



eClassroom

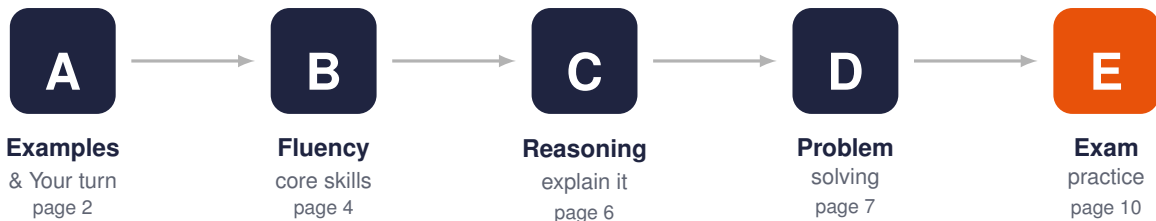
A-LEVEL MATHEMATICS • STATISTICS 1

Representations of Data

Worksheet

Pearson Edexcel International A Level (WST01) & A Level (9MA0) • Calculator and statistical tables allowed: show every step

Free Representations of Data worksheet for A-Level and International A-Level Maths, with full worked solutions



Tap a section to jump straight to it

By the end of this worksheet you can: use frequency density to read and draw histograms • find quartiles from stem-and-leaf diagrams • identify outliers and draw box plots • describe skewness • *Stretch:* histograms with no scale, skewness measures, ranges of values, and comparisons with named statistics.

Which questions are for me?

CORE The methods every A-level student must secure. Start here.

STRETCH Histograms without a scale, outliers you must name, box plots that must not change, and comparisons with numbers.

Every section has both, clearly labelled. Questions get harder as you go, and every skill in Sections A to D is needed in Section E.

Track your progress

Section	B	C	D	E
Core marks	23	10	13	21
Stretch marks	16	14	17	12
My score				

Full worked solutions

Exam questions by topic

Statistics 1 exam questions

Formula sheet builder

Name

Class

Date



Join students already acing their exams.

Book your FREE lesson at www.eclassroom.org.uk

A

Worked examples & Your turn

Read each example carefully. Cover it up, then answer the question beside it using the same method.

CORE

Secure the method

Worked example A1: histogram bars

The class 10–15 (frequency 30) is drawn 2 cm wide and 9 cm tall. Find the bar for 15–25 (frequency 40)

Frequency density: 10–15 has $\frac{30}{5} = 6$; 15–25 has $\frac{40}{10} = 4$

Width: 0.4 cm per unit, so 10×0.4 ; height: 1.5 cm per unit of density, so 4×1.5

4 cm wide, 6 cm tall

Your turn A1

The class 0–4 (frequency 12) is drawn 2 cm wide and 6 cm tall. Find the bar for 4–10 (frequency 27)

Worked example A2: quartiles from a stem-and-leaf diagram

11 values in order: 12, 15, 17, 21, 22, 25, 28, 30, 31, 36, 44. Find the quartiles.

$\frac{n}{4} = 2.75$: round up to the 3rd value; $\frac{n}{2} = 5.5$: the 6th;

$\frac{3n}{4} = 8.25$: the 9th

$Q_1 = 17$, $Q_2 = 25$, $Q_3 = 31$

Your turn A2

13 values in order: 34, 38, 41, 41, 45, 47, 50, 52, 55, 58, 60, 63, 71. Find the quartiles.

Worked example A3: outliers

$Q_1 = 20$ and $Q_3 = 32$. Outliers lie beyond $1.5(Q_3 - Q_1)$ from the quartiles. The data run from 5 to 53. Name any outlier.

$1.5 \times 12 = 18$: the limits are $20 - 18 = 2$ and $32 + 18 = 50$

5 is above 2; 53 is above 50

53 is an outlier

Your turn A3

$Q_1 = 46$ and $Q_3 = 58$, with the same rule. The data run from 25 to 74. Name any outlier.



STRETCH

Worked example A4: counting from a histogram

Classes 0–10, 10–20, 20–40 have frequencies 20, 35, 30. Estimate how many values lie between 12 and 26

Fraction of each class inside the range: $\frac{8}{10}(35) + \frac{6}{20}(30)$

37

Your turn A4

Classes 5–10, 10–15, 15–30 have frequencies 15, 24, 45. Estimate how many values lie between 8 and 20

Worked example A5: skewness from quartiles

$Q_1 = 14$, $Q_2 = 18$, $Q_3 = 26$. Describe the skewness

and evaluate $\frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$

$$Q_3 - Q_2 = 8 > Q_2 - Q_1 = 4$$

$$\frac{26 - 36 + 14}{12} = \frac{4}{12}$$

positive skew, 0.333

Your turn A5

$Q_1 = 30$, $Q_2 = 38$, $Q_3 = 43$. Describe the skewness and evaluate the same measure.

Worked example A6: comparing distributions

Group A: median 42, IQR 15. Group B: median 36, IQR 9. Compare the two groups.

A location and a spread, with values, in context.

A higher on average (42 vs 36)
 B more consistent (IQR 9 vs 15)

Your turn A6

Group C: median 25, IQR 6. Group D: median 31, IQR 11. Compare the two groups.

Worked example A7: a histogram with no scale

The bar for 20–30 has area 5 cm^2 and represents 40 people. All the bars have total area 18 cm^2 . How many people are shown?

Area is proportional to frequency: 1 cm^2 stands for $\frac{40}{5} = 8$ people.

$18 \times 8 = 144$

Your turn A7

The bar for 0–5 has area 3 cm^2 and represents 24 items. All the bars have total area 13.5 cm^2 . How many items are shown?



B**Fluency**

Each question is a designed sequence. From one part to the next, ask yourself: what changed, and what did it do to the answer?

CORE*Secure the method*

- B1** (a) The class 10–14 has frequency 36. Find its frequency density.
(b) The class 20–25 has frequency density 3.2. Find its frequency.
(c) A class has frequency 45 and frequency density 1.5. Find its width.
(d) The class 10–14 is drawn 1 cm wide and 4.5 cm tall. Find the width and height of the bar for 14–20, frequency 30 [4 marks]
- B2** The 21 values are 22, 25, 27, 28, 30, 31, 33, 33, 34, 36, 37, 38, 40, 41, 41, 43, 45, 46, 48, 52, 55
(a) Find Q_2 (b) Find Q_1
(c) Find Q_3 (d) Find the interquartile range. [4 marks]
- B3** (a) $Q_1 = 12$ and $Q_3 = 20$. Find the limits $Q_1 - 1.5(Q_3 - Q_1)$ and $Q_3 + 1.5(Q_3 - Q_1)$
(b) For the data in (a), decide whether 31 is an outlier.
(c) Outliers are values more than 2 standard deviations from the mean. The mean is 50 and the sd is 6. Is 64 an outlier?
(d) An outlier is removed from a data set. State, with a reason, which changes more: the mean or the median. [4 marks]
- B4** A box plot shows: lowest value 8, $Q_1 = 15$, median 19, $Q_3 = 27$, highest value 40
(a) Find the range. (b) Find the interquartile range.
(c) Describe the skewness, with a reason. (d) What percentage of the data are above 15? [4 marks]
- B5** Classes 0–5, 5–10, 10–20, 20–40 have frequencies 10, 25, 30, 15
(a) Estimate the number of values between 8 and 16
(b) Estimate the number of values above 32
(c) Find the frequency density of each class. [3 marks]



B6 (a) $Q_1 = 22$, $Q_2 = 26$, $Q_3 = 34$. Describe the skewness.

(b) Evaluate $\frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$ for the data in (a).

(c) The mean is 31, the mode is 25 and the sd is 8. Evaluate $\frac{\text{mean} - \text{mode}}{\text{sd}}$

(d) Using the median from (a), evaluate $\frac{3(\text{mean} - \text{median})}{\text{sd}}$ [4 marks]

STRETCH

B7 A histogram has bars 0–10, 10–15, 15–20, 20–30, 30–50 with heights 1.2, 3.6, 4.0, 1.8, 0.4 cm. The class 0–10 has frequency 18

(a) Find the frequency of 10–15

(b) Find the total frequency.

(c) Estimate the number between 12 and 24

(d) Estimate the median.

[4 marks]

B8 The 15 values are 11, 14, 15, 17, 18, 19, 20, 21, 22, 22, 24, 25, 27, 29, 41

(a) Find the quartiles.

(b) Using $1.5(Q_3 - Q_1)$, find the outlier limits and name any outlier.

(c) State where the upper whisker of the box plot ends.

(d) Find the mean with and without the outlier, and the median without it.

[4 marks]

B9 Two machines fill bags. Machine *P*: median 502 g, IQR 6 g, positive skew. Machine *Q*: median 498 g, IQR 3 g, symmetrical.

(a) Compare the averages.

(b) Compare the spreads.

(c) Which machine is more consistent?

(d) Compare the shapes.

[4 marks]

B10 20 values have mean 15 and sd 3. One value, 27, is found to be a recording error and is removed.

(a) Find $\sum x$ and $\sum x^2$ for the 20 values.

(b) Find the new mean.

(c) Find the new standard deviation.

(d) Explain why removing the value changes the sd so much.

[4 marks]



C**Reasoning**

Show your working and explain your thinking. A correct answer without a reason does not score full marks.

CORE*Secure the method*

- C1** Classes 0–10 and 10–30 have frequencies 20 and 30. Amy draws the 10–30 bar taller than the 0–10 bar. Explain her error, and say which bar should be taller. [2 marks]
- C2** In a back-to-back stem-and-leaf diagram, the left side reads 7 5 2 | 3. Ben writes these values as 73, 53, 23. Explain his error, and write the values in order. [2 marks]
- C3** The rule is: an outlier is greater than $Q_3 + 1.0(Q_3 - Q_1)$. With $Q_1 = 40$ and $Q_3 = 52$, Cara uses $1.5(Q_3 - Q_1)$ and says 66 is not an outlier. Explain her error, and decide correctly. [3 marks]
- C4** $Q_1 = 20$, $Q_3 = 30$ and the largest values are 41, 44 and 52. Dev draws the upper whisker to 45, the outlier limit. Explain his error, and say where the whisker should end and what else is drawn. [3 marks]

STRETCH

- C5** Eli compares two classes' test marks by writing "class A did better and was more spread out". Explain why this would not score full marks, and rewrite it for medians 61 and 55 and IQRs 18 and 12 [3 marks]
- C6** $Q_1 = 40$, $Q_2 = 48$, $Q_3 = 52$. Fay says the data are positively skewed because $Q_2 - Q_1 > Q_3 - Q_2$. Explain her error, and describe the skewness correctly. [3 marks]
- C7** Classes 0–10 and 10–30 have frequencies 20 and 30. To estimate the number below 15, Gus adds $20 + 30 = 50$. Explain his error, and give the correct estimate. [3 marks]
- C8** (a) Explain why, for any data set, $-1 \leq \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1} \leq 1$ when $Q_3 > Q_1$ (3)
- (b) Show that the measure is 0 when Q_2 is halfway between Q_1 and Q_3 , and explain what this tells you. (2)
- [5 marks]



D

Problem solving

You decide the method. Read the whole question first, and show every step.

CORE

Secure the method

- D1** A company tests 120 batteries and records the lifetime, in hours, of each one. The histogram in Figure 1 shows the results.

(a) Find the number of batteries with a lifetime of 30 to 35 hours, and the number with a lifetime of 40 to 50 hours.

(2)

(b) Estimate the number of batteries with a lifetime between 32 hours and 46 hours.

(1)

Frequency density

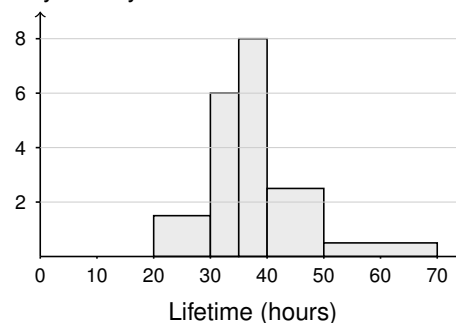


Figure 1

[3 marks]

- D2** A fisheries officer measures the length, in cm, of each of 23 fish caught in a river. The lengths are shown in the stem-and-leaf diagram in Figure 2. An outlier is a length greater than $Q_3 + 1.5(Q_3 - Q_1)$ or less than $Q_1 - 1.5(Q_3 - Q_1)$

(a) Find the median and the quartiles of these lengths.

(2)

Stem	Leaf	
1	8	(1)
2	4 6 7 9	(4)
3	0 1 1 2 3 4 4 5 6 7 8 9	(12)
4	0 1 2 4	(4)
5	6 8	(2)

Key: 3 | 2 means 32 cm

Figure 2

(b) Show that there are exactly two outliers in these lengths, stating their values. You must show your working clearly.

(2)

[4 marks]

- D3** A courier company records the masses, in kg, of 120 parcels. The box plot in Figure 3 summarises the masses.

(a) Use linear interpolation to estimate the number of these parcels with a mass of more than 2.2 kg.

(2)

(b) The company classes a parcel as an outlier if its mass is greater than $Q_3 + 1.5(Q_3 - Q_1)$. Find this limit, and explain whether any of these parcels is an outlier.

(1)

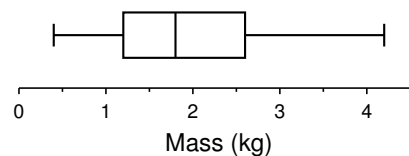


Figure 3

[3 marks]



D4 The table shows the marks of 11 students in a history test.

Student	A	B	C	D	E	F	G	H	I	J	K
Mark	38	41	44	45	47	48	50	52	55	57	79

An outlier is defined as a value greater than $Q_3 + 1.5(Q_3 - Q_1)$ or less than $Q_1 - 1.5(Q_3 - Q_1)$. The teacher says that the mark of student K is an outlier. Show that the teacher is correct. You must find and state the quartiles you use. [3 marks]

STRETCH

D5 The histogram in Figure 4 shows the times, in minutes, that 100 customers spent in a museum. No vertical scale is shown. The museum knows that 24 of these customers spent between 10 and 20 minutes in the museum.

(a) Estimate the number of customers who spent between 15 and 35 minutes in the museum.

(2)

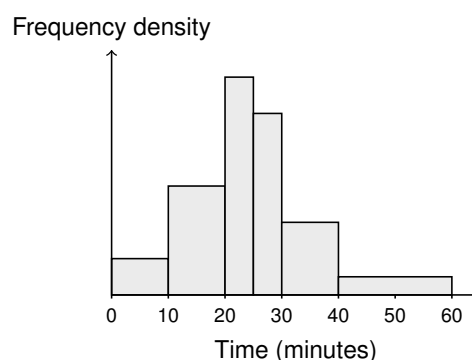


Figure 4

(b) Use linear interpolation to estimate the median time spent in the museum by these 100 customers.

(2)

[4 marks]

D6 The back-to-back stem-and-leaf diagram in Figure 5 shows the times, in minutes, that 15 customers spent in shop A and 15 customers spent in shop B. For shop B, $Q_1 = 36$, $Q_2 = 44$ and $Q_3 = 51$

Totals	Shop A		Shop B	Totals
(0)		2	4 9	(2)
(3)	8 5 1	3	3 6 8 9	(4)
(7)	9 7 6 5 3 2 0	4	1 4 4 6 8	(5)
(4)	8 5 2 0	5	1 3 7	(3)
(1)	1	6		(0)
(0)		7	0	(1)

Key: 5 | 4 | 1 means 45 minutes for shop A and 41 minutes for shop B

Figure 5

(a) Find the lower quartile, the median and the upper quartile for shop A

(2)

(b) Compare the times spent in the two shops. You should state the values of the statistics you use and interpret them in context.

(2)

[4 marks]



D7 The 11 values below have been put in order, and are used to draw a box plot.

12, 15, 18, 20, 23, 25, 27, 30, 33, 35, 41

Two further values, a and $2a$, are then added to the data. The box plot for all 13 values is exactly the same as the box plot for the original 11 values, and there are still no outliers.

Find the range of possible values of a . You must show your working clearly and give your answer as an inequality. [4 marks]

D8 The table shows the distances, x km, travelled to work by 140 employees of a company.

Distance (km)	0–	5–	10–	20–	30–50
Frequency	18	42	50	24	6

A histogram is drawn to represent these data. The height of the tallest bar is 12.6 cm.

(a) Find the height of the second tallest bar.

(2)

(b) Given that $Q_1 = 7.02$, $Q_2 = 12.0$ and $Q_3 = 19.0$, evaluate $\frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$ and describe the skewness of the distances.

(2)

(c) Comment on whether a normal distribution would be a suitable model for these distances.

(1)

[5 marks]



E**Exam practice**

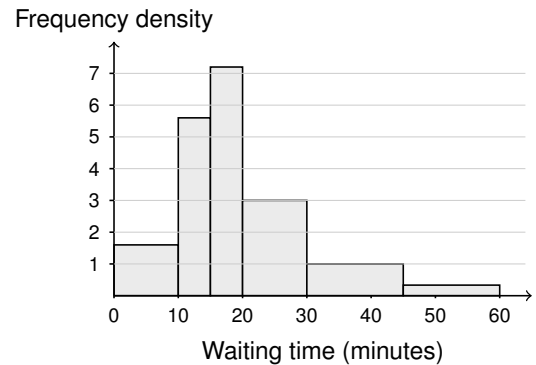
Full-size exam questions in the style of WST01. Write your answers in the spaces, and show every step.

CORE*Exam questions for everyone*

- E1** A clinic records the time, in minutes, that each patient waits before being seen. The histogram in Figure 6 shows the waiting times of the 130 patients who visited the clinic on one day.

(a) Use the histogram to complete the table below.

Time	0–	10–	15–	20–	30–	45–60
Freq.	16			30		5

(2)**Figure 6**

(b) On a second copy of this histogram, the bar for the class 10–15 is 1 cm wide and 5.6 cm tall. Find the width and the height of the bar for the class 30–45

(2)

(c) Use linear interpolation to estimate the median and the interquartile range of the waiting times of these 130 patients.

(3)

(d) Use your answers to part (c) to describe the skewness of the waiting times of these patients, giving a reason for your answer.

(1)

(e) Two of these 130 patients are chosen at random. Estimate the probability that both of the patients chosen waited for more than 30 minutes before being seen.

(2)**(Total for Question E1 is 10 marks)**

E2 A teacher asks each of the 19 students in group *A* how many minutes they spent looking at a screen on one particular school evening. The results are shown in the stem-and-leaf diagram in Figure 7. An outlier is a value greater than $Q_3 + 1.5(Q_3 - Q_1)$ or less than $Q_1 - 1.5(Q_3 - Q_1)$

(a) Find the median and the quartiles of these times.

(2)

Stem	Leaf	
4	5	(1)
5	2 8	(2)
6	0 3 5 8	(4)
7	0 2 4 5 8	(5)
8	0 3 5 8	(4)
9	2 5	(2)
10		(0)
11		(0)
12		(0)
13	8	(1)

Key: 7 | 4 means 74 minutes

Figure 7

(b) Show that there is exactly one outlier in the screen times of group *A*, stating its value. You must show your working clearly.

(2)

The grid in Figure 8 shows a box plot of the screen times of a second group, group *B*

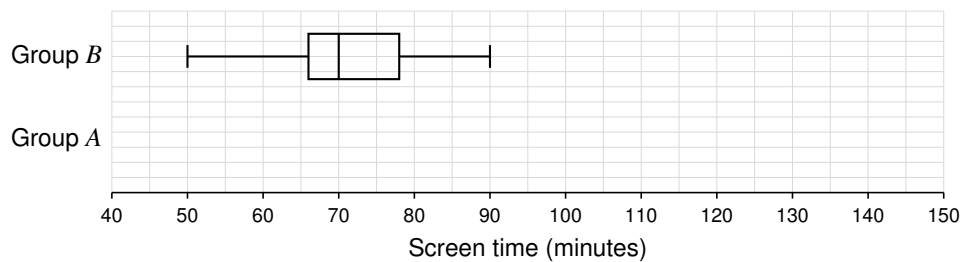


Figure 8

(c) On the grid in Figure 8, draw a box plot for the screen times of group *A*

(3)

(d) Compare the screen times of the students in the two groups. You should state the values of any statistics you use and interpret them in context.

(2)

(e) Describe the skewness of the screen times of each of the two groups, giving a reason for your answer in each case.

(2)

(Total for Question E2 is 11 marks)



STRETCH

- E3** Two couriers, *A* and *B*, deliver parcels in a city. The box plots in Figure 9 summarise the delivery times, in minutes, of a random sample of 160 deliveries made by courier *A* and a random sample of 160 deliveries made by courier *B* during the same month. For each courier, an outlier is defined as a value greater than $Q_3 + 1.5(Q_3 - Q_1)$ or less than $Q_1 - 1.5(Q_3 - Q_1)$

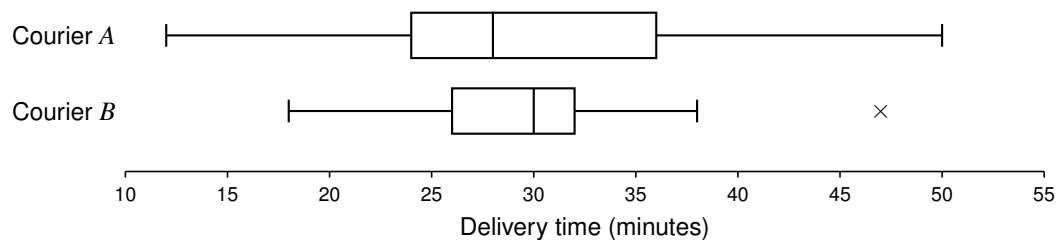


Figure 9

- (a) Use linear interpolation to estimate the number of deliveries made by courier *A* that took less than 30 minutes.

(2)

- (b) Show that the delivery time marked with a cross for courier *B* is an outlier, and that none of the other delivery times for courier *B* is an outlier. You must show your working clearly.

(2)

- (c) The manager of the company says that courier *B* is more reliable than courier *A*. Use statistics from the box plots in Figure 9 to comment on the manager's claim, stating the values of any statistics you use.

(2)

- (d) Evaluate $\frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$ for each courier, and describe the skewness of each courier's delivery times.

(3)

- (e) The manager finds that the outlier for courier *B* was a recording error, and that the correct delivery time was 38 minutes. State, giving a reason, what change this correction makes to the box plot for courier *B*

(2)

- (f) State, giving a reason based on your answer to part (d), whether a normal distribution would be a suitable model for the delivery times of courier *A*

(1)

(Total for Question E3 is 12 marks)



Want more exam practice? The **Exam Questions by Topic** booklets have 10 more full exam questions on Statistics 1 Chapter 3, Representations of Data, with a full mark scheme.

[Open the booklet](#)

Finished? Mark your work with the **full worked solutions**, fill in your scores on the front page, then practise real exam questions on the **Statistics 1 exam questions page**.



Join students already acing their exams.

Book your FREE lesson at www.eclassroom.org.uk