Pearson Edexcel AS Mathematics 8MA0

Statistics – Representation of Data

Time allowed: 45 minutes

School: www.CasperYC.club

Name:

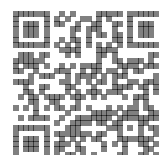**Teacher:**

Question	Points	Score
1	8	
2	10	
3	13	
4	9	
5	5	
6	5	
Total:	50	

How I can achieve better:

-
-
-

1. A travel agent sells flights to different destinations from Southstead airport. The distance of the destination from the airport is denoted d, where d is measured in 100 km units, so that $d = 2.2$ represents a distance of 220 km. Values of d and the associated fare £f are recorded for a random sample of 6 destinations.

| Destination	A	B	C	D	E	F
d (100 km)	2.2	4.0	6.0	2.5	8.0	5.0
f (£)	18	20	25	23	32	28

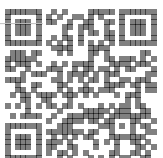(a) Using the axes below, complete a scatter diagram to illustrate this information. [2]
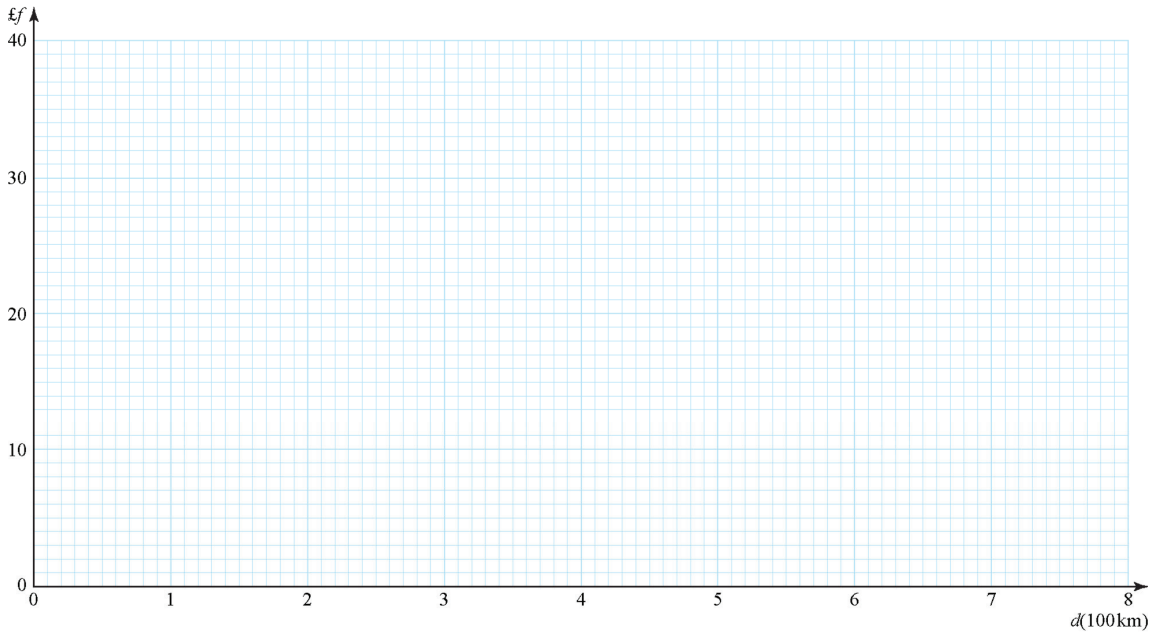
(b) Explain why a linear model may be appropriate to describe the relationship between f and d. [1]

(c) State which of f and d should be considered the response variable. [1]

(d) Use a line of best fit to estimate a fare £f for a flight to a destination which is 700 km away. [2]

(e) Comment on the reliability of your estimate, giving a reason for your answer. [1]

Jane is planning her holiday and wishes to fly from Southstead airport to a destination 180 km away.

(f) State if it is sensible for Jane to estimate the fare of her flight using the scatter graph, giving a reason for your answer. [1]

Total: 8

(Q1 continued . . .)
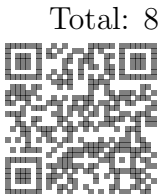
2. A random sample of distances travelled to work for 120 commuters from a train station in Devon is recorded. The distances travelled, to the nearest mile, are summarised below.

Distance	Number of commuters
$0-9$	10
$10-19$	19
$20-29$	43
$30-39$	25
$40-49$	8
$50-59$	6
$60-69$	5
$70-79$	3
$80-89$	1

For this distribution:

(a) estimate the median. [2]

The mid-point of each class was represented by x and its corresponding frequency by f.

The mid-point of the lowest class was taken to be 4.75 giving:

$$\sum fx = 3552.5 \quad \text{and} \quad \sum fx^2 = 138043.125$$

(b) Estimate the mean and the standard deviation of this distribution. [3]

(c) Explain why the median is less than the mean for these data. [1]

(d) For a second random sample of 120 commuters travelling to work from a train station in Greater London, the mean distance travelled to work is 15.6 miles with standard deviation 21.2 miles. Compare the measures of location and spread for the distance travelled to work for the two samples, giving possible reasons for any differences. [4]
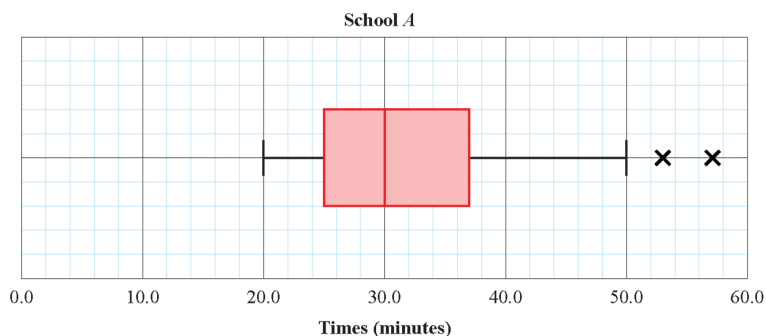
Total: 10

(Q2 continued . . .)
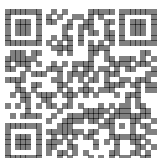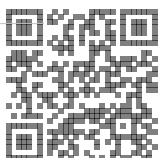
3. Children from schools A and B took part in a fun run for charity. The times, to the nearest minute, taken by the children from school A are summarised in the figure below.
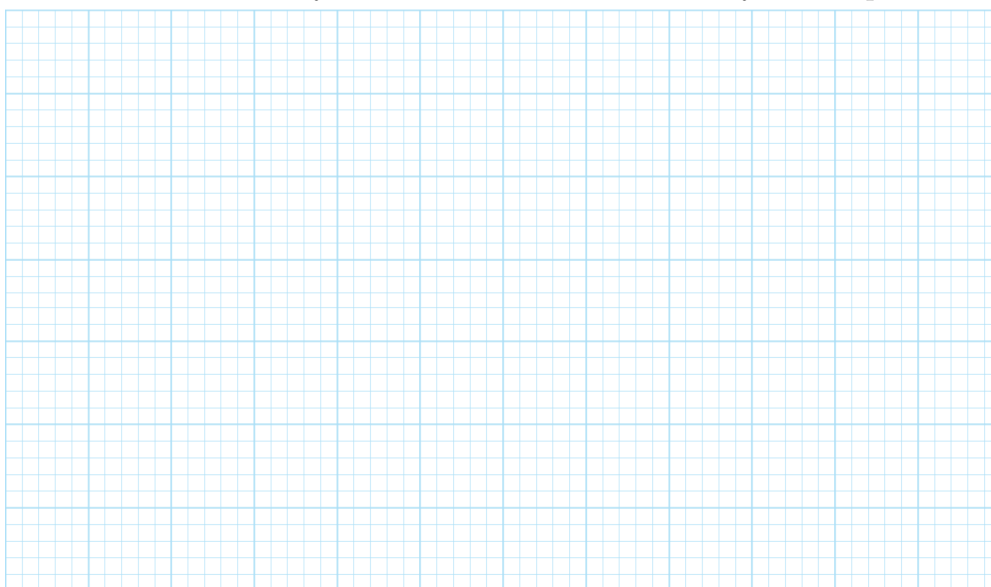
School A

Times (minutes)

(a) i. Write down the time by which 75% of the children in school A had completed the run.   [1]

   ii. State the name given to this value.   [1]

(b) Explain what the two crosses (X) represent on the box plot above. Interpret these in context.   [2]

For school B the least time taken by any of the children was 25 minutes and the longest time was 55 minutes. The three quartiles were $30, 37$ and 50 minutes respectively.
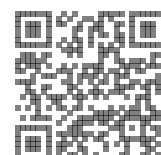
(c) Use $Q_3 + 1.5(Q_3 - Q_1)$ and $Q_1 - 1.5(Q_3 - Q_1)$ to determine if there are any outliers.   [3]
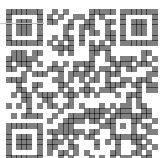   Give a reason for your conclusion.

(d) Draw a horizontal box plot to represent the data from school B so the distribution of the times taken for the fun run by children in school B can easily be compared to school A.   [3]

(e) Compare and contrast the two distributions in context.   [3]

Total: 13

(Q3 continued . . .)

4. Data relating to the lifetimes (to the nearest hour) of a random sample of 200 light bulbs from   [9]
   the production line of a manufacturer were summarised in a grouped frequency table. The mid-
   point of each class in the table was represented by x and the corresponding frequency for that
   class by f. The data were then coded using:

$$y = \frac{x - 755.0}{2.5}$$

and summarised as follows:

$$\sum fy = -467, \qquad \sum fy^2 = 9179.$$

Calculate estimates of the mean and the standard deviation of the lifetimes of this sample of
bulbs.

(Q4 continued ...)
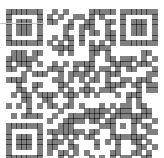
5. Cotinine is a chemical that is made by the body from nicotine which is found in cigarette smoke. A doctor tested the blood of 12 patients, who claimed to smoke one packet of cigarettes per day, for cotinine. The results are shown below.

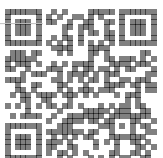| Patient	A	B	C	D	E	F	G	H	I	J	K	L
Cotinine level, x (ng/ml)	160	390	169	175	125	420	171	250	210	258	186	243

(a) The lower quartile of these data is 170 ng/ml. Find the upper quartile. [2]

A doctor suspects that some of his patients have been smoking more than one packet of cigarettes per day. He decides to use $Q_3 + 1.5(Q_3 - Q_1)$ to determine if any of the cotinine results are far enough away from the upper quartile to be considered as outliers.

(b) Identify which patient(s) the doctor suspects may have been smoking more than one packet [3] of cigarettes per day on this basis. Show your working clearly.

Total: 5

(Q5 continued . . .)
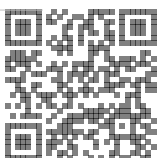
6. A couple wish to go on a holiday sometime in July or August. They cannot decide whether to   [5]
   visit Perth in Australia or Beijing in China. The weather will be a deciding factor. The box plots
   show the daily mean temperature and daily rainfall for Perth and Beijing in July and August
   2015. Based on the box plots, state three observations that may help the couple decide whether
   to spend their holiday in Perth or Beijing.

Daily mean temperature for Perth and Beijing, July and August 2015

Daily Rainfall for Perth and Beijing, July and August 2015

(Q6 continued . . .)