

DATA PRESENTATION & INTERPRETATION

AS Unit 2: Applied Mathematics A

Section B: Statistics

Past paper and specimen questions

Total marks available 80 (approximately 1 hour 40 minutes)

1. A small chapel was open to visitors for 55 days during the summer of 2015. The table summarises the daily numbers of visitors.

Number of visitors	Number of days
20 or fewer	1
21	2
22	3
23	6
24	8
25	10
26	13
27	7
28	2
29	1
30 or more	2
Total	55

- (a) For these data:
- (i) state the modal value; [1 mark]
- (ii) find values for the median and the interquartile range. [2 marks]
- (b) Name one measure of average **and** one measure of spread that cannot be calculated exactly from the data in the table. [2 marks]
- (c) Reference to the raw data revealed that the 3 unknown exact values in the table were 13, 37 and 58.
- Making use of this additional information, together with the data in the table, calculate the value of **each** of the two measures that you named in part (b). [3 marks]

(AQA 2016)

2. Bernie recorded the durations, to the nearest minute, of 23 telephone calls made from his landline. The recorded durations, in numerical order, are as follows.

5 5 5 5 10 10 10 13 14 15 16 17 18 19 20 21 22 22 23 30 30 35 95

- (a) For these data, give a reason why:
 - (i) the mode is **not** a suitable measure of average;
 - (ii) the range is **not** a suitable measure of spread. [2 marks]
- (b) Determine values for the median and the interquartile range of the 23 durations. [2 marks]
- (c) Calculate values for the mean and the standard deviation of the 23 durations. [3 marks]
- (d) Give **two** reasons why the measures determined in part (b) might be more appropriate than those calculated in part (c) for summarising the 23 durations. [2 marks]

(AQA 2018)

3. The heights, in centimetres, of a sample of 80 men were measured. For each man, his height in **excess of 175 cm** was recorded. The results are summarised in the table.

Height (cm)	Number of men
0 – 2	5
2 – 4	9
4 – 6	11
6 – 8	21
8 – 10	17
10 – 12	10
12 – 14	7
Total	80

- (a) Calculate estimates for the mean and the variance of the heights in the table. [4 marks]
- (b) Hence find estimates for the mean and the variance of the actual heights of the 80 men. [2 marks]
- (c) Given that 1 foot is equal to 30.48 cm, find, in **feet**, estimates for the mean and the variance of the actual heights of the 80 men. Give your answers to three significant figures. [3 marks]

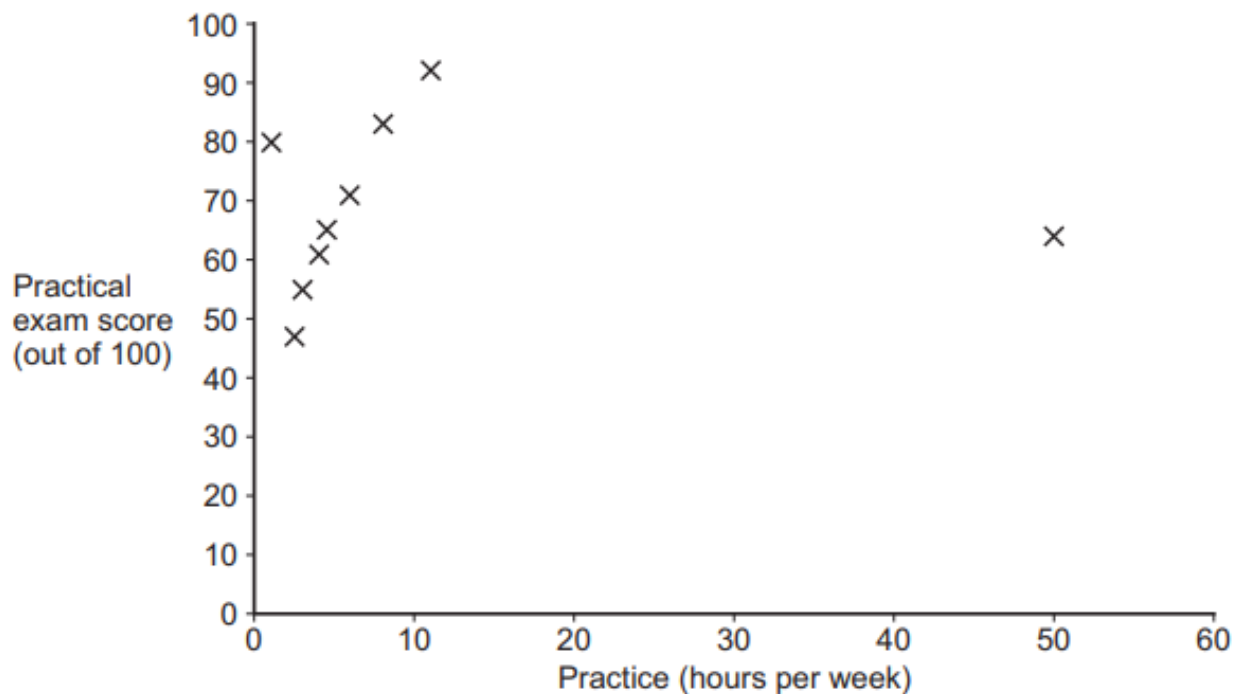
(AQA 2017)

4. Jennie is a piano teacher who teaches nine pupils.

She records how many hours per week they practice the piano along with their most recent practical exam score.

Student	Practice (hours per week)	Practical exam score (out of 100)
Donovan	50	64
Vazquez	6	71
Higgins	3	55
Begum	2.5	47
Collins	1	80
Coldbridge	4	61
Nedbalek	4.5	65
Carter	8	83
White	11	92

She plots a scatter diagram of this data, as shown below.



(a) Identify two possible outliers by name, giving a possible explanation for the position on the scatter diagram of each outlier. **[4 marks]**

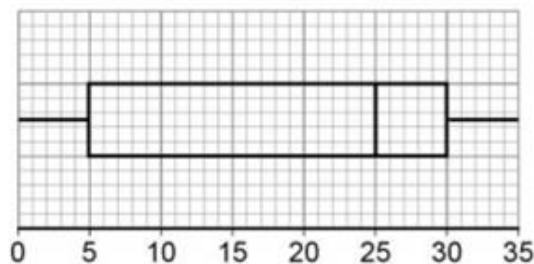
(b) Jennie discards the two outliers.

(b) (i) Describe the correlation shown by the scatter diagram for the remaining points. **[1 mark]**

(b) (ii) Interpret this correlation in the context of the question. **[1 mark]**

(AQA 2018)

5. The boxplot below represents the time spent in hours by students revising for a history exam.



(a) Use the information in the boxplot to state the value of a measure of central tendency of the revision times, stating clearly which measure you are using. **[1 mark]**

(b) Use the information in the boxplot to explain why the distribution of revision times is negatively skewed. **[1 mark]**

(AQA Sample)

6. (a) The weekly incomes, rounded to the nearest pound, of households on a certain street in Belfast are given below:

221 248 251 255 259 263 264 272 291 297 374

- (i) Calculate the mean and standard deviation for these weekly incomes. [4]

Outliers are defined to be more than $1.5 \times (Q_3 - Q_1)$ above Q_3 or
more than $1.5 \times (Q_3 - Q_1)$ below Q_1

- (ii) Show that £374 is an outlier. [3]

Further investigations suggested that the weekly income of £374 was recorded incorrectly and should in fact have been £347

Without carrying out any further investigation, state what effect, if any, this change would have on:

- (iii) the standard deviation, [1]
- (iv) the interquartile range. [1]

- (b) The histogram in **Fig. 4** below shows the weekly incomes, in £, of a random sample of households in a neighbourhood.

Histogram showing weekly incomes

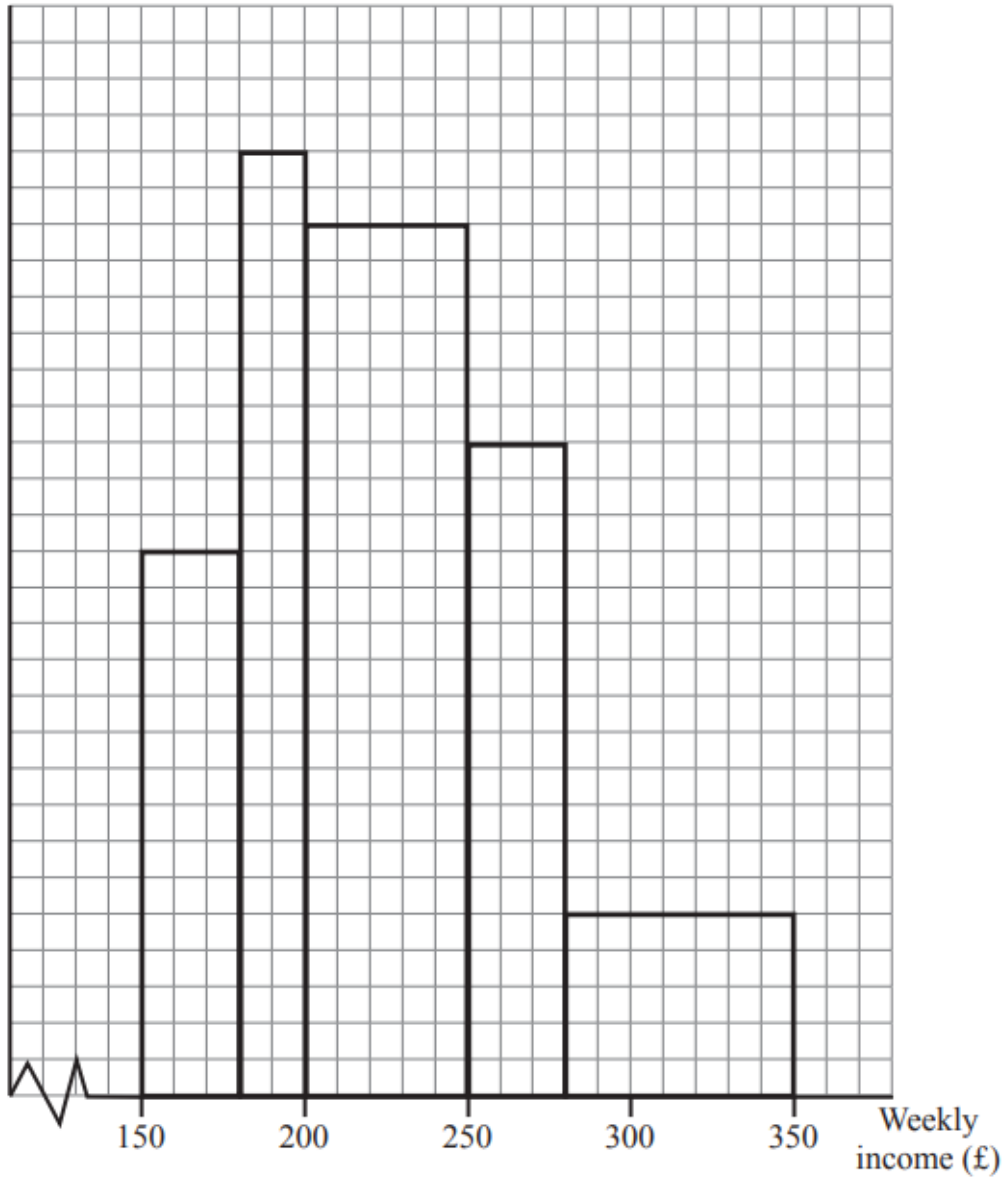


Fig. 4

- (i) Complete the table below. [1]

Weekly incomes (£)	Number of households
150 –	45
180 –	52
200 –	120
250 –	
280 –	35
350 –	0

- (ii) Estimate the probability that a household chosen at random received a weekly income of more than £230 [2]

(CEA 2019)

7.

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.

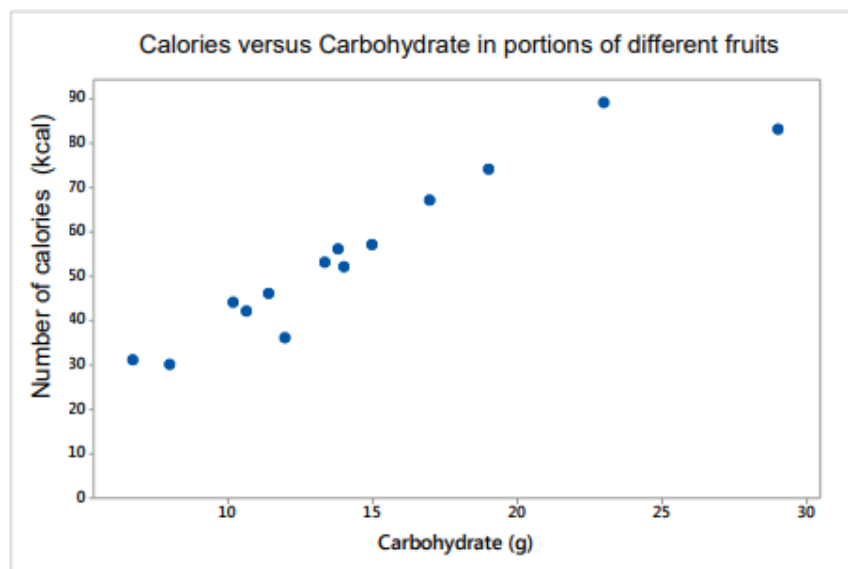
- (c) State, giving a reason, whether or not the calculations in part (b) support Thomas' belief. (2)

(EDEXCEL SPECIMEN)

8. A researcher wishes to investigate the relationship between the amount of carbohydrate and the number of calories in different fruits. He compiles a list of 90 different fruits, e.g. apricots, kiwi fruits, raspberries.

As he does not have enough time to collect data for each of the 90 different fruits, he decides to select a simple random sample of 14 different fruits from the list. For each fruit selected, he then uses a dieting website to find the number of calories (kcal) and the amount of carbohydrate (g) in a typical adult portion (e.g. a whole apple, a bunch of 10 grapes, half a cup of strawberries). He enters these data into a spreadsheet for analysis.

- (a) Explain how the random number function on a calculator could be used to select this sample of 14 different fruits. [3]
- (b) The scatter graph represents 'Number of calories' against 'Carbohydrate' for the sample of 14 different fruits.
- (i) Describe the correlation between 'Number of calories' and 'Carbohydrate'. [1]
- (ii) Interpret the correlation between 'Number of calories' and 'Carbohydrate' in this context. [1]



- (c) The equation of the regression line for this dataset is:

$$\text{'Number of calories'} = 12.4 + 2.9 \times \text{'Carbohydrate'}$$

- (i) Interpret the gradient of the regression line in this context. [1]
- (ii) Explain why it is reasonable for the regression line to have a non-zero intercept in this context. [1]

(WJEC SPECIMEN)

9. Gareth has a keen interest in pop music. He recently read the following claim in a music magazine.

In the pop industry most songs on the radio are not longer than three minutes.

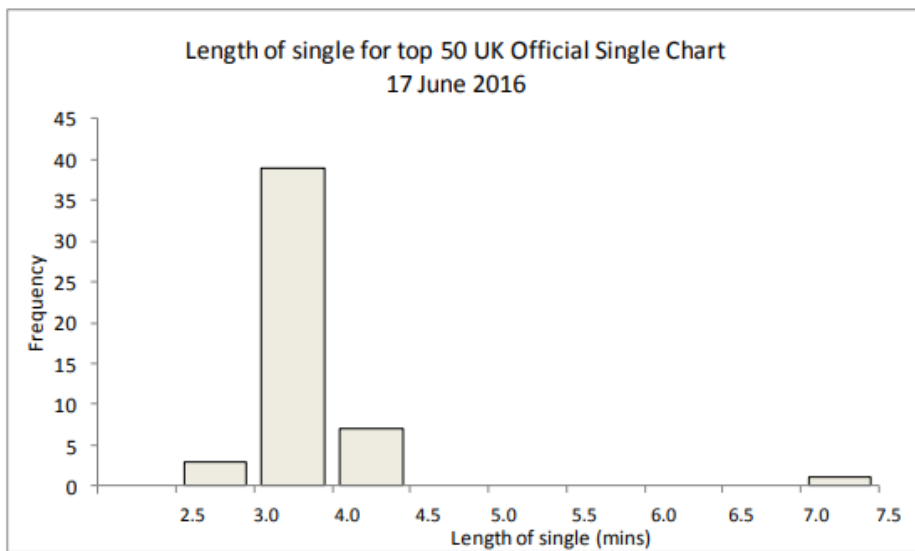
- (a) He decided to investigate this claim by recording the lengths of the top 50 singles in the UK Official Singles Chart for the week beginning 17 June 2016. (A 'single' in this context is one digital audio track.)

Comment on the suitability of this sample to investigate the magazine's claim. [1]

- (b) Gareth recorded the data in the table below.

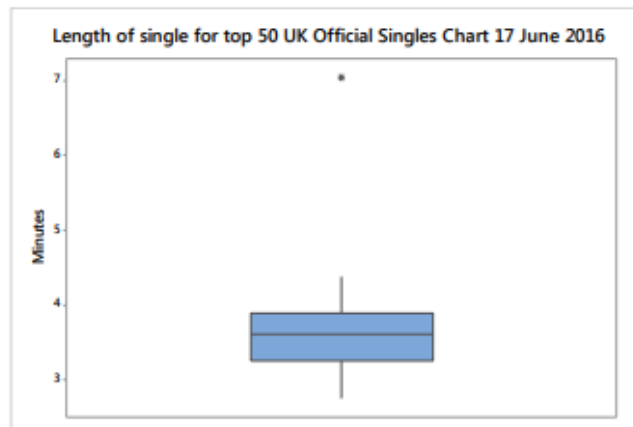
Length of singles for top 50 UK Official Chart singles, 17 June 2016									
2.5–(3.0)	3.0–(3.5)	3.5–(4.0)	4.0–(4.5)	4.5–(5.0)	5.0–(5.5)	5.5–(6.0)	6.0–(6.5)	6.5–(7.0)	7.0–(7.5)
3	17	22	7	0	0	0	0	0	1

He used these data to produce a graph of the distributions of the lengths of singles



State two corrections that Gareth needs to make to the histogram so that it accurately represents the data in the table. [2]

(c) Gareth also produced a box plot of the lengths of singles.



He sees that there is one obvious outlier.

- (i) What will happen to the mean if the outlier is removed?
- (ii) What will happen to the standard deviation if the outlier is removed? [2]

(d) Gareth decided to remove the outlier. He then produced a table of summary statistics.

- (i) Use the appropriate statistics from the table to show, by calculation, that the maximum value for the length of a single is not an outlier.

Summary statistics Length of single for top 50 UK Official Singles Chart (minutes)								
Length of single	N	Mean	Standard deviation	Minimum	Lower quartile	Median	Upper quartile	Maximum
	49	3.57	0.393	2.77	3.26	3.60	3.89	4.38

- (ii) State, with a reason, whether these statistics support the magazine's claim. [4]

(e) Gareth also calculated summary statistics for the lengths of 30 singles selected at random from his personal collection.

Summary statistics Length of single for Gareth's random sample of 30 singles (minutes)								
Length of single	N	Mean	Standard deviation	Minimum	Lower quartile	Median	Upper quartile	Maximum
	30	3.13	0.364	2.58	2.73	2.92	3.22	3.95

Compare and contrast the distribution of lengths of singles in Gareth's personal collection with the distribution in the top 50 UK Official Singles Chart. [3]

(WJEC SPECIMEN)

10. A researcher records the blood glucose levels of a group of females. The results, in mmol/litre are represented in the box plot in Fig. 2 below.

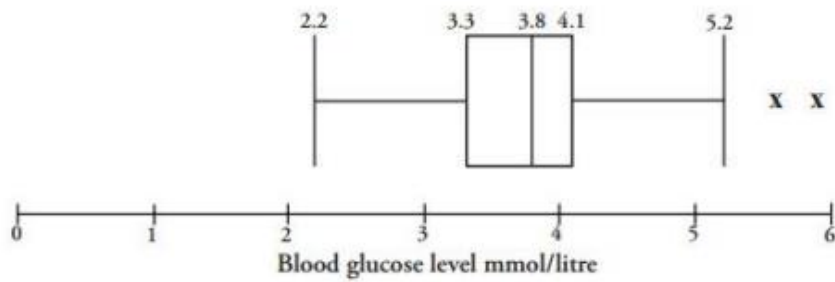


Fig.2

The researcher decides it is appropriate to clean the data by removing any outliers, which are indicated by **x** on the box plot.

If the outliers are removed from the original set of results, describe the effect this will have on

(i) the median, [1]

(ii) the standard deviation. [1]

(CEA SPECIMEN)

11.

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

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

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

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

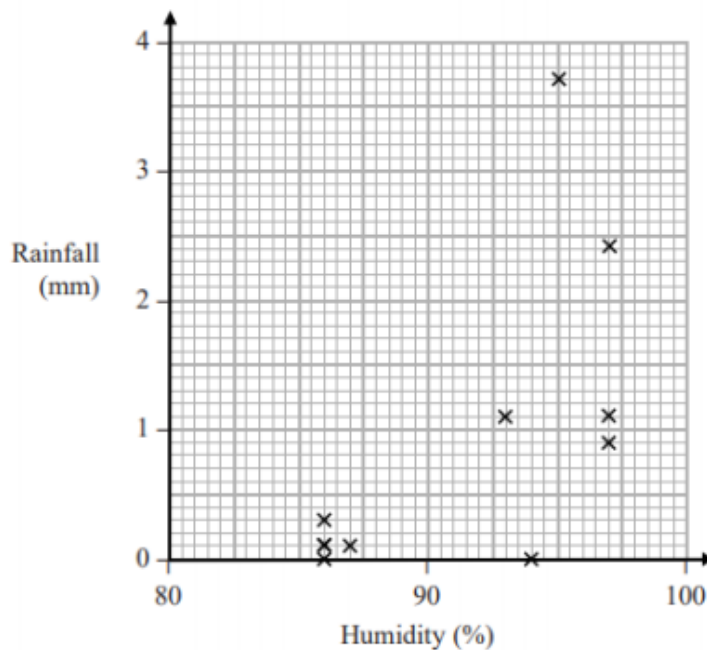
(ii) exclude

this day's reading.

(2)

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

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)