# AS Topics

Measures of Location and Spread

Representations of Data (Histograms, Box Plots)

Correlation

Probability

Binomial Distribution

Many statistics exam questions cover several topics. Be careful!

**Contains**:
AS SAMs
AS 2018
AS 2019
AS 2020
AS 2021
AS 2022
AS 2023

# A2 Topics

Regression and Correlation

Conditional Probability

The Normal Distribution

**Contains**:
A2 SAMs
A2 2018
A2 2019
A2 2020
A2 2021
A2 2022
A2 2023

# Measures of Location and Spread

# AS SAMs

## Location and Spread

**1.** Sara is investigating the variation in daily maximum gust, $t$ kn, for Camborne in June and July 1987.

She used the large data set to select a sample of size 20 from the June and July data for 1987. Sara selected the first value using a random number from 1 to 4 and then selected every third value after that.

(a) State the sampling technique Sara used.

**(1)**

(b) From your knowledge of the large data set, explain why this process may not generate a sample of size 20.

**(1)**

The data Sara collected are summarised as follows

$$n = 20 \qquad \sum t = 374 \qquad \sum t^2 = 7600$$

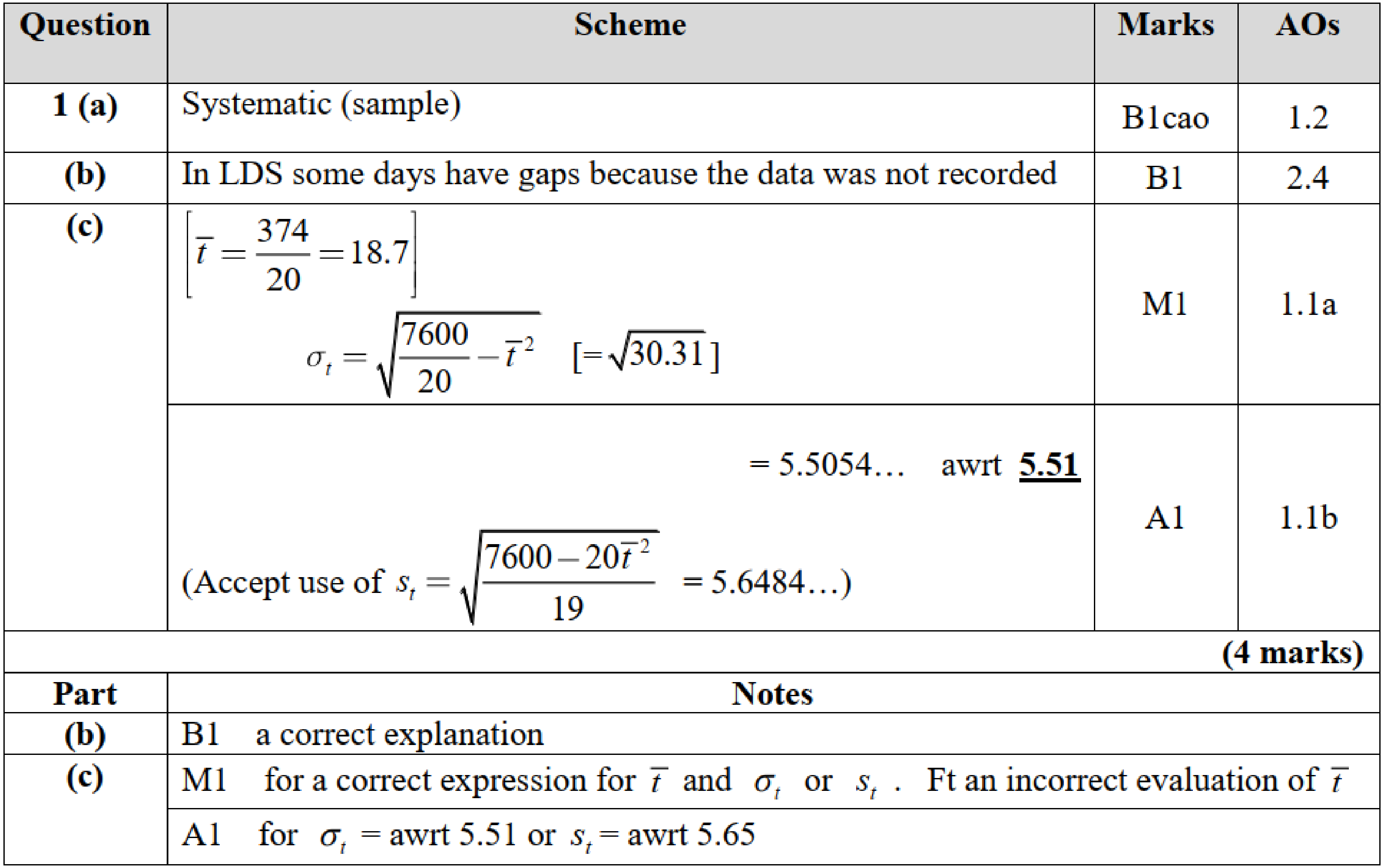
(c) Calculate the standard deviation.

**(2)**

HOME

**AS SAMs**

## Location and Spread



| Question | Scheme | Marks | AOs |
|---|---|---|---|
| **1 (a)** | Systematic (sample) | B1cao | 1.2 |
| **(b)** | In LDS some days have gaps because the data was not recorded | B1 | 2.4 |
| **(c)** | $\left[\bar{t}=\frac{374}{20}=18.7\right]$<br>$\sigma_t=\sqrt{\frac{7600}{20}-\bar{t}^2}\quad[=\sqrt{30.31}\,]$ | M1 | 1.1a |
| | $=5.5054\ldots$ awrt **5.51**<br>(Accept use of $s_t=\sqrt{\frac{7600-20\bar{t}^2}{19}}$ $=5.6484\ldots$) | A1 | 1.1b |
| | | | **(4 marks)** |

| Part | Notes |
|---|---|
| **(b)** | B1 a correct explanation |
| **(c)** | M1 for a correct expression for $\bar{t}$ and $\sigma_t$ or $s_t$ . Ft an incorrect evaluation of $\bar{t}$ |
| | A1 for $\sigma_t$ = awrt 5.51 or $s_t$ = awrt 5.65 |

HOME

**4.** Sara was studying the relationship between rainfall, $r$ mm, and humidity, $h$ %, in the UK. She takes a random sample of 11 days from May 1987 for Leuchars from the large data set.

She obtained the following results.

| $h$ | 93 | 86 | 95 | 97 | 86 | 94 | 97 | 97 | 87 | 97 | 86 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $r$ | 1.1 | 0.3 | 3.7 | 20.6 | 0 | 0 | 2.4 | 1.1 | 0.1 | 0.9 | 0.1 |

Sara examined the rainfall figures and found

$$Q_1 = 0.1 \qquad Q_2 = 0.9 \qquad Q_3 = 2.4$$

A value that is more than 1.5 times the interquartile range (IQR) above $Q_3$ is called an outlier.

(a) Show that $r = 20.6$ is an outlier.

(1)

(b) Give a reason why Sara might (i) include

(ii) exclude

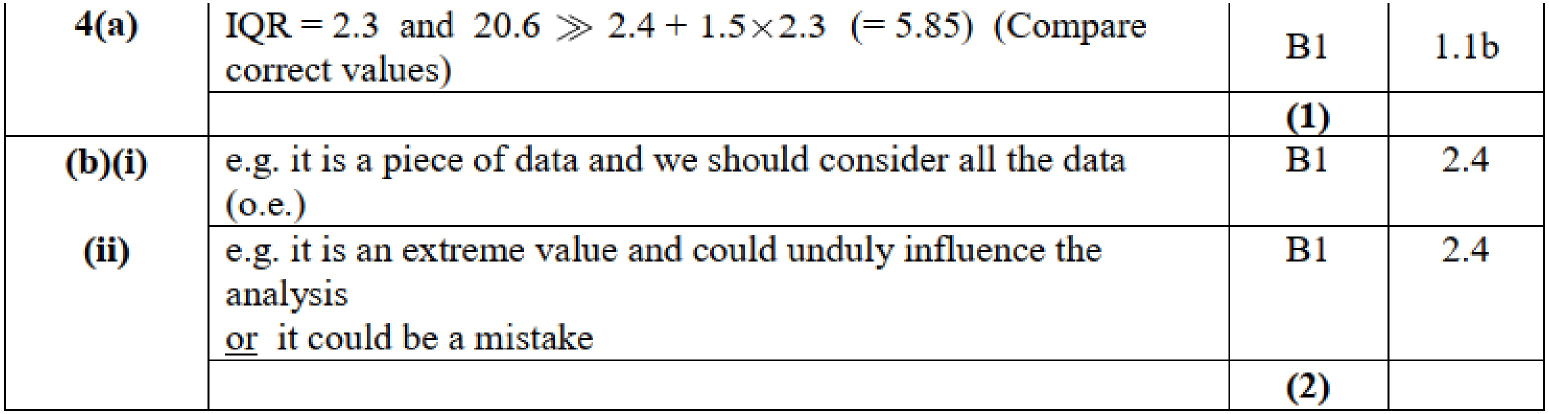
this day's reading.

(2)

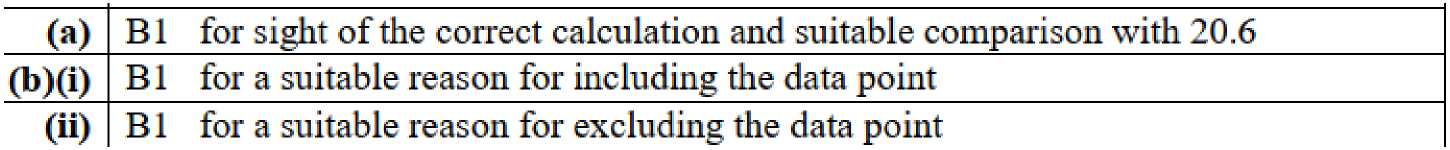
HOME

**AS SAMs**

Location and Spread



| | | | |
|---|---|---|---|
| **4(a)** | IQR = 2.3 and 20.6 $\gg$ 2.4 + 1.5×2.3 (= 5.85) (Compare correct values) | B1 | 1.1b |
| | | **(1)** | |
| **(b)(i)** | e.g. it is a piece of data and we should consider all the data (o.e.) | B1 | 2.4 |
| **(ii)** | e.g. it is an extreme value and could unduly influence the analysis<br>or it could be a mistake | B1 | 2.4 |
| | | **(2)** | |



| | | |
|---|---|---|
| **(a)** | B1 | for sight of the correct calculation and suitable comparison with 20.6 |
| **(b)(i)** | B1 | for a suitable reason for including the data point |
| **(ii)** | B1 | for a suitable reason for excluding the data point |

HOME

# AS 2018

## Location and Spread

**4.** Helen is studying the daily mean wind speed for Camborne using the large data set from 1987. The data for one month are summarised in Table 1 below.

| **Windspeed** | n/a | 6 | 7 | 8 | 9 | 11 | 12 | 13 | 14 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Frequency** | 13 | 2 | 3 | 2 | 2 | 3 | 1 | 2 | 1 | 2 |

**Table 1**

(a) Calculate the mean for these data.

(1)

(b) Calculate the standard deviation for these data and state the units.

(2)

The means and standard deviations of the daily mean wind speed for the other months from the large data set for Camborne in 1987 are given in Table 2 below. The data are not in month order.

| **Month** | *A* | *B* | *C* | *D* | *E* |
|---|---|---|---|---|---|
| **Mean** | 7.58 | 8.26 | 8.57 | 8.57 | 11.57 |
| **Standard Deviation** | 2.93 | 3.89 | 3.46 | 3.87 | 4.64 |

**Table 2**

(c) Using your knowledge of the large data set, suggest, giving a reason, which month had a mean of 11.57

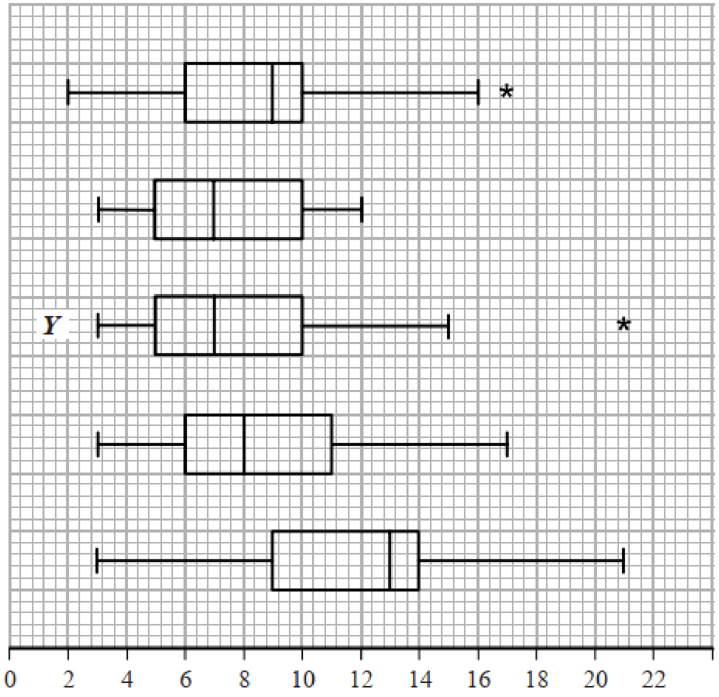
(2)

The data for these months are summarised in the box plots on the opposite page. They are not in month order or the same order as in Table 2.

(d) (i) State the meaning of the * symbol on some of the box plots.

(ii) Suggest, giving your reasons, which of the months in Table 2 is most likely to be summarised in the box plot marked *Y*.

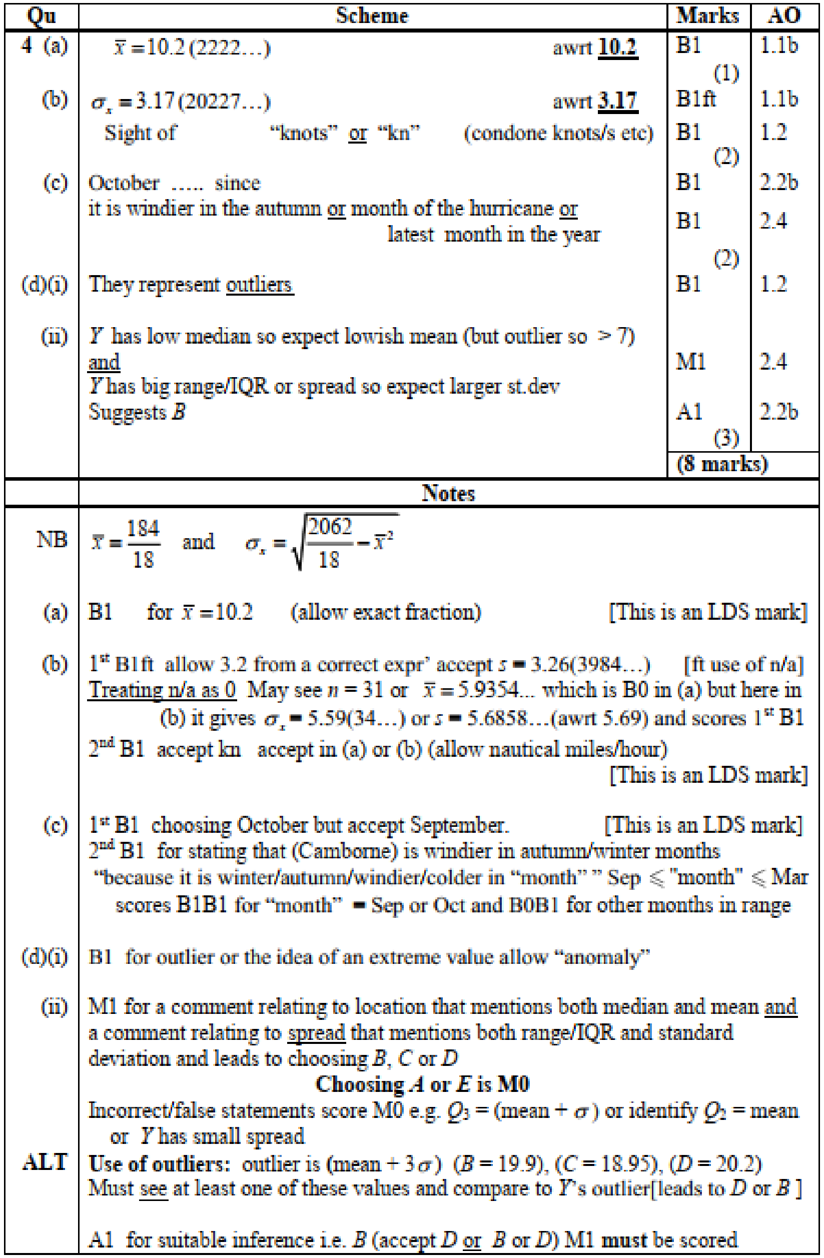
(3)



HOME

# AS 2018

## Location and Spread



| Qu | Scheme | Marks | AO |
|---|---|---|---|
| 4 (a) | $\bar{x} = 10.2\,(2222\ldots)$ awrt **<u>10.2</u>** | B1<br>(1) | 1.1b |
| (b) | $\sigma_x = 3.17\,(20227\ldots)$ awrt **<u>3.17</u>** | B1ft | 1.1b |
| | Sight of "knots" <u>or</u> "kn" (condone knots/s etc) | B1<br>(2) | 1.2 |
| (c) | October ..... since | B1 | 2.2b |
| | it is windier in the autumn <u>or</u> month of the hurricane <u>or</u> latest month in the year | B1<br>(2) | 2.4 |
| (d)(i) | They represent <u>outliers</u> | B1 | 1.2 |
| (ii) | *Y* has low median so expect lowish mean (but outlier so > 7) <u>and</u> | | |
| | *Y* has big range/IQR or spread so expect larger st.dev | M1 | 2.4 |
| | Suggests *B* | A1<br>(3) | 2.2b |
| | | **(8 marks)** | |

| | Notes |
|---|---|
| NB | $\bar{x} = \frac{184}{18}$ and $\sigma_x = \sqrt{\frac{2062}{18} - \bar{x}^2}$ |
| (a) | B1 for $\bar{x} = 10.2$ (allow exact fraction) [This is an LDS mark] |
| (b) | 1st B1ft allow 3.2 from a correct expr' accept $s = 3.26(3984\ldots)$ [ft use of n/a]<br><u>Treating n/a as 0</u> May see $n = 31$ or $\bar{x} = 5.9354$... which is B0 in (a) but here in (b) it gives $\sigma_x = 5.59(34\ldots)$ or $s = 5.6858\ldots$(awrt 5.69) and scores 1st B1<br>2nd B1 accept kn accept in (a) or (b) (allow nautical miles/hour) [This is an LDS mark] |
| (c) | 1st B1 choosing October but accept September. [This is an LDS mark]<br>2nd B1 for stating that (Camborne) is windier in autumn/winter months<br>"because it is winter/autumn/windier/colder in "month" " Sep $\leqslant$ "month" $\leqslant$ Mar scores B1B1 for "month" = Sep or Oct and B0B1 for other months in range |
| (d)(i) | B1 for outlier or the idea of an extreme value allow "anomaly" |
| (ii) | M1 for a comment relating to location that mentions both median and mean <u>and</u> a comment relating to <u>spread</u> that mentions both range/IQR and standard deviation and leads to choosing *B*, *C* or *D*<br>**Choosing *A* or *E* is M0**<br>Incorrect/false statements score M0 e.g. $Q_3$ = (mean + $\sigma$) or identify $Q_2$ = mean or *Y* has small spread |
| ALT | **Use of outliers:** outlier is (mean + 3$\sigma$) ($B = 19.9$), ($C = 18.95$), ($D = 20.2$)<br>Must <u>see</u> at least one of these values and compare to *Y*'s outlier[leads to *D* or *B* ]<br>A1 for suitable inference i.e. *B* (accept *D* <u>or</u> *B* or *D*) M1 **must** be scored |

HOME

# AS 2019

## Location and Spread

**4.** Joshua is investigating the daily total rainfall in Hurn for May to October 2015

Using the information from the large data set, Joshua wishes to calculate the mean of the daily total rainfall in Hurn for May to October 2015

(a) Using your knowledge of the large data set, explain why Joshua needs to clean the data before calculating the mean.

**(1)**

Using the information from the large data set, he produces the grouped frequency table below.

| Daily total rainfall ($r$ mm) | Frequency | Midpoint ($x$ mm) |
|---|---|---|
| $0 \leqslant r < 0.5$ | 121 | 0.25 |
| $0.5 \leqslant r < 1.0$ | 10 | 0.75 |
| $1.0 \leqslant r < 5.0$ | 24 | 3.0 |
| $5.0 \leqslant r < 10.0$ | 12 | 7.5 |
| $10.0 \leqslant r < 30.0$ | 17 | 20.0 |

You may use $\sum fx = 539.75$ and $\sum fx^2 = 7704.1875$

(b) Use linear interpolation to calculate an estimate for the upper quartile of the daily total rainfall.

**(2)**

(c) Calculate an estimate for the standard deviation of the daily total rainfall in Hurn for May to October 2015

**(2)**

(d) (i) State the assumption involved with using class midpoints to calculate an estimate of a mean from a grouped frequency table.

(ii) Using your knowledge of the large data set, explain why this assumption does not hold in this case.

(iii) State, giving a reason, whether you would expect the actual mean daily total rainfall in Hurn for May to October 2015 to be larger than, smaller than or the same as an estimate based on the grouped frequency table.

**(3)**

HOME

# AS 2019

## Location and Spread

| Part | Working or answer an examiner might expect to see | Mark | Notes |
|---|---|---|---|
| (a) | Trace data needs to be converted to numbers before the calculation can be carried out | B1 | This mark is given for a valid explanation |
| (b) | $(1 +)\ \dfrac{138-131}{24} \times 4$ | M1 | This mark is given for a methods to find an estimate of the upper quartile |
| | $= 2.17$ | A1 | This mark is given for a correct estimate of the upper quartile |
| (c) | $\sigma = \sqrt{\dfrac{7704.1875}{184} - \left(\dfrac{539.75}{184}\right)^2}$ | M1 | This mark is given for using the formula for standard deviation to find an estimate for the standard deviation of the total daily rainfall |
| | $= 5.77$ | A1 | This mark is given for a correct estimate for the standard deviation of the total daily rainfall |
| (d)(i) | Using class midpoints to estimate the mean assumes that the values are uniformly distributed in each class | B1 | This mark is given for an explanation that the data is assumed to be spread evenly across each class |
| (d)(ii) | The assumption does not hold since the majority of the data in the first class are 0 | B1 | This mark is given for a valid explanation by the assumption does not hold |
| (d)(iii) | The actual mean is likely to be smaller than the estimate; the first group has more values at 0 and close to 0 | B1 | This mark is given for a correct inference based on knowledge of the Large Data Set |

HOME

# AS 2020

# Location and Spread

**2.** Jerry is studying visibility for Camborne using the large data set June 1987.

The table below contains two extracts from the large data set.

It shows the daily maximum relative humidity and the daily mean visibility.

| **Date** | **Daily Maximum Relative Humidity** | **Daily Mean Visibility** |
|---|---|---|
| **Units** | **%** | |
| 10/06/1987 | 90 | 5300 |
| 28/06/1987 | 100 | 0 |

(The units for Daily Mean Visibility are deliberately omitted.)

Given that daily mean visibility is given to the nearest 100,

(a) write down the range of distances in metres that corresponds to the recorded value 0 for the daily mean visibility.

**(1)**

Jerry drew the following scatter diagram, Figure 2, and calculated some statistics using the June 1987 data for Camborne from the large data set.

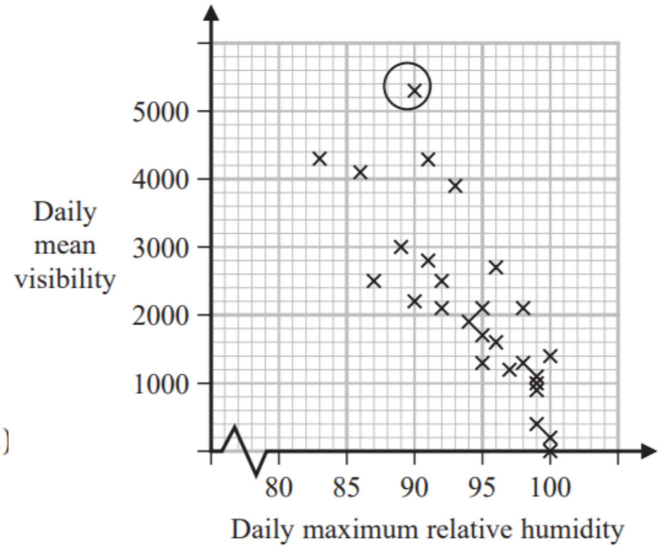


**Figure 2**

| | $Q_1$ | IQR |
|---|---|---|
| **Daily mean visibility** | 1100 | 1600 |
| **Daily maximum relative humidity (%)** | 92 | 8 |

Jerry defines an outlier as a value that is more than 1.5 times the interquartile range above $Q_3$ or more than 1.5 times the interquartile range below $Q_1$.

(b) Show that the point circled on the scatter diagram is an outlier for visibility.

**(2)**

(c) Interpret the correlation between the daily mean visibility and the daily maximum relative humidity.

**(1)**

HOME

**AS 2020**

Location and Spread

Jerry drew the following scatter diagram, Figure 3, using the June 1987 data for Camborne from the large data set, but forgot to label the $x$–axis.

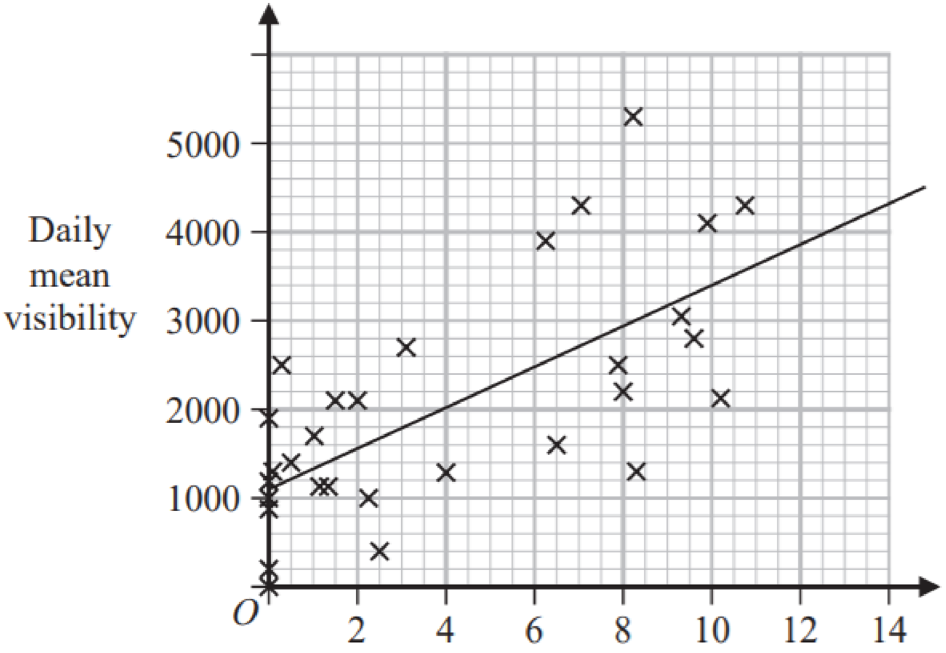


**Figure 3**

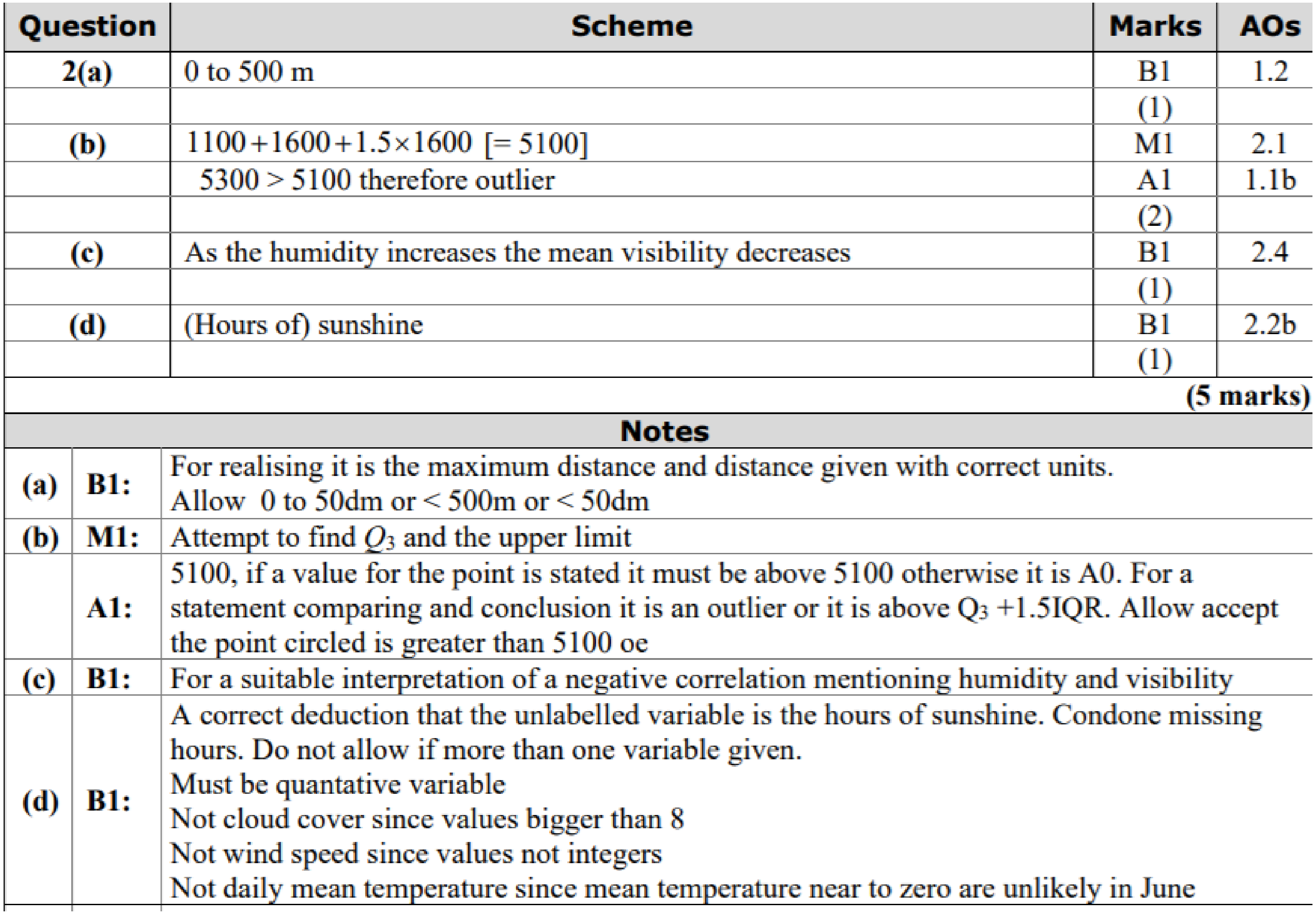
(d) Using your knowledge of the large data set, suggest which variable the $x$-axis on this scatter diagram represents.

**(1)**

HOME

AS 2020

Location and Spread



| Question | Scheme | Marks | AOs |
|---|---|---|---|
| 2(a) | 0 to 500 m | B1 | 1.2 |
| | | (1) | |
| (b) | $1100+1600+1.5\times1600\ [=5100]$ | M1 | 2.1 |
| | $5300>5100$ therefore outlier | A1 | 1.1b |
| | | (2) | |
| (c) | As the humidity increases the mean visibility decreases | B1 | 2.4 |
| | | (1) | |
| (d) | (Hours of) sunshine | B1 | 2.2b |
| | | (1) | |
| | | | (5 marks) |

**Notes**

| | | |
|---|---|---|
| (a) | B1: | For realising it is the maximum distance and distance given with correct units.<br>Allow 0 to 50dm or < 500m or < 50dm |
| (b) | M1: | Attempt to find $Q_3$ and the upper limit |
| | A1: | 5100, if a value for the point is stated it must be above 5100 otherwise it is A0. For a statement comparing and conclusion it is an outlier or it is above $Q_3$ +1.5IQR. Allow accept the point circled is greater than 5100 oe |
| (c) | B1: | For a suitable interpretation of a negative correlation mentioning humidity and visibility |
| (d) | B1: | A correct deduction that the unlabelled variable is the hours of sunshine. Condone missing hours. Do not allow if more than one variable given.<br>Must be quantative variable<br>Not cloud cover since values bigger than 8<br>Not wind speed since values not integers<br>Not daily mean temperature since mean temperature near to zero are unlikely in June |

HOME

# AS 2020

## Location and Spread

4. A lake contains three different types of carp.

There are an estimated 450 mirror carp, 300 leather carp and 850 common carp.

Tim wishes to investigate the health of the fish in the lake.

He decides to take a sample of 160 fish.

(a) Give a reason why stratified random sampling cannot be used. **(1)**

(b) Explain how a sample of size 160 could be taken to ensure that the estimated populations of each type of carp are fairly represented.

You should state the name of the sampling method used. **(2)**

As part of the health check, Tim weighed the fish.

His results are given in the table below.

| **Weight ($w$ kg)** | **Frequency (f)** | **Midpoint ($m$ kg)** |
|---|---|---|
| $2 \leqslant w < 3.5$ | 8 | 2.75 |
| $3.5 \leqslant w < 4$ | 32 | 3.75 |
| $4 \leqslant w < 4.5$ | 64 | 4.25 |
| $4.5 \leqslant w < 5$ | 40 | 4.75 |
| $5 \leqslant w < 6$ | 16 | 5.5 |

(You may use $\sum fm = 692$ and $\sum fm^2 = 3053$)

(c) Calculate an estimate for the standard deviation of the weight of the carp. **(2)**

Tim realised that he had transposed the figures for 2 of the weights of the fish.

He had recorded in the table 2.3 instead of 3.2 and 4.6 instead of 6.4

(d) Without calculating a new estimate for the standard deviation, state what effect

(i) using the correct figure of 3.2 instead of 2.3

(ii) using the correct figure of 6.4 instead of 4.6

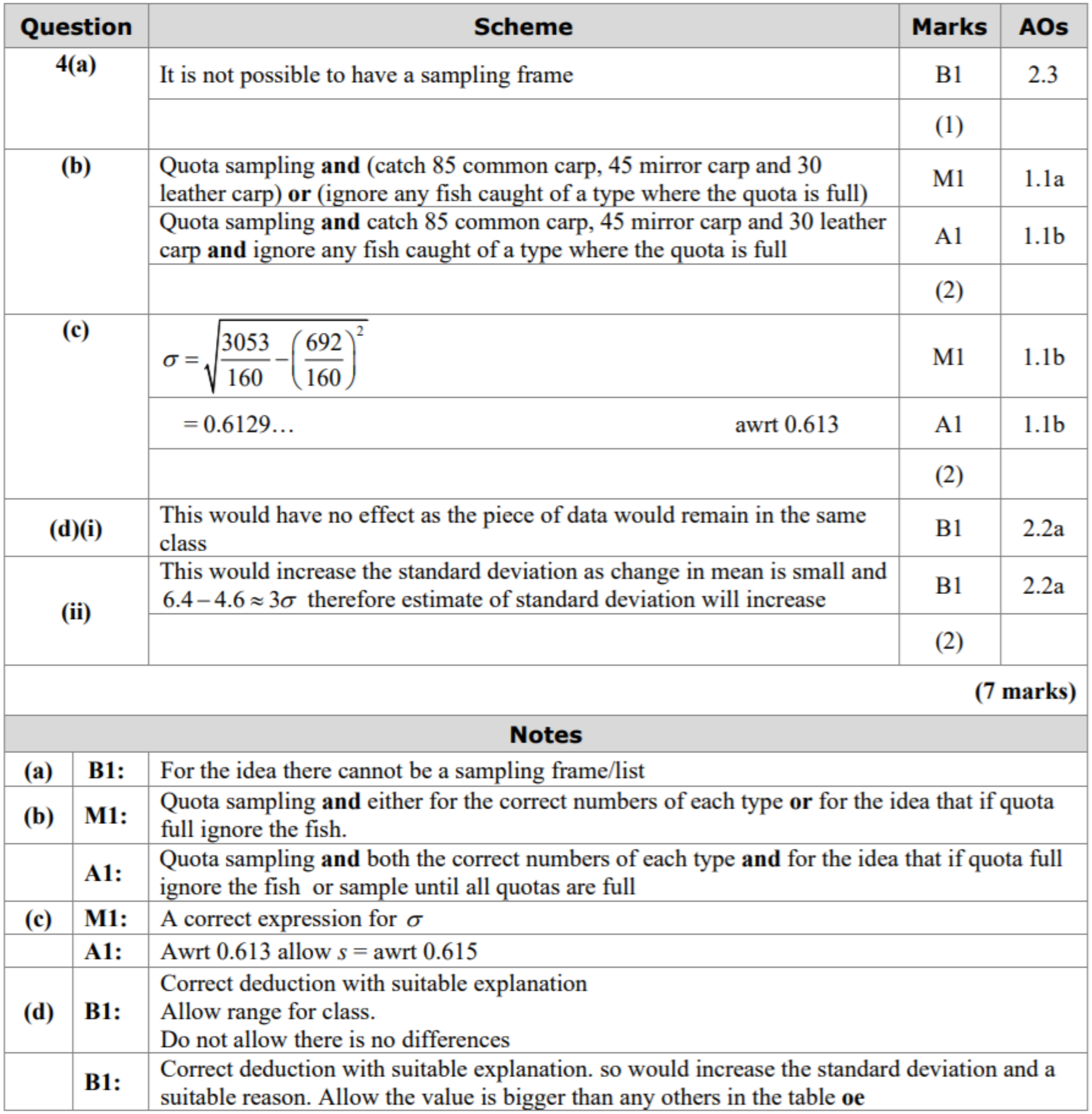
would have on your estimated standard deviation.

Give a reason for each of your answers. **(2)**

HOME

# AS 2020

## Location and Spread



| Question | Scheme | Marks | AOs |
|---|---|---|---|
| 4(a) | It is not possible to have a sampling frame | B1 | 2.3 |
| | | (1) | |
| (b) | Quota sampling **and** (catch 85 common carp, 45 mirror carp and 30 leather carp) **or** (ignore any fish caught of a type where the quota is full) | M1 | 1.1a |
| | Quota sampling **and** catch 85 common carp, 45 mirror carp and 30 leather carp **and** ignore any fish caught of a type where the quota is full | A1 | 1.1b |
| | | (2) | |
| (c) | $\sigma = \sqrt{\frac{3053}{160} - \left(\frac{692}{160}\right)^2}$ | M1 | 1.1b |
| | $= 0.6129\ldots$ awrt 0.613 | A1 | 1.1b |
| | | (2) | |
| (d)(i) | This would have no effect as the piece of data would remain in the same class | B1 | 2.2a |
| (ii) | This would increase the standard deviation as change in mean is small and $6.4 - 4.6 \approx 3\sigma$ therefore estimate of standard deviation will increase | B1 | 2.2a |
| | | (2) | |
| | | | **(7 marks)** |

| Notes | | |
|---|---|---|
| (a) | B1: | For the idea there cannot be a sampling frame/list |
| (b) | M1: | Quota sampling **and** either for the correct numbers of each type **or** for the idea that if quota full ignore the fish. |
| | A1: | Quota sampling **and** both the correct numbers of each type **and** for the idea that if quota full ignore the fish or sample until all quotas are full |
| (c) | M1: | A correct expression for $\sigma$ |
| | A1: | Awrt 0.613 allow $s$ = awrt 0.615 |
| (d) | B1: | Correct deduction with suitable explanation<br>Allow range for class.<br>Do not allow there is no differences |
| | B1: | Correct deduction with suitable explanation. so would increase the standard deviation and a suitable reason. Allow the value is bigger than any others in the table **oe** |

HOME

## AS 2021

## Location and Spread

**2.** The partially completed table and partially completed histogram give information about the ages of passengers on an airline.

There were no passengers aged 90 or over.

| **Age ($x$ years)** | $0 \leqslant x < 5$ | $5 \leqslant x < 20$ | $20 \leqslant x < 40$ | $40 \leqslant x < 65$ | $65 \leqslant x < 80$ | $80 \leqslant x < 90$ |
|---|---|---|---|---|---|---|
| **Frequency** | 5 | 45 | 90 | | | 1 |

(a) Complete the histogram. *(on the next slide)*

**(3)**

(b) Use linear interpolation to estimate the median age.

**(4)**

An outlier is defined as a value greater than $Q_3 + 1.5 \times$ interquartile range.

Given that $Q_1 = 27.3$ and $Q_3 = 58.9$

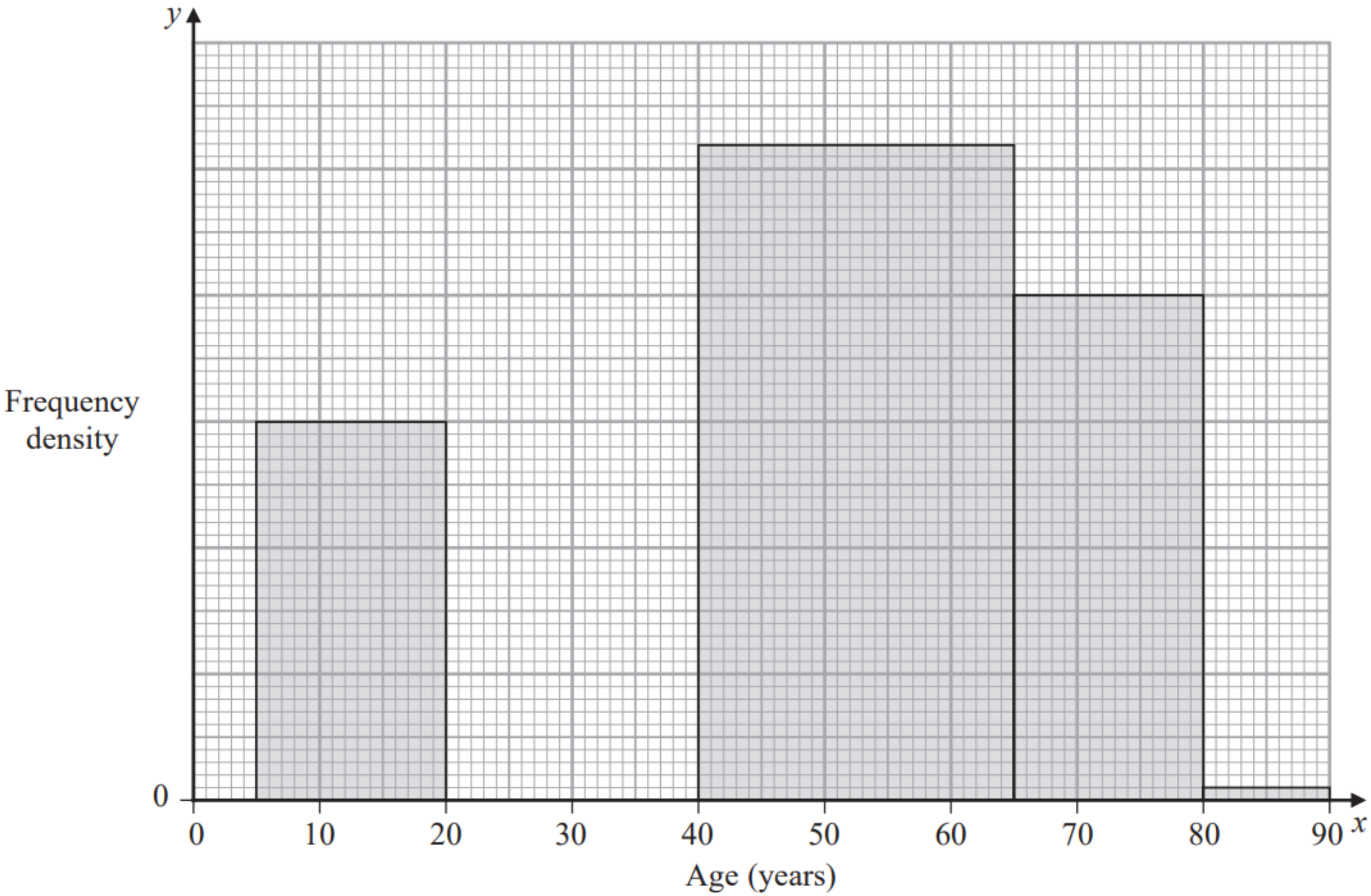
(c) determine, giving a reason, whether or not the oldest passenger could be considered as an outlier.

**(2)**

HOME

# AS 2021

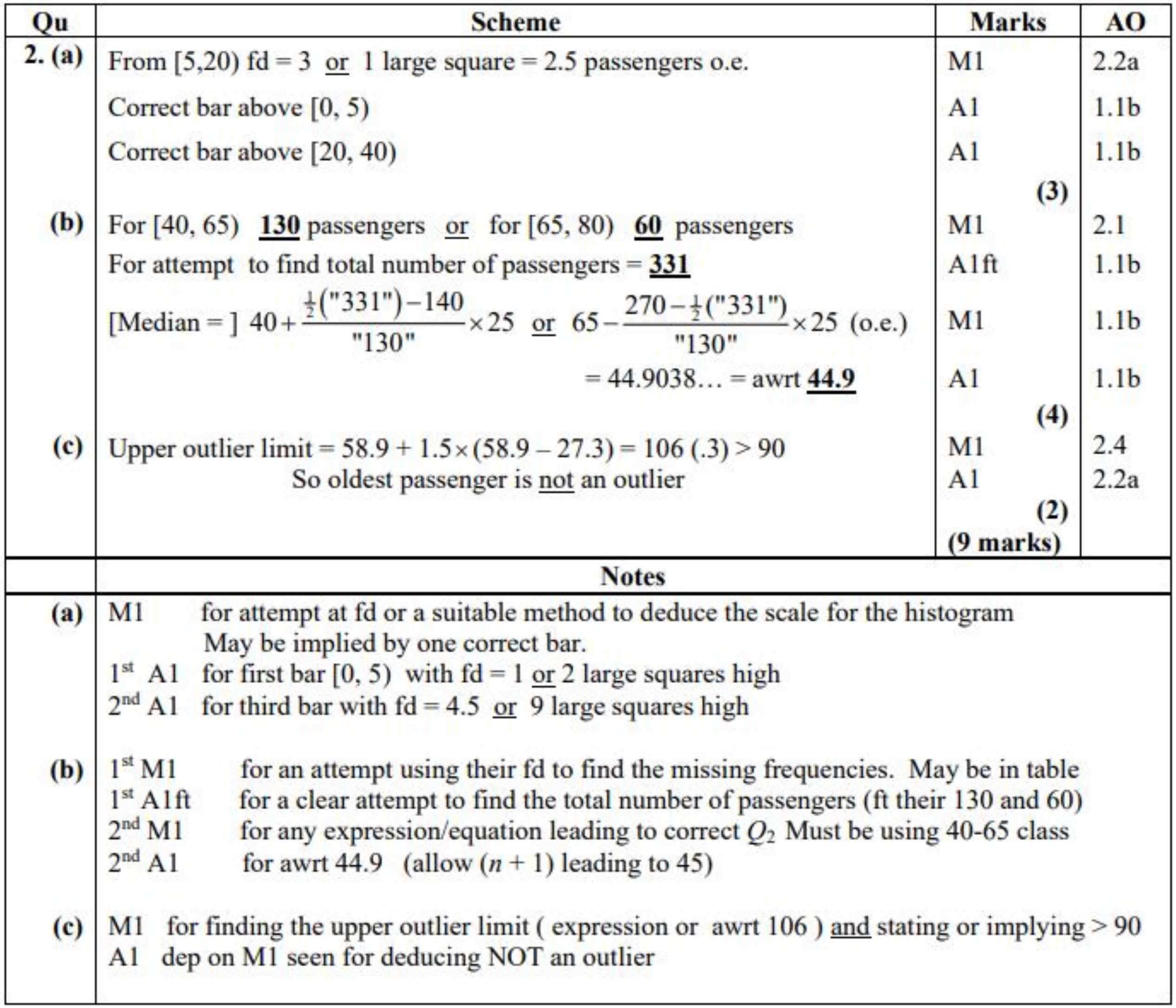
Location and Spread



HOME

# AS 2021

## Location and Spread



| Qu | Scheme | Marks | AO |
|---|---|---|---|
| 2. (a) | From [5,20) fd = 3 <u>or</u> 1 large square = 2.5 passengers o.e. | M1 | 2.2a |
| | Correct bar above [0, 5) | A1 | 1.1b |
| | Correct bar above [20, 40) | A1 | 1.1b |
| | | **(3)** | |
| (b) | For [40, 65) **<u>130</u>** passengers <u>or</u> for [65, 80) **<u>60</u>** passengers | M1 | 2.1 |
| | For attempt to find total number of passengers = **<u>331</u>** | A1ft | 1.1b |
| | [Median = ] $40+\frac{\frac{1}{2}(\text{"331"})-140}{\text{"130"}}\times 25$ <u>or</u> $65-\frac{270-\frac{1}{2}(\text{"331"})}{\text{"130"}}\times 25$ (o.e.) | M1 | 1.1b |
| | = 44.9038… = awrt **<u>44.9</u>** | A1 | 1.1b |
| | | **(4)** | |
| (c) | Upper outlier limit = $58.9 + 1.5\times(58.9 - 27.3) = 106\ (.3) > 90$<br>So oldest passenger is <u>not</u> an outlier | M1<br>A1 | 2.4<br>2.2a |
| | | **(2)** | |
| | | **(9 marks)** | |

| | Notes |
|---|---|
| (a) | M1 for attempt at fd or a suitable method to deduce the scale for the histogram<br>May be implied by one correct bar.<br>1st A1 for first bar [0, 5) with fd = 1 <u>or</u> 2 large squares high<br>2nd A1 for third bar with fd = 4.5 <u>or</u> 9 large squares high |
| (b) | 1st M1 for an attempt using their fd to find the missing frequencies. May be in table<br>1st A1ft for a clear attempt to find the total number of passengers (ft their 130 and 60)<br>2nd M1 for any expression/equation leading to correct $Q_2$ Must be using 40-65 class<br>2nd A1 for awrt 44.9 (allow $(n + 1)$ leading to 45) |
| (c) | M1 for finding the upper outlier limit ( expression or awrt 106 ) <u>and</u> stating or implying > 90<br>A1 dep on M1 seen for deducing NOT an outlier |

HOME

# AS 2021

Location and Spread

**3.** Helen is studying one of the qualitative variables from the large data set for Heathrow from 2015.

She started with the data from 3rd May and then took every 10th reading.

There were only 3 different outcomes with the following frequencies

| **Outcome** | *A* | *B* | *C* |
|---|---|---|---|
| **Frequency** | 16 | 2 | 1 |

(a) State the sampling technique Helen used.

**(1)**

(b) From your knowledge of the large data set

(i) suggest which variable was being studied,

(ii) state the name of outcome *A*.

**(2)**

George is also studying the same variable from the large data set for Heathrow from 2015. He started with the data from 5th May and then took every 10th reading and obtained the following

| **Outcome** | *A* | *B* | *C* |
|---|---|---|---|
| **Frequency** | 16 | 1 | 1 |

Helen and George decided they should examine all of the data for this variable for Heathrow from 2015 and obtained the following

| **Outcome** | *A* | *B* | *C* |
|---|---|---|---|
| **Frequency** | 155 | 26 | 3 |

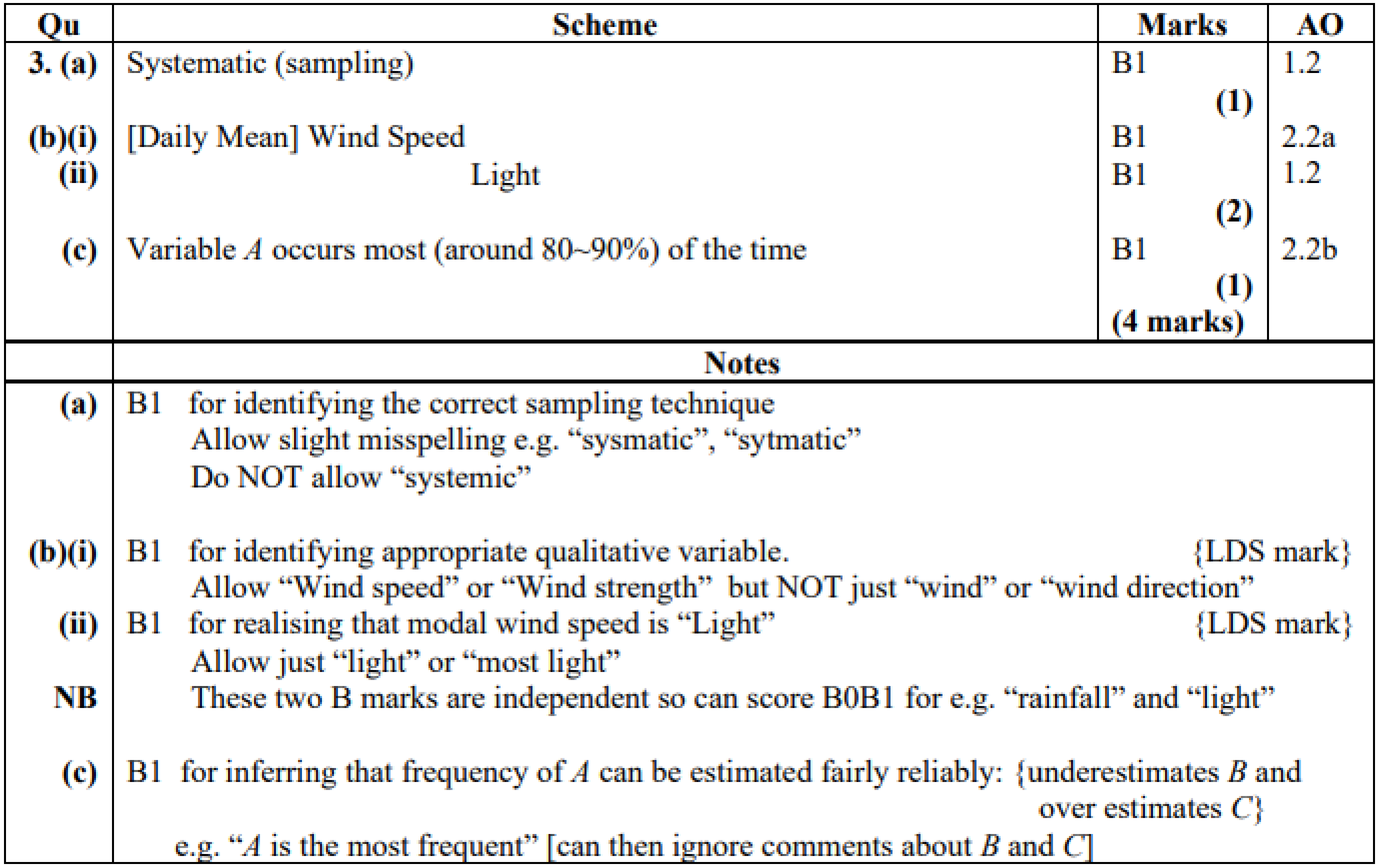
(c) State what inference Helen and George could reliably make from their original samples about the outcomes of this variable at Heathrow, for the period covered by the large data set in 2015.

**(1)**

HOME

**AS 2021**

## Location and Spread

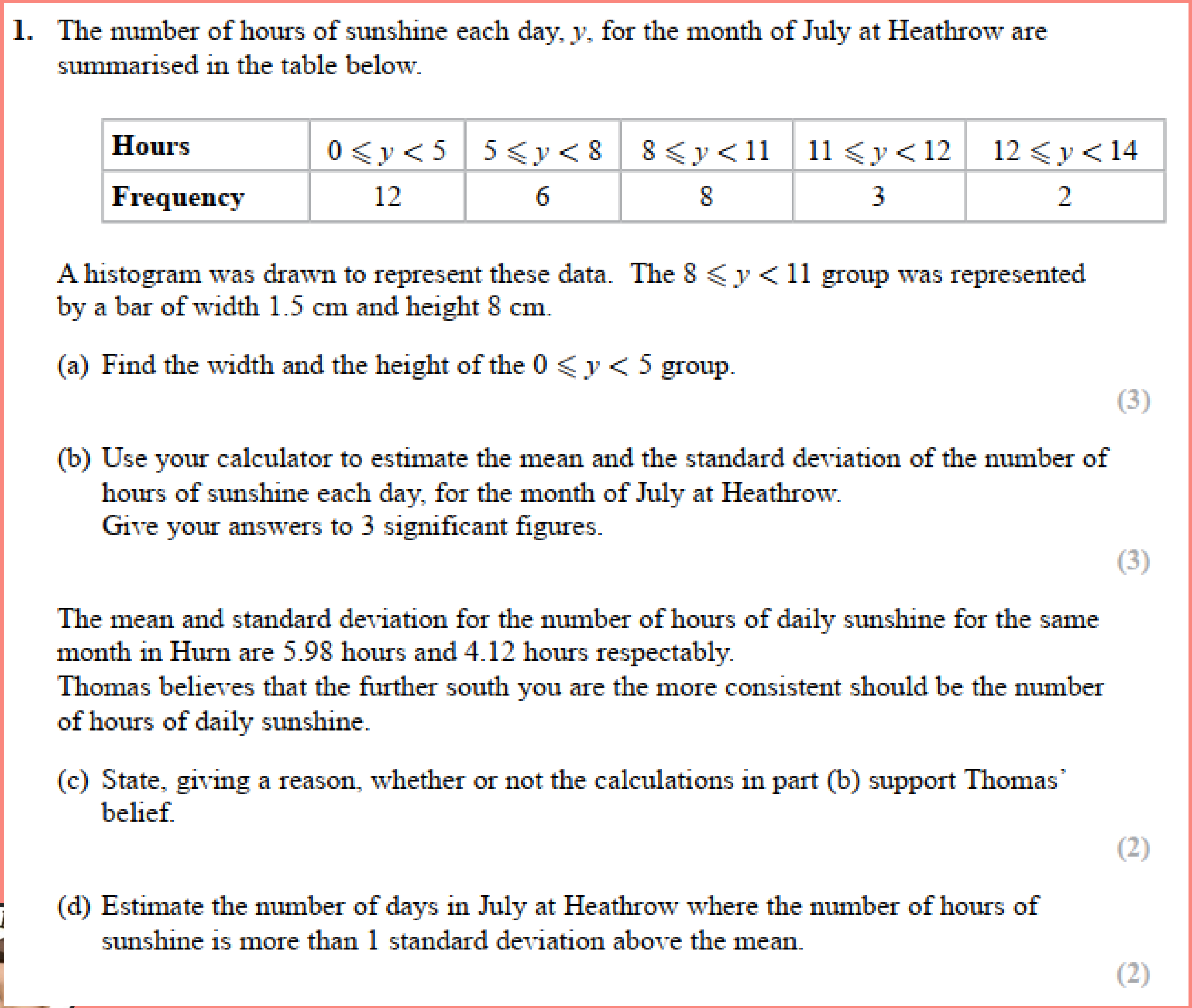


| Qu | Scheme | Marks | AO |
|---|---|---|---|
| **3. (a)** | Systematic (sampling) | B1<br>**(1)** | 1.2 |
| **(b)(i)** | [Daily Mean] Wind Speed | B1 | 2.2a |
| **(ii)** | Light | B1<br>**(2)** | 1.2 |
| **(c)** | Variable *A* occurs most (around 80~90%) of the time | B1<br>**(1)**<br>**(4 marks)** | 2.2b |
| | **Notes** | | |
| **(a)** | B1 for identifying the correct sampling technique<br>Allow slight misspelling e.g. "sysmatic", "sytmatic"<br>Do NOT allow "systemic" | | |
| **(b)(i)** | B1 for identifying appropriate qualitative variable. {LDS mark}<br>Allow "Wind speed" or "Wind strength" but NOT just "wind" or "wind direction" | | |
| **(ii)** | B1 for realising that modal wind speed is "Light" {LDS mark}<br>Allow just "light" or "most light" | | |
| **NB** | These two B marks are independent so can score B0B1 for e.g. "rainfall" and "light" | | |
| **(c)** | B1 for inferring that frequency of *A* can be estimated fairly reliably: {underestimates *B* and over estimates *C*}<br>e.g. "*A* is the most frequent" [can then ignore comments about *B* and *C*] | | |

HOME

# A2 SAMs

## Location and Spread



**1.** The number of hours of sunshine each day, $y$, for the month of July at Heathrow are summarised in the table below.

| **Hours** | $0 \leqslant y < 5$ | $5 \leqslant y < 8$ | $8 \leqslant y < 11$ | $11 \leqslant y < 12$ | $12 \leqslant y < 14$ |
|---|---|---|---|---|---|
| **Frequency** | 12 | 6 | 8 | 3 | 2 |

A histogram was drawn to represent these data. The $8 \leqslant y < 11$ group was represented by a bar of width 1.5 cm and height 8 cm.

(a) Find the width and the height of the $0 \leqslant y < 5$ group.

**(3)**

(b) Use your calculator to estimate the mean and the standard deviation of the number of hours of sunshine each day, for the month of July at Heathrow.
Give your answers to 3 significant figures.

**(3)**

The mean and standard deviation for the number of hours of daily sunshine for the same month in Hurn are 5.98 hours and 4.12 hours respectably.
Thomas believes that the further south you are the more consistent should be the number of hours of daily sunshine.

(c) State, giving a reason, whether or not the calculations in part (b) support Thomas' belief.

**(2)**

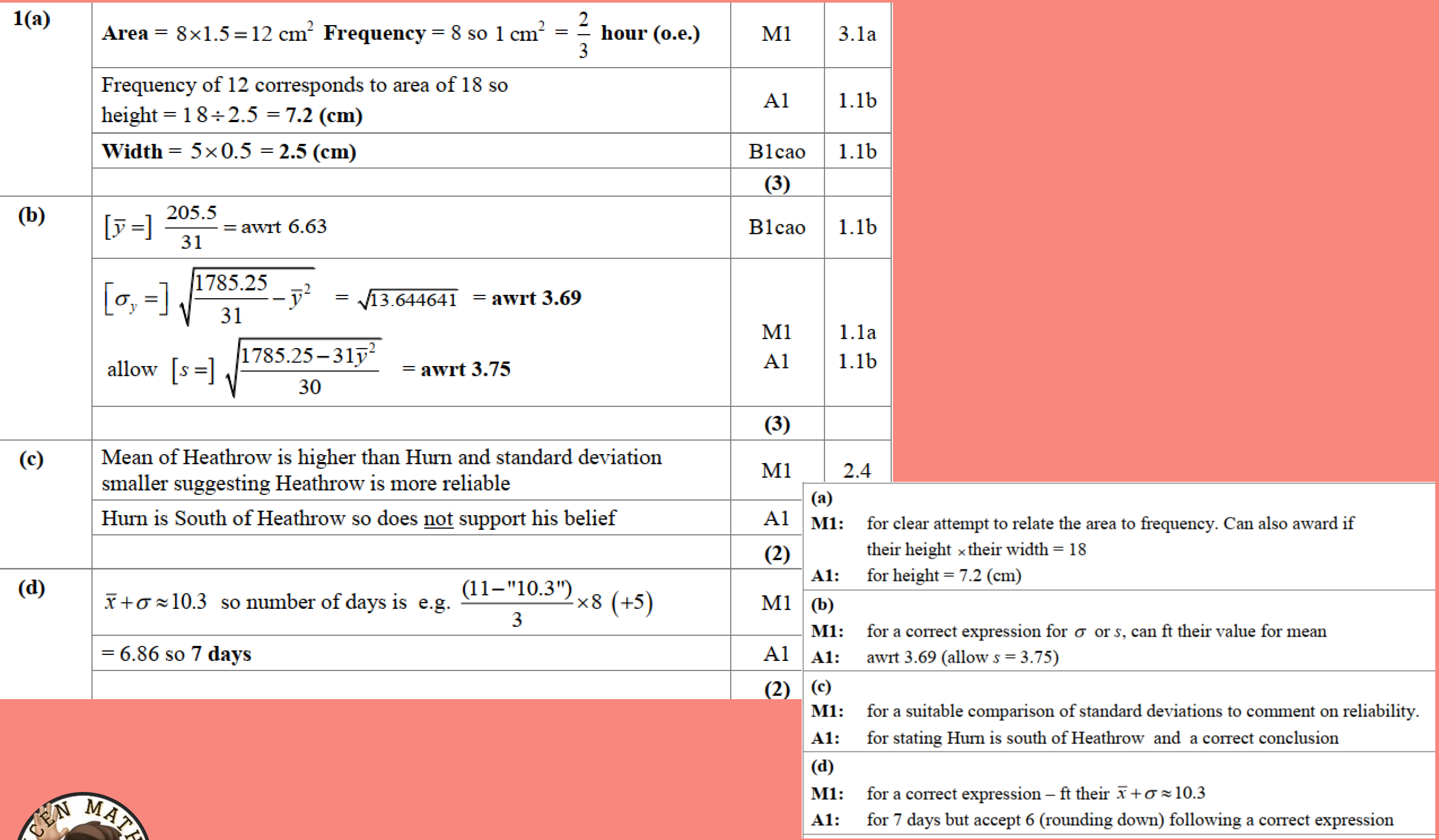
(d) Estimate the number of days in July at Heathrow where the number of hours of sunshine is more than 1 standard deviation above the mean.

**(2)**

HOME

# A2 SAMs

## Location and Spread



| | | | |
|---|---|---|---|
| 1(a) | **Area** = $8\times1.5=12\text{ cm}^2$ **Frequency** = 8 so $1\text{ cm}^2=\frac{2}{3}$ **hour (o.e.)** | M1 | 3.1a |
| | Frequency of 12 corresponds to area of 18 so<br>height = $18\div2.5$ = **7.2 (cm)** | A1 | 1.1b |
| | **Width** = $5\times0.5$ = **2.5 (cm)** | B1cao | 1.1b |
| | | **(3)** | |
| (b) | $[\bar{y}=]\ \frac{205.5}{31}$ = awrt 6.63 | B1cao | 1.1b |
| | $[\sigma_y=]\ \sqrt{\frac{1785.25}{31}-\bar{y}^2}=\sqrt{13.644641}$ = **awrt 3.69**<br>allow $[s=]\ \sqrt{\frac{1785.25-31\bar{y}^2}{30}}$ = **awrt 3.75** | M1<br>A1 | 1.1a<br>1.1b |
| | | **(3)** | |
| (c) | Mean of Heathrow is higher than Hurn and standard deviation smaller suggesting Heathrow is more reliable | M1 | 2.4 |
| | Hurn is South of Heathrow so does not support his belief | A1 | |
| | | **(2)** | |
| (d) | $\bar{x}+\sigma\approx10.3$ so number of days is e.g. $\frac{(11-\text{"}10.3\text{"})}{3}\times8\ (+5)$ | M1 | |
| | = 6.86 so **7 days** | A1 | |
| | | **(2)** | |

**(a)**

**M1:** for clear attempt to relate the area to frequency. Can also award if their height × their width = 18

**A1:** for height = 7.2 (cm)

**(b)**

**M1:** for a correct expression for $\sigma$ or $s$, can ft their value for mean

**A1:** awrt 3.69 (allow $s$ = 3.75)

**(c)**

**M1:** for a suitable comparison of standard deviations to comment on reliability.

**A1:** for stating Hurn is south of Heathrow and a correct conclusion

**(d)**

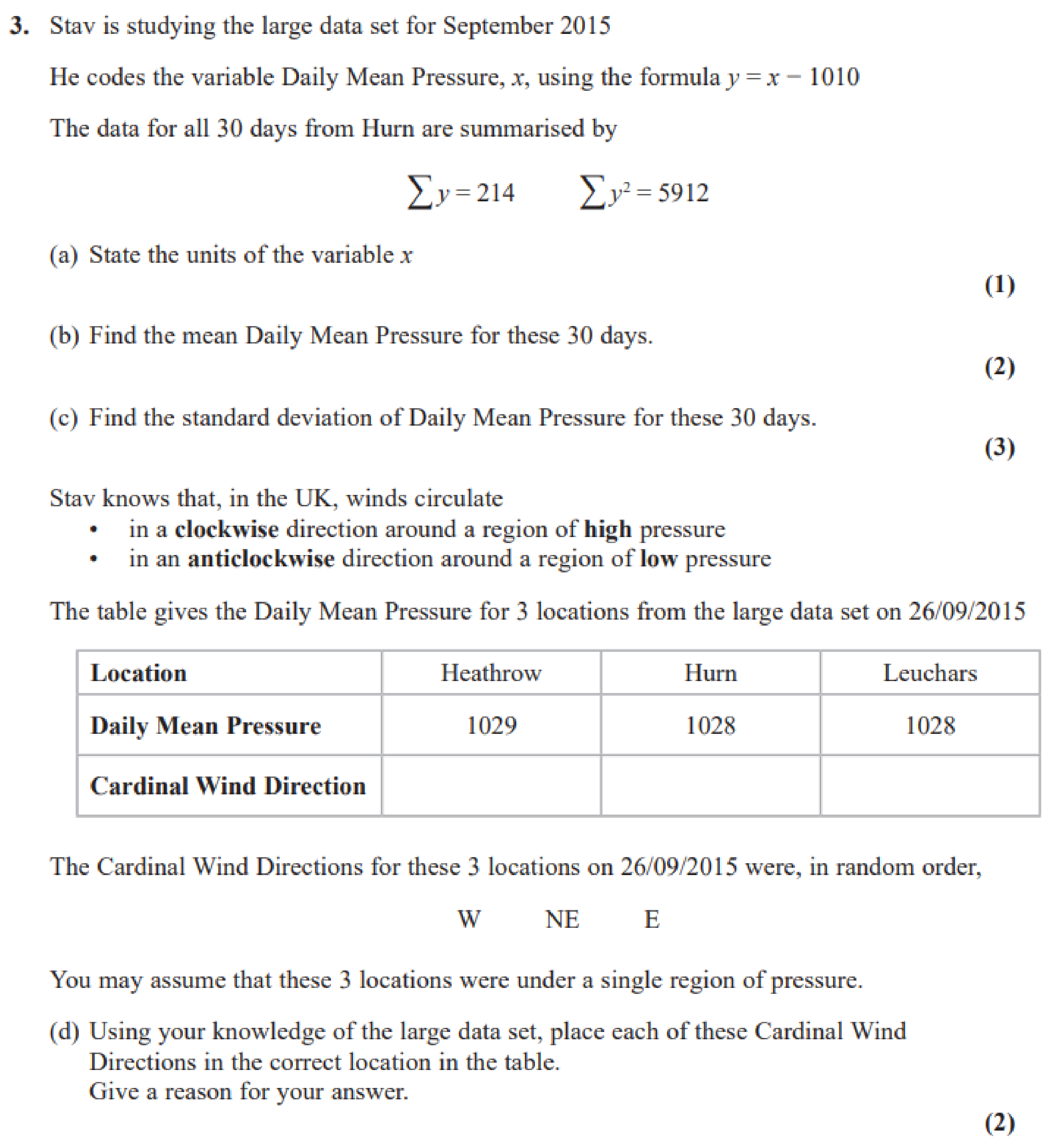
**M1:** for a correct expression – ft their $\bar{x}+\sigma\approx10.3$

**A1:** for 7 days but accept 6 (rounding down) following a correct expression

HOME

# A2 2021

Location and Spread



**3.** Stav is studying the large data set for September 2015

He codes the variable Daily Mean Pressure, $x$, using the formula $y = x - 1010$

The data for all 30 days from Hurn are summarised by

$$\sum y = 214 \qquad \sum y^2 = 5912$$

(a) State the units of the variable $x$

**(1)**

(b) Find the mean Daily Mean Pressure for these 30 days.

**(2)**

(c) Find the standard deviation of Daily Mean Pressure for these 30 days.

**(3)**

Stav knows that, in the UK, winds circulate

- in a **clockwise** direction around a region of **high** pressure
- in an **anticlockwise** direction around a region of **low** pressure

The table gives the Daily Mean Pressure for 3 locations from the large data set on 26/09/2015

| **Location** | Heathrow | Hurn | Leuchars |
|---|---|---|---|
| **Daily Mean Pressure** | 1029 | 1028 | 1028 |
| **Cardinal Wind Direction** | | | |

The Cardinal Wind Directions for these 3 locations on 26/09/2015 were, in random order,

W NE E

You may assume that these 3 locations were under a single region of pressure.

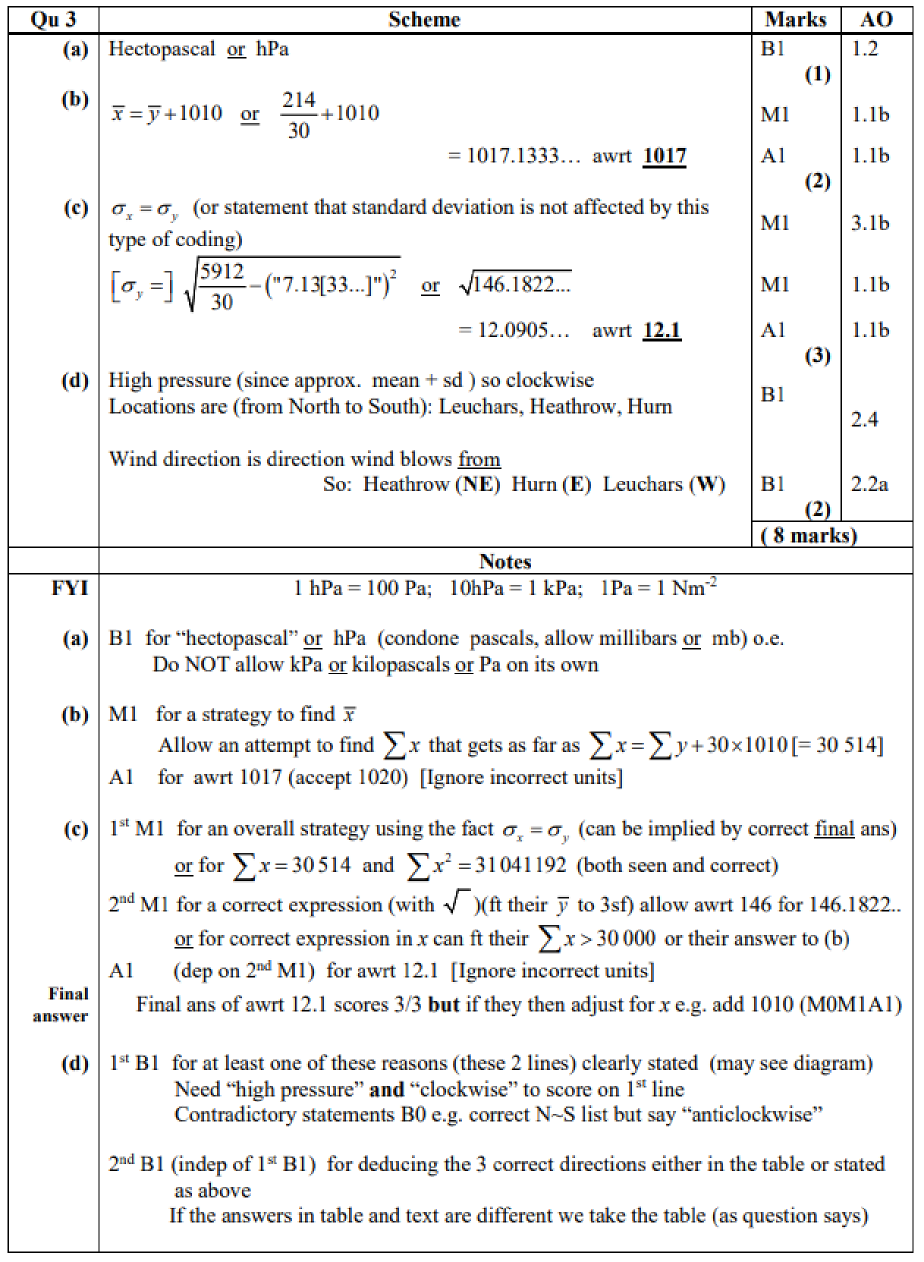
(d) Using your knowledge of the large data set, place each of these Cardinal Wind Directions in the correct location in the table.
Give a reason for your answer.

**(2)**

HOME

# A2 2021

# Location and Spread



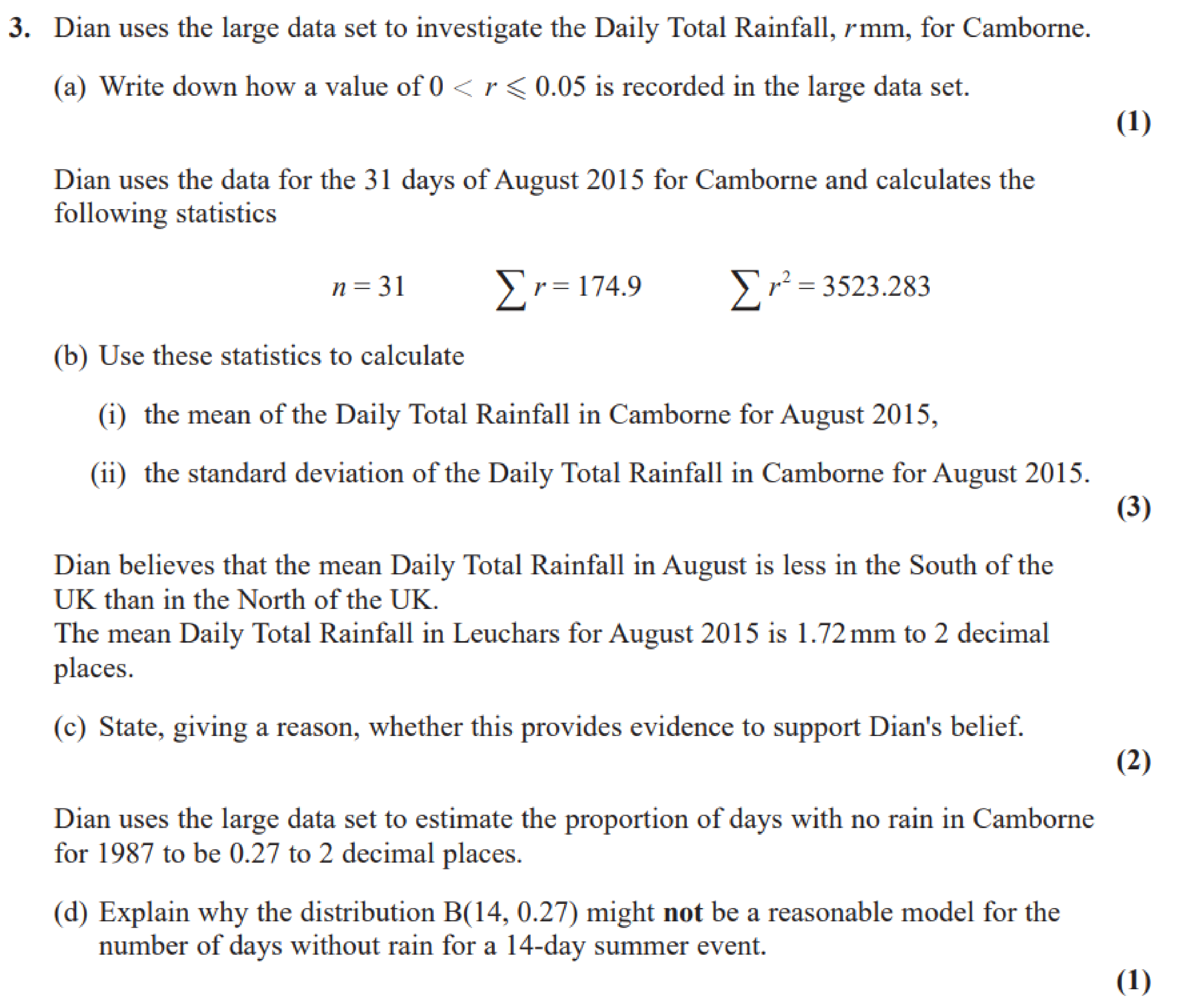
| Qu 3 | Scheme | Marks | AO |
|---|---|---|---|
| (a) | Hectopascal or hPa | B1 | 1.2 |
| | | (1) | |
| (b) | $\bar{x}=\bar{y}+1010$ or $\frac{214}{30}+1010$ | M1 | 1.1b |
| | $= 1017.1333\ldots$ awrt **1017** | A1 | 1.1b |
| | | (2) | |
| (c) | $\sigma_x=\sigma_y$ (or statement that standard deviation is not affected by this type of coding) | M1 | 3.1b |
| | $\left[\sigma_y=\right]\sqrt{\frac{5912}{30}-(\text{"}7.13[33...]\text{"})^2}$ or $\sqrt{146.1822...}$ | M1 | 1.1b |
| | $= 12.0905\ldots$ awrt **12.1** | A1 | 1.1b |
| | | (3) | |
| (d) | High pressure (since approx. mean + sd ) so clockwise<br>Locations are (from North to South): Leuchars, Heathrow, Hurn | B1 | 2.4 |
| | Wind direction is direction wind blows from<br>So: Heathrow (**NE**) Hurn (**E**) Leuchars (**W**) | B1 | 2.2a |
| | | (2) | |
| | | **( 8 marks)** | |

| | Notes |
|---|---|
| FYI | 1 hPa = 100 Pa; 10hPa = 1 kPa; 1Pa = 1 Nm$^{-2}$ |
| (a) | B1 for "hectopascal" or hPa (condone pascals, allow millibars or mb) o.e.<br>Do NOT allow kPa or kilopascals or Pa on its own |
| (b) | M1 for a strategy to find $\bar{x}$<br>Allow an attempt to find $\sum x$ that gets as far as $\sum x=\sum y+30\times1010\,[=30\,514]$<br>A1 for awrt 1017 (accept 1020) [Ignore incorrect units] |
| (c) | 1st M1 for an overall strategy using the fact $\sigma_x=\sigma_y$ (can be implied by correct final ans)<br>or for $\sum x=30\,514$ and $\sum x^2=31\,041\,192$ (both seen and correct)<br>2nd M1 for a correct expression (with $\sqrt{\ }$)(ft their $\bar{y}$ to 3sf) allow awrt 146 for 146.1822..<br>or for correct expression in $x$ can ft their $\sum x>30\,000$ or their answer to (b)<br>A1 (dep on 2nd M1) for awrt 12.1 [Ignore incorrect units] |
| **Final answer** | Final ans of awrt 12.1 scores 3/3 **but** if they then adjust for $x$ e.g. add 1010 (M0M1A1) |
| (d) | 1st B1 for at least one of these reasons (these 2 lines) clearly stated (may see diagram)<br>Need "high pressure" **and** "clockwise" to score on 1st line<br>Contradictory statements B0 e.g. correct N~S list but say "anticlockwise"<br>2nd B1 (indep of 1st B1) for deducing the 3 correct directions either in the table or stated as above<br>If the answers in table and text are different we take the table (as question says) |

HOME

**A2 2022**

## Location and Spread



**3.** Dian uses the large data set to investigate the Daily Total Rainfall, $r$ mm, for Camborne.

(a) Write down how a value of $0 < r \leqslant 0.05$ is recorded in the large data set.

**(1)**

Dian uses the data for the 31 days of August 2015 for Camborne and calculates the following statistics

$$n = 31 \qquad \sum r = 174.9 \qquad \sum r^2 = 3523.283$$

(b) Use these statistics to calculate

(i) the mean of the Daily Total Rainfall in Camborne for August 2015,

(ii) the standard deviation of the Daily Total Rainfall in Camborne for August 2015.

**(3)**

Dian believes that the mean Daily Total Rainfall in August is less in the South of the UK than in the North of the UK.
The mean Daily Total Rainfall in Leuchars for August 2015 is 1.72 mm to 2 decimal places.

(c) State, giving a reason, whether this provides evidence to support Dian's belief.

**(2)**

Dian uses the large data set to estimate the proportion of days with no rain in Camborne for 1987 to be 0.27 to 2 decimal places.

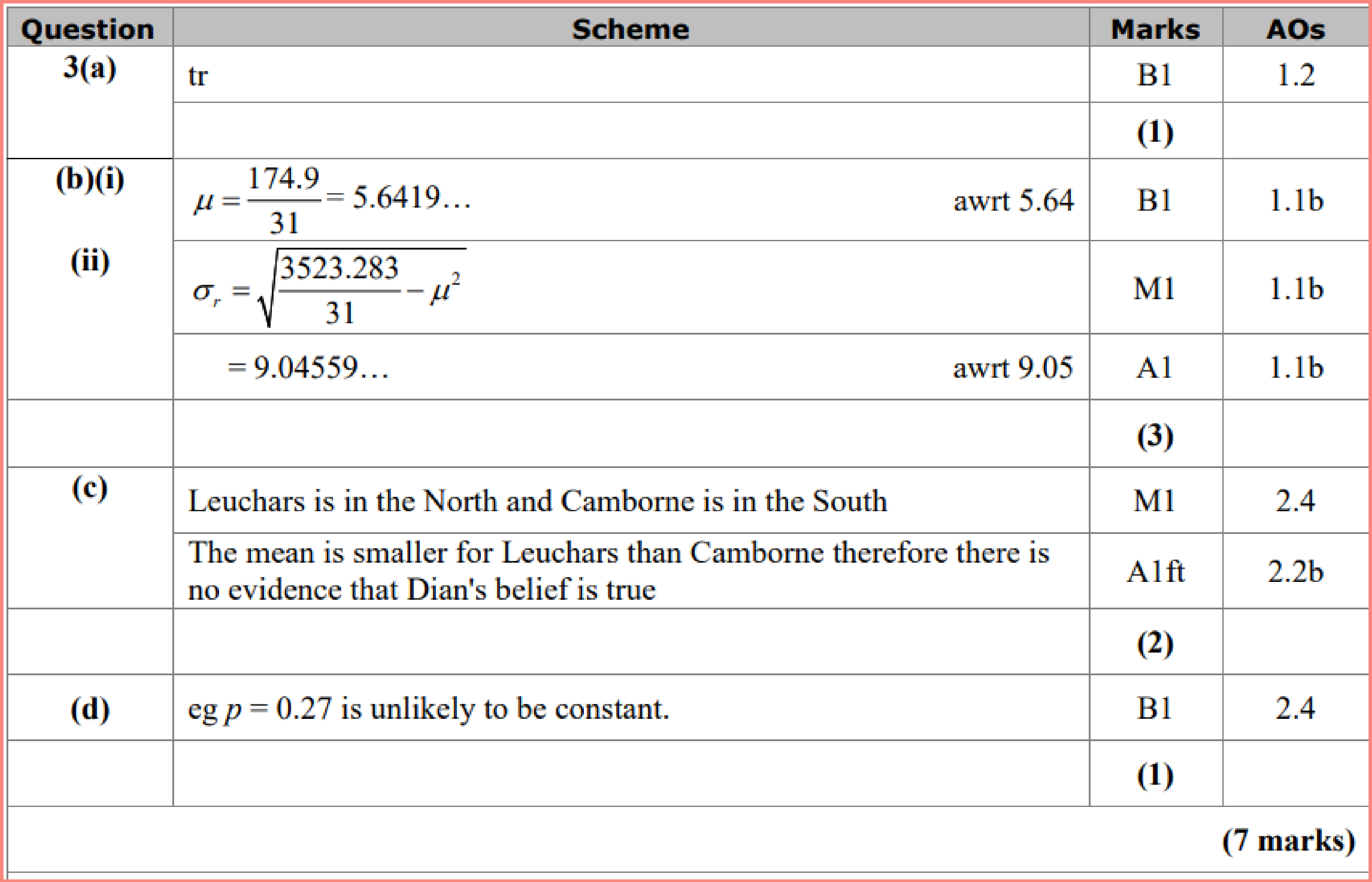
(d) Explain why the distribution B(14, 0.27) might **not** be a reasonable model for the number of days without rain for a 14-day summer event.

**(1)**

HOME

# A2 2022

## Location and Spread

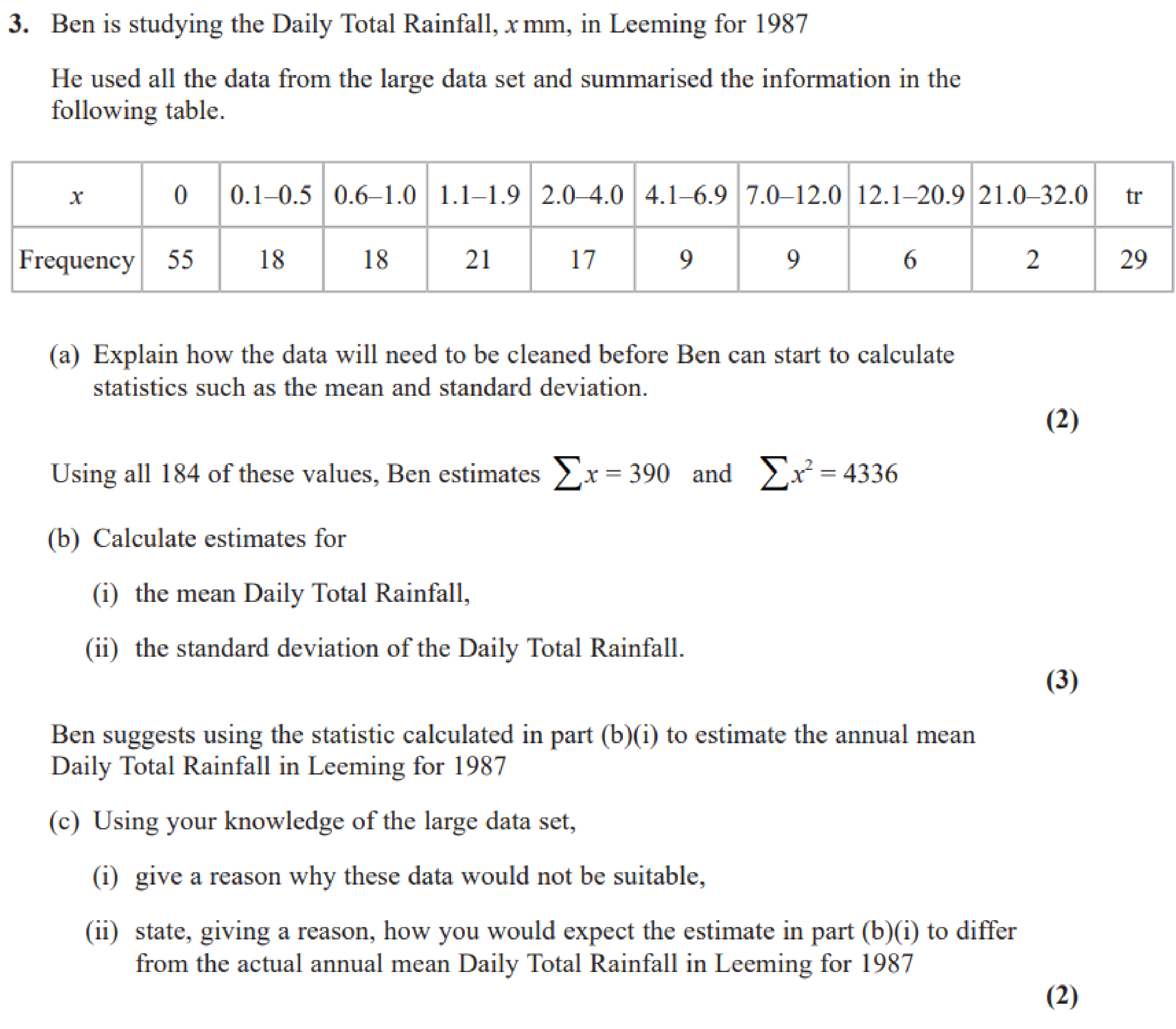


| Question | Scheme | | Marks | AOs |
|---|---|---|---|---|
| 3(a) | tr | | B1 | 1.2 |
| | | | **(1)** | |
| (b)(i) | $\mu = \frac{174.9}{31} = 5.6419\ldots$ | awrt 5.64 | B1 | 1.1b |
| (ii) | $\sigma_r = \sqrt{\frac{3523.283}{31} - \mu^2}$ | | M1 | 1.1b |
| | $= 9.04559\ldots$ | awrt 9.05 | A1 | 1.1b |
| | | | **(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)** |

HOME

**A2 2023**

**Location and Spread**



**3.** Ben is studying the Daily Total Rainfall, $x$ mm, in Leeming for 1987

He used all the data from the large data set and summarised the information in the following table.

| $x$ | 0 | 0.1–0.5 | 0.6–1.0 | 1.1–1.9 | 2.0–4.0 | 4.1–6.9 | 7.0–12.0 | 12.1–20.9 | 21.0–32.0 | tr |
|---|---|---|---|---|---|---|---|---|---|---|
| Frequency | 55 | 18 | 18 | 21 | 17 | 9 | 9 | 6 | 2 | 29 |

(a) Explain how the data will need to be cleaned before Ben can start to calculate statistics such as the mean and standard deviation.

**(2)**

Using all 184 of these values, Ben estimates $\sum x = 390$ and $\sum x^2 = 4336$

(b) Calculate estimates for

(i) the mean Daily Total Rainfall,

(ii) the standard deviation of the Daily Total Rainfall.

**(3)**

Ben suggests using the statistic calculated in part (b)(i) to estimate the annual mean Daily Total Rainfall in Leeming for 1987

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

(i) give a reason why these data would not be suitable,

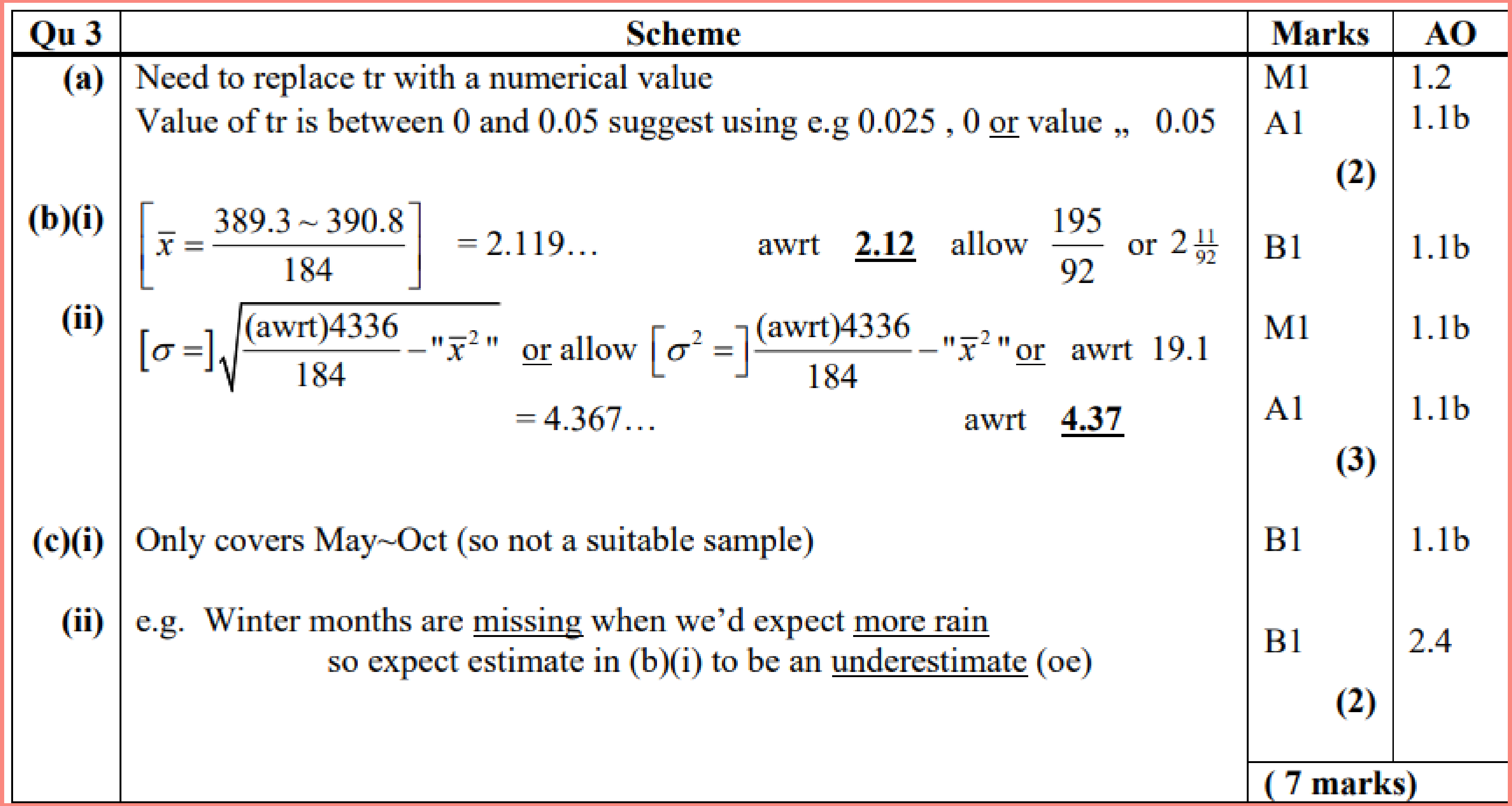
(ii) state, giving a reason, how you would expect the estimate in part (b)(i) to differ from the actual annual mean Daily Total Rainfall in Leeming for 1987

**(2)**

HOME

**A2 2023**

Location and Spread

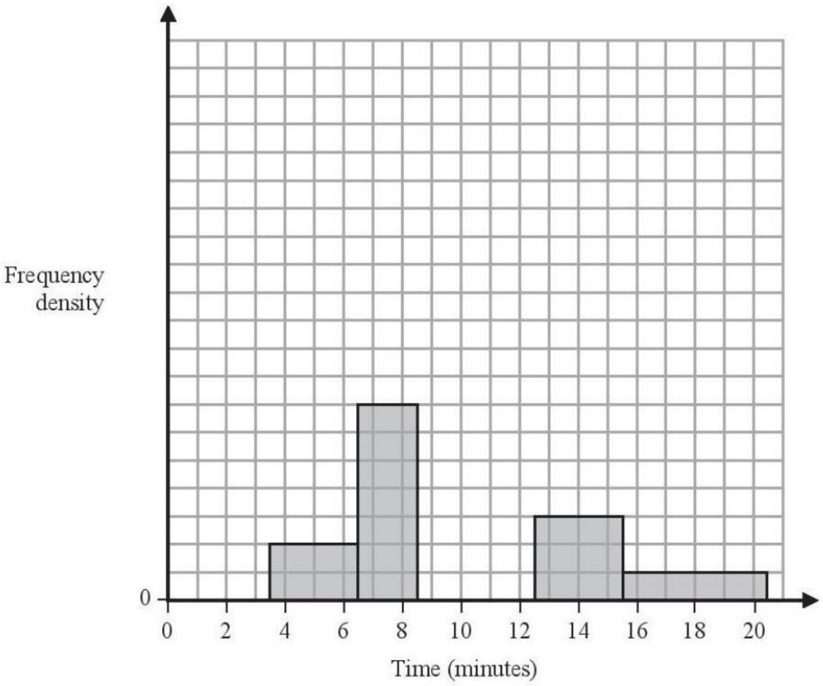


| Qu 3 | Scheme | Marks | AO |
|---|---|---|---|
| (a) | Need to replace tr with a numerical value | M1 | 1.2 |
| | Value of tr is between 0 and 0.05 suggest using e.g 0.025 , 0 or value „ 0.05 | A1 | 1.1b |
| | | **(2)** | |
| (b)(i) | $\left[\bar{x} = \frac{389.3 \sim 390.8}{184}\right] = 2.119\ldots$ awrt **2.12** allow $\frac{195}{92}$ or $2\frac{11}{92}$ | B1 | 1.1b |
| (ii) | $[\sigma =]\sqrt{\frac{(\text{awrt})4336}{184} - "\bar{x}^2"}$ or allow $\left[\sigma^2 =\right]\frac{(\text{awrt})4336}{184} - "\bar{x}^2"$ or awrt 19.1 | M1 | 1.1b |
| | $= 4.367\ldots$ awrt **4.37** | A1 | 1.1b |
| | | **(3)** | |
| (c)(i) | Only covers May~Oct (so not a suitable sample) | B1 | 1.1b |
| (ii) | e.g. Winter months are missing when we'd expect more rain<br>so expect estimate in (b)(i) to be an underestimate (oe) | B1 | 2.4 |
| | | **(2)** | |
| | | **( 7 marks)** | |

HOME

# Representations of Data (Histograms, Box Plots)

**2.** The partially completed histogram and the partially completed table show the time, to the nearest minute, that a random sample of motorists were delayed by roadworks on a stretch of motorway.



| **Delay (minutes)** | **Number of motorists** |
|---|---|
| 4 – 6 | 6 |
| 7 – 8 | |
| 9 | 17 |
| 10 – 12 | 45 |
| 13 – 15 | 9 |
| 16 – 20 | |

Estimate the percentage of these motorists who were delayed by the roadworks for between 8.5 and 13.5 minutes.

**(5)**

HOME

**AS SAMs**

## Representations of Data

| Question | Scheme | Marks | AOs |
|---|---|---|---|
| **2** | $17+45+\frac{1}{3}\times 9 \quad [=65]$ | M1 | 2.2a |
| | $(7-8)\ \ \underline{\mathbf{14}}\ \ \underline{\text{or}}\ \ (16-20)\ \ \underline{\mathbf{5}}$<br>[Values may be seen in the table] | M1<br>A1 | 3.1a<br>1.1b |
| | Percentage of motorists is $\dfrac{\text{"65"}}{6+\text{"14"}+17+45+9+\text{"5"}}\times 100$ | M1 | 3.1b |
| | $= \underline{\mathbf{67.7\%}}$ | A1 | 1.1b |
| | | | **(5 marks)** |

| Part | Notes |
|---|---|
| | 1st M1 for a fully correct expression for the number of motorists in the interval |
| | 2nd M1 for clear use of frequency density in (4-6) or (13-15) cases to establish the fd scale. Then use of area to find frequency in one of the missing cases. |
| | 1st A1 for both correct values seen |
| | 3rd M1 for realising that total is required and attempting a correct expression for % |
| | 2nd A1 for awrt 67.7% |

HOME

# AS 2018

## Representations of Data

**4.** Helen is studying the daily mean wind speed for Camborne using the large data set from 1987. The data for one month are summarised in Table 1 below.

| **Windspeed** | n/a | 6 | 7 | 8 | 9 | 11 | 12 | 13 | 14 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Frequency** | 13 | 2 | 3 | 2 | 2 | 3 | 1 | 2 | 1 | 2 |

**Table 1**

(a) Calculate the mean for these data.

(1)

(b) Calculate the standard deviation for these data and state the units.

(2)

The means and standard deviations of the daily mean wind speed for the other months from the large data set for Camborne in 1987 are given in Table 2 below. The data are not in month order.

| **Month** | *A* | *B* | *C* | *D* | *E* |
|---|---|---|---|---|---|
| **Mean** | 7.58 | 8.26 | 8.57 | 8.57 | 11.57 |
| **Standard Deviation** | 2.93 | 3.89 | 3.46 | 3.87 | 4.64 |

**Table 2**

(c) Using your knowledge of the large data set, suggest, giving a reason, which month had a mean of 11.57

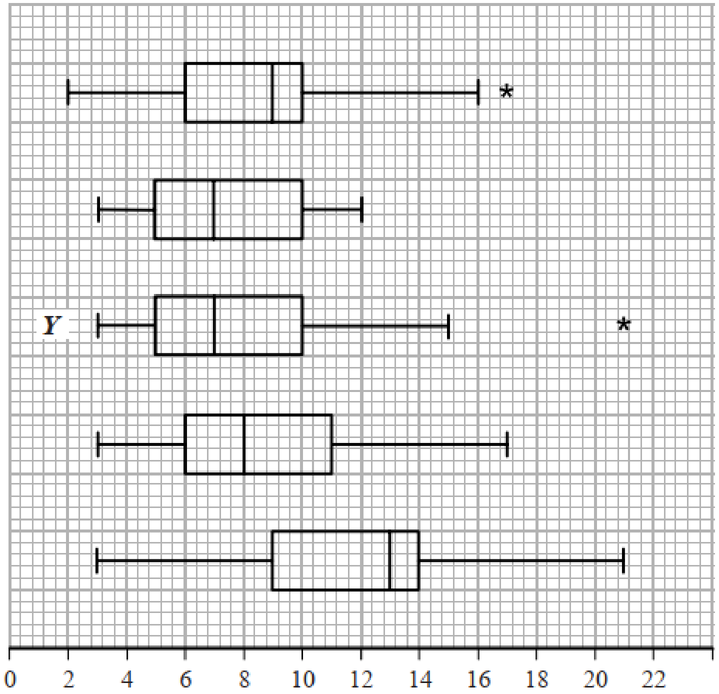
(2)

The data for these months are summarised in the box plots on the opposite page. They are not in month order or the same order as in Table 2.

(d) (i) State the meaning of the * symbol on some of the box plots.

(ii) Suggest, giving your reasons, which of the months in Table 2 is most likely to be summarised in the box plot marked *Y*.

(3)



HOME

# AS 2018

## Representations of Data

| Qu | Scheme | Marks | AO |
|---|---|---|---|
| 4 (a) | $\bar{x} = 10.2\,(2222\ldots)$ awrt **10.2** | B1<br>(1) | 1.1b |
| (b) | $\sigma_x = 3.17\,(20227\ldots)$ awrt **3.17** | B1ft | 1.1b |
| | Sight of "knots" or "kn" (condone knots/s etc) | B1<br>(2) | 1.2 |
| (c) | October ….. since | B1 | 2.2b |
| | it is windier in the autumn or month of the hurricane or latest month in the year | B1<br>(2) | 2.4 |
| (d)(i) | They represent outliers | B1 | 1.2 |
| (ii) | $Y$ has low median so expect lowish mean (but outlier so $> 7$)<br>and<br>$Y$ has big range/IQR or spread so expect larger st.dev | M1 | 2.4 |
| | Suggests $B$ | A1<br>(3) | 2.2b |
| | | **(8 marks)** | |

| | Notes |
|---|---|
| NB | $\bar{x} = \frac{184}{18}$ and $\sigma_x = \sqrt{\frac{2062}{18} - \bar{x}^2}$ |
| (a) | B1 for $\bar{x} = 10.2$ (allow exact fraction) [This is an LDS mark] |
| (b) | 1st B1ft allow 3.2 from a correct expr' accept $s = 3.26(3984\ldots)$ [ft use of n/a]<br>Treating n/a as 0 May see $n = 31$ or $\bar{x} = 5.9354\ldots$ which is B0 in (a) but here in (b) it gives $\sigma_x = 5.59(34\ldots)$ or $s = 5.6858\ldots$(awrt 5.69) and scores 1st B1<br>2nd B1 accept kn accept in (a) or (b) (allow nautical miles/hour)<br>[This is an LDS mark] |
| (c) | 1st B1 choosing October but accept September. [This is an LDS mark]<br>2nd B1 for stating that (Camborne) is windier in autumn/winter months<br>"because it is winter/autumn/windier/colder in "month" " Sep $\leqslant$ "month" $\leqslant$ Mar scores B1B1 for "month" = Sep or Oct and B0B1 for other months in range |
| (d)(i) | B1 for outlier or the idea of an extreme value allow "anomaly" |
| (ii) | M1 for a comment relating to location that mentions both median and mean and a comment relating to spread that mentions both range/IQR and standard deviation and leads to choosing $B$, $C$ or $D$<br>**Choosing $A$ or $E$ is M0**<br>Incorrect/false statements score M0 e.g. $Q_3 = (\text{mean} + \sigma)$ or identify $Q_2 = \text{mean}$ or $Y$ has small spread |
| ALT | **Use of outliers:** outlier is $(\text{mean} + 3\sigma)$ $(B = 19.9)$, $(C = 18.95)$, $(D = 20.2)$<br>Must see at least one of these values and compare to $Y$'s outlier[leads to $D$ or $B$ ]<br><br>A1 for suitable inference i.e. $B$ (accept $D$ or $B$ or $D$) M1 **must** be scored |

HOME

AS 2020

Representations of Data

1.

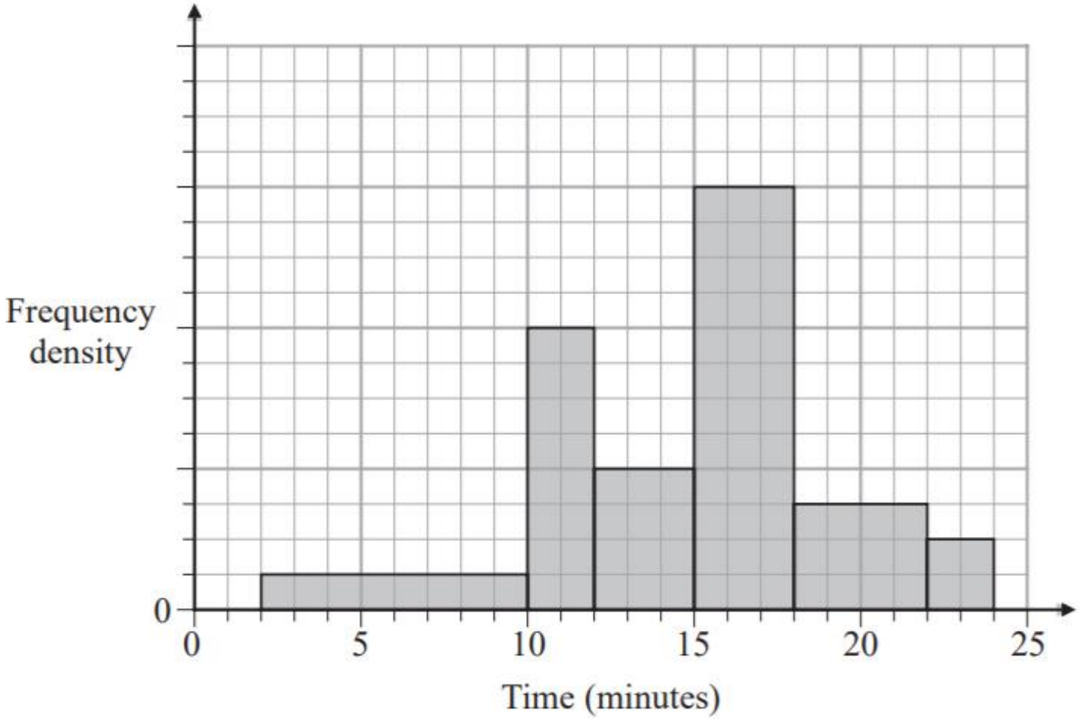


**Figure 1**

The histogram in Figure 1 shows the times taken to complete a crossword by a random sample of students.

The number of students who completed the crossword in more than 15 minutes is 78

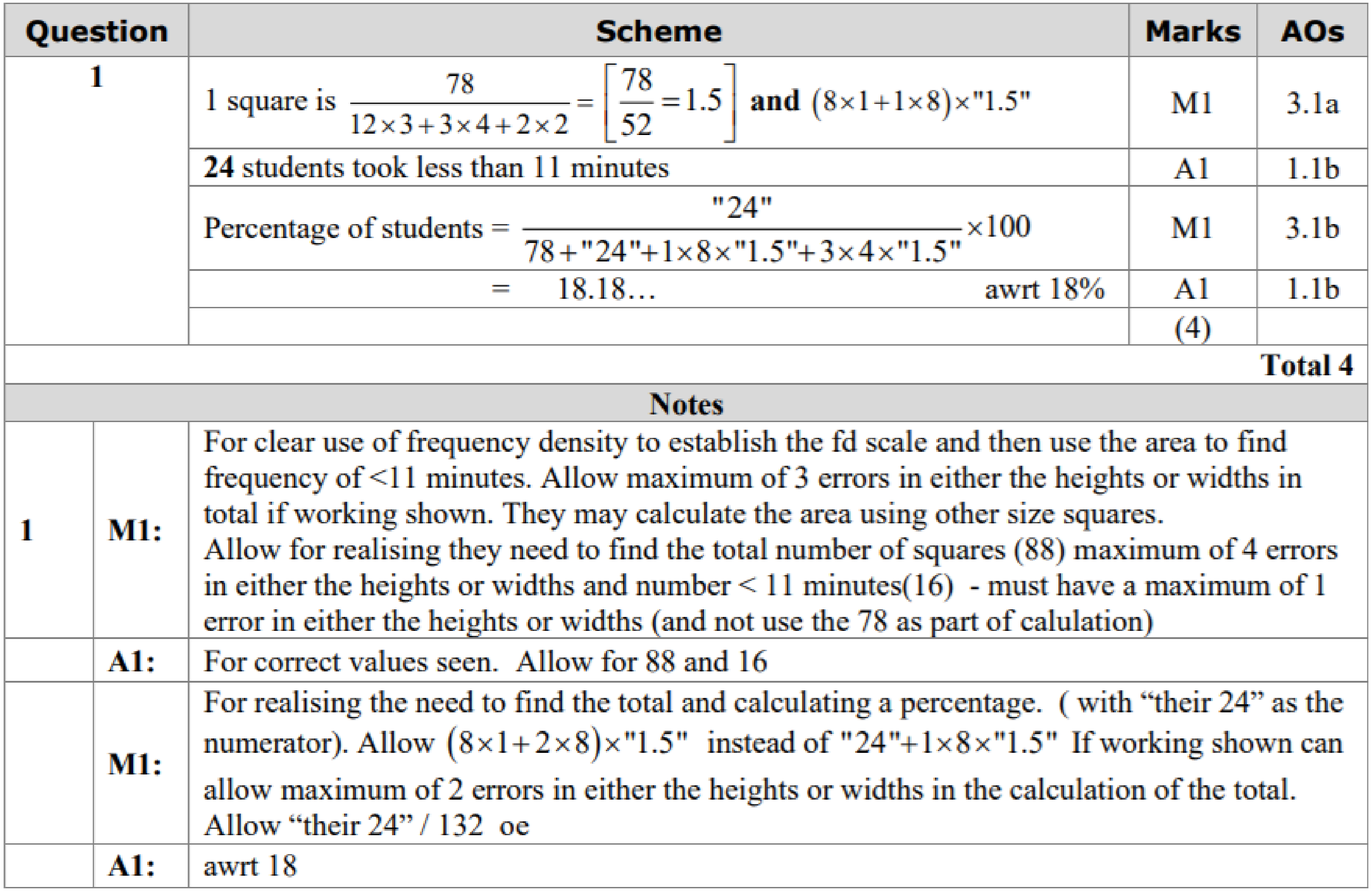
Estimate the percentage of students who took less than 11 minutes to complete the crossword.

**(4)**

HOME

# AS 2020

## Representations of Data



| Question | Scheme | Marks | AOs |
|---|---|---|---|
| 1 | 1 square is $\frac{78}{12\times3+3\times4+2\times2}=\left[\frac{78}{52}=1.5\right]$ **and** $(8\times1+1\times8)\times"1.5"$ | M1 | 3.1a |
| | **24** students took less than 11 minutes | A1 | 1.1b |
| | Percentage of students = $\frac{"24"}{78+"24"+1\times8\times"1.5"+3\times4\times"1.5"}\times100$ | M1 | 3.1b |
| | $=$ 18.18… awrt 18% | A1 | 1.1b |
| | | (4) | |
| | | | **Total 4** |

**Notes**

| | | |
|---|---|---|
| 1 | **M1:** | For clear use of frequency density to establish the fd scale and then use the area to find frequency of <11 minutes. Allow maximum of 3 errors in either the heights or widths in total if working shown. They may calculate the area using other size squares.<br>Allow for realising they need to find the total number of squares (88) maximum of 4 errors in either the heights or widths and number < 11 minutes(16) - must have a maximum of 1 error in either the heights or widths (and not use the 78 as part of calulation) |
| | **A1:** | For correct values seen. Allow for 88 and 16 |
| | **M1:** | For realising the need to find the total and calculating a percentage. ( with "their 24" as the numerator). Allow $(8\times1+2\times8)\times"1.5"$ instead of $"24"+1\times8\times"1.5"$ If working shown can allow maximum of 2 errors in either the heights or widths in the calculation of the total.<br>Allow "their 24" / 132 oe |
| | **A1:** | awrt 18 |

HOME

# AS 2021

## Representations of Data

**2.** The partially completed table and partially completed histogram give information about the ages of passengers on an airline.

There were no passengers aged 90 or over.

| **Age ($x$ years)** | $0 \leqslant x < 5$ | $5 \leqslant x < 20$ | $20 \leqslant x < 40$ | $40 \leqslant x < 65$ | $65 \leqslant x < 80$ | $80 \leqslant x < 90$ |
|---|---|---|---|---|---|---|
| **Frequency** | 5 | 45 | 90 | | | 1 |

(a) Complete the histogram. *(on the next slide)*

**(3)**

(b) Use linear interpolation to estimate the median age.

**(4)**

An outlier is defined as a value greater than $Q_3 + 1.5 \times$ interquartile range.

Given that $Q_1 = 27.3$ and $Q_3 = 58.9$

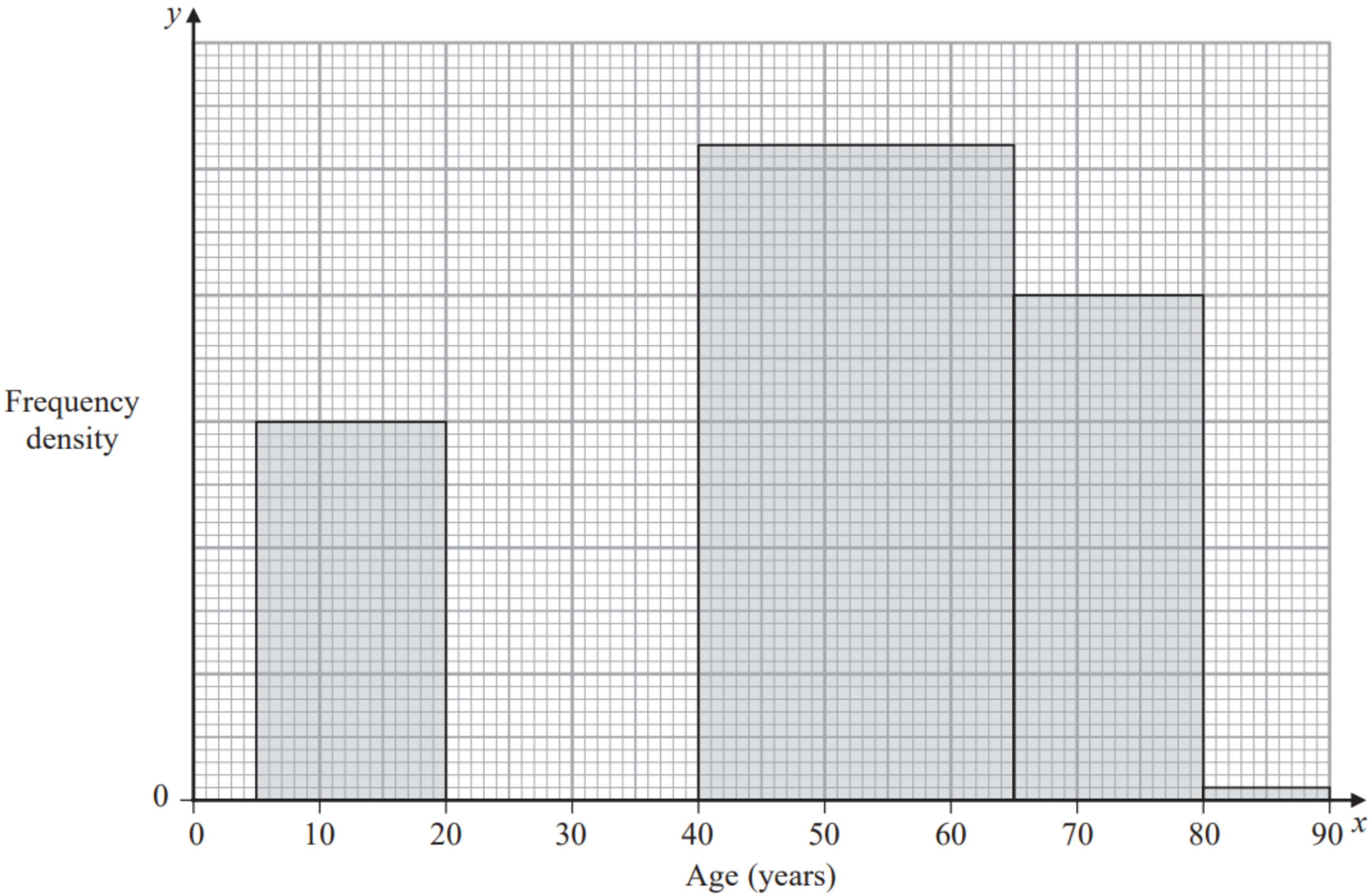
(c) determine, giving a reason, whether or not the oldest passenger could be considered as an outlier.

**(2)**

HOME

# AS 2021

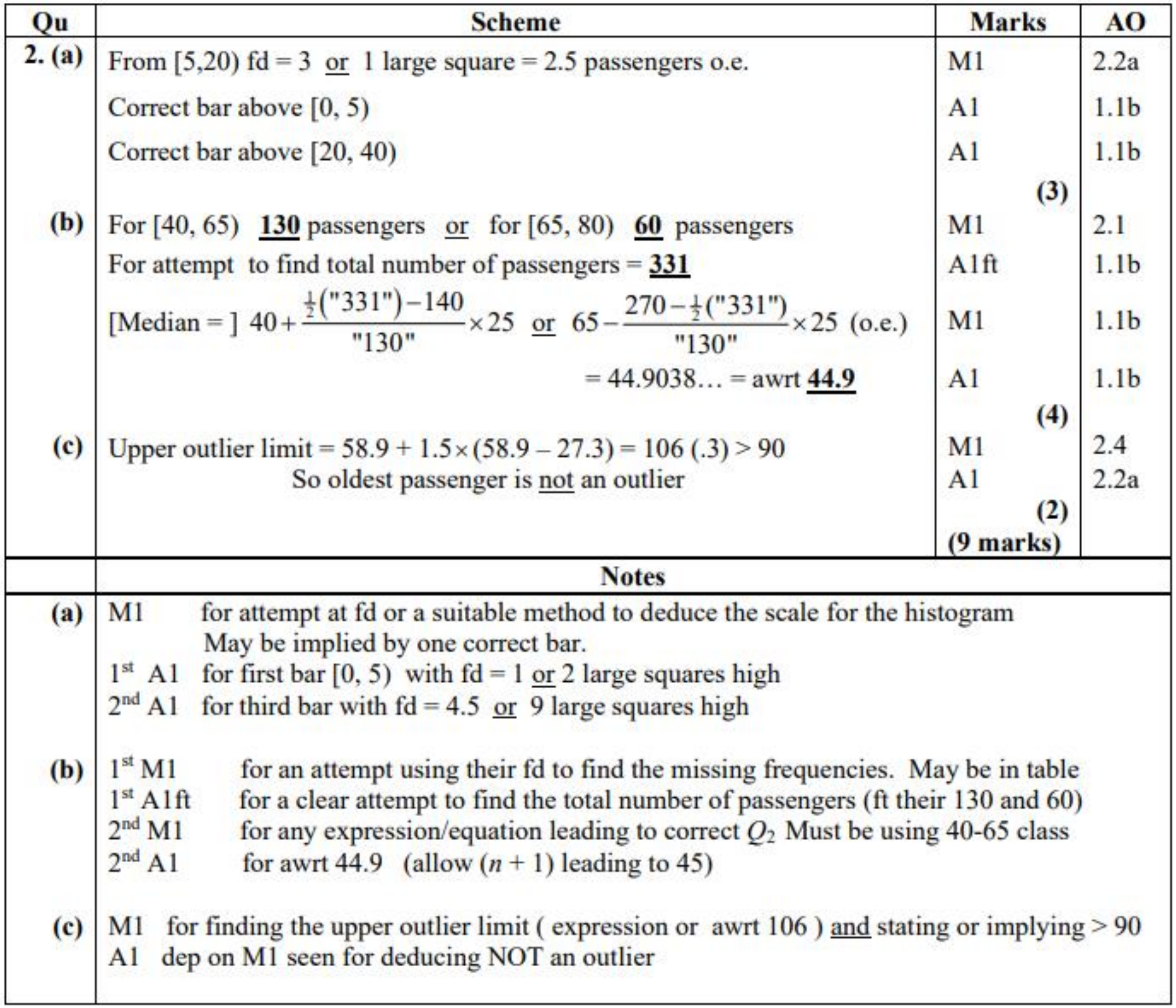
Representations of Data



HOME

# AS 2021

## Representations of Data



| Qu | Scheme | Marks | AO |
|---|---|---|---|
| 2. (a) | From [5,20) fd = 3 or 1 large square = 2.5 passengers o.e. | M1 | 2.2a |
| | Correct bar above [0, 5) | A1 | 1.1b |
| | Correct bar above [20, 40) | A1 | 1.1b |
| | | (3) | |
| (b) | For [40, 65) **130** passengers or for [65, 80) **60** passengers | M1 | 2.1 |
| | For attempt to find total number of passengers = **331** | A1ft | 1.1b |
| | [Median = ] $40+\frac{\frac{1}{2}(\text{"331"})-140}{\text{"130"}}\times 25$ or $65-\frac{270-\frac{1}{2}(\text{"331"})}{\text{"130"}}\times 25$ (o.e.) | M1 | 1.1b |
| | = 44.9038… = awrt **44.9** | A1 | 1.1b |
| | | (4) | |
| (c) | Upper outlier limit = $58.9 + 1.5\times(58.9 - 27.3) = 106\ (.3) > 90$<br>So oldest passenger is not an outlier | M1<br>A1 | 2.4<br>2.2a |
| | | (2) | |
| | | (9 marks) | |

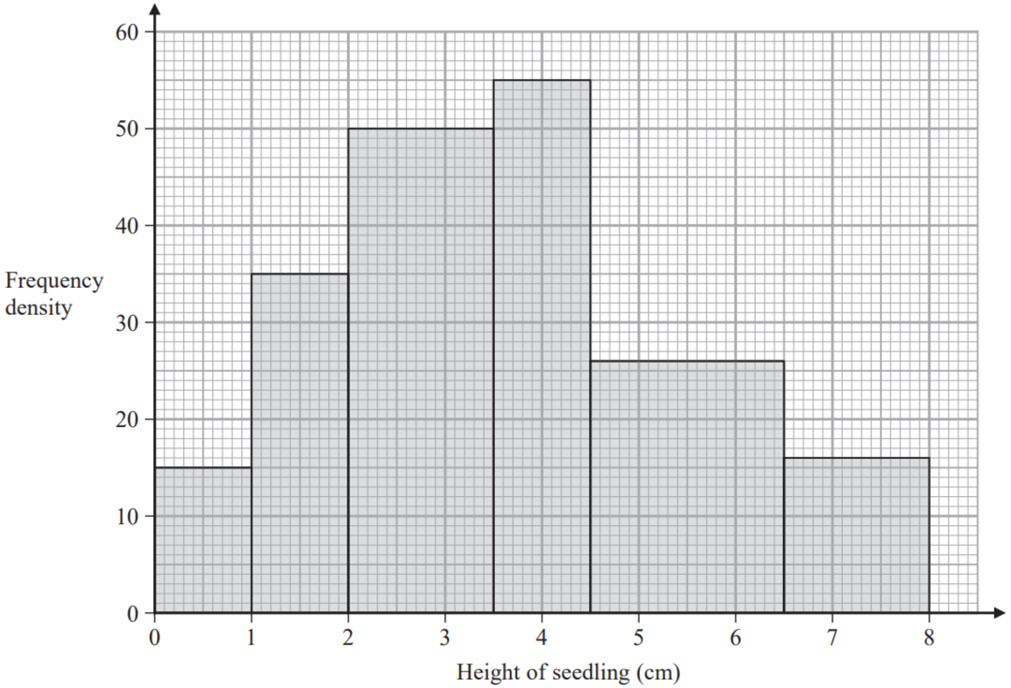
| | Notes |
|---|---|
| (a) | M1 for attempt at fd or a suitable method to deduce the scale for the histogram<br>May be implied by one correct bar.<br>1st A1 for first bar [0, 5) with fd = 1 or 2 large squares high<br>2nd A1 for third bar with fd = 4.5 or 9 large squares high |
| (b) | 1st M1 for an attempt using their fd to find the missing frequencies. May be in table<br>1st A1ft for a clear attempt to find the total number of passengers (ft their 130 and 60)<br>2nd M1 for any expression/equation leading to correct $Q_2$ Must be using 40-65 class<br>2nd A1 for awrt 44.9 (allow $(n + 1)$ leading to 45) |
| (c) | M1 for finding the upper outlier limit ( expression or awrt 106 ) and stating or implying > 90<br>A1 dep on M1 seen for deducing NOT an outlier |

HOME

# AS 2022

Representations of Data

**3.** The histogram summarises the heights of 256 seedlings two weeks after they were planted.



(a) Use linear interpolation to estimate the median height of the seedlings.

**(4)**

Chris decides to model the **frequency density** for these 256 seedlings by a curve with equation

$$y = kx(8 - x) \qquad 0 \leqslant x \leqslant 8$$

where $k$ is a constant.

(b) Find the value of $k$

**(3)**

Using this model,

(c) write down the median height of the seedlings.

**(1)**

HOME

# AS 2022

## Representations of Data

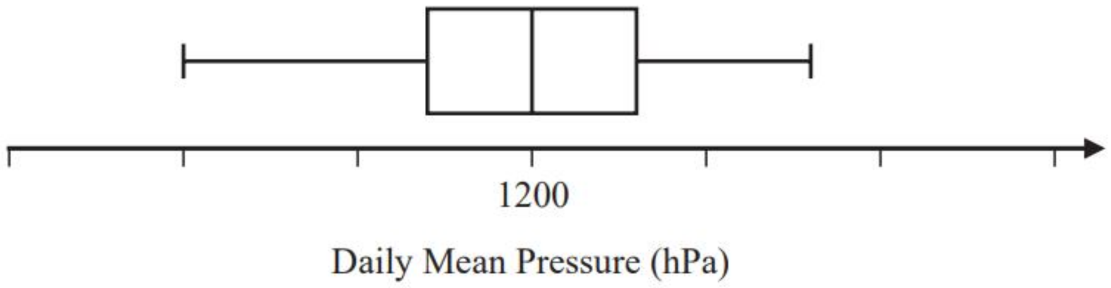
<table>
<tr><th>Qu</th><th>Scheme</th><th>Mark</th><th>AO</th></tr>
<tr><td>3. (a)</td><td><table><tr><th>Class</th><th>Frequency</th><th>Cum. Frequency</th></tr><tr><td>0 – 1</td><td>15</td><td>15</td></tr><tr><td>1 – 2</td><td>35</td><td>50</td></tr><tr><td>2 – 3.5</td><td>75</td><td>125</td></tr><tr><td>3.5 – 4.5</td><td>55</td><td>180</td></tr></table></td><td>M1<br>A1</td><td>2.1<br>1.1b</td></tr>
<tr><td></td><td>$[Q_2 =](3.5)+\frac{\frac{256}{2}-\text{"125"}}{\text{"55"}}\times(4.5-3.5)$ <u>or</u> $(4.5)-\frac{\text{"180"}-\frac{256}{2}}{\text{"55"}}\times 1$</td><td>M1</td><td>2.1</td></tr>
<tr><td></td><td>$= 3.5545\ldots\ldots$ awrt <b><u>3.55</u></b></td><td>A1<br>(4)</td><td>1.1b</td></tr>
<tr><td>(b)</td><td>Need area under curve to be 256 so $\int_{(0)}^{(8)} kx(8-x)\,\mathrm{d}x = 256$</td><td>M1</td><td>3.1a</td></tr>
<tr><td></td><td>$k\left[4x^2-\frac{x^3}{3}\right]_{(0)}^{(8)} = 256$</td><td>M1</td><td>1.1b</td></tr>
<tr><td></td><td>$\left\{k\left[4\times8^2-\frac{8}{3}\times8^2\right]=256\Rightarrow\right\}$ $\underline{\boldsymbol{k=3}}$</td><td>A1<br>(3)</td><td>1.1b</td></tr>
<tr><td>(c)</td><td>[By symmetry median = ] <b><u>4</u></b></td><td>B1<br>(1)</td><td>2.2a</td></tr>
<tr><td></td><td></td><td>(8 marks)</td><td></td></tr>
</table>

HOME

4. Jiang is studying the variable Daily Mean Pressure from the large data set.

He drew the following box and whisker plot for these data for one of the months for one location using a linear scale but

- he failed to label all the values on the scale
- he gave an incorrect value for the median



Using your knowledge of the large data set, suggest a suitable value for

(a) the median,

**(1)**

(b) the range.

**(1)**

(*You are not expected to have memorised values from the large data set. The question is simply looking for sensible answers.*)

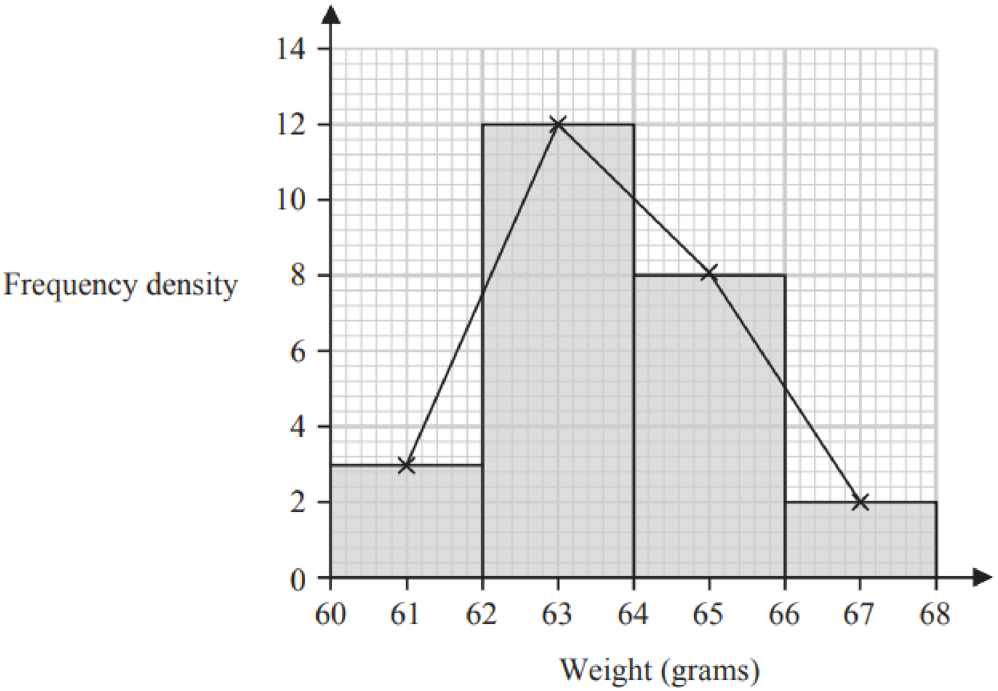
# AS 2022

Representations of Data

| | | | |
|---|---|---|---|
| **4. (a)** | Accept 990 to 1030 inclusive | B1<br>**(1)** | 1.1b |
| **(b)** | Any range between 10 and 50 inclusive | B1<br>**(1)** | 1.1b |
| | | **(2 marks)** | |
| | **Notes** | | |
| **(a)** | B1 (Median pressures usually around 1000~1020) [LDS mark] | | |
| **(b)** | B1 Any answer in this range [LDS mark]<br>Allow answers in the form $a \sim b$ where $\lvert b - a \rvert$ is between 10 and 50<br>Also allow the case where both $a$ and $b$ are in [10, 50] | | |

HOME

**1.** The histogram and its frequency polygon below give information about the weights, in grams, of 50 plums.



(a) Show that an estimate for the mean weight of the 50 plums is 63.72 grams.

**(2)**

(b) Calculate an estimate for the standard deviation of the 50 plums.

**(2)**

Later it was discovered that the scales used to weigh the plums were broken.

Each plum actually weighs 5 grams less than originally thought.

(c) State the effect this will have on the estimate of the standard deviation in part (b). Give a reason for your answer.

**(1)**

**AS 2023**

Representations of Data

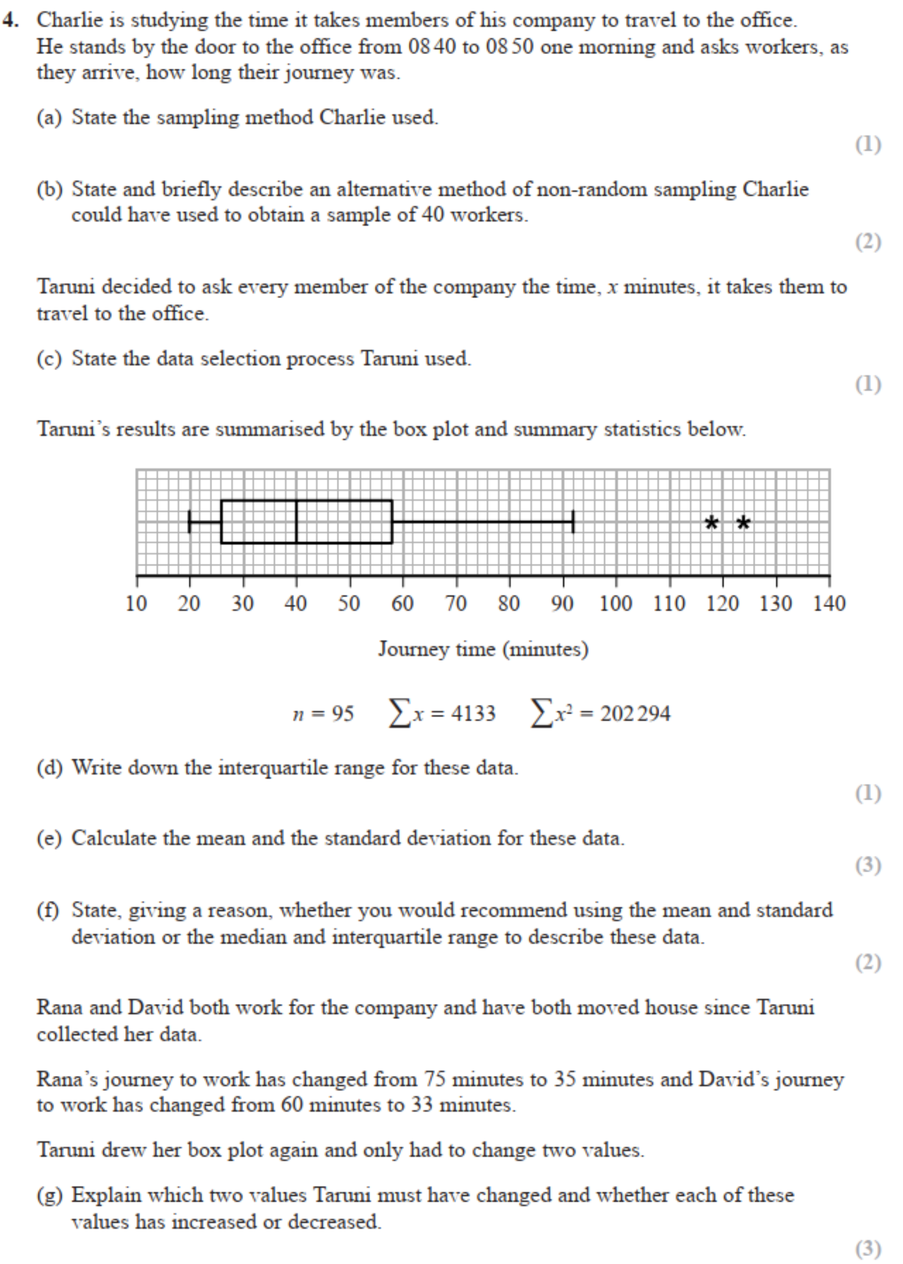
| | | | |
|---|---|---|---|
| **1(a)** | $61\times(2\times3),\quad 63\times(2\times12),\quad 65\times(2\times8),\quad 67\times(2\times2)$ | M1 | 2.1 |
| | $\frac{61\times(2\times3)+63\times(2\times12)+65\times(2\times8)+67\times(2\times2)}{50}=63.72*$ | A1*cso | 1.1b |
| | | **(2)** | |
| **(b)** | $\sqrt{\frac{61^2\times6+63^2\times24+65^2\times16+67^2\times4}{50}-63.72^2}$ | M1 | 1.1b |
| | $=\sqrt{2.5216}=1.58795...$ $\quad$ = awrt **<u>1.59</u>** | A1 | 1.1b |
| | | **(2)** | |
| **(c)** | No effect (oe) since…e.g.<br>• since addition/subtraction does not affect the standard deviation (only multiplication and division do)<br>• the weights will have the same spread<br>• the distance of each weight from the mean will not have changed<br>• they all change by the same amount | B1 | 2.4 |
| | | **(1)** | |

**(5 marks)**

HOME

# A2 2018

# Representations of Data



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.

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

(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.

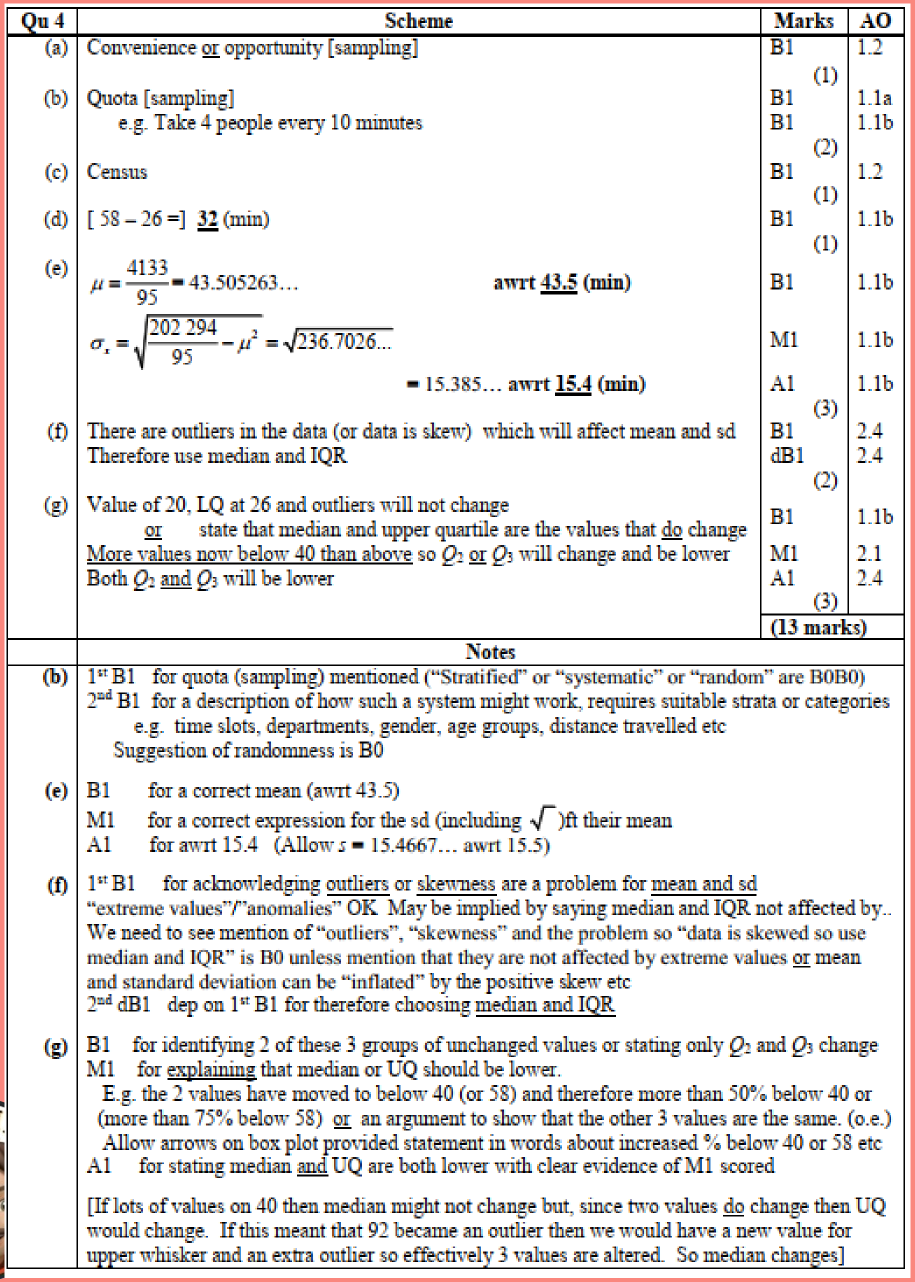
(g) Explain which two values Taruni must have changed and whether each of these values has increased or decreased.

(3)

HOME

# A2 2018

## Representations of Data



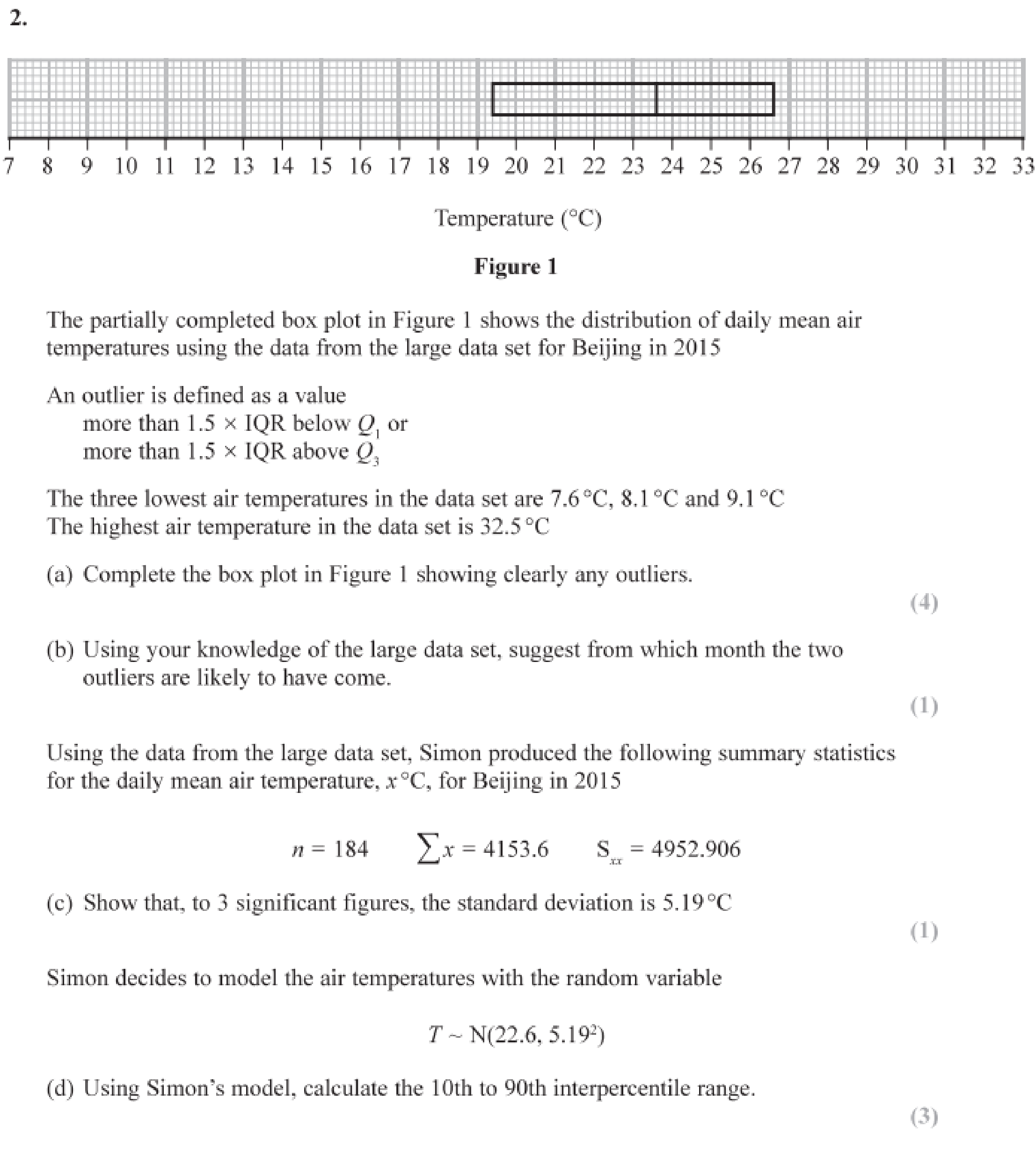
| Qu 4 | Scheme | Marks | AO |
|---|---|---|---|
| (a) | Convenience or opportunity [sampling] | B1<br>(1) | 1.2 |
| (b) | Quota [sampling] | B1 | 1.1a |
| | e.g. Take 4 people every 10 minutes | B1<br>(2) | 1.1b |
| (c) | Census | B1<br>(1) | 1.2 |
| (d) | [ 58 – 26 =] **32** (min) | B1<br>(1) | 1.1b |
| (e) | $\mu = \frac{4133}{95} = 43.505263\ldots$    **awrt 43.5 (min)** | B1 | 1.1b |
| | $\sigma_x = \sqrt{\frac{202\,294}{95} - \mu^2} = \sqrt{236.7026\ldots}$ | M1 | 1.1b |
| | $= 15.385\ldots$ **awrt 15.4 (min)** | A1<br>(3) | 1.1b |
| (f) | There are outliers in the data (or data is skew) which will affect mean and sd | B1 | 2.4 |
| | Therefore use median and IQR | dB1<br>(2) | 2.4 |
| (g) | Value of 20, LQ at 26 and outliers will not change<br>or state that median and upper quartile are the values that do change | B1 | 1.1b |
| | More values now below 40 than above so $Q_2$ or $Q_3$ will change and be lower | M1 | 2.1 |
| | Both $Q_2$ and $Q_3$ will be lower | A1<br>(3) | 2.4 |
| | | **(13 marks)** | |

| | Notes |
|---|---|
| **(b)** | 1st B1 for quota (sampling) mentioned ("Stratified" or "systematic" or "random" are B0B0)<br>2nd B1 for a description of how such a system might work, requires suitable strata or categories<br>e.g. time slots, departments, gender, age groups, distance travelled etc<br>Suggestion of randomness is B0 |
| **(e)** | B1 for a correct mean (awrt 43.5)<br>M1 for a correct expression for the sd (including $\sqrt{\ }$)ft their mean<br>A1 for awrt 15.4 (Allow $s = 15.4667\ldots$ awrt 15.5) |
| **(f)** | 1st B1 for acknowledging outliers or skewness are a problem for mean and sd<br>"extreme values"/"anomalies" OK May be implied by saying median and IQR not affected by..<br>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 or mean and standard deviation can be "inflated" by the positive skew etc<br>2nd dB1 dep on 1st B1 for therefore choosing median and IQR |
| **(g)** | B1 for identifying 2 of these 3 groups of unchanged values or stating only $Q_2$ and $Q_3$ change<br>M1 for explaining that median or UQ should be lower.<br>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) or an argument to show that the other 3 values are the same. (o.e.)<br>Allow arrows on box plot provided statement in words about increased % below 40 or 58 etc<br>A1 for stating median and UQ are both lower with clear evidence of M1 scored<br>[If lots of values on 40 then median might not change but, since two values do 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] |

HOME

# A2 2019

Representations of Data



**2.**

7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33
Temperature (°C)

**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 × IQR below $Q_1$ or
more than 1.5 × 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, $x$ °C, for Beijing in 2015

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

(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

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

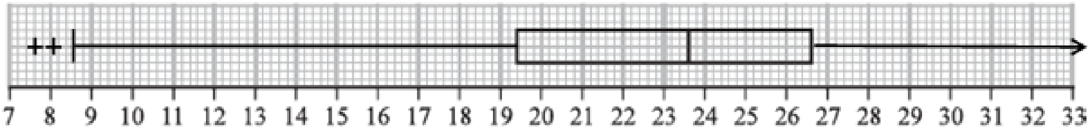
(d) Using Simon's model, calculate the 10th to 90th interpercentile range.

(3)

HOME

# A2 2019

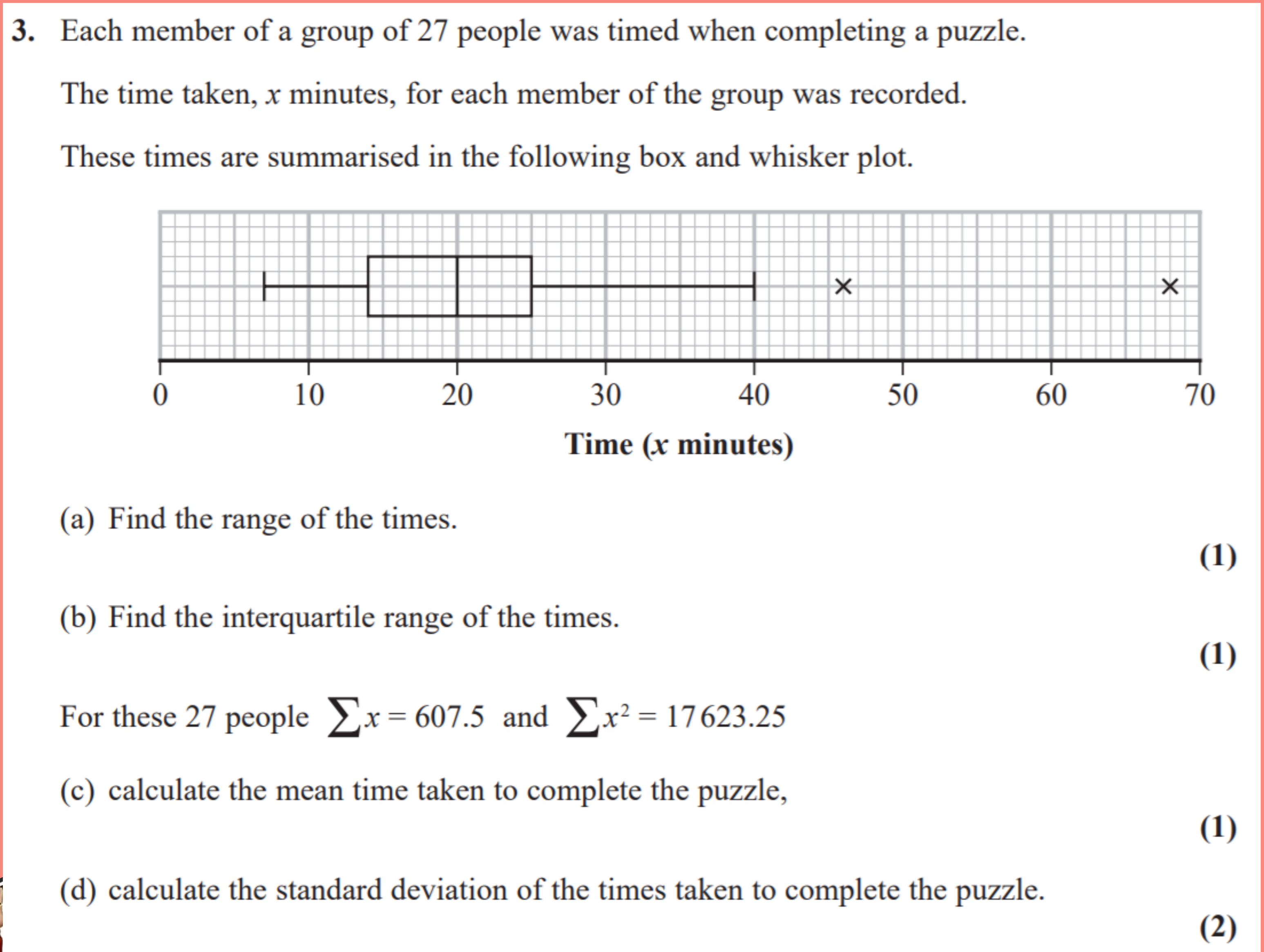
## Representations of Data

| | | | |
|---|---|---|---|
| (a) | IQR = 2.6. – 19.4 = 7.2 | B1 | This mark is given for finding the interquartile range |
| | 19.4 – (1.5 × 7.2) = 8.6<br>19.4 + (1.5 × 7.2) = 37.4 | M1 | This mark is given for a method find the values for the whiskers of the boxplot |
| |  | | |
| | | A1 | This mark is given for plotting the correct whisker (8.6) on the boxplot |
| | | A1 | This mark is given for plotting the two correct outliers 7.6 °C and 8.1 °C |
| (b) | October (since it is the month with the coldest temperatures between May and October in Beijing) | B1 | This mark is given for a correct suggestion with a supporting reason. |
| (c) | $\sigma = \sqrt{\frac{S_{xx}}{n}} = \sqrt{\frac{4952.906}{184}} = \sqrt{26.92} = 5.19$ | B1 | This mark is given for showing the calculation for the standard deviation to three significant figures |
| (d) | z = (±) 1.2816 | B1 | This mark is given for identifying the z-value for the 10th and 90th percentiles (from tables or calculator) |
| | $2 \times z \times 5.19$ | M1 | This mark is given for a method to find the interpercentile range between the 10th and 90th value |
| | = 13.303 | A1 | This mark is given for finding a correct interpercentile range between the 10th and 90th value |

HOME

# A2 2020 (Part 1)

Representations of Data



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.

0
10
20
30
40
50
60
70
Time ($x$ minutes)

(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)**

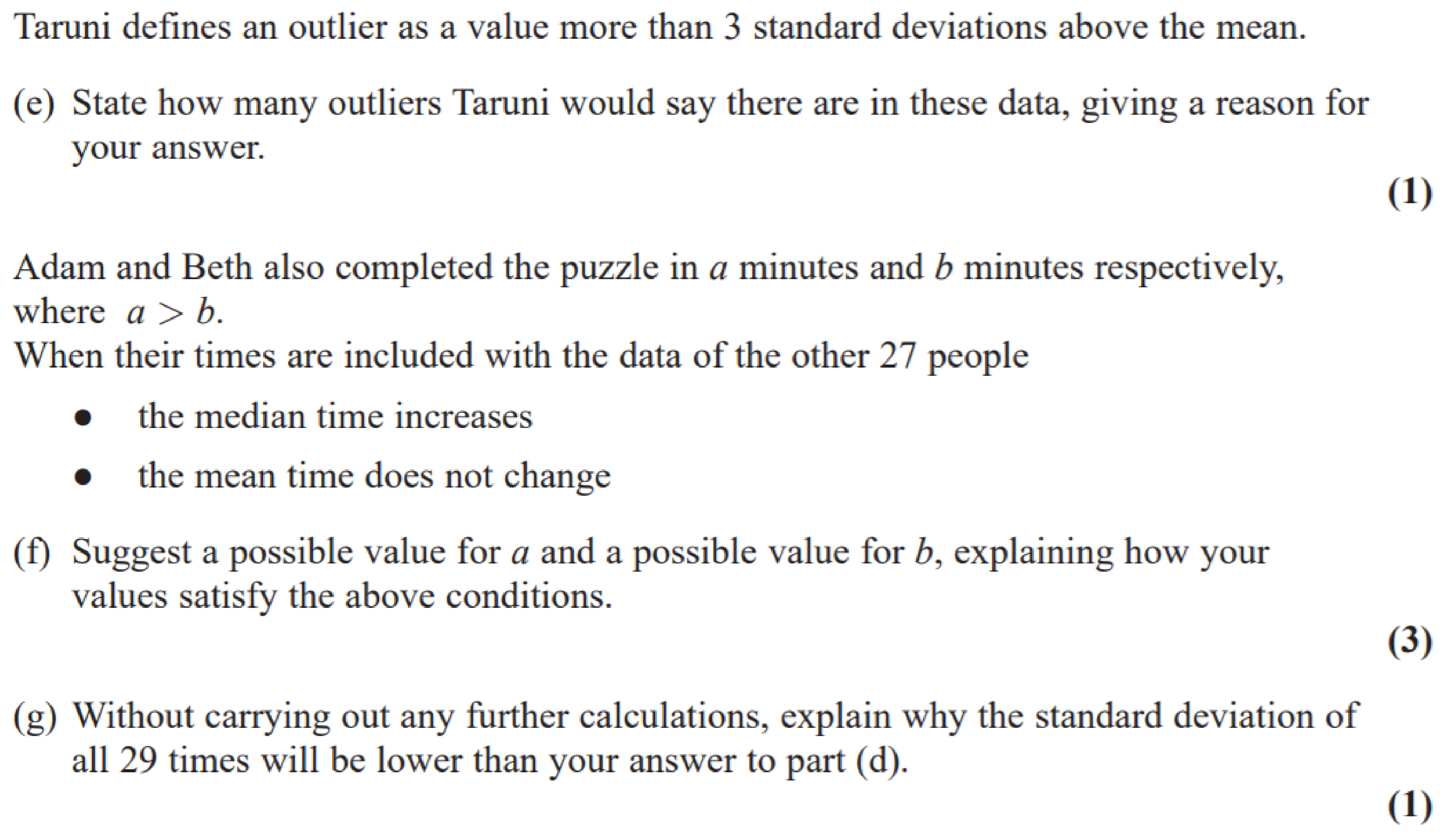
(d) calculate the standard deviation of the times taken to complete the puzzle.

**(2)**

HOME

# A2 2020 (Part 2)

Representations of Data



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)**

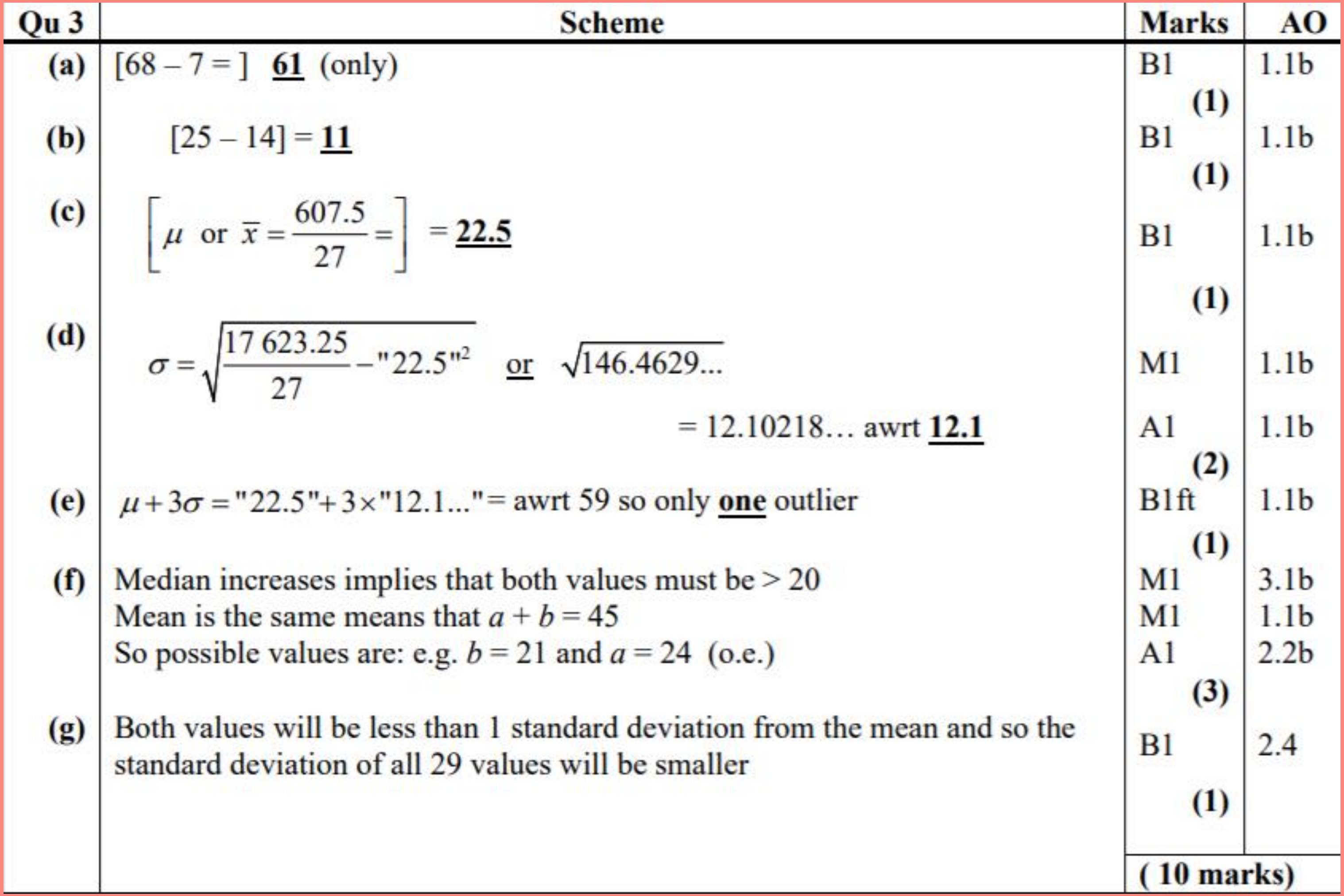
(g) Without carrying out any further calculations, explain why the standard deviation of all 29 times will be lower than your answer to part (d).

**(1)**

HOME

# A2 2020

Representations of Data

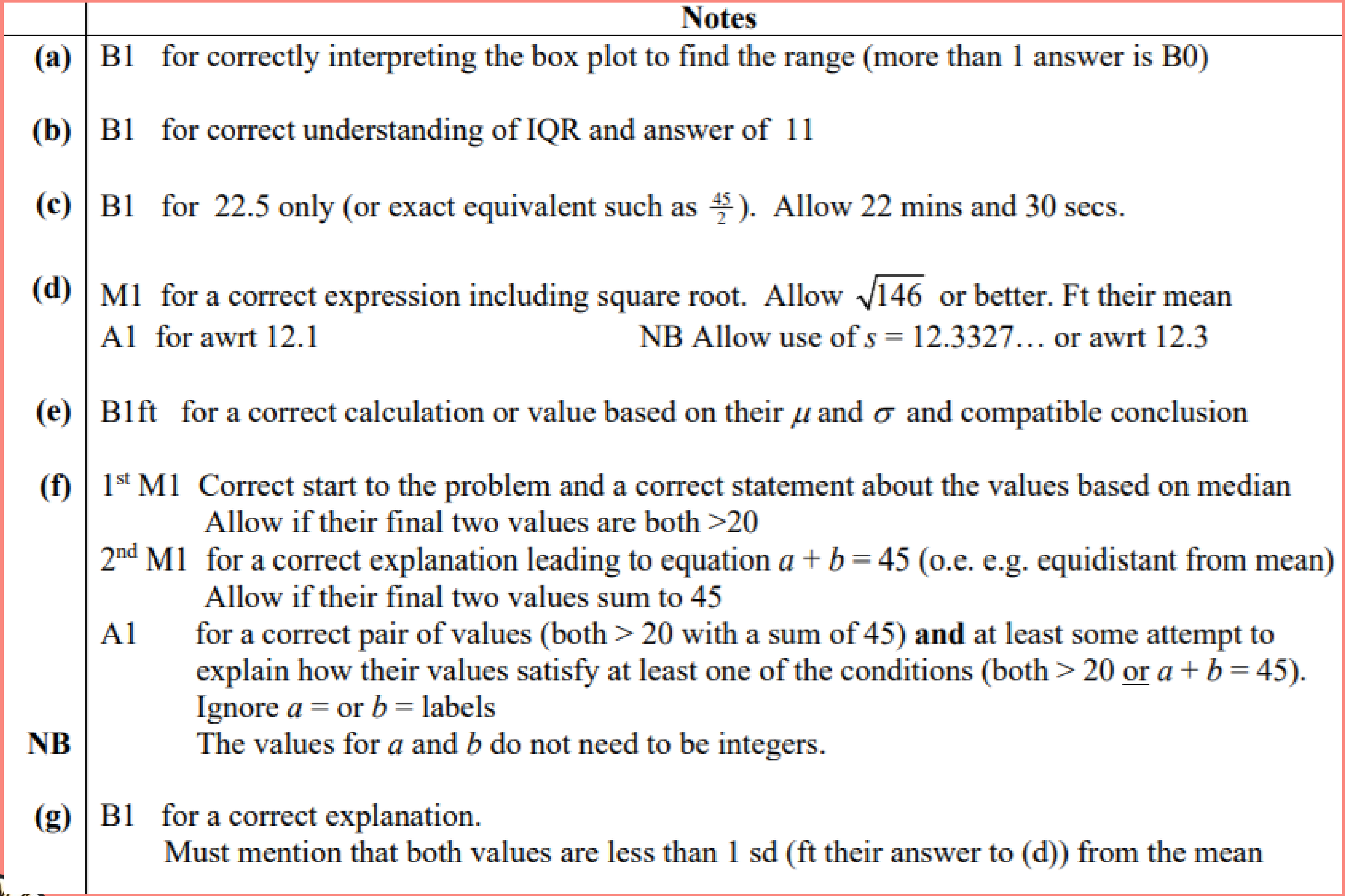


| Qu 3 | Scheme | Marks | AO |
|---|---|---|---|
| (a) | [68 – 7 = ] **61** (only) | B1<br>**(1)** | 1.1b |
| (b) | [25 – 14] = **11** | B1<br>**(1)** | 1.1b |
| (c) | $\left[\mu \text{ or } \bar{x} = \frac{607.5}{27} = \right] = \underline{\mathbf{22.5}}$ | B1<br>**(1)** | 1.1b |
| (d) | $\sigma = \sqrt{\frac{17\,623.25}{27} - \text{"}22.5\text{"}^2}$ or $\sqrt{146.4629...}$ | M1 | 1.1b |
|  | = 12.10218… awrt **12.1** | A1<br>**(2)** | 1.1b |
| (e) | $\mu + 3\sigma = \text{"}22.5\text{"} + 3 \times \text{"}12.1...\text{"} =$ awrt 59 so only **one** outlier | B1ft<br>**(1)** | 1.1b |
| (f) | Median increases implies that both values must be > 20 | M1 | 3.1b |
|  | Mean is the same means that $a + b = 45$ | M1 | 1.1b |
|  | So possible values are: e.g. $b = 21$ and $a = 24$ (o.e.) | A1<br>**(3)** | 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<br>**(1)** | 2.4 |
|  |  | **( 10 marks)** |  |

HOME

A2 2020

# Representations of Data



| | 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<br>A1 for awrt 12.1 NB Allow use of $s = 12.3327\ldots$ or awrt 12.3 |
| (e) | B1ft for a correct calculation or value based on their $\mu$ and $\sigma$ and compatible conclusion |
| (f) | 1st M1 Correct start to the problem and a correct statement about the values based on median<br>Allow if their final two values are both >20<br>2nd M1 for a correct explanation leading to equation $a + b = 45$ (o.e. e.g. equidistant from mean)<br>Allow if their final two values sum to 45<br>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 or $a + b = 45$).<br>Ignore $a =$ or $b =$ labels |
| NB | The values for $a$ and $b$ do not need to be integers. |
| (g) | B1 for a correct explanation.<br>Must mention that both values are less than 1 sd (ft their answer to (d)) from the mean |

# A2 2023 (Part 1)

Representations of Data

**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)**

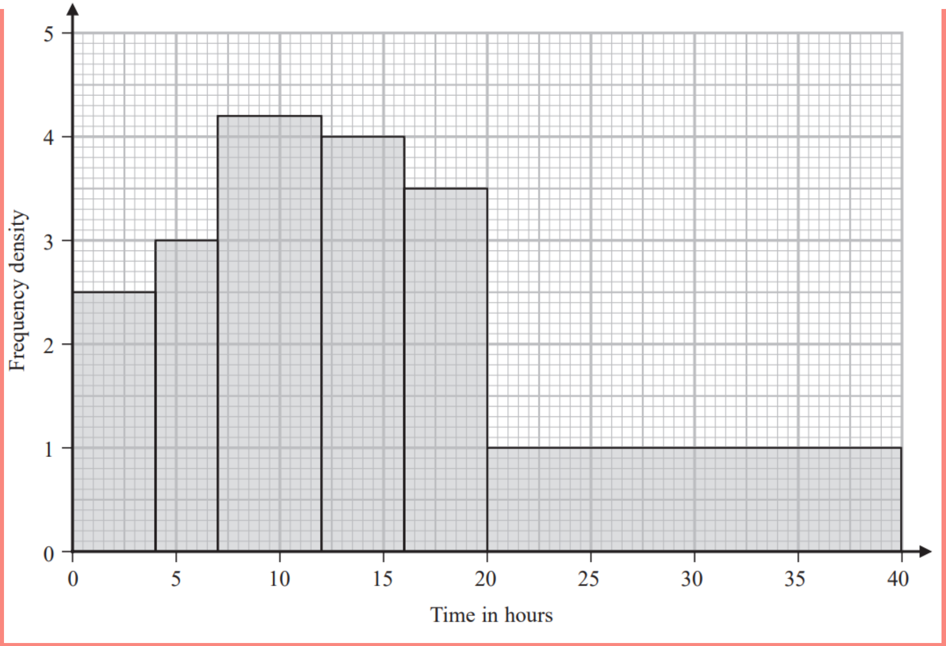
For these 90 patients the time spent in hospital following the operation had

- 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.

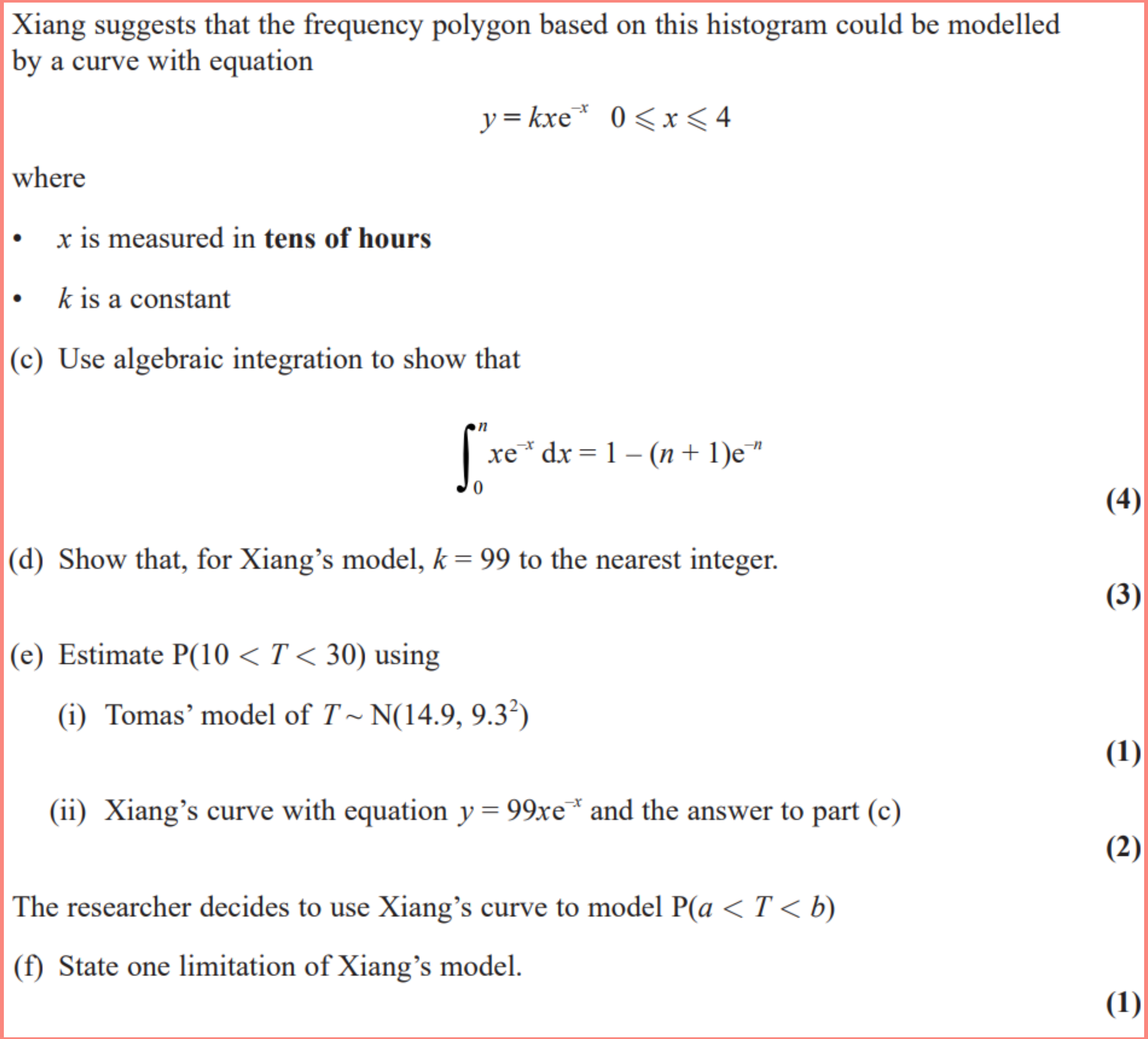
**(1)**



HOME

# A2 2023 (Part 2)

Representations of Data



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

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

where

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

(c) Use algebraic integration to show that

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

**(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 $P(a < T < b)$

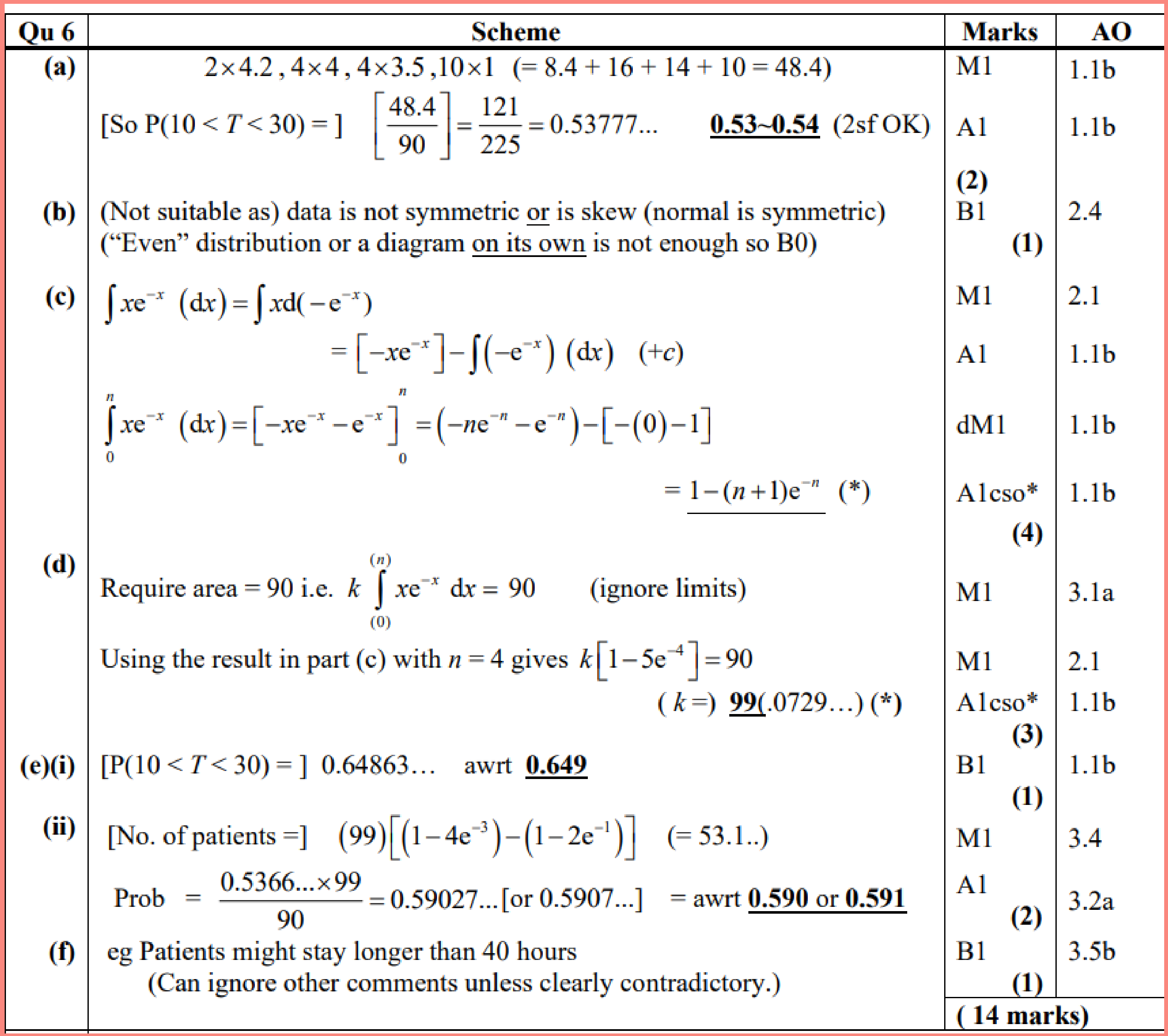
(f) State one limitation of Xiang's model.

**(1)**

HOME

# A2 2023

Representations of Data



| Qu 6 | Scheme | Marks | AO |
|---|---|---|---|
| (a) | $2\times4.2\,,4\times4\,,4\times3.5\,,10\times1$ $(=8.4+16+14+10=48.4)$ | M1 | 1.1b |
| | [So $P(10<T<30)=$ ] $\left[\frac{48.4}{90}\right]=\frac{121}{225}=0.53777...$ **<u>0.53~0.54</u>** (2sf OK) | A1 | 1.1b |
| | | **(2)** | |
| (b) | (Not suitable as) data is not symmetric <u>or</u> is skew (normal is symmetric)<br>("Even" distribution or a diagram <u>on its own</u> is not enough so B0) | B1<br>**(1)** | 2.4 |
| (c) | $\int xe^{-x}\ (dx)=\int xd(-e^{-x})$ | M1 | 2.1 |
| | $=\left[-xe^{-x}\right]-\int\left(-e^{-x}\right)(dx)\quad(+c)$ | A1 | 1.1b |
| | $\int_0^n xe^{-x}\ (dx)=\left[-xe^{-x}-e^{-x}\right]_0^n=\left(-ne^{-n}-e^{-n}\right)-\left[-(0)-1\right]$ | dM1 | 1.1b |
| | $=\underline{1-(n+1)e^{-n}}\quad(*)$ | A1cso* | 1.1b |
| | | **(4)** | |
| (d) | Require area = 90 i.e. $k\int_{(0)}^{(n)} xe^{-x}\ dx=90$ (ignore limits) | M1 | 3.1a |
| | Using the result in part (c) with $n=4$ gives $k\left[1-5e^{-4}\right]=90$ | M1 | 2.1 |
| | ( $k=$ ) **<u>99</u>**(.0729…) (*) | A1cso*<br>**(3)** | 1.1b |
| (e)(i) | [$P(10<T<30)=$ ] 0.64863… awrt **<u>0.649</u>** | B1<br>**(1)** | 1.1b |
| (ii) | [No. of patients =] $(99)\left[\left(1-4e^{-3}\right)-\left(1-2e^{-1}\right)\right]$ (= 53.1..) | M1 | 3.4 |
| | Prob $=\frac{0.5366...\times99}{90}=0.59027...$[or 0.5907...] = awrt **<u>0.590 or 0.591</u>** | A1<br>**(2)** | 3.2a |
| (f) | eg Patients might stay longer than 40 hours<br>(Can ignore other comments unless clearly contradictory.) | B1<br>**(1)** | 3.5b |
| | | **( 14 marks)** | |

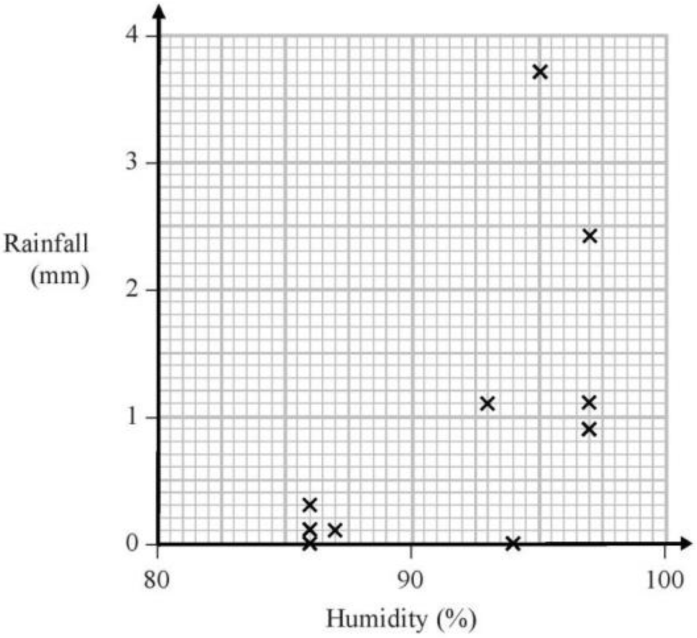
HOME

# Correlation

# AS SAMs

## Correlation

Sara decided to exclude this day's reading and drew the following scatter diagram for the remaining 10 days' values of $r$ and $h$.



(c) Give an interpretation of the correlation between rainfall and humidity.

**(1)**

The equation of the regression line of $r$ on $h$ for these 10 days is $r = -12.8 + 0.15h$

(d) Give an interpretation of the gradient of this regression line.

**(1)**

(e) (i) Comment on the suitability of Sara's sampling method for this study.

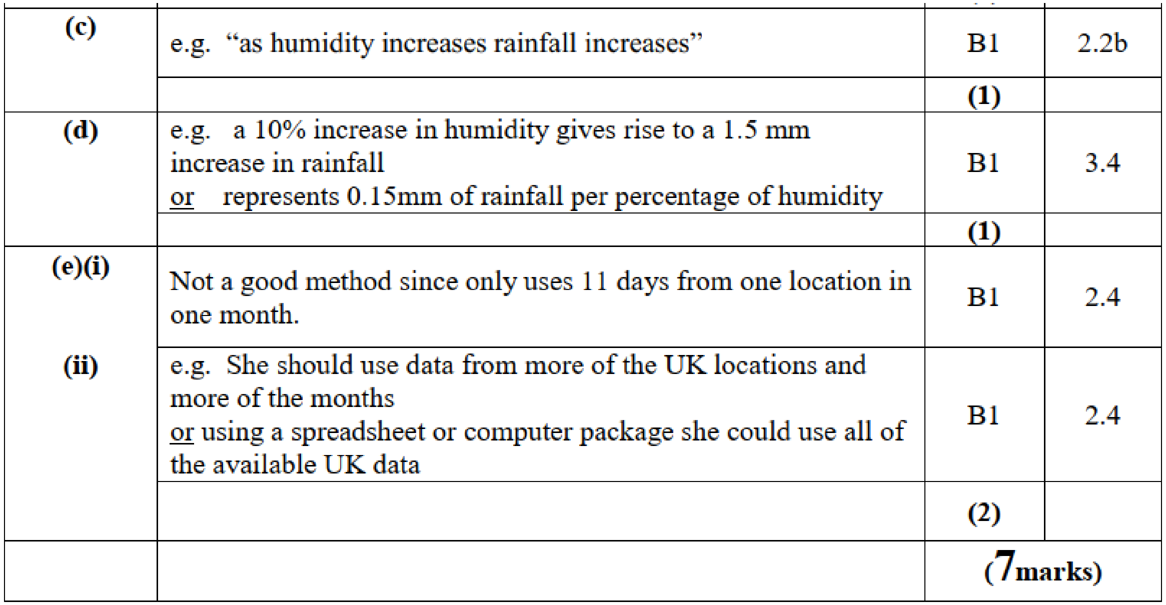
(ii) Suggest how Sara could make better use of the large data set for her study.

**(2)**

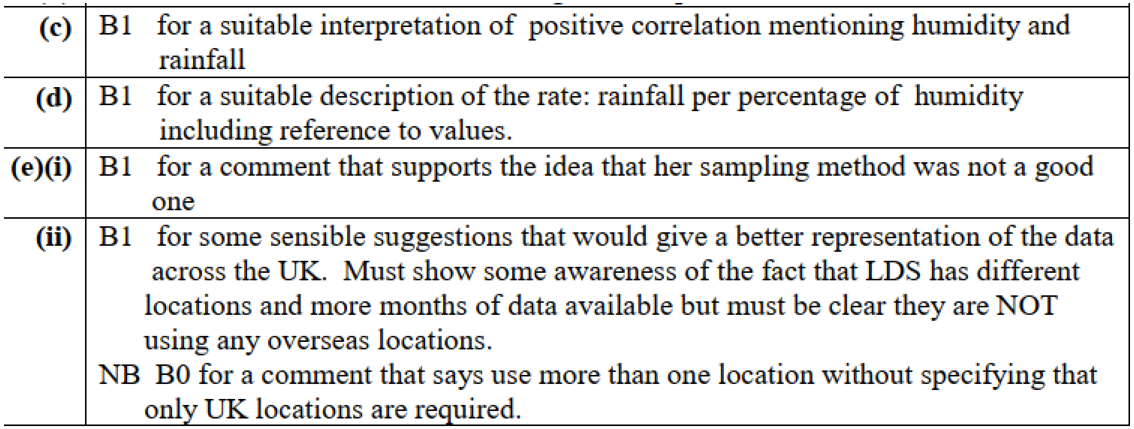
HOME

# AS SAMs

## Correlation



| | | | |
|---|---|---|---|
| **(c)** | e.g. "as humidity increases rainfall increases" | B1 | 2.2b |
| | | **(1)** | |
| **(d)** | e.g. a 10% increase in humidity gives rise to a 1.5 mm increase in rainfall<br>or represents 0.15mm of rainfall per percentage of humidity | B1 | 3.4 |
| | | **(1)** | |
| **(e)(i)** | Not a good method since only uses 11 days from one location in one month. | B1 | 2.4 |
| **(ii)** | e.g. She should use data from more of the UK locations and more of the months<br>or using a spreadsheet or computer package she could use all of the available UK data | B1 | 2.4 |
| | | **(2)** | |
| | | **(7marks)** | |



| | |
|---|---|
| **(c)** | B1 for a suitable interpretation of positive correlation mentioning humidity and rainfall |
| **(d)** | B1 for a suitable description of the rate: rainfall per percentage of humidity including reference to values. |
| **(e)(i)** | B1 for a comment that supports the idea that her sampling method was not a good one |
| **(ii)** | B1 for some sensible suggestions that would give a better representation of the data across the UK. Must show some awareness of the fact that LDS has different locations and more months of data available but must be clear they are NOT using any overseas locations.<br>NB B0 for a comment that says use more than one location without specifying that only UK locations are required. |

HOME

## AS 2018

Correlation

1. A company is introducing a job evaluation scheme. Points ($x$) will be awarded to each job based on the qualifications and skills needed and the level of responsibility. Pay (£$y$) will then be allocated to each job according to the number of points awarded.

Before the scheme is introduced, a random sample of 8 employees was taken and the linear regression equation of pay on points was $y = 4.5x - 47$

(a) Describe the correlation between points and pay.

**(1)**

(b) Give an interpretation of the gradient of this regression line.

**(1)**

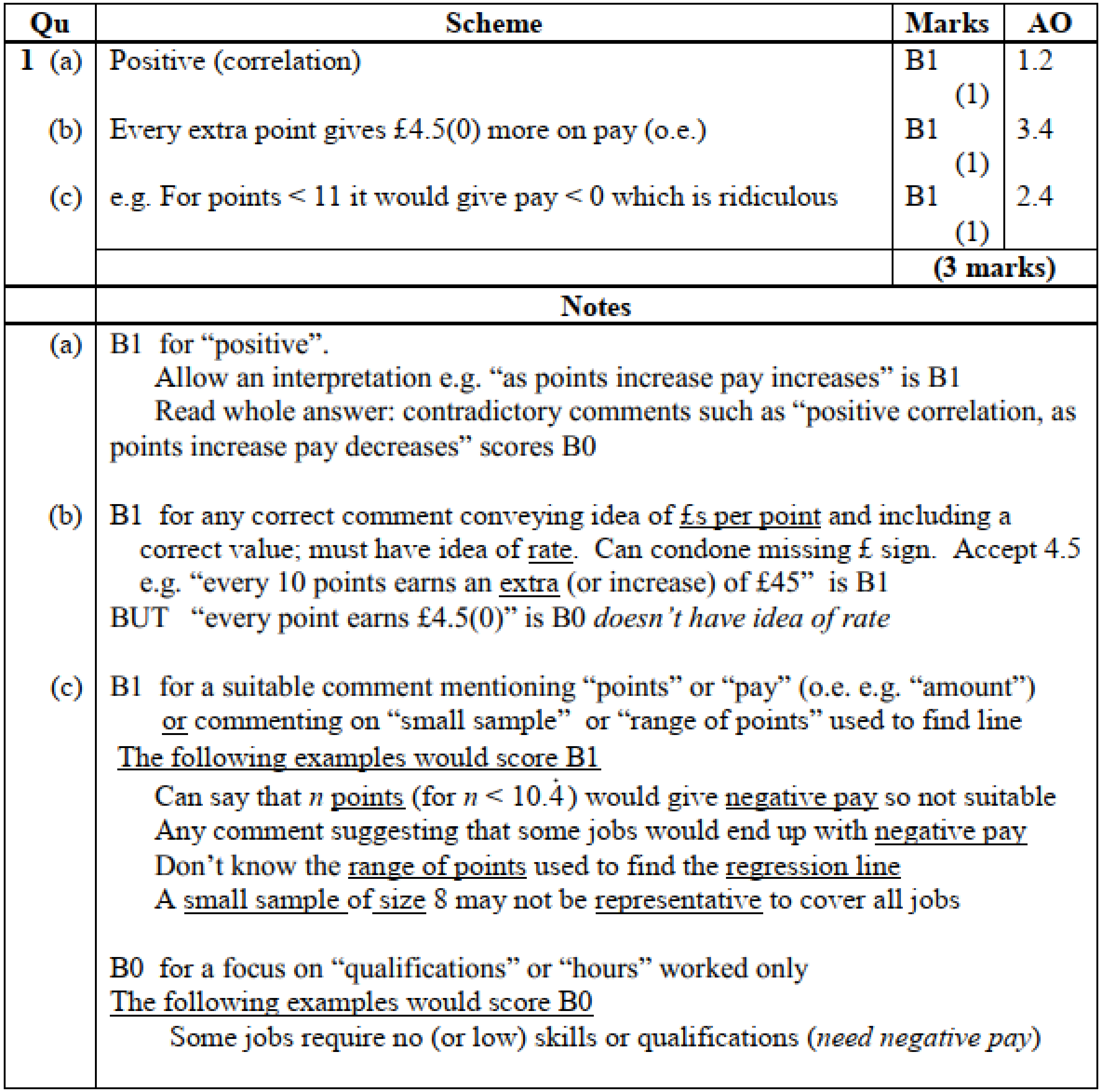
(c) Explain why this model might not be appropriate for all jobs in the company.

**(1)**

HOME

# AS 2018

## Correlation

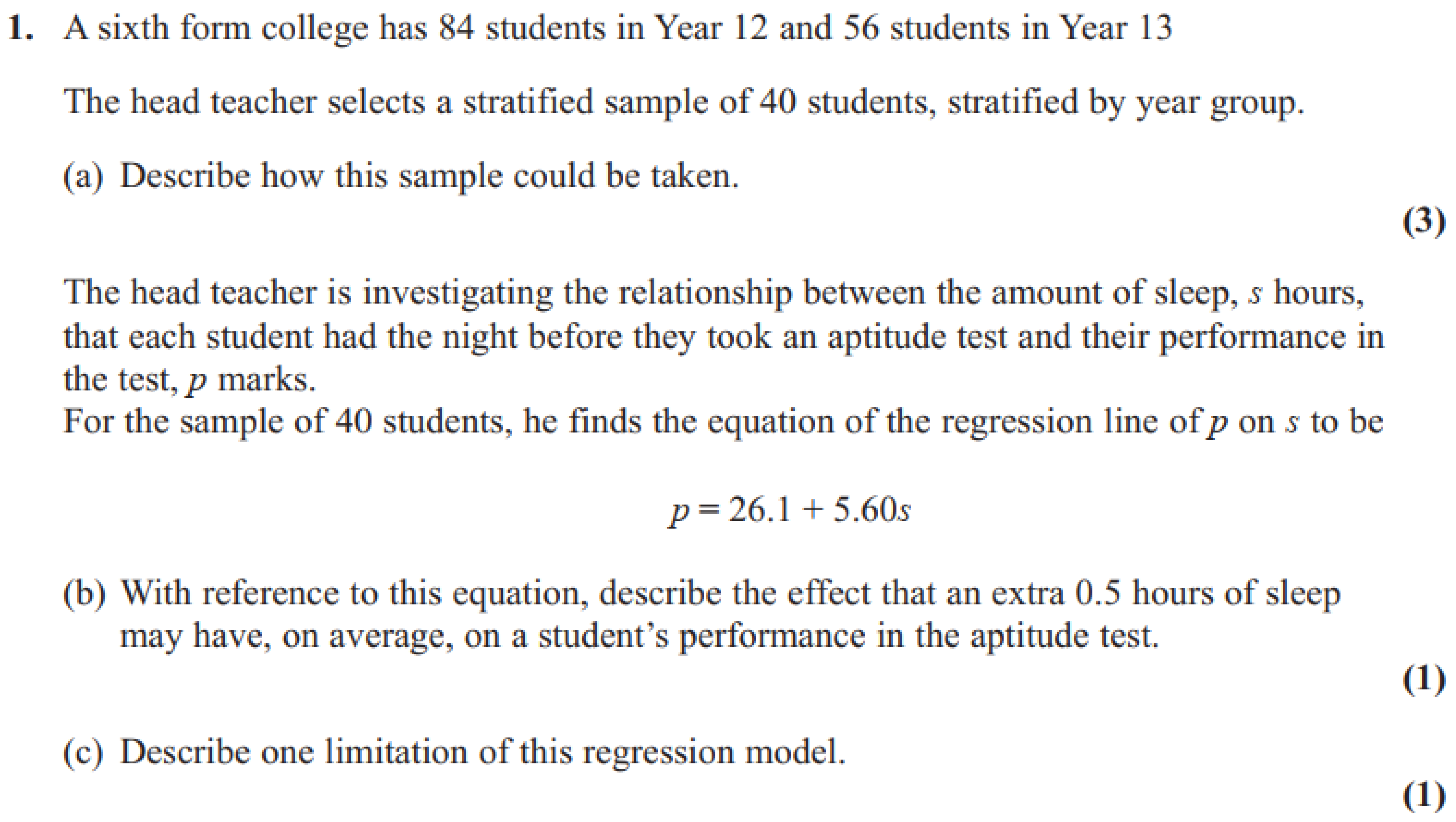


| Qu | Scheme | Marks | AO |
|---|---|---|---|
| 1 (a) | Positive (correlation) | B1<br>(1) | 1.2 |
| (b) | Every extra point gives £4.5(0) more on pay (o.e.) | B1<br>(1) | 3.4 |
| (c) | e.g. For points < 11 it would give pay < 0 which is ridiculous | B1<br>(1) | 2.4 |
| | | (3 marks) | |
| | Notes | | |
| (a) | B1 for "positive".<br>Allow an interpretation e.g. "as points increase pay increases" is B1<br>Read whole answer: contradictory comments such as "positive correlation, as points increase pay decreases" scores B0 | | |
| (b) | B1 for any correct comment conveying idea of £s per point and including a correct value; must have idea of rate. Can condone missing £ sign. Accept 4.5<br>e.g. "every 10 points earns an extra (or increase) of £45" is B1<br>BUT "every point earns £4.5(0)" is B0 *doesn't have idea of rate* | | |
| (c) | B1 for a suitable comment mentioning "points" or "pay" (o.e. e.g. "amount") or commenting on "small sample" or "range of points" used to find line<br>The following examples would score B1<br>Can say that $n$ points (for $n < 10.\dot{4}$) would give negative pay so not suitable<br>Any comment suggesting that some jobs would end up with negative pay<br>Don't know the range of points used to find the regression line<br>A small sample of size 8 may not be representative to cover all jobs<br><br>B0 for a focus on "qualifications" or "hours" worked only<br>The following examples would score B0<br>Some jobs require no (or low) skills or qualifications (*need negative pay*) | | |

HOME

## AS 2019

## Correlation



**1.** A sixth form college has 84 students in Year 12 and 56 students in Year 13

The head teacher selects a stratified sample of 40 students, stratified by year group.

(a) Describe how this sample could be taken.

**(3)**

The head teacher is investigating the relationship between the amount of sleep, $s$ hours, that each student had the night before they took an aptitude test and their performance in the test, $p$ marks.
For the sample of 40 students, he finds the equation of the regression line of $p$ on $s$ to be

$$p = 26.1 + 5.60s$$

(b) With reference to this equation, describe the effect that an extra 0.5 hours of sleep may have, on average, on a student's performance in the aptitude test.

**(1)**

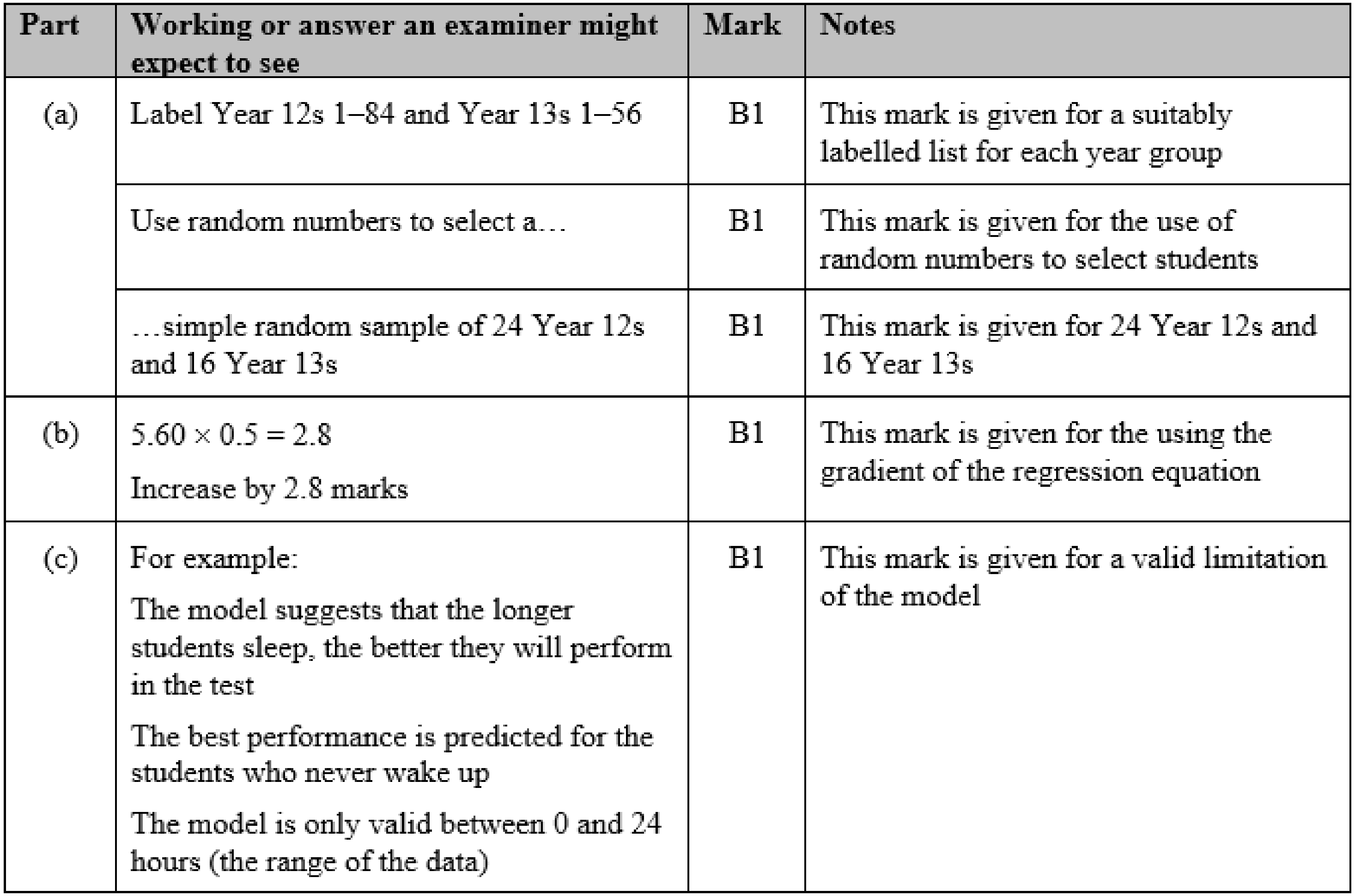
(c) Describe one limitation of this regression model.

**(1)**

HOME

# AS 2019

## Correlation



| Part | Working or answer an examiner might expect to see | Mark | Notes |
|---|---|---|---|
| (a) | Label Year 12s 1–84 and Year 13s 1–56 | B1 | This mark is given for a suitably labelled list for each year group |
| | Use random numbers to select a… | B1 | This mark is given for the use of random numbers to select students |
| | …simple random sample of 24 Year 12s and 16 Year 13s | B1 | This mark is given for 24 Year 12s and 16 Year 13s |
| (b) | $5.60 \times 0.5 = 2.8$<br>Increase by 2.8 marks | B1 | This mark is given for the using the gradient of the regression equation |
| (c) | For example:<br>The model suggests that the longer students sleep, the better they will perform in the test<br>The best performance is predicted for the students who never wake up<br>The model is only valid between 0 and 24 hours (the range of the data) | B1 | This mark is given for a valid limitation of the model |

HOME

# Correlation

1. The relationship between two variables $p$ and $t$ is modelled by the regression line with equation

$$p = 22 - 1.1\,t$$

The model is based on observations of the independent variable, $t$, between 1 and 10

(a) Describe the correlation between $p$ and $t$ implied by this model.

**(1)**

Given that $p$ is measured in centimetres and $t$ is measured in days,

(b) state the units of the gradient of the regression line.

**(1)**

Using the model,

(c) calculate the change in $p$ over a 3-day period.

**(2)**

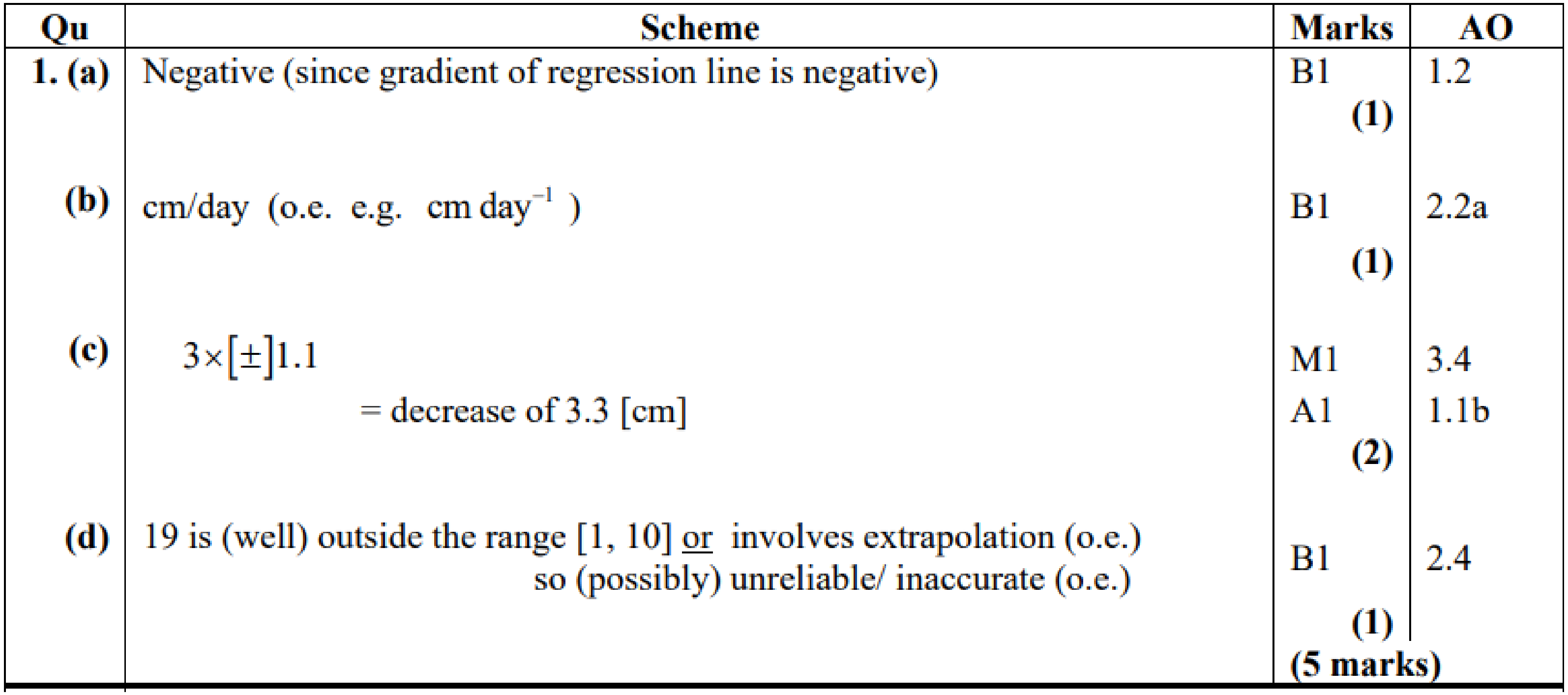
Tisam uses this model to estimate the value of $p$ when $t = 19$

(d) Comment, giving a reason, on the reliability of this estimate.

**(1)**

# AS 2022

Correlation



| Qu | Scheme | Marks | AO |
|---|---|---|---|
| **1. (a)** | Negative (since gradient of regression line is negative) | B1 | 1.2 |
| | | **(1)** | |
| **(b)** | cm/day (o.e. e.g. cm day$^{-1}$ ) | B1 | 2.2a |
| | | **(1)** | |
| **(c)** | $3\times[\pm]1.1$ | M1 | 3.4 |
| | = decrease of 3.3 [cm] | A1 | 1.1b |
| | | **(2)** | |
| **(d)** | 19 is (well) outside the range [1, 10] or involves extrapolation (o.e.) so (possibly) unreliable/ inaccurate (o.e.) | B1 | 2.4 |
| | | **(1)** | |
| | | **(5 marks)** | |

HOME

# AS 2023 (Part 1)

Correlation

**2.** Fred and Nadine are investigating whether there is a linear relationship between Daily Mean Pressure, $p$ hPa, and Daily Mean Air Temperature, $t$ °C, in Beijing using the 2015 data from the large data set.

Fred randomly selects one month from the data set and draws the scatter diagram in Figure 1 using the data from that month.

The scale has been left off the horizontal axis.

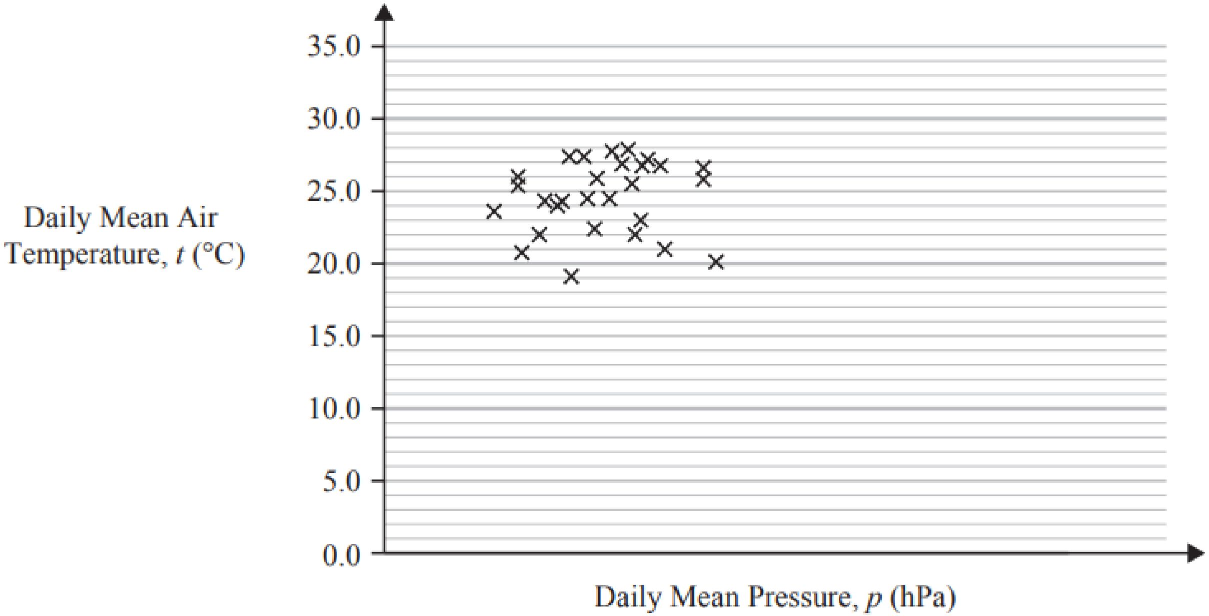


**Figure 1**

(a) Describe the correlation shown in Figure 1.

**(1)**

HOME

# AS 2023 (Part 2)

Correlation

Nadine chooses to use all of the data for Beijing from 2015 and draws the scatter diagram in Figure 2.

She uses the same scales as Fred.

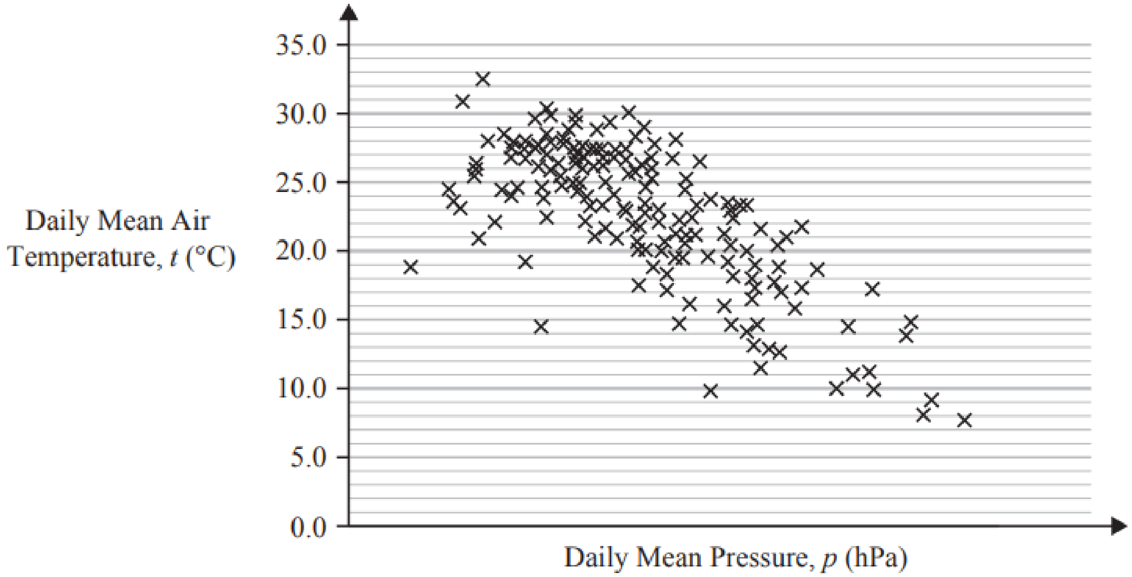


**Figure 2**

(b) Explain, in context, what Nadine can infer about the relationship between $p$ and $t$ using the information shown in Figure 2.

**(1)**

(c) Using your knowledge of the large data set, state a value of $p$ for which interpolation can be used with Figure 2 to predict a value of $t$.

**(1)**

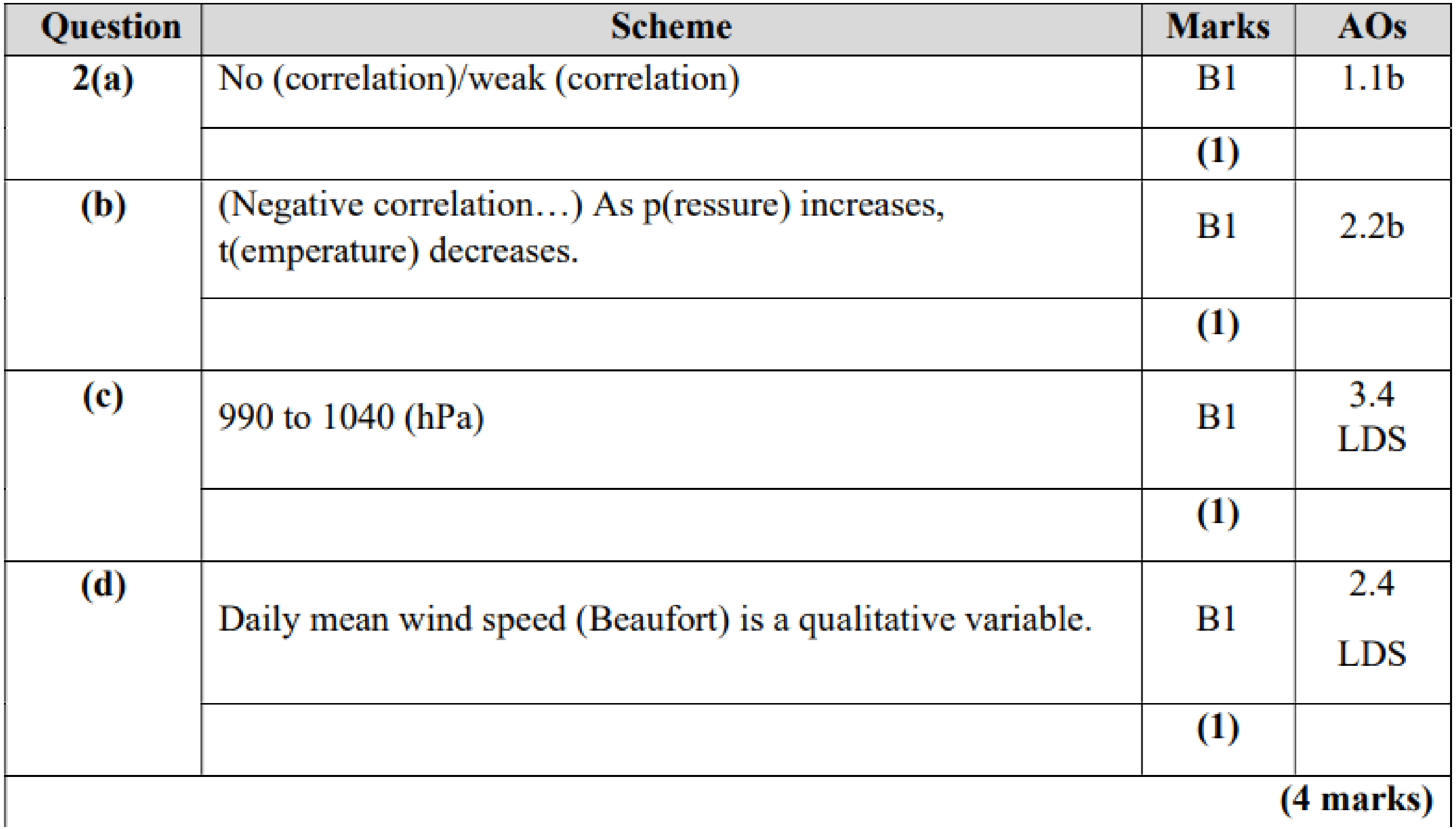
(d) Using your knowledge of the large data set, explain why it is not meaningful to look for a linear relationship between Daily Mean Wind Speed (Beaufort Conversion) and Daily Mean Air Temperature in Beijing in 2015.

**(1)**

HOME

# AS 2023

Correlation



| Question | Scheme | Marks | AOs |
|---|---|---|---|
| 2(a) | No (correlation)/weak (correlation) | B1 | 1.1b |
| | | **(1)** | |
| **(b)** | (Negative correlation…) As p(ressure) increases, t(emperature) decreases. | B1 | 2.2b |
| | | **(1)** | |
| **(c)** | 990 to 1040 (hPa) | B1 | 3.4<br>LDS |
| | | **(1)** | |
| **(d)** | Daily mean wind speed (Beaufort) is a qualitative variable. | B1 | 2.4<br>LDS |
| | | **(1)** | |
| | | | **(4 marks)** |

HOME

# Probability

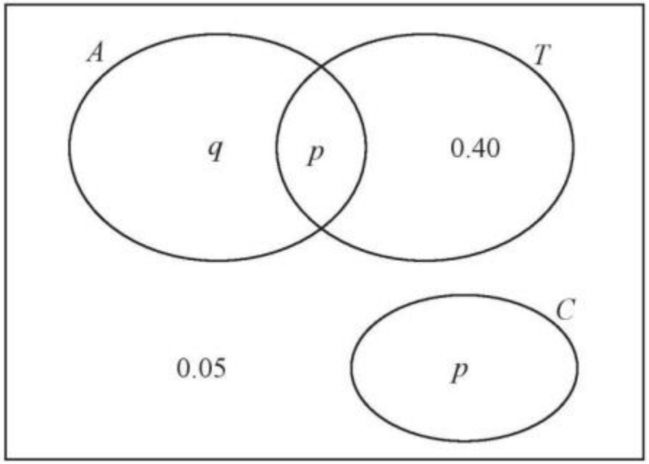
**3.** The Venn diagram shows the probabilities for students at a college taking part in various sports.

$A$ represents the event that a student takes part in Athletics.

$T$ represents the event that a student takes part in Tennis.

$C$ represents the event that a student takes part in Cricket.

$p$ and $q$ are probabilities.



The probability that a student selected at random takes part in Athletics or Tennis is 0.75

(a) Find the value of $p$.

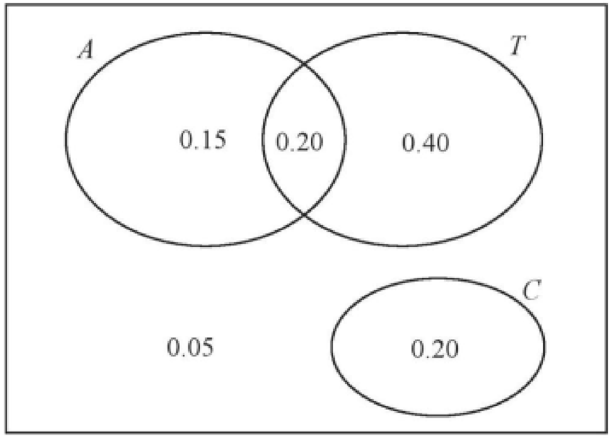
(1)

(b) State, giving a reason, whether or not the events $A$ and $T$ are statistically independent. Show your working clearly.

(3)

(c) Find the probability that a student selected at random does not take part in Athletics or Cricket.

(1)

| Question | Scheme | Marks | AOs |
|---|---|---|---|
| 3 (a) | $p = [1 - 0.75 - 0.05 =]$ **<u>0.20</u>** | B1 | 1.1b |
| | | **(1)** | |
| (b) | $q =$ **<u>0.15</u>** | B1ft | 1.1b |
| | P(*A*) = 0.35  P(*T*) = 0.6  P(*A* and *T*) = 0.20<br>P(*A*)×P(*T*) = 0.21 | M1 | 2.1 |
| | Since 0.20 ≠ 0.21 therefore *A* and *T* are **not** independent | A1 | 2.4 |
| | | **(3)** | |
| |  | | |
| (c) | P(not [*A* or *C*]) = **<u>0.45</u>** | B1 | 1.1b |
| | | **(1)** | |
| | | | **(5 marks)** |

| Part | Notes |
|---|---|
| (a) | B1cao for $p = 0.20$ |
| (b) | B1ft for use of their $p$ and P(*A* or *T*) to find $q$ i.e. 0.75 – "$p$" – 0.40 <u>or</u> $q = 0.15$ |
| | M1 for the statement of all probabilities required for a suitable test and sight of any appropriate calculations required. |
| | A1 All probabilities correct, correct comparison and suitable comment. |
| (c) | B1cao for 0.45 |

HOME

**AS 2018**

Probability

**2.** A factory buys 10% of its components from supplier $A$, 30% from supplier $B$ and the rest from supplier $C$. It is known that 6% of the components it buys are faulty.

Of the components bought from supplier $A$, 9% are faulty and of the components bought from supplier $B$, 3% are faulty.

(a) Find the percentage of components bought from supplier $C$ that are faulty.

**(3)**

A component is selected at random.

(b) Explain why the event "the component was bought from supplier $B$" is not statistically independent from the event "the component is faulty".

**(1)**

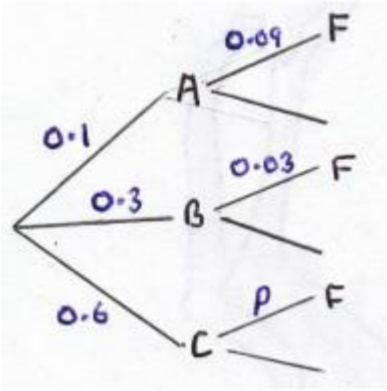
HOME

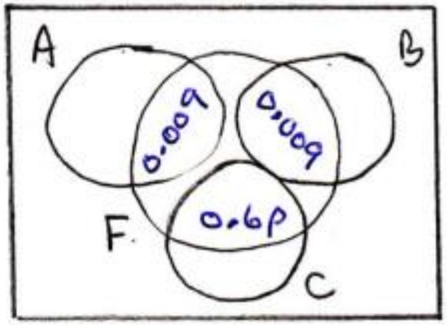
# AS 2018

## Probability

| Qu | Scheme | Marks | AO |
|---|---|---|---|
| **2** (a) | [Let $p = \mathrm{P}(F \mid C)$] | | |
| | Tree diagram or some other method to find an equation for $p$ | M1 | 2.1 |
| | $0.1\times 0.09+0.3\times 0.03+0.6\times p=0.06$ | A1 | 1.1b |
| | $p=0.07$ i.e. 7% | A1 | 1.1b |
| | | (3) | |
| (b) | e.g. $\mathrm{P}(B \text{ and } F) = 0.3\times 0.03 = 0.009$ but<br>$\mathrm{P}(B)\times\mathrm{P}(F) = 0.3\times 0.06 = 0.018$<br>These are not equal so not independent | B1 | 2.4 |
| | | (1) | |
| | | **(4 marks)** | |

| | Notes |
|---|---|
| (a) | M1 for selecting a suitable method to find the missing probability<br>e.g. sight of tree diagram with 0.1, 0.3, 0.6 and 0.09, 0.03, $p$ suitably placed<br>e.g. sight of VD with 0.009 for $A\cap F$ and $B\cap F$ and $0.6p$ suitably placed<br>or attempt an equation with at least one correct numerical and one "$p$" product (not necessarily correct) on LHS<br>or for sight of $0.06-(0.009+0.009)$ (o.e. e.g. $6-1.8 = 4.2\%$)<br>1st A1 for a correct equation for $p$ (May be implied by a correct answer)<br>or for the expression $\dfrac{0.06-(0.009+0.009)}{0.6}$ (o.e.)<br>2nd A1 for 7% ( accept 0.07)<br>**Correct Ans:** Provided there is no incorrect working seen award 3/3<br>e.g. may just see tree diagram with 0.07 for $p$ (probably from trial and improv') |
| (b) | B1 for a suitable explanation…may talk about 2nd branches on tree diagram and point out that $0.03 \neq 0.06$ but need some supporting calculation/words<br>Can condone incorrect use of set notation (it is not on AS spec) provided the rest of the calculations and words are correct. |



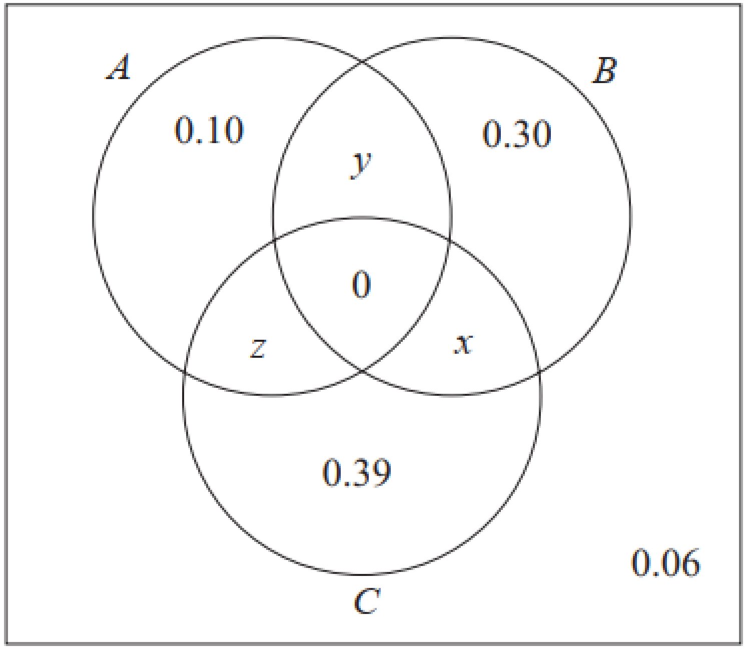


HOME

**AS 2019**

Probability

2. The Venn diagram shows three events, $A$, $B$ and $C$, and their associated probabilities.



Events $B$ and $C$ are mutually exclusive.
Events $A$ and $C$ are independent.

Showing your working, find the value of $x$, the value of $y$ and the value of $z$.

**(5)**

HOME

# AS 2019

## Probability

| Part | Working or answer an examiner might expect to see | Mark | Notes |
|---|---|---|---|
| | Events *B* and *c* are mutually exclusive so<br>$x = 0$ | B1 | This mark is given for deducing that $x = 0$ |
| | $\mathrm{P}(A) = 0.1 + z + y$<br>$\mathrm{P}(C) = 0.93 + z\ [+ x]$<br>$\mathrm{P}(A \text{ and } C) = z$ | M1 | This mark is given for identifying the probabilities required for independence |
| | $\mathrm{P}(A \text{ and } C) = \mathrm{P}(A) \times \mathrm{P}(C)$<br>$(0.1 + z + y) \times (0.39 + z + 0) = z$ | M1 | This mark is given for using independence |
| | $\sum p = 1$<br>$0.06 + 0.3 + 0.39 + 0.1 + z + y + 0 = 1$ | M1 | This mark is given for using the fact that the sum of probabilities sum to 1 |
| | $y + z = 0.15$<br>$z = (0.1 + 0.15) \times (0.39 + z)$<br>$z = 0.975 + 0.25z$<br>$0.75z = 0.0975$<br>$z = \dfrac{0.0975}{0.75} = 0.13$<br>$y = 0.15 - 0.13 = 0.02$ | A1 | This mark is given for finding the values of *y* and *z* |

HOME

**AS 2020**

Probability

**3.** In a game, a player can score 0, 1, 2, 3 or 4 points each time the game is played.

The random variable $S$, representing the player's score, has the following probability distribution where $a$, $b$ and $c$ are constants.

| $s$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $\mathrm{P}(S = s)$ | $a$ | $b$ | $c$ | 0.1 | 0.15 |

The probability of scoring less than 2 points is twice the probability of scoring at least 2 points.

Each game played is independent of previous games played.

John plays the game twice and adds the two scores together to get a total.

Calculate the probability that the total is 6 points.

**(6)**

HOME

**AS 2020**

Probability

| Question | Scheme | Marks | AOs |
|---|---|---|---|
| **3** | Overall method | M1 | 2.1 |
| | $a+b=2c+0.5$ oe **or** $a+b=2(1-a-b)$ | B1 | 2.2a |
| | $a+b+c=0.75$ oe | B1 | 1.1b |
| | $3c=0.25$ $\left[c=0.0833\ldots \text{ or } \frac{1}{12}\right]$ | M1 | 1.1b |
| | P(scoring 2,4 or 4,2 or 3,3) = $2\times"\frac{1}{12}"\times0.15+0.1^2$ | M1 | 3.1b |
| | $=0.035$ oe | A1cso | 1.1b |
| | | (6) | |

**(6 marks)**

| Notes | | |
|---|---|---|
| 3 | **M1:** | A fully correct method with all the required steps. For gaining 2 correct equations with at least one correct(allow if unsimplified). Attempting to solve to find a value of $c$ followed by **correct method** to find the probability |
| | **B1:** | Forming a correct equation from the information given in the question |
| | **B1:** | A correct equation using the sum of the probabilities equals 1 |
| | **M1:** | Correct method for solving 2 equations to find $c$ Implied by $c=\frac{1}{12}$ |
| | **M1:** | Recognising the ways to get a total of 6. Condone missing arrangments or repeats. Do not ignore extras written unless ignored in the calculation. May be implied by<br>$m\times"\frac{1}{12}"\times0.15+n\times0.1^2$ where $m$ and $n$ are positive integers |
| | **A1cso:** | Cao 0.035, $\frac{7}{200}$ oe |

HOME

**AS 2020**

Probability

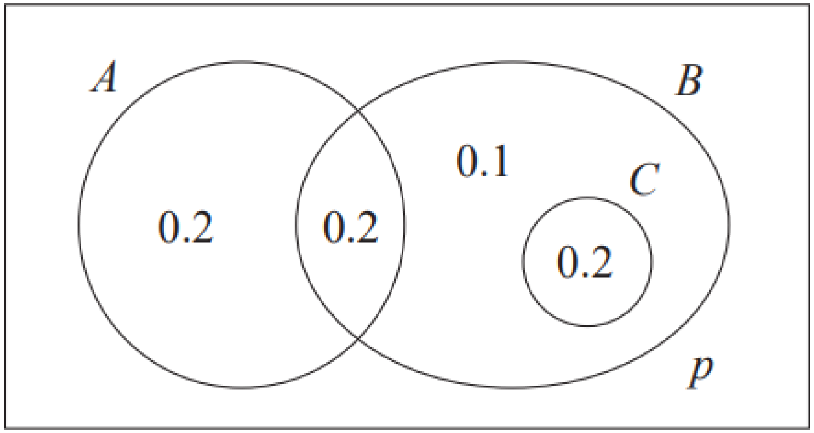
| Question | Scheme | Marks | AOs |
|---|---|---|---|
| **3** | Overall method | M1 | 2.1 |
| | $a+b=2c+0.5$ oe **or** $a+b=2(1-a-b)$ | B1 | 2.2a |
| | $a+b+c=0.75$ oe | B1 | 1.1b |
| | $3c=0.25 \quad \left[c=0.0833\ldots \text{ or } \frac{1}{12}\right]$ | M1 | 1.1b |
| | P(scoring 2,4 or 4,2 or 3,3) = $2\times"\frac{1}{12}"\times 0.15+0.1^2$ | M1 | 3.1b |
| | $=0.035$ oe | A1cso | 1.1b |
| | | (6) | |

**(6 marks)**

| Notes | | |
|---|---|---|
| 3 | **M1:** | A fully correct method with all the required steps. For gaining 2 correct equations with at least one correct(allow if unsimplified). Attempting to solve to find a value of *c* followed by **correct method** to find the probability |
| | **B1:** | Forming a correct equation from the information given in the question |
| | **B1:** | A correct equation using the sum of the probabilities equals 1 |
| | **M1:** | Correct method for solving 2 equations to find *c* Implied by $c=\frac{1}{12}$ |
| | **M1:** | Recognising the ways to get a total of 6. Condone missing arrangments or repeats. Do not ignore extras written unless ignored in the calculation. May be implied by $m\times"\frac{1}{12}"\times 0.15+n\times 0.1^2$ where *m* and *n* are positive integers |
| | **A1cso:** | Cao 0.035, $\frac{7}{200}$ oe |

HOME

**1.**



The Venn diagram, where $p$ is a probability, shows the 3 events $A$, $B$ and $C$ with their associated probabilities.

(a) Find the value of $p$.

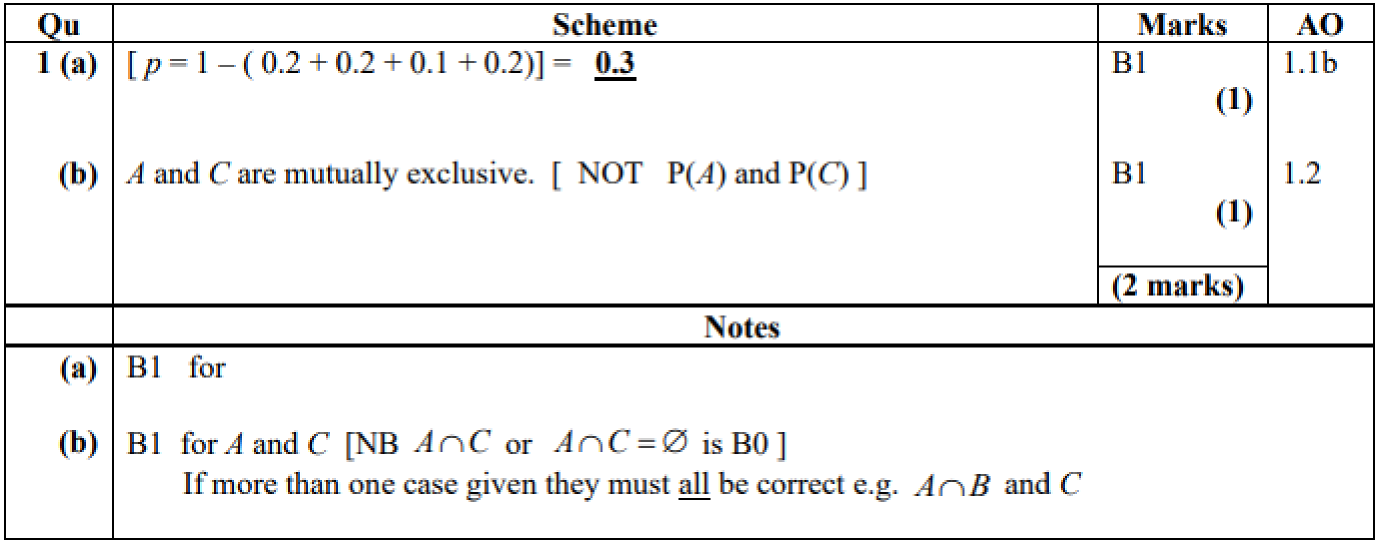
**(1)**

(b) Write down a pair of mutually exclusive events from $A$, $B$ and $C$.

**(1)**

**AS 2021**

**Probability**



| Qu | Scheme | Marks | AO |
|---|---|---|---|
| 1 (a) | [ $p = 1 - (0.2 + 0.2 + 0.1 + 0.2)$] = $\underline{\mathbf{0.3}}$ | B1<br>**(1)** | 1.1b |
| (b) | $A$ and $C$ are mutually exclusive. [ NOT P($A$) and P($C$) ] | B1<br>**(1)** | 1.2 |
| | | **(2 marks)** | |
| | **Notes** | | |
| (a) | B1 for | | |
| (b) | B1 for $A$ and $C$ [NB $A \cap C$ or $A \cap C = \varnothing$ is B0 ]<br>If more than one case given they must <u>all</u> be correct e.g. $A \cap B$ and $C$ | | |

HOME

**5.** Two bags, **A** and **B**, each contain balls which are either red or yellow or green.

Bag **A** contains 4 red, 3 yellow and $n$ green balls.
Bag **B** contains 5 red, 3 yellow and 1 green ball.

A ball is selected at random from bag **A** and placed into bag **B**.
A ball is then selected at random from bag **B** and placed into bag **A**.

The probability that bag **A** now contains an equal number of red, yellow and green balls is $p$.

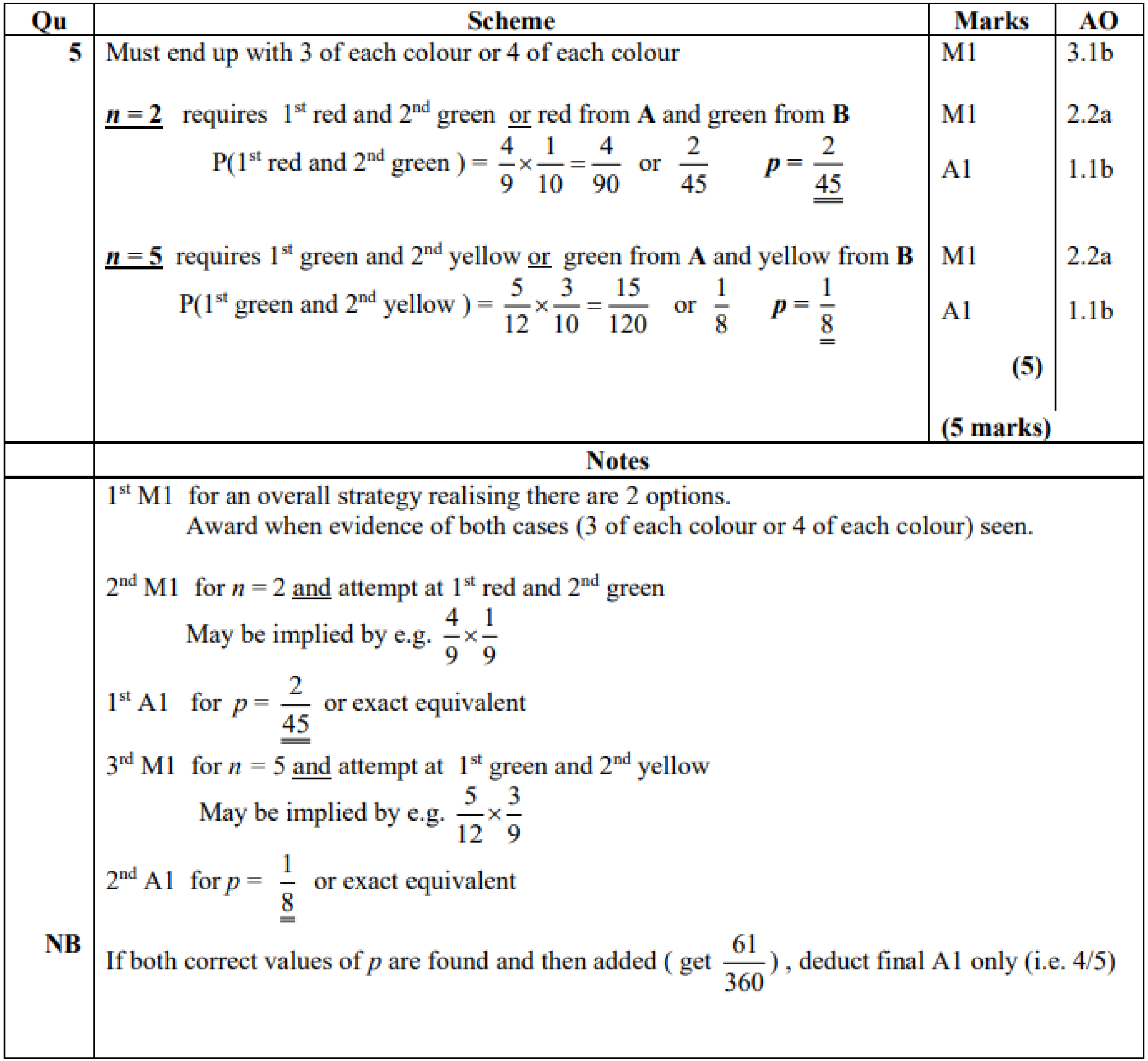
Given that $p > 0$, find the possible values of $n$ and $p$.

**(5)**

HOME

# AS 2021

## Probability



| Qu | Scheme | Marks | AO |
|---|---|---|---|
| 5 | Must end up with 3 of each colour or 4 of each colour | M1 | 3.1b |
| | $\underline{n=2}$ requires 1st red and 2nd green or red from **A** and green from **B** | M1 | 2.2a |
| | P(1st red and 2nd green ) = $\frac{4}{9}\times\frac{1}{10}=\frac{4}{90}$ or $\frac{2}{45}$ $\quad p=\underline{\underline{\frac{2}{45}}}$ | A1 | 1.1b |
| | $\underline{n=5}$ requires 1st green and 2nd yellow or green from **A** and yellow from **B** | M1 | 2.2a |
| | P(1st green and 2nd yellow ) = $\frac{5}{12}\times\frac{3}{10}=\frac{15}{120}$ or $\frac{1}{8}$ $\quad p=\underline{\underline{\frac{1}{8}}}$ | A1 | 1.1b |
| | | (5) | |
| | | (5 marks) | |

| | Notes |
|---|---|
| | 1st M1 for an overall strategy realising there are 2 options.<br>Award when evidence of both cases (3 of each colour or 4 of each colour) seen. |
| | 2nd M1 for $n=2$ and attempt at 1st red and 2nd green<br>May be implied by e.g. $\frac{4}{9}\times\frac{1}{9}$ |
| | 1st A1 for $p=\underline{\underline{\frac{2}{45}}}$ or exact equivalent |
| | 3rd M1 for $n=5$ and attempt at 1st green and 2nd yellow<br>May be implied by e.g. $\frac{5}{12}\times\frac{3}{9}$ |
| | 2nd A1 for $p=\underline{\underline{\frac{1}{8}}}$ or exact equivalent |
| NB | If both correct values of $p$ are found and then added ( get $\frac{61}{360}$) , deduct final A1 only (i.e. 4/5) |

HOME

# AS 2022

Probability

**5.** Manon has two biased spinners, one red and one green.

The random variable $R$ represents the score when the red spinner is spun.
The random variable $G$ represents the score when the green spinner is spun.

The probability distributions for $R$ and $G$ are given below.

| $r$ | 2 | 3 |
|---|---|---|
| $\mathrm{P}(R = r)$ | $\frac{1}{4}$ | $\frac{3}{4}$ |

| $g$ | 1 | 4 |
|---|---|---|
| $\mathrm{P}(G = g)$ | $\frac{2}{3}$ | $\frac{1}{3}$ |

Manon spins each spinner once and adds the two scores.

(a) Find the probability that

(i) the sum of the two scores is 7

(ii) the sum of the two scores is less than 4

**(3)**

The random variable $X = mR + nG$ where $m$ and $n$ are integers.

$$\mathrm{P}(X = 20) = \frac{1}{6} \qquad \text{and} \qquad \mathrm{P}(X = 50) = \frac{1}{4}$$

(b) Find the value of $m$ and the value of $n$

**(5)**

HOME

**AS 2022**

## Probability

| Qu | Scheme | Mark | AO |
|---|---|---|---|
| **5. (a)(i)** | Require $R = 3$ and $G = 4$ so probability is $\frac{3}{4}\times\frac{1}{3}$ | M1 | 2.1 |
| | $= \underline{\frac{1}{4}}$ or $\underline{\mathbf{0.25}}$ | A1 | 1.1b |
| **(ii)** | [$R$ must be 2 and $G = 1$so $\frac{1}{4}\times\frac{2}{3}$ ] $= \underline{\frac{1}{6}}$ | A1 | 1.1b |
| | | **(3)** | |
| **(b)** | $\text{P}(X = 50) = 0.25$ must mean $R = 3$ and $G = 4$ | M1 | 3.1a |
| | so $3m + 4n = 50$ | A1 | 1.1b |
| | $\text{P}(X = 20) = \frac{1}{6} \Rightarrow R = 2, G = 1$ so $2m + n = 20$ | A1 | 2.1 |
| | Solving: $3m + 4(20 - 2m) = 50$ (o.e.) | M1 | 1.1b |
| | $\underline{\boldsymbol{m = 6}}$ and $\underline{\boldsymbol{n = 8}}$ | A1 | 3.2a |
| | | **(5)** | |
| | | **(8 marks)** | |

HOME

# AS 2023

## Probability

**3.** In an after-school club, students can choose to take part in Art, Music, both or neither.

There are 45 students that attend the after-school club. Of these

- 25 students take part in Art
- 12 students take part in both Art and Music
- the number of students that take part in Music is $x$

(a) Find the range of possible values of $x$

**(2)**

One of the 45 students is selected at random.

Event $A$ is the event that the student selected takes part in Art.

Event $M$ is the event that the student selected takes part in Music.

(b) Determine whether or not it is possible for the events $A$ and $M$ to be independent.

**(4)**

HOME

**AS 2023**

**Probability**

| Que. | Scheme | Marks | AOs |
|---|---|---|---|
| 3(a) | $45 - 25 = 20$ or e.g. '$25 \leqslant 13 + 12 + y \leqslant 45$' | M1 | 2.1 |
| | $12 \leqslant x \leqslant 32$ | A1 | 1.1b |
| | | (2) | |
| (b) | To be independent $\mathrm{P}(A)\times\mathrm{P}(M)=\mathrm{P}(A \text{ and } M)$ | M1 | 1.1a |
| | $\mathrm{P}(M)=\frac{\mathrm{P}(A\text{ and }M)}{\mathrm{P}(A)}=\frac{\frac{12}{45}}{\frac{25}{45}}=\frac{12}{25}$ or $\frac{25}{45}\times\mathrm{P}(M)=\frac{12}{45}$<br>or $\frac{25}{45}\times\frac{x}{45}=\frac{12}{45}$ or $\frac{25}{45}\times\frac{12+y}{45}=\frac{12}{45}$ | A1 | 2.1 |
| | The number of students taking part in music would be $\frac{12}{25}\times 45 = 21.6$<br>The number of students taking part in music but not art would be $y = 9.6$ | A1 | 1.1b |
| | …so it is not possible for $A$ and $M$ to be independent (since it must be a whole number). | A1 | 2.2a |
| | | (4) | |
| | | | (6 marks) |

HOME

**AS 2023**

Probability

**5.** Julia selects 3 letters at random, one at a time without replacement, from the word

VARIANCE

The discrete random variable $X$ represents the number of times she selects a letter A.

(a) Find the complete probability distribution of $X$.

**(5)**

Yuki selects 10 letters at random, one at a time **with** replacement, from the word

DEVIATION

(b) Find the probability that he selects the letter E at least 4 times.

**(3)**

HOME

**AS 2023**

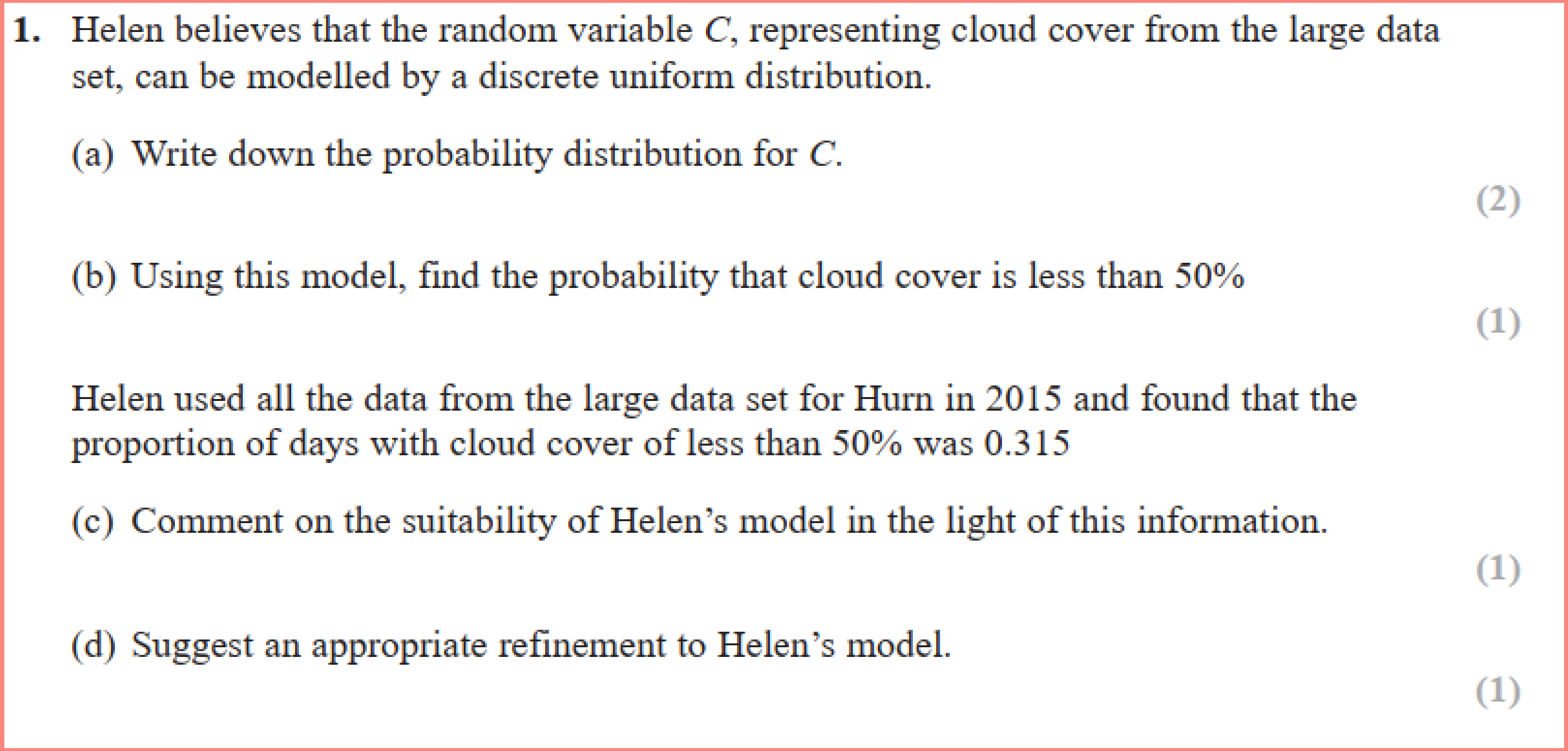
Probability

<table>
<tr><th>Que.</th><th>Scheme</th><th>Marks</th><th>AOs</th></tr>
<tr><td>5(a)</td><td>$X = 0,\ 1,\ 2$ only</td><td>B1</td><td>3.1b</td></tr>
<tr><td></td><td>$[\text{P}(X=0)=]\dfrac{6}{8}\times\dfrac{5}{7}\times\dfrac{4}{6}$</td><td>M1</td><td>1.1b</td></tr>
<tr><td></td><td>$[\text{P}(X=1)=]3\times\dfrac{2}{8}\times\dfrac{6}{7}\times\dfrac{5}{6}$ or<br>$[\text{P}(X=2)=]3\times\dfrac{2}{8}\times\dfrac{1}{7}\times\dfrac{6}{6}$</td><td>M1</td><td>2.1</td></tr>
<tr><td></td><td><table><tr><td>$x$</td><td>0</td><td>1</td><td>2</td></tr><tr><td>$\text{P}(X=x)$</td><td>$\dfrac{5}{14}$</td><td>$\dfrac{15}{28}$</td><td>$\dfrac{3}{28}$</td></tr></table></td><td>A1<br>A1</td><td>1.1b<br>1.1b</td></tr>
<tr><td></td><td></td><td>(5)</td><td></td></tr>
<tr><td>(b)</td><td>$J \sim \text{B}(10,\ \tfrac{1}{9})$</td><td>M1</td><td>3.1b</td></tr>
<tr><td></td><td>$\text{P}(J \geqslant 4) = 1 - \text{P}(J \leqslant 3)$ or<br>$\text{P}(J \geqslant 4) = \text{P}(J=4) + \text{P}(J=5) + \ldots + \text{P}(J=10)$ or<br>$1 - 0.981(57\ldots)$</td><td>M1</td><td>3.4</td></tr>
<tr><td></td><td>= awrt 0.0184</td><td>A1</td><td>1.1b</td></tr>
<tr><td></td><td></td><td>(3)</td><td></td></tr>
<tr><td colspan="4">(8 marks)</td></tr>
</table>

HOME

A2 2018

Probability



1. Helen believes that the random variable $C$, representing cloud cover from the large data set, can be modelled by a discrete uniform distribution.

(a) Write down the probability distribution for $C$.

(2)

(b) Using this model, find the probability that cloud cover is less than 50%

(1)

Helen used all the data from the large data set for Hurn in 2015 and found that the proportion of days with cloud cover of less than 50% was 0.315

(c) Comment on the suitability of Helen's model in the light of this information.

(1)

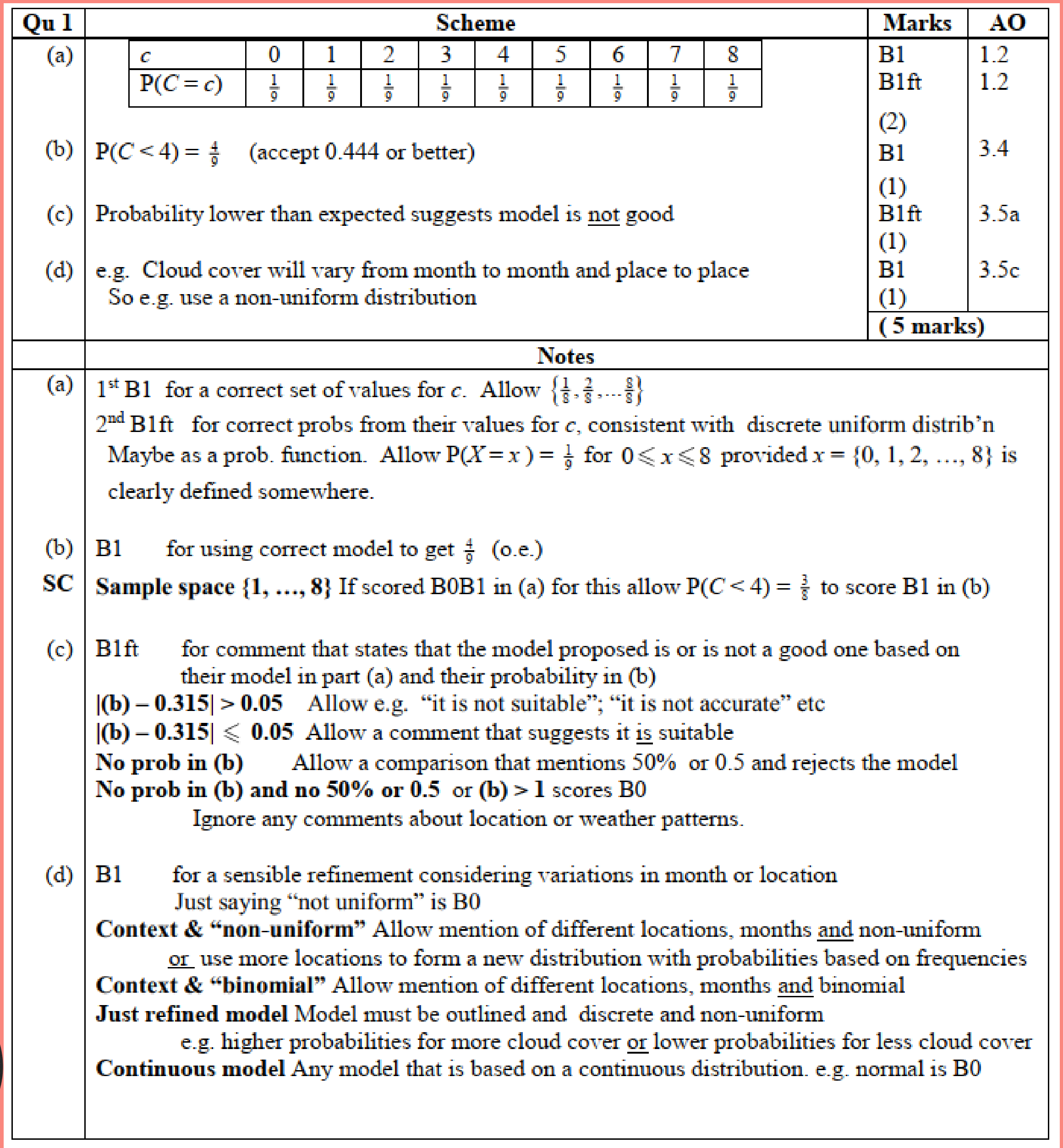
(d) Suggest an appropriate refinement to Helen's model.

(1)

HOME

# A2 2018

## Probability



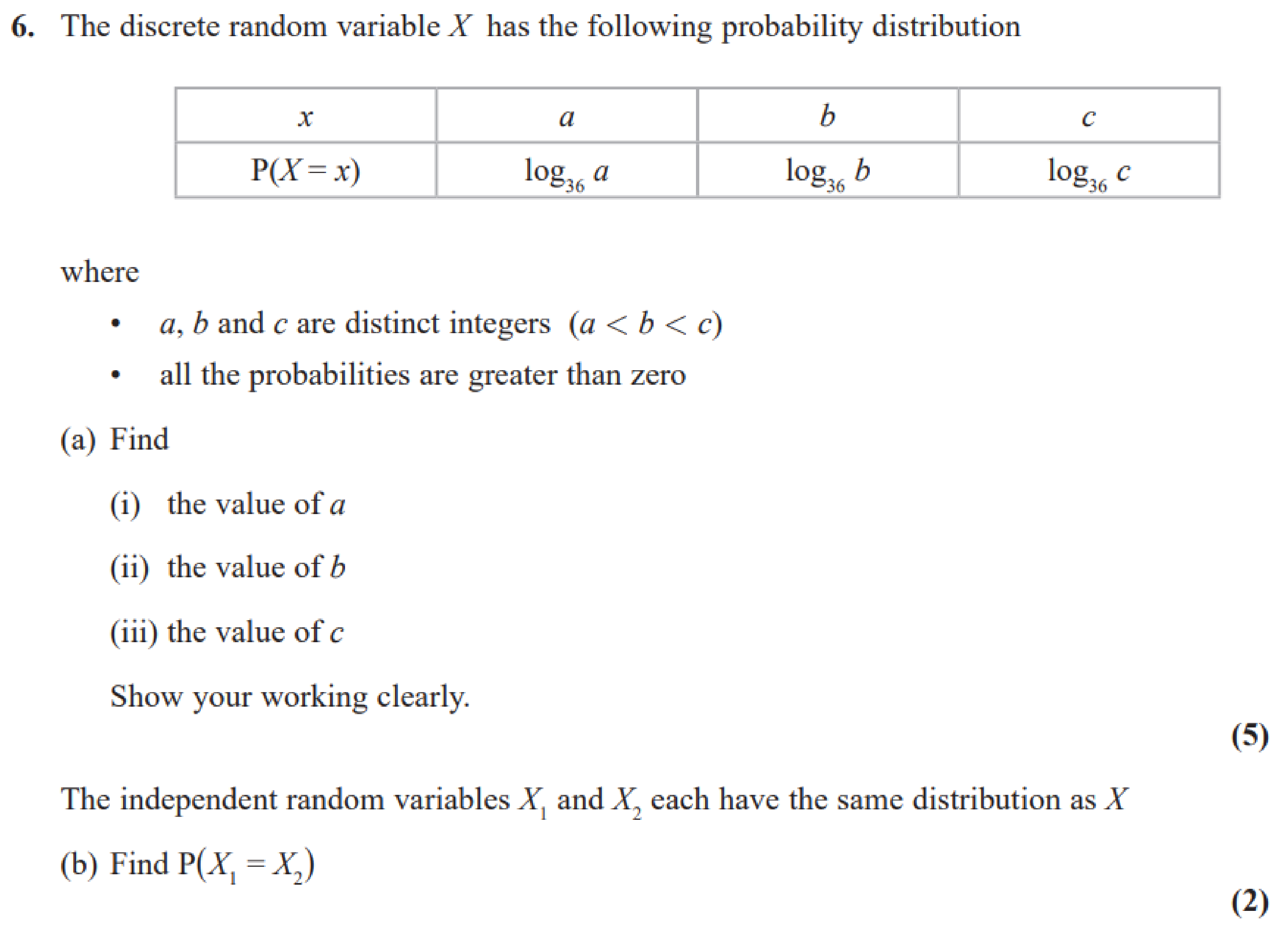
| Qu 1 | Scheme | Marks | AO |
|---|---|---|---|
| (a) | $\begin{array}{|c|c|c|c|c|c|c|c|c|c|} \hline c & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \hline \mathrm{P}(C=c) & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \hline \end{array}$ | B1<br>B1ft<br>(2) | 1.2<br>1.2 |
| (b) | $\mathrm{P}(C<4)=\frac{4}{9}$ (accept 0.444 or better) | B1<br>(1) | 3.4 |
| (c) | Probability lower than expected suggests model is not good | B1ft<br>(1) | 3.5a |
| (d) | e.g. Cloud cover will vary from month to month and place to place<br>So e.g. use a non-uniform distribution | B1<br>(1) | 3.5c |
| | | **( 5 marks)** | |

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

HOME

## A2 2021

Probability



**6.** The discrete random variable $X$ has the following probability distribution

| $x$ | $a$ | $b$ | $c$ |
|---|---|---|---|
| $\mathrm{P}(X = x)$ | $\log_{36} a$ | $\log_{36} b$ | $\log_{36} c$ |

where

- $a$, $b$ and $c$ are distinct integers $(a < b < c)$
- all the probabilities are greater than zero

(a) Find

(i) the value of $a$

(ii) the value of $b$

(iii) the value of $c$

Show your working clearly.

**(5)**

The independent random variables $X_1$ and $X_2$ each have the same distribution as $X$

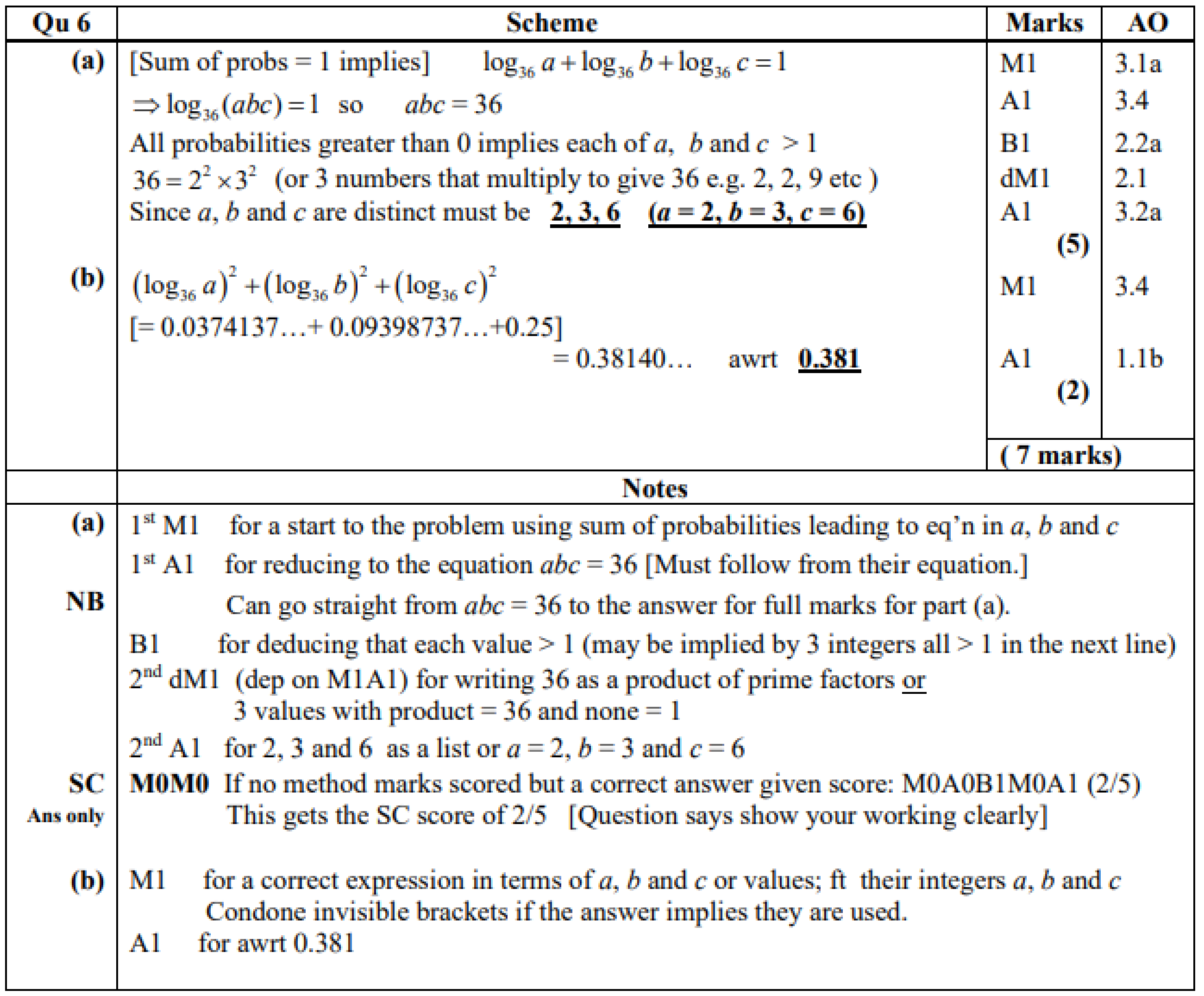
(b) Find $\mathrm{P}(X_1 = X_2)$

**(2)**

HOME

# A2 2021

Probability



| Qu 6 | Scheme | Marks | AO |
|---|---|---|---|
| (a) | [Sum of probs = 1 implies] $\log_{36} a+\log_{36} b+\log_{36} c=1$ | M1 | 3.1a |
| | $\Rightarrow \log_{36}(abc)=1$ so $abc=36$ | A1 | 3.4 |
| | All probabilities greater than 0 implies each of $a$, $b$ and $c$ $>1$ | B1 | 2.2a |
| | $36=2^2\times3^2$ (or 3 numbers that multiply to give 36 e.g. 2, 2, 9 etc ) | dM1 | 2.1 |
| | Since $a$, $b$ and $c$ are distinct must be **2, 3, 6** (***a* = 2, *b* = 3, *c* = 6**) | A1 | 3.2a |
| | | **(5)** | |
| (b) | $(\log_{36} a)^2+(\log_{36} b)^2+(\log_{36} c)^2$ | M1 | 3.4 |
| | [= 0.0374137…+ 0.09398737…+0.25] | | |
| | = 0.38140… awrt **0.381** | A1 | 1.1b |
| | | **(2)** | |
| | | **( 7 marks)** | |
| | **Notes** | | |
| (a) | $1^{st}$ M1 for a start to the problem using sum of probabilities leading to eq'n in $a$, $b$ and $c$<br>$1^{st}$ A1 for reducing to the equation $abc=36$ [Must follow from their equation.] | | |
| **NB** | Can go straight from $abc=36$ to the answer for full marks for part (a).<br>B1 for deducing that each value > 1 (may be implied by 3 integers all > 1 in the next line)<br>$2^{nd}$ dM1 (dep on M1A1) for writing 36 as a product of prime factors or 3 values with product = 36 and none = 1<br>$2^{nd}$ A1 for 2, 3 and 6 as a list or $a=2$, $b=3$ and $c=6$ | | |
| **SC**<br>**Ans only** | **M0M0** If no method marks scored but a correct answer given score: M0A0B1M0A1 (2/5)<br>This gets the SC score of 2/5 [Question says show your working clearly] | | |
| (b) | M1 for a correct expression in terms of $a$, $b$ and $c$ or values; ft their integers $a$, $b$ and $c$<br>Condone invisible brackets if the answer implies they are used.<br>A1 for awrt 0.381 | | |

HOME

# The Binomial Distribution

# AS SAMs

## The Binomial Distribution

**5.** (a) The discrete random variable $X \sim \mathrm{B}(40, 0.27)$

Find $\mathrm{P}(X \geqslant 16)$

**(2)**

Past records suggest that 30% of customers who buy baked beans from a large supermarket buy them in single tins. A new manager suspects that there has been a change in the proportion of customers who buy baked beans in single tins. A random sample of 20 customers who had bought baked beans was taken.

(b) Write down the hypotheses that should be used to test the manager's suspicion.

**(1)**

(c) Using a 10% level of significance, find the critical region for a two-tailed test to answer the manager's suspicion. You should state the probability of rejection in each tail, which should be less than 0.05

**(3)**

(d) Find the actual significance level of a test based on your critical region from part (c).

**(1)**

One afternoon the manager observes that 12 of the 20 customers who bought baked beans, bought their beans in single tins.

(e) Comment on the manager's suspicion in the light of this observation.

**(1)**

Later it was discovered that the local scout group visited the supermarket that afternoon to buy food for their camping trip.

(f) Comment on the validity of the model used to obtain the answer to part (e), giving a reason for your answer.

**(1)**

HOME

# AS SAMs

# The Binomial Distribution

| | | | |
|---|---|---|---|
| 5(a) | $P(X \geqslant 16) = 1 - P(X \leqslant 15)$ | M1 | 1.1b |
| | $= 1 - 0.949077\ldots \qquad =$ awrt $\underline{\mathbf{0.0509}}$ | A1 | 1.1b |
| | | **(2)** | |
| **(b)** | $H_0 : p = 0.3 \qquad H_1 : p \neq 0.3$ (Both correct in terms of $p$ or $\pi$) | B1 | 2.5 |
| | | **(1)** | |
| **(c)** | $[Y \sim B(20, 0.3)]$ sight of $P(Y \leqslant 2) = 0.0355$ or $P(Y \leqslant 9) = 0.9520$ | M1 | 2.1 |
| | Critical region is $\{Y \leqslant \mathbf{2}\}$ or (o.e.) | A1 | 1.1b |
| | $\{Y \geqslant \mathbf{10}\}$ (o.e.) | A1 | 1.1b |
| | | **(3)** | |
| **(d)** | $[0.0355 + (1 - 0.9520)] = 0.0835$ <u>or</u> $\underline{\mathbf{8.35\%}}$ | B1ft | 1.1b |
| | | **(1)** | |
| **(e)** | (Assuming that the 20 customers represent a random sample then) 12 is in the CR so the manager's suspicion is supported | B1ft | 3.2a |
| | | **(1)** | |
| **(f)** | e.g. (e) requires the 20 customers to be a random sample or independent and the members of the scout group may invalidate this so binomial distribution would not be valid (and conclusion in (e) is probably not valid) | B1 | 3.5a |
| | | **(1)** | |

| Part | Notes |
|---|---|
| **(a)** | M1 for dealing with $P(X \geqslant 16)$ – they need to use cumulative prob. function on calc. |
| | A1 awrt 0.0509 (from calculator) |
| **(b)** | B1 for both hypotheses in terms of $p$ or $\pi$ and $H_1$ must be 2-tail |
| **(c)** | M1 for correct use of tables to find probability associated with critical value. |
| | 1st A1 for the correct lower limit of the CR. Do not award for $P(Y \leqslant 2)$ |
| | 2nd A1 for the correct upper limit. |
| **(d)** | B1ft ft on their 0.0355 and (1 – their 0.9520) provided each probability is less than 0.05 |
| **(e)** | B1ft for a comment that relates 12 to their CR and makes a consistent comment relating this to the manager's suspicion |
| **(f)** | B1 for a comment that: gives a suitable reason based on lack of independence <u>or</u> the sample not being random <u>so</u> the binomial model is not valid |

HOME

**AS 2018**

## The Binomial Distribution

**3.** Naasir is playing a game with two friends. The game is designed to be a game of chance so that the probability of Naasir winning each game is $\frac{1}{3}$

Naasir and his friends play the game 15 times.

(a) Find the probability that Naasir wins

(i) exactly 2 games,

(ii) more than 5 games.

**(3)**

Naasir claims he has a method to help him win more than $\frac{1}{3}$ of the games. To test this claim, the three of them played the game again 32 times and Naasir won 16 of these games.

(b) Stating your hypotheses clearly, test Naasir's claim at the 5% level of significance.

**(4)**

HOME

# AS 2018

## The Binomial Distribution

| Qu | Scheme | Marks | AO |
|---|---|---|---|
| **3** (a) | Let $N$ = the number of games Naasir wins $N \sim \mathrm{B}(15, \frac{1}{3})$ | M1 | 3.3 |
| (i) | $\mathrm{P}(N = 2) = 0.059946\ldots$ **awrt 0.0599** | A1 | 1.1b |
| (ii) | $\mathrm{P}(N > 5) = 1 - \mathrm{P}(N \leqslant 5) = 0.38162\ldots$ **awrt 0.382** | A1 | 1.1b |
| | | (3) | |
| (b) | $\mathrm{H}_0 : p = \frac{1}{3}$ $\mathrm{H}_1 : p > \frac{1}{3}$ | B1 | 2.5 |
| | Let $X$ = the number of games Naasir wins $X \sim \mathrm{B}(32, \frac{1}{3})$ | M1 | 3.3 |
| | $\mathrm{P}(X \geqslant 16) = 1 - \mathrm{P}(X \leqslant 15) = 0.03765$ $(< 0.05)$ | A1 | 3.4 |
| | [Significant result so reject $\mathrm{H}_0$ (the null model) and conclude:]<br>There is evidence to support Naasir's claim (o.e.) | A1 | 3.5a |
| | | (4) | |
| | | **(7 marks)** | |

| | Notes |
|---|---|
| (a) | M1 for selecting a binomial model with correct $n$ or $p$<br>Award for sight of $\mathrm{B}(15, \frac{1}{3})$ (o.e. e.g. in words) or implied by 1 correct answer<br>1st A1 for awrt 0.0599 (from a calculator). Allow 0.05995<br>2nd A1 for awrt 0.382 (from a calculator) |
| (b) | B1 for correctly stating both hypotheses in terms of $p$ or $\pi$<br>Accept $p = 0.\dot{3}$ or any exact equivalent. $\mathrm{H}_1 : p \geqslant \frac{1}{3}$ is B0<br>M1 for selecting a suitable model to use for the test.<br>Award for sight of $\mathrm{B}(32, \frac{1}{3})$ (o.e. e.g. in words) or implied by 0.03765<br>1st A1 for use of the model to calculate an appropriate probability using calc.<br>Sight of $\mathrm{P}(X \geqslant 16)$ **and** answer awrt 0.0377 |
| **ALT** | **CR** May use CR so award 1st A1 for CR of $X \geqslant 16$ must have seen some probabilities though: 1 of $\mathrm{P}(X \leqslant 15) = 0.9623$ or $\mathrm{P}(X \leqslant 14) = 0.9224$ or $0.9223$<br><br>2nd A1 for conclusion in context that there is support for Naasir's claim<br>Must mention "Naasir" or "his" and "claim" or "method" (o.e.)<br>or e.g. probability of winning a game is $> \frac{1}{3}$ or has increased<br>Dependent on M1 and 1st A1 but can ignore hypotheses. |
| **SC** | **Use of 0.3 for $\frac{1}{3}$**<br>If used 0.3 instead of $\frac{1}{3}$ in (a) and score M0A0A0 can condone use of 0.3 in (b)<br>1st A1 ft needs $\mathrm{P}(X \geqslant 16) = 0.0138$<br>or CR of $X \geqslant 15$ and sight of 1 of $\mathrm{P}(X \geqslant 15) = 0.0327$ or $\mathrm{P}(X \geqslant 14) = 0.0694$<br>2nd A1 as before with 0.3 instead $\frac{1}{3}$ (if appropriate) |

HOME

**5.** A biased spinner can only land on one of the numbers 1, 2, 3 or 4. The random variable $X$ represents the number that the spinner lands on after a single spin and $P(X = r) = P(X = r + 2)$ for $r = 1, 2$

Given that $P(X = 2) = 0.35$

(a) find the complete probability distribution of $X$.

**(2)**

Ambroh spins the spinner 60 times.

(b) Find the probability that more than half of the spins land on the number 4
Give your answer to 3 significant figures.

**(3)**

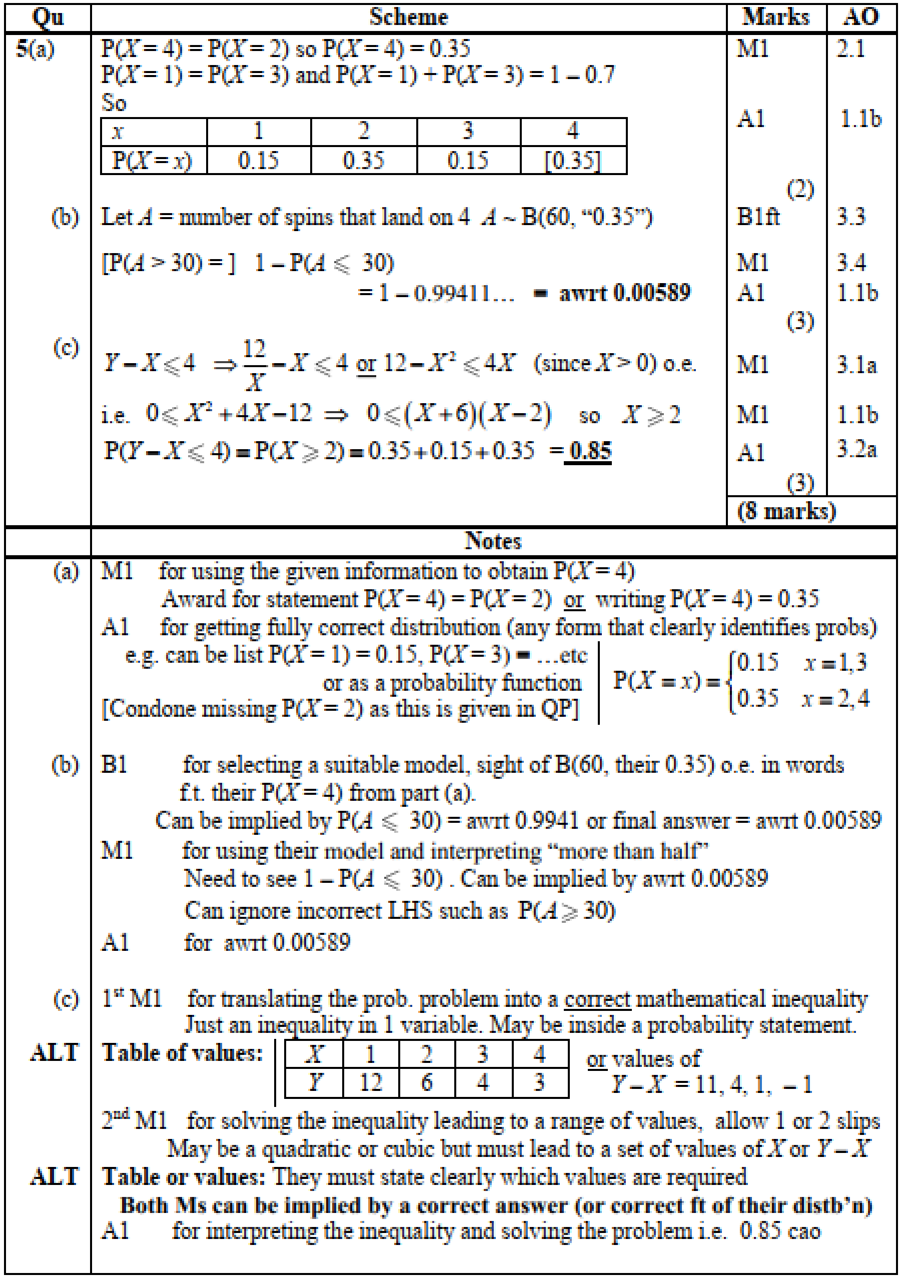
The random variable $Y = \dfrac{12}{X}$

(c) Find $P(Y - X \leqslant 4)$

**(3)**

# AS 2018

# The Binomial Distribution



| Qu | Scheme | Marks | AO |
|---|---|---|---|
| 5(a) | $P(X = 4) = P(X = 2)$ so $P(X = 4) = 0.35$ | M1 | 2.1 |
| | $P(X = 1) = P(X = 3)$ and $P(X = 1) + P(X = 3) = 1 - 0.7$ | | |
| | So | A1 | 1.1b |

| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P(X = x)$ | 0.15 | 0.35 | 0.15 | [0.35] |

| Qu | Scheme | Marks | AO |
|---|---|---|---|
| | | (2) | |
| (b) | Let $A$ = number of spins that land on 4  $A \sim B(60, \text{“}0.35\text{”})$ | B1ft | 3.3 |
| | $[P(A > 30) = ]\ \ 1 - P(A \leqslant 30)$ | M1 | 3.4 |
| | $= 1 - 0.99411\ldots$ = **awrt 0.00589** | A1 | 1.1b |
| | | (3) | |
| (c) | $Y - X \leqslant 4 \ \Rightarrow \frac{12}{X} - X \leqslant 4$ or $12 - X^2 \leqslant 4X$ (since $X > 0$) o.e. | M1 | 3.1a |
| | i.e. $0 \leqslant X^2 + 4X - 12 \ \Rightarrow\ 0 \leqslant (X+6)(X-2)$ so $X \geqslant 2$ | M1 | 1.1b |
| | $P(Y - X \leqslant 4) = P(X \geqslant 2) = 0.35 + 0.15 + 0.35 = \underline{\mathbf{0.85}}$ | A1 | 3.2a |
| | | (3) | |
| | | (8 marks) | |

| | Notes |
|---|---|
| (a) | M1 for using the given information to obtain $P(X = 4)$<br>Award for statement $P(X = 4) = P(X = 2)$ or writing $P(X = 4) = 0.35$ |
| | A1 for getting fully correct distribution (any form that clearly identifies probs)<br>e.g. can be list $P(X = 1) = 0.15$, $P(X = 3) =$ …etc<br>or as a probability function $P(X = x) = \begin{cases} 0.15 & x = 1, 3 \\ 0.35 & x = 2, 4 \end{cases}$<br>[Condone missing $P(X = 2)$ as this is given in QP] |
| (b) | B1 for selecting a suitable model, sight of B(60, their 0.35) o.e. in words<br>f.t. their $P(X = 4)$ from part (a).<br>Can be implied by $P(A \leqslant 30)$ = awrt 0.9941 or final answer = awrt 0.00589 |
| | M1 for using their model and interpreting “more than half”<br>Need to see $1 - P(A \leqslant 30)$ . Can be implied by awrt 0.00589<br>Can ignore incorrect LHS such as $P(A \geqslant 30)$ |
| | A1 for awrt 0.00589 |
| (c) | 1st M1 for translating the prob. problem into a correct mathematical inequality<br>Just an inequality in 1 variable. May be inside a probability statement. |
| ALT | **Table of values:** $\begin{array}{c\|c\|c\|c\|c} X & 1 & 2 & 3 & 4 \\ \hline Y & 12 & 6 & 4 & 3 \end{array}$ or values of $Y - X = 11, 4, 1, -1$ |
| | 2nd M1 for solving the inequality leading to a range of values, allow 1 or 2 slips<br>May be a quadratic or cubic but must lead to a set of values of $X$ or $Y - X$ |
| ALT | **Table or values:** They must state clearly which values are required |
| | **Both Ms can be implied by a correct answer (or correct ft of their distb’n)** |
| | A1 for interpreting the inequality and solving the problem i.e. 0.85 cao |

HOME

**AS 2019**

The Binomial Distribution

**3.** A fair 5-sided spinner has sides numbered 1, 2, 3, 4 and 5

The spinner is spun once and the score of the side it lands on is recorded.

(a) Write down the name of the distribution that can be used to model the score of the side it lands on.

**(1)**

The spinner is spun 28 times.

The random variable $X$ represents the number of times the spinner lands on 2

(b) (i) Find the probability that the spinner lands on 2 at least 7 times.

(ii) Find $\mathrm{P}(4 \leqslant X < 8)$

**(5)**

HOME

AS 2019

## The Binomial Distribution

**Question 3 (Total 6 marks)**

| Part | Working or answer an examiner might expect to see | Mark | Notes |
|---|---|---|---|
| (a) | Discrete uniform distribution | B1 | This mark is given for the correct distribution |
| (b)(i) | $X \sim \mathrm{B}(28, 0.2)$ | B1 | This mark is given for identifying the correct model |
| | $\mathrm{P}(X \geq 7) = 1 - \mathrm{P}(X \leq 6)$<br>$= 1 - 0.6784$ | M1 | This mark is given for a method to find the probability |
| | $= 0.322$ | A1 | This mark is given for finding the correct probability |
| (b)(ii) | $\mathrm{P}(4 \leq X < 8) = \mathrm{P}(X \leq 7) - \mathrm{P}(X \leq 3)$<br>$= 0.818 - 0.160$ | M1 | This mark is given for a method to find the probability |
| | $= 0.658$ | A1 | This mark is given for finding the correct probability |

HOME

# AS 2019

## The Binomial Distribution

**5.** Past records show that 15% of customers at a shop buy chocolate. The shopkeeper believes that moving the chocolate closer to the till will increase the proportion of customers buying chocolate.

After moving the chocolate closer to the till, a random sample of 30 customers is taken and 8 of them are found to have bought chocolate.

Julie carries out a hypothesis test, at the 5% level of significance, to test the shopkeeper's belief.

Julie's hypothesis test is shown below.

$H_0 : p = 0.15$

$H_1 : p \geqslant 0.15$

Let $X$ = the number of customers who buy chocolate.

$X \sim B(30, 0.15)$

$P(X = 8) = 0.0420$

$0.0420 < 0.05$ so reject $H_0$

There is sufficient evidence to suggest that the proportion of customers buying chocolate has increased.

(a) Identify the first two errors that Julie has made in her hypothesis test.
**(2)**

(b) Explain whether or not these errors will affect the conclusion of her hypothesis test. Give a reason for your answer.
**(1)**

(c) Find, using a 5% level of significance, the critical region for a one-tailed test of the shopkeeper's belief. The probability in the tail should be less than 0.05
**(2)**

(d) Find the actual level of significance of this test.
**(1)**

HOME

# AS 2019

## The Binomial Distribution

**Question 5 (Total 6 marks)**

| Part | Working or answer an examiner might expect to see | Mark | Notes |
|---|---|---|---|
| (a) | $H_1$: $p \geq 0$ is incorrect; it should be $H_1$: $p > 0$ | B1 | This mark is given for identifying the error in the alternative hypothesis |
| | The calculation of the test statistic $P(X = 8)$ is incorrect; it should be $P(X \geq 8)$ | B1 | This mark is given for identifying the error in the test statistic |
| (b) | The errors will affect the conclusion as the null hypothesis should not be rejected since $P(X \geq 8)$ [= 0.0698] is greater than 0.05 | B1 | This mark is given for a correct explanation |
| (c) | $P(X \leq 8) = 0.9722 > 0.95$ or<br>$P(X \geq 9) = 0.0278 < 0.05$ | M1 | This mark is given for the use of tables or calculator to find the probability associated with the critical value with B(30. 0.15) |
| | Critical region: $\{X \geq 9\}$ | A1 | This mark is given for finding the correct critical region |
| (d) | 0.0278 | B1 | This mark is given for finding the correct level of significance of the test |

HOME

## AS 2020

## The Binomial Distribution

**5.** Afrika works in a call centre.

She assumes that calls are independent and knows, from past experience, that on each sales call that she makes there is a probability of $\frac{1}{6}$ that it is successful.

Afrika makes 9 sales calls.

(a) Calculate the probability that at least 3 of these sales calls will be successful.

**(2)**

The probability of Afrika making a successful sales call is the same each day.

Afrika makes 9 sales calls on each of 5 different days.

(b) Calculate the probability that at least 3 of the sales calls will be successful on exactly 1 of these days.

**(2)**

Rowan works in the same call centre as Afrika and believes he is a more successful salesperson.

To check Rowan's belief, Afrika monitors the next 35 sales calls Rowan makes and finds that 11 of the sales calls are successful.

(c) Stating your hypotheses clearly test, at the 5% level of significance, whether or not there is evidence to support Rowan's belief.

**(4)**

HOME

# The Binomial Distribution

| Question | Scheme | Marks | AOs |
|---|---|---|---|
| 5(a) | Let $C$ = the number of successful calls. $C$ □ $\mathrm{B}\left(9,\frac{1}{6}\right)$ | M1 | 3.3 |
| | $\mathrm{P}(C\geq 3)=1-\mathrm{P}(C\leq 2)=0.1782\ldots$ **awrt 0.178** | A1 | 1.1b |
| | | (2) | |
| (b) | Let $X$ = the number of occasions when at least 3 calls are successful.<br>$\mathrm{P}(X=1)=5\times(\text{"}0.1782\ldots\text{"})\times(\text{"}0.8217\ldots\text{"})^4$ | M1 | 1.1b |
| | $=0.4061\ldots$ **awrt 0.406** | A1 | 1.1b |
| | | (2) | |
| (c) | $\mathrm{H}_0: p=\frac{1}{6}$ $\quad \mathrm{H}_1: p>\frac{1}{6}$ | B1 | 2.5 |
| | Let $R$ = the number of successful calls $R$ □ $\mathrm{B}\left(35,\frac{1}{6}\right)$ | M1 | 3.3 |
| | $\mathrm{P}(R\geq 11)=1-\mathrm{P}(R\leq 10)=0.02\ldots$ | A1 | 3.4 |
| | There is sufficient evidence to support that **Rowan** has more successful sales calls than Afrika. | A1 | 2.2b |
| | | (4) | |
| | | | **(8 marks)** |

**Notes**

| | | |
|---|---|---|
| **5(a)** | **M1:** | For selecting the right model |
| | **A1:** | awrt 0.178 |
| **(b)** | **M1:** | For $5\times(\text{"their}(a)\text{"})\times(\text{"}1-\text{their}(a)\text{"})^4$ |
| | **A1:** | awrt 0.406 |
| **(c)** | **B1:** | for correctly stating both hypotheses in terms of $p$ or $\pi$ Accept $p=0.1\dot{6}$ |
| | **M1:** | For selecting a suitable model. May be implied by a correct probability or CR |
| | **A1:** | Correct probability statement and answer of 0.02 or better (0.02318…)<br>(CR $R\geq 11$ and either $\mathrm{P}(R\leq 9)=0.9450$ or $\mathrm{P}(R\leq 10)=0.9768$ or $1-\mathrm{P}(R\leq 10)=0.0232$) |
| | **A1:** | Dependent on M1A1 but can ignore hypotheses. For conclusion in context supporting **Rowan's** belief / **Rowan** is a better sales person |
| | | Do not accept Rowan can reject H₀ |

HOME

AS 2021

The Binomial Distribution

**4.** A nursery has a sack containing a large number of coloured beads of which 14% are coloured red.

Aliya takes a random sample of 18 beads from the sack to make a bracelet.

(a) State a suitable binomial distribution to model the number of red beads in Aliya's bracelet.

**(1)**

(b) Use this binomial distribution to find the probability that

(i) Aliya has just 1 red bead in her bracelet,

(ii) there are at least 4 red beads in Aliya's bracelet.

**(3)**

(c) Comment on the suitability of a binomial distribution to model this situation.

**(1)**

After several children have used beads from the sack, the nursery teacher decides to test whether or not the proportion of red beads in the sack has changed.
She takes a random sample of 75 beads and finds 4 red beads.

(d) Stating your hypotheses clearly, use a 5% significance level to carry out a suitable test for the teacher.

**(4)**

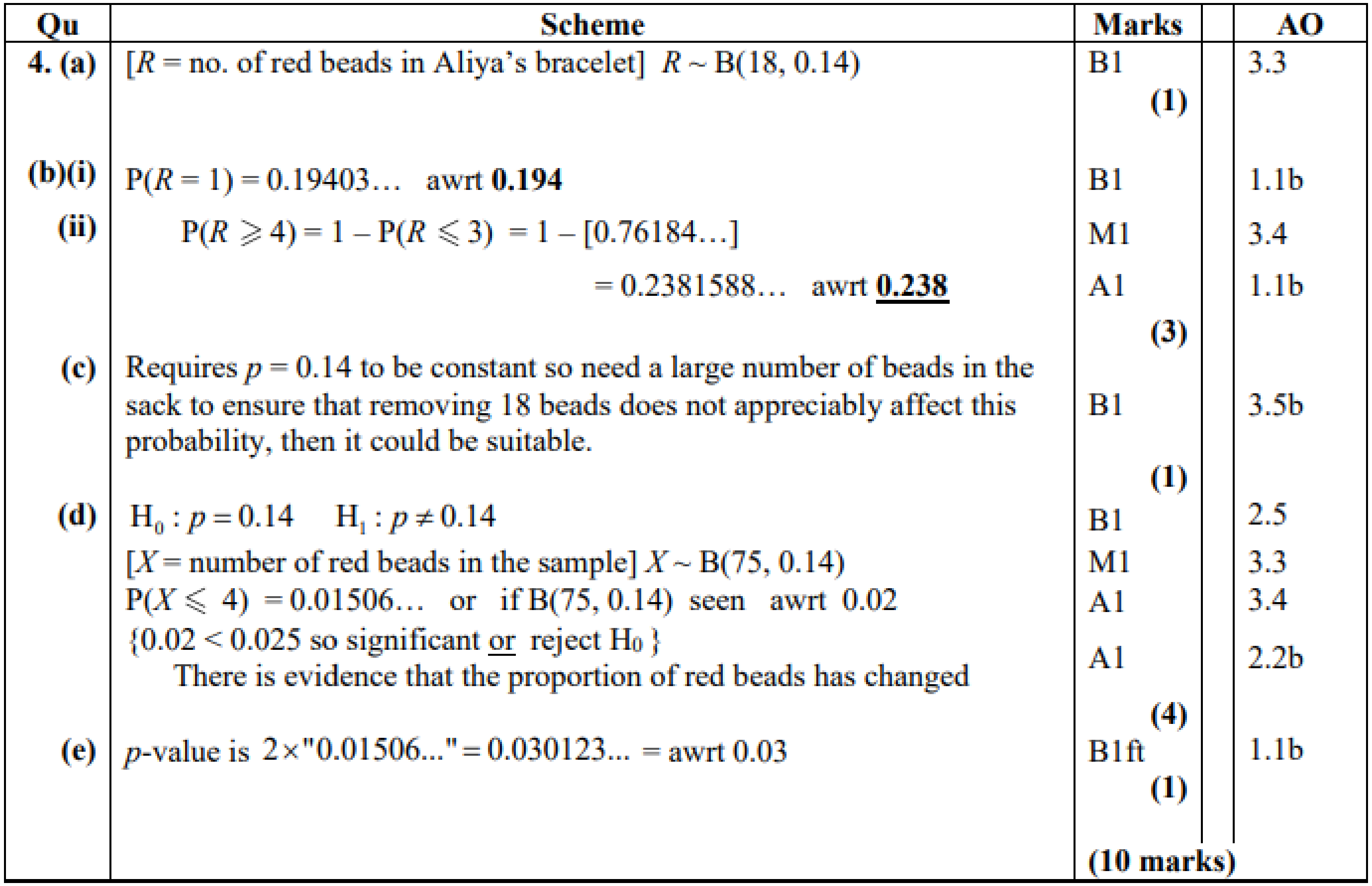
(e) Find the *p*-value in this case.

**(1)**

HOME

# AS 2021

## The Binomial Distribution



| Qu | Scheme | Marks | | AO |
|---|---|---|---|---|
| 4. (a) | [$R$ = no. of red beads in Aliya's bracelet] $R \sim \mathrm{B}(18, 0.14)$ | B1 | | 3.3 |
| | | (1) | | |
| (b)(i) | $\mathrm{P}(R = 1) = 0.19403\ldots$ awrt **0.194** | B1 | | 1.1b |
| (ii) | $\mathrm{P}(R \geqslant 4) = 1 - \mathrm{P}(R \leqslant 3) = 1 - [0.76184\ldots]$ | M1 | | 3.4 |
| | $= 0.2381588\ldots$ awrt **0.238** | A1 | | 1.1b |
| | | (3) | | |
| (c) | Requires $p = 0.14$ to be constant so need a large number of beads in the sack to ensure that removing 18 beads does not appreciably affect this probability, then it could be suitable. | B1 | | 3.5b |
| | | (1) | | |
| (d) | $\mathrm{H}_0 : p = 0.14 \quad \mathrm{H}_1 : p \neq 0.14$ | B1 | | 2.5 |
| | [$X$ = number of red beads in the sample] $X \sim \mathrm{B}(75, 0.14)$ | M1 | | 3.3 |
| | $\mathrm{P}(X \leqslant 4) = 0.01506\ldots$ or if $\mathrm{B}(75, 0.14)$ seen awrt 0.02 | A1 | | 3.4 |
| | {$0.02 < 0.025$ so significant or reject $\mathrm{H}_0$}<br>There is evidence that the proportion of red beads has changed | A1 | | 2.2b |
| | | (4) | | |
| (e) | $p$-value is $2 \times "0.01506..." = 0.030123...$ = awrt 0.03 | B1ft | | 1.1b |
| | | (1) | | |
| | | (10 marks) | | |

HOME

**AS 2022**

**The Binomial Distribution**

**2.** A manufacturer of sweets knows that 8% of the bags of sugar delivered from supplier *A* will be damp.
A random sample of 35 bags of sugar is taken from supplier *A*.

(a) Using a suitable model, find the probability that the number of bags of sugar that are damp is

(i) exactly 2

(ii) more than 3

**(3)**

Supplier *B* claims that when it supplies bags of sugar, the proportion of bags that are damp is less than 8%

The manufacturer takes a random sample of 70 bags of sugar from supplier *B* and finds that only 2 of the bags are damp.

(b) Carry out a suitable test to assess supplier *B*'s claim.
You should state your hypotheses clearly and use a 10% level of significance.

**(4)**

HOME

# AS 2022

## The Binomial Distribution

| Qu | Scheme | Mark | AO |
|---|---|---|---|
| 2. (a) | [$D$ = number of bags that are damp] $D \sim \mathrm{B}(35, 0.08)$ NB $0.08 = \frac{2}{25}$ | M1 | 3.3 |
| (i) | $\mathrm{P}(D = 2) = 0.2430497\ldots$ awrt **0.243** | A1 | 3.4 |
| (ii) | $\mathrm{P}(D > 3) = [1 - \mathrm{P}(D \leqslant 3) = 1 - 0.69397\ldots] = 0.30602\ldots$ awrt **0.306** | A1 | 1.1b |
| | | (3) | |
| (b) | $\mathrm{H}_0 : p = 0.08$ $\mathrm{H}_1 : p < 0.08$ | B1 | 2.5 |
| | [$X \sim$] $\mathrm{B}(70, 0.08)$ | M1 | 2.1 |
| | $[\mathrm{P}(X \leqslant 2)] = 0.0739756\ldots$ awrt **0.074** | A1 | 1.1b |
| | [$0.074 < 0.10$ so significant, reject $\mathrm{H}_0$ so…] | | |
| | there is evidence to support supplier *B*'s claim (o.e.) | A1 | 2.2b |
| | | (4) | |
| | | (7 marks) | |

HOME

**AS 2023**

**The Binomial Distribution**

**4.** Past information shows that 25% of adults in a large population have a particular allergy.

Rylan believes that the proportion that has the allergy differs from 25%

He takes a random sample of 50 adults from the population.

Rylan carries out a test of the null hypothesis $H_0$: $p = 0.25$ using a 5% level of significance.

(a) Write down the alternative hypothesis for Rylan's test.

**(1)**

(b) Find the critical region for this test.
You should state the probability associated with each tail, which should be as close to 2.5% as possible.

**(4)**

(c) State the actual probability of incorrectly rejecting $H_0$ for this test.

**(1)**

Rylan finds that 10 of the adults in his sample have the allergy.

(d) State the conclusion of Rylan's hypothesis test.

**(1)**

HOME

**AS 2023**

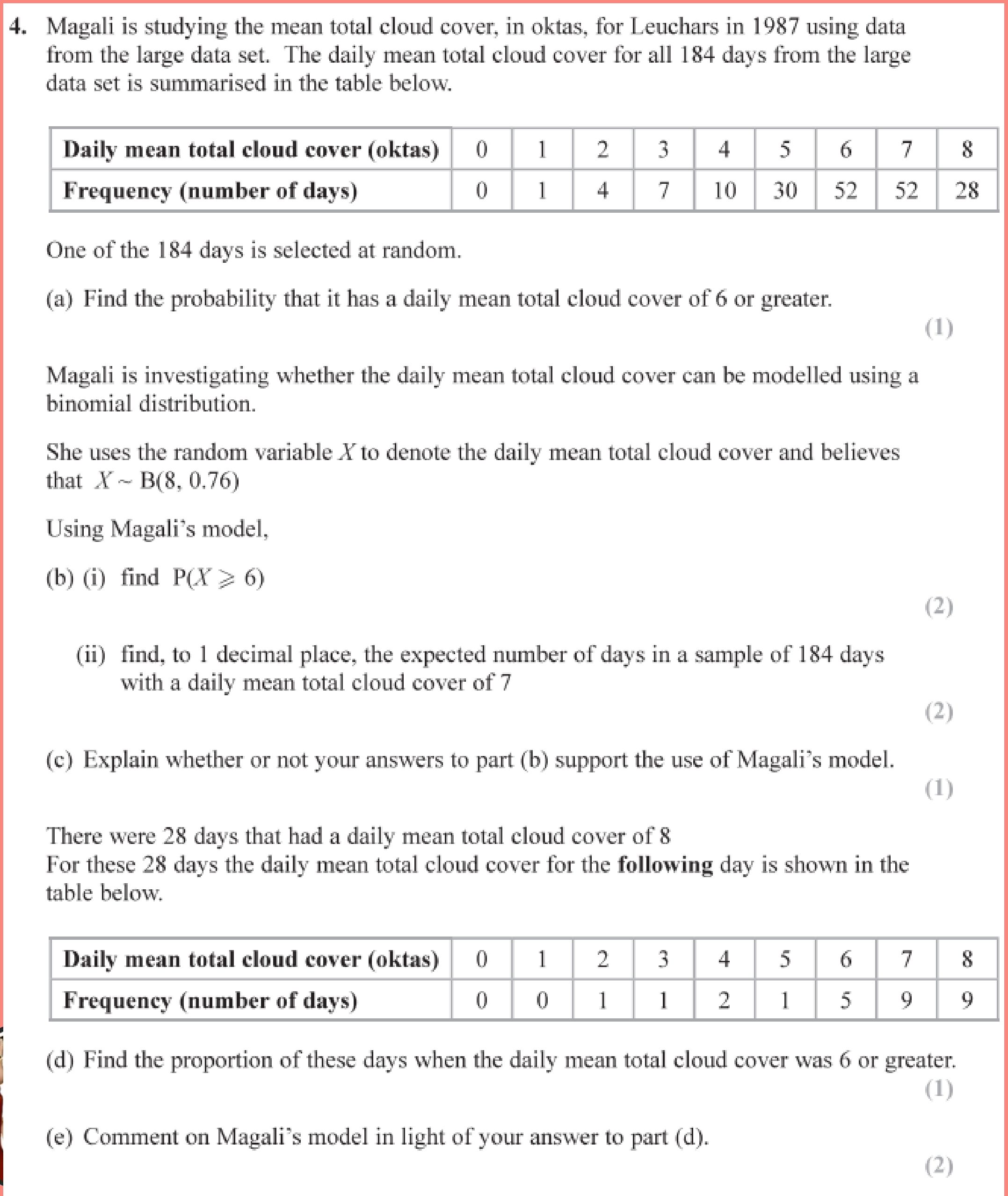
## The Binomial Distribution

| | | | |
|---|---|---|---|
| **4(a)** | [$H_1$ :] $p \neq 0.25$ | B1 | 2.5 |
| | | **(1)** | |
| **(b)** | $X$~B(50, 0.25) | B1 | 3.3 |
| | [P($X \leqslant 6$) =]0.0194 or [P($X \leqslant 18$) =]0.9713 or<br>[P($X \geqslant 19$) =]0.0287<br>or $X \leqslant 6$ or $X \geqslant 19$ | M1 | 3.4 |
| | [P($X \leqslant 6$) =]awrt 0.0194 and [P($X \geqslant 19$) =]awrt 0.0287 | A1 | 1.1b |
| | CR: $X \leqslant 6$ or $X \geqslant 19$ | A1 | 1.1b |
| | | **(4)** | |
| **(c)** | [0.0194 + 0.0287 =] awrt 0.048 | B1ft | 1.1b |
| | | **(1)** | |
| **(d)** | (Do not reject $H_0$,) there is insufficient evidence to suggest that the **proportion** of those with the **allergy** differs from 25%/**Rylan's belief** not supported | B1 | 2.2b |
| | | **(1)** | |
| | | **(7 marks)** | |

HOME

# A2 2019

## The Binomial Distribution



**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 $P(X \geqslant 6)$

(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)

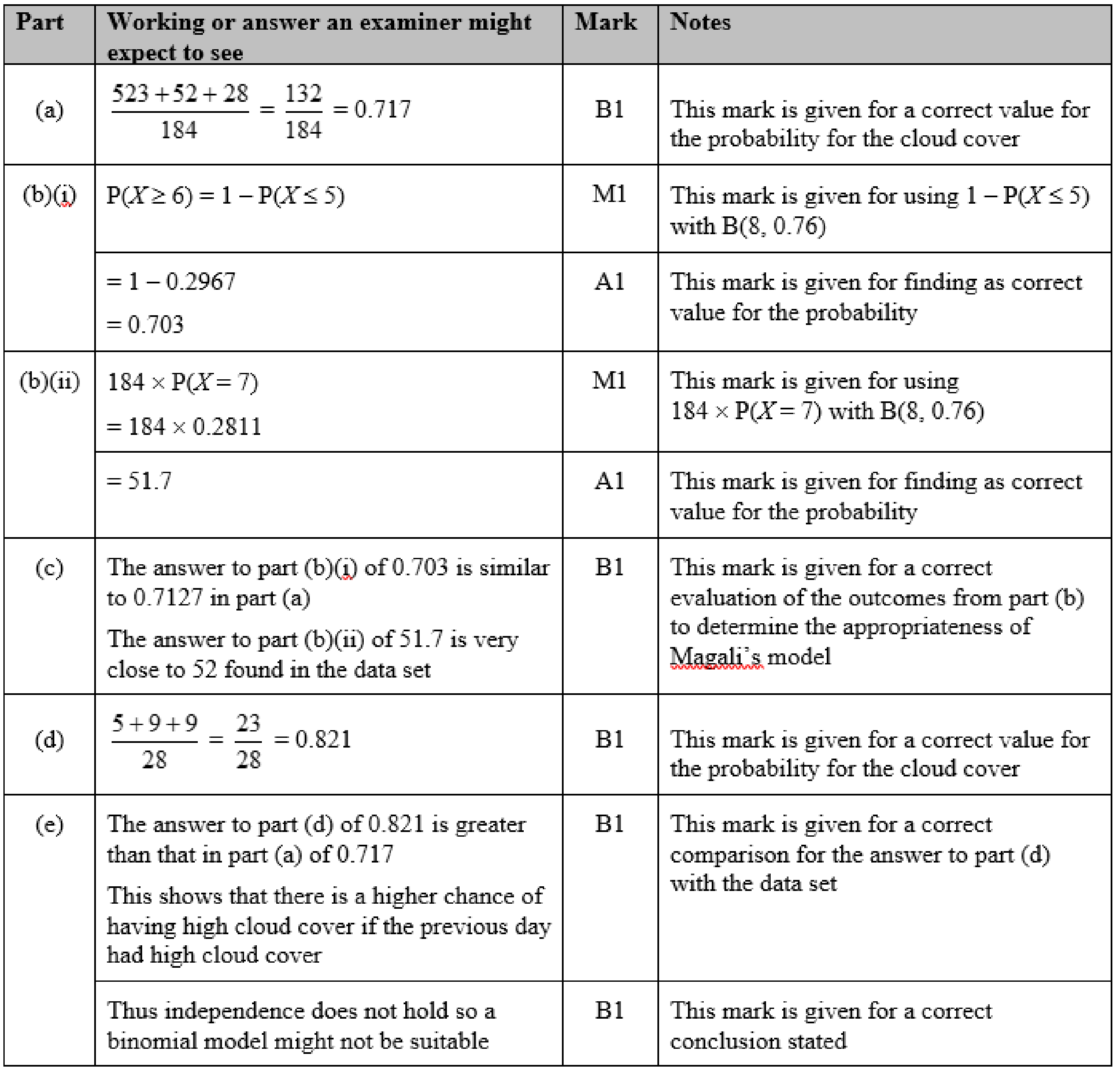
(e) Comment on Magali's model in light of your answer to part (d).

(2)

HOME

# A2 2019

The Binomial Distribution

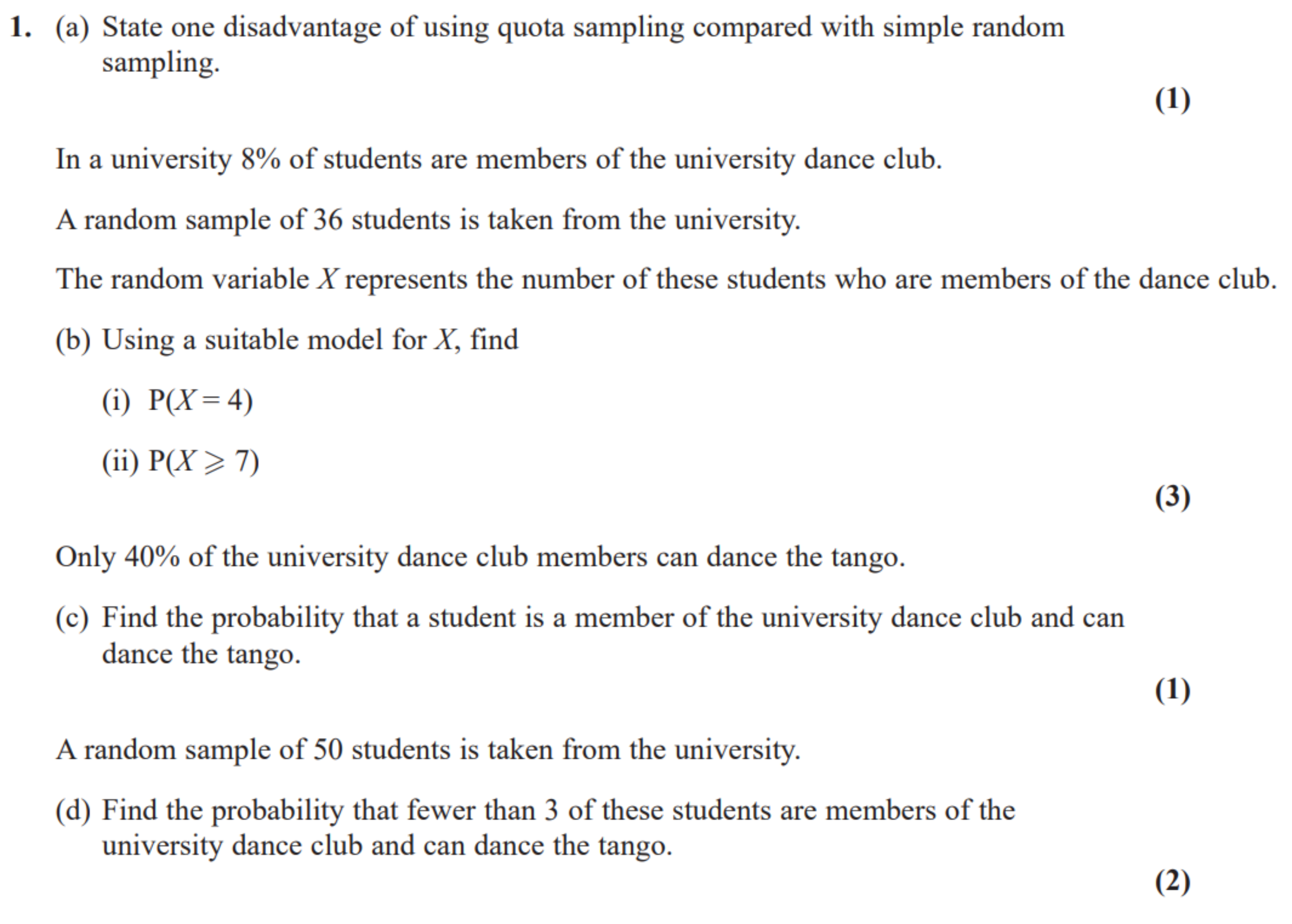


| Part | Working or answer an examiner might expect to see | Mark | Notes |
|---|---|---|---|
| (a) | $\dfrac{523+52+28}{184} = \dfrac{132}{184} = 0.717$ | B1 | This mark is given for a correct value for the probability for the cloud cover |
| (b)(i) | $\mathrm{P}(X \geq 6) = 1 - \mathrm{P}(X \leq 5)$ | M1 | This mark is given for using $1 - \mathrm{P}(X \leq 5)$ with B(8, 0.76) |
| | $= 1 - 0.2967$<br>$= 0.703$ | A1 | This mark is given for finding as correct value for the probability |
| (b)(ii) | $184 \times \mathrm{P}(X = 7)$<br>$= 184 \times 0.2811$ | M1 | This mark is given for using $184 \times \mathrm{P}(X = 7)$ with B(8, 0.76) |
| | $= 51.7$ | A1 | This mark is given for finding as correct value for the probability |
| (c) | The answer to part (b)(i) of 0.703 is similar to 0.7127 in part (a)<br>The answer to part (b)(ii) of 51.7 is very close to 52 found in the data set | B1 | This mark is given for a correct evaluation of the outcomes from part (b) to determine the appropriateness of Magali's model |
| (d) | $\dfrac{5+9+9}{28} = \dfrac{23}{28} = 0.821$ | B1 | This mark is given for a correct value for the probability for the cloud cover |
| (e) | The answer to part (d) of 0.821 is greater than that in part (a) of 0.717<br>This shows that there is a higher chance of having high cloud cover if the previous day had high cloud cover | B1 | This mark is given for a correct comparison for the answer to part (d) with the data set |
| | Thus independence does not hold so a binomial model might not be suitable | B1 | This mark is given for a correct conclusion stated |

HOME

A2 2021

## The Binomial Distribution



1. (a) State one disadvantage of using quota sampling compared with simple random sampling.

**(1)**

In a university 8% of students are members of the university dance club.

A random sample of 36 students is taken from the university.

The random variable $X$ represents the number of these students who are members of the dance club.

(b) Using a suitable model for $X$, find

(i) $\mathrm{P}(X = 4)$

(ii) $\mathrm{P}(X \geqslant 7)$

**(3)**

Only 40% of the university dance club members can dance the tango.

(c) Find the probability that a student is a member of the university dance club and can dance the tango.

**(1)**

A random sample of 50 students is taken from the university.

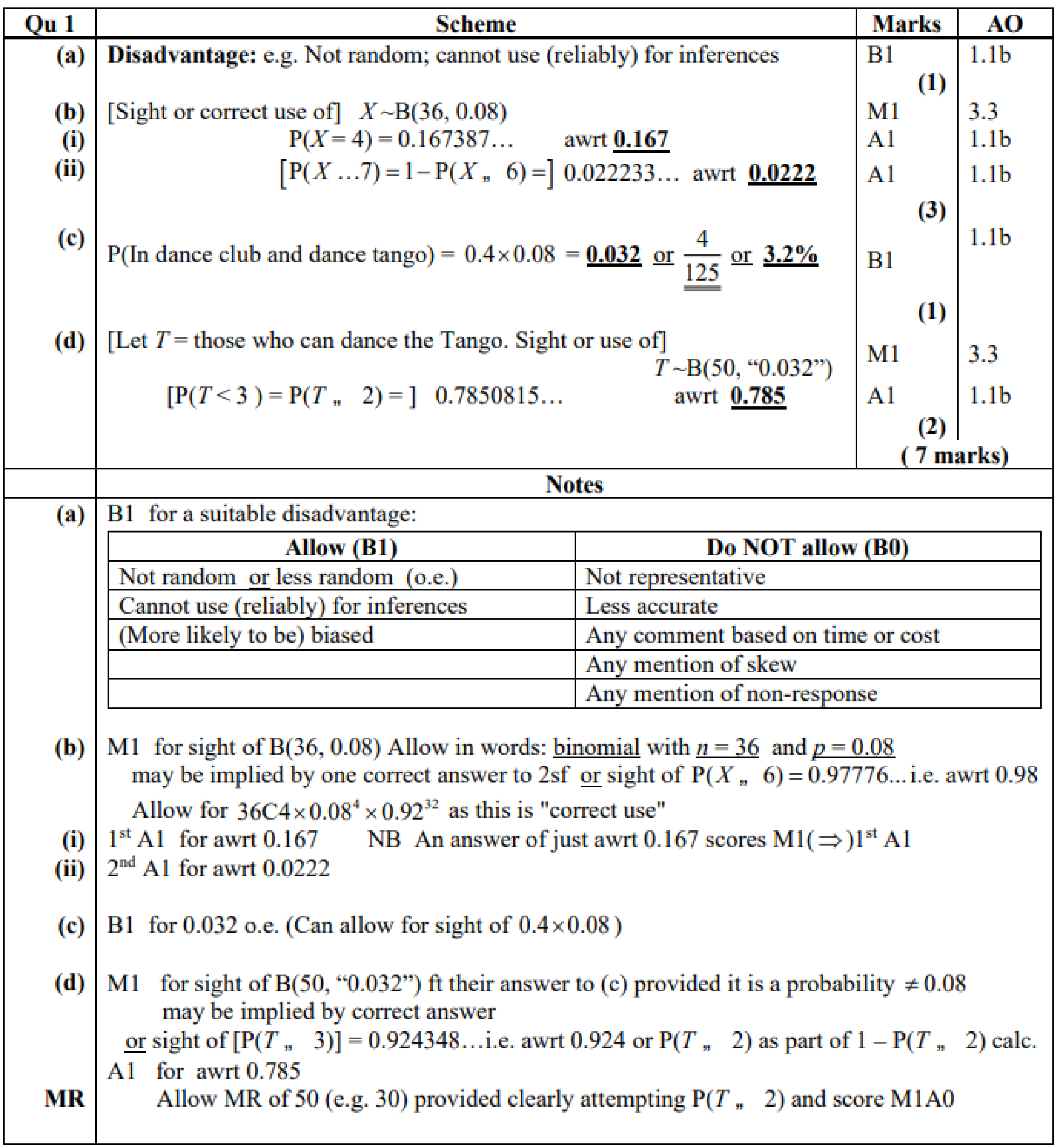
(d) Find the probability that fewer than 3 of these students are members of the university dance club and can dance the tango.

**(2)**

HOME

A2 2021

# The Binomial Distribution



| Qu 1 | Scheme | Marks | AO |
|---|---|---|---|
| (a) | **Disadvantage:** e.g. Not random; cannot use (reliably) for inferences | B1<br>**(1)** | 1.1b |
| (b) | [Sight or correct use of] $X \sim B(36, 0.08)$ | M1 | 3.3 |
| (i) | P($X$ = 4) = 0.167387… awrt **0.167** | A1 | 1.1b |
| (ii) | [P($X$ …7) = 1 − P($X$ „ 6) =] 0.022233… awrt **0.0222** | A1<br>**(3)** | 1.1b |
| (c) | P(In dance club and dance tango) = 0.4×0.08 = **0.032** or $\frac{4}{125}$ or **3.2%** | B1<br>**(1)** | 1.1b |
| (d) | [Let $T$ = those who can dance the Tango. Sight or use of]<br>$T \sim$B(50, "0.032") | M1 | 3.3 |
| | [P($T$ < 3 ) = P($T$ „ 2) = ] 0.7850815… awrt **0.785** | A1<br>**(2)** | 1.1b |
| | | **( 7 marks)** | |

**Notes**

**(a)** B1 for a suitable disadvantage:

| Allow (B1) | Do NOT allow (B0) |
|---|---|
| Not random or less random (o.e.) | Not representative |
| Cannot use (reliably) for inferences | Less accurate |
| (More likely to be) biased | Any comment based on time or cost |
| | Any mention of skew |
| | Any mention of non-response |

**(b)** M1 for sight of B(36, 0.08) Allow in words: binomial with $n = 36$ and $p = 0.08$
may be implied by one correct answer to 2sf or sight of P($X$ „ 6) = 0.97776... i.e. awrt 0.98
Allow for $36C4 \times 0.08^4 \times 0.92^{32}$ as this is "correct use"

**(i)** 1st A1 for awrt 0.167 NB An answer of just awrt 0.167 scores M1($\Rightarrow$)1st A1

**(ii)** 2nd A1 for awrt 0.0222

**(c)** B1 for 0.032 o.e. (Can allow for sight of $0.4 \times 0.08$ )

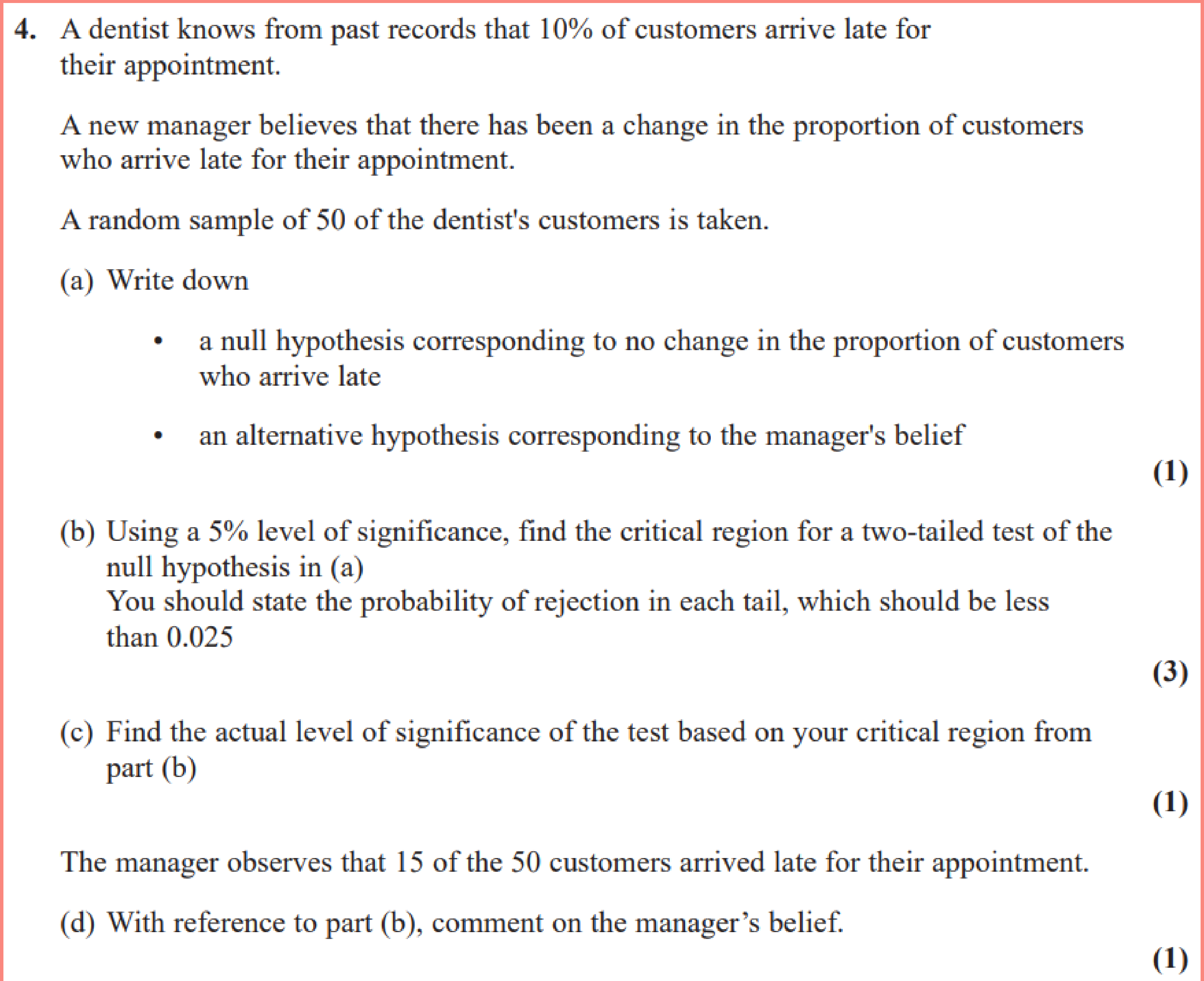
**(d)** M1 for sight of B(50, "0.032") ft their answer to (c) provided it is a probability $\neq 0.08$
may be implied by correct answer
or sight of [P($T$ „ 3)] = 0.924348…i.e. awrt 0.924 or P($T$ „ 2) as part of 1 – P($T$ „ 2) calc.
A1 for awrt 0.785

**MR** Allow MR of 50 (e.g. 30) provided clearly attempting P($T$ „ 2) and score M1A0

HOME

**A2 2022**

## The Binomial Distribution



4. A dentist knows from past records that 10% of customers arrive late for their appointment.

A new manager believes that there has been a change in the proportion of customers who arrive late for their appointment.

A random sample of 50 of the dentist's customers is taken.

(a) Write down

- a null hypothesis corresponding to no change in the proportion of customers who arrive late
- an alternative hypothesis corresponding to the manager's belief

**(1)**

(b) Using a 5% level of significance, find the critical region for a two-tailed test of the null hypothesis in (a)
You should state the probability of rejection in each tail, which should be less than 0.025

**(3)**

(c) Find the actual level of significance of the test based on your critical region from part (b)

**(1)**

The manager observes that 15 of the 50 customers arrived late for their appointment.

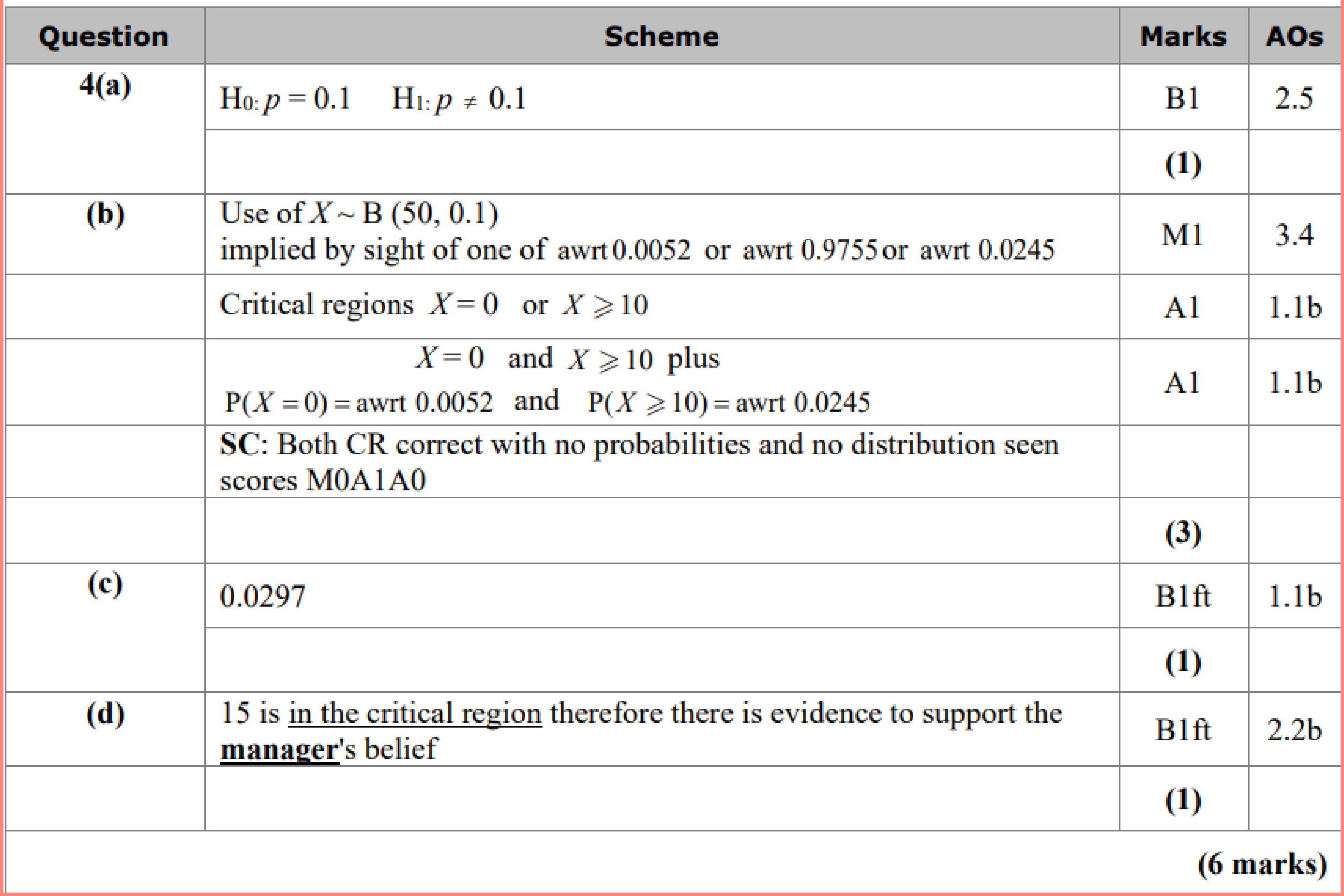
(d) With reference to part (b), comment on the manager's belief.

**(1)**

HOME

# A2 2022

## The Binomial Distribution

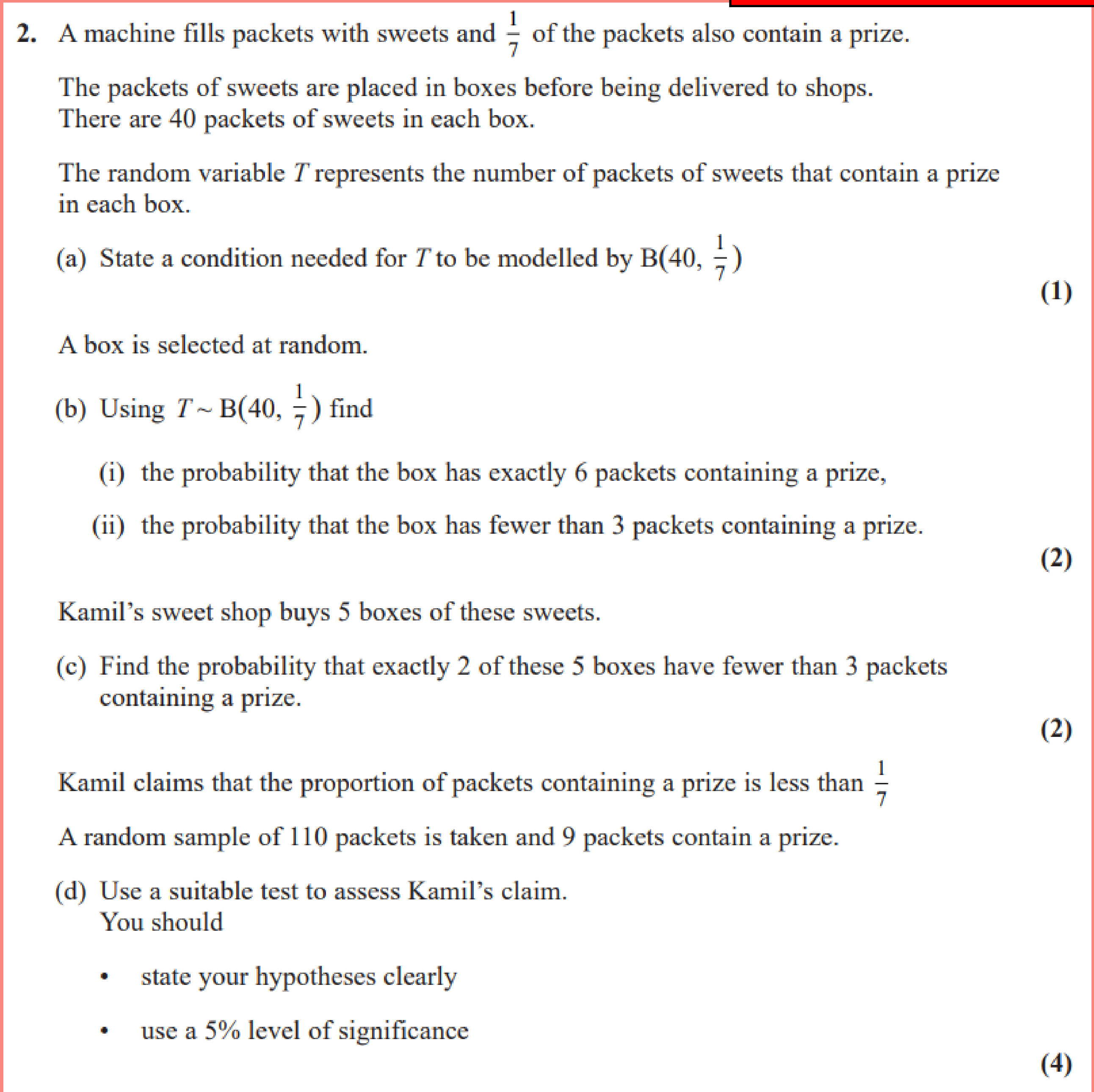


| Question | Scheme | Marks | AOs |
|---|---|---|---|
| 4(a) | $H_0: p = 0.1 \quad H_1: p \neq 0.1$ | B1 | 2.5 |
| | | **(1)** | |
| **(b)** | Use of $X \sim B(50, 0.1)$<br>implied by sight of one of awrt 0.0052 or awrt 0.9755 or awrt 0.0245 | M1 | 3.4 |
| | Critical regions $X = 0$ or $X \geqslant 10$ | A1 | 1.1b |
| | $X = 0$ and $X \geqslant 10$ plus<br>$P(X = 0) =$ awrt 0.0052 and $P(X \geqslant 10) =$ awrt 0.0245 | A1 | 1.1b |
| | **SC**: Both CR correct with no probabilities and no distribution seen scores M0A1A0 | | |
| | | **(3)** | |
| **(c)** | 0.0297 | B1ft | 1.1b |
| | | **(1)** | |
| **(d)** | 15 is in the critical region therefore there is evidence to support the **manager**'s belief | B1ft | 2.2b |
| | | **(1)** | |
| | | | **(6 marks)** |

HOME

**A2 2023**

## The Binomial Distribution



**2.** A machine fills packets with sweets and $\frac{1}{7}$ of the packets also contain a prize.

The packets of sweets are placed in boxes before being delivered to shops.
There are 40 packets of sweets in each box.

The random variable $T$ represents the number of packets of sweets that contain a prize in each box.

(a) State a condition needed for $T$ to be modelled by $\mathrm{B}(40, \frac{1}{7})$

**(1)**

A box is selected at random.

(b) Using $T \sim \mathrm{B}(40, \frac{1}{7})$ find

(i) the probability that the box has exactly 6 packets containing a prize,

(ii) the probability that the box has fewer than 3 packets containing a prize.

**(2)**

Kamil's sweet shop buys 5 boxes of these sweets.

(c) Find the probability that exactly 2 of these 5 boxes have fewer than 3 packets containing a prize.

**(2)**

Kamil claims that the proportion of packets containing a prize is less than $\frac{1}{7}$

A random sample of 110 packets is taken and 9 packets contain a prize.

(d) Use a suitable test to assess Kamil's claim.
You should

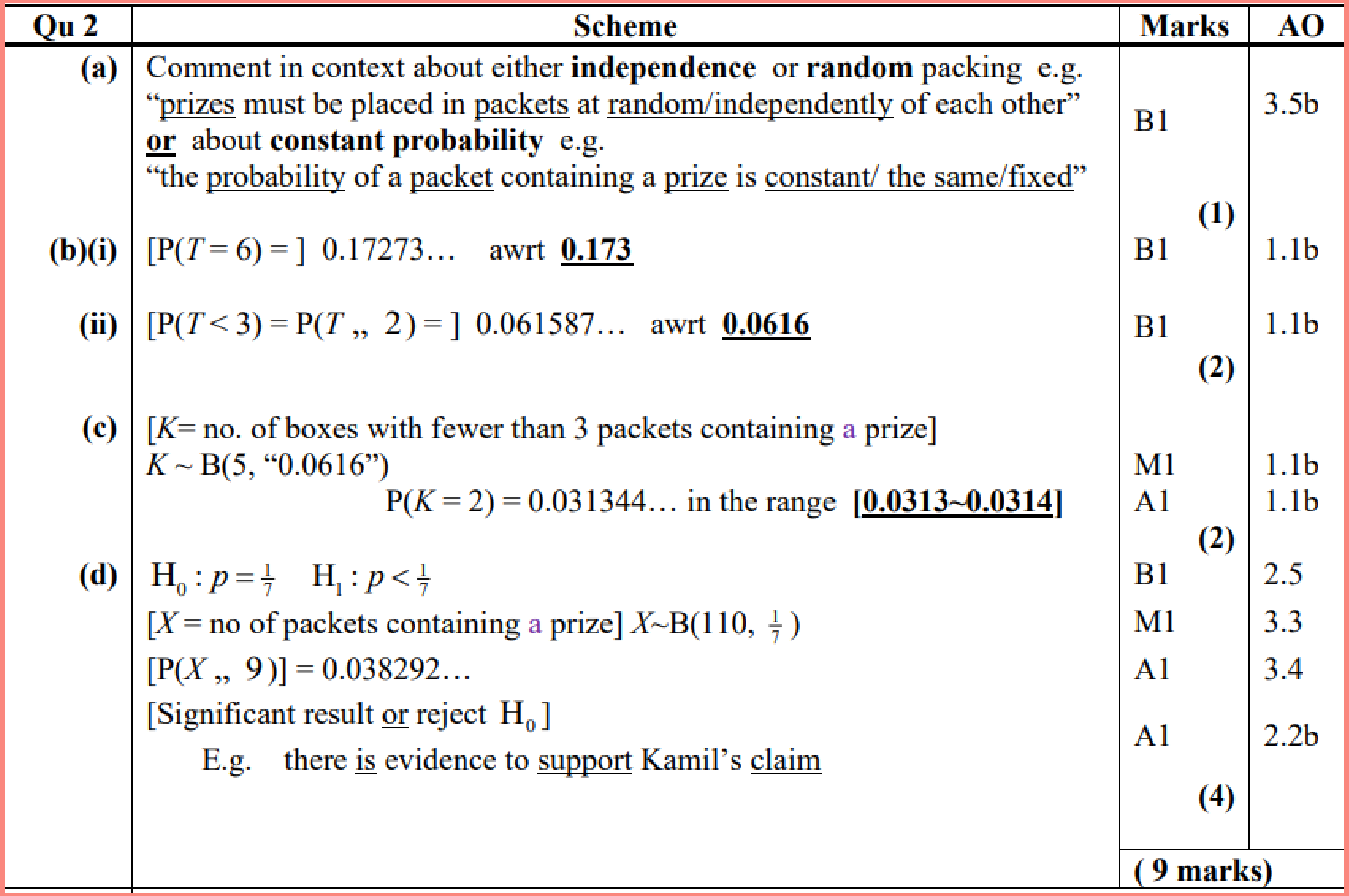
- state your hypotheses clearly
- use a 5% level of significance

**(4)**

HOME

# A2 2023

## The Binomial Distribution



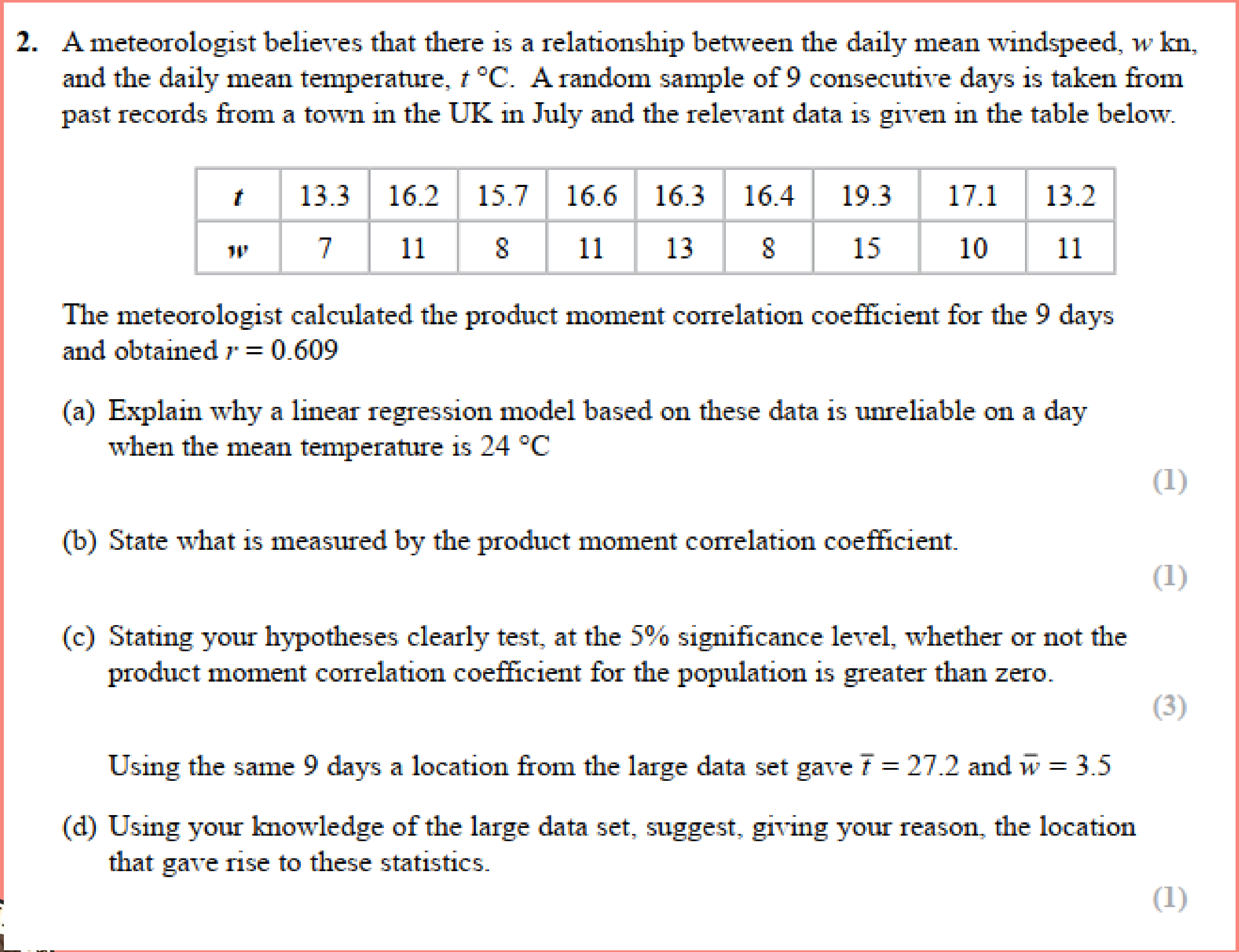
| Qu 2 | Scheme | Marks | AO |
|---|---|---|---|
| **(a)** | Comment in context about either **independence** or **random** packing e.g.<br>"prizes must be placed in packets at random/independently of each other"<br>**or** about **constant probability** e.g.<br>"the probability of a packet containing a prize is constant/ the same/fixed" | B1<br>**(1)** | 3.5b |
| **(b)(i)** | [P($T = 6$) = ] 0.17273… awrt **0.173** | B1 | 1.1b |
| **(ii)** | [P($T < 3$) = P($T$ ,, 2) = ] 0.061587… awrt **0.0616** | B1<br>**(2)** | 1.1b |
| **(c)** | [$K$= no. of boxes with fewer than 3 packets containing a prize]<br>$K \sim$ B(5, "0.0616") | M1 | 1.1b |
| | P($K = 2$) = 0.031344… in the range **[0.0313~0.0314]** | A1<br>**(2)** | 1.1b |
| **(d)** | $\text{H}_0 : p = \frac{1}{7}$ $\text{H}_1 : p < \frac{1}{7}$ | B1 | 2.5 |
| | [$X$ = no of packets containing a prize] $X$~B(110, $\frac{1}{7}$ ) | M1 | 3.3 |
| | [P($X$ ,, 9)] = 0.038292… | A1 | 3.4 |
| | [Significant result or reject $\text{H}_0$]<br>E.g. there is evidence to support Kamil's claim | A1<br>**(4)** | 2.2b |
| | | **( 9 marks)** | |

HOME

# Regression and Correlation

# A2 SAMs

## Regression and Correlation



**2.** A meteorologist believes that there is a relationship between the daily mean windspeed, $w$ kn, and the daily mean temperature, $t$ °C. A random sample of 9 consecutive days is taken from past records from a town in the UK in July and the relevant data is given in the table below.

| $t$ | 13.3 | 16.2 | 15.7 | 16.6 | 16.3 | 16.4 | 19.3 | 17.1 | 13.2 |
|---|---|---|---|---|---|---|---|---|---|
| $w$ | 7 | 11 | 8 | 11 | 13 | 8 | 15 | 10 | 11 |

The meteorologist calculated the product moment correlation coefficient for the 9 days and obtained $r = 0.609$

(a) Explain why a linear regression model based on these data is unreliable on a day when the mean temperature is 24 °C

(1)

(b) State what is measured by the product moment correlation coefficient.

(1)

(c) Stating your hypotheses clearly test, at the 5% significance level, whether or not the product moment correlation coefficient for the population is greater than zero.

(3)

Using the same 9 days a location from the large data set gave $\bar{t} = 27.2$ and $\bar{w} = 3.5$

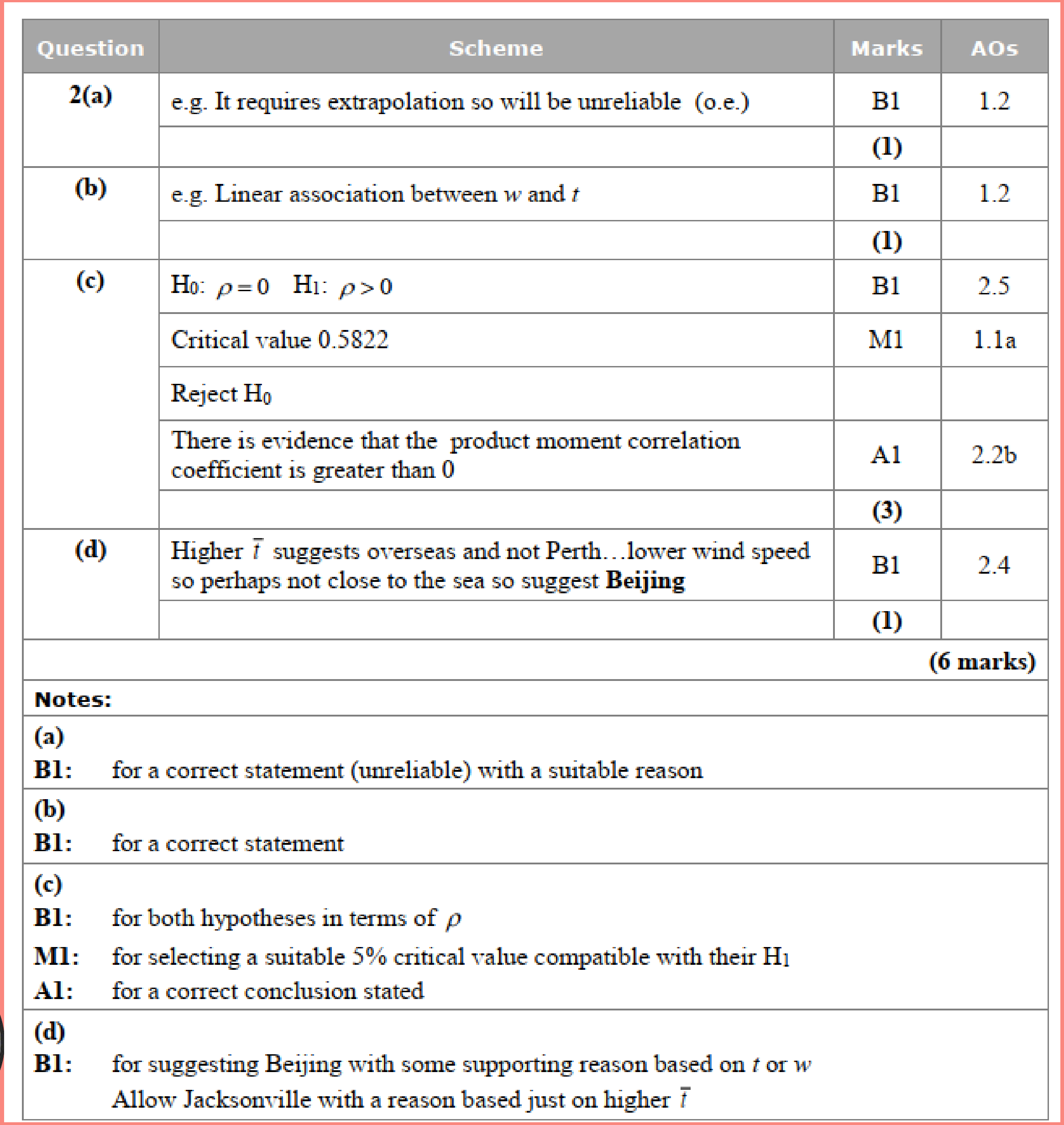
(d) Using your knowledge of the large data set, suggest, giving your reason, the location that gave rise to these statistics.

(1)

HOME

# A2 SAMs

## Regression and Correlation



| Question | Scheme | Marks | AOs |
|---|---|---|---|
| 2(a) | e.g. It requires extrapolation so will be unreliable (o.e.) | B1 | 1.2 |
| | | **(1)** | |
| (b) | e.g. Linear association between $w$ and $t$ | B1 | 1.2 |
| | | **(1)** | |
| (c) | $H_0$: $\rho = 0$ $H_1$: $\rho > 0$ | B1 | 2.5 |
| | Critical value 0.5822 | M1 | 1.1a |
| | Reject $H_0$ | | |
| | There is evidence that the product moment correlation coefficient is greater than 0 | A1 | 2.2b |
| | | **(3)** | |
| (d) | Higher $\bar{t}$ suggests overseas and not Perth…lower wind speed so perhaps not close to the sea so suggest **Beijing** | B1 | 2.4 |
| | | **(1)** | |
| | | | **(6 marks)** |

**Notes:**

**(a)**
**B1:** for a correct statement (unreliable) with a suitable reason

**(b)**
**B1:** for a correct statement

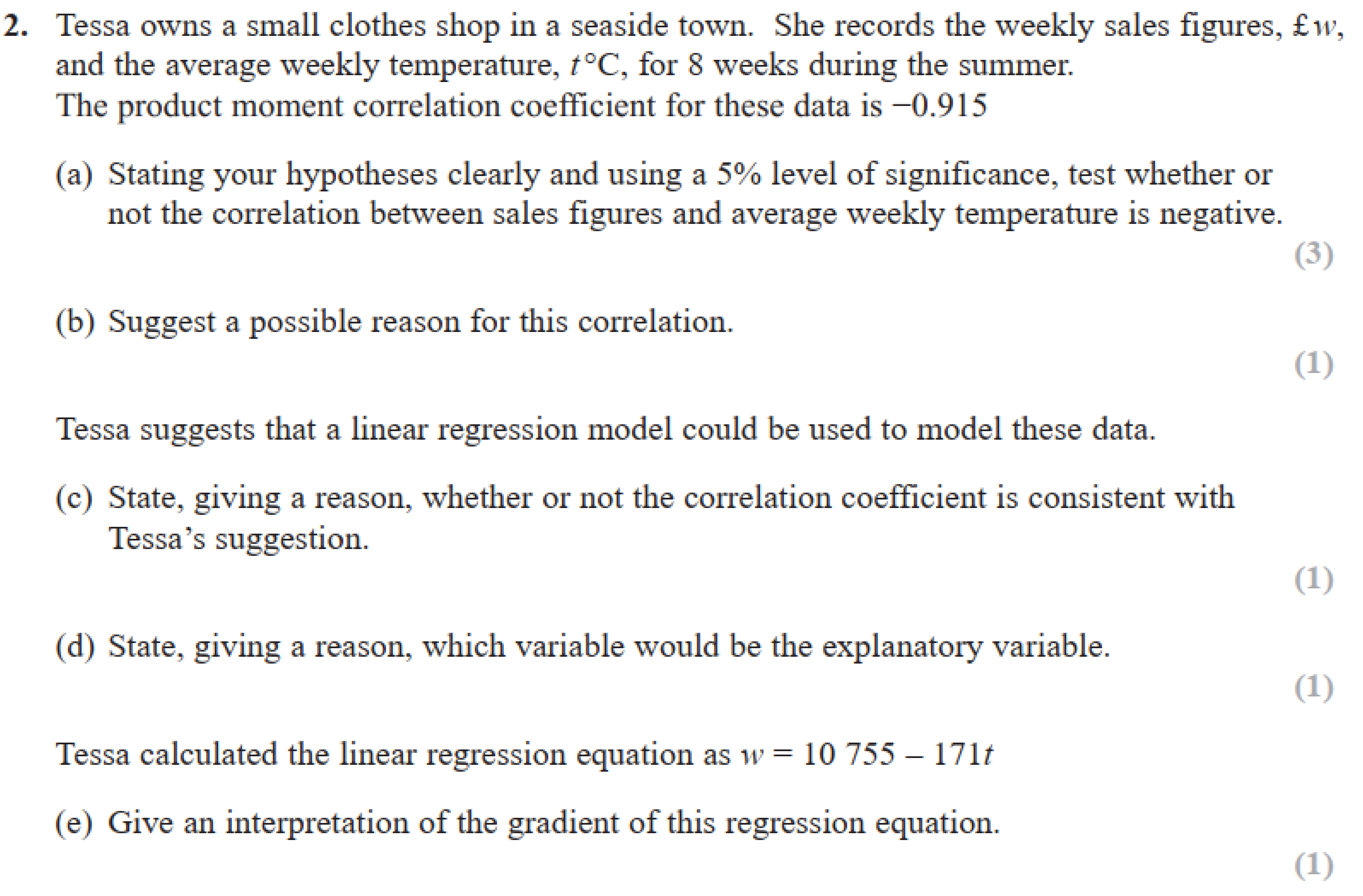
**(c)**
**B1:** for both hypotheses in terms of $\rho$
**M1:** for selecting a suitable 5% critical value compatible with their $H_1$
**A1:** for a correct conclusion stated

**(d)**
**B1:** for suggesting Beijing with some supporting reason based on $t$ or $w$
Allow Jacksonville with a reason based just on higher $\bar{t}$

HOME

**A2 2018**

Regression and Correlation



**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 $w = 10\,755 - 171t$

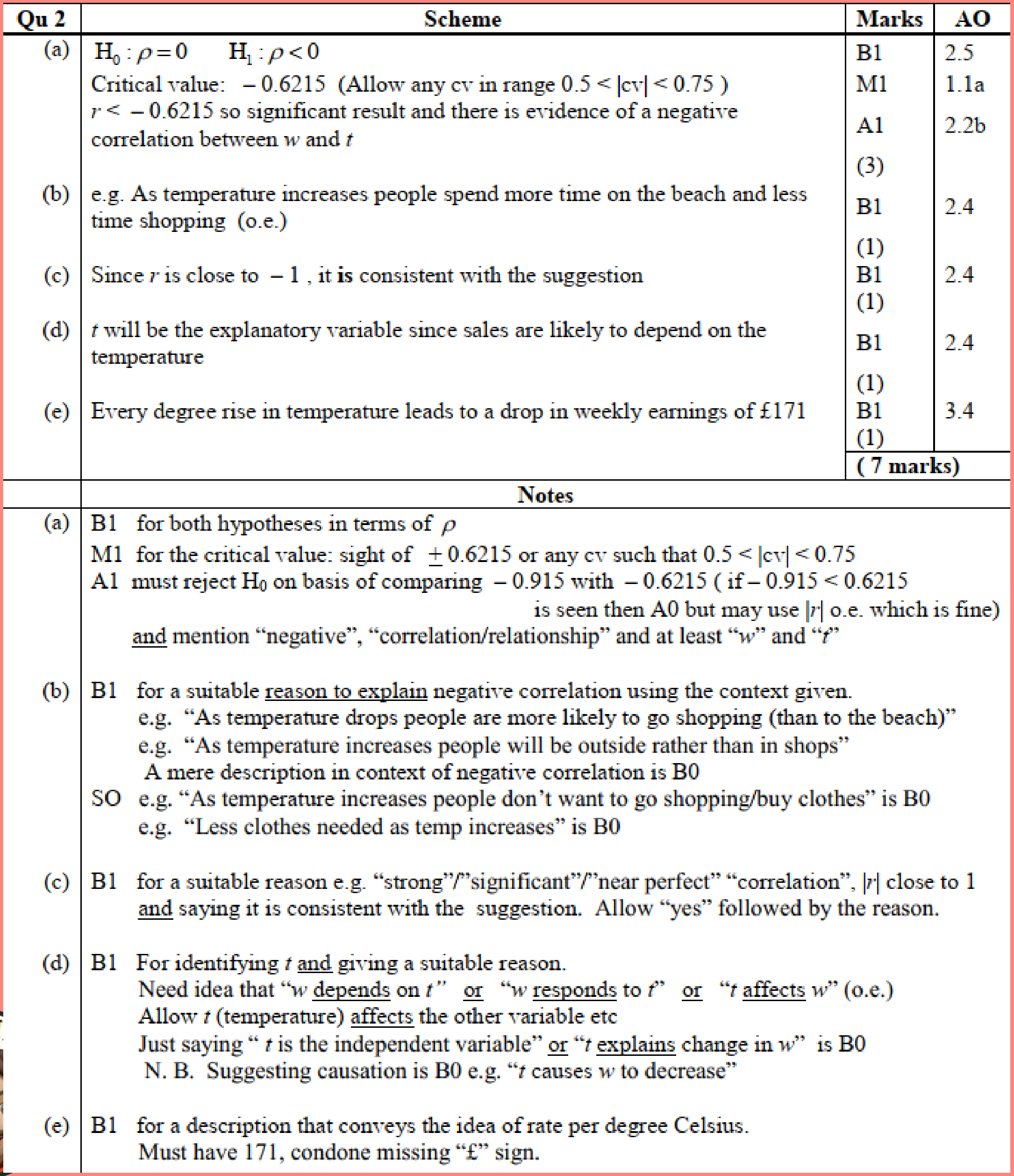
(e) Give an interpretation of the gradient of this regression equation.

**(1)**

HOME

# A2 2018

## Regression and Correlation

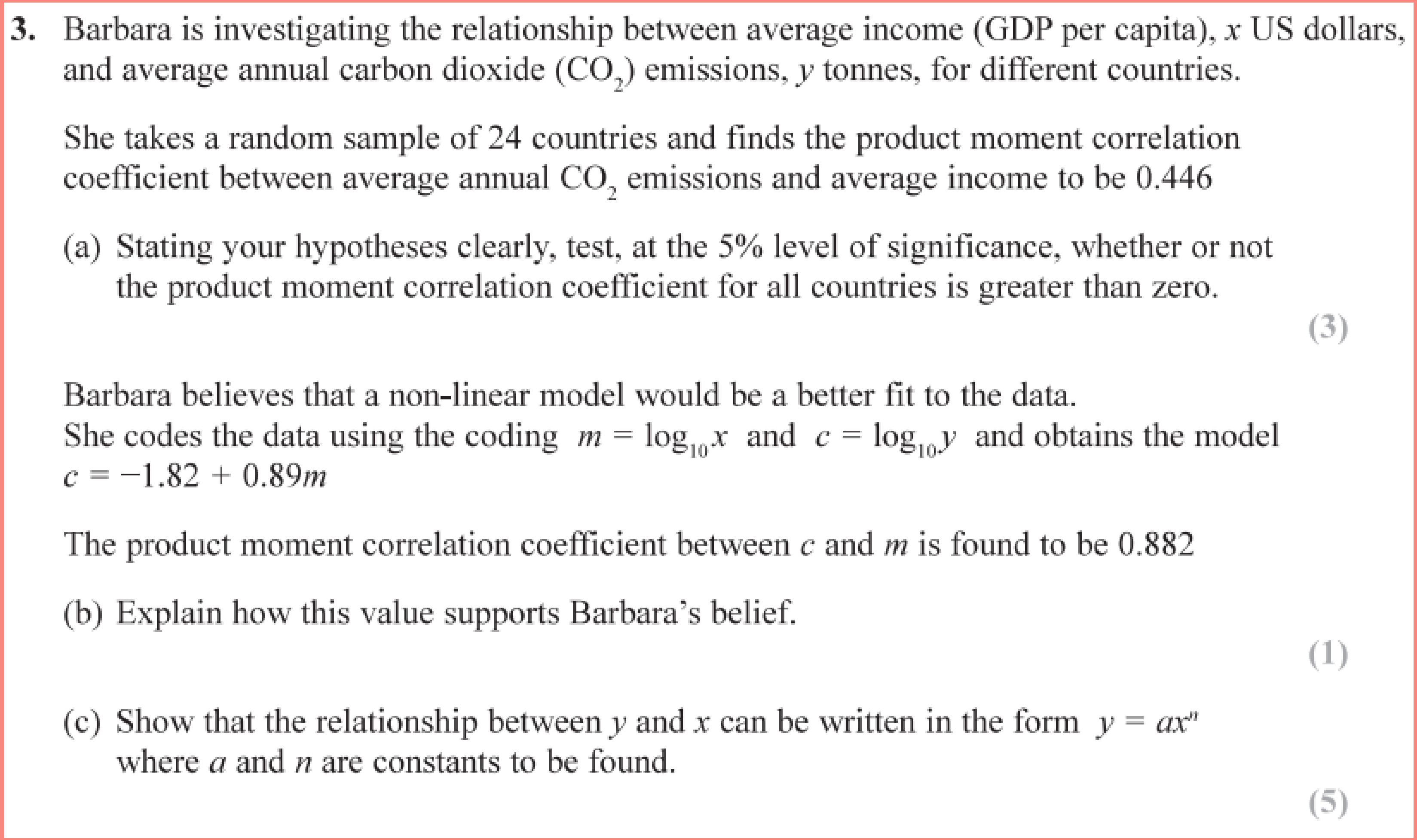


| Qu 2 | Scheme | Marks | AO |
|---|---|---|---|
| (a) | $H_0 : \rho = 0 \quad H_1 : \rho < 0$ | B1 | 2.5 |
| | Critical value: $-0.6215$ (Allow any cv in range $0.5 < \lvert cv \rvert < 0.75$ ) | M1 | 1.1a |
| | $r < -0.6215$ so significant result and there is evidence of a negative correlation between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

HOME

**A2 2019**

**Regression and Correlation**



**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)**

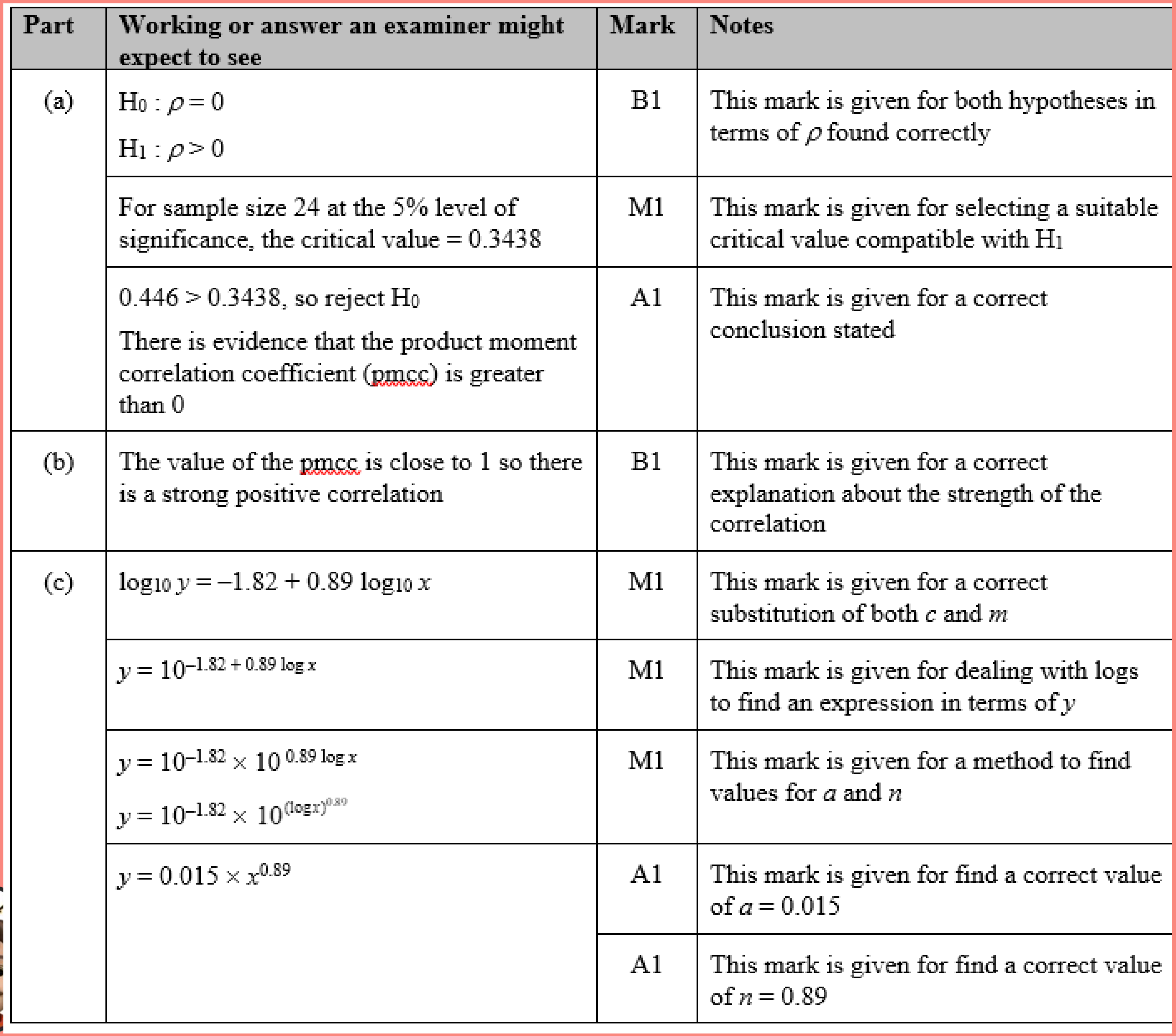
(c) Show that the relationship between $y$ and $x$ can be written in the form $y = ax^n$ where $a$ and $n$ are constants to be found.

**(5)**

HOME

# A2 2019

## Regression and Correlation



| Part | Working or answer an examiner might expect to see | Mark | Notes |
|---|---|---|---|
| (a) | $H_0 : \rho = 0$<br>$H_1 : \rho > 0$ | B1 | This mark is given for both hypotheses in terms of $\rho$ found correctly |
| | For sample size 24 at the 5% level of significance, the critical value = 0.3438 | M1 | This mark is given for selecting a suitable critical value compatible with $H_1$ |
| | $0.446 > 0.3438$, so reject $H_0$<br>There is evidence that the product moment correlation coefficient (pmcc) is greater than 0 | A1 | This mark is given for a correct conclusion stated |
| (b) | The value of the pmcc is close to 1 so there is a strong positive correlation | B1 | This mark is given for a correct explanation about the strength of the correlation |
| (c) | $\log_{10} y = -1.82 + 0.89 \log_{10} x$ | M1 | This mark is given for a correct substitution of both $c$ and $m$ |
| | $y = 10^{-1.82 + 0.89 \log x}$ | M1 | This mark is given for dealing with logs to find an expression in terms of $y$ |
| | $y = 10^{-1.82} \times 10^{0.89 \log x}$<br>$y = 10^{-1.82} \times 10^{(\log x)^{0.89}}$ | M1 | This mark is given for a method to find values for $a$ and $n$ |
| | $y = 0.015 \times x^{0.89}$ | A1 | This mark is given for find a correct value of $a = 0.015$ |
| | | A1 | This mark is given for find a correct value of $n = 0.89$ |

HOME

# A2 2020

Regression and Correlation

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.

(a) Describe the correlation.

**(1)**

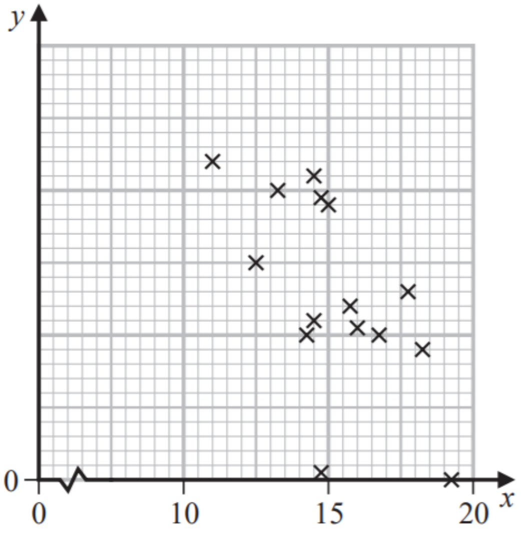


**Figure 1**

The variable on the $x$-axis is Daily Mean Temperature measured in °C.

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

(i) suggest which variable is on the $y$-axis,

(ii) 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%

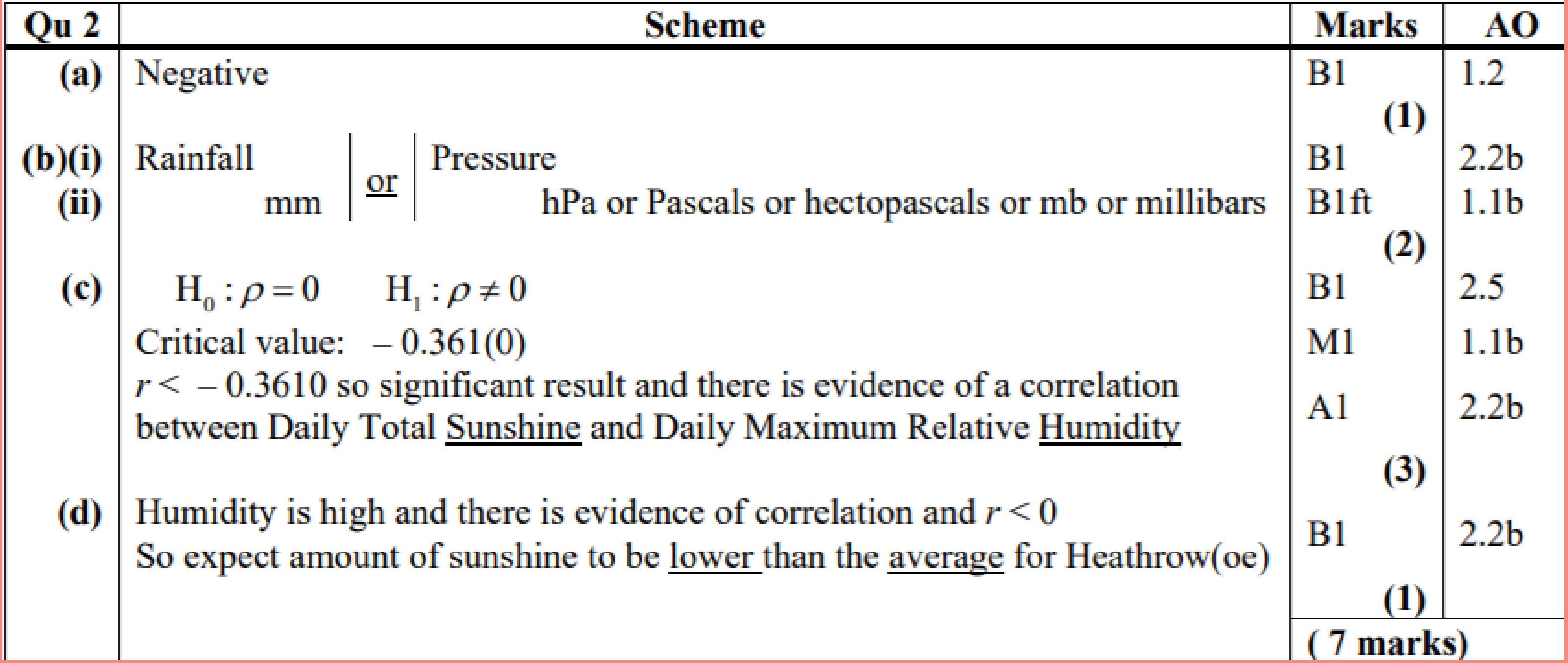
(d) Comment on the number of hours of sunshine you would expect on that day, giving a reason for your answer.

**(1)**

HOME

# A2 2020

## Regression and Correlation

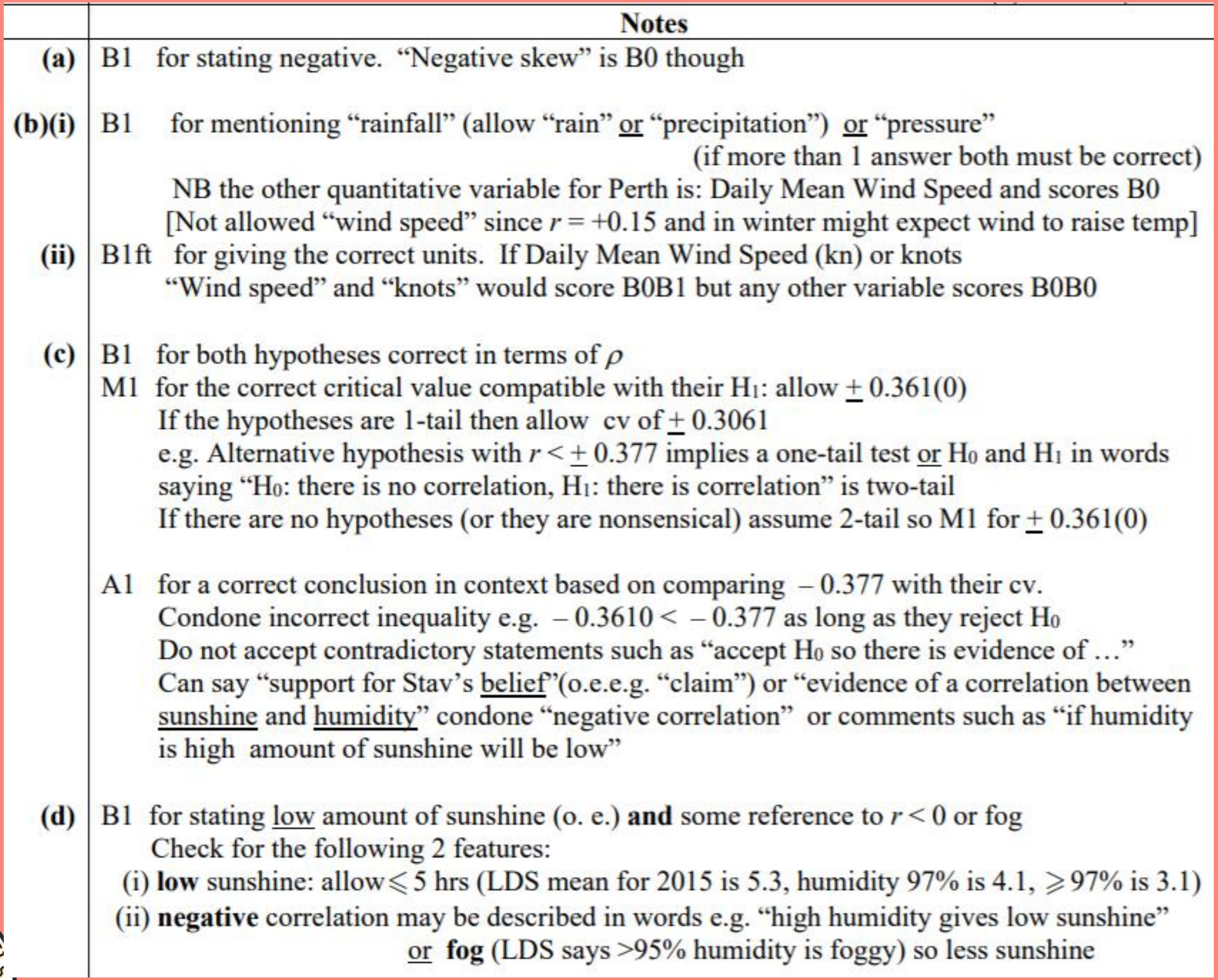


| Qu 2 | Scheme | Marks | AO |
|---|---|---|---|
| (a) | Negative | B1<br>(1) | 1.2 |
| (b)(i)<br>(ii) | Rainfall<br>mm<br>or<br>Pressure<br>hPa or Pascals or hectopascals or mb or millibars | B1<br>B1ft<br>(2) | 2.2b<br>1.1b |
| (c) | $H_0 : \rho = 0$ $H_1 : \rho \neq 0$ | B1 | 2.5 |
| | Critical value: $-0.361(0)$ | M1 | 1.1b |
| | $r < -0.3610$ so significant result and there is evidence of a correlation between Daily Total Sunshine and Daily Maximum Relative Humidity | A1<br>(3) | 2.2b |
| (d) | Humidity is high and there is evidence of correlation and $r < 0$<br>So expect amount of sunshine to be lower than the average for Heathrow(oe) | B1<br>(1) | 2.2b |
| | | (7 marks) | |

HOME

**A2 2020**

Regression and Correlation



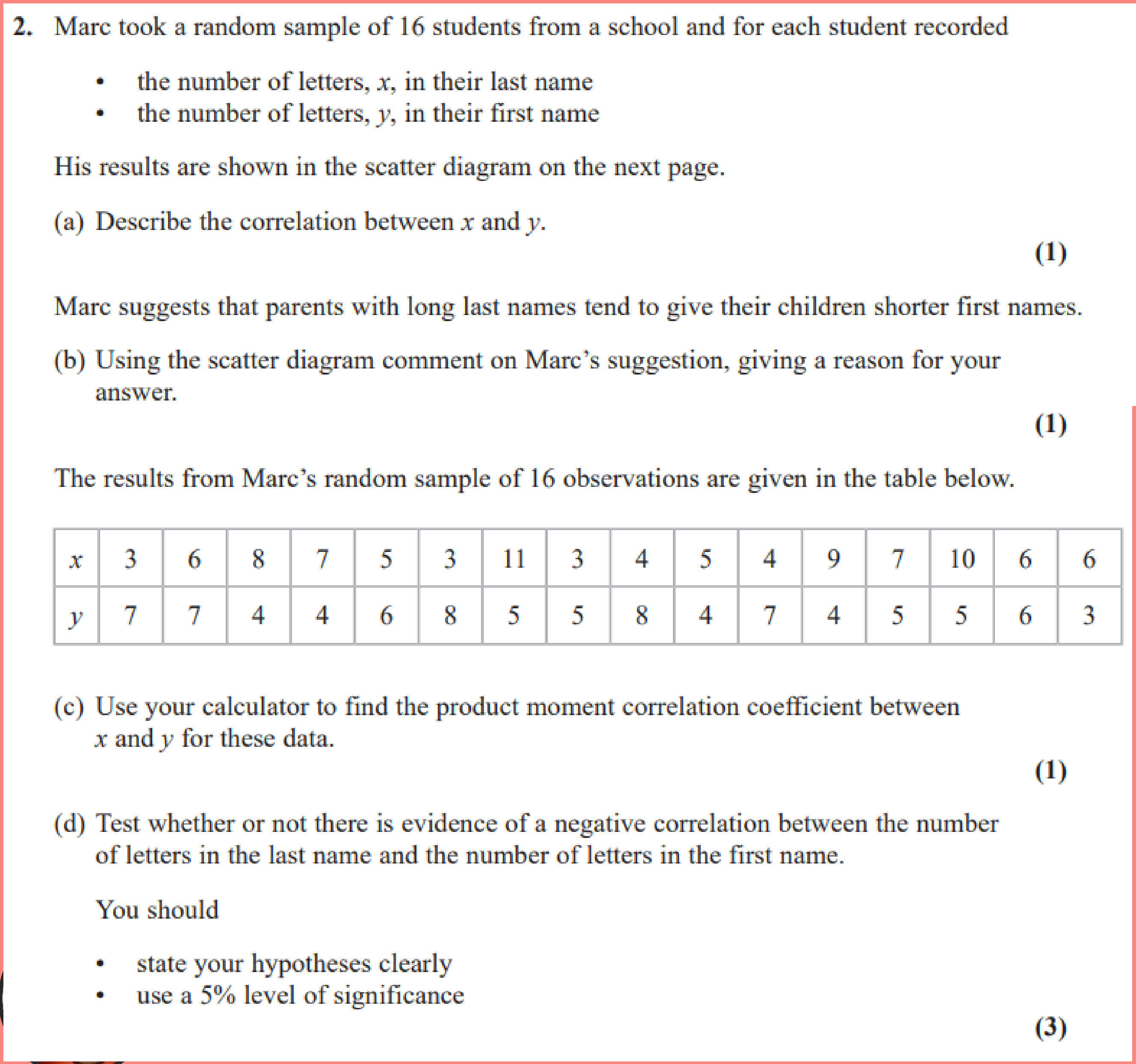
| | Notes |
|---|---|
| **(a)** | B1 for stating negative. "Negative skew" is B0 though |
| **(b)(i)** | B1 for mentioning "rainfall" (allow "rain" or "precipitation") or "pressure"<br>(if more than 1 answer both must be correct)<br>NB the other quantitative variable for Perth is: Daily Mean Wind Speed and scores B0<br>[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<br>"Wind speed" and "knots" would score B0B1 but any other variable scores B0B0 |
| **(c)** | B1 for both hypotheses correct in terms of $\rho$<br>M1 for the correct critical value compatible with their $H_1$: allow $\pm$ 0.361(0)<br>If the hypotheses are 1-tail then allow cv of $\pm$ 0.3061<br>e.g. Alternative hypothesis with $r < \pm 0.377$ implies a one-tail test or $H_0$ and $H_1$ in words saying "$H_0$: there is no correlation, $H_1$: there is correlation" is two-tail<br>If there are no hypotheses (or they are nonsensical) assume 2-tail so M1 for $\pm$ 0.361(0)<br><br>A1 for a correct conclusion in context based on comparing $-0.377$ with their cv.<br>Condone incorrect inequality e.g. $-0.3610 < -0.377$ as long as they reject $H_0$<br>Do not accept contradictory statements such as "accept $H_0$ so there is evidence of …"<br>Can say "support for Stav's belief"(o.e.e.g. "claim") or "evidence of a correlation between sunshine and humidity" condone "negative correlation" or comments such as "if humidity is high amount of sunshine will be low" |
| **(d)** | B1 for stating low amount of sunshine (o. e.) **and** some reference to $r < 0$ or fog<br>Check for the following 2 features:<br>(i) **low** sunshine: allow $\leqslant$ 5 hrs (LDS mean for 2015 is 5.3, humidity 97% is 4.1, $\geqslant$ 97% is 3.1)<br>(ii) **negative** correlation may be described in words e.g. "high humidity gives low sunshine"<br>or **fog** (LDS says >95% humidity is foggy) so less sunshine |

BICEN MATHS

HOME

# A2 2021

## Regression and Correlation



**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.

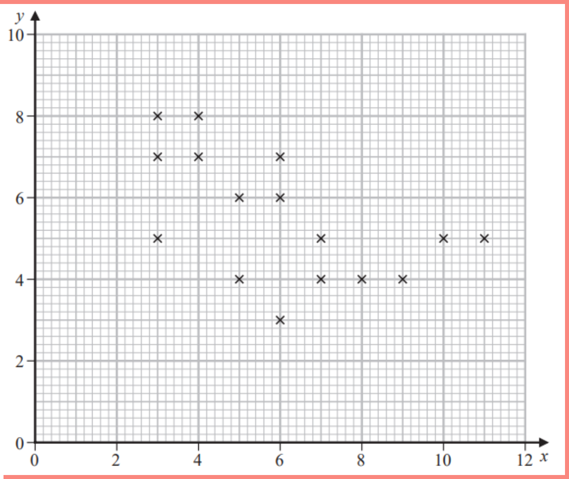
**(1)**

(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

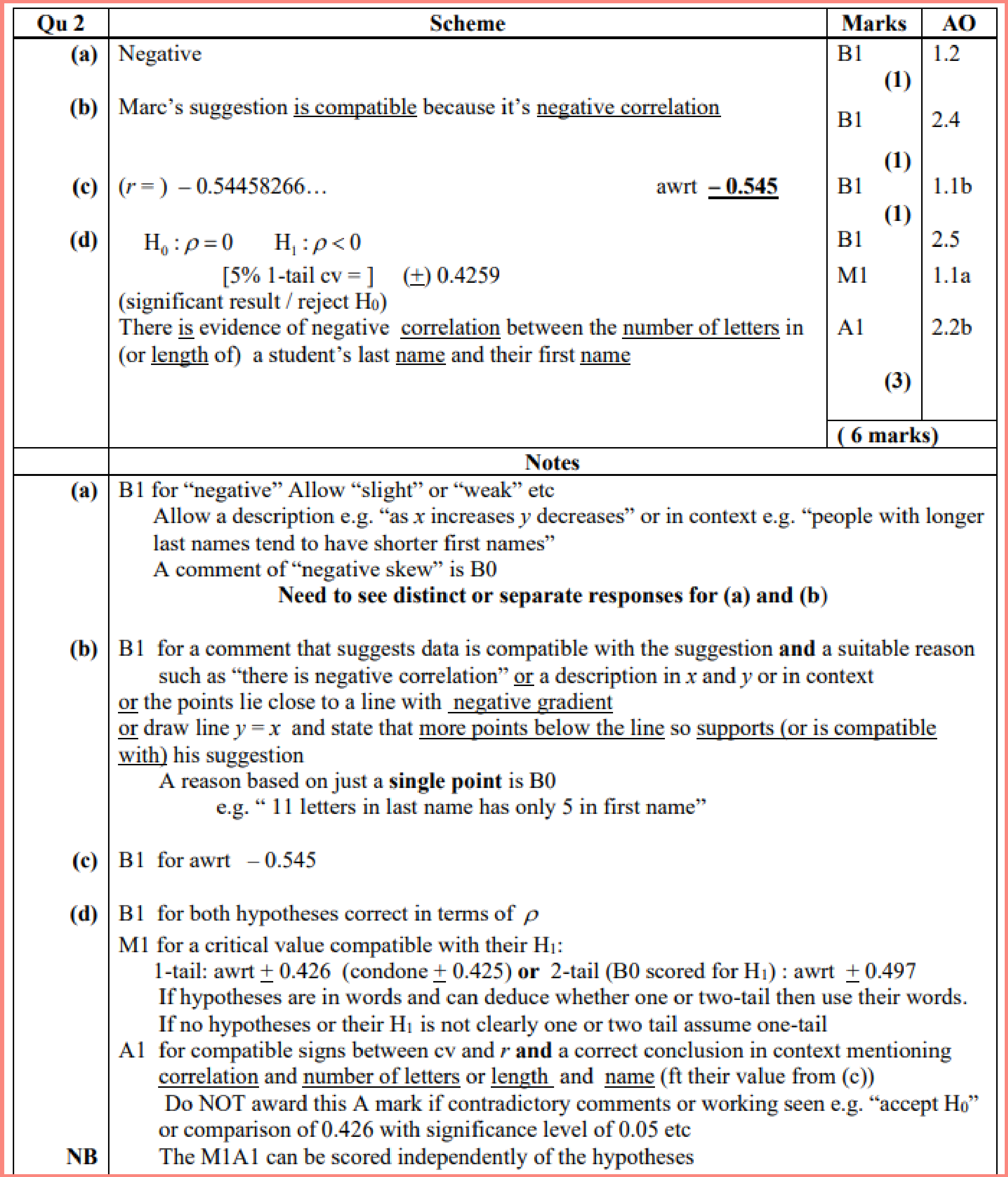
**(3)**



HOME

# A2 2021

# Regression and Correlation



| Qu 2 | Scheme | Marks | AO |
|---|---|---|---|
| (a) | Negative | B1<br>(1) | 1.2 |
| (b) | Marc's suggestion is compatible because it's negative correlation | B1<br>(1) | 2.4 |
| (c) | ($r =$) $-0.54458266\ldots$ awrt $\mathbf{-0.545}$ | B1<br>(1) | 1.1b |
| (d) | $H_0 : \rho = 0$ $H_1 : \rho < 0$ | B1 | 2.5 |
| | [5% 1-tail cv = ] ($\pm$) 0.4259 | M1 | 1.1a |
| | (significant result / reject $H_0$)<br>There is evidence of negative correlation between the number of letters in (or length of) a student's last name and their first name | A1 | 2.2b |
| | | (3) | |
| | | ( 6 marks) | |

| | Notes |
|---|---|
| (a) | B1 for "negative" Allow "slight" or "weak" etc<br>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"<br>A comment of "negative skew" is B0<br>**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" or a description in $x$ and $y$ or in context<br>or the points lie close to a line with negative gradient<br>or draw line $y = x$ and state that more points below the line so supports (or is compatible with) his suggestion<br>A reason based on just a **single point** is B0<br>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 $\rho$<br>M1 for a critical value compatible with their $H_1$:<br>1-tail: awrt $\pm$ 0.426 (condone $\pm$ 0.425) **or** 2-tail (B0 scored for $H_1$) : awrt $\pm$ 0.497<br>If hypotheses are in words and can deduce whether one or two-tail then use their words.<br>If no hypotheses or their $H_1$ is not clearly one or two tail assume one-tail<br>A1 for compatible signs between cv and $r$ **and** a correct conclusion in context mentioning correlation and number of letters or length and name (ft their value from (c))<br>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 |

HOME

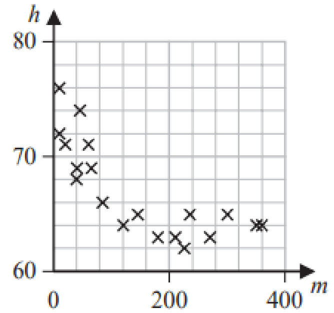
A2 2022

Regression and Correlation

**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

$$h = am^k$$

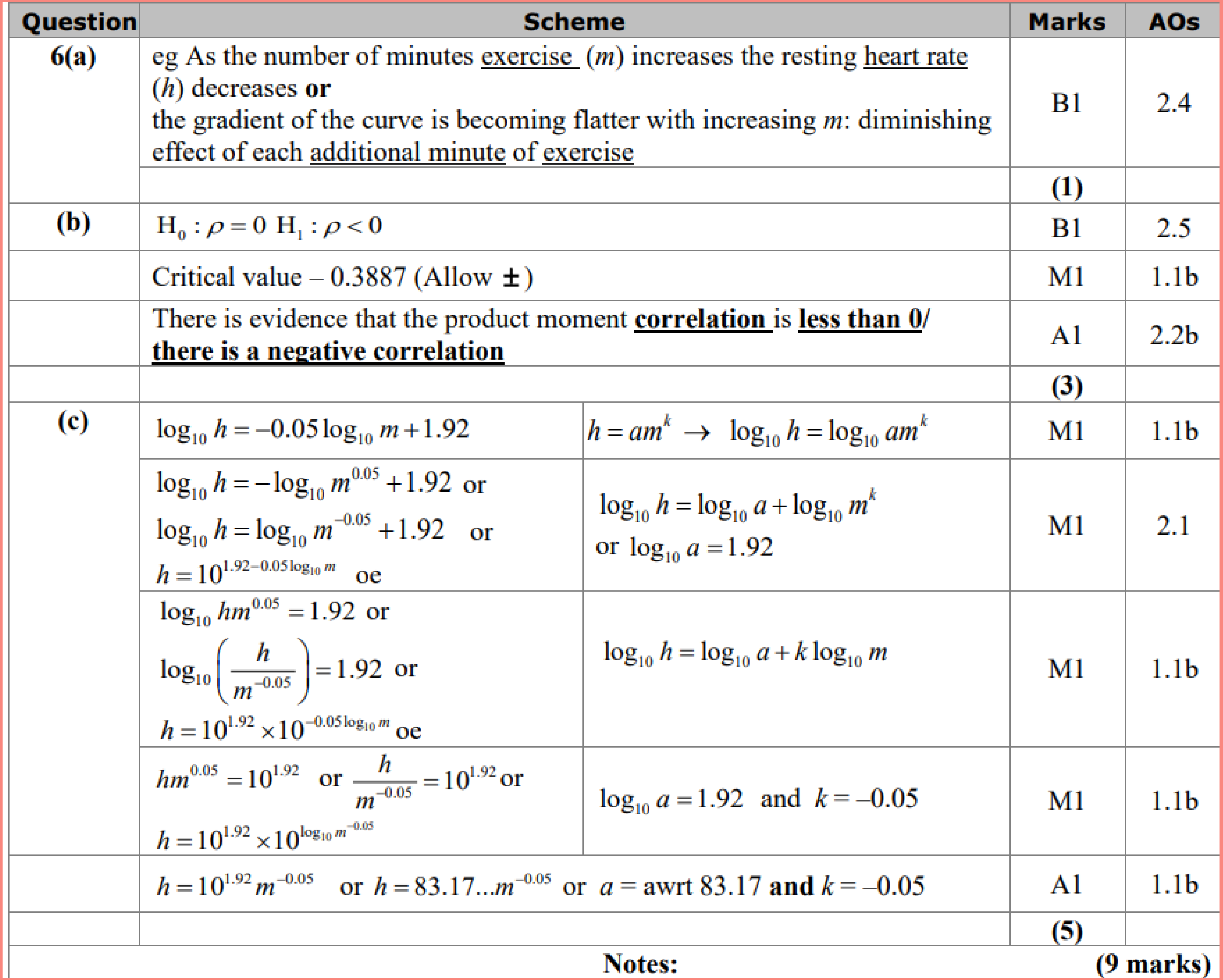
where $a$ and $k$ are constants to be found.

**(5)**

HOME

# A2 2022

## Regression and Correlation



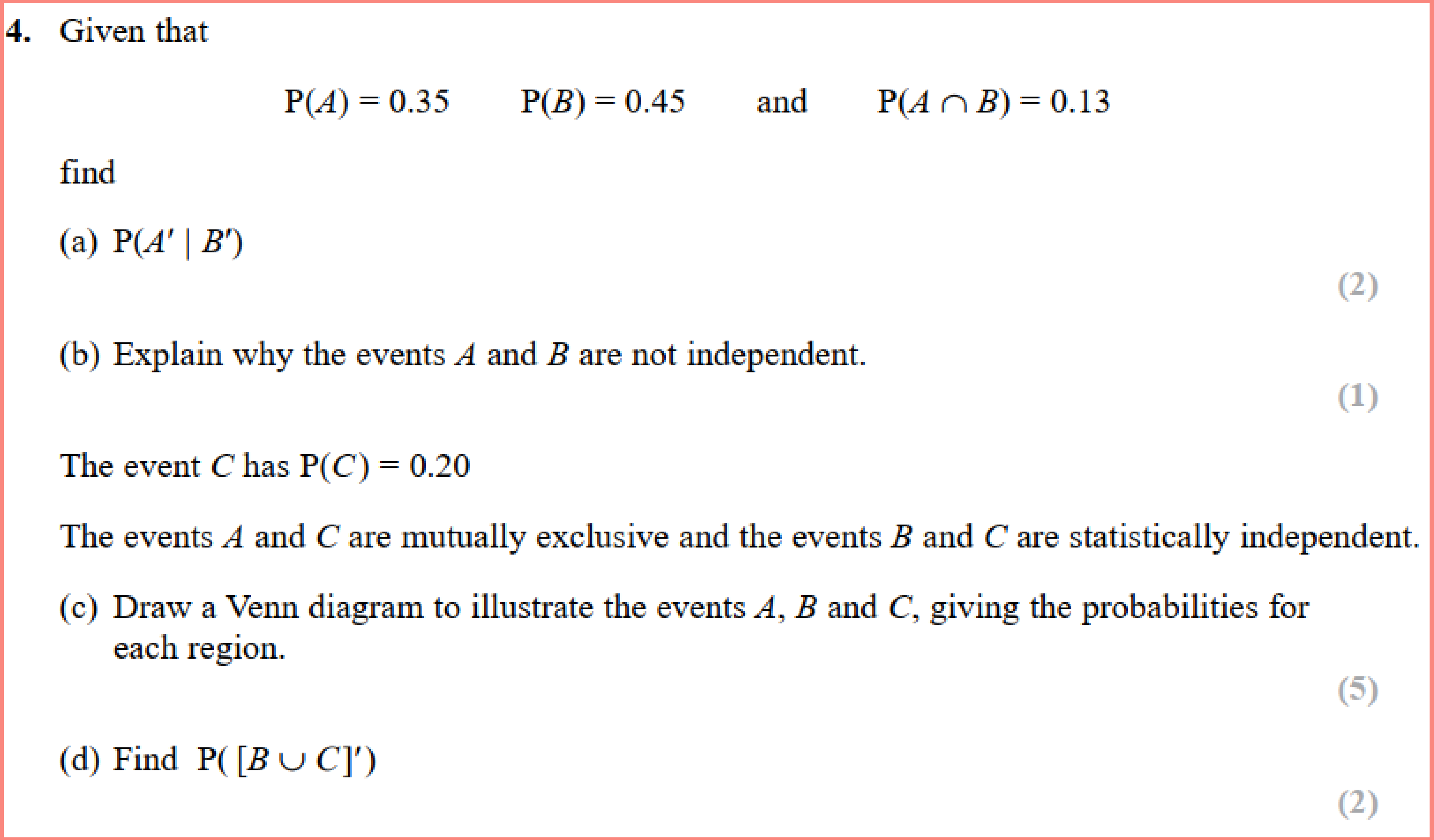
| Question | Scheme | | Marks | AOs |
|---|---|---|---|---|
| 6(a) | eg As the number of minutes exercise ($m$) increases the resting heart rate ($h$) decreases **or**<br>the gradient of the curve is becoming flatter with increasing $m$: diminishing effect of each additional minute of exercise | | B1 | 2.4 |
| | | | **(1)** | |
| (b) | $\mathrm{H}_0 : \rho = 0 \ \mathrm{H}_1 : \rho < 0$ | | B1 | 2.5 |
| | Critical value – 0.3887 (Allow ±) | | M1 | 1.1b |
| | There is evidence that the product moment **correlation** is **less than 0/ there is a negative correlation** | | 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<br>$\log_{10} h = \log_{10} m^{-0.05} + 1.92$ or<br>$h = 10^{1.92 - 0.05\log_{10} m}$ oe | $\log_{10} h = \log_{10} a + \log_{10} m^k$<br>or $\log_{10} a = 1.92$ | M1 | 2.1 |
| | $\log_{10} hm^{0.05} = 1.92$ or<br>$\log_{10}\left(\dfrac{h}{m^{-0.05}}\right) = 1.92$ or<br>$h = 10^{1.92} \times 10^{-0.05\log_{10} m}$ oe | $\log_{10} h = \log_{10} a + k\log_{10} m$ | M1 | 1.1b |
| | $hm^{0.05} = 10^{1.92}$ or $\dfrac{h}{m^{-0.05}} = 10^{1.92}$ or<br>$h = 10^{1.92} \times 10^{\log_{10} m^{-0.05}}$ | $\log_{10} a = 1.92$ and $k = -0.05$ | M1 | 1.1b |
| | $h = 10^{1.92} m^{-0.05}$ or $h = 83.17\ldots m^{-0.05}$ or $a =$ awrt 83.17 **and** $k = -0.05$ | | A1 | 1.1b |
| | | | **(5)** | |
| | **Notes:** | | | **(9 marks)** |

HOME

# Conditional Probability

A2 SAMs

# Conditional Probability



**4.** Given that

$$P(A) = 0.35 \qquad P(B) = 0.45 \qquad \text{and} \qquad P(A \cap B) = 0.13$$

find

(a) $P(A' \mid B')$

(2)

(b) Explain why the events $A$ and $B$ are not independent.

(1)

The event $C$ has $P(C) = 0.20$

The events $A$ and $C$ are mutually exclusive and the events $B$ and $C$ are statistically independent.

(c) Draw a Venn diagram to illustrate the events $A$, $B$ and $C$, giving the probabilities for each region.

(5)

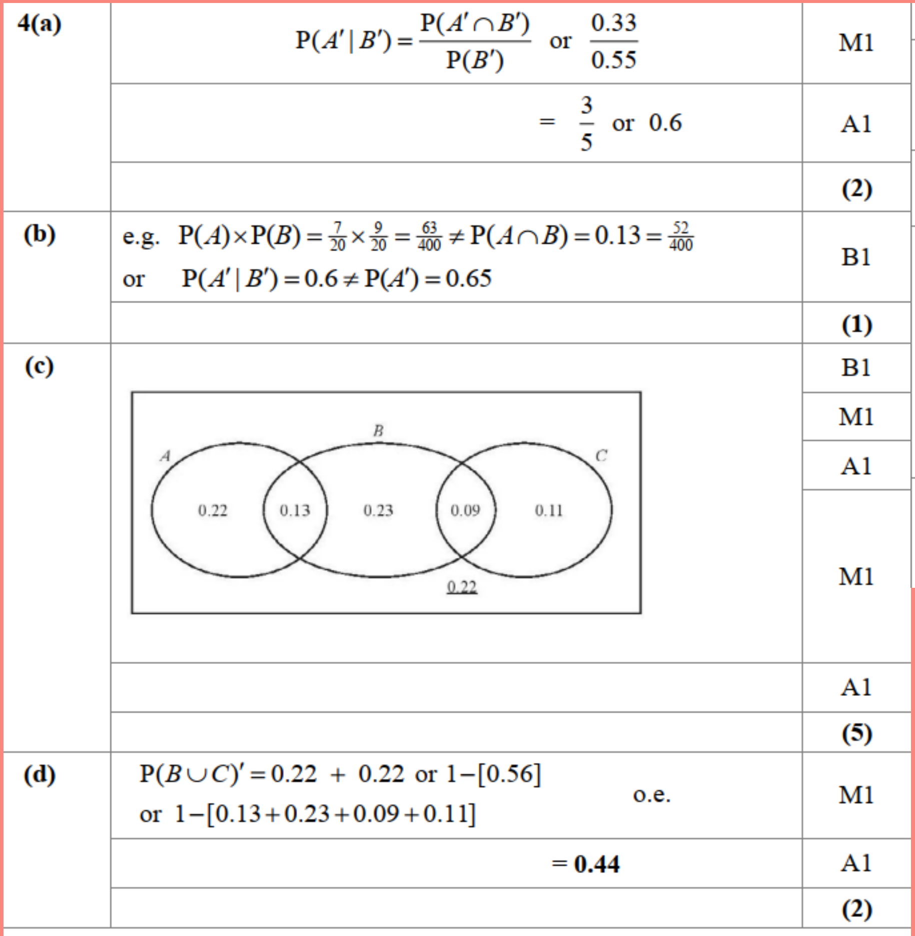
(d) Find $P([B \cup C]')$

(2)

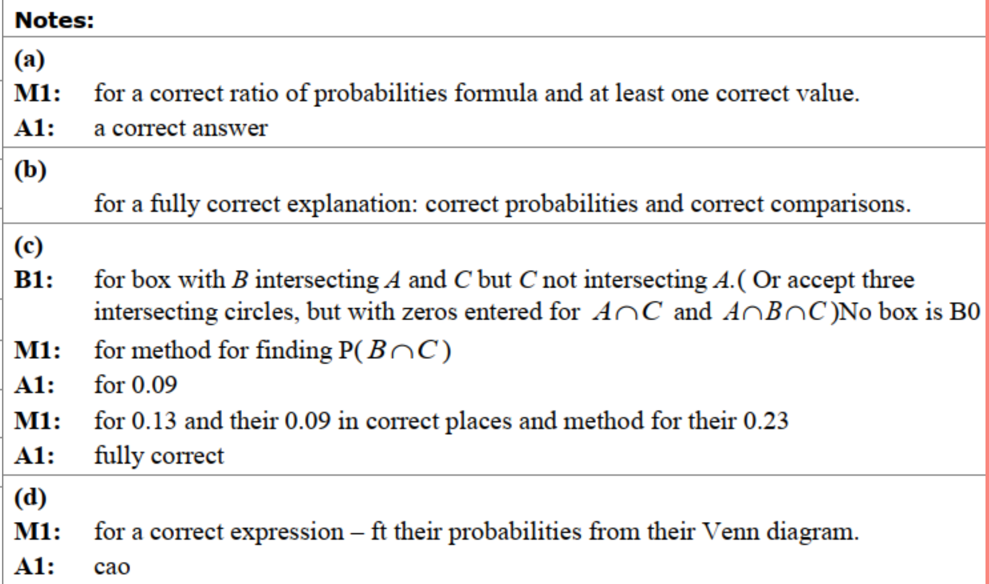
HOME

# A2 SAMs

## Conditional Probability



| | | |
|---|---|---|
| **4(a)** | $P(A' \mid B') = \dfrac{P(A' \cap B')}{P(B')}$ or $\dfrac{0.33}{0.55}$ | M1 |
| | $= \dfrac{3}{5}$ or $0.6$ | A1 |
| | | **(2)** |
| **(b)** | e.g. $P(A) \times P(B) = \frac{7}{20} \times \frac{9}{20} = \frac{63}{400} \neq P(A \cap B) = 0.13 = \frac{52}{400}$<br>or $P(A' \mid B') = 0.6 \neq P(A') = 0.65$ | B1 |
| | | **(1)** |
| **(c)** |  A<br>B<br>C<br>0.22<br>0.13<br>0.23<br>0.09<br>0.11<br>0.22 | B1<br>M1<br>A1<br>M1 |
| | | A1 |
| | | **(5)** |
| **(d)** | $P(B \cup C)' = 0.22 + 0.22$ or $1-[0.56]$ o.e.<br>or $1-[0.13+0.23+0.09+0.11]$ | M1 |
| | $=$ **0.44** | A1 |
| | | **(2)** |



**Notes:**

**(a)**

**M1:** for a correct ratio of probabilities formula and at least one correct value.

**A1:** a correct answer

**(b)**

for a fully correct explanation: correct probabilities and correct comparisons.

**(c)**

**B1:** for box with $B$ intersecting $A$ and $C$ but $C$ not intersecting $A$.( Or accept three intersecting circles, but with zeros entered for $A \cap C$ and $A \cap B \cap C$ )No box is B0

**M1:** for method for finding P( $B \cap C$ )

**A1:** for 0.09

**M1:** for 0.13 and their 0.09 in correct places and method for their 0.23

**A1:** fully correct

**(d)**

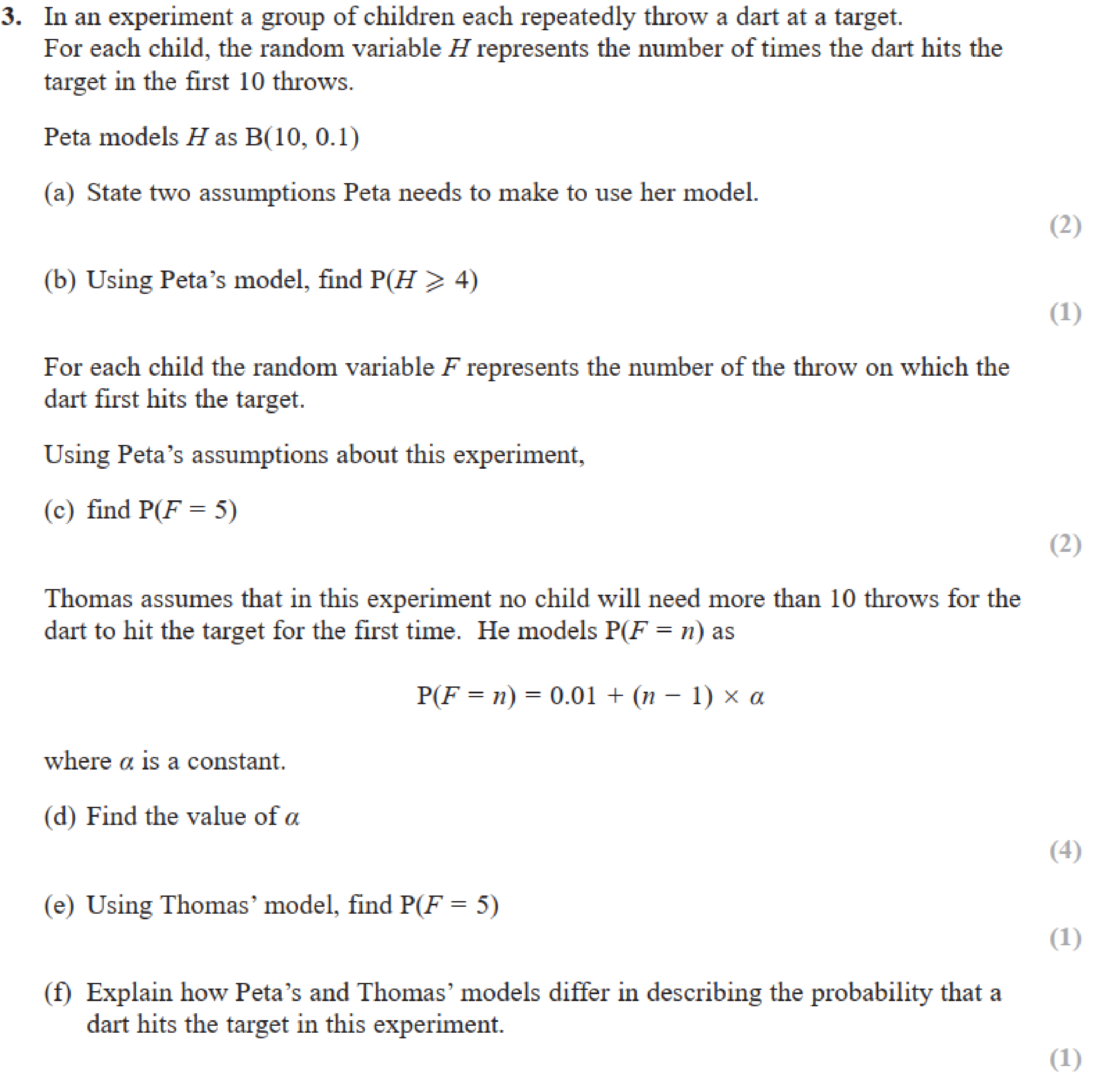
**M1:** for a correct expression – ft their probabilities from their Venn diagram.

**A1:** cao

HOME

# A2 2018

Conditional Probability



**3.** In an experiment a group of children each repeatedly throw a dart at a target. For each child, the random variable $H$ represents the number of times the dart hits the target in the first 10 throws.

Peta models $H$ as B(10, 0.1)

(a) State two assumptions Peta needs to make to use her model.

**(2)**

(b) Using Peta's model, find $\mathrm{P}(H \geqslant 4)$

**(1)**

For each child the random variable $F$ represents the number of the throw on which the dart first hits the target.

Using Peta's assumptions about this experiment,

(c) find $\mathrm{P}(F = 5)$

**(2)**

Thomas assumes that in this experiment no child will need more than 10 throws for the dart to hit the target for the first time. He models $\mathrm{P}(F = n)$ as

$$\mathrm{P}(F = n) = 0.01 + (n - 1) \times \alpha$$

where $\alpha$ is a constant.

(d) Find the value of $\alpha$

**(4)**

(e) Using Thomas' model, find $\mathrm{P}(F = 5)$

**(1)**

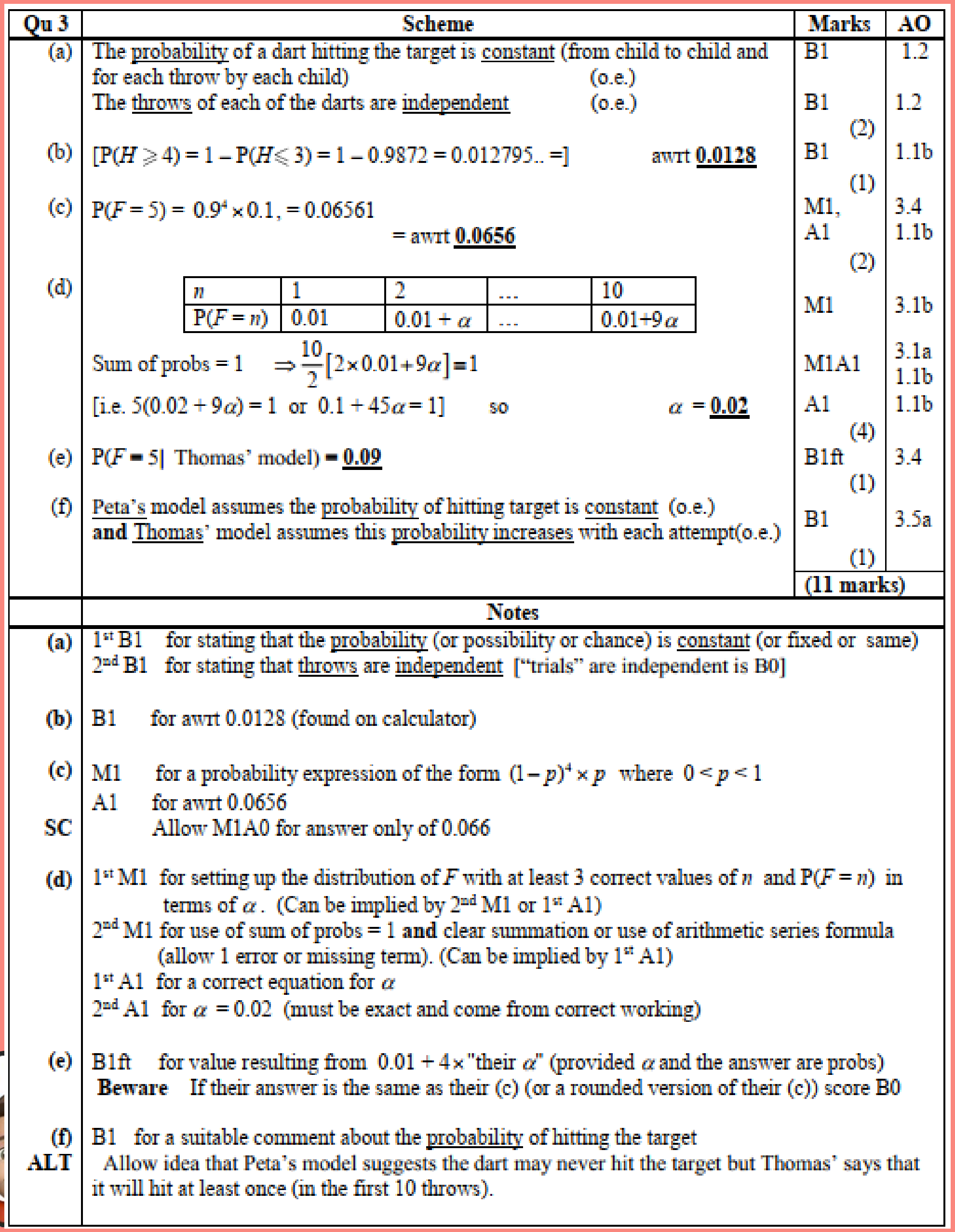
(f) Explain how Peta's and Thomas' models differ in describing the probability that a dart hits the target in this experiment.

**(1)**

HOME

# A2 2018

## Conditional Probability



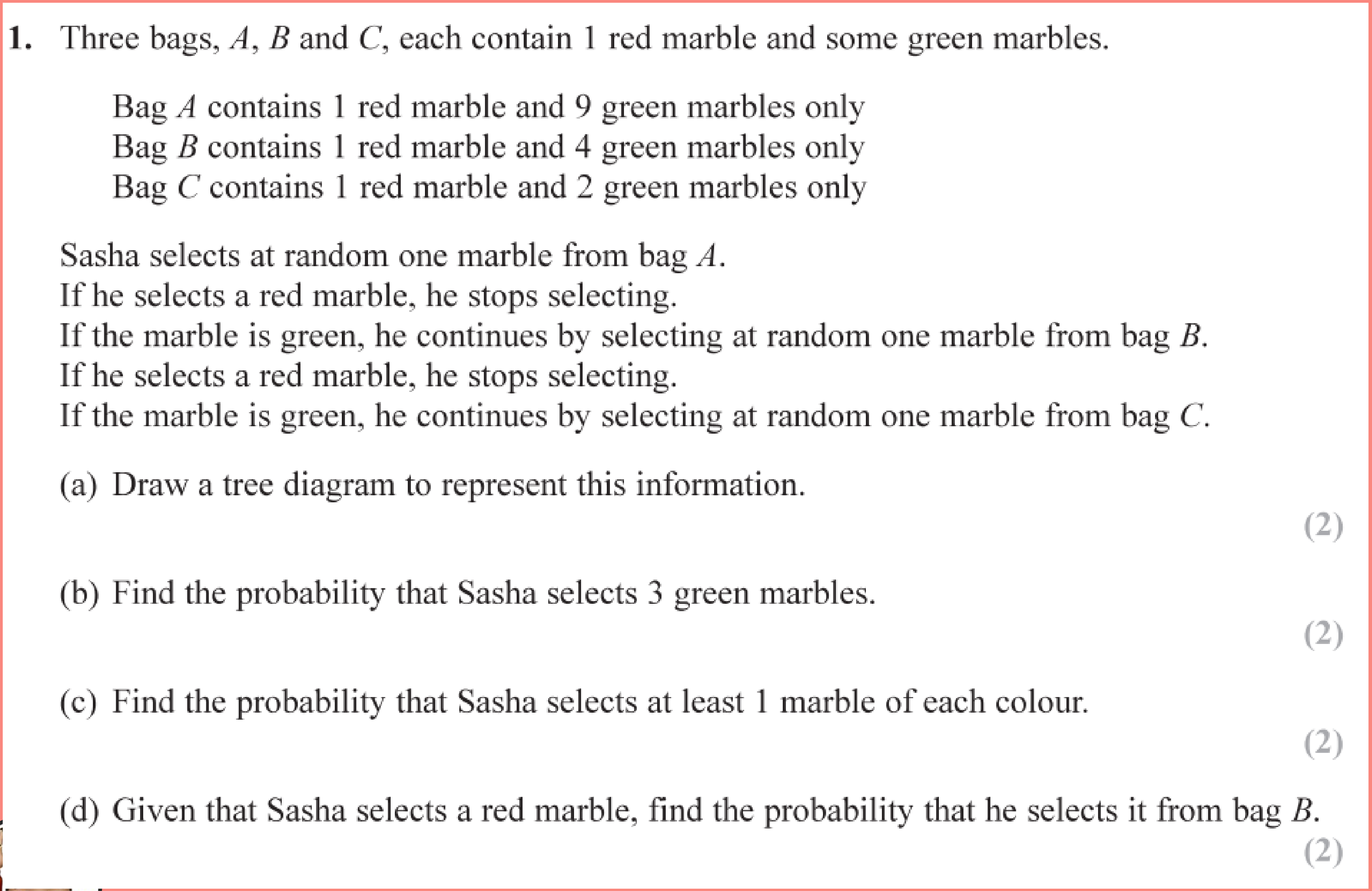
| Qu 3 | Scheme | Marks | AO |
|---|---|---|---|
| (a) | The probability of a dart hitting the target is constant (from child to child and for each throw by each child) (o.e.) | B1 | 1.2 |
| | The throws of each of the darts are independent (o.e.) | B1<br>(2) | 1.2 |
| (b) | $[\mathrm{P}(H \geqslant 4) = 1 - \mathrm{P}(H \leqslant 3) = 1 - 0.9872 = 0.012795.. =]$ awrt **0.0128** | B1<br>(1) | 1.1b |
| (c) | $\mathrm{P}(F = 5) = 0.9^4 \times 0.1, = 0.06561$<br>= awrt **0.0656** | M1,<br>A1<br>(2) | 3.4<br>1.1b |
| (d) | $\begin{array}{ccccc} n & 1 & 2 & \ldots & 10 \\ \mathrm{P}(F = n) & 0.01 & 0.01 + \alpha & \ldots & 0.01+9\alpha \end{array}$ | M1 | 3.1b |
| | Sum of probs = 1 $\Rightarrow \frac{10}{2}[2 \times 0.01 + 9\alpha] = 1$ | M1A1 | 3.1a<br>1.1b |
| | [i.e. $5(0.02 + 9\alpha) = 1$ or $0.1 + 45\alpha = 1$] so $\alpha$ = **0.02** | A1<br>(4) | 1.1b |
| (e) | $\mathrm{P}(F = 5 \mid$ Thomas' model$)$ = **0.09** | B1ft<br>(1) | 3.4 |
| (f) | Peta's model assumes the probability of hitting target is constant (o.e.)<br>**and** Thomas' model assumes this probability increases with each attempt(o.e.) | B1<br>(1) | 3.5a |
| | | **(11 marks)** | |

| | Notes |
|---|---|
| **(a)** | 1st B1 for stating that the probability (or possibility or chance) is constant (or fixed or same)<br>2nd B1 for stating that throws are independent ["trials" are independent is B0] |
| **(b)** | B1 for awrt 0.0128 (found on calculator) |
| **(c)** | M1 for a probability expression of the form $(1-p)^4 \times p$ where $0 < p < 1$<br>A1 for awrt 0.0656 |
| **SC** | Allow M1A0 for answer only of 0.066 |
| **(d)** | 1st M1 for setting up the distribution of $F$ with at least 3 correct values of $n$ and $\mathrm{P}(F = n)$ in terms of $\alpha$. (Can be implied by 2nd M1 or 1st A1)<br>2nd M1 for use of sum of probs = 1 **and** clear summation or use of arithmetic series formula (allow 1 error or missing term). (Can be implied by 1st A1)<br>1st A1 for a correct equation for $\alpha$<br>2nd A1 for $\alpha = 0.02$ (must be exact and come from correct working) |
| **(e)** | B1ft for value resulting from $0.01 + 4 \times$ "their $\alpha$" (provided $\alpha$ and the answer are probs)<br>**Beware** If their answer is the same as their (c) (or a rounded version of their (c)) score B0 |
| **(f)**<br>**ALT** | B1 for a suitable comment about the probability of hitting the target<br>Allow idea that Peta's model suggests the dart may never hit the target but Thomas' says that it will hit at least once (in the first 10 throws). |

HOME

## A2 2019

Conditional Probability



1. Three bags, *A*, *B* and *C*, each contain 1 red marble and some green marbles.

   Bag *A* contains 1 red marble and 9 green marbles only
   Bag *B* contains 1 red marble and 4 green marbles only
   Bag *C* contains 1 red marble and 2 green marbles only

   Sasha selects at random one marble from bag *A*.
   If he selects a red marble, he stops selecting.
   If the marble is green, he continues by selecting at random one marble from bag *B*.
   If he selects a red marble, he stops selecting.
   If the marble is green, he continues by selecting at random one marble from bag *C*.

   (a) Draw a tree diagram to represent this information.
   **(2)**

   (b) Find the probability that Sasha selects 3 green marbles.
   **(2)**

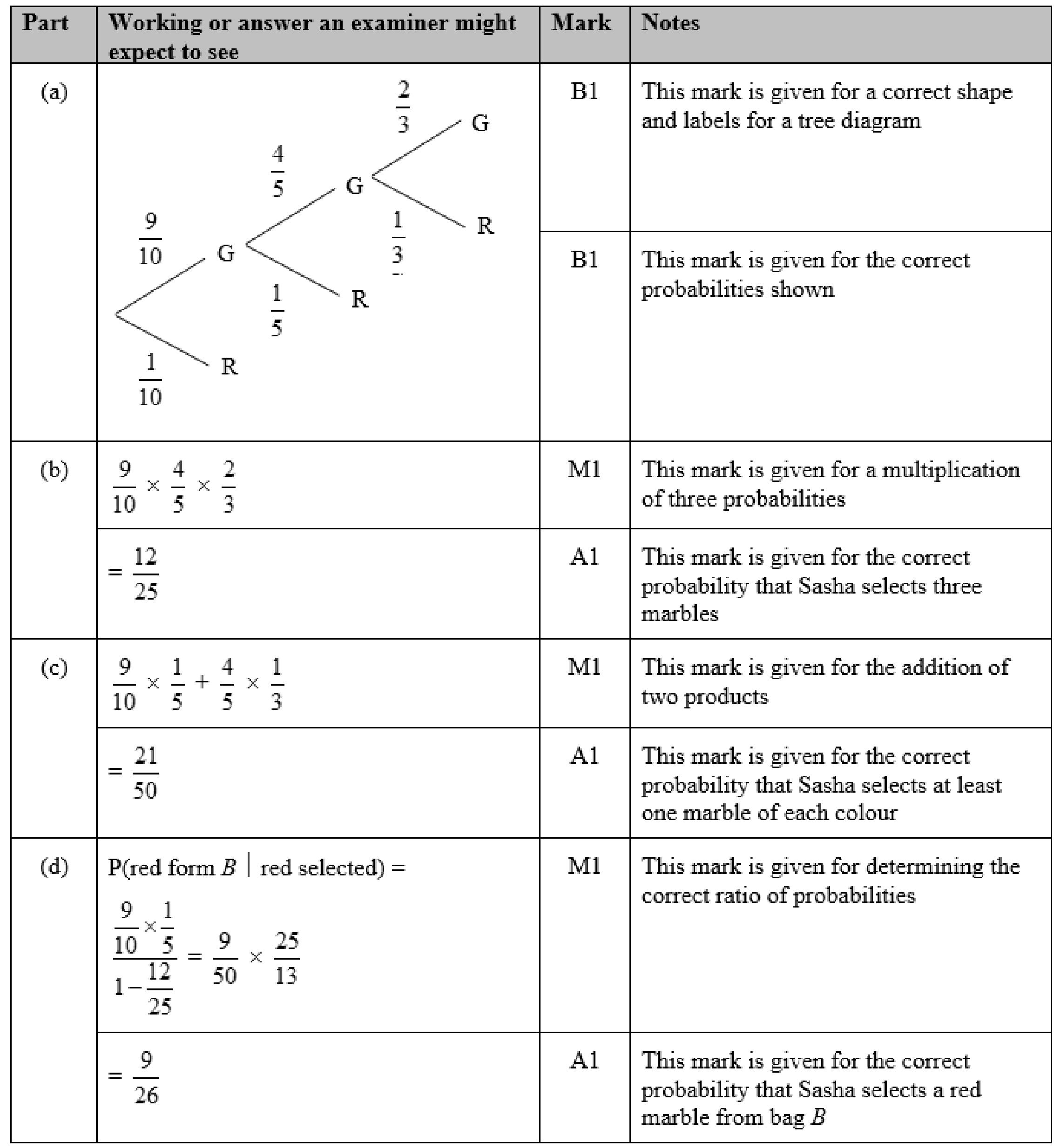
   (c) Find the probability that Sasha selects at least 1 marble of each colour.
   **(2)**

   (d) Given that Sasha selects a red marble, find the probability that he selects it from bag *B*.
   **(2)**

HOME

# A2 2019

## Conditional Probability



| Part | Working or answer an examiner might expect to see | Mark | Notes |
|---|---|---|---|
| (a) |  $\frac{9}{10}$<br>G<br>$\frac{1}{10}$<br>R<br>$\frac{4}{5}$<br>G<br>$\frac{1}{5}$<br>R<br>$\frac{2}{3}$<br>G<br>$\frac{1}{3}$<br>R | B1 | This mark is given for a correct shape and labels for a tree diagram |
| | | B1 | This mark is given for the correct probabilities shown |
| (b) | $\frac{9}{10} \times \frac{4}{5} \times \frac{2}{3}$ | M1 | This mark is given for a multiplication of three probabilities |
| | $= \frac{12}{25}$ | A1 | This mark is given for the correct probability that Sasha selects three marbles |
| (c) | $\frac{9}{10} \times \frac{1}{5} + \frac{4}{5} \times \frac{1}{3}$ | M1 | This mark is given for the addition of two products |
| | $= \frac{21}{50}$ | A1 | This mark is given for the correct probability that Sasha selects at least one marble of each colour |
| (d) | P(red form *B* \| red selected) =<br>$\dfrac{\frac{9}{10} \times \frac{1}{5}}{1 - \frac{12}{25}} = \frac{9}{50} \times \frac{25}{13}$ | M1 | This mark is given for determining the correct ratio of probabilities |
| | $= \frac{9}{26}$ | A1 | This mark is given for the correct probability that Sasha selects a red marble from bag *B* |

**A2 2020**

Conditional Probability

1. The Venn diagram shows the probabilities associated with four events, $A$, $B$, $C$ and $D$

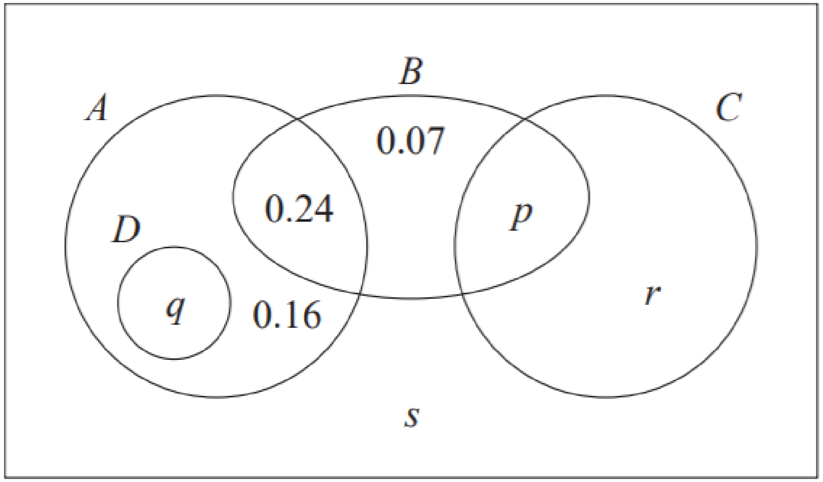
(a) Write down any pair of mutually exclusive events from $A$, $B$, $C$ and $D$

**(1)**

Given that P($B$) = 0.4

(b) find the value of $p$

**(1)**



Given also that $A$ and $B$ are independent

(c) find the value of $q$

**(2)**

Given further that $\mathrm{P}(B' \mid C) = 0.64$

(d) find

(i) the value of $r$

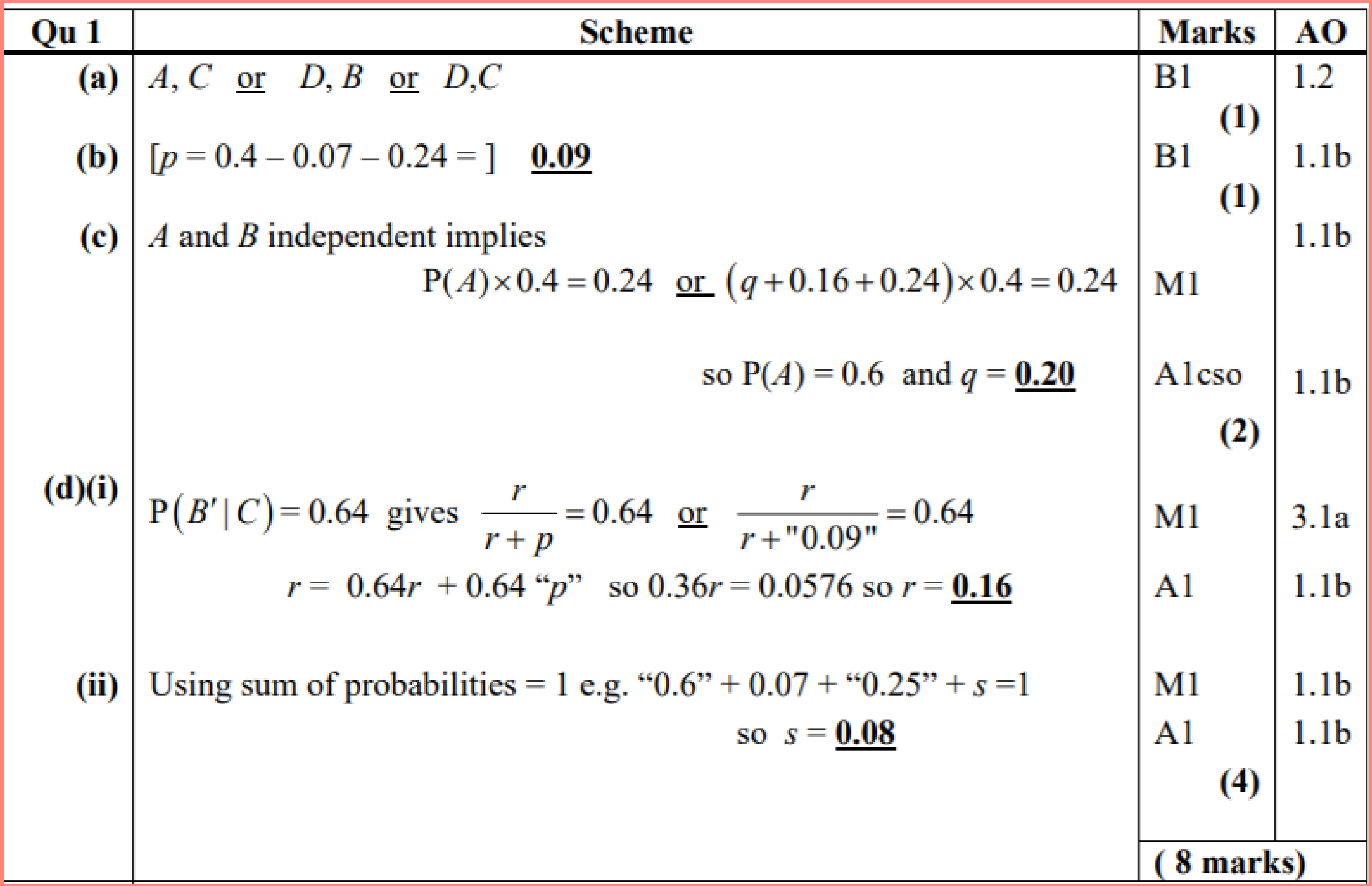
(ii) the value of $s$

**(4)**

HOME

# A2 2020

## Conditional Probability

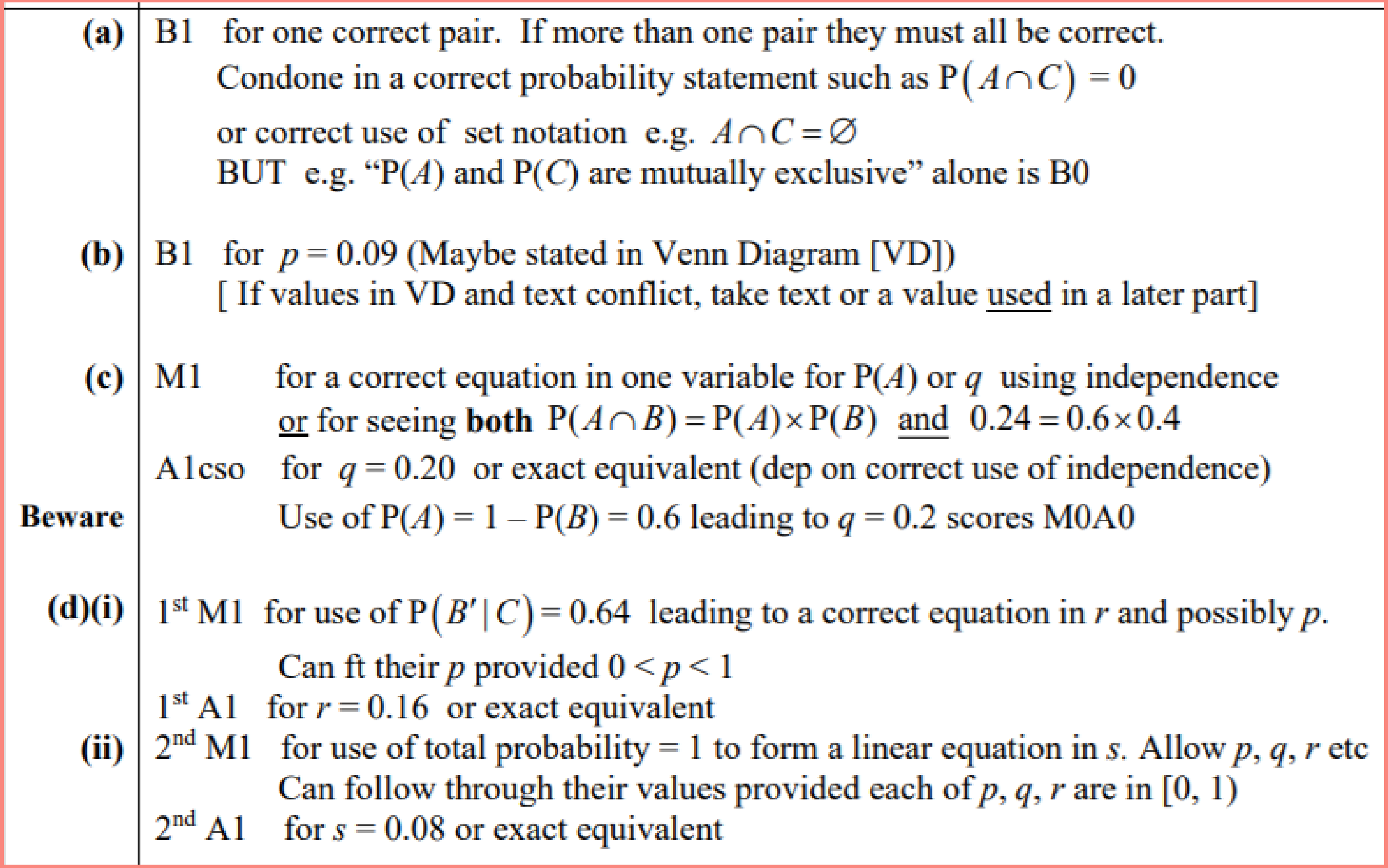


| Qu 1 | Scheme | Marks | AO |
|---|---|---|---|
| (a) | $A, C$ <u>or</u> $D, B$ <u>or</u> $D,C$ | B1<br>(1) | 1.2 |
| (b) | $[p = 0.4 - 0.07 - 0.24 = ]$ $\underline{\mathbf{0.09}}$ | B1<br>(1) | 1.1b |
| (c) | $A$ and $B$ independent implies | | 1.1b |
| | $\mathrm{P}(A)\times 0.4 = 0.24$ <u>or</u> $(q+0.16+0.24)\times 0.4 = 0.24$ | M1 | |
| | so $\mathrm{P}(A) = 0.6$ and $q = \underline{\mathbf{0.20}}$ | A1cso<br>(2) | 1.1b |
| (d)(i) | $\mathrm{P}(B' \mid C) = 0.64$ gives $\dfrac{r}{r+p} = 0.64$ <u>or</u> $\dfrac{r}{r+\text{"0.09"}} = 0.64$ | M1 | 3.1a |
| | $r = 0.64r + 0.64$ “$p$” so $0.36r = 0.0576$ so $r = \underline{\mathbf{0.16}}$ | A1 | 1.1b |
| (ii) | Using sum of probabilities = 1 e.g. “0.6” + 0.07 + “0.25” + $s$ =1 | M1 | 1.1b |
| | so $s = \underline{\mathbf{0.08}}$ | A1<br>(4) | 1.1b |
| | | **( 8 marks)** | |

HOME

**A2 2020**

## Conditional Probability



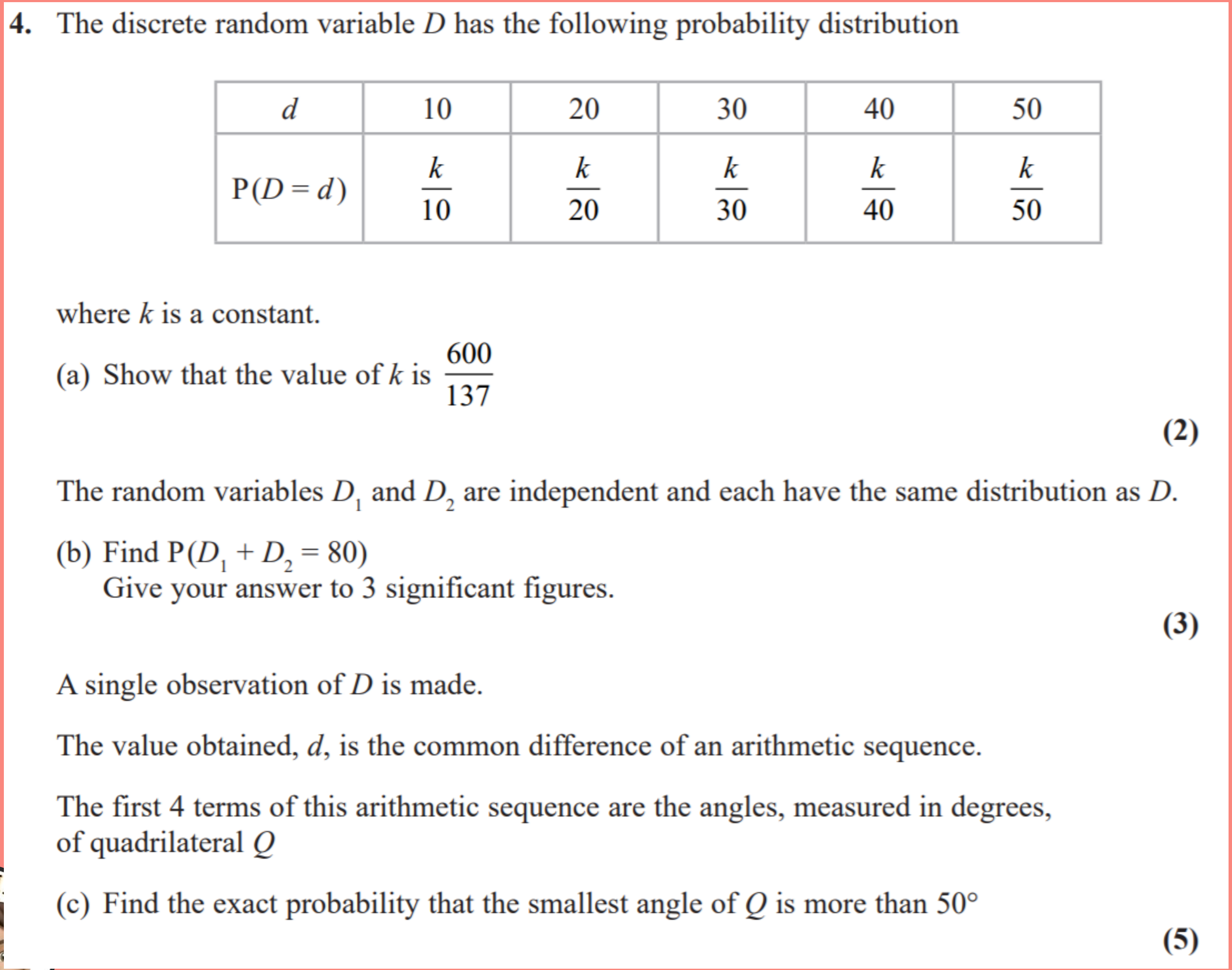
| | |
|---|---|
| **(a)** | B1 for one correct pair. If more than one pair they must all be correct.<br>Condone in a correct probability statement such as $\mathrm{P}(A \cap C) = 0$<br>or correct use of set notation e.g. $A \cap C = \varnothing$<br>BUT e.g. "P($A$) and P($C$) are mutually exclusive" alone is B0 |
| **(b)** | B1 for $p = 0.09$ (Maybe stated in Venn Diagram [VD])<br>[ If values in VD and text conflict, take text or a value used in a later part] |
| **(c)** | M1 for a correct equation in one variable for P($A$) or $q$ using independence<br>or for seeing **both** $\mathrm{P}(A \cap B) = \mathrm{P}(A) \times \mathrm{P}(B)$ and $0.24 = 0.6 \times 0.4$<br>A1cso for $q = 0.20$ or exact equivalent (dep on correct use of independence) |
| **Beware** | Use of P($A$) = 1 – P($B$) = 0.6 leading to $q = 0.2$ scores M0A0 |
| **(d)(i)** | 1st M1 for use of $\mathrm{P}(B' \mid C) = 0.64$ leading to a correct equation in $r$ and possibly $p$.<br>Can ft their $p$ provided $0 < p < 1$<br>1st A1 for $r = 0.16$ or exact equivalent |
| **(ii)** | 2nd M1 for use of total probability = 1 to form a linear equation in $s$. Allow $p$, $q$, $r$ etc<br>Can follow through their values provided each of $p$, $q$, $r$ are in [0, 1)<br>2nd A1 for $s = 0.08$ or exact equivalent |

HOME

# A2 2020

Conditional Probability

*Note: this is included in A2 section as it contains A2 Pure Series content. The Statistics is AS standard.*



**4.** The discrete random variable $D$ has the following probability distribution

| $d$ | 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|
| $P(D = d)$ | $\frac{k}{10}$ | $\frac{k}{20}$ | $\frac{k}{30}$ | $\frac{k}{40}$ | $\frac{k}{50}$ |

where $k$ is a constant.

(a) Show that the value of $k$ is $\frac{600}{137}$

**(2)**

The random variables $D_1$ and $D_2$ are independent and each have the same distribution as $D$.

(b) Find $P(D_1 + D_2 = 80)$
Give your answer to 3 significant figures.

**(3)**

A single observation of $D$ is made.

The value obtained, $d$, is the common difference of an arithmetic sequence.

The first 4 terms of this arithmetic sequence are the angles, measured in degrees, of quadrilateral $Q$

(c) Find the exact probability that the smallest angle of $Q$ is more than 50°

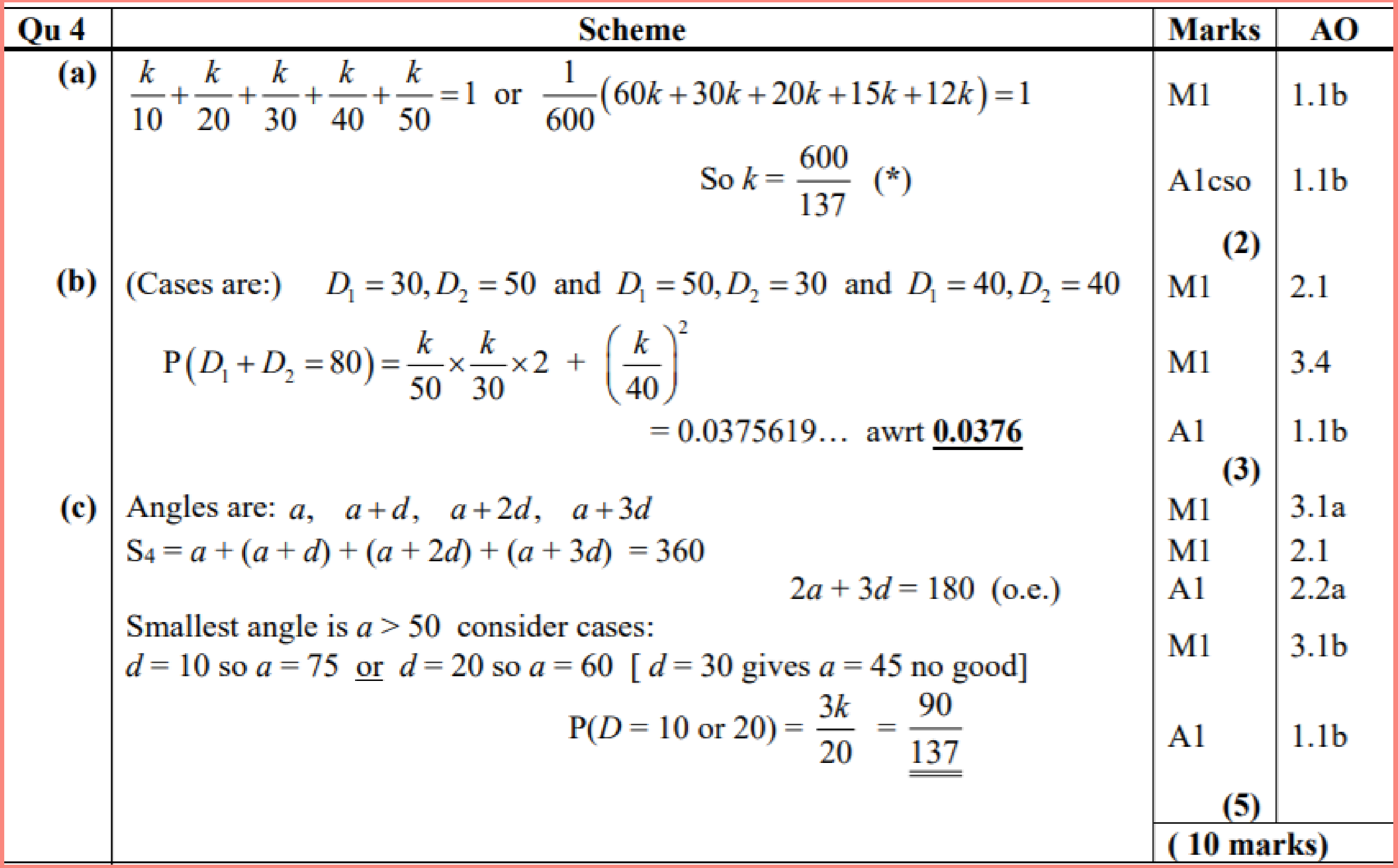
**(5)**

HOME

# A2 2020

Conditional Probability

*Note: this is included in A2 section as it contains A2 Pure Series content. The Statistics is AS standard.*

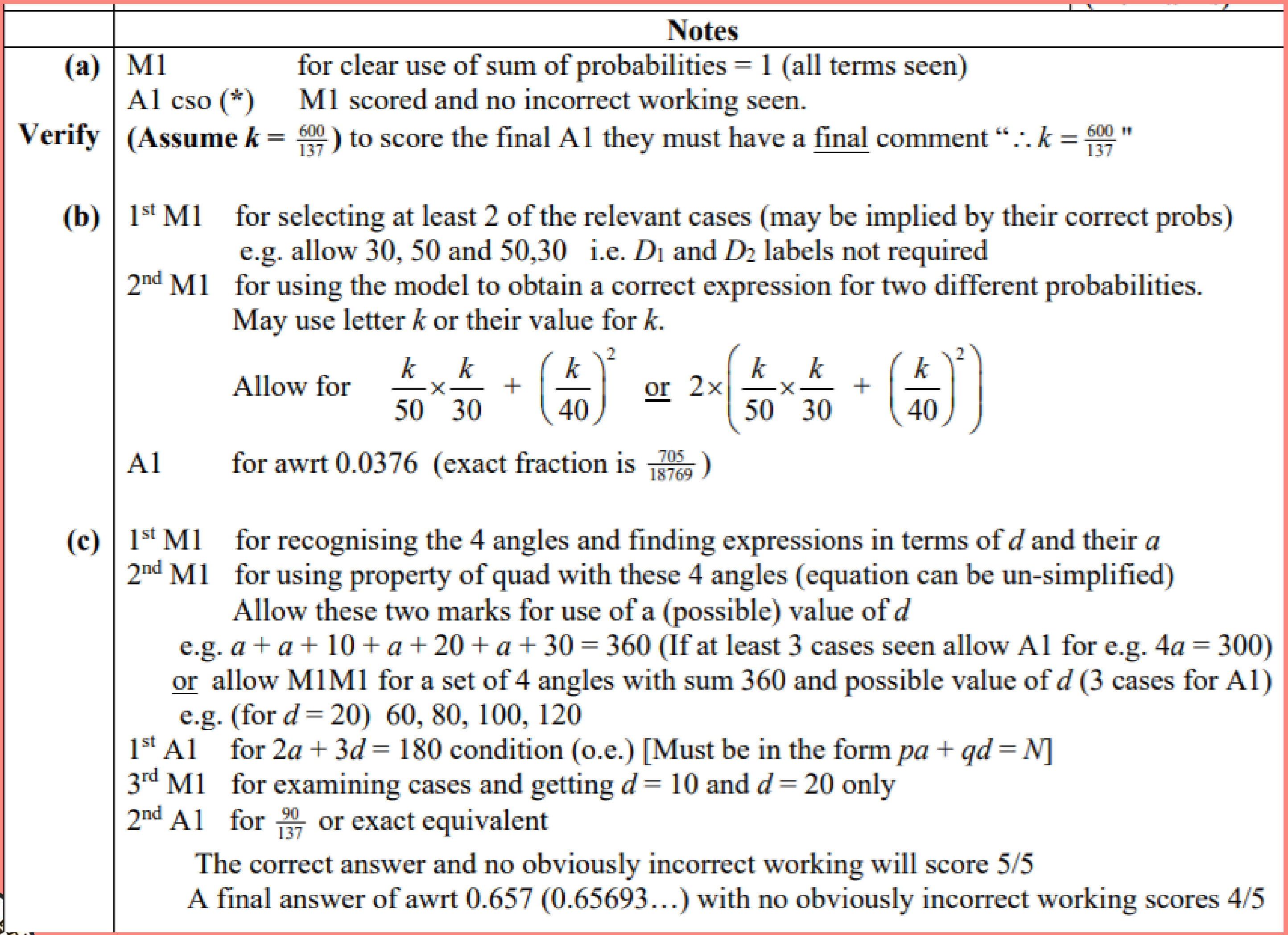


| Qu 4 | Scheme | Marks | AO |
|---|---|---|---|
| (a) | $\frac{k}{10}+\frac{k}{20}+\frac{k}{30}+\frac{k}{40}+\frac{k}{50}=1$ or $\frac{1}{600}\left(60k+30k+20k+15k+12k\right)=1$ | M1 | 1.1b |
| | So $k=\frac{600}{137}$ (*) | A1cso | 1.1b |
| | | **(2)** | |
| (b) | (Cases are:) $D_1=30, D_2=50$ and $D_1=50, D_2=30$ and $D_1=40, D_2=40$ | M1 | 2.1 |
| | $\mathrm{P}\left(D_1+D_2=80\right)=\frac{k}{50}\times\frac{k}{30}\times 2+\left(\frac{k}{40}\right)^2$ | M1 | 3.4 |
| | $=0.0375619\ldots$ awrt $\underline{\mathbf{0.0376}}$ | A1 | 1.1b |
| | | **(3)** | |
| (c) | Angles are: $a, \quad a+d, \quad a+2d, \quad a+3d$ | M1 | 3.1a |
| | $S_4=a+(a+d)+(a+2d)+(a+3d)=360$ | M1 | 2.1 |
| | $2a+3d=180$ (o.e.) | A1 | 2.2a |
| | Smallest angle is $a>50$ consider cases:<br>$d=10$ so $a=75$ or $d=20$ so $a=60$ [ $d=30$ gives $a=45$ no good] | M1 | 3.1b |
| | $\mathrm{P}(D=10 \text{ or } 20)=\frac{3k}{20}=\underline{\underline{\frac{90}{137}}}$ | A1 | 1.1b |
| | | **(5)** | |
| | | **( 10 marks)** | |

# A2 2020

## Conditional Probability

*Note: this is included in A2 section as it contains A2 Pure Series content. The Statistics is AS standard.*



| | Notes |
|---|---|
| **(a)**<br><br>**Verify** | M1 for clear use of sum of probabilities = 1 (all terms seen)<br>A1 cso (*) M1 scored and no incorrect working seen.<br>**(Assume $k = \frac{600}{137}$)** to score the final A1 they must have a final comment "$\therefore k = \frac{600}{137}$" |
| **(b)** | 1st M1 for selecting at least 2 of the relevant cases (may be implied by their correct probs) e.g. allow 30, 50 and 50,30 i.e. $D_1$ and $D_2$ labels not required<br>2nd M1 for using the model to obtain a correct expression for two different probabilities. May use letter $k$ or their value for $k$.<br>Allow for $\frac{k}{50}\times\frac{k}{30} + \left(\frac{k}{40}\right)^2$ or $2\times\left(\frac{k}{50}\times\frac{k}{30} + \left(\frac{k}{40}\right)^2\right)$<br>A1 for awrt 0.0376 (exact fraction is $\frac{705}{18769}$) |
| **(c)** | 1st M1 for recognising the 4 angles and finding expressions in terms of $d$ and their $a$<br>2nd M1 for using property of quad with these 4 angles (equation can be un-simplified) Allow these two marks for use of a (possible) value of $d$<br>e.g. $a + a + 10 + a + 20 + a + 30 = 360$ (If at least 3 cases seen allow A1 for e.g. $4a = 300$)<br>or allow M1M1 for a set of 4 angles with sum 360 and possible value of $d$ (3 cases for A1)<br>e.g. (for $d = 20$) 60, 80, 100, 120<br>1st A1 for $2a + 3d = 180$ condition (o.e.) [Must be in the form $pa + qd = N$]<br>3rd M1 for examining cases and getting $d = 10$ and $d = 20$ only<br>2nd A1 for $\frac{90}{137}$ or exact equivalent<br>The correct answer and no obviously incorrect working will score 5/5<br>A final answer of awrt 0.657 (0.65693…) with no obviously incorrect working scores 4/5 |

HOME

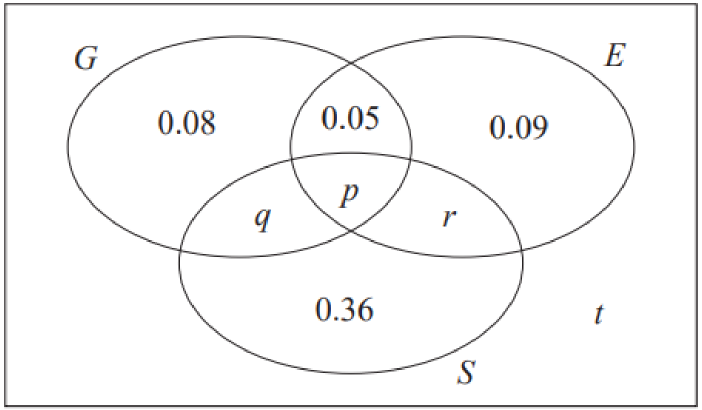
# A2 2021

Conditional Probability

4. A large college produces three magazines.
One magazine is about green issues, one is about equality and one is about sports.
A student at the college is selected at random and the events $G$, $E$ and $S$ are defined as follows

$G$ is the event that the student reads the magazine about green issues
$E$ is the event that the student reads the magazine about equality
$S$ is the event that the student reads the magazine about sports

The Venn diagram, where $p$, $q$, $r$ and $t$ are probabilities, gives the probability for each subset.



(a) Find the proportion of students in the college who read exactly one of these magazines.

**(1)**

No students read all three magazines and $\mathrm{P}(G) = 0.25$

(b) Find

(i) the value of $p$

(ii) the value of $q$

**(3)**

Given that $\mathrm{P}(S \mid E) = \dfrac{5}{12}$

(c) find

(i) the value of $r$

(ii) the value of $t$

**(4)**

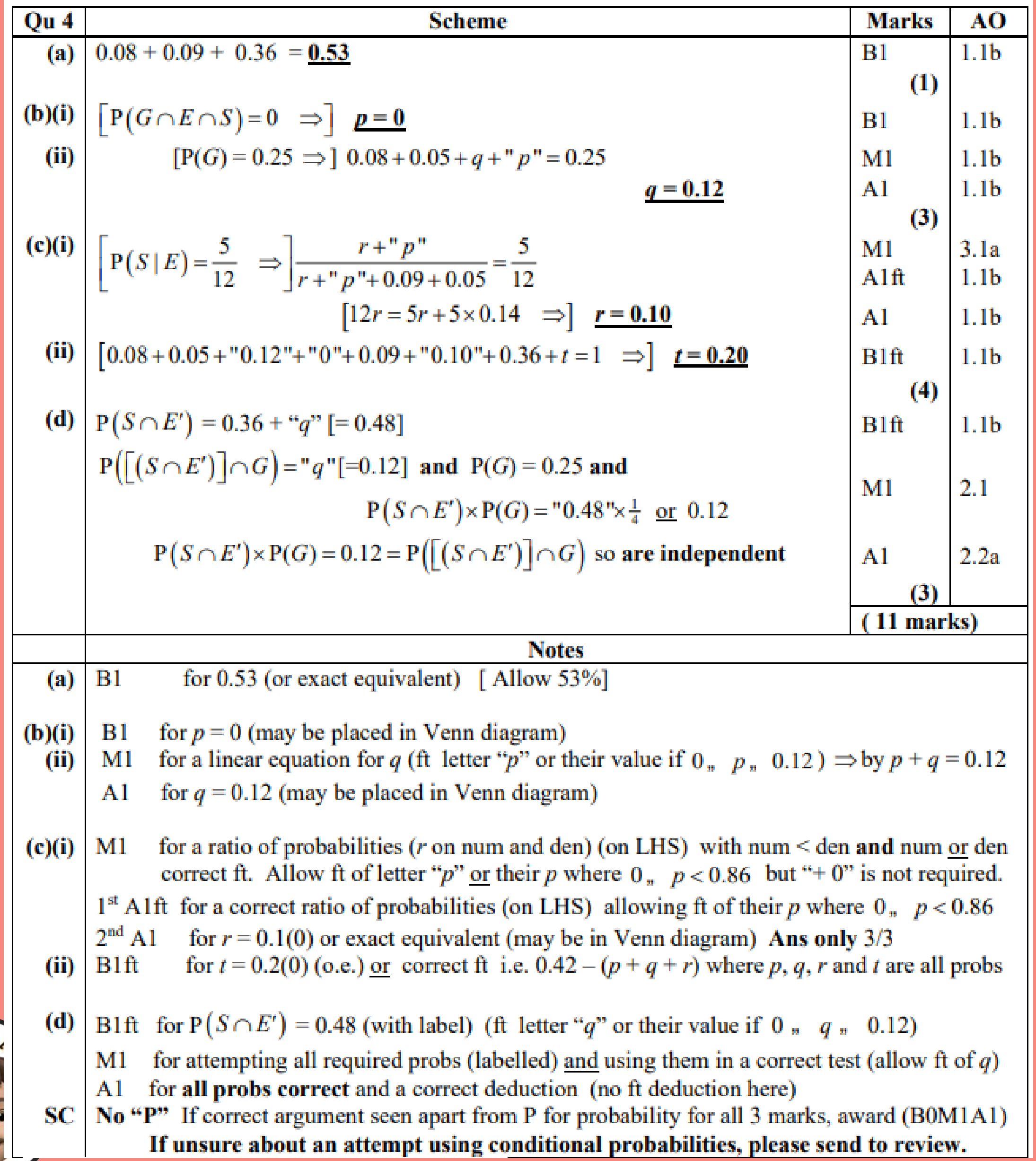
(d) Determine whether or not the events $(S \cap E')$ and $G$ are independent.
Show your working clearly.

**(3)**

HOME

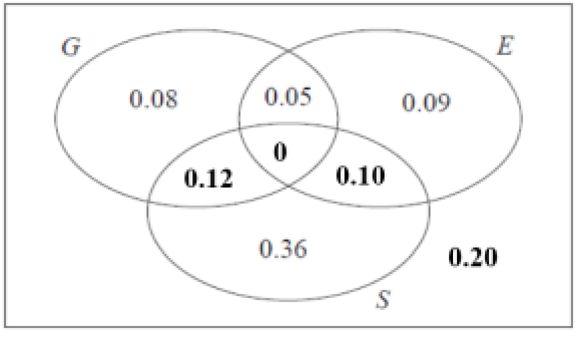
# A2 2021

## Conditional Probability



| Qu 4 | Scheme | Marks | AO |
|---|---|---|---|
| **(a)** | $0.08 + 0.09 + 0.36 = \underline{\mathbf{0.53}}$ | B1 | 1.1b |
| | | **(1)** | |
| **(b)(i)** | $\left[\mathrm{P}(G \cap E \cap S) = 0 \quad \Rightarrow\right] \quad \underline{\mathbf{p = 0}}$ | B1 | 1.1b |
| **(ii)** | $[\mathrm{P}(G) = 0.25 \Rightarrow]\ 0.08 + 0.05 + q + "p" = 0.25$ | M1 | 1.1b |
| | $\underline{\mathbf{q = 0.12}}$ | A1 | 1.1b |
| | | **(3)** | |
| **(c)(i)** | $\left[\mathrm{P}(S \mid E) = \frac{5}{12} \quad \Rightarrow\right] \frac{r + "p"}{r + "p" + 0.09 + 0.05} = \frac{5}{12}$ | M1<br>A1ft | 3.1a<br>1.1b |
| | $\left[12r = 5r + 5 \times 0.14 \quad \Rightarrow\right] \quad \underline{\mathbf{r = 0.10}}$ | A1 | 1.1b |
| **(ii)** | $\left[0.08 + 0.05 + "0.12" + "0" + 0.09 + "0.10" + 0.36 + t = 1 \quad \Rightarrow\right] \quad \underline{\mathbf{t = 0.20}}$ | B1ft | 1.1b |
| | | **(4)** | |
| **(d)** | $\mathrm{P}(S \cap E') = 0.36 +$ "$q$" $[= 0.48]$ | B1ft | 1.1b |
| | $\mathrm{P}\left(\left[(S \cap E')\right] \cap G\right) = "q"[=0.12]$ **and** $\mathrm{P}(G) = 0.25$ **and**<br>$\mathrm{P}(S \cap E') \times \mathrm{P}(G) = "0.48" \times \frac{1}{4}$ <u>or</u> $0.12$ | M1 | 2.1 |
| | $\mathrm{P}(S \cap E') \times \mathrm{P}(G) = 0.12 = \mathrm{P}\left(\left[(S \cap E')\right] \cap G\right)$ so **are independent** | A1 | 2.2a |
| | | **(3)** | |
| | | **( 11 marks)** | |

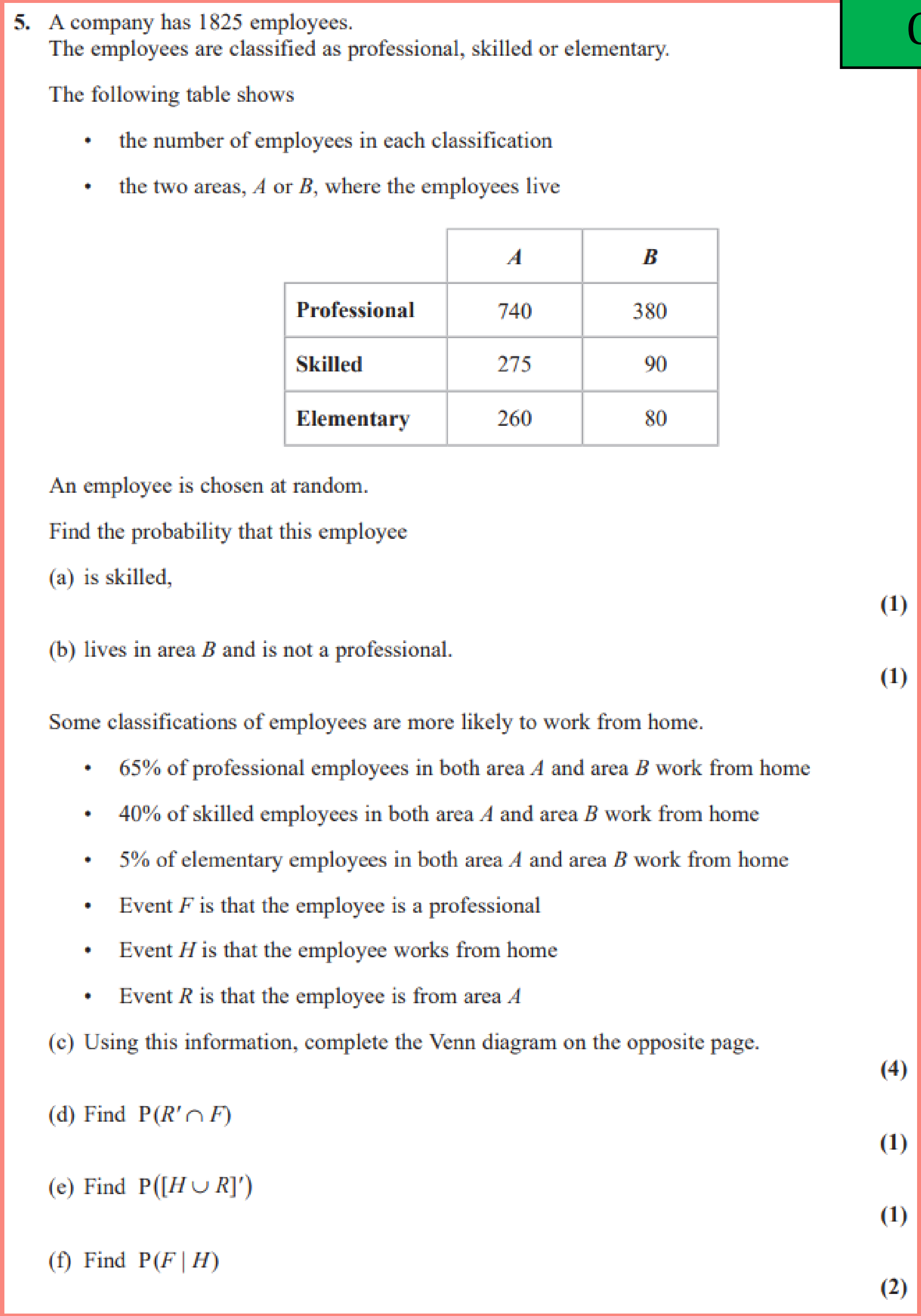
| | **Notes** |
|---|---|
| **(a)** | B1 for 0.53 (or exact equivalent) [ Allow 53%] |
| **(b)(i)** | B1 for $p = 0$ (may be placed in Venn diagram) |
| **(ii)** | M1 for a linear equation for $q$ (ft letter "$p$" or their value if $0$ „ $p$ „ $0.12$ ) $\Rightarrow$ by $p + q = 0.12$ |
| | A1 for $q = 0.12$ (may be placed in Venn diagram) |
| **(c)(i)** | M1 for a ratio of probabilities ($r$ on num and den) (on LHS) with num < den **and** num <u>or</u> den correct ft. Allow ft of letter "$p$" <u>or</u> their $p$ where $0$ „ $p < 0.86$ but "+ 0" is not required. |
| | 1st A1ft for a correct ratio of probabilities (on LHS) allowing ft of their $p$ where $0$ „ $p < 0.86$ |
| | 2nd A1 for $r = 0.1(0)$ or exact equivalent (may be in Venn diagram) **Ans only** 3/3 |
| **(ii)** | B1ft for $t = 0.2(0)$ (o.e.) <u>or</u> correct ft i.e. $0.42 - (p + q + r)$ where $p$, $q$, $r$ and $t$ are all probs |
| **(d)** | B1ft for $\mathrm{P}(S \cap E') = 0.48$ (with label) (ft letter "$q$" or their value if $0$ „ $q$ „ $0.12$) |
| | M1 for attempting all required probs (labelled) <u>and</u> using them in a correct test (allow ft of $q$) |
| | A1 for **all probs correct** and a correct deduction (no ft deduction here) |
| **SC** | **No "P"** If correct argument seen apart from P for probability for all 3 marks, award (B0M1A1)<br>**If unsure about an attempt using conditional probabilities, please send to review.** |



HOME

# A2 2022

# Conditional Probability



**5.** A company has 1825 employees.
The employees are classified as professional, skilled or elementary.

The following table shows

- the number of employees in each classification
- the two areas, $A$ or $B$, where the employees live

| | $A$ | $B$ |
|---|---|---|
| **Professional** | 740 | 380 |
| **Skilled** | 275 | 90 |
| **Elementary** | 260 | 80 |

An employee is chosen at random.

Find the probability that this employee

(a) is skilled, **(1)**

(b) lives in area $B$ and is not a professional. **(1)**

Some classifications of employees are more likely to work from home.

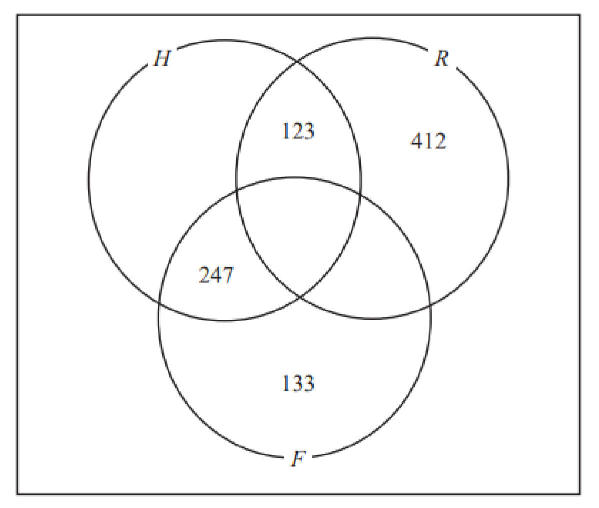
- 65% of professional employees in both area $A$ and area $B$ work from home
- 40% of skilled employees in both area $A$ and area $B$ work from home
- 5% of elementary employees in both area $A$ and area $B$ work from home
- Event $F$ is that the employee is a professional
- Event $H$ is that the employee works from home
- Event $R$ is that the employee is from area $A$

(c) Using this information, complete the Venn diagram on the opposite page. **(4)**

(d) Find $\mathrm{P}(R' \cap F)$ **(1)**

(e) Find $\mathrm{P}([H \cup R]')$ **(1)**

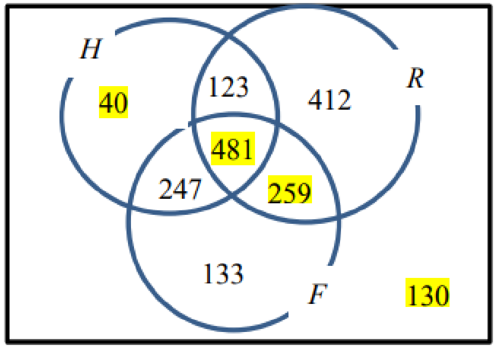
(f) Find $\mathrm{P}(F \mid H)$ **(2)**



HOME

# A2 2022

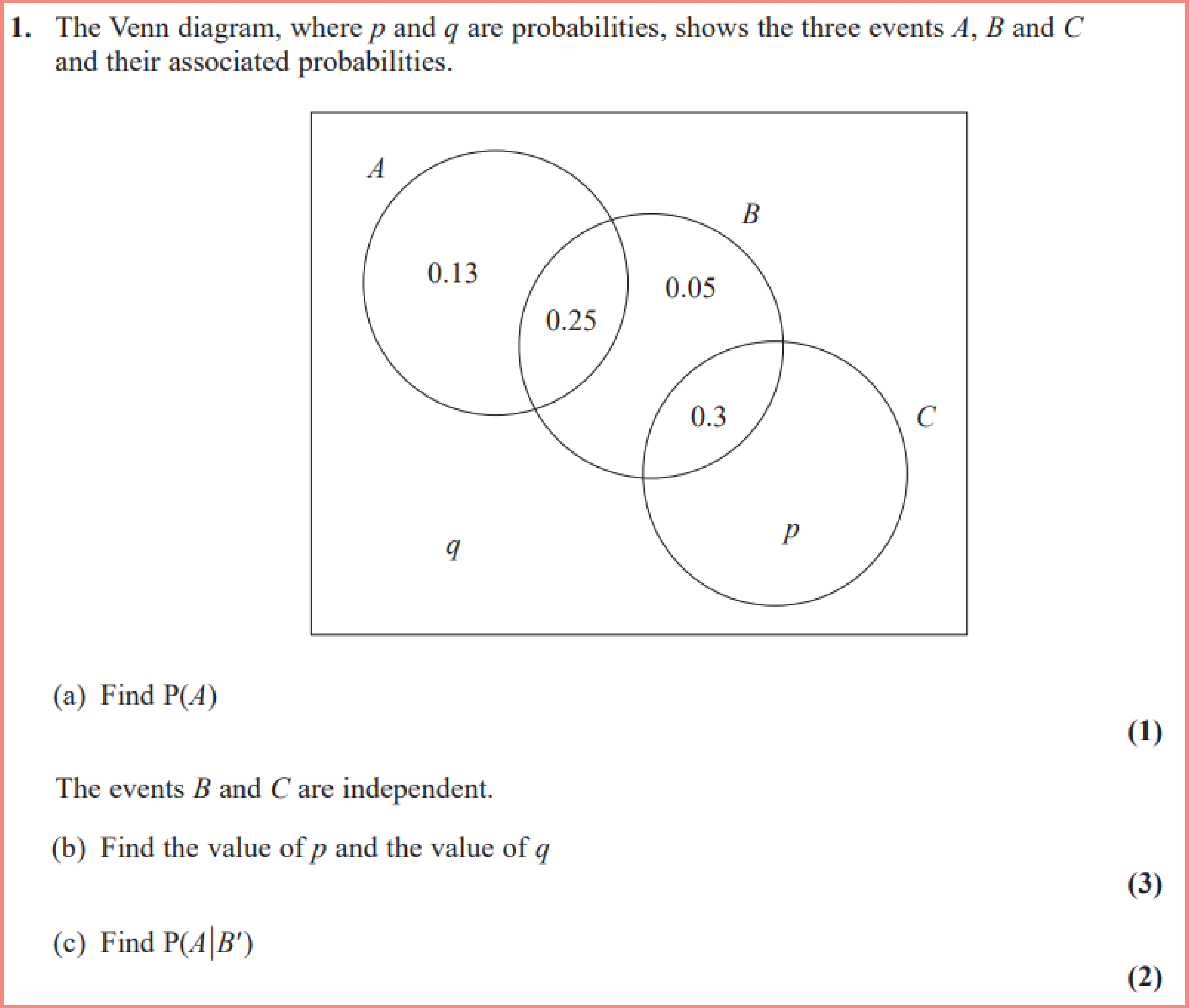
## Conditional Probability

| Question | Scheme | Marks | AOs |
|---|---|---|---|
| 5(a) | $\frac{365}{1825}$ or $\frac{1}{5}$ or 0.2 oe | B1 | 1.1b |
| | | **(1)** | |
| (b) | $\frac{170}{1825}$ or $\frac{34}{365}$ or awrt 0.093 | B1 | 1.1b |
| | | **(1)** | |
| (c) | $90\times0.4+80\times0.05[=40]$ or $90\times0.6+80\times0.95[=130]$ or $740\times0.65[=481]$ or $740\times0.35[=259]$<br><br> H<br>R<br>F<br>40<br>123<br>412<br>481<br>247<br>259<br>133<br>130 | M1<br>B1<br>B1<br>A1 | 3.1b<br>1.1b<br>1.1b<br>1.1b |
| | | **(4)** | |
| (d) | $P(R'\cap F)=\frac{380}{1825}\left[=\frac{76}{365}=0.208...\right]$ oe awrt 0.208 | B1 | 1.1b |
| | | **(1)** | |
| (e) | $\left[\frac{133\ +\ "130"}{1825}=\right]\frac{"263"}{1825}$ awrt 0.144 | B1ft | 1.1b |
| | | **(1)** | |
| (f) | $\frac{247+"481"}{247+"481"+123+"40"}$ | M1 | 3.4 |
| | $=\frac{728}{891}$ awrt 0.817 | A1 | 1.1b |
| | | **(2)** | |
| | **Notes:** | **(10 marks)** | |

HOME

# A2 2023

## Conditional Probability



1. The Venn diagram, where $p$ and $q$ are probabilities, shows the three events $A$, $B$ and $C$ and their associated probabilities.

A
B
C
0.13
0.25
0.05
0.3
$p$
$q$

(a) Find $\mathrm{P}(A)$

**(1)**

The events $B$ and $C$ are independent.

(b) Find the value of $p$ and the value of $q$

**(3)**

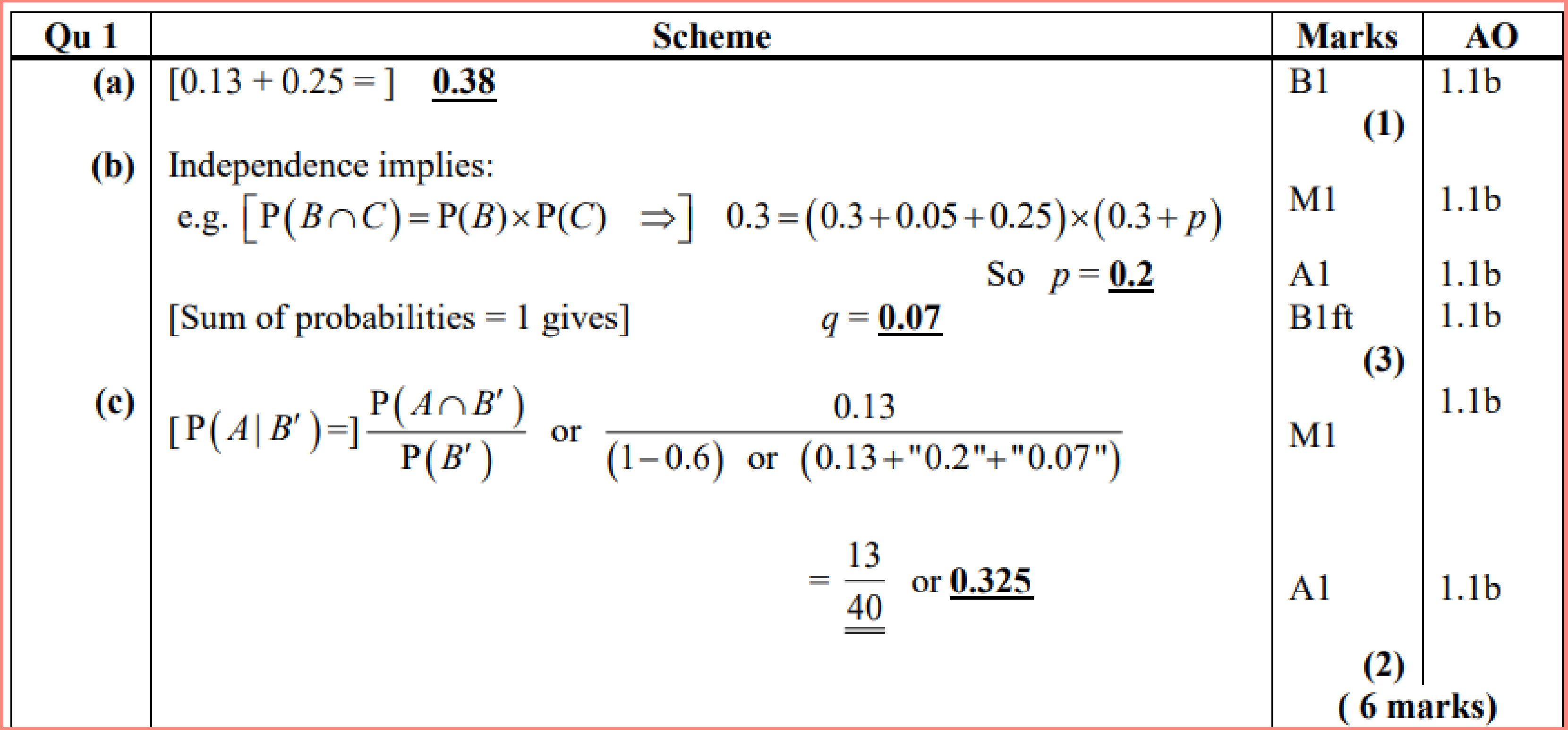
(c) Find $\mathrm{P}(A \mid B')$

**(2)**

HOME

# A2 2023

Conditional Probability

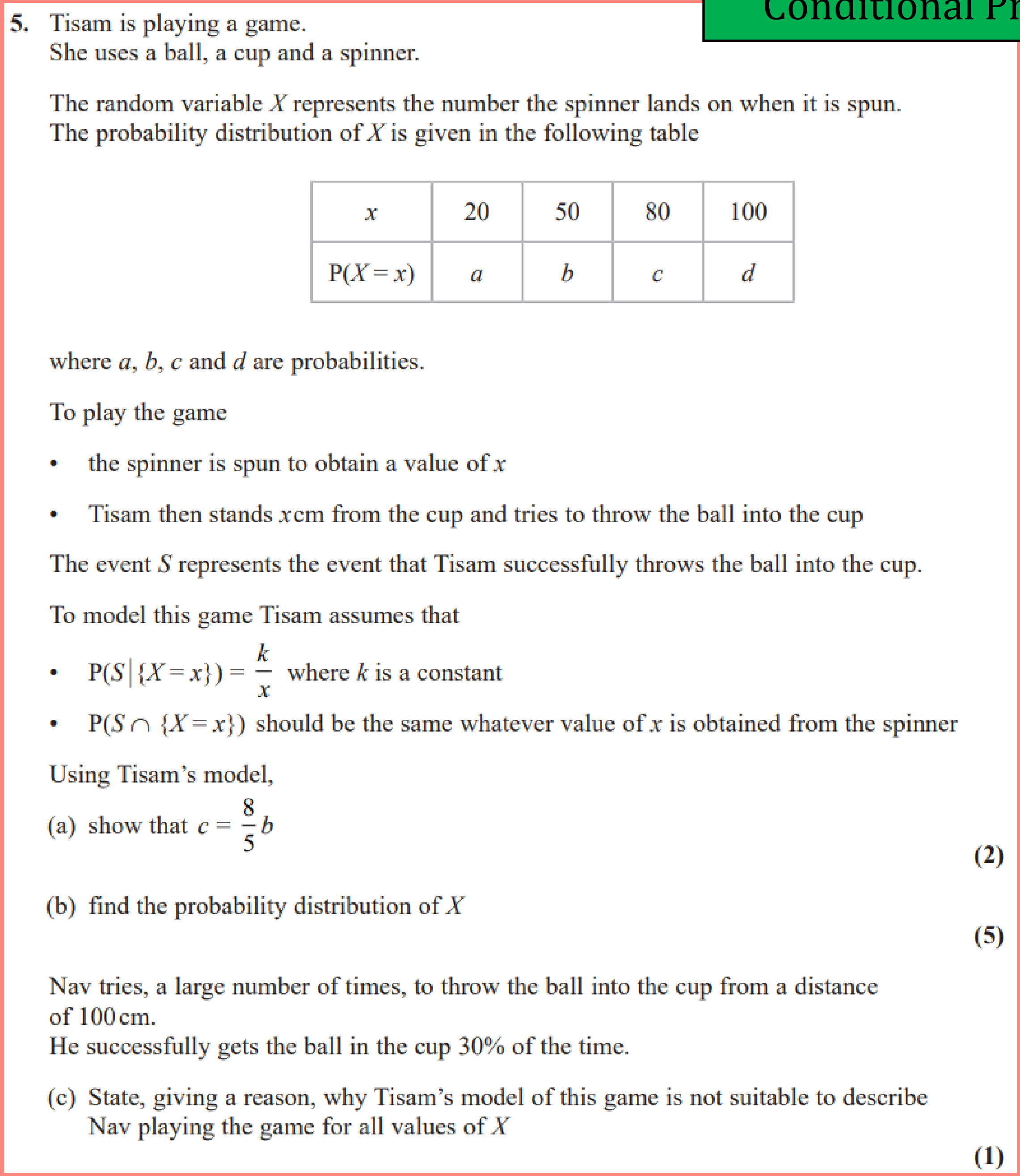


| Qu 1 | Scheme | Marks | AO |
|---|---|---|---|
| **(a)** | [0.13 + 0.25 = ] $\underline{\mathbf{0.38}}$ | B1<br>**(1)** | 1.1b |
| **(b)** | Independence implies:<br>e.g. $\left[\mathrm{P}(B \cap C) = \mathrm{P}(B) \times \mathrm{P}(C) \quad \Rightarrow\right] \quad 0.3 = (0.3 + 0.05 + 0.25) \times (0.3 + p)$ | M1 | 1.1b |
| | So $p = \underline{\mathbf{0.2}}$ | A1 | 1.1b |
| | [Sum of probabilities = 1 gives] $q = \underline{\mathbf{0.07}}$ | B1ft<br>**(3)** | 1.1b |
| **(c)** | $[\mathrm{P}(A \mid B') =] \dfrac{\mathrm{P}(A \cap B')}{\mathrm{P}(B')}$ or $\dfrac{0.13}{(1-0.6) \text{ or } (0.13 + "0.2" + "0.07")}$ | M1 | 1.1b |
| | $= \dfrac{13}{\underline{\underline{40}}}$ or $\underline{\mathbf{0.325}}$ | A1<br>**(2)**<br>**( 6 marks)** | 1.1b |

HOME

A2 2023

# Conditional Probability



**5.** Tisam is playing a game.
She uses a ball, a cup and a spinner.

The random variable $X$ represents the number the spinner lands on when it is spun.
The probability distribution of $X$ is given in the following table

| $x$ | 20 | 50 | 80 | 100 |
|---|---|---|---|---|
| $\mathrm{P}(X = x)$ | $a$ | $b$ | $c$ | $d$ |

where $a$, $b$, $c$ and $d$ are probabilities.

To play the game

- the spinner is spun to obtain a value of $x$
- Tisam then stands $x$ cm from the cup and tries to throw the ball into the cup

The event $S$ represents the event that Tisam successfully throws the ball into the cup.

To model this game Tisam assumes that

- $\mathrm{P}(S \mid \{X = x\}) = \dfrac{k}{x}$ where $k$ is a constant
- $\mathrm{P}(S \cap \{X = x\})$ should be the same whatever value of $x$ is obtained from the spinner

Using Tisam's model,

(a) show that $c = \dfrac{8}{5}b$

**(2)**

(b) find the probability distribution of $X$

**(5)**

Nav tries, a large number of times, to throw the ball into the cup from a distance of 100 cm.
He successfully gets the ball in the cup 30% of the time.

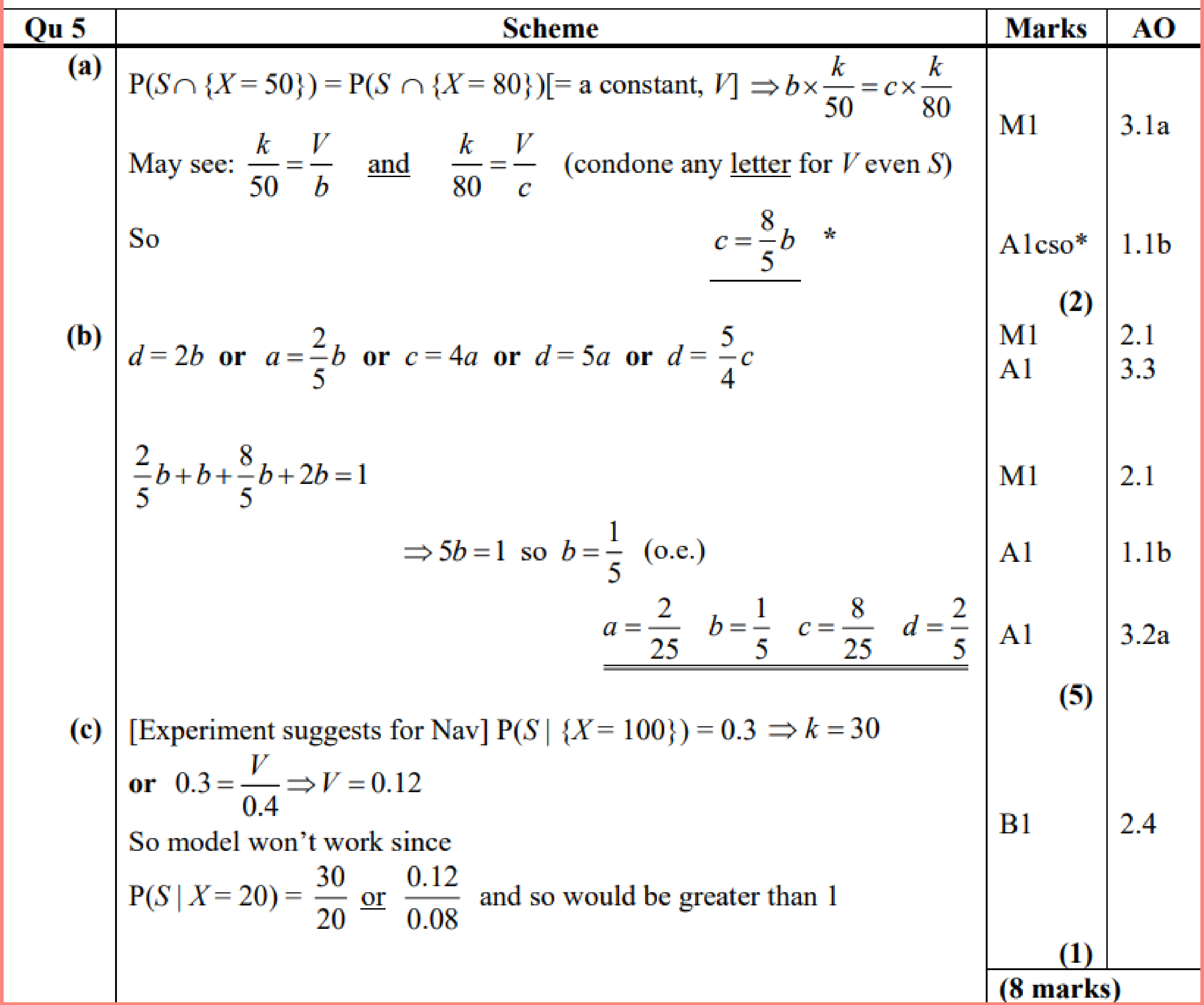
(c) State, giving a reason, why Tisam's model of this game is not suitable to describe Nav playing the game for all values of $X$

**(1)**

HOME

# A2 2023

## Conditional Probability



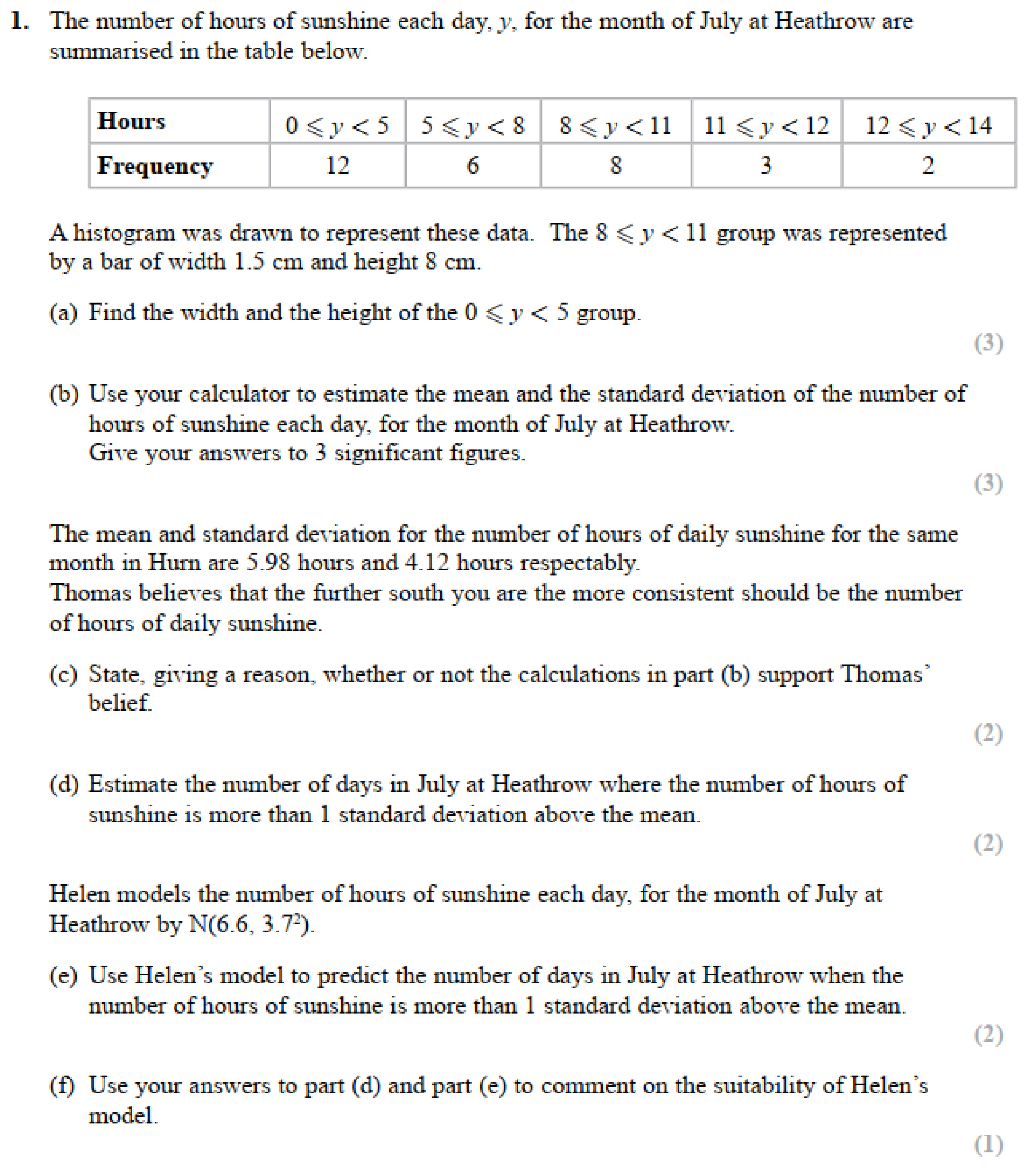
| Qu 5 | Scheme | Marks | AO |
|---|---|---|---|
| (a) | $\mathrm{P}(S\cap\{X=50\})=\mathrm{P}(S\cap\{X=80\})$[= a constant, $V$] $\Rightarrow b\times\frac{k}{50}=c\times\frac{k}{80}$<br>May see: $\frac{k}{50}=\frac{V}{b}$ <u>and</u> $\frac{k}{80}=\frac{V}{c}$ (condone any <u>letter</u> for $V$ even $S$) | M1 | 3.1a |
| | So $\underline{c=\frac{8}{5}b}$ * | A1cso* | 1.1b |
| | | (2) | |
| (b) | $d=2b$ **or** $a=\frac{2}{5}b$ **or** $c=4a$ **or** $d=5a$ **or** $d=\frac{5}{4}c$ | M1<br>A1 | 2.1<br>3.3 |
| | $\frac{2}{5}b+b+\frac{8}{5}b+2b=1$ | M1 | 2.1 |
| | $\Rightarrow 5b=1$ so $b=\frac{1}{5}$ (o.e.) | A1 | 1.1b |
| | $\underline{\underline{a=\frac{2}{25}\quad b=\frac{1}{5}\quad c=\frac{8}{25}\quad d=\frac{2}{5}}}$ | A1 | 3.2a |
| | | (5) | |
| (c) | [Experiment suggests for Nav] $\mathrm{P}(S\mid\{X=100\})=0.3\Rightarrow k=30$<br>**or** $0.3=\frac{V}{0.4}\Rightarrow V=0.12$<br>So model won't work since<br>$\mathrm{P}(S\mid X=20)=\frac{30}{20}$ <u>or</u> $\frac{0.12}{0.08}$ and so would be greater than 1 | B1 | 2.4 |
| | | (1) | |
| | | (8 marks) | |

HOME

# The Normal Distribution

# A2 SAMs

## The Normal Distribution



**1.** The number of hours of sunshine each day, $y$, for the month of July at Heathrow are summarised in the table below.

| **Hours** | $0 \leqslant y < 5$ | $5 \leqslant y < 8$ | $8 \leqslant y < 11$ | $11 \leqslant y < 12$ | $12 \leqslant y < 14$ |
|---|---|---|---|---|---|
| **Frequency** | 12 | 6 | 8 | 3 | 2 |

A histogram was drawn to represent these data. The $8 \leqslant y < 11$ group was represented by a bar of width 1.5 cm and height 8 cm.

(a) Find the width and the height of the $0 \leqslant y < 5$ group.

(3)

(b) Use your calculator to estimate the mean and the standard deviation of the number of hours of sunshine each day, for the month of July at Heathrow.
Give your answers to 3 significant figures.

(3)

The mean and standard deviation for the number of hours of daily sunshine for the same month in Hurn are 5.98 hours and 4.12 hours respectably.
Thomas believes that the further south you are the more consistent should be the number of hours of daily sunshine.

(c) State, giving a reason, whether or not the calculations in part (b) support Thomas' belief.

(2)

(d) Estimate the number of days in July at Heathrow where the number of hours of sunshine is more than 1 standard deviation above the mean.

(2)

Helen models the number of hours of sunshine each day, for the month of July at Heathrow by $\mathrm{N}(6.6, 3.7^2)$.

(e) Use Helen's model to predict the number of days in July at Heathrow when the number of hours of sunshine is more than 1 standard deviation above the mean.

(2)

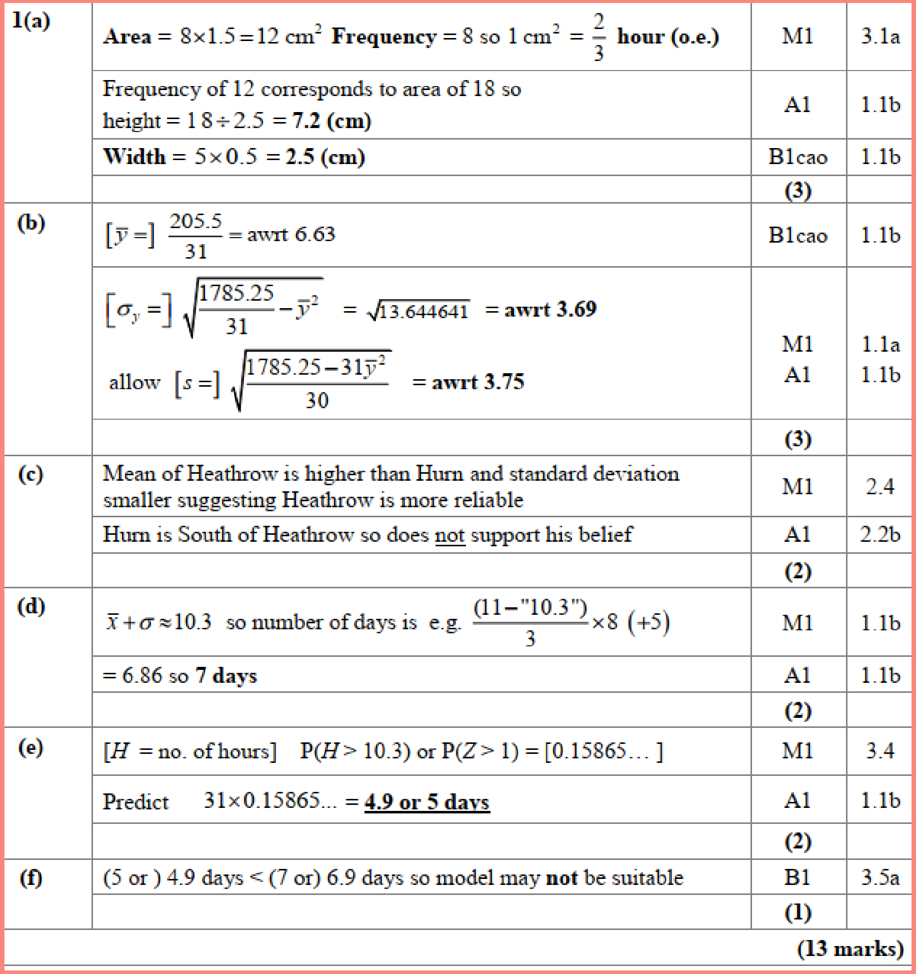
(f) Use your answers to part (d) and part (e) to comment on the suitability of Helen's model.

(1)

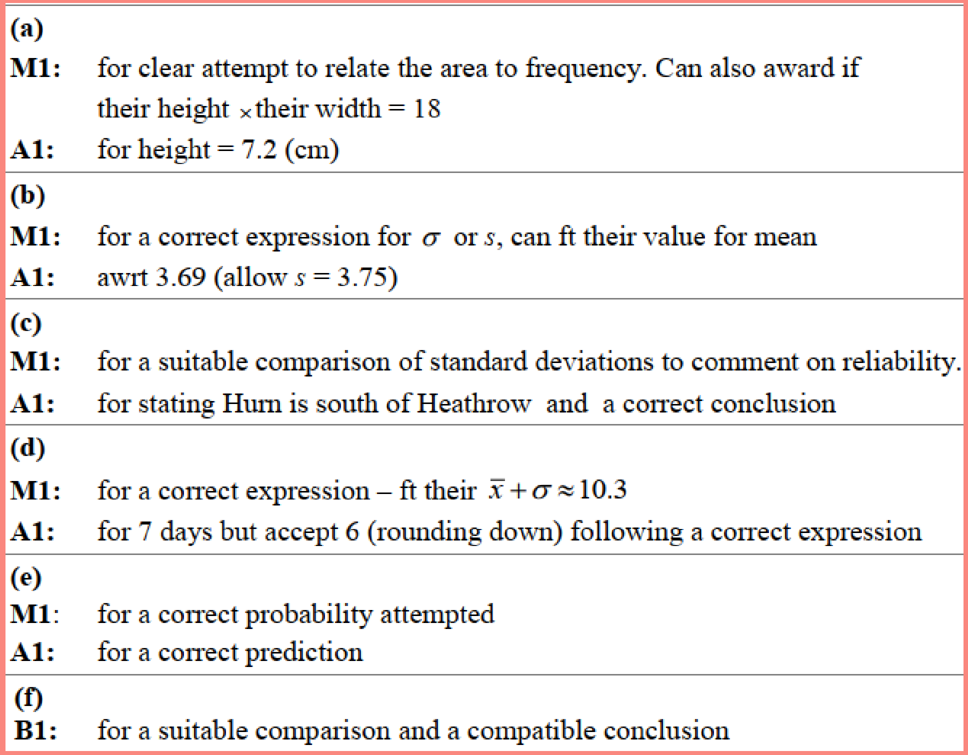
HOME

# A2 SAMs

## The Normal Distribution



| | | | |
|---|---|---|---|
| 1(a) | **Area** = $8\times1.5=12\text{ cm}^2$ **Frequency** = 8 so $1\text{ cm}^2=\frac{2}{3}$ **hour (o.e.)** | M1 | 3.1a |
| | Frequency of 12 corresponds to area of 18 so height = $18\div2.5$ = **7.2 (cm)** | A1 | 1.1b |
| | **Width** = $5\times0.5$ = **2.5 (cm)** | B1cao | 1.1b |
| | | **(3)** | |
| (b) | $[\bar{y}=]\ \frac{205.5}{31}$ = awrt 6.63 | B1cao | 1.1b |
| | $[\sigma_y=]\ \sqrt{\frac{1785.25}{31}-\bar{y}^2}=\sqrt{13.644641}$ = **awrt 3.69**<br>allow $[s=]\ \sqrt{\frac{1785.25-31\bar{y}^2}{30}}$ = **awrt 3.75** | M1<br>A1 | 1.1a<br>1.1b |
| | | **(3)** | |
| (c) | Mean of Heathrow is higher than Hurn and standard deviation smaller suggesting Heathrow is more reliable | M1 | 2.4 |
| | Hurn is South of Heathrow so does not support his belief | A1 | 2.2b |
| | | **(2)** | |
| (d) | $\bar{x}+\sigma\approx10.3$ so number of days is e.g. $\frac{(11-\text{"}10.3\text{"})}{3}\times8\ (+5)$ | M1 | 1.1b |
| | = 6.86 so **7 days** | A1 | 1.1b |
| | | **(2)** | |
| (e) | [$H$ = no. of hours] P($H$ > 10.3) or P($Z$ > 1) = [0.15865… ] | M1 | 3.4 |
| | Predict 31×0.15865… = **4.9 or 5 days** | A1 | 1.1b |
| | | **(2)** | |
| (f) | (5 or ) 4.9 days < (7 or) 6.9 days so model may **not** be suitable | B1 | 3.5a |
| | | **(1)** | |
| | | **(13 marks)** | |



**(a)**

**M1:** for clear attempt to relate the area to frequency. Can also award if their height × their width = 18

**A1:** for height = 7.2 (cm)

**(b)**

**M1:** for a correct expression for $\sigma$ or $s$, can ft their value for mean

**A1:** awrt 3.69 (allow $s$ = 3.75)

**(c)**

**M1:** for a suitable comparison of standard deviations to comment on reliability.

**A1:** for stating Hurn is south of Heathrow and a correct conclusion

**(d)**

**M1:** for a correct expression – ft their $\bar{x}+\sigma\approx10.3$

**A1:** for 7 days but accept 6 (rounding down) following a correct expression

**(e)**

**M1:** for a correct probability attempted

**A1:** for a correct prediction

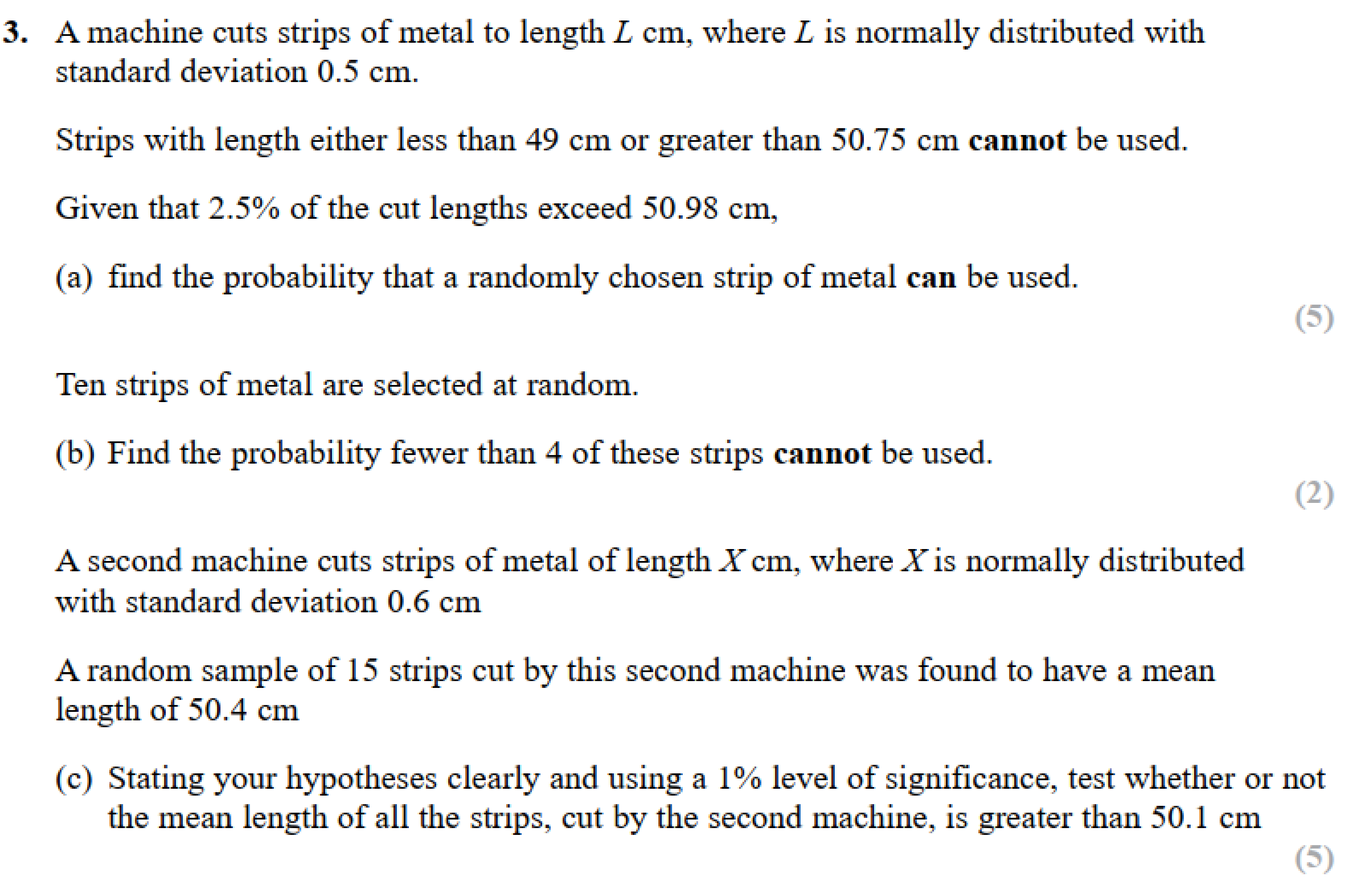
**(f)**

**B1:** for a suitable comparison and a compatible conclusion

HOME

# A2 SAMs

## The Normal Distribution



**3.** A machine cuts strips of metal to length $L$ cm, where $L$ is normally distributed with standard deviation 0.5 cm.

Strips with length either less than 49 cm or greater than 50.75 cm **cannot** be used.

Given that 2.5% of the cut lengths exceed 50.98 cm,

(a) find the probability that a randomly chosen strip of metal **can** be used.

(5)

Ten strips of metal are selected at random.

(b) Find the probability fewer than 4 of these strips **cannot** be used.

(2)

A second machine cuts strips of metal of length $X$ cm, where $X$ is normally distributed with standard deviation 0.6 cm

A random sample of 15 strips cut by this second machine was found to have a mean length of 50.4 cm

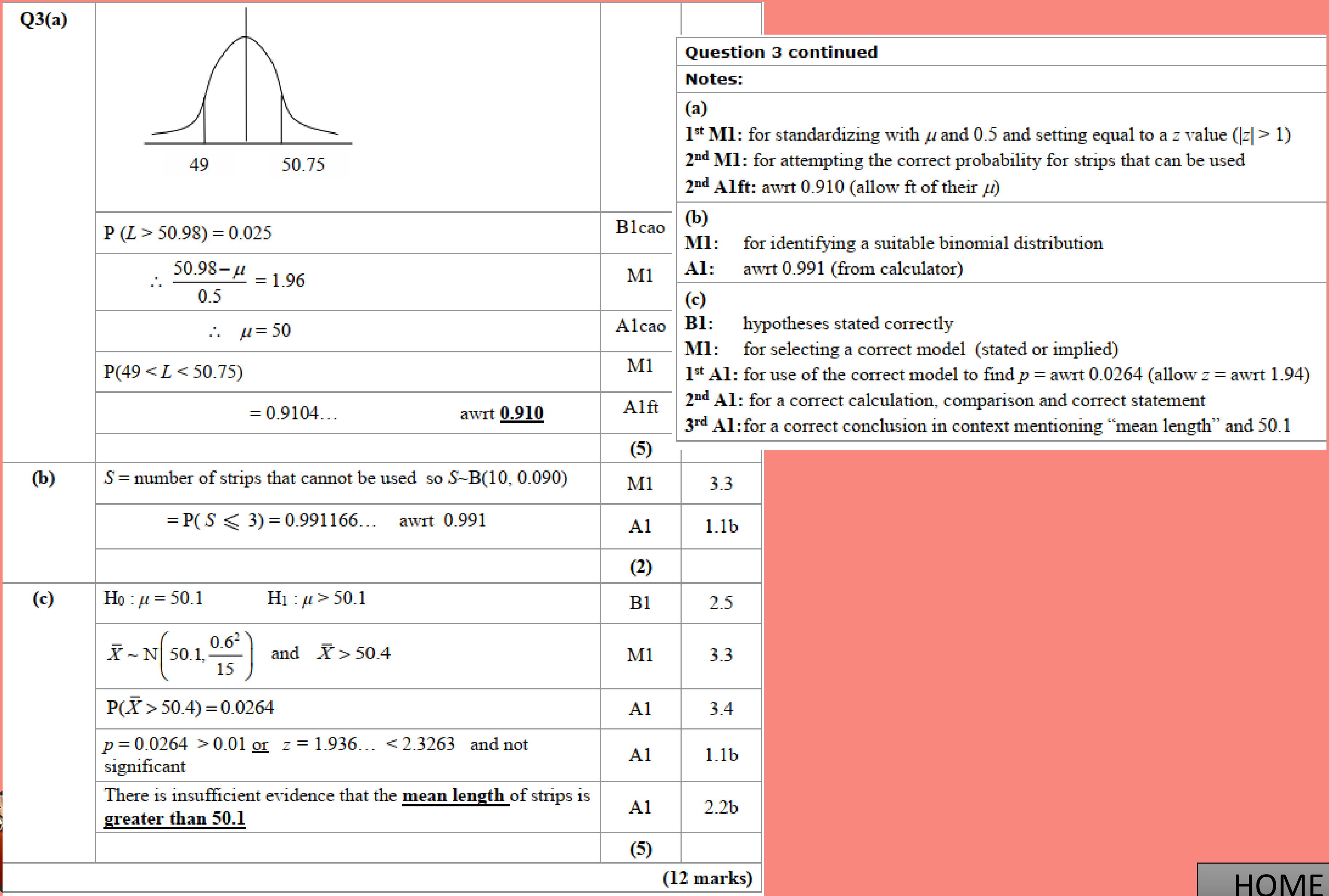
(c) Stating your hypotheses clearly and using a 1% level of significance, test whether or not the mean length of all the strips, cut by the second machine, is greater than 50.1 cm

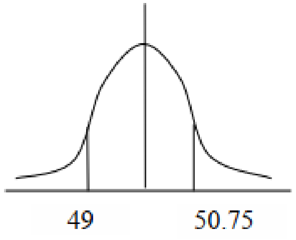
(5)

HOME

# A2 SAMs

## The Normal Distribution



| | | | |
|---|---|---|---|
| **Q3(a)** |  | | |
| | P $(L > 50.98) = 0.025$ | B1cao | |
| | $\therefore \dfrac{50.98-\mu}{0.5} = 1.96$ | M1 | |
| | $\therefore \quad \mu = 50$ | A1cao | |
| | $\text{P}(49 < L < 50.75)$ | M1 | |
| | $= 0.9104\ldots$ awrt **0.910** | A1ft | |
| | | **(5)** | |
| **(b)** | $S$ = number of strips that cannot be used so $S \sim \text{B}(10, 0.090)$ | M1 | 3.3 |
| | $= \text{P}(S \leqslant 3) = 0.991166\ldots$ awrt 0.991 | A1 | 1.1b |
| | | **(2)** | |
| **(c)** | $\text{H}_0 : \mu = 50.1$ $\quad \text{H}_1 : \mu > 50.1$ | B1 | 2.5 |
| | $\bar{X} \sim \text{N}\left(50.1, \dfrac{0.6^2}{15}\right)$ and $\bar{X} > 50.4$ | M1 | 3.3 |
| | $\text{P}(\bar{X} > 50.4) = 0.0264$ | A1 | 3.4 |
| | $p = 0.0264 > 0.01$ or $z = 1.936\ldots < 2.3263$ and not significant | A1 | 1.1b |
| | There is insufficient evidence that the **mean length** of strips is **greater than 50.1** | A1 | 2.2b |
| | | **(5)** | |
| | | | **(12 marks)** |

**Question 3 continued**

**Notes:**

**(a)**

**1st M1:** for standardizing with $\mu$ and 0.5 and setting equal to a $z$ value ($|z| > 1$)

**2nd M1:** for attempting the correct probability for strips that can be used

**2nd A1ft:** awrt 0.910 (allow ft of their $\mu$)

**(b)**

**M1:** for identifying a suitable binomial distribution

**A1:** awrt 0.991 (from calculator)

**(c)**

**B1:** hypotheses stated correctly

**M1:** for selecting a correct model (stated or implied)

**1st A1:** for use of the correct model to find $p$ = awrt 0.0264 (allow $z$ = awrt 1.94)

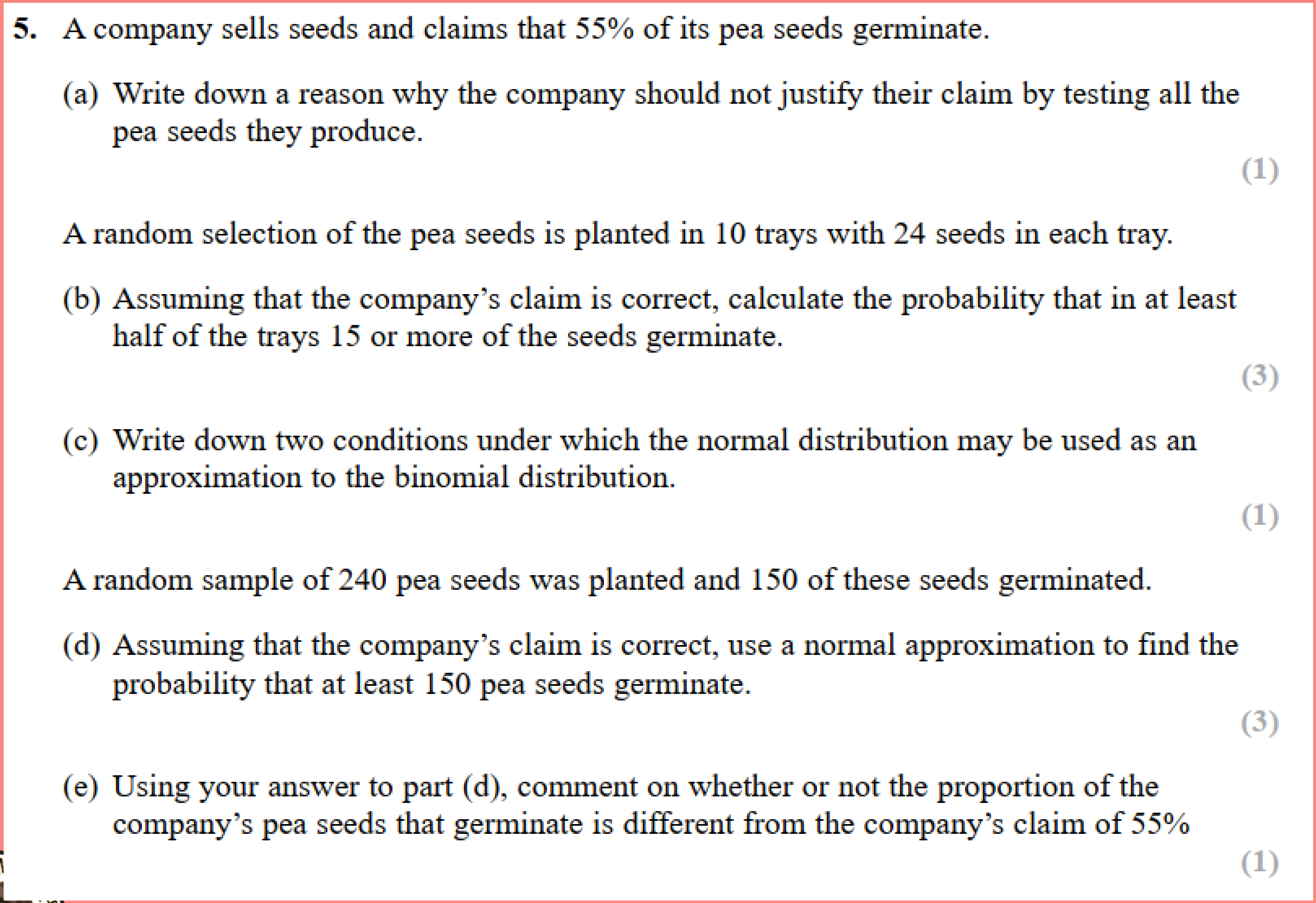
**2nd A1:** for a correct calculation, comparison and correct statement

**3rd A1:** for a correct conclusion in context mentioning "mean length" and 50.1

HOME

A2 SAMs

## The Normal Distribution



**5.** A company sells seeds and claims that 55% of its pea seeds germinate.

(a) Write down a reason why the company should not justify their claim by testing all the pea seeds they produce.

**(1)**

A random selection of the pea seeds is planted in 10 trays with 24 seeds in each tray.

(b) Assuming that the company's claim is correct, calculate the probability that in at least half of the trays 15 or more of the seeds germinate.

**(3)**

(c) Write down two conditions under which the normal distribution may be used as an approximation to the binomial distribution.

**(1)**

A random sample of 240 pea seeds was planted and 150 of these seeds germinated.

(d) Assuming that the company's claim is correct, use a normal approximation to find the probability that at least 150 pea seeds germinate.

**(3)**

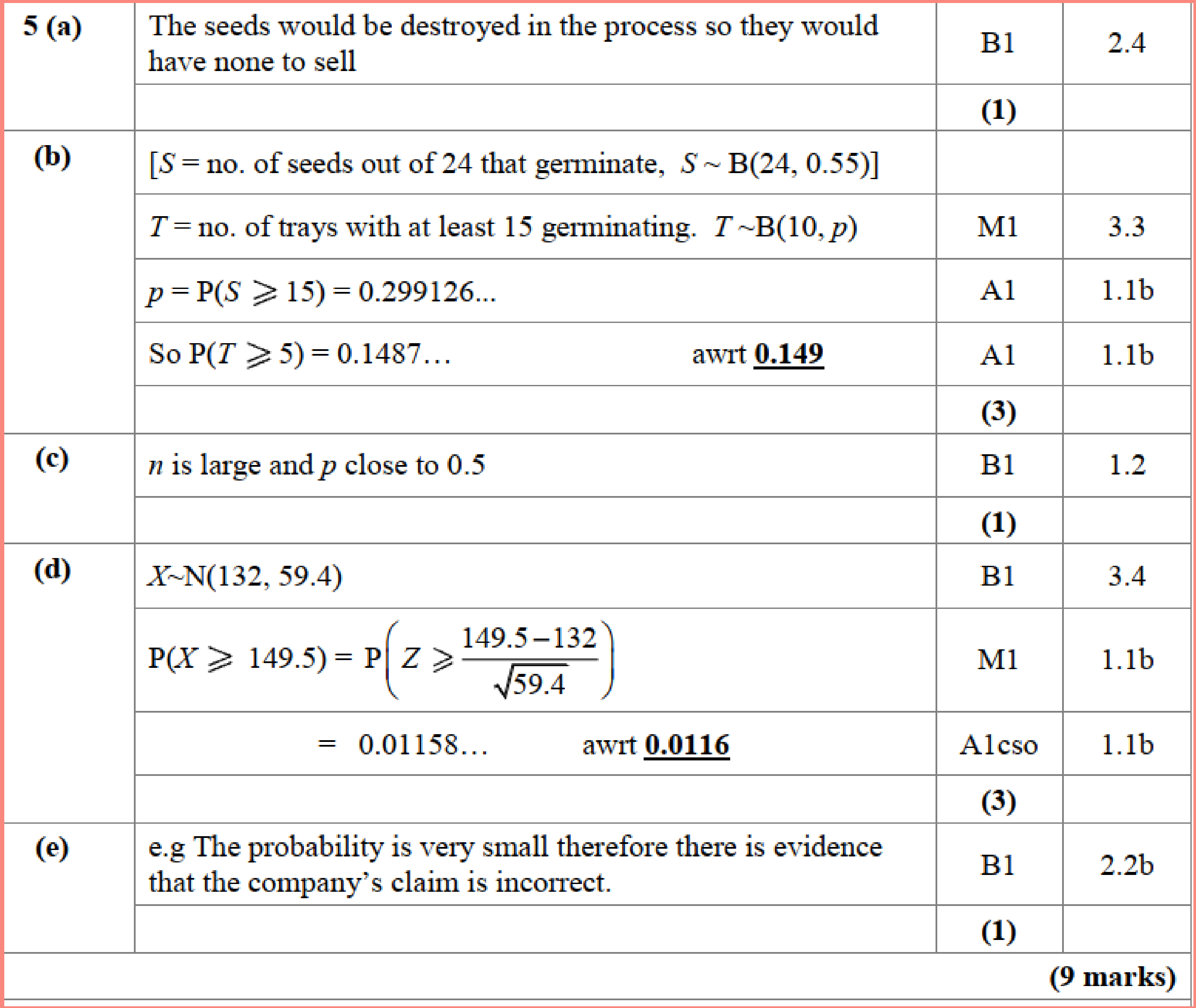
(e) Using your answer to part (d), comment on whether or not the proportion of the company's pea seeds that germinate is different from the company's claim of 55%

**(1)**

HOME

# A2 SAMs

## The Normal Distribution

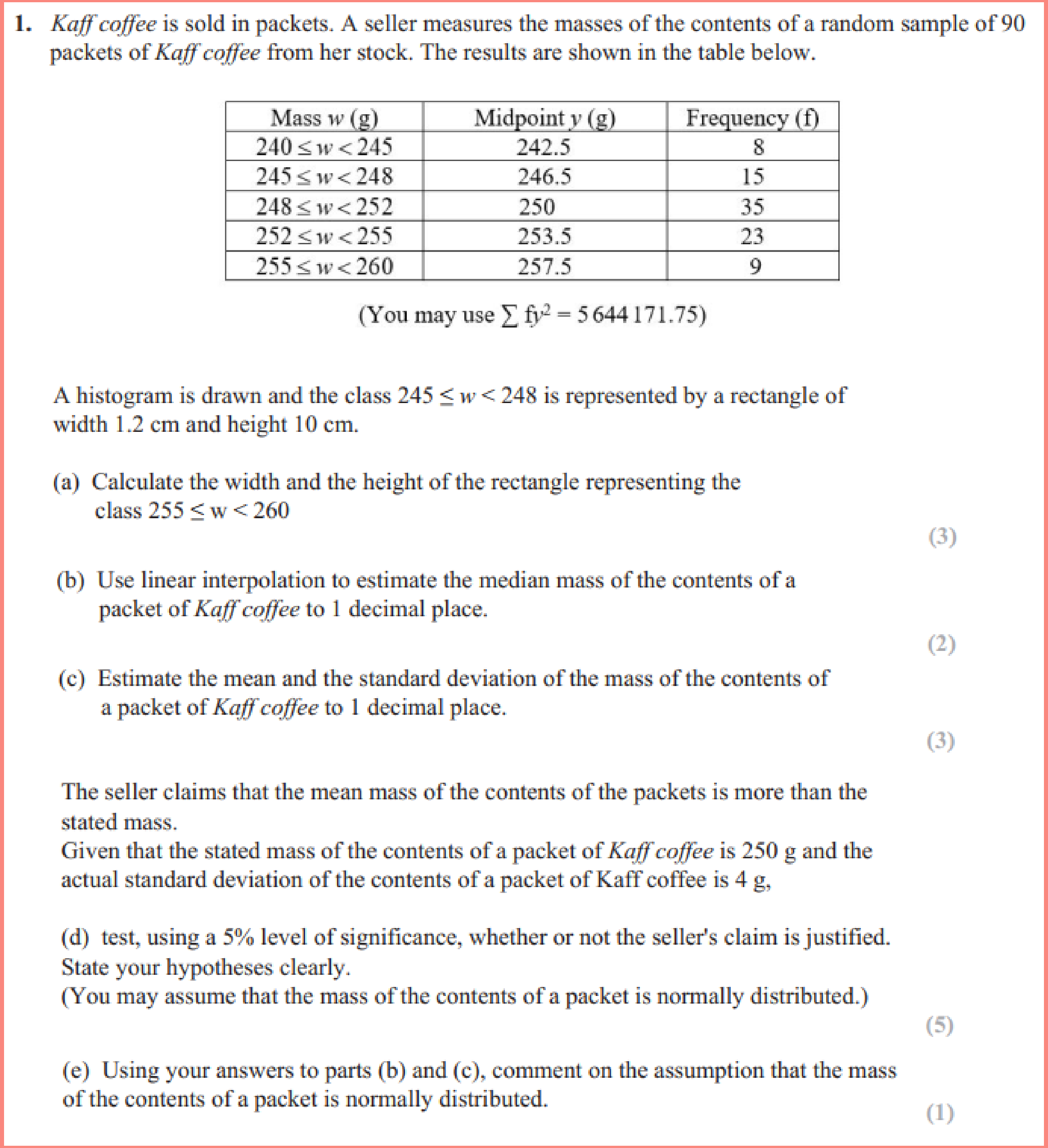


| | | | |
|---|---|---|---|
| **5 (a)** | The seeds would be destroyed in the process so they would have none to sell | B1 | 2.4 |
| | | **(1)** | |
| **(b)** | [$S$ = no. of seeds out of 24 that germinate, $S \sim \mathrm{B}(24, 0.55)$] | | |
| | $T$ = no. of trays with at least 15 germinating. $T \sim \mathrm{B}(10, p)$ | M1 | 3.3 |
| | $p = \mathrm{P}(S \geqslant 15) = 0.299126...$ | A1 | 1.1b |
| | So $\mathrm{P}(T \geqslant 5) = 0.1487\ldots$ awrt **0.149** | A1 | 1.1b |
| | | **(3)** | |
| **(c)** | $n$ is large and $p$ close to 0.5 | B1 | 1.2 |
| | | **(1)** | |
| **(d)** | $X \sim \mathrm{N}(132, 59.4)$ | B1 | 3.4 |
| | $\mathrm{P}(X \geqslant 149.5) = \mathrm{P}\left(Z \geqslant \frac{149.5-132}{\sqrt{59.4}}\right)$ | M1 | 1.1b |
| | $= 0.01158\ldots$ awrt **0.0116** | A1cso | 1.1b |
| | | **(3)** | |
| **(e)** | e.g The probability is very small therefore there is evidence that the company's claim is incorrect. | B1 | 2.2b |
| | | **(1)** | |
| | | | **(9 marks)** |

HOME

# A2 SAMs

## The Normal Distribution



**1.** *Kaff coffee* is sold in packets. A seller measures the masses of the contents of a random sample of 90 packets of *Kaff coffee* from her stock. The results are shown in the table below.

| Mass $w$ (g) | Midpoint $y$ (g) | Frequency (f) |
|---|---|---|
| $240 \le w < 245$ | 242.5 | 8 |
| $245 \le w < 248$ | 246.5 | 15 |
| $248 \le w < 252$ | 250 | 35 |
| $252 \le w < 255$ | 253.5 | 23 |
| $255 \le w < 260$ | 257.5 | 9 |

(You may use $\sum fy^2 = 5\,644\,171.75$)

A histogram is drawn and the class $245 \le w < 248$ is represented by a rectangle of width 1.2 cm and height 10 cm.

(a) Calculate the width and the height of the rectangle representing the class $255 \le \text{w} < 260$

(3)

(b) Use linear interpolation to estimate the median mass of the contents of a packet of *Kaff coffee* to 1 decimal place.

(2)

(c) Estimate the mean and the standard deviation of the mass of the contents of a packet of *Kaff coffee* to 1 decimal place.

(3)

The seller claims that the mean mass of the contents of the packets is more than the stated mass.
Given that the stated mass of the contents of a packet of *Kaff coffee* is 250 g and the actual standard deviation of the contents of a packet of Kaff coffee is 4 g,

(d) test, using a 5% level of significance, whether or not the seller's claim is justified.
State your hypotheses clearly.
(You may assume that the mass of the contents of a packet is normally distributed.)

(5)

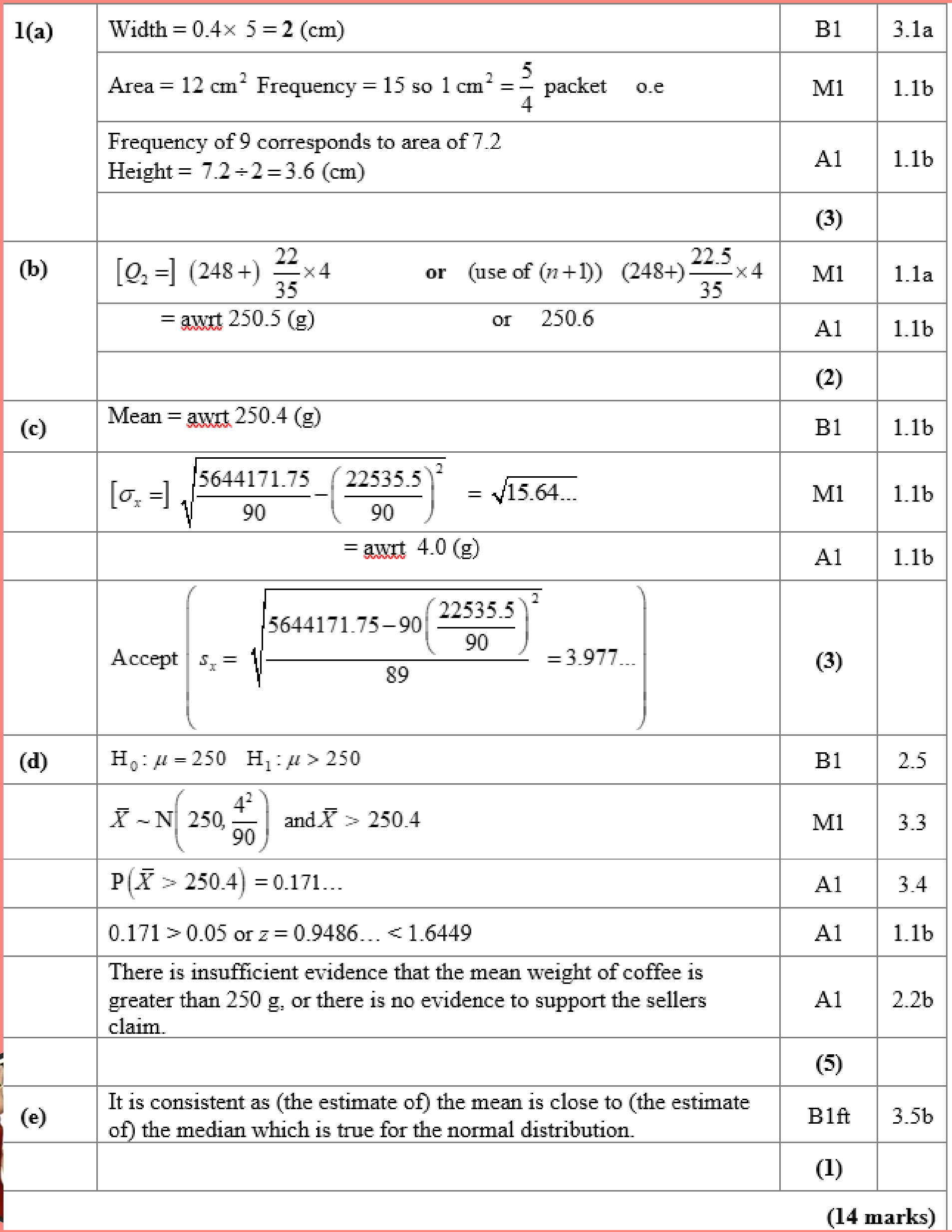
(e) Using your answers to parts (b) and (c), comment on the assumption that the mass of the contents of a packet is normally distributed.

(1)

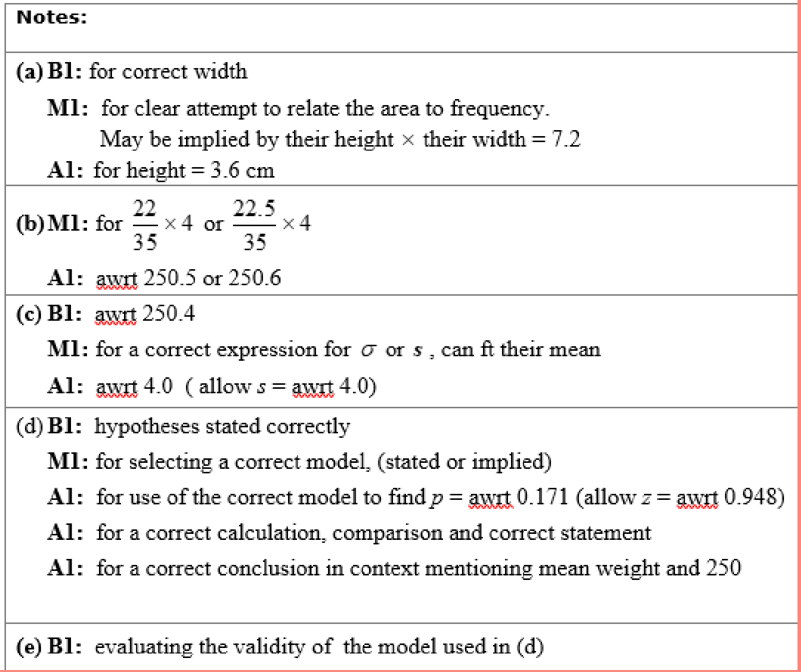
HOME

# A2 SAMs

## The Normal Distribution



| | | | |
|---|---|---|---|
| **1(a)** | Width = 0.4× 5 = **2** (cm) | B1 | 3.1a |
| | Area = 12 cm$^2$ Frequency = 15 so 1 cm$^2$ = $\frac{5}{4}$ packet o.e | M1 | 1.1b |
| | Frequency of 9 corresponds to area of 7.2<br>Height = $7.2 \div 2 = 3.6$ (cm) | A1 | 1.1b |
| | | **(3)** | |
| **(b)** | $[Q_2 =]\ (248+)\ \frac{22}{35}\times 4$ **or** (use of $(n+1)$) $(248+)\frac{22.5}{35}\times 4$ | M1 | 1.1a |
| | = awrt 250.5 (g) or 250.6 | A1 | 1.1b |
| | | **(2)** | |
| **(c)** | Mean = awrt 250.4 (g) | B1 | 1.1b |
| | $[\sigma_x =]\ \sqrt{\frac{5644171.75}{90} - \left(\frac{22535.5}{90}\right)^2} = \sqrt{15.64...}$ | M1 | 1.1b |
| | = awrt 4.0 (g) | A1 | 1.1b |
| | Accept $\left( s_x = \sqrt{\frac{5644171.75 - 90\left(\frac{22535.5}{90}\right)^2}{89}} = 3.977... \right)$ | **(3)** | |
| **(d)** | $\text{H}_0: \mu = 250 \quad \text{H}_1: \mu > 250$ | B1 | 2.5 |
| | $\bar{X} \sim \text{N}\left(250, \frac{4^2}{90}\right)$ and $\bar{X} > 250.4$ | M1 | 3.3 |
| | $\text{P}\left(\bar{X} > 250.4\right) = 0.171...$ | A1 | 3.4 |
| | $0.171 > 0.05$ or $z = 0.9486... < 1.6449$ | A1 | 1.1b |
| | There is insufficient evidence that the mean weight of coffee is greater than 250 g, or there is no evidence to support the sellers claim. | A1 | 2.2b |
| | | **(5)** | |
| **(e)** | It is consistent as (the estimate of) the mean is close to (the estimate of) the median which is true for the normal distribution. | B1ft | 3.5b |
| | | **(1)** | |
| | | **(14 marks)** | |



**Notes:**

**(a) B1:** for correct width

**M1:** for clear attempt to relate the area to frequency.
May be implied by their height × their width = 7.2

**A1:** for height = 3.6 cm

**(b) M1:** for $\frac{22}{35}\times 4$ or $\frac{22.5}{35}\times 4$

**A1:** awrt 250.5 or 250.6

**(c) B1:** awrt 250.4

**M1:** for a correct expression for $\sigma$ or $s$, can ft their mean

**A1:** awrt 4.0 ( allow $s$ = awrt 4.0)

(d) **B1:** hypotheses stated correctly

**M1:** for selecting a correct model, (stated or implied)

**A1:** for use of the correct model to find $p$ = awrt 0.171 (allow $z$ = awrt 0.948)

**A1:** for a correct calculation, comparison and correct statement

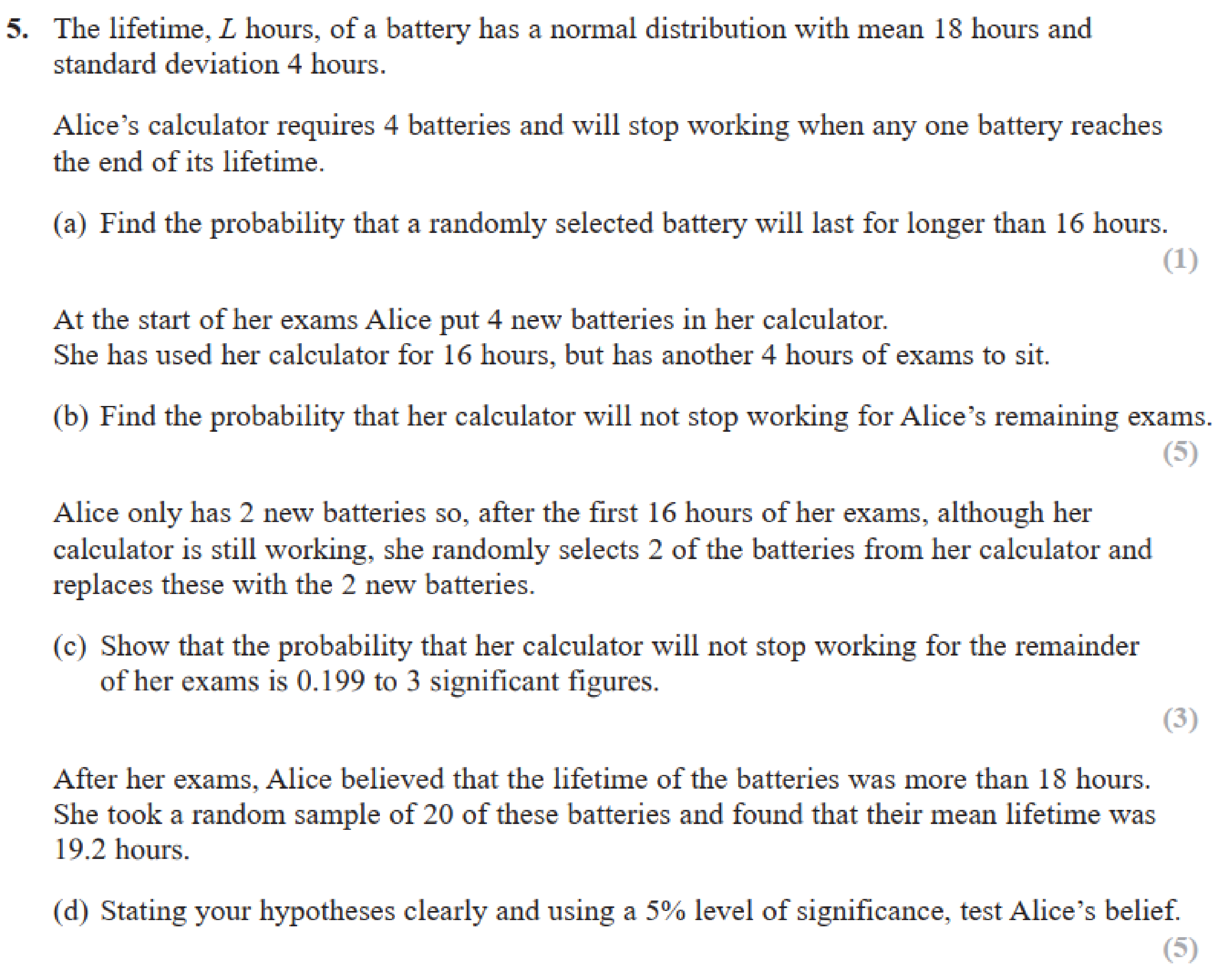
**A1:** for a correct conclusion in context mentioning mean weight and 250

**(e) B1:** evaluating the validity of the model used in (d)

HOME

# A2 2018

## The Normal Distribution



**5.** The lifetime, $L$ hours, of a battery has a normal distribution with mean 18 hours and standard deviation 4 hours.

Alice's calculator requires 4 batteries and will stop working when any one battery reaches the end of its lifetime.

(a) Find the probability that a randomly selected battery will last for longer than 16 hours.

**(1)**

At the start of her exams Alice put 4 new batteries in her calculator.
She has used her calculator for 16 hours, but has another 4 hours of exams to sit.

(b) Find the probability that her calculator will not stop working for Alice's remaining exams.

**(5)**

Alice only has 2 new batteries so, after the first 16 hours of her exams, although her calculator is still working, she randomly selects 2 of the batteries from her calculator and replaces these with the 2 new batteries.

(c) Show that the probability that her calculator will not stop working for the remainder of her exams is 0.199 to 3 significant figures.

**(3)**

After her exams, Alice believed that the lifetime of the batteries was more than 18 hours. She took a random sample of 20 of these batteries and found that their mean lifetime was 19.2 hours.

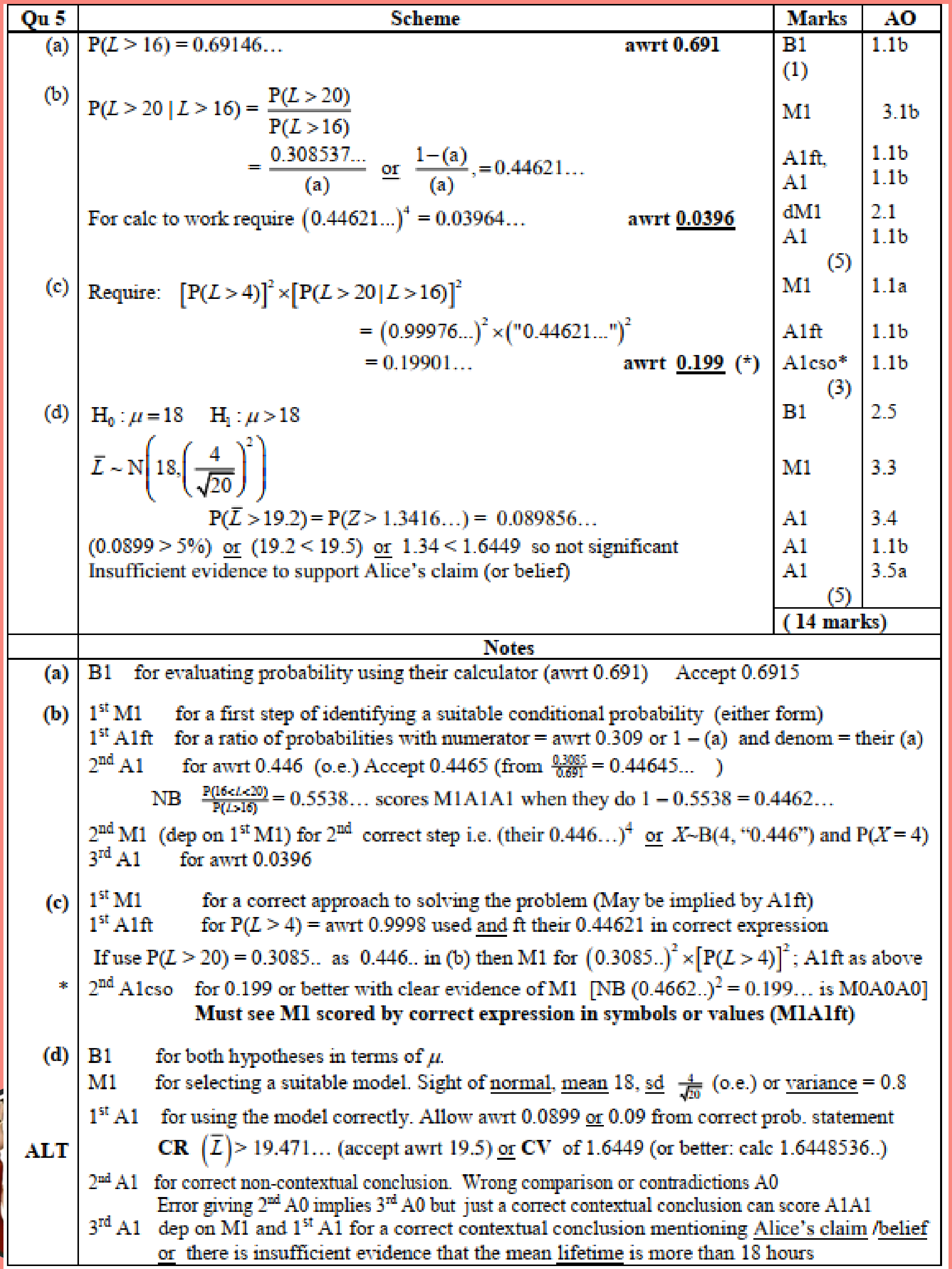
(d) Stating your hypotheses clearly and using a 5% level of significance, test Alice's belief.

**(5)**

HOME

# A2 2018

## The Normal Distribution



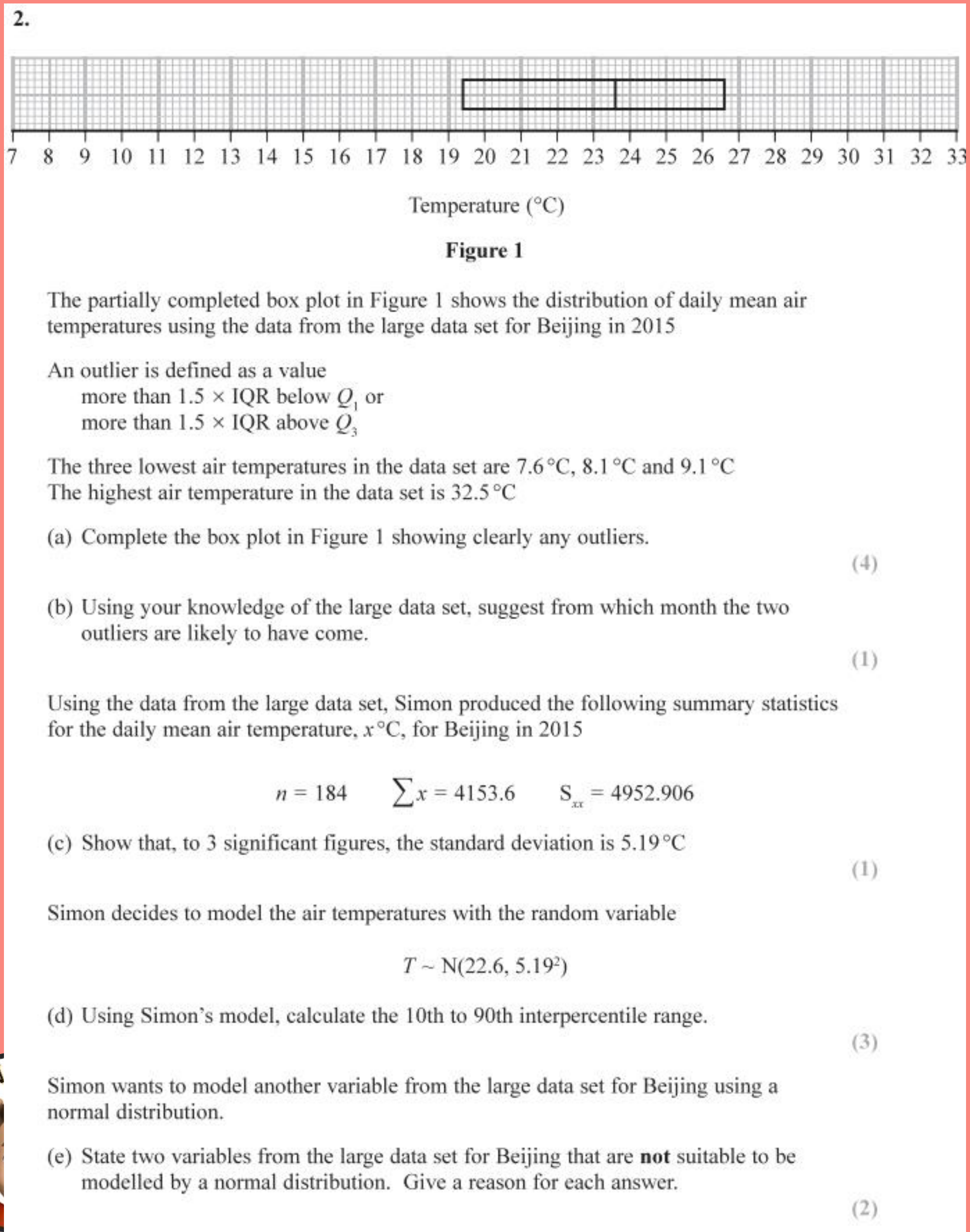
| Qu 5 | Scheme | Marks | AO |
|---|---|---|---|
| (a) | $P(L > 16) = 0.69146\ldots$ **awrt 0.691** | B1<br>(1) | 1.1b |
| (b) | $P(L > 20 \mid L > 16) = \dfrac{P(L > 20)}{P(L > 16)}$ | M1 | 3.1b |
| | $= \dfrac{0.308537\ldots}{(a)}$ <u>or</u> $\dfrac{1-(a)}{(a)}$, $= 0.44621\ldots$ | A1ft,<br>A1 | 1.1b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

HOME

# A2 2019

## The Normal Distribution


7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33

**2.**

Temperature (°C)

**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 × IQR below $Q_1$ or
more than 1.5 × 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, $x$ °C, for Beijing in 2015

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

(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

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

(d) Using Simon's model, calculate the 10th to 90th interpercentile range.
**(3)**

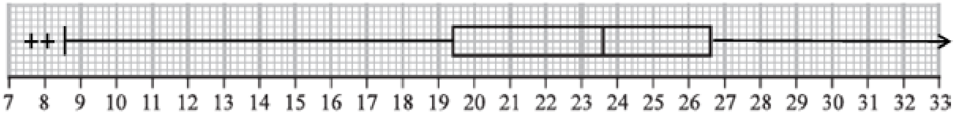
Simon wants to model another variable from the large data set for Beijing using a normal distribution.

(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)**

HOME

# A2 2019

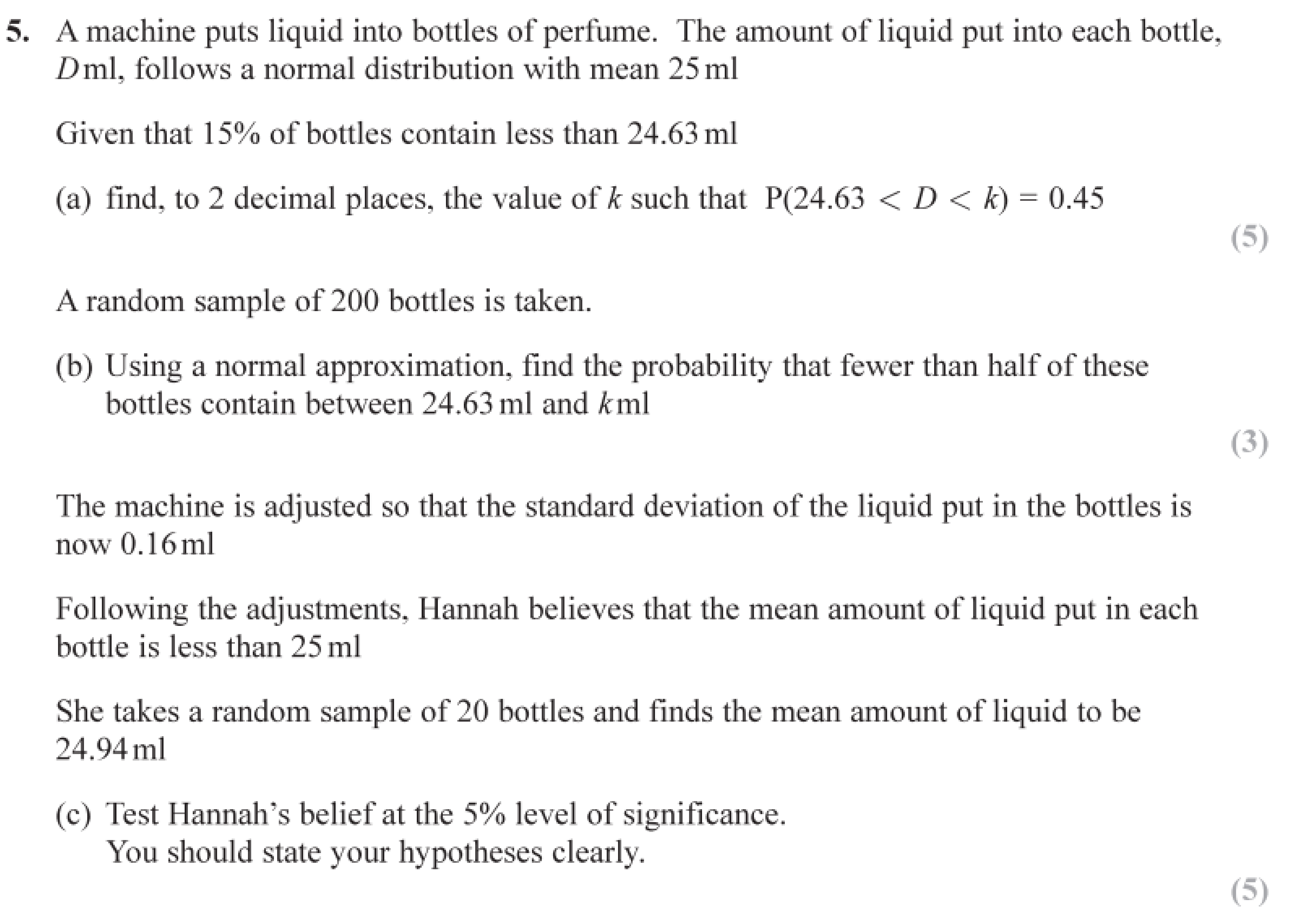
# The Normal Distribution

| | | | |
|---|---|---|---|
| (a) | IQR = 2.6. – 19.4 = 7.2 | B1 | This mark is given for finding the interquartile range |
| | 19.4 – (1.5 × 7.2) = 8.6<br>19.4 + (1.5 × 7.2) = 37.4 | M1 | This mark is given for a method find the values for the whiskers of the boxplot |
| |  | | |
| | | A1 | This mark is given for plotting the correct whisker (8.6) on the boxplot |
| | | A1 | This mark is given for plotting the two correct outliers 7.6 °C and 8.1 °C |
| (b) | October (since it is the month with the coldest temperatures between May and October in Beijing) | B1 | This mark is given for a correct suggestion with a supporting reason. |
| (c) | $\sigma = \sqrt{\frac{S_{xx}}{n}} = \sqrt{\frac{4952.906}{184}} = \sqrt{26.92} = 5.19$ | B1 | This mark is given for showing the calculation for the standard deviation to three significant figures |
| (d) | z = (±) 1.2816 | B1 | This mark is given for identifying the z-value for the 10th and 90th percentiles (from tables or calculator) |
| | $2 \times z \times 5.19$ | M1 | This mark is given for a method to find the interpercentile range between the 10th and 90th value |
| | = 13.303 | A1 | This mark is given for finding a correct interpercentile range between the 10th and 90th value |
| (e) | Daily wind speed (Beaufort) since it is qualitative data | B1 | This mark si given for stating a correct variable with a supporting reason |
| | Rainfall (since it is not symmetric) | B1 | This mark si given for stating a correct variable with a supporting reason |

HOME

**A2 2019**

## The Normal Distribution



**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 $\mathrm{P}(24.63 < D < k) = 0.45$

**(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

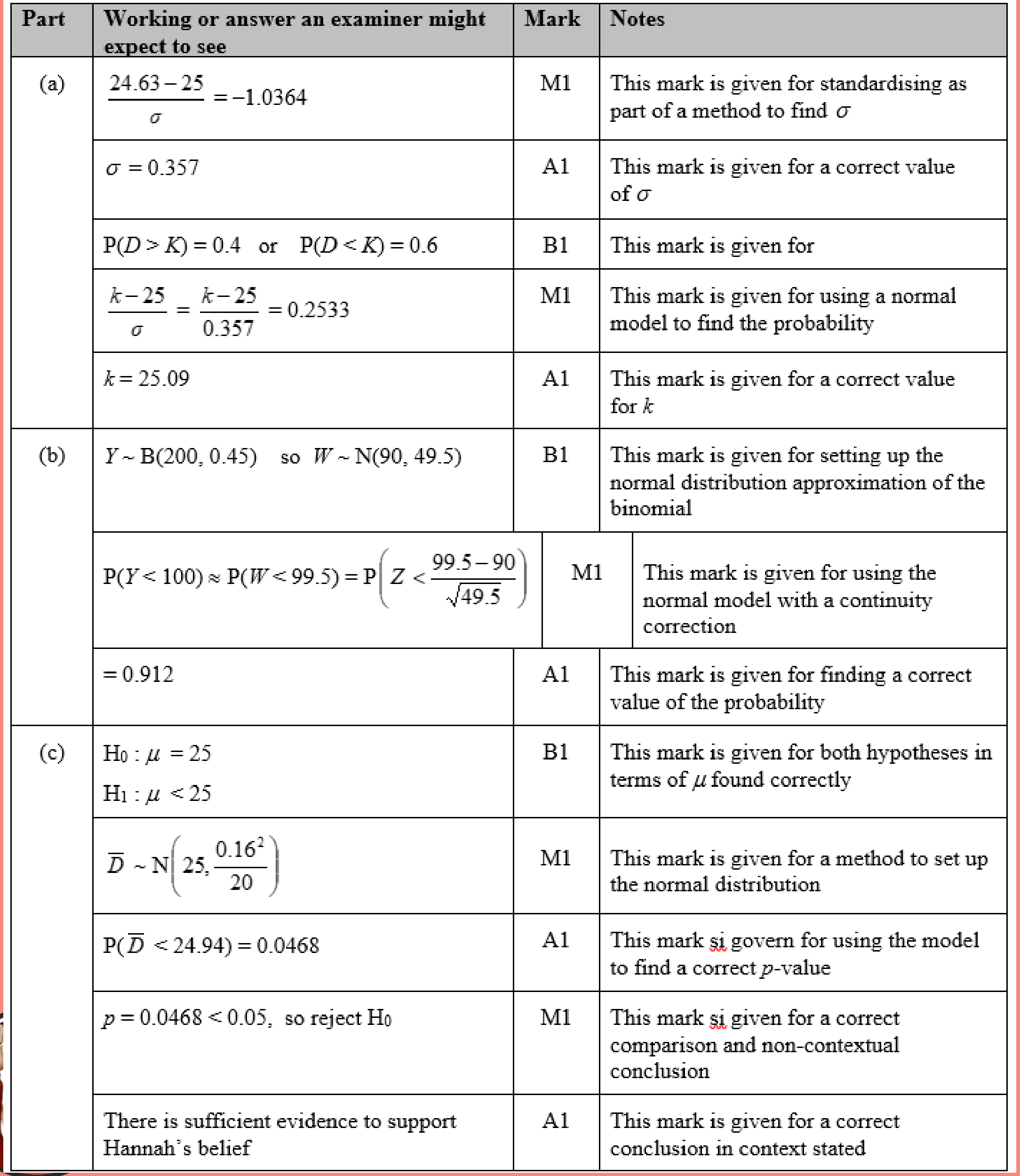
(c) Test Hannah's belief at the 5% level of significance.
You should state your hypotheses clearly.

**(5)**

HOME

# A2 2019

## The Normal Distribution

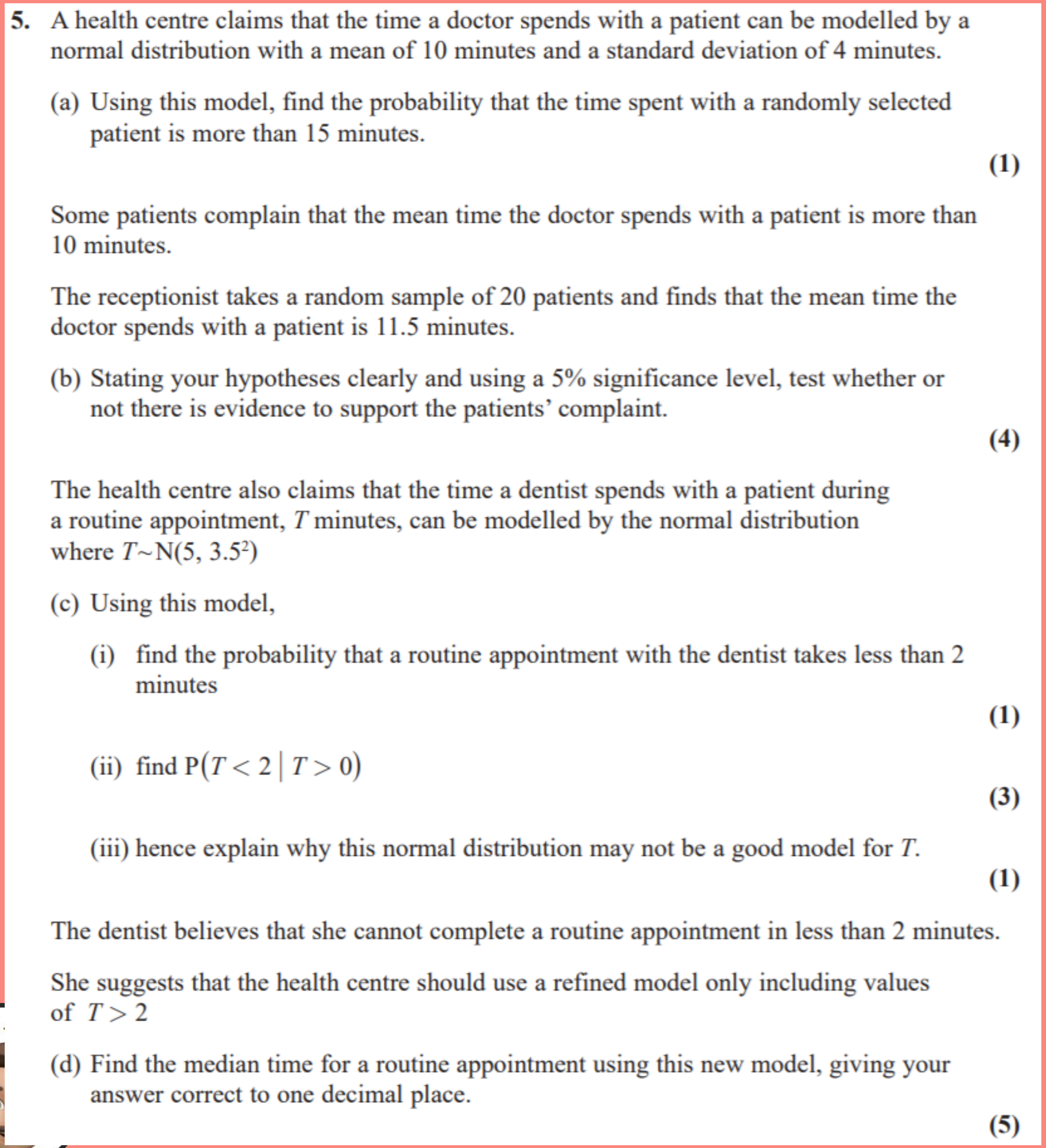


| Part | Working or answer an examiner might expect to see | Mark | Notes |
|---|---|---|---|
| (a) | $\frac{24.63-25}{\sigma} = -1.0364$ | M1 | This mark is given for standardising as part of a method to find $\sigma$ |
| | $\sigma = 0.357$ | A1 | This mark is given for a correct value of $\sigma$ |
| | $P(D > K) = 0.4$ or $P(D < K) = 0.6$ | B1 | This mark is given for |
| | $\frac{k-25}{\sigma} = \frac{k-25}{0.357} = 0.2533$ | M1 | This mark is given for using a normal model to find the probability |
| | $k = 25.09$ | A1 | This mark is given for a correct value for $k$ |
| (b) | $Y \sim B(200, 0.45)$ so $W \sim N(90, 49.5)$ | B1 | This mark is given for setting up the normal distribution approximation of the binomial |
| | $P(Y < 100) \approx P(W < 99.5) = P\left(Z < \frac{99.5-90}{\sqrt{49.5}}\right)$ | M1 | This mark is given for using the normal model with a continuity correction |
| | $= 0.912$ | A1 | This mark is given for finding a correct value of the probability |
| (c) | $H_0 : \mu = 25$<br>$H_1 : \mu < 25$ | B1 | This mark is given for both hypotheses in terms of $\mu$ found correctly |
| | $\bar{D} \sim N\left(25, \frac{0.16^2}{20}\right)$ | M1 | This mark is given for a method to set up the normal distribution |
| | $P(\bar{D} < 24.94) = 0.0468$ | A1 | This mark si govern for using the model to find a correct $p$-value |
| | $p = 0.0468 < 0.05$, so reject $H_0$ | M1 | This mark si given for a correct comparison and non-contextual conclusion |
| | There is sufficient evidence to support Hannah's belief | A1 | This mark is given for a correct conclusion in context stated |

HOME

# A2 2020

## The Normal Distribution



**5.** A health centre claims that the time a doctor spends with a patient can be modelled by a normal distribution with a mean of 10 minutes and a standard deviation of 4 minutes.

(a) Using this model, find the probability that the time spent with a randomly selected patient is more than 15 minutes.

**(1)**

Some patients complain that the mean time the doctor spends with a patient is more than 10 minutes.

The receptionist takes a random sample of 20 patients and finds that the mean time the doctor spends with a patient is 11.5 minutes.

(b) Stating your hypotheses clearly and using a 5% significance level, test whether or not there is evidence to support the patients' complaint.

**(4)**

The health centre also claims that the time a dentist spends with a patient during a routine appointment, $T$ minutes, can be modelled by the normal distribution where $T \sim \mathrm{N}(5, 3.5^2)$

(c) Using this model,

(i) find the probability that a routine appointment with the dentist takes less than 2 minutes

**(1)**

(ii) find $\mathrm{P}(T < 2 \mid T > 0)$

**(3)**

(iii) hence explain why this normal distribution may not be a good model for $T$.

**(1)**

The dentist believes that she cannot complete a routine appointment in less than 2 minutes.

She suggests that the health centre should use a refined model only including values of $T > 2$

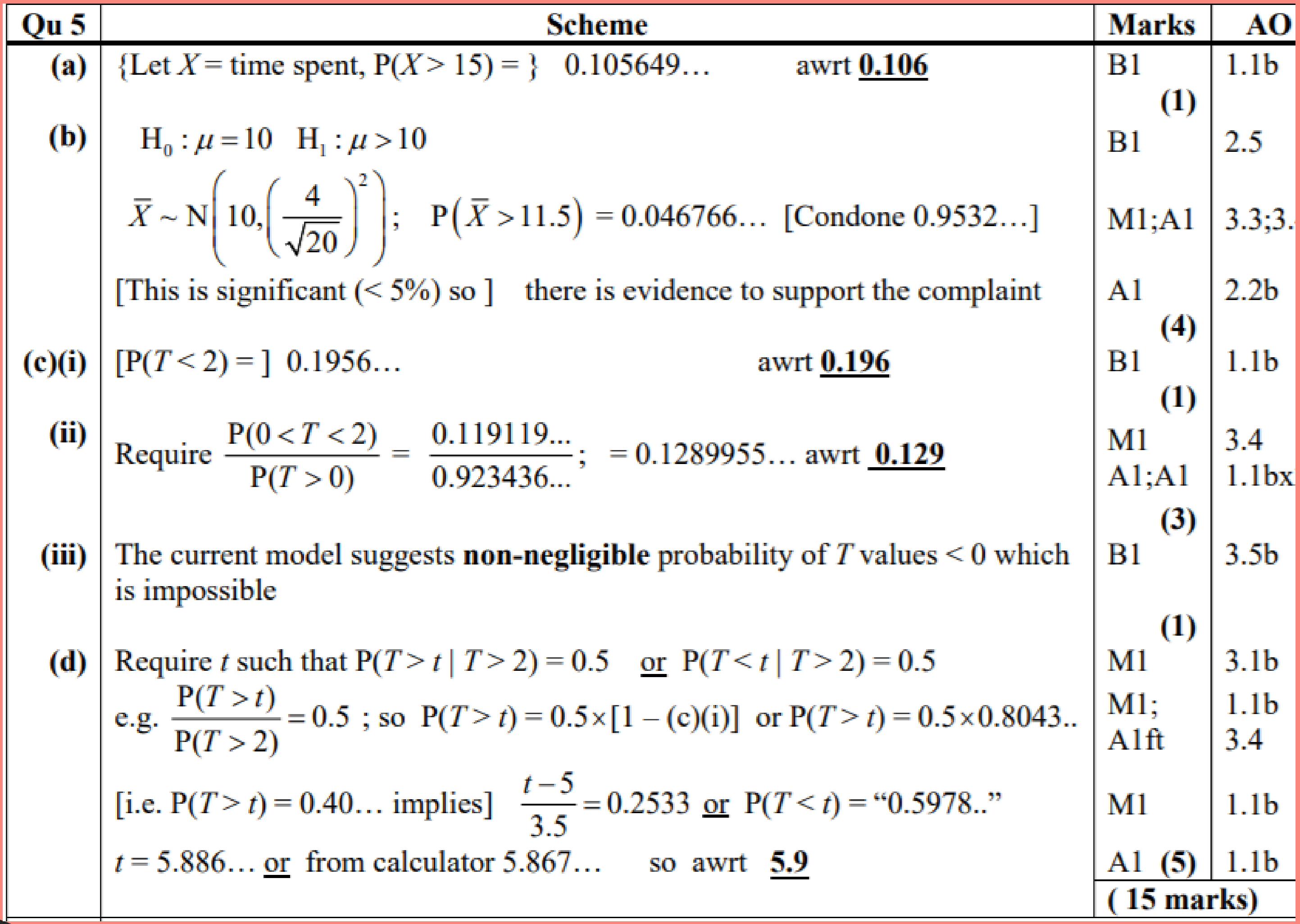
(d) Find the median time for a routine appointment using this new model, giving your answer correct to one decimal place.

**(5)**

HOME

# A2 2020

## The Normal Distribution

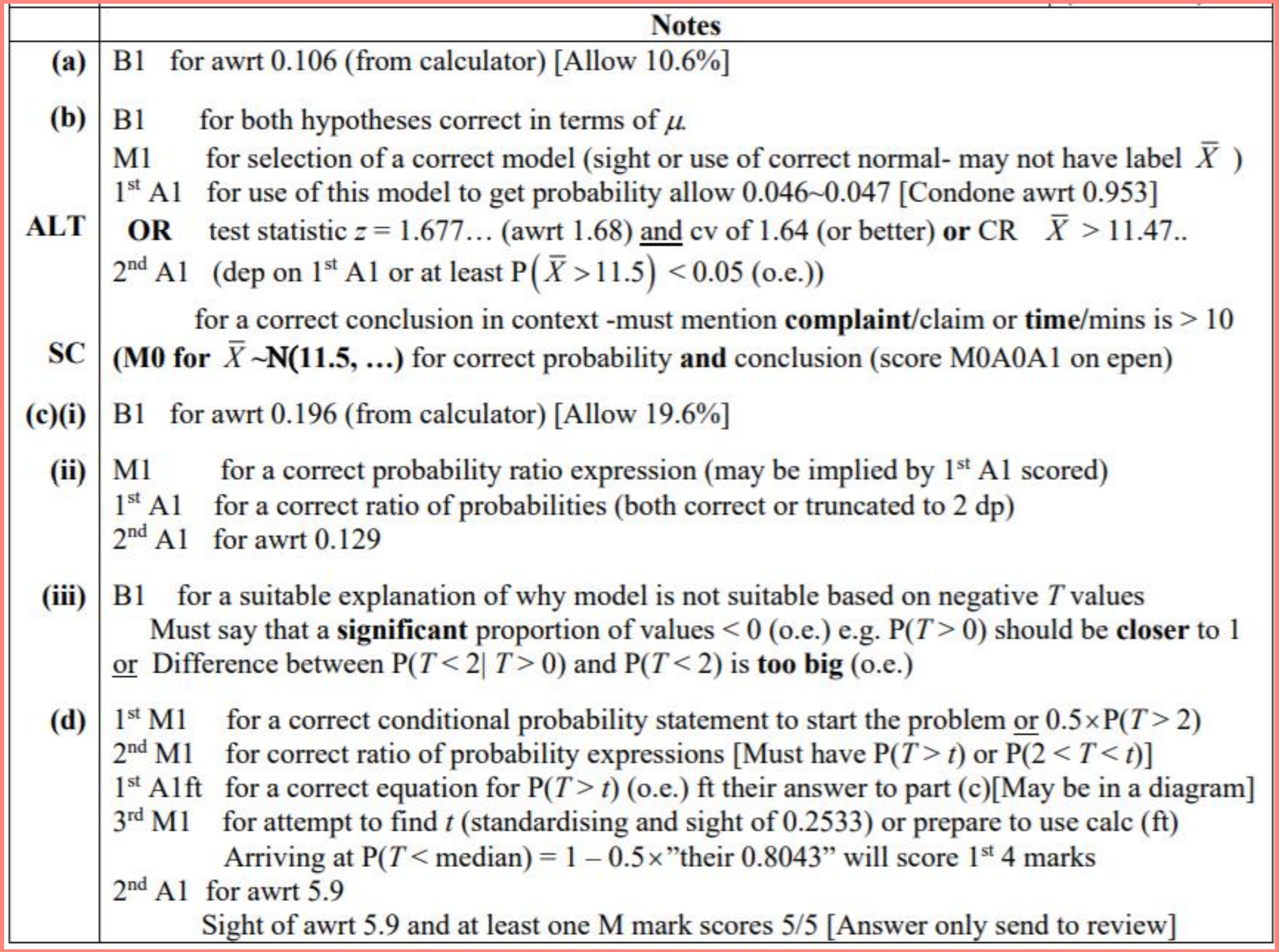


| Qu 5 | Scheme | Marks | AO |
|---|---|---|---|
| (a) | {Let $X$ = time spent, $P(X > 15)$ = } 0.105649… awrt **0.106** | B1 | 1.1b |
| | | (1) | |
| (b) | $H_0 : \mu = 10 \quad H_1 : \mu > 10$ | B1 | 2.5 |
| | $\bar{X} \sim N\left(10, \left(\frac{4}{\sqrt{20}}\right)^2\right)$; $P\left(\bar{X} > 11.5\right) = 0.046766\ldots$ [Condone 0.9532…] | M1;A1 | 3.3;3. |
| | [This is significant (< 5%) so ] there is evidence to support the complaint | A1 | 2.2b |
| | | (4) | |
| (c)(i) | [$P(T < 2) =$ ] 0.1956… awrt **0.196** | B1 | 1.1b |
| | | (1) | |
| (ii) | Require $\frac{P(0 < T < 2)}{P(T > 0)} = \frac{0.119119...}{0.923436...}$; $= 0.1289955\ldots$ awrt **0.129** | M1<br>A1;A1 | 3.4<br>1.1bx |
| | | (3) | |
| (iii) | The current model suggests **non-negligible** probability of $T$ values < 0 which is impossible | B1 | 3.5b |
| | | (1) | |
| (d) | Require $t$ such that $P(T > t \mid T > 2) = 0.5$ or $P(T < t \mid T > 2) = 0.5$ | M1 | 3.1b |
| | e.g. $\frac{P(T > t)}{P(T > 2)} = 0.5$ ; so $P(T > t) = 0.5 \times [1 - \text{(c)(i)}]$ or $P(T > t) = 0.5 \times 0.8043..$ | M1;<br>A1ft | 1.1b<br>3.4 |
| | [i.e. $P(T > t) = 0.40\ldots$ implies] $\frac{t - 5}{3.5} = 0.2533$ or $P(T < t) =$ "0.5978.." | M1 | 1.1b |
| | $t = 5.886\ldots$ or from calculator 5.867… so awrt **5.9** | A1 (5) | 1.1b |
| | | ( 15 marks) | |

HOME

# A2 2020

# The Normal Distribution

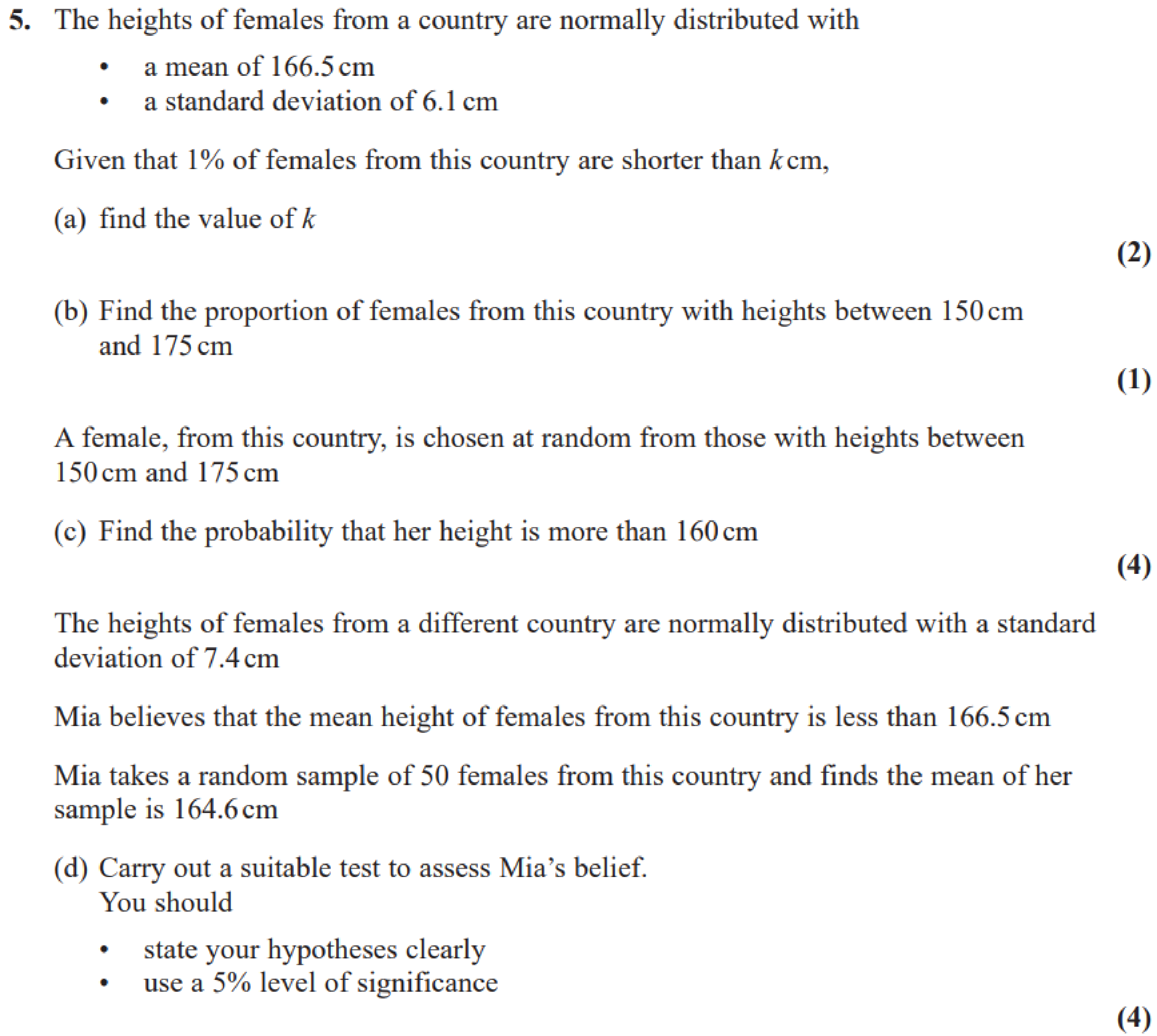


| | Notes |
|---|---|
| **(a)** | B1 for awrt 0.106 (from calculator) [Allow 10.6%] |
| **(b)** | B1 for both hypotheses correct in terms of $\mu$.<br>M1 for selection of a correct model (sight or use of correct normal- may not have label $\bar{X}$ )<br>1st A1 for use of this model to get probability allow 0.046~0.047 [Condone awrt 0.953] |
| **ALT** | **OR** test statistic $z = 1.677\ldots$ (awrt 1.68) and cv of 1.64 (or better) **or** CR $\bar{X} > 11.47..$<br>2nd A1 (dep on 1st A1 or at least $\mathrm{P}\left(\bar{X} > 11.5\right) < 0.05$ (o.e.))<br>for a correct conclusion in context -must mention **complaint**/claim or **time**/mins is > 10 |
| **SC** | **(M0 for $\bar{X}$~N(11.5, …)** for correct probability **and** conclusion (score M0A0A1 on epen) |
| **(c)(i)** | B1 for awrt 0.196 (from calculator) [Allow 19.6%] |
| **(ii)** | M1 for a correct probability ratio expression (may be implied by 1st A1 scored)<br>1st A1 for a correct ratio of probabilities (both correct or truncated to 2 dp)<br>2nd A1 for awrt 0.129 |
| **(iii)** | B1 for a suitable explanation of why model is not suitable based on negative $T$ values<br>Must say that a **significant** proportion of values < 0 (o.e.) e.g. $\mathrm{P}(T > 0)$ should be **closer** to 1<br>or Difference between $\mathrm{P}(T < 2 \mid T > 0)$ and $\mathrm{P}(T < 2)$ is **too big** (o.e.) |
| **(d)** | 1st M1 for a correct conditional probability statement to start the problem or $0.5\times\mathrm{P}(T > 2)$<br>2nd M1 for correct ratio of probability expressions [Must have $\mathrm{P}(T > t)$ or $\mathrm{P}(2 < T < t)$]<br>1st A1ft for a correct equation for $\mathrm{P}(T > t)$ (o.e.) ft their answer to part (c)[May be in a diagram]<br>3rd M1 for attempt to find $t$ (standardising and sight of 0.2533) or prepare to use calc (ft)<br>Arriving at $\mathrm{P}(T <$ median$) = 1 - 0.5\times$"their 0.8043" will score 1st 4 marks<br>2nd A1 for awrt 5.9<br>Sight of awrt 5.9 and at least one M mark scores 5/5 [Answer only send to review] |

HOME

## A2 2021

## The Normal Distribution



**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

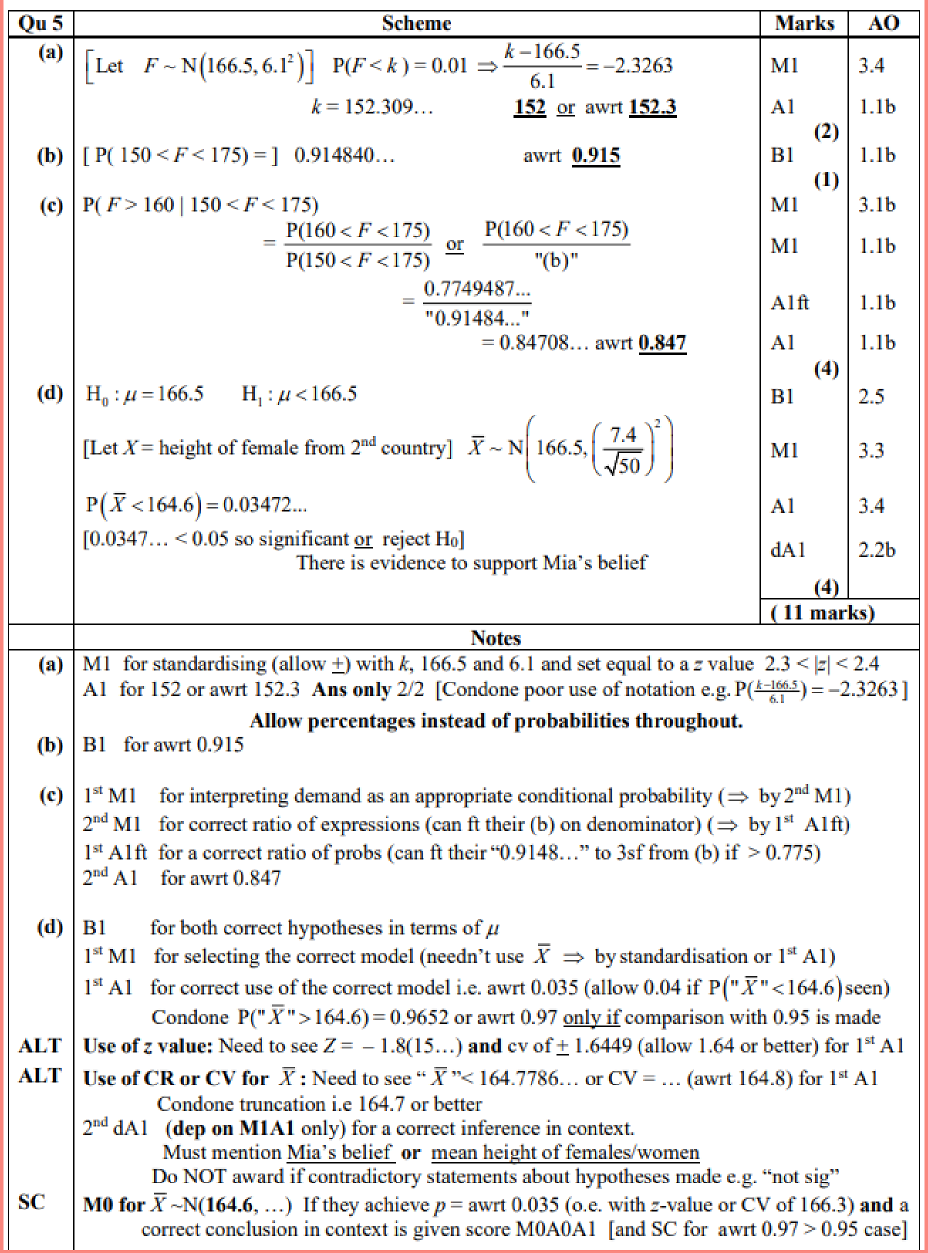
- state your hypotheses clearly
- use a 5% level of significance

**(4)**

HOME

# A2 2021

# The Normal Distribution



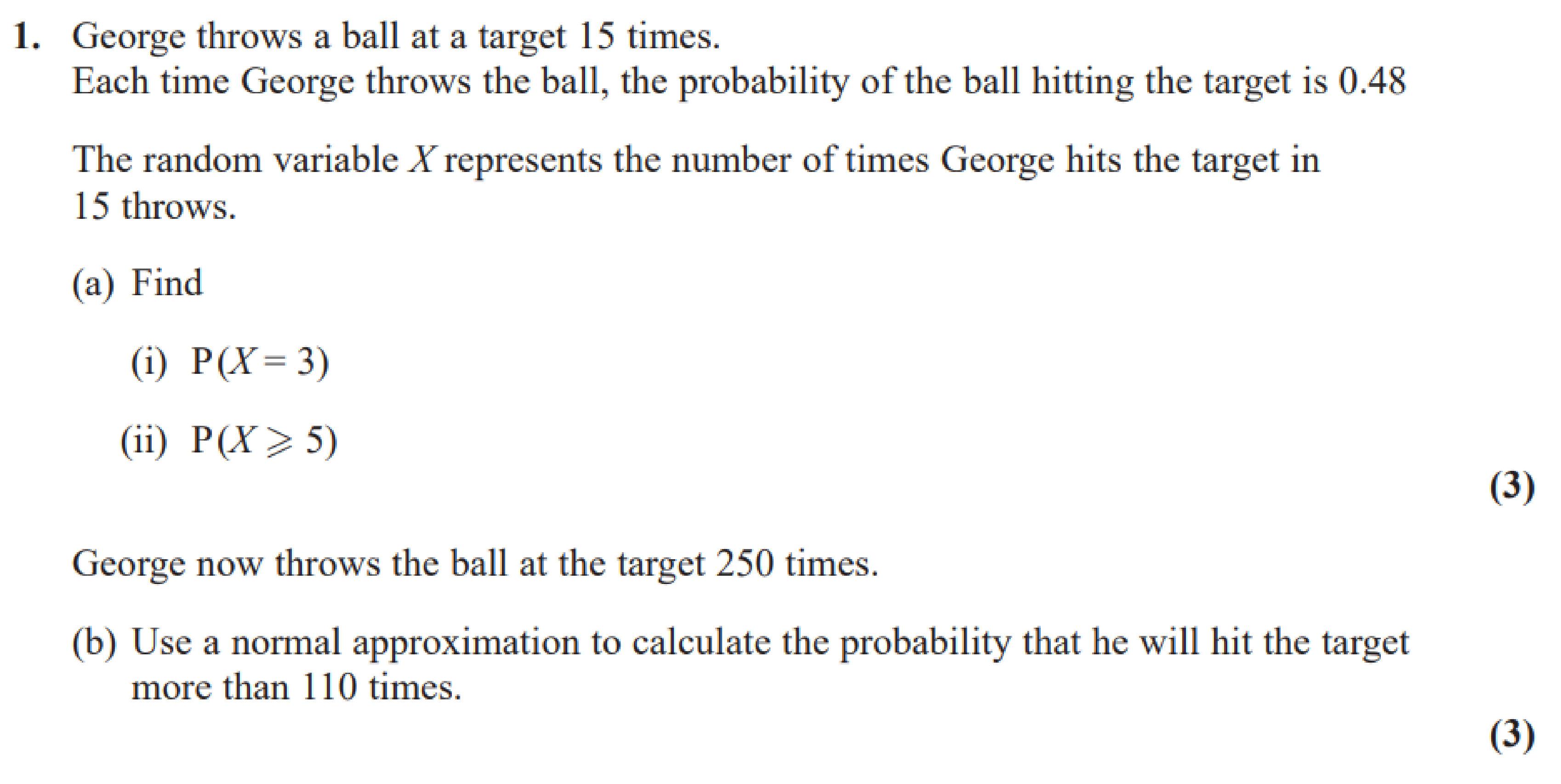
| Qu 5 | Scheme | Marks | AO |
|---|---|---|---|
| (a) | $\left[\text{Let } \ F \sim \text{N}\left(166.5, 6.1^2\right)\right]$ $\ \text{P}(F < k) = 0.01 \Rightarrow \dfrac{k-166.5}{6.1} = -2.3263$ | M1 | 3.4 |
| | $k = 152.309\ldots$ **152** or awrt **152.3** | A1 | 1.1b |
| | | (2) | |
| (b) | [ P( 150 < $F$ < 175) = ] 0.914840… awrt **0.915** | B1 | 1.1b |
| | | (1) | |
| (c) | P( $F$ > 160 \| 150 < $F$ < 175) | M1 | 3.1b |
| | $= \dfrac{\text{P}(160 < F < 175)}{\text{P}(150 < F < 175)}$ or $\dfrac{\text{P}(160 < F < 175)}{\text{"(b)"}}$ | M1 | 1.1b |
| | $= \dfrac{0.7749487\ldots}{\text{"}0.91484\ldots\text{"}}$ | A1ft | 1.1b |
| | = 0.84708… awrt **0.847** | A1 | 1.1b |
| | | (4) | |
| (d) | $\text{H}_0 : \mu = 166.5 \qquad \text{H}_1 : \mu < 166.5$ | B1 | 2.5 |
| | [Let $X$ = height of female from 2nd country] $\bar{X} \sim \text{N}\left(166.5, \left(\dfrac{7.4}{\sqrt{50}}\right)^2\right)$ | M1 | 3.3 |
| | $\text{P}\left(\bar{X} < 164.6\right) = 0.03472\ldots$ | A1 | 3.4 |
| | [0.0347… < 0.05 so significant or reject H$_0$]<br>There is evidence to support Mia's belief | dA1 | 2.2b |
| | | (4) | |
| | | ( 11 marks) | |

| | Notes |
|---|---|
| (a) | M1 for standardising (allow ±) with $k$, 166.5 and 6.1 and set equal to a $z$ value $2.3 < \lvert z \rvert < 2.4$<br>A1 for 152 or awrt 152.3 **Ans only** 2/2 [Condone poor use of notation e.g. $\text{P}(\frac{k-166.5}{6.1}) = -2.3263$ ]<br>**Allow percentages instead of probabilities throughout.** |
| (b) | B1 for awrt 0.915 |
| (c) | 1st M1 for interpreting demand as an appropriate conditional probability (⇒ by 2nd M1)<br>2nd M1 for correct ratio of expressions (can ft their (b) on denominator) (⇒ by 1st A1ft)<br>1st A1ft for a correct ratio of probs (can ft their "0.9148…" to 3sf from (b) if > 0.775)<br>2nd A1 for awrt 0.847 |
| (d) | B1 for both correct hypotheses in terms of $\mu$<br>1st M1 for selecting the correct model (needn't use $\bar{X}$ ⇒ by standardisation or 1st A1)<br>1st A1 for correct use of the correct model i.e. awrt 0.035 (allow 0.04 if $\text{P}\left(\text{"}\bar{X}\text{"} < 164.6\right)$ seen)<br>Condone $\text{P}(\text{"}\bar{X}\text{"} > 164.6) = 0.9652$ or awrt 0.97 only if comparison with 0.95 is made |
| ALT | **Use of $z$ value:** Need to see $Z = -1.8(15\ldots)$ **and** cv of ± 1.6449 (allow 1.64 or better) for 1st A1 |
| ALT | **Use of CR or CV for $\bar{X}$ :** Need to see "$\bar{X}$"< 164.7786… or CV = … (awrt 164.8) for 1st A1<br>Condone truncation i.e 164.7 or better<br>2nd dA1 (**dep on M1A1** only) for a correct inference in context.<br>Must mention Mia's belief **or** mean height of females/women<br>Do NOT award if contradictory statements about hypotheses made e.g. "not sig" |
| SC | **M0 for $\bar{X}$ ~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] |

HOME

**A2 2022**

## The Normal Distribution



**1.** George throws a ball at a target 15 times.
Each time George throws the ball, the probability of the ball hitting the target is 0.48

The random variable $X$ represents the number of times George hits the target in 15 throws.

(a) Find

(i) $P(X = 3)$

(ii) $P(X \geqslant 5)$

**(3)**

George now throws the ball at the target 250 times.

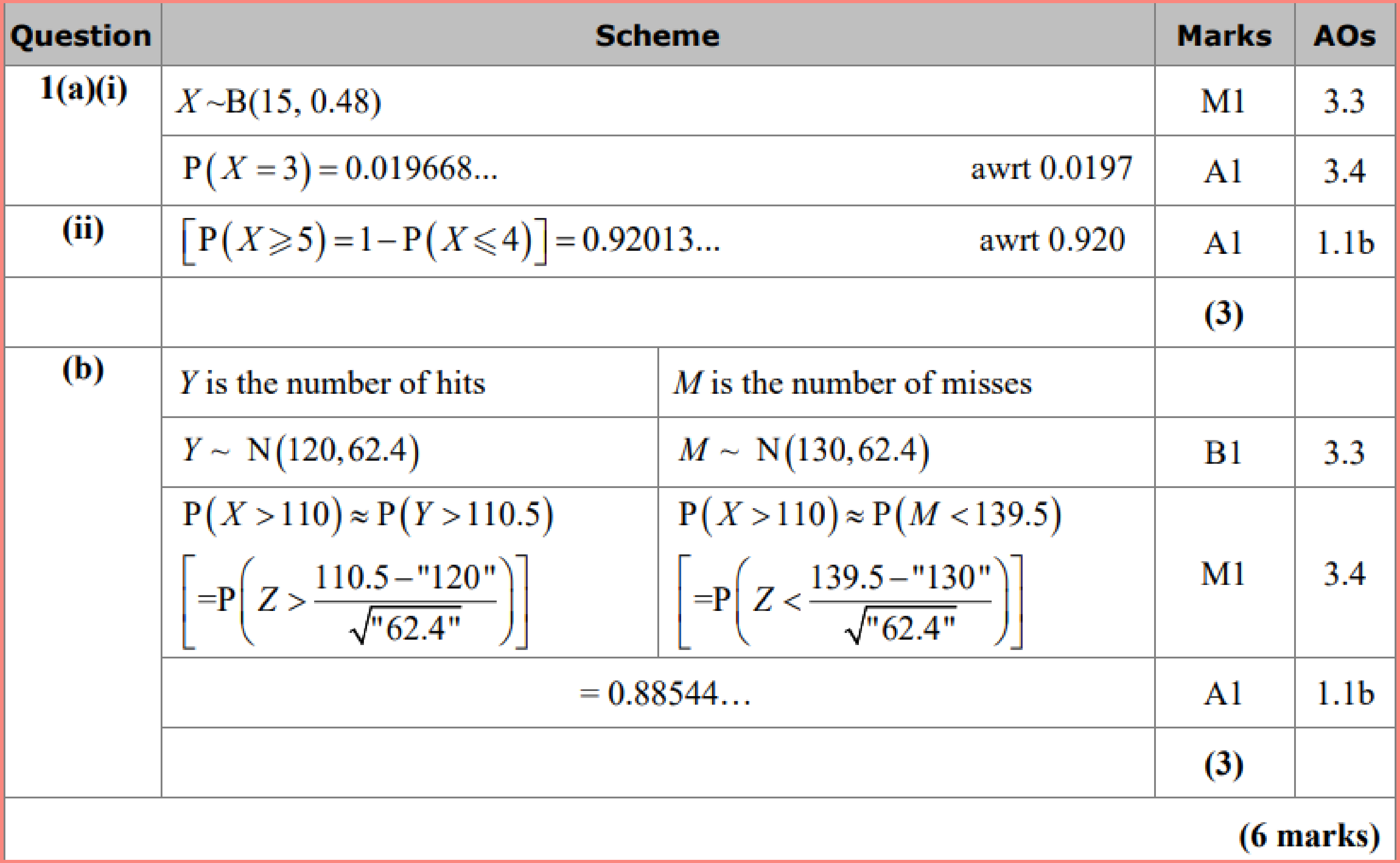
(b) Use a normal approximation to calculate the probability that he will hit the target more than 110 times.

**(3)**

HOME

# A2 2022

## The Normal Distribution

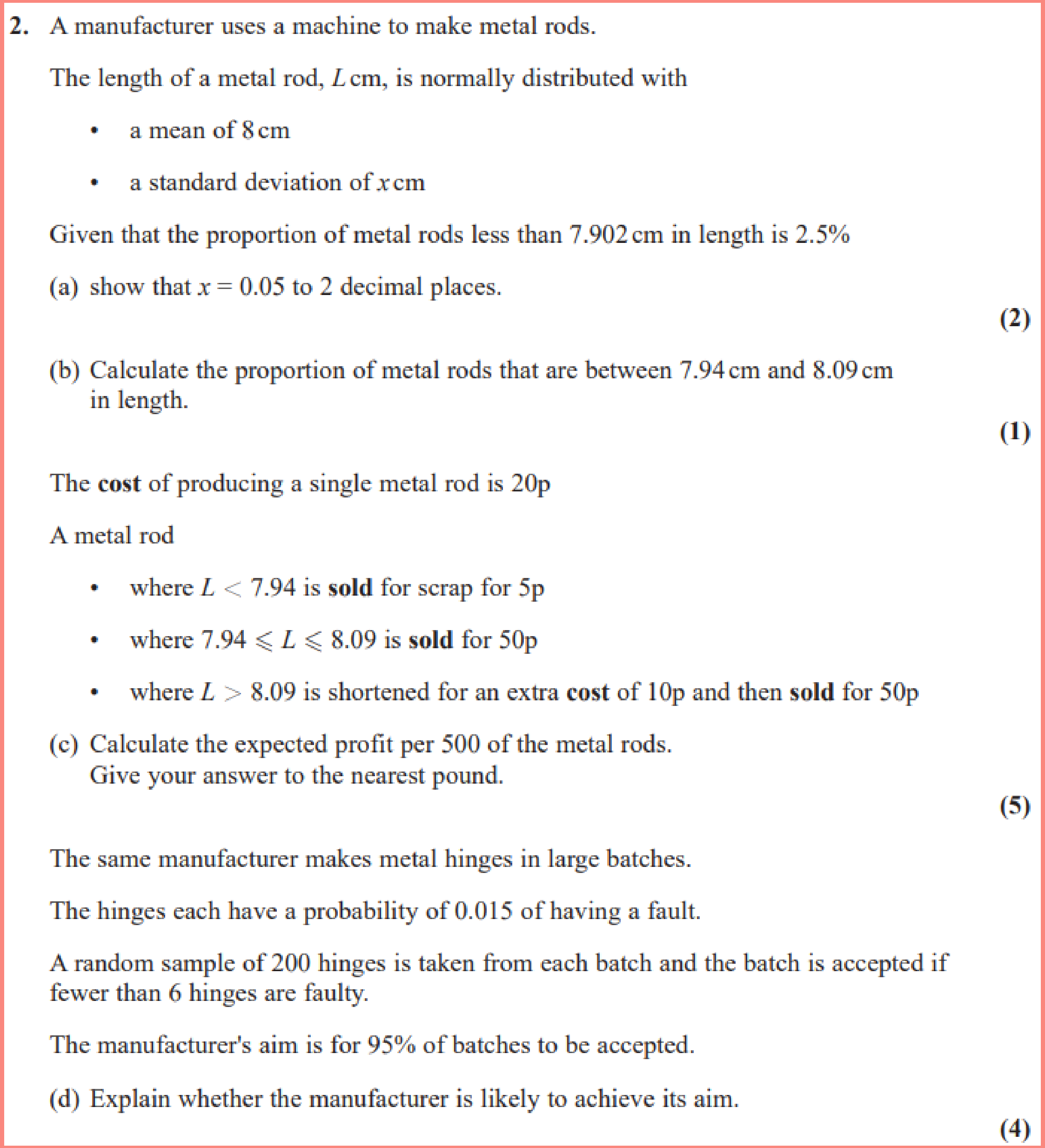


| Question | Scheme | | Marks | AOs |
|---|---|---|---|---|
| **1(a)(i)** | $X \sim B(15, 0.48)$ | | M1 | 3.3 |
| | $P(X = 3) = 0.019668...$ awrt 0.0197 | | A1 | 3.4 |
| **(ii)** | $\left[P(X \geqslant 5) = 1 - P(X \leqslant 4)\right] = 0.92013...$ awrt 0.920 | | A1 | 1.1b |
| | | | **(3)** | |
| **(b)** | $Y$ is the number of hits | $M$ is the number of misses | | |
| | $Y \sim N(120, 62.4)$ | $M \sim N(130, 62.4)$ | B1 | 3.3 |
| | $P(X > 110) \approx P(Y > 110.5)$<br>$\left[=P\left(Z > \frac{110.5 - "120"}{\sqrt{"62.4"}}\right)\right]$ | $P(X > 110) \approx P(M < 139.5)$<br>$\left[=P\left(Z < \frac{139.5 - "130"}{\sqrt{"62.4"}}\right)\right]$ | M1 | 3.4 |
| | $= 0.88544\ldots$ | | A1 | 1.1b |
| | | | **(3)** | |
| | | | **(6 marks)** | |

HOME

A2 2022

The Normal Distribution



**2.** A manufacturer uses a machine to make metal rods.

The length of a metal rod, $L$ cm, is normally distributed with

- a mean of 8 cm
- a standard deviation of $x$ cm

Given that the proportion of metal rods less than 7.902 cm in length is 2.5%

(a) show that $x = 0.05$ to 2 decimal places.

**(2)**

(b) Calculate the proportion of metal rods that are between 7.94 cm and 8.09 cm in length.

**(1)**

The **cost** of producing a single metal rod is 20p

A metal rod

- where $L < 7.94$ is **sold** for scrap for 5p
- where $7.94 \leqslant L \leqslant 8.09$ is **sold** for 50p
- where $L > 8.09$ is shortened for an extra **cost** of 10p and then **sold** for 50p

(c) Calculate the expected profit per 500 of the metal rods.
Give your answer to the nearest pound.

**(5)**

The same manufacturer makes metal hinges in large batches.

The hinges each have a probability of 0.015 of having a fault.

A random sample of 200 hinges is taken from each batch and the batch is accepted if fewer than 6 hinges are faulty.

The manufacturer's aim is for 95% of batches to be accepted.

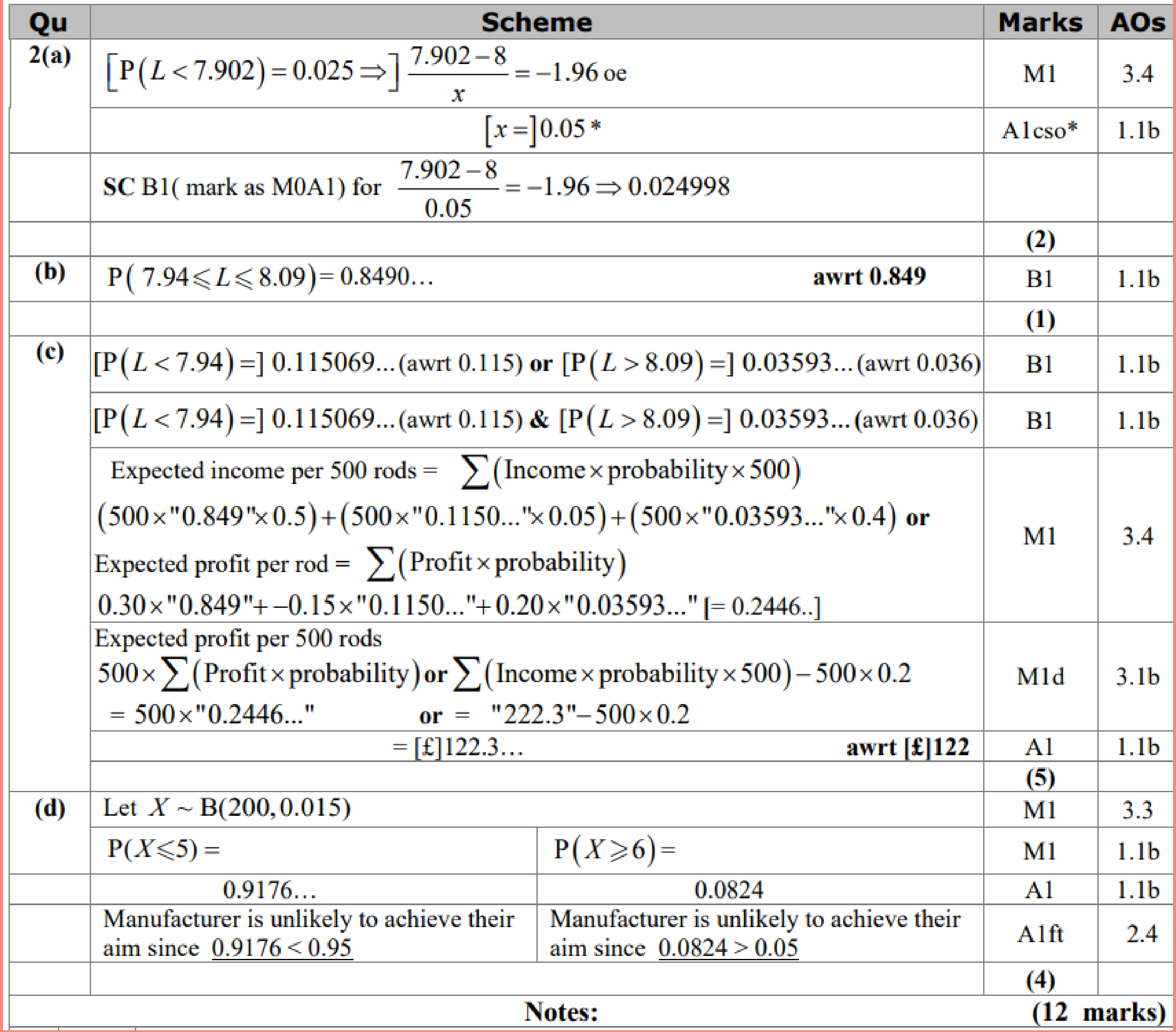
(d) Explain whether the manufacturer is likely to achieve its aim.

**(4)**

HOME

# A2 2022

## The Normal Distribution

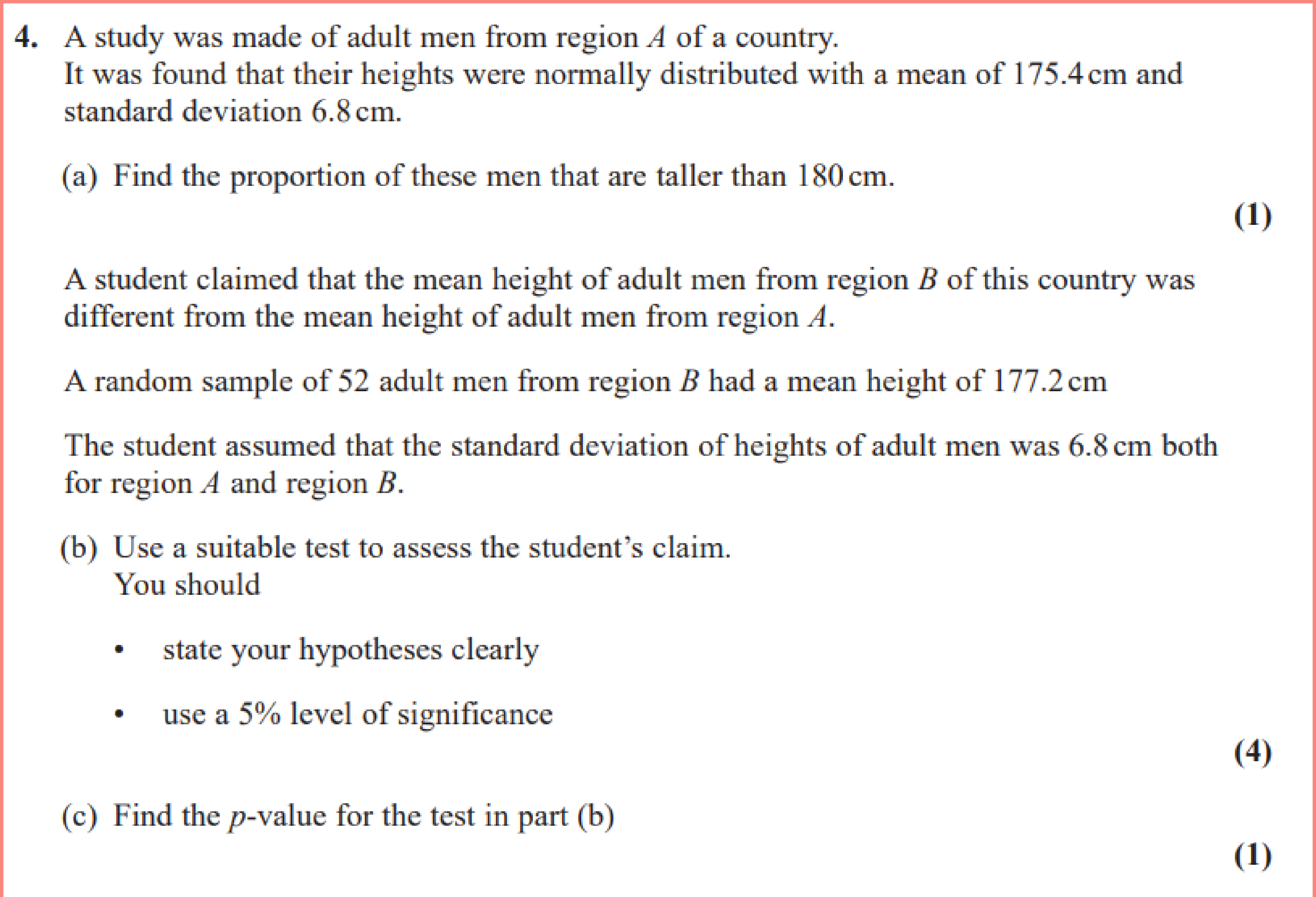


| Qu | Scheme | | Marks | AOs |
|---|---|---|---|---|
| 2(a) | $\left[\mathrm{P}(L<7.902)=0.025\Rightarrow\right]\ \dfrac{7.902-8}{x}=-1.96$ oe | | M1 | 3.4 |
| | $[x=]0.05$ * | | A1cso* | 1.1b |
| | **SC** B1( mark as M0A1) for $\dfrac{7.902-8}{0.05}=-1.96\Rightarrow 0.024998$ | | | |
| | | | **(2)** | |
| **(b)** | $\mathrm{P}(7.94\leqslant L\leqslant 8.09)=0.8490\ldots$ **awrt 0.849** | | B1 | 1.1b |
| | | | **(1)** | |
| **(c)** | $[\mathrm{P}(L<7.94)=]\ 0.115069\ldots$ (awrt 0.115) **or** $[\mathrm{P}(L>8.09)=]\ 0.03593\ldots$ (awrt 0.036) | | B1 | 1.1b |
| | $[\mathrm{P}(L<7.94)=]\ 0.115069\ldots$ (awrt 0.115) **&** $[\mathrm{P}(L>8.09)=]\ 0.03593\ldots$ (awrt 0.036) | | B1 | 1.1b |
| | Expected income per 500 rods = $\sum(\text{Income}\times\text{probability}\times 500)$<br>$(500\times\text{"}0.849\text{"}\times 0.5)+(500\times\text{"}0.1150\ldots\text{"}\times 0.05)+(500\times\text{"}0.03593\ldots\text{"}\times 0.4)$ **or**<br>Expected profit per rod = $\sum(\text{Profit}\times\text{probability})$<br>$0.30\times\text{"}0.849\text{"}+-0.15\times\text{"}0.1150\ldots\text{"}+0.20\times\text{"}0.03593\ldots\text{"}$ $[=0.2446..]$ | | M1 | 3.4 |
| | Expected profit per 500 rods<br>$500\times\sum(\text{Profit}\times\text{probability})$ **or** $\sum(\text{Income}\times\text{probability}\times 500)-500\times 0.2$<br>$=500\times\text{"}0.2446\ldots\text{"}$ **or** $=\text{"}222.3\text{"}-500\times 0.2$ | | M1d | 3.1b |
| | $=[£]122.3\ldots$ **awrt [£]122** | | A1 | 1.1b |
| | | | **(5)** | |
| **(d)** | Let $X\sim\mathrm{B}(200,0.015)$ | | M1 | 3.3 |
| | $\mathrm{P}(X\leqslant 5)=$ | $\mathrm{P}(X\geqslant 6)=$ | M1 | 1.1b |
| | $0.9176\ldots$ | $0.0824$ | A1 | 1.1b |
| | Manufacturer is unlikely to achieve their aim since $\underline{0.9176<0.95}$ | Manufacturer is unlikely to achieve their aim since $\underline{0.0824>0.05}$ | A1ft | 2.4 |
| | | | **(4)** | |
| | **Notes:** | | **(12 marks)** | |

HOME

A2 2023

The Normal Distribution



**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)**

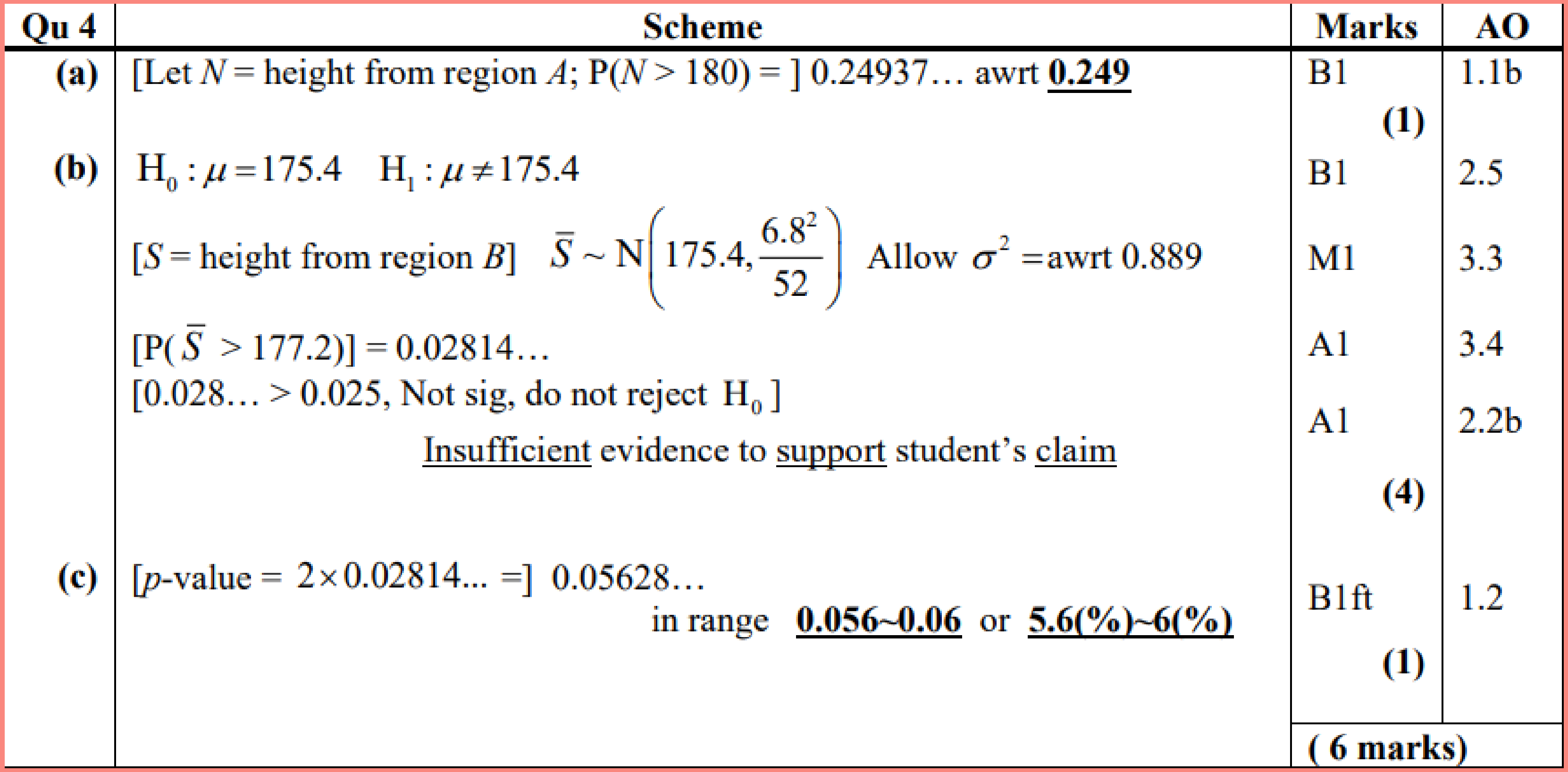
(c) Find the *p*-value for the test in part (b)

**(1)**

HOME

A2 2023

## The Normal Distribution



| Qu 4 | Scheme | Marks | AO |
|---|---|---|---|
| (a) | [Let $N$ = height from region $A$; P($N$ > 180) = ] 0.24937… awrt **0.249** | B1 | 1.1b |
| | | (1) | |
| (b) | $\mathrm{H}_0 : \mu = 175.4 \quad \mathrm{H}_1 : \mu \neq 175.4$ | B1 | 2.5 |
| | [$S$ = height from region $B$] $\bar{S} \sim \mathrm{N}\left(175.4, \frac{6.8^2}{52}\right)$ Allow $\sigma^2$ =awrt 0.889 | M1 | 3.3 |
| | [P($\bar{S}$ > 177.2)] = 0.02814… | A1 | 3.4 |
| | [0.028… > 0.025, Not sig, do not reject $\mathrm{H}_0$]<br>Insufficient evidence to support student's claim | A1 | 2.2b |
| | | (4) | |
| (c) | [$p$-value = $2 \times 0.02814...$ =] 0.05628…<br>in range **0.056~0.06** or **5.6(%)~6(%)** | B1ft | 1.2 |
| | | (1) | |
| | | ( 6 marks) | |

HOME