Additional Assessment Materials

Summer 2021

Pearson Edexcel GCE in As Mathematics

8MA0_21 (Public release version)

Resource Set 1: Topic 2

Data presentation

Pearson: helping people progress, everywhere

Pearson aspires to be the world's leading learning company. Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Additional Assessment Materials, Summer 2021

General guidance to Additional Assessment Materials for use in 2021

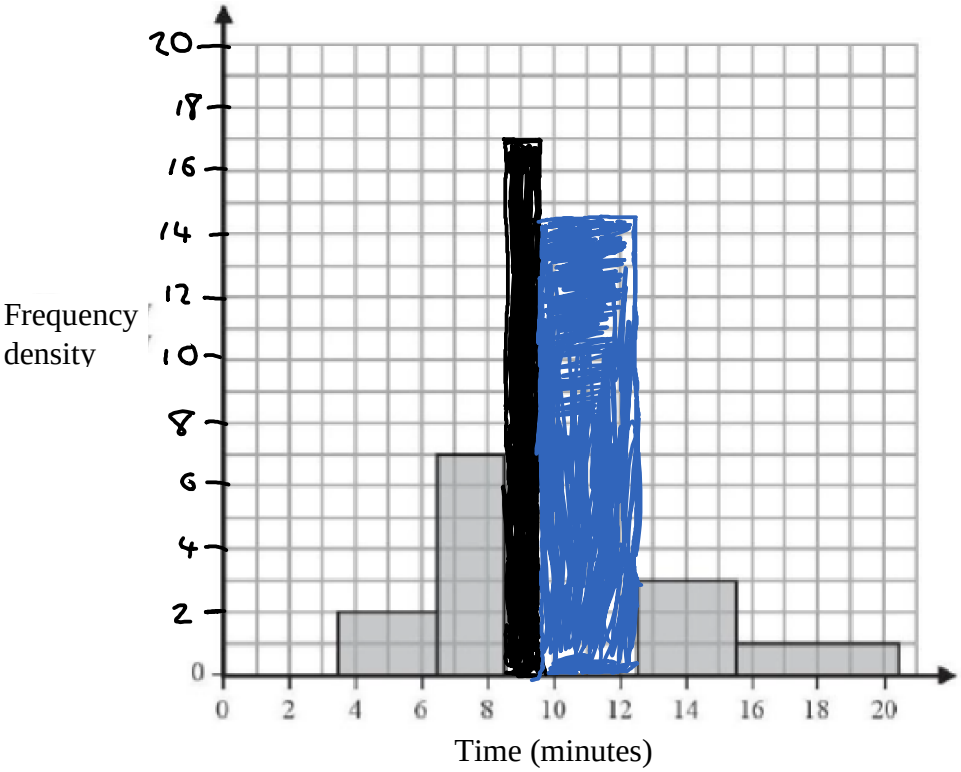
Context

- Additional Assessment Materials are being produced for GCSE, AS and A levels (with the exception of Art and Design).
- The Additional Assessment Materials presented in this booklet are an optional part of the range of evidence teachers may use when deciding on a candidate's grade.
- 2021 Additional Assessment Materials have been drawn from previous examination materials, namely past papers.
- Additional Assessment Materials have come from past papers both published (those materials available publicly) and unpublished (those currently under padlock to our centres) presented in a different format to allow teachers to adapt them for use with candidate.

Purpose

- The purpose of this resource to provide qualification-specific sets/groups of questions covering the knowledge, skills and understanding relevant to this Pearson qualification.
- This document should be used in conjunction with the mapping guidance which will map content and/or skills covered within each set of questions.
- These materials are only intended to support the summer 2021 series.

1. 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	14
9	17
10 – 12	45
13 – 15	9
16 – 20	5

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

(5)

Between 4–6 minutes is actually 3.5–6.5 minutes

Grid area = 6 blocks

so height = frequency density = 2

Now label frequency density column going up in 2.

For 7–8 minutes → 6.5 – 8.5
width = 2 area = 7 × 2 = 14 Frequency

For 16–20 minutes → 15.5 – 20.5
width = 5 area = 5 × 1 = 5 frequency

Total number of motorists

6 + 14 + 17 + 45 + 9 + 5 = 96

Number delayed 8.5 – 13.5 minutes = 17 + 45 + 3
= 65

(65/96) × 100 = 67.7 %

(Total for Question 1 is 5 marks)

---

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)

$0 \leq d < 50$

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

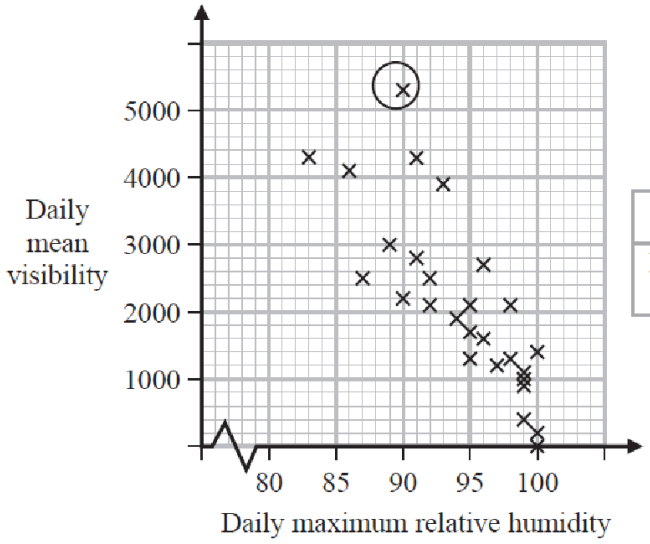


Figure 2

	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)

$Q_1 - (1.5 \times IQR) \rightarrow 1100 - (1.5 \times 1600) = -1300$

$Q_3 + (1.5 \times IQR) \rightarrow 2700 + (1.5 \times 1600) = 5100$

$5300 > 5100$ so its an outlier above Q_3.

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

(1)

Negative correlation

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.

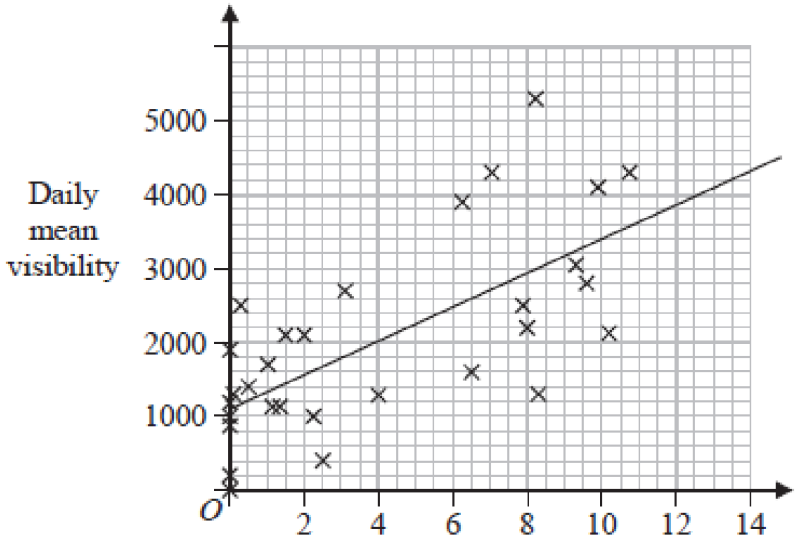


Figure 3

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

(1)

Daily mean windspeed

(Total for Question 2 is 5 marks)

---

3. 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)

Positive correlation

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

(1)

Pay increases by £4.5 for every extra point

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

(1)

It a worker gets less than 11 points they will get negative pay according to this model. Cant happen in real life.

(Total for Question 3 is 3 marks)

4.

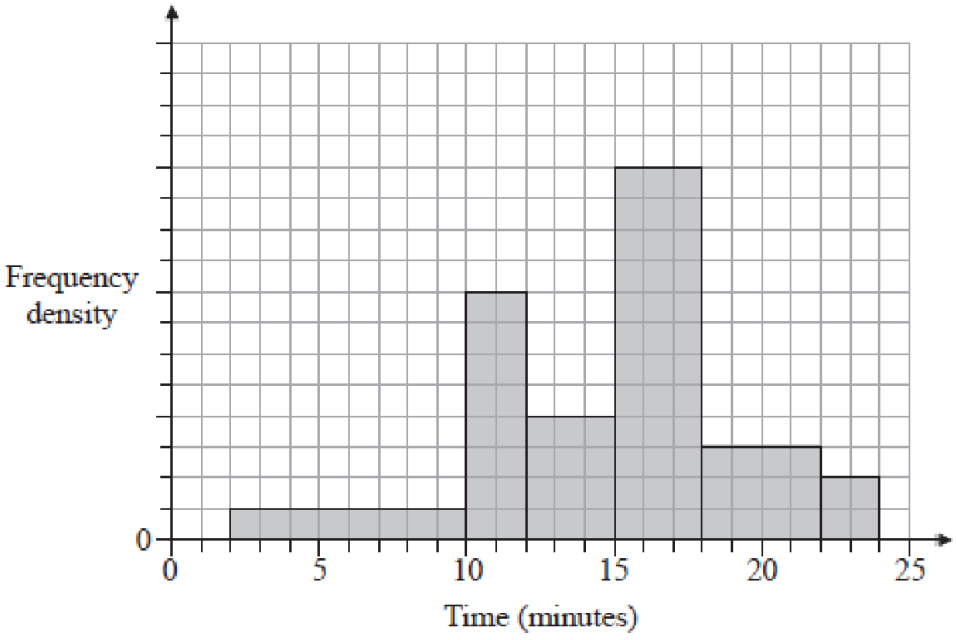


Figure 1

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)

above 15 minutes = 78 students

amount of blocks above 15 minutes = 52 blocks

78 students = 52 blocks → 78/52 = 1.5 student per block

Boxes for less than 11 minutes = 16

16 × 1.5 = 24 students took less than 11 minutes

Total students = 78 + 24 + 30 = 132

(24/132) × 100 = 18.18...%

(Total for Question 4 is 4 marks)

---

5. 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.

Some rainfall values are trace. These need to be removed or converted to numerical values. **(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 $\sum f x = 539.75$ and $\sum f x^2 = 7704.1875$

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

(2)

$\Sigma f = 184 \quad Q_3 = \frac{3 \times 184}{4} = 138^{th}$ value

138th value falls in $1.0 \leq r < 5.0$ group

$Q_3 = 1 + (7/24 \times 4) = 2.17$ mm

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

(2)

$$\sigma = \sqrt{\frac{\Sigma f x^2}{\Sigma f} - \left(\frac{\Sigma f x}{\Sigma f}\right)^2} \rightarrow \sqrt{\frac{7704.1875}{184} - \left(\frac{559.75}{184}\right)^2}$$

$= 5.77$

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

Using class midpoints to estimate the mean assumes that the values are uniformly distributed.

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

May to October contains mostly Summer months. As a result most values in $0 \leq r < 0.5$ will be 0. So the assumption wont hold.

(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)

Following on from answer to (ii) the actual mean is likely to be smaller.

(Total for Question 5 is 8 marks)