IB MATHEMATICS / APPLICATIONS AND INTERPRETATION SL

# Representing data

Frequency tables, histograms, cumulative frequency graphs and box plots

| | |
|---|---|
| **Learning goal** | Choose, draw and interpret statistical displays to compare data in context. |
| **Syllabus link** | AI SL 4.2 · Statistics and probability |
| **Structure** | 15 short responses, 15 extended responses and 20 Paper 1-style questions. |
| **Technology** | Use a graphic display calculator where helpful. Have it available for Section 3. |

**Instructions** Work in your notebook and use the grids provided. Give units. When reading a graph, show your guide lines or record the cumulative frequency you used.

## SECTION 01 Short response Q1–15

### 1. From a list to a frequency table

Twelve students report how many buses they take on their journey to school.

$$0,\ 1,\ 2,\ 1,\ 0,\ 3,\ 1,\ 2,\ 2,\ 0,\ 1,\ 1$$

Construct a frequency table for the number of buses, including all values from 0 to 3.

### 2. On a class boundary

A parcel depot groups masses, $m$ kg, in equal classes. Two consecutive classes are $5 \le m < 10$ and $10 \le m < 15$. Write down the next class and state which class contains a parcel weighing exactly 15 kg.

### 3. Combining classes

The heights of 25 seedlings are recorded after two weeks.

| Height $h$ (cm) | $0 < h \le 4$ | $4 < h \le 8$ | $8 < h \le 12$ | $12 < h \le 16$ |
|---|---|---|---|---|
| Frequency | 3 | 8 | 9 | 5 |

Find the percentage of seedlings with heights greater than 4 cm and at most 12 cm.

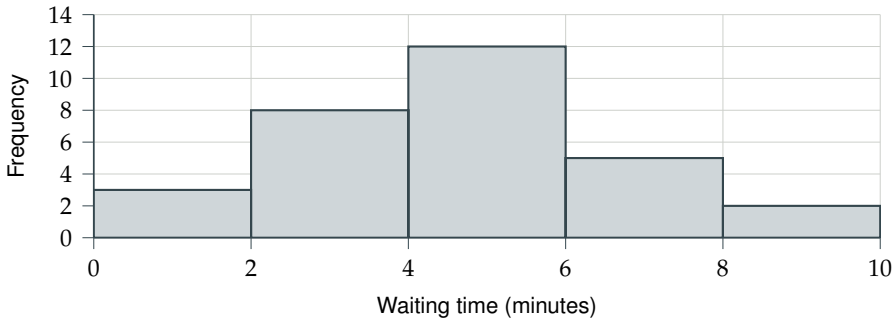
### 4. A missing frequency

An art teacher sorts 36 sketches into four groups. The frequencies are 7, $k$, $2k$ and 5. Find how many sketches are in the third group.

SECTION 01 / Short response Questions 5–9

### 5. Reading a frequency histogram

A cinema records waiting times at a ticket kiosk. Each class includes its lower endpoint and excludes its upper endpoint.



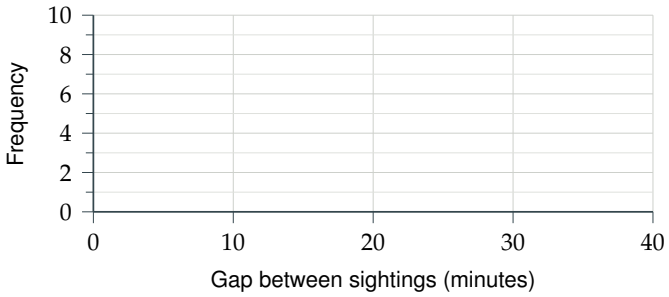

**(a)** State how many customers waited at least 6 minutes.

**(b)** Write down the waiting-time interval containing the most customers.

### 6. Drawing a histogram

A camera trap records the time between successive sightings of deer. Draw a frequency histogram on the grid using the table.

| Gap $t$ (min) | Frequency |
|---|---|
| $0 < t \leq 10$ | 2 |
| $10 < t \leq 20$ | 6 |
| $20 < t \leq 30$ | 8 |
| $30 < t \leq 40$ | 4 |




### 7. Recovering a class frequency

Of 50 orders, 18 were ready in at most 8 minutes and 41 were ready in at most 12 minutes. How many were ready in more than 8 minutes but at most 12 minutes?

### 8. Which point belongs on the graph?

A 5-minute class $0 < t \leq 5$ contains 7 observations. The next class $5 < t \leq 10$ contains 11. Which point belongs on the cumulative frequency graph at the end of the second class? Justify your choice.

$(7.5, 18)$ $(10, 11)$ $(10, 18)$

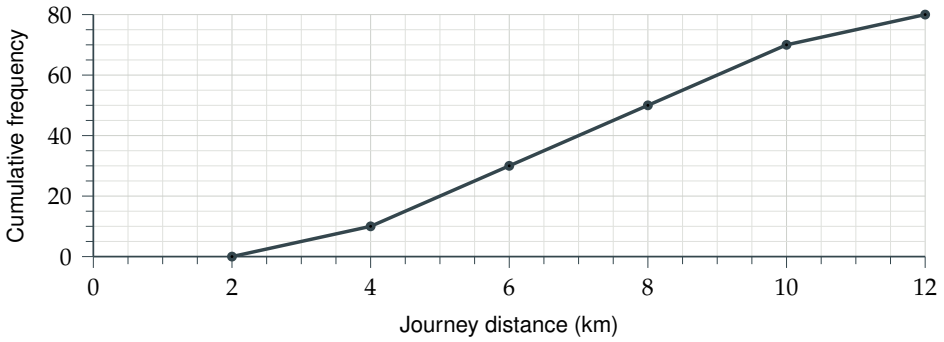
### 9. Interpreting a percentile

For 160 bicycle repairs, the 85th percentile of the repair time is 48 hours. Estimate how many repairs took longer than 48 hours.

SECTION 01 / Short response Questions 10–15

### 10. Quartiles from a graph

The graph shows the distances of 80 cycle journeys. Estimate the lower quartile, median and upper quartile. Hence find the interquartile range.




### 11. The longest journeys

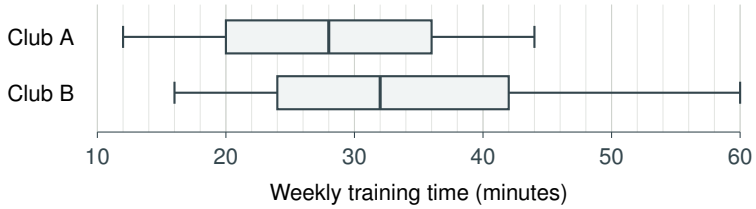
Use the graph in Question 10 to estimate the distance exceeded by only the longest 10% of journeys.

### 12. Observed extremes

Use the two ends of the graph in Question 10 to estimate the range. The shortest recorded journey was 2.3 km and the longest was 11.8 km. How does the exact range compare with your estimate?

### 13. Comparing two climbing clubs

The box plots show weekly training times for members of two climbing clubs. Compare the clubs using both the median and the interquartile range.




### 14. Judging symmetry

Could a normal distribution be a plausible model for Club A's training times? Refer to the box and whiskers in Question 13.

### 15. Where does a whisker end?

For a set of lap times, $Q_1 = 16$ s and $Q_3 = 24$ s. The two largest times are 36 s and 37 s. Using the $1.5 \times$ IQR rule, state which value should be shown with a cross and where the upper whisker should end.

## SECTION 02 Extended response Q16–30

### 16. Climbing trials

Each of 32 climbers attempts the same five routes. The table shows the number of routes each climber completes.

| Routes completed | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Climbers | 3 | 7 | 9 | $k$ | 4 | 2 |

**(a)** Find $k$ and the number of climbers who complete at most two routes.

**(b)** The coach invites any climber who completes at least three routes to a second trial. What percentage of the group receives an invitation?

### 17. Lunchtime queues

A canteen records waiting times for 40 customers. The table gives cumulative counts.

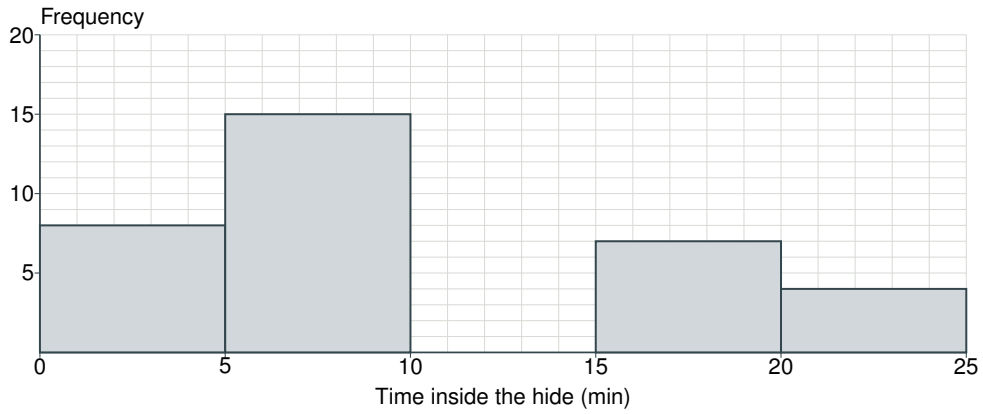
| Waiting time $t$ (min) | $t < 5$ | $t < 10$ | $t < 15$ | $t < 20$ | $t < 25$ |
|---|---|---|---|---|---|
| Number of customers | 4 | 17 | 29 | 36 | 40 |

**(a)** Find the frequencies for the five successive 5-minute intervals, starting at 0.

**(b)** The manager says that at least half the customers waited less than 10 minutes. Use the data to assess this claim.

### 18. A bird hide

A ranger times 50 visits to a bird hide. In this frequency histogram the bar for $10 \le t < 15$ has been omitted.



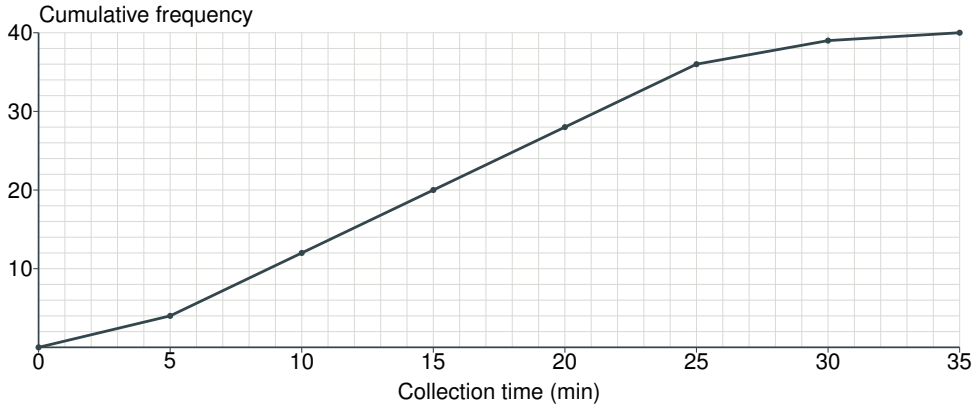

**(a)** Complete the histogram.

**(b)** Find the percentage of visits lasting at least 10 minutes but less than 20 minutes.

SECTION 02 / Extended response Questions 19–21

### 19. A parcel collection point

The graph summarises grouped collection times for 40 parcels. The shortest and longest recorded times were 2 and 34 minutes respectively.




**(a)** Use the graph to estimate the lower and upper quartiles, and hence the interquartile range.

**(b)** Find the exact range. Explain why this value can be stated exactly while your quartile readings are estimates.

### 20. A rehearsal

Sixteen musicians record how many times they restart a rehearsal passage. Each dot represents one musician.




**(a)** Find the median number of restarts. Show how you located it.

**(b)** What percentage of musicians restarted the passage at least four times?

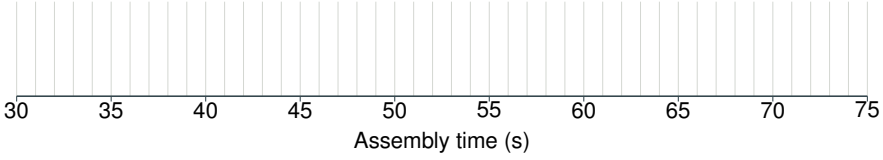
### 21. Assembly times

Twelve assembly times, in seconds, are shown in order.

$$32,\ 35,\ 36,\ 36,\ 38,\ 41,\ 43,\ 44,\ 48,\ 48,\ 50,\ 72$$

**(a)** Find the quartiles and median, using the medians of the lower and upper halves.

**(b)** Use the $1.5 \times$ IQR rule and construct a box-and-whisker diagram on the scale below. Mark any outlier with a cross.



SECTION 02 / Extended response Questions 22–24

### 22. A fish survey

The table groups the lengths of 30 fish. Two measurements were copied incorrectly. One fish recorded as 14.0 cm was actually 14.1 cm, and one recorded as 16.0 cm was actually 16.1 cm.

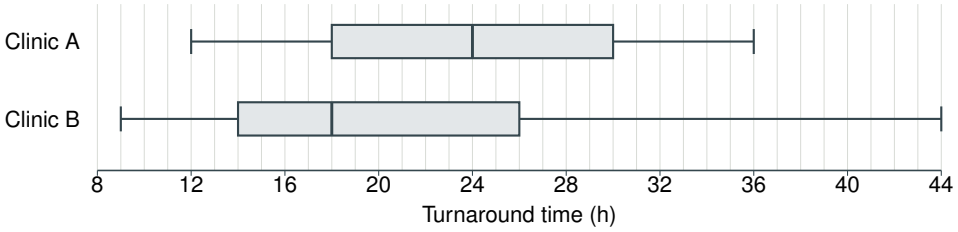
| Length $L$ (cm) | $10 < L \leq 12$ | $12 < L \leq 14$ | $14 < L \leq 16$ | $16 < L \leq 18$ |
|---|---|---|---|---|
| Recorded frequency | 4 | 13 | 9 | 4 |

**(a)** Write down the four corrected frequencies.

**(b)** Find the percentage of the fish whose corrected lengths exceed 14 cm.

### 23. Two clinics

The diagrams show test-result turnaround times at two clinics, on the same scale.




**(a)** Which clinic typically returns results sooner? Support your answer with both medians.

**(b)** Compare the spread of the middle 50% and the overall spread of the two samples.

**(c)** For which clinic is a normal model more plausible? Refer to the shapes of both diagrams.

### 24. Grading fruit by mass

A grower weighs 80 pieces of fruit. For estimates within a class, assume the masses are evenly spread.

| Mass $m$ (g) | $100 \leq m < 120$ | $120 \leq m < 140$ | $140 \leq m < 160$ | $160 \leq m < 180$ |
|---|---|---|---|---|
| Frequency | 9 | 27 | 32 | 12 |

**(a)** Estimate the 75th percentile of the masses.

**(b)** Estimate how many pieces would qualify for a label requiring a mass of at least 150 g.

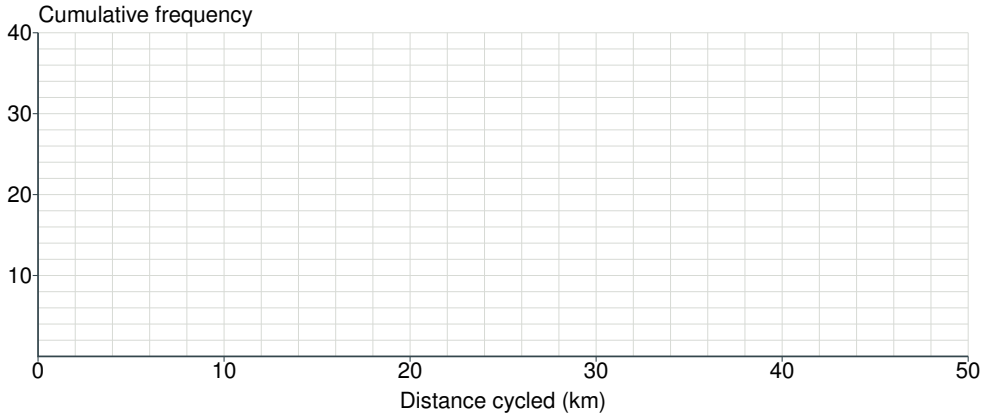
SECTION 02 / Extended response Questions 25–27

### 25. A charity ride

Forty cyclists report how far they rode during a charity event.

| Distance $d$ (km) | $0 \le d < 10$ | $10 \le d < 20$ | $20 \le d < 30$ | $30 \le d < 40$ | $40 \le d < 50$ |
|---|---|---|---|---|---|
| Frequency | 5 | 11 | 14 | 6 | 4 |

**(a)** Construct the cumulative frequency graph on the grid. Join consecutive points with straight segments.




**(b)** Use your graph to estimate the percentage of cyclists who rode less than 25 km.

### 26. Do matching box plots tell the whole story?

Two teams each record 12 finishing times in minutes, listed in ascending order.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Team A | 2 | 4 | 6 | 6 | 8 | 10 | 10 | 12 | 14 | 14 | 16 | 18 |
| Team B | 2 | 6 | 6 | 6 | 6 | 8 | 12 | 14 | 14 | 14 | 14 | 18 |

**(a)** Using medians of halves for the quartiles, show that the two teams have identical five-number summaries.

**(b)** A coach claims that an equal number of runners on each team must have finished in exactly 10 minutes. Assess the claim using the data.

### 27. Ferry departures

The target is a delay of less than 8 minutes. A port manager argues that East performed better because it had more departures meeting the target.

| Delay $d$ (min) | $0 \le d < 4$ | $4 \le d < 8$ | $8 \le d < 12$ | $12 \le d < 16$ |
|---|---|---|---|---|
| East terminal | 14 | 10 | 10 | 6 |
| West terminal | 10 | 12 | 6 | 2 |

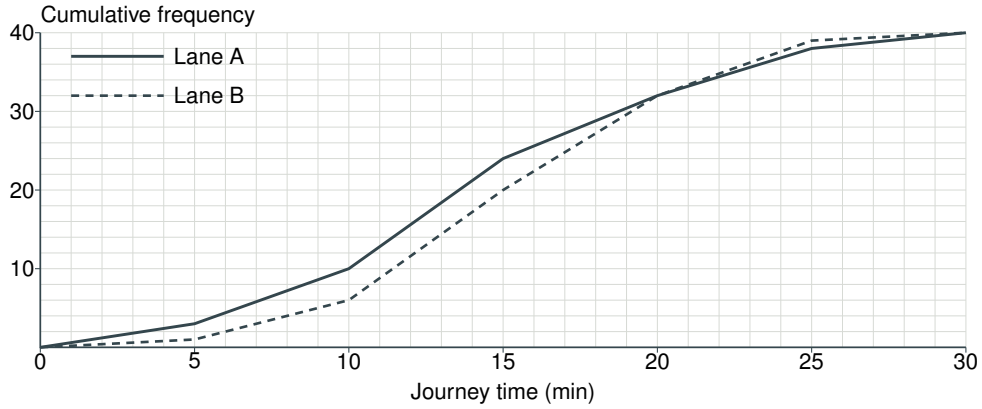
**(a)** Find the number of departures meeting the target at each terminal.

**(b)** Use percentages to evaluate the manager's argument.

SECTION 02 / Extended response Questions 28–30

### 28. Cycle lanes

Each lane was monitored for 40 journeys. At each time on the horizontal axis, the graph shows the cumulative number of journeys taking less than that time.




**(a)** Estimate the median journey time on each lane. Which lane has the shorter median and by about how much?

**(b)** The two graphs meet at $(20, 32)$. Interpret this point by giving a percentage in the context of the journeys.

### 29. Choosing a display for a museum

A museum records visit durations for four exhibitions. Choose a display for each request and justify your choice. You may use equal-width frequency histograms, cumulative frequency graphs or box-and-whisker diagrams.

**(a)** The booking team wants the time by which about 85% of visitors have finished the audio exhibition.

**(b)** A curator suspects that most visitors to the sculpture exhibition either stop briefly or stay a long time, with few visits of intermediate length.

**(c)** The director wants one figure comparing the typical visit duration and the spread of the middle half of visit durations across all four exhibitions.

### 30. What changes when samples are combined?

At site A, 90% of 40 sound readings are below 70 decibels. At site B, 70% of 60 readings are below 70 decibels. No reading is exactly 70 decibels.

**(a)** Find the percentage of the combined readings below 70 decibels.

**(b)** A technician simply averages 90% and 70%. Explain when this method would give the correct combined percentage for these two rates.

**(c)** At which site does 70 decibels lie above the upper quartile? Justify your answer using the percentile positions.

## SECTION 03 Paper 1 practice Q31–50

### 31. Locker use [Maximum mark: 5]

The table shows the number of times each of 60 students used a school locker on Monday.

| Number of uses | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Number of students | 8 | 15 | $k$ | 9 | 4 |

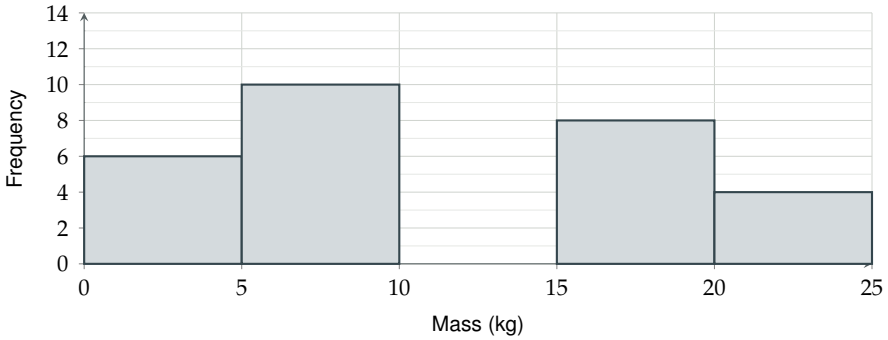
**(a)** Find $k$. **[2]**

**(b)** Find the percentage of students who used a locker at least twice. **[2]**

**(c)** A student who used a locker twice is selected. State whether this student's number of uses is above, below or equal to the median. **[1]**

### 32. Parcels at a sorting depot [Maximum mark: 6]

A sorting depot weighs a consignment of parcels. Exactly 30% weigh more than 15 kg. The bar for $10 < m \leq 15$ is missing. Each class includes its upper boundary.




**(a)** Use the two rightmost bars to find the total number of parcels. **[2]**

**(b)** Hence complete the histogram. **[2]**

**(c)** A supervisor claims that at least half the parcels weighed 10 kg or less. Use the histogram to assess this claim. **[2]**

### 33. Two bus routes [Maximum mark: 5]

The table summarises journey times, in minutes, over 80 mornings on each route.

| Route | Minimum | $Q_1$ | Median | $Q_3$ | Maximum |
|---|---|---|---|---|---|
| East | 20 | 28 | 31 | 34 | 47 |
| West | 22 | 25 | 29 | 37 | 42 |

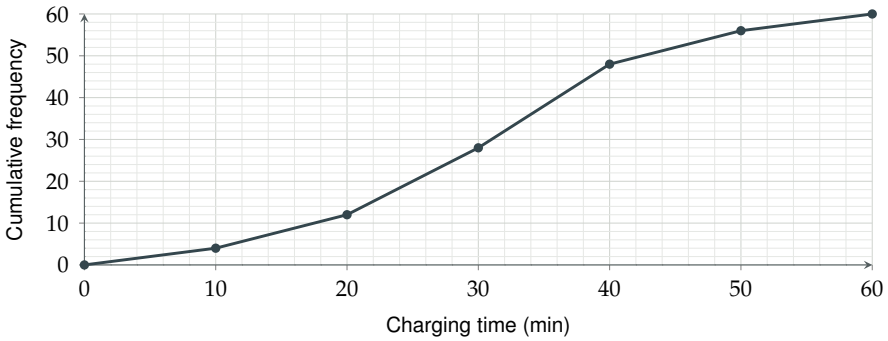
**(a)** Identify the route with the shorter typical journey. Support your answer. **[2]**

**(b)** Which route has the more consistent middle half of journey times? Justify your answer with calculations. **[3]**

SECTION 03 / Paper 1 practice Questions 34–36

## 34. Time spent charging [Maximum mark: 6]

A car park records 60 charging sessions. The graph gives the number lasting at most $t$ minutes.




**(a)** Estimate the median charging time. **[2]**

**(b)** A fee applies to the longest 20% of sessions. Estimate the time at which the fee should start. **[2]**

**(c)** Find the number of sessions lasting more than 20 minutes and at most 40 minutes. **[2]**

## 35. Reading at home [Maximum mark: 7]

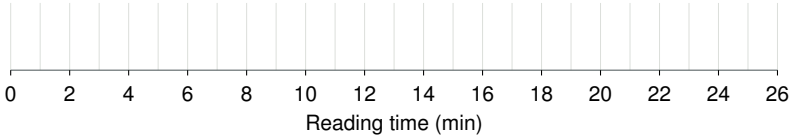
Twelve students recorded their reading times, in minutes, on one evening.

$$5,\ 7,\ 8,\ 8,\ 9,\ 10,\ 10,\ 11,\ 12,\ 12,\ 14,\ 24$$

**(a)** Find the median and interquartile range. **[2]**

**(b)** Use the $1.5 \times$ IQR rule to decide whether 24 minutes is an outlier. Show your working. **[2]**

**(c)** Draw a box and whisker diagram on the scale, indicating any outliers with a cross. **[3]**




## 36. A spreadsheet error [Maximum mark: 5]

The table should show cumulative totals for 48 deliveries. Exactly one entry is wrong.

| Delivery time, $t$ (min) | $t \leq 10$ | $t \leq 20$ | $t \leq 30$ | $t \leq 40$ |
|---|---|---|---|---|
| Cumulative frequency | 7 | 28 | 25 | 48 |

**(a)** Explain how the table shows that an error was made. **[1]**

**(b)** The driver confirms that 12 deliveries took more than 20 minutes and at most 30 minutes, and that the total at 20 minutes is correct. Find the corrected entry. **[2]**

**(c)** Hence find the percentage of deliveries taking more than 30 minutes. **[2]**

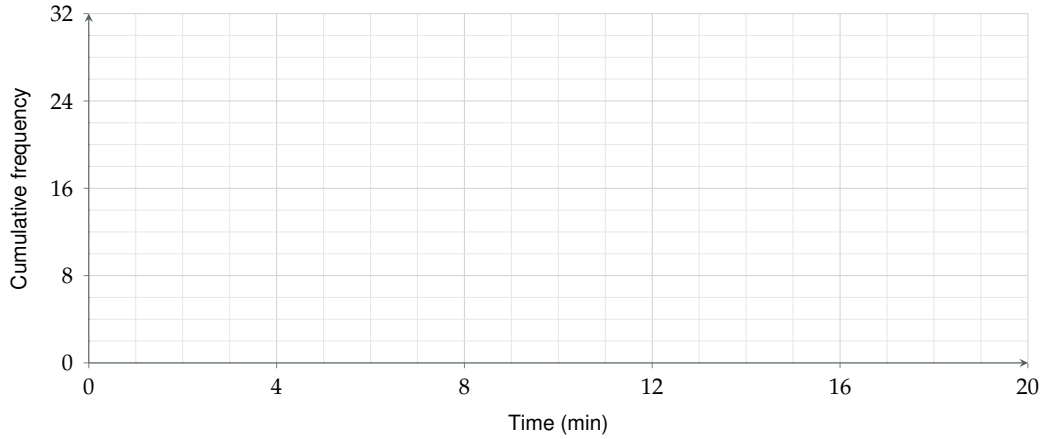
SECTION 03 / Paper 1 practice Questions 37–39

## 37. A gallery sound installation [Maximum mark: 7]

A curator records how long 32 visitors spend at an installation. All times are positive.

| Time, $t$ (min) | $0 < t \leq 4$ | $4 < t \leq 8$ | $8 < t \leq 12$ | $12 < t \leq 16$ | $16 < t \leq 20$ |
|---|---|---|---|---|---|
| Frequency | 4 | 8 | 12 | 6 | 2 |

**(a)** On the grid, draw a cumulative frequency graph. Join your points with straight lines. **[3]**




**(b)** Use your graph to estimate the median time. **[2]**

**(c)** One full performance lasts 12 minutes. Find the percentage of visitors who stayed longer than one full performance. **[2]**

## 38. Which runner is more consistent? [Maximum mark: 5]

Two runners record their times, in seconds, for the same hill sprint.

| Runner | Minimum | $Q_1$ | Median | $Q_3$ | Maximum |
|---|---|---|---|---|---|
| Lina | 48 | 52 | 54 | 56 | 65 |
| Arun | 49 | 51 | 54 | 59 | 62 |

**(a)** Use the ranges to identify the more consistent runner over all recorded sprints. **[2]**

**(b)** The coach only wants to compare the middle half of the times. Explain whether this changes the choice in part (a). **[2]**

**(c)** Compare the runners' median times. **[1]**

## 39. Correcting a class boundary [Maximum mark: 5]

The counts below include one misclassified reading. A temperature of exactly 8°C was entered into the third class. All other entries are correct.

| Temperature, $T$ (°C) | $0 < T \leq 4$ | $4 < T \leq 8$ | $8 < T \leq 12$ | $12 < T \leq 16$ |
|---|---|---|---|---|
| Recorded frequency | 3 | 9 | 12 | 6 |

**(a)** Write down the two corrected class frequencies. **[2]**

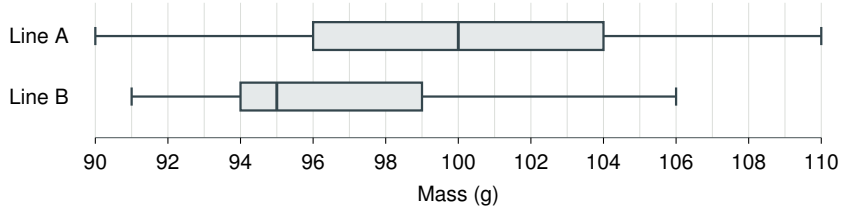
**(b)** Find the corrected percentage of readings that were at most 8°C. **[2]**

**(c)** State the effect of the correction on the cumulative frequency at 12°C. **[1]**

SECTION 03 / Paper 1 practice Questions 40–42

## 40. Checking a filling machine [Maximum mark: 6]

The diagrams summarize the masses of sealed packets from two production lines.




**(a)** Find the difference between the median masses. **[2]**

**(b)** Determine which line has the smaller interquartile range. Give both values. **[2]**

**(c)** For which line is a normal model more plausible? Justify your choice using the box and whiskers. **[2]**

## 41. A theatre queue [Maximum mark: 6]

The table gives the waiting times of 50 customers.

| Wait, $w$ (min) | $0 < w \leq 4$ | $4 < w \leq 8$ | $8 < w \leq 12$ | $12 < w \leq 16$ |
|---|---|---|---|---|
| Frequency | 6 | 14 | 20 | 10 |

**(a)** Find the number of customers who waited at most eight minutes. **[1]**

**(b)** Assuming waits are evenly distributed within each class, estimate the number who waited at most nine minutes. **[3]**

**(c)** Use your estimate to find the percentage who waited more than nine minutes. **[2]**

## 42. Machine stoppages [Maximum mark: 6]

A factory records how many times a machine stops during each of 30 shifts. There are twice as many shifts with one stop as shifts with three stops.

| Number of stops | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| Number of shifts | 7 | $p$ | 8 | $q$ |

**(a)** Find $p$ and $q$. **[3]**

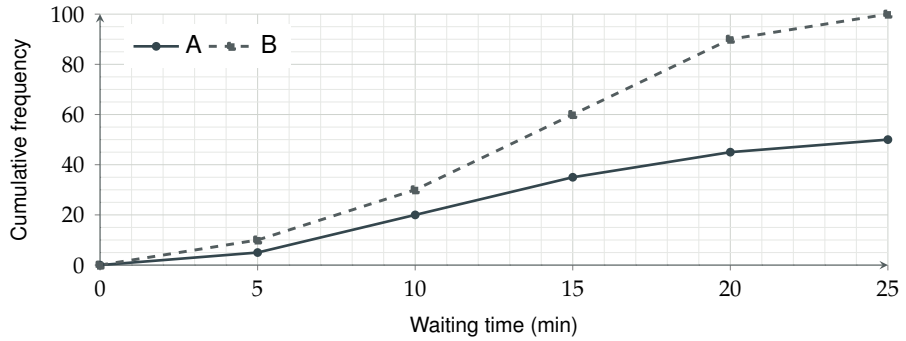
**(b)** Determine the median number of stops per shift. **[2]**

**(c)** Write down the number of shifts with more than one stop. **[1]**

SECTION 03 / Paper 1 practice Questions 43–45

## 43. Two clinic waiting rooms [Maximum mark: 7]

The graph shows waiting times on one morning at clinic A (50 patients) and clinic B (100 patients).




**(a)** For each clinic, estimate the percentage of patients seen within 10 minutes. **[3]**

**(b)** An administrator says that B was quicker because it saw more patients within 10 minutes. Use your answers to assess this claim. **[2]**

**(c)** Estimate the 90th percentile of waiting time at each clinic. **[2]**

## 44. Stream temperatures [Maximum mark: 5]

A surveyor groups stream temperatures into classes of width 2°C. The lowest occupied class is $8 < T \leq 10$ and the highest is $18 < T \leq 20$.

**(a)** A colleague reports an exact range of 12°C. Explain why the grouped data do not support this claim. **[2]**

**(b)** The surveyor retrieves the lowest and highest readings, 8.4°C and 19.7°C. Calculate the range. **[1]**

**(c)** A faulty thermometer added 0.6°C to every reading. Find the range after correction and explain your answer. **[2]**

## 45. Awarding a badge [Maximum mark: 5]

A club gives a badge for completing at least $k$ puzzles. Exactly 40 members take part.

| Puzzles completed | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Number of members | 3 | 7 | 12 | 10 | 8 |

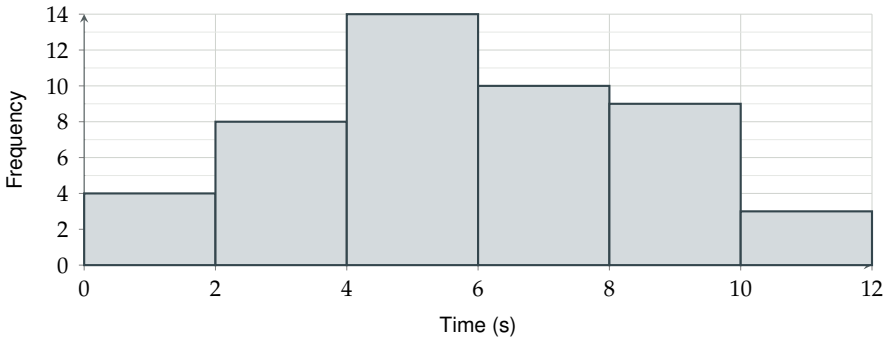
**(a)** For $k = 3$, find the percentage who receive a badge. **[2]**

**(b)** The club wants badges to go to exactly the top quarter. Decide whether any integer $k$ can achieve this. Support your answer. **[3]**

SECTION 03 / Paper 1 practice Questions 46–48

## 46. Rebinning a histogram [Maximum mark: 5]

A technician records how long 48 lamps take to reach full brightness. The histogram has six classes of width two seconds. Each class includes its upper boundary.




**(a)** The times are regrouped into classes $0 < t \leq 4$, $4 < t \leq 8$ and $8 < t \leq 12$. Find the three new frequencies. **[2]**

**(b)** Use the *original* histogram to find the percentage taking more than six seconds. **[2]**

**(c)** Explain why that percentage cannot be recovered exactly from the three new frequencies alone. **[1]**

## 47. A missing survey total [Maximum mark: 6]

The table gives cumulative frequencies for the distances travelled to school by a group of students. Exactly 75% of students travel at most 6 km.

| Distance, $d$ (km) | $d \leq 2$ | $d \leq 4$ | $d \leq 6$ | $d \leq 8$ | $d \leq 10$ |
|---|---|---|---|---|---|
| Cumulative frequency | 12 | 33 | 54 | 68 | $N$ |

**(a)** Find the total number of students, $N$. All distances are at most 10 km. **[2]**

**(b)** Find the number who travel more than 4 km but at most 8 km. **[2]**

**(c)** Identify the class interval containing the median distance. Justify your answer. **[2]**

## 48. An express-lane target [Maximum mark: 6]

A supermarket wants at least 70% of customers to wait at most three minutes. It records these waits for 80 customers.

| Wait, $t$ (min) | $0 < t \leq 2$ | $2 < t \leq 4$ | $4 < t \leq 6$ | $6 < t \leq 8$ |
|---|---|---|---|---|
| Frequency | 24 | 40 | 12 | 4 |

**(a)** Calculate how many of the 80 customers must wait at most three minutes to meet the target. **[1]**

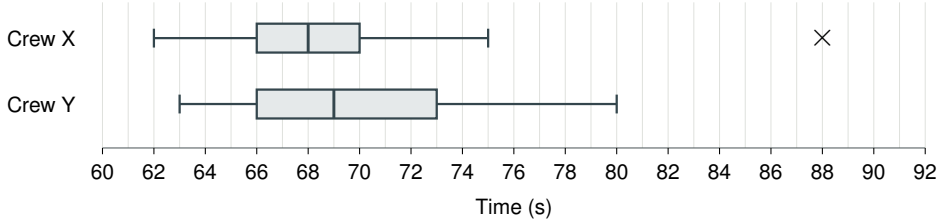
**(b)** Assuming waits are evenly distributed within the second class, estimate the percentage who waited at most three minutes. Hence assess the target. **[3]**

**(c)** Using only the table, find the least and the greatest possible numbers who waited at most three minutes. **[2]**

SECTION 03 / Paper 1 practice Questions 49–50

## 49. Two kayak crews [Maximum mark: 7]

Two crews record how long it takes to paddle a fixed course. Each cross marks an outlier. All times are in seconds.



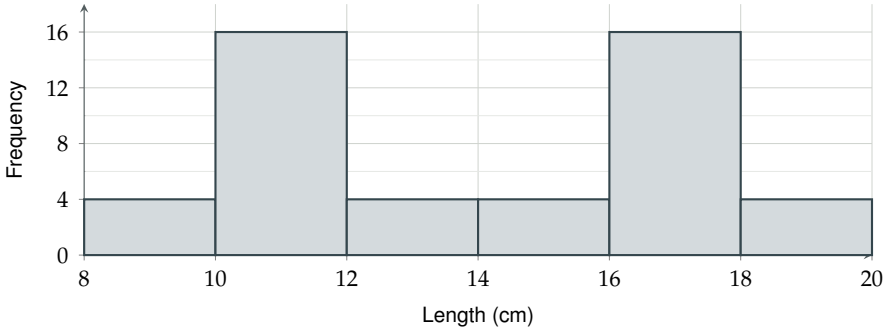

**(a)** Find the range of all the times recorded by crew X. **[2]**

**(b)** Show, using the interquartile range, why the crossed value for X is an outlier. **[3]**

**(c)** A coach wants the crew with the lower median and smaller interquartile range. Use the diagrams to choose a crew and support your decision. **[2]**

## 50. Fish in a hatchery [Maximum mark: 11]

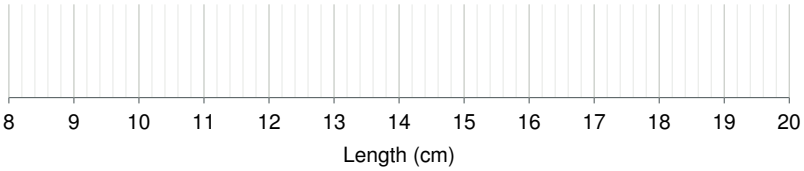
A biologist measures the lengths of 48 fish. The histogram uses classes of width 2 cm. Each class includes its upper boundary.




**(a)** Write down the cumulative frequencies at lengths of 12 cm and 16 cm. **[2]**

**(b)** Assume lengths are evenly distributed within each class. Estimate the lower quartile, median and upper quartile. **[4]**

**(c)** The smallest and largest fish measured 8.4 cm and 19.6 cm. Using your estimates from part (b), draw a box and whisker diagram on the scale below. **[3]**

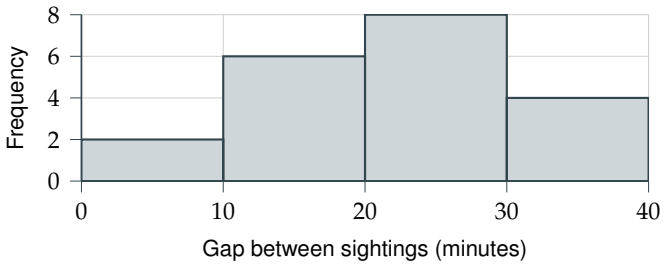



**(d)** The biologist suggests a normal model because the quartiles are equally spaced about the median. Use the histogram to assess this suggestion. **[2]**

# Answer key and solution outlines

## Section 01 Short response — Questions 1–15

1. For 0, 1, 2 and 3 buses, the frequencies are 3, 5, 3 and 1 respectively. They sum to 12 students.
2. $15 \leq m < 20$. A mass of 15 kg lies in this class because the lower endpoint is included.
3. $(8+9)/25 = 17/25 = 68\%$.
4. $7 + k + 2k + 5 = 36$, so $3k = 24$ and $k = 8$. There are $2k = 16$ sketches in the third group.
5. **(a)** $5 + 2 = 7$ customers. **(b)** $4 \leq t < 6$ minutes, with 12 customers.
6. Draw four touching bars of width 10 minutes, with successive heights 2, 6, 8 and 4.



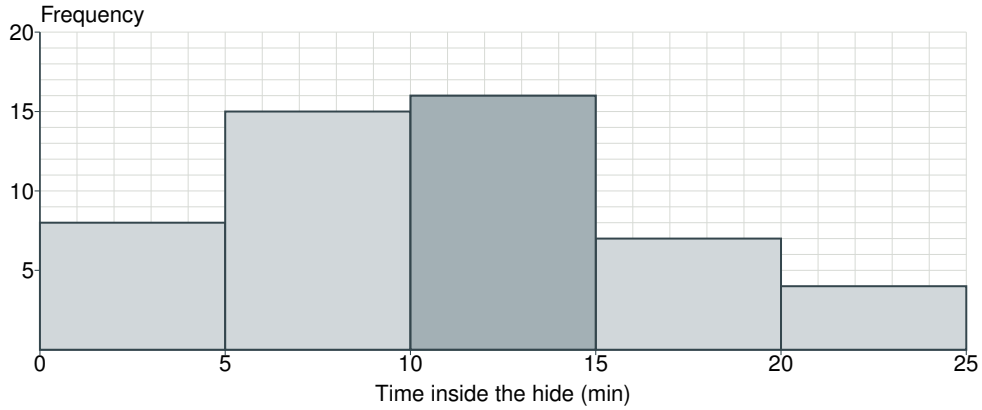

7. $41 - 18 = 23$ orders.
8. $(10, 18)$. The upper boundary is 10 minutes and the accumulated count there is $7 + 11 = 18$.
9. The upper 15% contains about $0.15(160) = 24$ repairs.
10. Read at cumulative frequencies $20, 40, 60$. The estimates are $Q_1 = 5$ km, median $= 7$ km and $Q_3 = 9$ km, giving IQR $= 9 - 5 = 4$ km.
11. The 90th percentile is read at cumulative frequency $0.90(80) = 72$. The distance is about 10.4 km.
12. The graph spans class boundaries from about 2 to 12 km, suggesting a range of about 10 km. The exact range is $11.8 - 2.3 = 9.5$ km, which is 0.5 km smaller.
13. Club B has a higher median weekly training time (32 minutes versus 28 minutes for A). The middle half of Club B's times is slightly more spread out. Its IQR is $42 - 24 = 18$ minutes, compared with $36 - 20 = 16$ minutes for A.
14. Yes. The median is central in the box (28 is midway between 20 and 36) and both whiskers have length 8 minutes. This symmetry is consistent with a normal model, although a box plot alone cannot establish normality.
15. The IQR is 8 s and the upper fence is $24 + 1.5(8) = 36$ s. Mark 37 s with a cross. The upper whisker ends at 36 s, the largest value at or below the fence.

## Section 02 Extended response — Questions 16–23

**16.** **(a)** $k = 32 - (3+7+9+4+2) = 7$. At most two routes: $3+7+9 = 19$ climbers. **(b)** $(7+4+2)/32 \times 100 = 40.625\%$.

**17.** **(a)** $4,\ 17-4,\ 29-17,\ 36-29,\ 40-36$, giving $4, 13, 12, 7, 4$. **(b)** $17/40 = 42.5\%$ waited less than 10 minutes, so the claim is false.

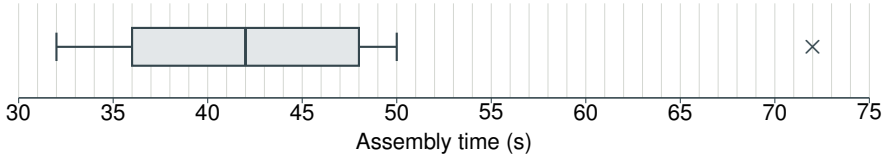
**18.** **(a)** The missing bar has frequency $50 - (8+15+7+4) = 16$. **(b)** $(16+7)/50 \times 100 = 46\%$.




**19.** **(a)** Read at cumulative frequencies 10 and 30. $Q_1 \approx 8.75$ min and $Q_3 \approx 21.25$ min, giving IQR $\approx 12.5$ min. Allow reasonable graph readings. **(b)** The range is $34 - 2 = 32$ min, using the recorded extremes. The graph endpoints 0 and 35 are class boundaries, and the quartiles are estimated between plotted points.

**20.** **(a)** The 8th and 9th results are both 2, so the median is 2 restarts. **(b)** $(3+0+1)/16 \times 100 = 25\%$.

**21.** **(a)** $Q_1 = (36+36)/2 = 36$, median $= (41+43)/2 = 42$, $Q_3 = (48+48)/2 = 48$ s. **(b)** IQR $= 12$ s. Fences are 18 and 66 s. Whiskers end at 32 and 50 s. Mark 72 s with a cross.



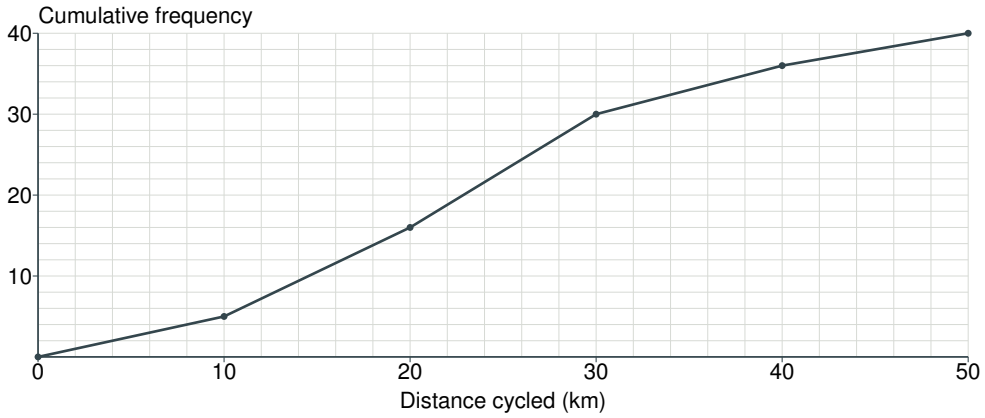

**22.** **(a)** $4, 12, 9, 5$. One fish leaves and one enters the third class. **(b)** $(9+5)/30 \times 100 = 46\frac{2}{3}\%$ (about 46.7%).

**23.** **(a)** Clinic B has the shorter median, 18 h compared with 24 h. **(b)** Both IQRs are 12 h ($30-18$ and $26-14$). Clinic B has the larger range, $44 - 9 = 35$ h compared with $36 - 12 = 24$ h. **(c)** Clinic A is more consistent with a normal model. Its median is central and its whiskers have equal lengths. Clinic B has a longer upper whisker and its median is closer to the lower quartile.

## Section 02 Extended response Questions 24–30

**24.** **(a)** The 60th cumulative position lies in the $140 \leq m < 160$ class. Estimate $140 + 20(60 - 36)/32 = 155$ g. **(b)** Half the 32 in that class and all 12 in the top class qualify, giving about $16 + 12 = 28$ pieces.

**25.** **(a)** Join $(0,0), (10,5), (20,16), (30,30), (40,36), (50,40)$. **(b)** At 25 km, the estimated count is $(16 + 30)/2 = 23$. The percentage is about $23/40 \times 100 = 57.5\%$.




**26.** **(a)** Both have minimum 2, $Q_1 = (6 + 6)/2 = 6$, $Q_3 = (14 + 14)/2 = 14$ and maximum 18 min. Medians are $(10 + 10)/2 = (8 + 12)/2 = 10$ min. **(b)** Two Team A runners and no Team B runners took exactly 10 minutes, so the claim is false.

**27.** **(a)** East: $14 + 10 = 24$. West: $10 + 12 = 22$. **(b)** East met the target on $24/40 = 60\%$ of departures and West on $22/30 \approx 73.3\%$. West performed better relative to its total number of departures.

**28.** **(a)** Read at cumulative frequency 20. Lane A takes about 13.6 min and lane B 15 min. Lane A's median is about 1.4 min shorter. Accept reasonable readings. **(b)** On each lane, $32/40 = 80\%$ of journeys took less than 20 minutes.

**29.** **(a)** Use a cumulative frequency graph. At $0.85N$ on its vertical axis, read off the time for the audio exhibition. **(b)** A frequency histogram can show two peaks with a trough between them. **(c)** Four box plots on one scale show medians and IQRs directly. Accept other displays if the explanation shows how they answer the request.

**30.** **(a)** $0.90(40) + 0.70(60) = 36 + 42 = 78$ out of 100, so 78%. **(b)** The simple average gives each site equal weight. With these distinct rates it is correct when the sample sizes are equal. **(c)** Site A only. There, 90% of readings fall below 70 dB, placing it above $Q_3$. At B the proportion is 70%, below the 75% needed.

## Section 03 Paper 1 practice — Questions 31–38

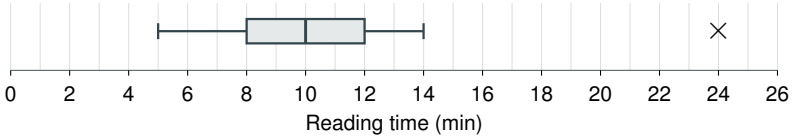
**31.** **(a)** $k = 60 - (8 + 15 + 9 + 4) = 24$. **(b)** $(24 + 9 + 4)/60 \times 100 \approx 61.7\%$. **(c)** Equal. Positions 30 and 31 both represent two uses (positions 24–47).

**32.** **(a)** The two rightmost bars give $8 + 4 = 12$ parcels. Since this is 30% of the total, there are $12/0.30 = 40$ parcels. **(b)** The missing frequency is $40 - (6 + 10 + 8 + 4) = 12$. Draw a bar from 10 to 15 of height 12. **(c)** Only $6 + 10 = 16$ out of 40 parcels (40%) weighed at most 10 kg, so the claim is false.

**33.** **(a)** West has the shorter median journey, 29 min compared with 31 min. **(b)** East has IQR $34 - 28 = 6$ min, compared with $37 - 25 = 12$ min for West. East's middle half is more consistent.

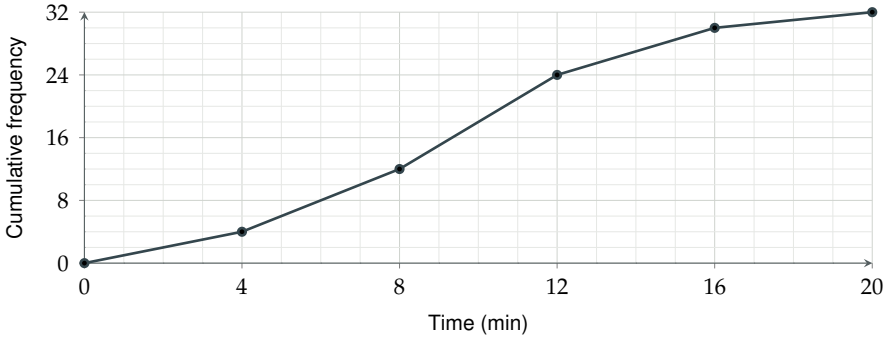
**34.** **(a)** The median is read at cumulative frequency 30, giving about 31 min. **(b)** $0.80(60) = 48$. Read across from 48 to obtain 40 min. **(c)** $48 - 12 = 36$ sessions.

**35.** **(a)** Median $= 10$ min, $Q_1 = 8$, $Q_3 = 12$, so IQR $= 4$ min. **(b)** Upper fence $= 12 + 1.5(4) = 18$ min. Since $24 > 18$, it is an outlier. **(c)** The whiskers end at 5 and 14. The cross marks 24.




**36.** **(a)** A cumulative total cannot decrease from 28 to 25. **(b)** The total at 30 min should be $28 + 12 = 40$. **(c)** $8/48 \times 100 \approx 16.7\%$.

**37.** **(a)** Start at $(0, 0)$ and plot totals 4, 12, 24, 30, 32 at 4, 8, 12, 16, 20 min respectively. Join with straight segments as shown. **(b)** Read across from 16 to estimate 9.3 min. **(c)** $(32 - 24)/32 \times 100 = 25\%$.




**38.** **(a)** Lina's range is $65 - 48 = 17$ s and Arun's is $62 - 49 = 13$ s. Arun is more consistent by range.
**(b)** Yes. Lina's IQR is $56 - 52 = 4$ s against Arun's $59 - 51 = 8$ s. Lina's middle half varies less.
**(c)** Both have median 54 s.

## Section 03 Paper 1 practice — Questions 39–50

**39.** **(a)** The corrected frequencies for the second and third classes are 10 and 11 respectively. **(b)** $(3+10)/30 \times 100 \approx 43.3\%$. **(c)** No change. The reading counts towards the total at 12°C in either class.

**40.** **(a)** $100 - 95 = 5$ g. **(b)** A has IQR $104 - 96 = 8$ g, B has $99 - 94 = 5$ g. B's IQR is smaller. **(c)** A is more plausible. Its median is centred within the box and its whiskers have equal lengths. B has a longer upper whisker and its median lies near the lower quartile.

**41.** **(a)** $6 + 14 = 20$. **(b)** Nine minutes is one-quarter of the way from 8 to 12. Estimate $20 + \frac{1}{4}(20) = 25$ customers. **(c)** $(50 - 25)/50 \times 100 = 50\%$.

**42.** **(a)** $p + q = 30 - 7 - 8 = 15$, $p = 2q$, so $q = 5$ and $p = 10$. **(b)** Positions 15 and 16 lie between cumulative totals 7 and 17. The median is one stop. **(c)** $8 + 5 = 13$ shifts.

**43.** **(a)** A has $20/50 = 40\%$ and B has $30/100 = 30\%$. **(b)** B saw 30 patients within 10 min, compared with 20 at A, but had twice as many patients. A saw the larger fraction within 10 min, so these counts do not support the claim. **(c)** Read at $0.90(50) = 45$ for A and $0.90(100) = 90$ for B. Both give 20 min.

**44.** **(a)** The boundaries 8 and 20 do not identify the observed extrema. In particular, the minimum exceeds 8°C and the maximum is at most 20°C, so the range is less than 12°C. **(b)** $19.7 - 8.4 = 11.3$°C. **(c)** 11.3°C. Subtracting 0.6 from both extrema leaves their difference unchanged.

**45.** **(a)** $(10+8)/40 \times 100 = 45\%$. **(b)** The top quarter comprises 10 members. With $k = 4$, only eight qualify, whereas $k = 3$ gives 18. No integer threshold can select exactly 10 because of the tie at three puzzles.

**46.** **(a)** $4 + 8 = 12$, $14 + 10 = 24$, $9 + 3 = 12$. **(b)** $(10 + 9 + 3)/48 \times 100 \approx 45.8\%$. **(c)** The new middle class straddles six seconds. Its frequency does not identify how many of its lamps took more than six seconds.

**47.** **(a)** $0.75N = 54$, giving $N = 72$. **(b)** $68 - 33 = 35$. **(c)** The 36th and 37th distances lie in $4 < d \leq 6$ km, since $33 < 36 < 37 \leq 54$.

**48.** **(a)** $0.70(80) = 56$. **(b)** Estimate $24 + \frac{1}{2}(40) = 44$ customers, or $44/80 = 55\%$. The estimate falls short of the target. **(c)** The least is 24 and the greatest is 64, since anywhere from zero to 40 customers in the second class could have waited at most three minutes.

**49.** **(a)** The cross is an observed time. The range is $88 - 62 = 26$ s. **(b)** IQR $= 70 - 66 = 4$ s and upper fence $= 70 + 1.5(4) = 76$ s. The observation at 88 s exceeds 76 s. **(c)** Choose X. Its median is 68 s compared with 69 s for Y, and its IQR is 4 s compared with $73 - 66 = 7$ s.

**50.** **(a)** At 12 cm the total is $4 + 16 = 20$. At 16 cm it is $20 + 4 + 4 = 28$. **(b)** The quartile levels are 12, 24 and 36. From 10 to 12 cm the cumulative count rises from 4 to 20, so $Q_1 \approx 11$ cm. At 14 cm the count is 24, so the estimated median is 14 cm. From 16 to 18 cm it rises from 28 to 44, so $Q_3 \approx 17$ cm. **(c)** Draw the box from 11 to 17 cm, with the median at 14 cm. Draw the whiskers to 8.4 and 19.6 cm. **(d)** The histogram has two peaks, at 10–12 cm and 16–18 cm, with a trough in the centre. A normal model would have a single central peak.

