Name	Date started	Target end date

REVISE
.wales

3.21 — Estimating mean & median from grouped data

Spec 4.2.9, 4.2.11, 4.2.12, 4.2.13, 4.2.14 — Unit 3 (calculator allowed)

When data is grouped into classes, exact values are lost. Estimate the mean using class midpoints, identify the modal class, and find the median (or any percentile) by linear interpolation or from cumulative frequency. Sourced from legacy WJEC GCSE Mathematics and Mathematics-Numeracy papers, organised for revision under the 2025 spec.

2025 SPECIFICATION

Estimated time for entire question pack: ~39 minutes

*Derived from the GCSE Higher pace of **~1.5 min/mark** (26 marks across 9 questions).*

*You are advised to **not** attempt to complete all of this in one sitting.*

ABOUT THIS QUESTION PACK

This is a **focused single-topic practice pack**, not a single mock paper. Questions are organised against the 2025 specification. Questions are ordered chronologically by sitting, with custom-written and SAM questions at the end.

INSTRUCTIONS

Use black ink or black ball-point pen. Show all working — method marks are awarded for clear setup.

A calculator is allowed on every question in this pack (Unit 3 is the calculator-allowed paper).

Estimating mean & median from grouped data — what the new spec asks

WJEC GCSE Mathematics (first teaching 2025) · Unit 3: calculator-allowed.

Modal class & class midpoints 4.2.9

- Identify the modal class from a grouped frequency table.
- Use the class midpoint $\frac{a+b}{2}$ to represent values in a class.
- Recognise that values in a grouped table cannot recover the exact original data.

Estimating the mean from grouped data 4.2.11

- Calculate $\bar{x} \approx \frac{\sum f x_{\text{mid}}}{\sum f}$ for grouped data.
- Show working in extra columns of the frequency table.
- State the answer explicitly as an estimate.

Median & quartiles by linear interpolation 4.2.12

- Identify the median class via cumulative frequency.
- Apply linear interpolation: $L + \frac{n/2 - F}{f} w$.
- Apply the same technique to estimate quartiles and percentiles.

Cumulative frequency diagrams 4.2.13

- Plot cumulative frequency against the *upper class boundary*.
- Read off the median and quartiles from the curve.
- Use the curve to estimate the proportion of data above or below a given value.

Comparing grouped data sets 4.2.14

- Compare distributions by quoting an estimate of average and a measure of spread.
- Refer back to the context, not just numerical differences.
- Recognise the limitations of estimates from grouped data.

Crib sheet »

Estimating mean & median from grouped data in one page

Quick-reference notes — revisit before each question. Don't use during the questions.

Why 'estimate'?

Grouped data only tells us each value falls in a class — not the exact value.

So the mean we calculate is an *estimate*, never the true mean of the original raw data.

Class midpoints

For a class $a \le x < b$, the midpoint is $\frac{a+b}{2}$.

The midpoint represents every value in that class for the purposes of estimating the mean.

Watch out for inclusive class boundaries (e.g. 10–19 has midpoint 14.5 if continuous).

Estimating the mean

$$\bar{x} \approx \frac{\sum f x_{\text{mid}}}{\sum f}$$

For each class: multiply midpoint by frequency; sum the products; divide by total frequency.

Lay the working out in extra columns of the frequency table.

Modal class

The class with the highest frequency.

You can't pick a single 'modal value' from grouped data — only the modal *class*.

Median class

Locate the $\frac{n+1}{2}$-th data point (or $\frac{n}{2}$ for large n) using the cumulative frequency.

The class containing this position is the median class.

Linear interpolation for the median

$$\text{median} \approx L + \frac{\frac{n}{2} - F}{f} w$$

L = lower bound of median class, F = cumulative freq *before* the median class, f = freq of median class, w = class width.

Cumulative frequency

Running total of frequencies from the start of the table.

Plot at the *upper class boundary*; join with a smooth curve to read off the median and quartiles.

Comparing distributions

Compare two grouped data sets by quoting an estimate of the mean (or median) *and* a measure of spread (range or IQR).

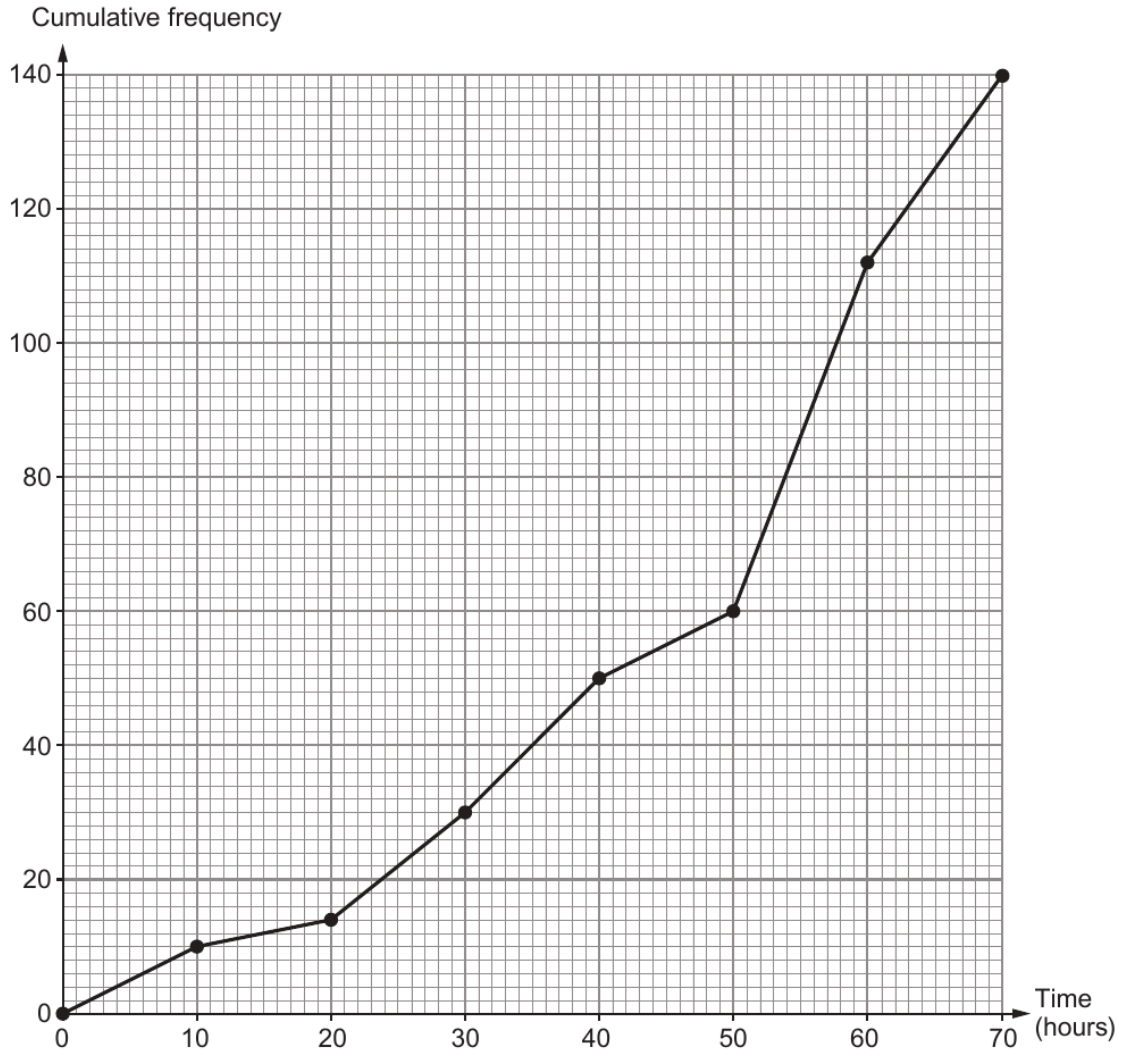
Always interpret in context, not just numerically.

Common traps

- Using class limits instead of midpoints in $\sum fx$.
- Forgetting the result is an *estimate* — mark schemes want that word.
- Plotting cumulative frequency at the midpoint instead of the upper bound.
- Picking a single 'mode' from grouped data instead of the modal class.

Examiner only

4. *(a)* 140 girls were asked how long they spent revising for their GCSE examinations. The cumulative frequency diagram shows the results.



(i) Estimate the median time the girls spent revising.
Circle your answer. [1]

35 hours 40 hours 48 hours 52 hours 70 hours

(ii) Calculate the number of girls who spent between 40 and 50 hours revising.
Circle your answer. [1]

0 girls 5 girls 10 girls 15 girls 20 girls

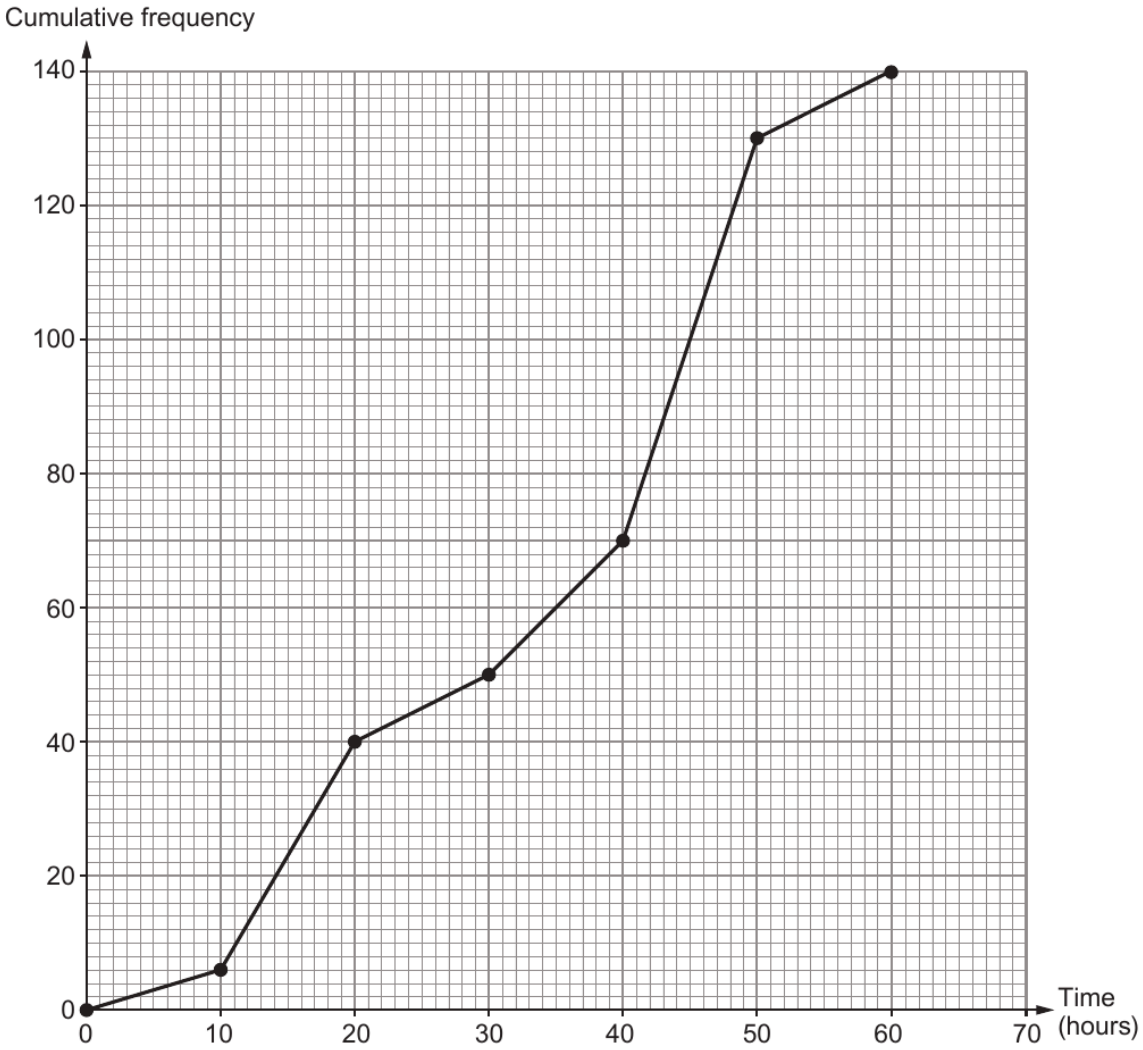
Examiner only

(iii) Circle either TRUE or FALSE for each of the following statements. [2]

25 girls spent between 30 and 50 hours revising.	TRUE	FALSE
No girls spent more than 80 hours revising.	TRUE	FALSE
The modal group is between 50 and 60 hours spent revising.	TRUE	FALSE
20 girls spent more than 60 hours revising.	TRUE	FALSE

Examiner only

(b) 140 boys were asked how long they spent revising for their GCSE examinations. The cumulative frequency diagram below shows the results.



Examiner only

Trefor makes two statements.

1. The boys' interquartile range is greater than the girls' interquartile range.

2. On average, boys spent more time revising.

Are both Trefor's statements correct?
Show calculations and give reasons to support your answers. [4]

Statement 1: .....

.....

.....

.....

.....

.....

Statement 2: .....

.....

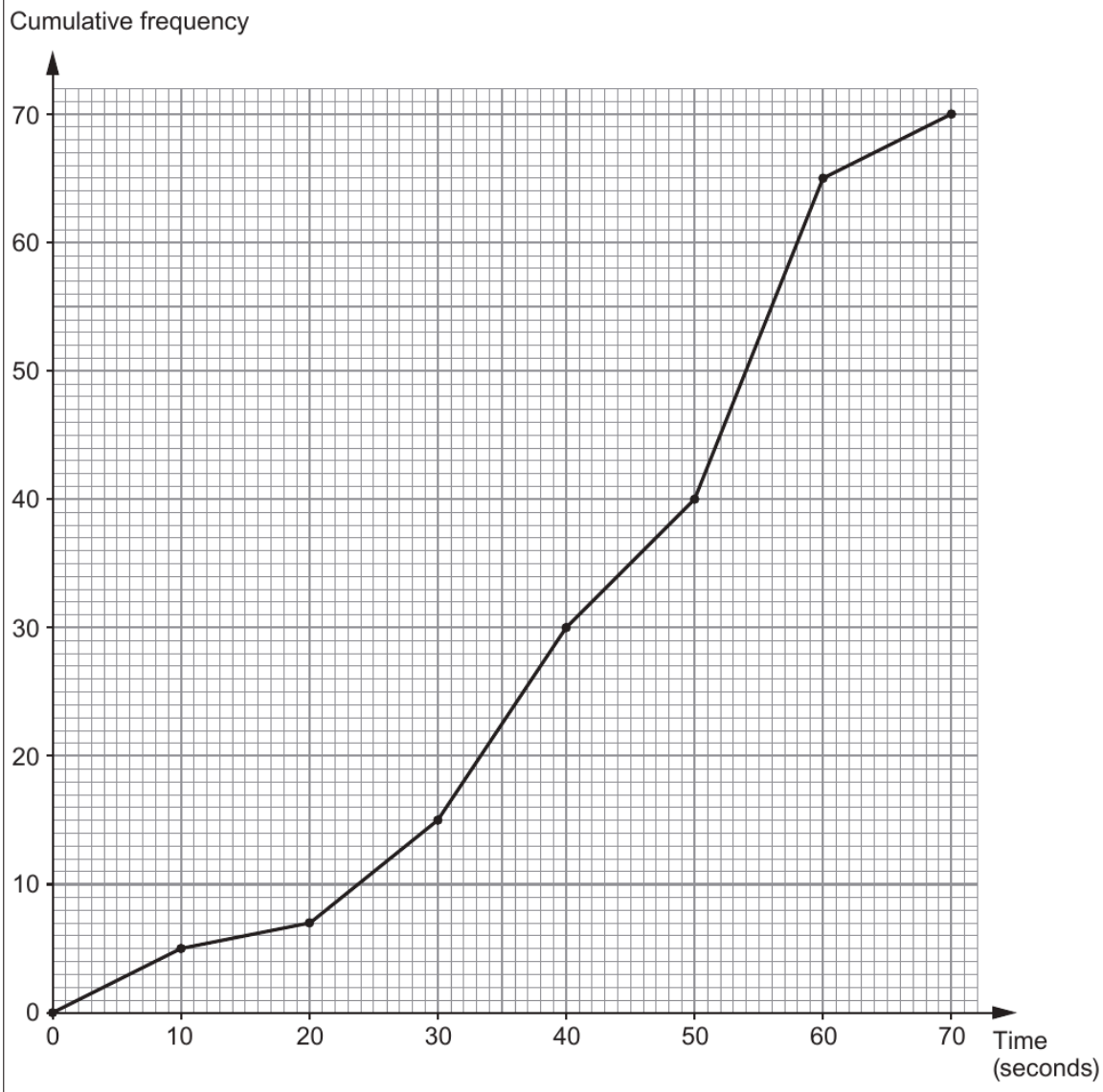
.....

.....

.....

Examiner only

5. *Cambria Airlines* has planes that can carry up to 70 passengers.
For safety, the crew practise the emergency exit procedures with a group of 70 passengers.
Every 10 seconds the safety officer records the total number of passengers who have left the plane.
He has displayed the results in the cumulative frequency diagram shown below.



(a) Estimate the median time taken by the passengers to leave the plane. [1]

..... seconds

Examiner only

(b) How many passengers took more than 50 seconds to leave the plane?
Circle your answer. [1]

10 20 30 40 50

(c) *Cambria Airlines* has a policy that states the following.

'In the event of an emergency exit procedure, at least 90% of the 70 passengers must have left the plane within 1 minute.'

Did the practice emergency exit procedure meet the requirements of the airline's policy?
You must show all your working. [4]

Examiner only

6. Dr Khan and her daughter Faryl have different opinions about the mean temperature in their hallway.

Dr Khan and Faryl recorded the temperature in the hallway at 4 p.m. each day during the 30 days of April.

(a) In her note pad, Dr Khan summarised the temperatures in a grouped frequency table.

Unfortunately, Dr Khan has torn the page containing the table from her note pad and has lost some of the original data.

Temperature, t (°C)	Number of days
$20 \leqslant t < 21$	4
$21 \leqslant t < 22$	8
$22 \leqslant t < 23$	8
$23 \leqslant t < 24$	

Calculate an estimate of the mean temperature at 4 p.m. for these 30 days in Dr Khan's hallway. [5]

Estimate of the mean temperature at 4 p.m. for April in the hallway is ..... °C

Examiner only

(b) What assumption have you made in calculating an estimate of the mean temperature at 4 p.m. for April in Dr Khan's hallway? [1]

(c) Faryl recorded the same temperatures as her mother at 4 p.m. each day during April. She found that the **actual** mean temperature in the hallway during April was **lower** than the correctly calculated estimate of the mean.

Explain how this can be true. [1]

Examiner only

2. *(a)* A biased coin is thrown 100 times.
The number of heads thrown is recorded after 20 throws, 40 throws, 60 throws, 80 throws and 100 throws.

Some of the results are recorded in the relative frequency table below.

Complete the table. [2]

Number of throws	20	40	60	80	100
Number of heads	11	18	24	30	
Relative frequency	0·55	0·45		0·375	0·37

.....

.....

.....

.....

Examiner only

(b) *In this part of the question, you will be assessed on the quality of your organisation, communication and accuracy in writing.*

5 7 8 11 14 17 17 19 26 28

The sum of the ten numbers shown above is 152.
The numbers are displayed in the grouped frequency table shown below.

Number	0 - 9	10 - 19	20 - 29
Frequency	3	5	2

Consider the estimated mean calculated from the table and the actual mean of the ten numbers.
Calculate the difference between these two values.
You must show all your working. [5 + 2 OCW]

Examiner only

4. *(a)* Which one of the following options describes $2x + 5y$?
Circle your answer. [1]

an equation a formula an expression

an inequality none of these

(b) Which one of the following options describes $3x - 2 = 7$?
Circle your answer. [1]

an equation a formula an expression

an inequality none of these

5. Data for different values of t are shown in the table below.

t	Frequency
$0 \leqslant t < 5$	8
$5 \leqslant t < 10$	0
$10 \leqslant t < 15$	7
$15 \leqslant t < 20$	5

Calculate an estimate for the mean value of t. [4]

Examiner only

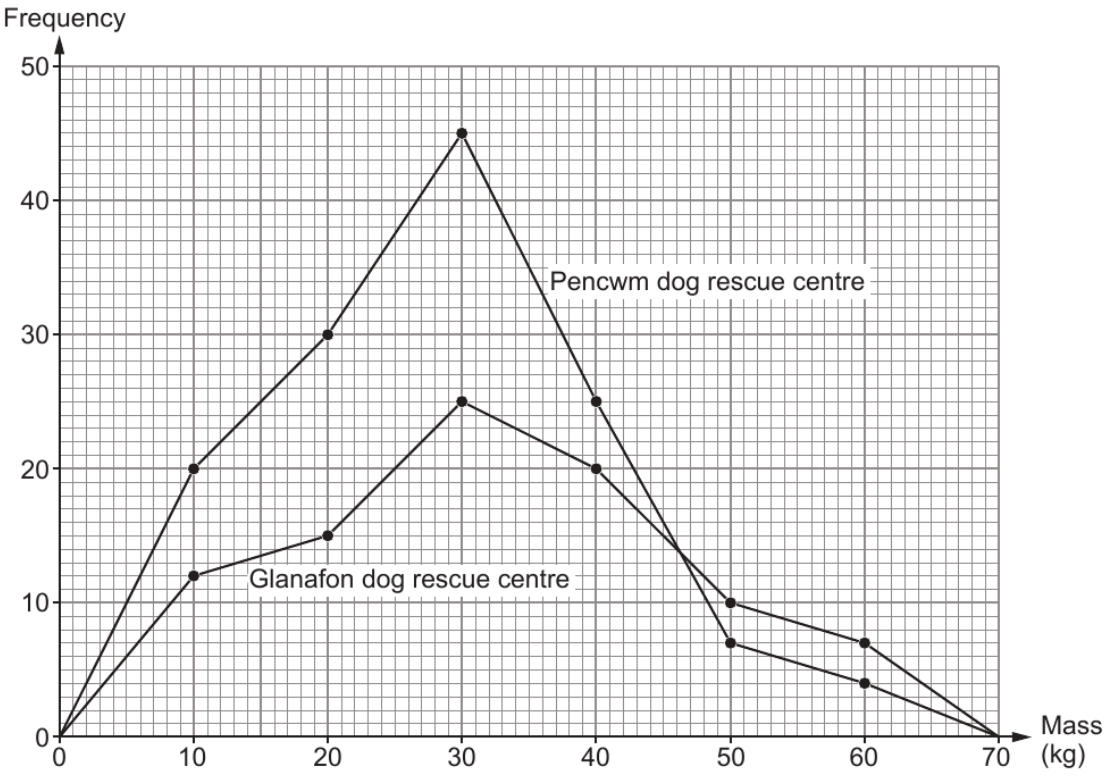
2. Glanafon and Pencwm dog rescue centres take in unwanted dogs.

The mass of each dog in the two dog rescue centres was recorded.

Groups of width 10 kg were used:

$5\,\text{kg} \leqslant \text{mass} < 15\,\text{kg}, \quad 15\,\text{kg} \leqslant \text{mass} < 25\,\text{kg}, \quad \ldots, \quad 55\,\text{kg} \leqslant \text{mass} < 65\,\text{kg}$

The results are shown in the frequency polygons below.



(a) Doreen, Rory and Muzhir look at these frequency polygons.

(i) Doreen says,

"The modal group of the masses of dogs in each dog rescue centre is the same."

Is Doreen correct?

Yes ☐ No ☐ Can't tell ☐

You must give a reason for your answer. [1]

.....

.....

.....

Examiner only

(ii) Rory says,
"28 of the dogs in Pencwm each have a mass of 18 kg."

Is Rory correct?

Yes ☐ No ☐ Can't tell ☐

You must give a reason for your answer. [1]

.....

.....

.....

(iii) Muzhir says,
"There is a higher proportion of dogs that are heavier than 35 kg in Glanafon than in Pencwm."

Without doing any calculations, decide if Muzhir is correct.

Correct ☐ Incorrect ☐ Can't tell ☐

You must give a reason for your answer. [1]

.....

.....

.....

(b) The estimate of the mean mass of the dogs in Glanafon was 32·5 kg.
How much less was the estimate of the mean mass of the dogs in Pencwm?
You must show all your working. [5]

.....

.....

.....

.....

.....

.....

.....

Estimate of the mean mass of the dogs in Pencwm is ..... kg less than in Glanafon.

Examiner only

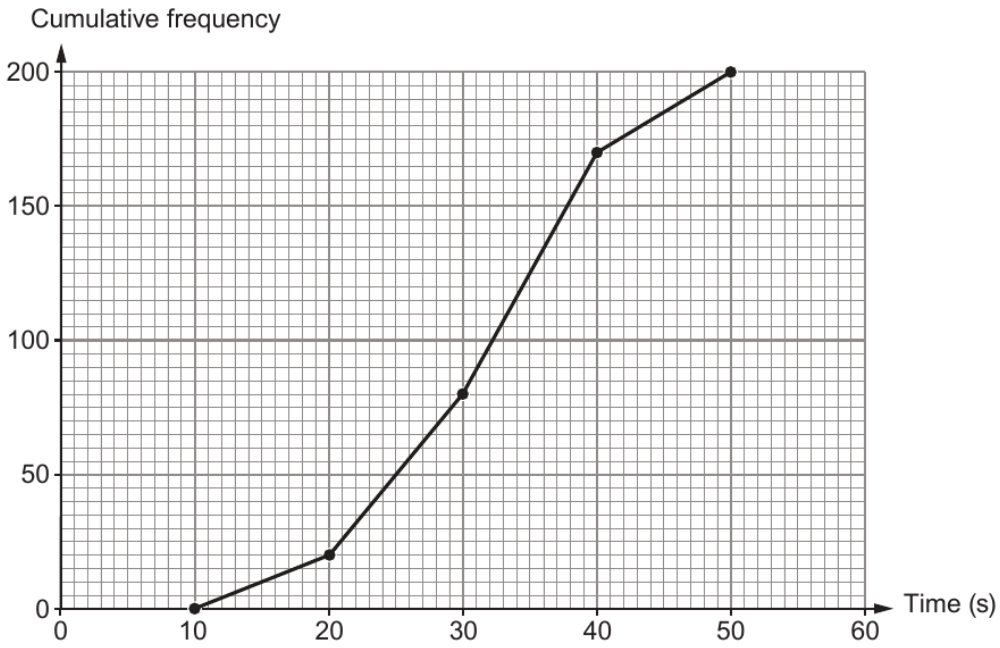
9. The grouped frequency table below shows data for 32 values of t.
One of the frequencies in the table is represented by a.

t	Frequency
$20 \leqslant t < 30$	2
$30 \leqslant t < 40$	8
$40 \leqslant t < 50$	4
$50 \leqslant t < 60$	a
$60 \leqslant t < 70$	3
$70 \leqslant t < 80$	5

Find the estimated mean value of t for the 32 values. [5]

Examiner only

4. (a) On 1st June last year, 200 customers used cash to pay at Shop Lil.
The cumulative frequency diagram represents the time each of these 200 customers waited to be given change at the checkout.



(i) How many of these customers waited between 30 and 50 seconds for their change? [2]

.....

.....

(ii) Use the graph to estimate the median time these 200 customers waited for their change.
Circle your answer. [1]

24 seconds 32 seconds 38 seconds 80 seconds 100 seconds

(iii) Calculate the fraction of these 200 customers who waited 40 seconds or longer for their change.
Give your answer in its simplest form. [2]

.....

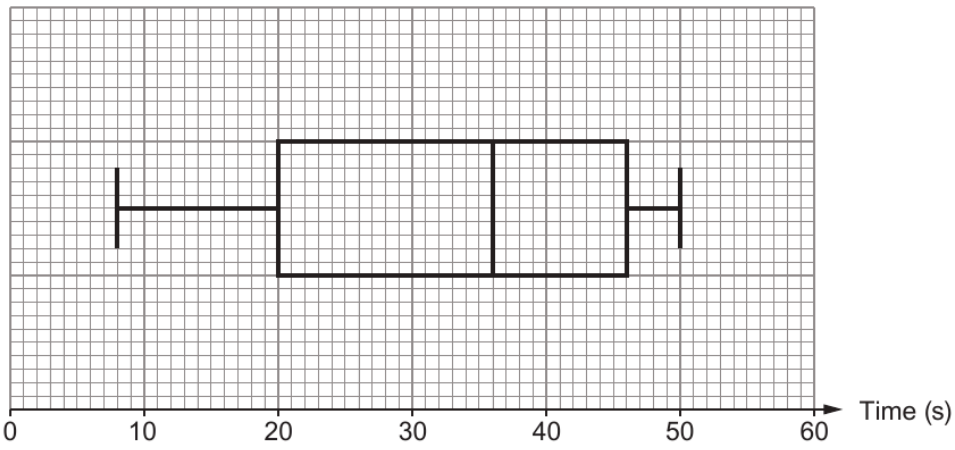
.....

.....

.....

Examiner only

(b) On 1st June this year, the manager at Shop Lil drew a box-and-whisker plot of the times 200 customers waited for their change at the checkout.



Based on the results of these 200 customers, the manager made the following statements. Complete the statements.

(i) "On 1st June this year, 50% of our customers were given their change in ..... seconds or less." [1]

(ii) "On 1st June this year, the interquartile range of the times taken to give customers their change was ..... seconds." [2]

.....

.....

(c) Consider the 50 customers waiting the **longest** times to get their change on 1st June last year and this year.
Has the speed of giving change at the checkout improved since last year?

Yes ☐ No ☐

You must give a reason for your answer. [1]

.....

.....

.....

.....

3310U501
09

Examiner only

2. Eleri is investigating whether people working in offices are happy with the processing speed of their office computer.
She considers the data she needs to collect.

Eleri includes the following two questions in her questionnaire.
For each question, write down one set of possible groups she could use as answer options.
[2]

Question 1: How many days per month do you work in your office?

Possible groups for answer:

.....

.....

.....

Question 2: How happy are you with the processing speed of your office computer?

Possible groups for answer:

.....

.....

.....

3310U501
05