



Additional Assessment Materials

Summer 2021

Pearson Edexcel GCE in Mathematics

9MA0 (Applied) (Public release version)

Resource Set 1: Topic 2

Data Presentation and Interpretation

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

All the material in this publication is copyright

© Pearson Education Ltd 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- Additional Assessment Materials have come from past papers both published (those materials available publicly) and unpublished (those currently under padlock to our centres) presented in a different format to allow teachers to adapt them for use with candidate.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 A random sample of 15 days is taken from the large data set for Perth in June and July 1987. The scatter diagram in Figure 1 displays the values of two of the variables for these 15 days.

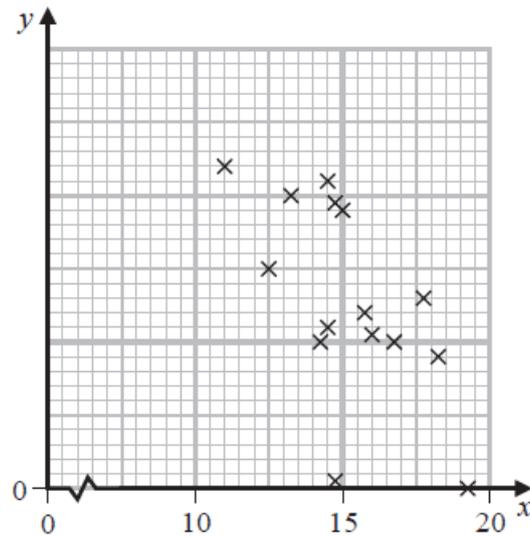


Figure 1

- (a) Describe the correlation. (1)

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

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

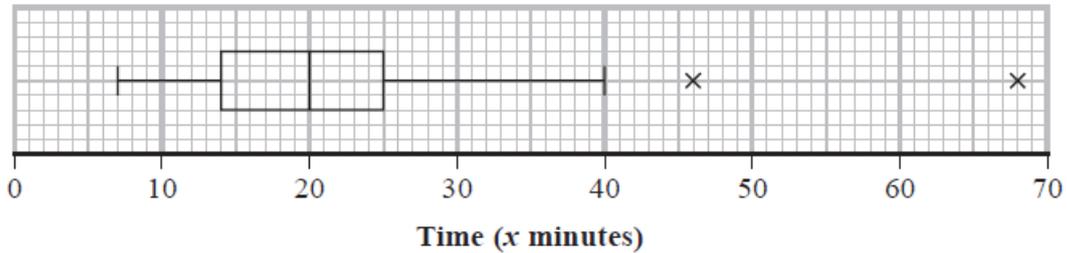
(1)

(Total for Question 1 is 7 marks)

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

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

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



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

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

For these 27 people $\sum x = 607.5$ and $\sum x^2 = 17623.25$

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

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

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

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

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

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

- the median time increases
- the mean time does not change

(f) Suggest a possible value for a and a possible value for b , explaining how your values satisfy the above conditions. (3)

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

(1)

(Total for Question 2 is 10 marks)

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

Hours	$0 \leq y < 5$	$5 \leq y < 8$	$8 \leq y < 11$	$11 \leq y < 12$	$12 \leq y < 14$
Frequency	12	6	8	3	2

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

- (a) Find the width and the height of the $0 \leq y < 5$ 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 respectively. 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 $N(6.6, 3.7^2)$.

- (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)

(Total for Question 3 is 13 marks)

4. The partially completed table below summarises the times taken by 120 job applicants to complete a task.

Time, t (minutes)	$5 < t \leq 7$	$7 < t \leq 10$	$10 < t \leq 14$	$14 < t \leq 18$	$18 < t \leq 30$
Frequency	10	23	51		

A histogram is drawn. The bar representing the $5 < t \leq 7$ has a width of 1 cm and a height of 5 cm.

- (a) Given that the bar representing the group $14 < t \leq 18$ has a height of 4 cm, find the frequency of this group. (2)
- (b) Showing your working, estimate the mean time taken by the 120 job applicants. (3)

The lower quartile of the times is 9.6 minutes and the upper quartile of the times is 15.5 minutes.

For these data, an outlier is classified as any value greater than $Q_3 + 1.5 \times \text{IQR}$.

- (c) Showing your working, explain whether or not any of the times taken by these 120 job applicants might be classified as outliers. (2)

Candidates with the fastest 5% of times for the task are given interviews.

- (d) Estimate the time taken by a job applicant, below which they might be given an interview. (2)

(Total for Question 4 is 9 marks)

5.

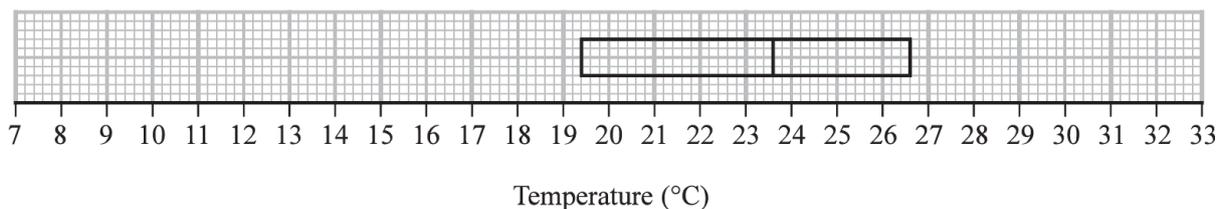


Figure 1

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

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

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

- (a) Complete the box plot in Figure 1 showing clearly any outliers (4)
- (b) Using your knowledge of the large data set, suggest from which month the two outliers are likely to have come. (1)

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

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

- (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

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

- (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)

(Total for Question 5 is 11 marks)

6. 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.

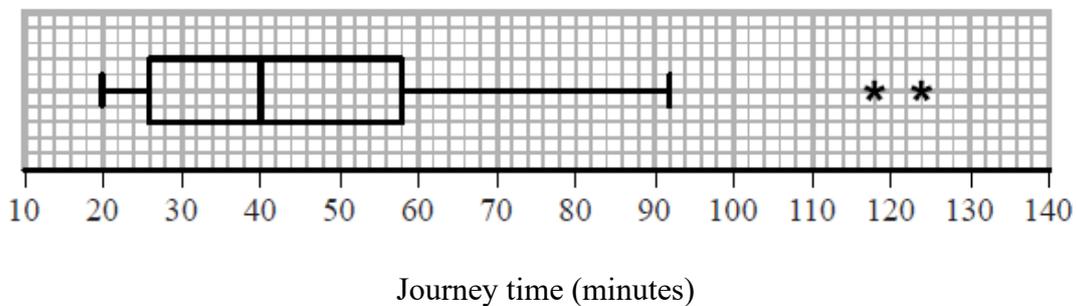
(a) State the sampling method Charlie used. (1)

(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.



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

(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)

(Total for Question 6 is 13 marks)