ awrt } \underline{1017}$	(1) M1 A1	1.1b 1.1b
(c)	$\sigma_x = \sigma_y$ (or statement that standard deviation is not affected by this type of coding) $[\sigma_y =] \sqrt{\frac{5912}{30} - ("7.13[33...]")^2}$ <u>or</u> $\sqrt{146.1822...}$ $= 12.0905... \text{ awrt } \underline{12.1}$	(2) M1 M1 A1	3.1b 1.1b 1.1b
(d)	High pressure (since approx. mean + sd) so clockwise Locations are (from North to South): Leuchars, Heathrow, Hurn Wind direction is direction wind blows <u>from</u> So: Heathrow (NE) Hurn (E) Leuchars (W)	B1 B1 (2)	2.4 2.2a
(8 marks)			
Notes			
FYI	1 hPa = 100 Pa; 10hPa = 1 kPa; 1Pa = 1 Nm ⁻²		
(a)	B1 for “hectopascal” <u>or</u> hPa (condone pascals, allow millibars <u>or</u> mb) o.e. Do NOT allow kPa <u>or</u> kilopascals <u>or</u> Pa on its own		
(b)	M1 for a strategy to find $\bar{x}$ Allow an attempt to find $\sum x$ that gets as far as $\sum x = \sum y + 30 \times 1010 [= 30\,514]$ A1 for awrt 1017 (accept 1020) [Ignore incorrect units]		
(c)	1 st M1 for an overall strategy using the fact $\sigma_x = \sigma_y$ (can be implied by correct <u>final</u> ans) <u>or</u> for $\sum x = 30\,514$ and $\sum x^2 = 31\,041\,192$ (both seen and correct) 2 nd M1 for a correct expression (with $\sqrt{}$) (ft their $\bar{y}$ to 3sf) allow awrt 146 for 146.1822.. <u>or</u> for correct expression in x can ft their $\sum x > 30\,000$ or their answer to (b) A1 (dep on 2 nd M1) for awrt 12.1 [Ignore incorrect units] Final ans of awrt 12.1 scores 3/3 but if they then adjust for x e.g. add 1010 (M0M1A1)		
Final answer	(d) 1 st B1 for at least one of these reasons (these 2 lines) clearly stated (may see diagram) Need “high pressure” and “clockwise” to score on 1 st line Contradictory statements B0 e.g. correct N~S list but say “anticlockwise” 2 nd B1 (indep of 1 st B1) for deducing the 3 correct directions either in the table or stated as above If the answers in table and text are different we take the table (as question says)		



Question	Scheme	Marks	AOs
4 (a)	$\frac{132}{184} = 0.71739...$ awrt 0.717	B1	1.1b
		(1)	
(b)(i)	$P(X \geq 6) = 1 - P(X \leq 5)$ or $P([X =]6) + P([X =]7) + P([X =]8)$	M1	3.4
	$= 1 - 0.296722...$ awrt 0.703	A1	1.1b
		(2)	
(b)(ii)	$184 \times P(X = 7)$ [= $184 \times 0.2811...$]	M1	1.1b
	$= 51.7385...$ awrt 51.7	A1	1.1b
		(2)	
(c)	Part (a) and part (b)(i) are similar and the expected number of 7s (51.7 or 0.281) matches with the number of 7s found in the data set (52 or 0.283) so Magali's model is supported.	B1ft	3.5a
		(1)	
(d)	$\frac{23}{28} = 0.82142...$ awrt 0.821	B1	1.1b
		(1)	
(e)	Any one of... - Part (d)/'0.821' differs from part (a)/(b)(i)/(0.7...) - there is a greater/different probability of high cloud cover/more likely to have high cloud cover if the previous day had high cloud cover independence(o.e.) does not hold	B1	2.4
	...therefore Magali's (binomial) model may not be suitable.	dB1	3.5a
		(2)	
(9 marks)			
Notes			
	Allow fractions, decimals or percentages throughout this question.		
(a)	Allow equivalent fraction, e.g. $\frac{33}{46}$		
(b)(i)	M1: for writing or using $1 - P(X \leq 5)$ or $P(X = 6) + P(X = 7) + P(X = 8)$ A1: awrt 0.703 (correct answer scores 2 out of 2)		
(b)(ii)	M1: for $184 \times P(X = 7)$ o.e. e.g., $184 \times [P(X \leq 7) - P(X \leq 6)]$ A1: awrt 51.7		
(c)	B1ft: comparing '0.717' with '0.703' <u>and</u> '51.7' or '0.281' with 52 or 0.283 and concluding that Magali's model is supported (must be comparing prob. with prob. <u>and</u> days with days). Allow not supported or mixed conclusions if consistent with their f.t. answers in (a) and (b)		
(e)	B1: Any bullet point dB1: (dep on previous B1) for Magali's model may not be suitable (o.e.) Condone not accurate for not suitable SC: part (d) is similar to part (a)/(b)(i) and a compatible conclusion (i.e. Magali's model is supported) to score B1B1.		

Qu 5	Scheme	Marks	AO
(a)	$[P(H > 1.6) =] 0.091211... = \text{awrt } \underline{0.0912}$	B1 (1)	1.1b
(b)	Need H and T to be independent or events $\{H > 1.6\}$ and $\{T < 300\}$ are independent	B1 (1)	2.4
(c)	$[P(T < 300) =] 0.124(2816...)$ Prob both is: $"0.0912..." \times "0.124..."$ $= 0.011335... = \text{awrt } \underline{0.0113}$	M1 M1 A1 (3)	3.4 1.1b 1.1b
(d)	$\frac{16.3 - \mu}{\sigma} = -0.5244(0051...) , \frac{29 - \mu}{\sigma} = 1.2816$ (calc: 1.28155156...) e.g. $29 - 16.3 = \sigma("1.2816" - "-0.5244")$ $\sigma = 7.032115... = \text{awrt } \underline{7.03}$ $\mu = 19.9876... = \underline{19.95} \leq \mu \leq \underline{20.0}$	M1M1 M1 A1 A1 (5)	3.1a 1.1b 1.1b 1.1b 3.2a
(10 marks)			
Notes			
(a)	B1 for awrt 0.0912 (from calculator)		
(b)	B1 for a suitable reason mentioning or implying H and T are independent Allow: e.g. “they”/ “each event”/ “ $P(H)$ and $P(T)$ ”/ “the variables” and “independent” B0 for “the results” / “the values” are independent. Ignore other comments that are not incorrect or contradictory.		
(c)	1 st M1 for using model for T to attempt to find $P(T < 300)$ e.g. sight of 0.124 or better or sight of $\pm \left(\frac{300 - 330}{26} \right)$ or $\pm \left(\frac{5 - 5.5}{0.433...} \right)$ or $Z = \pm 1.15(3...)$ 2 nd M1 for multiplying their two probabilities together ft part (a) and their $P(T < 300)$ provided both values are probabilities. NB M0M1 is possible here A1 for awrt 0.0113 [Correct answer with no incorrect working 3/3]		
(d)	1 st M1 for standardising 16.3 and setting equal to z value where $0.5 < z < 0.6$ 2 nd M1 for standardising 29 and setting equal to z value where $1 < z < 1.5$ 3 rd M1 dep on 1 st or 2 nd M1 for solving their two linear eq'ns – reach an eq'n in one variable May be implied by sight of $\sigma = 7$ (or better) or $\mu = 20$ (or better) For 1st A mark we must also see one of – 0.5244 or 1.2816 (or better) used in their equ'ns OR both z values correct to 3dp i.e. – 0.524 and 1.282 1 st A1 for $\sigma = \text{awrt } 7.03$ (but see 3 rd case below) 2 nd A1 for $\mu = \text{in } [19.95, 20.0]$ (i.e shouldn't see something rounding down to 20.0) Allow 20 from equations with suitable z values (see examples below)		
NB	Use of – 0.524 and 1.28 [would give 7.0399... and 19.988...] and scores M3A0A1 Use of – 0.524 and 1.2816 [would give 7.033... and 19.99...] and scores M3A1A1 Use of – 0.5244 and 1.28 [would give 7.038... and 19.99 ...] and scores M3A1A1 Both z values correct to 3dp i.e. – 0.524 and 1.282 [should give 7.032 and 19.984] scores A1A1		

Qu 5	Scheme	Marks	AO
(a)	$\left[\text{Let } F \sim N(166.5, 6.1^2) \right] P(F < k) = 0.01 \Rightarrow \frac{k - 166.5}{6.1} = -2.3263$ $k = 152.309... \quad \underline{152} \text{ or awrt } \underline{152.3}$	M1 A1 (2)	3.4 1.1b
(b)	$[P(150 < F < 175) =] \quad 0.914840... \quad \text{awrt } \underline{0.915}$	B1 (1)	1.1b
(c)	$P(F > 160 \mid 150 < F < 175)$ $= \frac{P(160 < F < 175)}{P(150 < F < 175)} \quad \text{or} \quad \frac{P(160 < F < 175)}{"(b)"} \\ = \frac{0.7749487...}{"0.914840..."} \\ = 0.84708... \text{ awrt } \underline{0.847}$	M1 M1 A1ft A1 (4)	3.1b 1.1b 1.1b 1.1b
(d)	$H_0: \mu = 166.5 \quad H_1: \mu < 166.5$ $[\text{Let } \bar{X} = \text{height of female from 2}^{\text{nd}} \text{ country}] \quad \bar{X} \sim N\left(166.5, \left(\frac{7.4}{\sqrt{50}}\right)^2\right)$ $P(\bar{X} < 164.6) = 0.03472...$ $[0.0347... < 0.05 \text{ so significant or reject } H_0]$ There is evidence to support Mia's belief	B1 M1 A1 dA1 (4)	2.5 3.3 3.4 2.2b
Notes			
(a)	M1 for standardising (allow $\pm$) with k , 166.5 and 6.1 and set equal to a z value $2.3 < z < 2.4$ A1 for 152 or awrt 152.3 Ans only 2/2 [Condone poor use of notation e.g. $P(\frac{k-166.5}{6.1}) = -2.3263$] Allow percentages instead of probabilities throughout.		
(b)	B1 for awrt 0.915		
(c)	1 st M1 for interpreting demand as an appropriate conditional probability ($\Rightarrow$ by 2 nd M1) 2 nd M1 for correct ratio of expressions (can fit their (b) on denominator) ($\Rightarrow$ by 1 st A1ft) 1 st A1ft for a correct ratio of probs (can fit their "0.9148..." to 3sf from (b) if > 0.775) 2 nd A1 for awrt 0.847		
(d)	B1 for both correct hypotheses in terms of μ 1 st M1 for selecting the correct model (needn't use $\bar{X} \Rightarrow$ by standardisation or 1 st A1) 1 st A1 for correct use of the correct model i.e. awrt 0.035 (allow 0.04 if $P(\bar{X} < 164.6)$ seen) Condone $P(\bar{X} > 164.6) = 0.9652$ or awrt 0.97 <u>only if</u> comparison with 0.95 is made		
ALT	Use of z value: Need to see $Z = -1.8(15...)$ and cv of ± 1.6449 (allow 1.64 or better) for 1 st A1		
ALT	Use of CR or CV for $\bar{X}$: Need to see " $\bar{X}$ " $> 164.7786...$ or CV = ... (awrt 164.8) for 1 st A1 Condone truncation i.e. 164.7 or better 2 nd dA1 (dep on M1A1 only) for a correct inference in context. Must mention <u>Mia's belief</u> or <u>mean height of females/women</u> Do NOT award if contradictory statements about hypotheses made e.g. "not sig"		
SC	M0 for $\bar{X} \sim N(164.6, ...)$ If they achieve $p =$ awrt 0.035 (o.e. with z -value or CV of 166.3) and a correct conclusion in context is given score M0A0A1 [and SC for awrt 0.97 > 0.95 case]		

Qu	Scheme		Marks	AOs
2(a)	$\left[P(L < 7.902) = 0.025 \Rightarrow \right] \frac{7.902 - 8}{x} = -1.96$ oe		M1	3.4
	$[x =] 0.05^*$		A1cso*	1.1b
	SC B1(mark as M0A1) for $\frac{7.902 - 8}{0.05} = -1.96 \Rightarrow 0.024998$			
			(2)	
(b)	$P(7.94 \leq L \leq 8.09) = 0.8490\dots$ awrt 0.849		B1	1.1b
			(1)	
(c)	$[P(L < 7.94) =] 0.115069\dots$ (awrt 0.115) or $[P(L > 8.09) =] 0.03593\dots$ (awrt 0.036)		B1	1.1b
	$[P(L < 7.94) =] 0.115069\dots$ (awrt 0.115) & $[P(L > 8.09) =] 0.03593\dots$ (awrt 0.036)		B1	1.1b
	Expected income per 500 rods = $\sum(\text{Income} \times \text{probability} \times 500)$ $(500 \times "0.849" \times 0.5) + (500 \times "0.1150\dots" \times 0.05) + (500 \times "0.03593\dots" \times 0.4)$ or Expected profit per rod = $\sum(\text{Profit} \times \text{probability})$ $0.30 \times "0.849" + -0.15 \times "0.1150\dots" + 0.20 \times "0.03593\dots" [= 0.2446\dots]$		M1	3.4
	Expected profit per 500 rods $500 \times \sum(\text{Profit} \times \text{probability})$ or $\sum(\text{Income} \times \text{probability} \times 500) - 500 \times 0.2$ $= 500 \times "0.2446\dots"$ or $= "222.3" - 500 \times 0.2$		M1d	3.1b
	$= [£]122.3\dots$ awrt [£]122		A1	1.1b
			(5)	
(d)	Let $X \sim B(200, 0.015)$		M1	3.3
	$P(X \leq 5) =$	$P(X \geq 6) =$	M1	1.1b
	0.9176...	0.0824	A1	1.1b
	Manufacturer is unlikely to achieve their aim since <u>0.9176 < 0.95</u>	Manufacturer is unlikely to achieve their aim since <u>0.0824 > 0.05</u>	A1ft	2.4
			(4)	
Notes: (12 marks)				
(a)	M1	Using the normal distribution to set up equation. Allow σ for x and awrt ± 1.96		
	A1*	cso For a correct expression for x followed by 0.05 or 0.05000... No incorrect working seen		
(b)	B1	awrt 0.849		
(c)	B1	awrt 0.115 (Implied by awrt 57.5 for number of rods) or awrt 0.036 (Implied by awrt 18 for number of rods)		
	B1	awrt 0.115 (Implied by awrt 57.5 for number of rods) and awrt 0.036 (Implied by awrt 18 for number of rods)		
	M1	Correct method to find the total income of 500 rods. Attempt at all 3 with at least two correct and no extras or Correct method to find sum of all three profits with at least two of 30, -15 or 20 correct. May work in pence but need to be consistent. Allow awrt 24.5 or 0.245		
	M1d	Dep on previous method for finding profit for 500 rods. May work in pence but need to be consistent. Allow "0.2446..." $\times 500$ or "their income" for 500 rods $- 500 \times 0.2$ (accept 499 or 501)		
	A1	All previous marks must be awarded for awrt 122 awrt 12200p NB if uses any integer values for numbers of rods then it is A0 other than for 18 for $L > 8.09$		
(d)	M1	Selecting the appropriate model. May be seen or used. Allow B(200,0.985) or Po(3) Condone B(0.015, 200) or B(0.985, 200).		
	M1	Writing or using $P(X \leq 5)$ Do not accept $P(X < 6)$ unless found $P(X \leq 5)$	Writing or using $P(X \geq 6)$ Do not accept $P(X > 5)$ unless found $P(X \geq 6)$	
	A1	0.92 (Poisson 0.916...)	0.08 or better	
	A1ft	Need at least one of the method marks to be awarded. Correct conclusion with the comparison (may be in words). Ft "their $p = 0.9176\dots$ " as long as $p > 0.9$ If "their 0.9176..." < 0.95 must ... be unlikely... If "their 0.9176..." > 0.95 they must say ... be likely... To fit the alternative then $p < 0.1$		

Qu 4	Scheme	Marks	AO
(a)	Convenience <u>or</u> opportunity [sampling]	B1	1.2
(b)	Quota [sampling] e.g. Take 4 people every 10 minutes	B1 (1)	1.1a 1.1b
(c)	Census	B1 (2)	1.2
(d)	[58 – 26 =] <u>32</u> (min)	B1 (1)	1.1b
(e)	$\mu = \frac{4133}{95} = 43.505263\dots$ $\sigma_x = \sqrt{\frac{202\,294}{95} - \mu^2} = \sqrt{236.7026\dots}$ $= 15.385\dots$	awrt <u>43.5</u> (min) M1 A1 (3)	1.1b 1.1b 1.1b
(f)	There are outliers in the data (or data is skew) which will affect mean and sd Therefore use median and IQR	B1 dB1 (2)	2.4 2.4
(g)	Value of 20, LQ at 26 and outliers will not change <u>or</u> state that median and upper quartile are the values that <u>do</u> change <u>More values now below 40 than above</u> so Q_2 <u>or</u> Q_3 will change and be lower Both Q_2 <u>and</u> Q_3 will be lower	B1 M1 A1 (3)	1.1b 2.1 2.4
(13 marks)			
Notes			
(b)	1 st B1 for quota (sampling) mentioned (“Stratified” or “systematic” or “random” are B0B0) 2 nd B1 for a description of how such a system might work, requires suitable strata or categories e.g. time slots, departments, gender, age groups, distance travelled etc Suggestion of randomness is B0		
(e)	B1 for a correct mean (awrt 43.5) M1 for a correct expression for the sd (including $\sqrt{}$)ft their mean A1 for awrt 15.4 (Allow $s = 15.4667\dots$ awrt 15.5)		
(f)	1 st B1 for acknowledging <u>outliers</u> or <u>skewness</u> are a problem for <u>mean and sd</u> “extreme values”/“anomalies” OK May be implied by saying median and IQR not affected by.. We need to see mention of “outliers”, “skewness” and the problem so “data is skewed so use median and IQR” is B0 unless mention that they are not affected by extreme values <u>or</u> mean and standard deviation can be “inflated” by the positive skew etc 2 nd dB1 dep on 1 st B1 for therefore choosing <u>median and IQR</u>		
(g)	B1 for identifying 2 of these 3 groups of unchanged values or stating only Q_2 and Q_3 change M1 for <u>explaining</u> that median or UQ should be lower. E.g. the 2 values have moved to below 40 (or 58) and therefore more than 50% below 40 or (more than 75% below 58) <u>or</u> an argument to show that the other 3 values are the same. (o.e.) Allow arrows on box plot provided statement in words about increased % below 40 or 58 etc A1 for stating median <u>and</u> UQ are both lower with clear evidence of M1 scored [If lots of values on 40 then median might not change but, since two values <u>do</u> change then UQ would change. If this meant that 92 became an outlier then we would have a new value for upper whisker and an extra outlier so effectively 3 values are altered. So median changes]		

Question	Scheme	Marks	AOs
5(a)	$\frac{24.63 - 25}{\sigma} = -1.0364$	M1	3.1b
	$[\sigma =]0.357$ (must come from compatible signs)	A1	1.1b
	$P(D > k) = 0.4$ or $P(D < k) = 0.6$	B1	1.1b
	$\frac{k - 25}{0.357} = 0.2533$	M1	3.4
	$k = \text{awrt } \underline{25.09}$	A1	1.1b
		(5)	
(b)	$[Y \sim B(200, 0.45) \rightarrow] W \sim N(90, 49.5)$	B1	3.3
	$P(Y < 100) \approx P(W < 99.5) \left[= P\left(Z < \frac{99.5 - 90}{\sqrt{49.5}}\right) \right]$	M1	3.4
	$= 0.9115...$ awrt <u>0.912</u>	A1	1.1b
		(3)	
(c)	$H_0 : \mu = 25$ $H_1 : \mu < 25$	B1	2.5
	$[\bar{D} \sim N\left(25, \frac{0.16^2}{20}\right)]$	M1	3.3
	$P(\bar{D} < 24.94) [= P(Z < -1.677...)] = 0.046766...$	A1	3.4
	$p = 0.047 < 0.05$ or $z = -1.677... < -1.6449$ or $24.94 < 24.94115...$ or reject H_0 / in the critical region/significant	M1	1.1b
	There is sufficient evidence to support <u>Hannah's belief</u> .	A1	2.2b
		(5)	
(13 marks)			
Notes			
(a)	M1: for standardising 24.63, 25 and ' σ ' (ignore label) and setting = to z where $1 < z < 2$ A1: $[\sigma =]$ awrt 0.36. Do not award this mark if signs are not compatible. B1: for either correct probability statement (may be implied by correct answer) this mark may be scored for a correct region shown on a diagram M1: for a correct expression with $z =$ awrt 0.253 (may be implied by correct answer) A1: awrt 25.09 (Correct answer with no incorrect working scores 5 out of 5)		
(b)	B1: setting up normal distribution approximation of binomial $N(90, 49.5)$ (may be implied by a correct answer) Look out for e.g. $\sigma = \frac{3\sqrt{22}}{2}$ or $\sigma =$ awrt 7.04 M1: attempting a probability using a continuity correction i.e. $P(W < 100.5)$, $P(W < 99.5)$ or $P(W < 98.5)$ condone $\leq$ (The continuity correction may be seen in a standardisation). A1: awrt 0.912 [Note: 0.911299... from binomial scores 0 out of 3]		
(c)	B1: for both hypotheses in terms of μ M1: selecting suitable model must see $N(\text{ormal})$, mean 25, $sd = \frac{0.16}{\sqrt{20}}$ (o.e.) or $var = \frac{4}{3125}$ (o.e.) Condone $N(25, \frac{0.16}{\sqrt{20}})$ if $\frac{0.16}{\sqrt{20}}$ then used as s.d. A1: p value = awrt 0.047 or test statistic awrt -1.68 or CV awrt 24.941 (any of these values imply the M1 provided they do not come from Normal mean = 24.94) M1: a correct comparison (including compatible signs) or correct non-contextual conclusion (f.t. their p value, test statistic or critical value in the comparison) M1 may be implied by a correct contextual statement NB Any contradictory non contextual statements/comparisons score M0A0 e.g. ' $p < 0.05$, not significant' A1: correct conclusion in context mentioning <u>Hannah's belief</u> or the mean <u>amount/liquid</u> in each bottle is now <u>less than 25ml</u> (dep on M1A1M1)		

Qu 5	Scheme	Marks	AO
(a)	$P(L > 16) = 0.69146...$ awrt 0.691	B1 (1)	1.1b
(b)	$P(L > 20 L > 16) = \frac{P(L > 20)}{P(L > 16)}$ $= \frac{0.308537...}{(a)} \text{ or } \frac{1-(a)}{(a)}, = 0.44621...$ For calc to work require $(0.44621...) ^4 = 0.03964...$ awrt 0.0396	M1 A1ft, A1 dM1 A1 (5)	3.1b 1.1b 1.1b 2.1 1.1b
(c)	Require: $[P(L > 4)]^2 \times [P(L > 20 L > 16)]^2$ $= (0.99976...) ^2 \times ("0.44621...")^2$ $= 0.19901...$ awrt 0.199 (*)	M1 A1ft A1cso* (3)	1.1a 1.1b 1.1b
(d)	$H_0 : \mu = 18 \quad H_1 : \mu > 18$ $\bar{L} \sim N\left(18, \left(\frac{4}{\sqrt{20}}\right)^2\right)$ $P(\bar{L} > 19.2) = P(Z > 1.3416...) = 0.089856...$ (0.0899 > 5%) or (19.2 < 19.5) or $1.34 < 1.6449$ so not significant Insufficient evidence to support Alice's claim (or belief)	B1 M1 A1 A1 A1 (5)	2.5 3.3 3.4 1.1b 3.5a
(14 marks)			
Notes			
(a)	B1 for evaluating probability using their calculator (awrt 0.691) Accept 0.6915		
(b)	1st M1 for a first step of identifying a suitable conditional probability (either form) 1st A1ft for a ratio of probabilities with numerator = awrt 0.309 or $1 - (a)$ and denom = their (a) 2nd A1 for awrt 0.446 (o.e.) Accept 0.4465 (from $\frac{0.3085}{0.691} = 0.44645...$) NB $\frac{P(16 < L < 20)}{P(L > 16)} = 0.5538...$ scores M1A1A1 when they do $1 - 0.5538 = 0.4462...$ 2nd M1 (dep on 1st M1) for 2nd correct step i.e. (their 0.446...) ⁴ or $X \sim B(4, "0.446")$ and $P(X = 4)$ 3rd A1 for awrt 0.0396		
(c)	1st M1 for a correct approach to solving the problem (May be implied by A1ft) 1st A1ft for $P(L > 4) =$ awrt 0.9998 used and ft their 0.44621 in correct expression If use $P(L > 20) = 0.3085..$ as 0.446.. in (b) then M1 for $(0.3085..) ^2 \times [P(L > 4)]^2$; A1ft as above * 2nd A1cso for 0.199 or better with clear evidence of M1 [NB $(0.4662..) ^2 = 0.199...$ is M0A0A0] Must see M1 scored by correct expression in symbols or values (M1A1ft)		
(d)	B1 for both hypotheses in terms of μ. M1 for selecting a suitable model. Sight of <u>normal</u>, <u>mean</u> 18, <u>sd</u> $\frac{4}{\sqrt{20}}$ (o.e.) or <u>variance</u> = 0.8 1st A1 for using the model correctly. Allow awrt 0.0899 or 0.09 from correct prob. statement		
ALT	CR $(\bar{L}) > 19.471...$ (accept awrt 19.5) or CV of 1.6449 (or better: calc 1.6448536..) 2nd A1 for correct non-contextual conclusion. Wrong comparison or contradictions A0 Error giving 2nd A0 implies 3rd A0 but just a correct contextual conclusion can score A1A1 3rd A1 dep on M1 and 1st A1 for a correct contextual conclusion mentioning <u>Alice's claim</u> /<u>belief</u> or there is insufficient evidence that the mean <u>lifetime</u> is more than 18 hours		



Qu 5	Scheme	Marks	AO
(a)	{ Let X = time spent, $P(X > 15) = \}$ 0.105649... awrt 0.106	B1 (1)	1.1b
(b)	$H_0: \mu = 10$ $H_1: \mu > 10$ $\bar{X} \sim N\left(10, \left(\frac{4}{\sqrt{20}}\right)^2\right)$; $P(\bar{X} > 11.5) = 0.046766...$ [Condone 0.9532...] [This is significant (< 5%) so] there is evidence to support the complaint	B1 M1;A1 A1 (4)	2.5 3.3;3.4 2.2b
(c)(i)	$[P(T < 2) =]$ 0.1956... awrt 0.196	B1 (1)	1.1b
(ii)	Require $\frac{P(0 < T < 2)}{P(T > 0)} = \frac{0.119119...}{0.923436...}$; = 0.1289955... awrt 0.129	M1 A1;A1 (3)	3.4 1.1bx2
(iii)	The current model suggests non-negligible probability of T values < 0 which is impossible	B1 (1)	3.5b
(d)	Require t such that $P(T > t T > 2) = 0.5$ or $P(T < t T > 2) = 0.5$ e.g. $\frac{P(T > t)}{P(T > 2)} = 0.5$; so $P(T > t) = 0.5 \times [1 - (c)(i)]$ or $P(T > t) = 0.5 \times 0.8043..$ [i.e. $P(T > t) = 0.40...$ implies] $\frac{t-5}{3.5} = 0.2533$ or $P(T < t) = "0.5978.."$ $t = 5.886...$ or from calculator 5.867... so awrt 5.9	M1 M1; A1ft M1 A1 (5)	3.1b 1.1b 3.4 1.1b 1.1b
(15 marks)			
Notes			
(a)	B1 for awrt 0.106 (from calculator) [Allow 10.6%]		
(b)	B1 for both hypotheses correct in terms of μ . M1 for selection of a correct model (sight or use of correct normal- may not have label $\bar{X}$) 1 st A1 for use of this model to get probability allow 0.046~0.047 [Condone awrt 0.953]		
ALT	OR test statistic $z = 1.677...$ (awrt 1.68) and cv of 1.64 (or better) or CR $\bar{X} > 11.47..$		
	2 nd A1 (dep on 1 st A1 or at least $P(\bar{X} > 11.5) < 0.05$ (o.e.))		
	for a correct conclusion in context -must mention complaint /claim or time /mins is > 10		
SC	(M0 for $\bar{X} \sim N(11.5, ...)$ for correct probability and conclusion (score M0A0A1 on open)		
(c)(i)	B1 for awrt 0.196 (from calculator) [Allow 19.6%]		
(ii)	M1 for a correct probability ratio expression (may be implied by 1 st A1 scored) 1 st A1 for a correct ratio of probabilities (both correct or truncated to 2 dp) 2 nd A1 for awrt 0.129		
(iii)	B1 for a suitable explanation of why model is not suitable based on negative T values Must say that a significant proportion of values < 0 (o.e.) e.g. $P(T > 0)$ should be closer to 1 or Difference between $P(T < 2 T > 0)$ and $P(T < 2)$ is too big (o.e.)		
(d)	1 st M1 for a correct conditional probability statement to start the problem or $0.5 \times P(T > 2)$ 2 nd M1 for correct ratio of probability expressions [Must have $P(T > t)$ or $P(2 < T < t)$] 1 st A1ft for a correct equation for $P(T > t)$ (o.e.) ft their answer to part (c)[May be in a diagram] 3 rd M1 for attempt to find t (standardising and sight of 0.2533) or prepare to use calc (ft) Arriving at $P(T < \text{median}) = 1 - 0.5 \times \text{their } 0.8043$ will score 1 st 4 marks 2 nd A1 for awrt 5.9 Sight of awrt 5.9 and at least one M mark scores 5/5 [Answer only send to review]		



Question	Scheme		Marks	AOs
2(a)	$\left[\bar{x} = \frac{6612}{120} \right] = \underline{55.1}$ (seconds)		B1	1.1b
			(1)	
2(b)	$[\sigma_x =] \sqrt{\frac{364\,902}{120} - ("55.1")^2}$ or $\sqrt{4.84}$		M1	1.1b
	= <u>2.2</u> (seconds)		A1	1.1b
			(2)	
2(c)	Coach <i>B</i> because standard deviation is greater		M1	2.4
	This means <i>B</i> has a greater range and therefore more extreme values		A1ft	2.2b
			(2)	
(5 marks)				
Notes: Ignore units in this question				
(a)	B1	for 55.1 or $55\frac{1}{10}$ only (may be seen in the question) ISW e.g. 55.1 followed by 55 etc Do not accept $\frac{551}{10}$ that scores B0		
(b)	M1	for a correct expression (values seen not just a formula)(including $\sqrt{}$)		
	A1	for 2.2 or $\frac{11}{5}$ o.e. (allow $s = 2.209\dots$ so awrt 2.21) Correct answer is 2 marks		
(c)	Ignore comments about different number of runners. Don't need to state means are the same. If there is no answer given for (b) then no marks can be scored in (c)			
“2.2”<3.6 SC “2.2”>3.6	B1	for coach <i>B</i> and states s.d. of <i>B</i> is greater o.e. Must compare. Allow implied choose <i>B</i> e.g. <i>B</i> has a greater range or <i>B</i> s values are more spread out or $3.6 > \text{“2.2”}$ or $\sigma_B > \text{“2.2”}$ May say s.d. of <i>A</i> is smaller which is equivalent to saying s.d. of <i>B</i> is larger		
	dB1	(dep on 1 st B1 scored) for clear reason why greater s.d. means choose coach <i>B</i> Idea of greater range and more extreme values (faster times) in <i>B</i> e.g. <i>B</i> has more (or greater chance of or greater % of) runners at lower times (may compare mean – 1 or 2 or 3 standard deviations to show that <i>B</i> has the quicker time) e.g. coach <i>A</i> (52.9) > (51.5) coach <i>B</i> so coach <i>B</i> likely to be quickest		
	NB	“coach <i>B</i> since <i>B</i> has greater range /s.d. and so more extreme values” scores B1B1		
	(A0 in (b) and answer > 3.6)			
	B1	equivalent reasons as above but choose <i>A</i>		
dB1	same rules as above but applied to <i>A</i>			

Qu 3	Scheme	Marks	AO
(a)	Rain[fall] (allow [Mean] Windspeed)	B1	1.2
(b)	$[\bar{x} =] 15.2239... = \text{awrt } \underline{15.2}$	B1	1.1b
	$\sigma_x = \sqrt{\frac{44\,695.4}{184} - "15.22.."^2} \text{ or } \sqrt{11.1(422...)}$	M1	1.1b
	$= 3.33800... \text{ awrt } \underline{3.34}$	A1	1.1b
(c)	<u>Mean is higher</u> than average OR a <u>summer/spring</u> month	M1	2.4
	If they say winter/autumn they must explain that these are hotter months for Perth.		
	[Perth is southern hemisphere or Australia so latest available] month is Oct	A1cso	2.2b
		(2)	
		(6 marks)	
	Notes		
	Answers may appear next to the question.		
(a)	B1 for Rain[fall] or precipitation (e.g. allow Daily Total [or Mean or max] Rainfall etc) (or allow Mean Windspeed or just “windspeed” BUT not max windspeed or “gust”) If they give more than one answer we take the last one . (NB Actual windspeed mean is 8.2, sd 2.38. No other quantitative variables available)		
(b)	B1 for awrt 15.2 (Do not accept fractions or mixed numbers) M1 for a correct expression including square root (ft their mean) May be implied by an answer of 3.3 or better. A1 for awrt 3.34 [Allow $s = 3.3471...$ i.e. awrt 3.35 if correct formula/expression is seen]		
(c)	If answer in (b)(i) > 19.4 and an attempt is made in (c) please send to review. M1 for a reason mentioning that mean or temperature is higher (o.e.) e.g. it is a <u>warmer/hotter</u> month is OK <u>or</u> sight of $19.4 > (\text{their}) 15.2$ Only ft their 15.2 if it is less than 19.4 OR suggesting a summer/spring month. Ignore incorrect statements that are irrelevant or don't contradict For incorrect statements that contradict score M0 A1cso dep on M1 scored for inferring October Must choose <u>just</u> October not a range like August~October (NB actual mean for Sep is 15.6 and sd 3.19 and this scores A0) Can accept for example “high mean so December” for M1A0		
SC	M1A1 for “October since Perth is in the southern hemisphere/Australia” M1A0 for “Sep <u>or</u> Nov <u>or</u> Dec <u>or</u> Jan <u>or</u> Feb and “Perth is in the southern hemisphere/Australia” M0A0 just “Perth is in the southern hemisphere/Australia” <u>without</u> a month M0A0		



Question	Scheme		Marks	AOs
2(a)	IQR = 26.6 – 19.4 [= 7.2]		B1	2.1
	19.4 – 1.5 × ‘7.2’ [= 8.6] or 26.6 + 1.5 × ‘7.2’ [= 37.4]		M1	1.1b
	Plotting one upper whisker to 32.5 and one lower whisker to 8.6 or 9.1		A1	1.1b
	Plotting 7.6 and 8.1 as the only two outliers		A1	1.1b
			(4)	
(b)	<u>October</u> (since it is the month with the coldest temperatures between May and October in Beijing)		B1	2.4
			(1)	
(c)	$[\sigma = \sqrt{\frac{4952.906}{184}}]$ or e.g. $[\sigma = \sqrt{\frac{S_{xx}}{n}} = 5.188...]$ [=5.19*]		B1cso*	1.1b
			(1)	
(d)	$z = (\pm) 1.28(16)$	$[P_{90} =]29.251... \text{ or } [P_{10} =]15.948...$	B1	3.1b
	$2 \times 1.2816 \times 5.19$	‘29.251...’ – ‘15.948...’	M1	1.1b
	= awrt 13.3		A1	1.1b
			(3)	
(e)	Daily mean <u>wind speed</u> /Beaufort conversion since it is <u>qualitative</u> <u>Rainfall</u> since it is not symmetric/lots of days with 0 rainfall		B1 B1	2.4 2.4
			(2)	
(11 marks)				
Notes				
(a)	B1: for a correct calculation for the IQR (implied by 10.8 or 8.6 or 37.4 seen) M1: for a complete method for either lower outlier limit or upper outlier limit (allow ft on their IQR) (may be implied by the 1 st A1 or a lower whisker at 8.6) A1: both whiskers plotted correctly (allow ½ square tolerance) A1: only two outliers plotted, 7.6 and 8.1 (must be disconnected from whisker) NOTE: A fully correct box plot with no incorrect working scores 4/4			
	(c)	B1cso*: Correct expression with square root or correct formula and 5.188 or better Allow a complete correct method finding $\sum x^2 = \text{awrt } 98720$ and $\sigma = \sqrt{\frac{98715.9...}{184} - \left(\frac{4153.6}{184}\right)^2}$		
(d)	B1: Identifying z-value for 10th or 90th percentile (allow awrt (±) 1.28) or for identifying $[P_{90} =]29.251... \text{ (awrt 29.3) or } [P_{10} =]15.948... \text{ (awrt 15.9)}$ (This may be implied by a correct answer awrt 13.3) M1: for $2 \times z \times 5.19$ where $1 < z < 2$ or for their $P_{90} - P_{10}$ where $25 < P_{90} < 35$ and $10 < P_{10} < 20$ A1: awrt 13.3			
	(e)	B1: for one variable identified and a correct supporting reason B1: for two variables identified and a correct supporting reason for each Allow any two of the following: • <u>Wind speed/Beaufort</u> since the data is <u>non-numeric</u> (o.e.). They need not mention Beaufort provided there is a description of the data as non-numeric (Do not allow wind direction/wind gust) • <u>Rainfall</u> as not symmetric/is skewed/is not bell shaped/lots of 0s /many days with no rain/mean≠mode or median • <u>Date</u> since each data value appears once/it is uniformly distributed • Daily mean <u>pressure</u> since it is not symmetric/is skewed/not bell shaped • Daily mean <u>wind speed</u> since it is not symmetric/is skewed/not bell shaped Do not allow ‘not continuous’ or ‘discrete’ as a supporting reason. Ignore extraneous non-contradicting statements		