IB MATHEMATICS / ANALYSIS AND APPROACHES SL & HL

Representing data

Frequency tables, histograms, cumulative frequency and box plots

Learning goal	Choose, draw and interpret a display to answer a question about data.
Syllabus link	AA SL 4.2 · Topic 4, Statistics and probability (SL and HL).
Structure	15 short responses, 15 extended responses, 20 Paper 1 questions and 10 additional challenges for HL (p. 15). Answers begin on p. 18.
Technology	Work without a calculator. Use a ruler for scales and straight-line graphs.

Instructions Work in your notebook and use the grids provided. Show your reasoning. Give exact answers unless an estimate is requested. For ordered lists, take the median of each half for the quartiles, excluding the overall median when the number of values is odd.

SECTION 1 | Short response

Q1–4

1. Twelve pupils count how many seeds germinated in their trays. Organize the counts in a discrete frequency table, showing every value from 0 to 3.

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

2. A laboratory groups lengths using $0 \le x < 5$, $5 \le x < 10$ and $10 \le x < 15$, in centimetres. Identify the classes containing 4.99, 5.00 and 10.00 cm, respectively.

3. A train carried the following numbers of bicycles on 20 journeys. How many journeys carried at least two bicycles?

Bicycles	0	1	2	3	4
Journeys	3	5	7	4	1

4. A student wants to draw a frequency histogram of lap times. She lists three classes as $10 < t < 20$, $20 < t < 30$ and $30 < t < 40$. Rewrite them to include every time from 10 seconds up to, but excluding, 40 seconds exactly once.

SECTION 1 | Short response Q5–10

5. The histogram shows how far 30 people walked during a break. Each class includes its lower boundary and excludes its upper boundary. How many walked at least 200 m but less than 400 m?



6. The frequencies in four consecutive time intervals are 5, 7, 0 and 8. Write down the four cumulative frequencies.

7. A cumulative frequency graph represents 80 measurements. At which cumulative frequency should you read the 65th percentile?

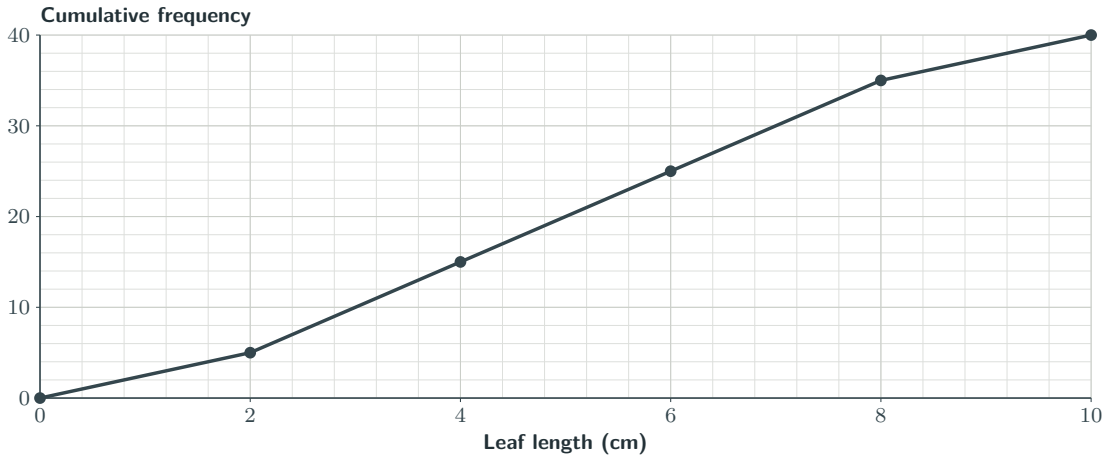
8. Nine scores in ascending order are $2, 4, 6, 7, 8, 11, 12, 14, 16$. Find the median and the interquartile range.

9. A box plot has whisker ends at 6 and 18 minutes and a single cross at 25 minutes. Write down the largest observed time and the range.

10. A response-time table starts with the class $5 \leq t < 10$ seconds, of frequency 3. No value is below 5. Give the first two points on a cumulative frequency graph using upper class boundaries.

SECTION 1 | Short response Q11–15

11. This cumulative frequency graph shows the lengths of 40 leaves. Use it to estimate the median and the 90th percentile.



12. A grouped table has 18 observations below 30 and 25 below 40. How many lie in $30 \le x < 40$?

13. The proposed cumulative frequencies of four consecutive classes are $8, 17, 15, 24$. Explain how you know at least one entry is wrong.

14. The box plot shows packing times at a warehouse. Read off the lower quartile and give the interval containing the middle 50% of times.



15. A box plot has no outliers. Its whiskers have equal length and its median lies halfway between the quartiles. What does this suggest about symmetry? Could the data be normally distributed? Explain briefly.

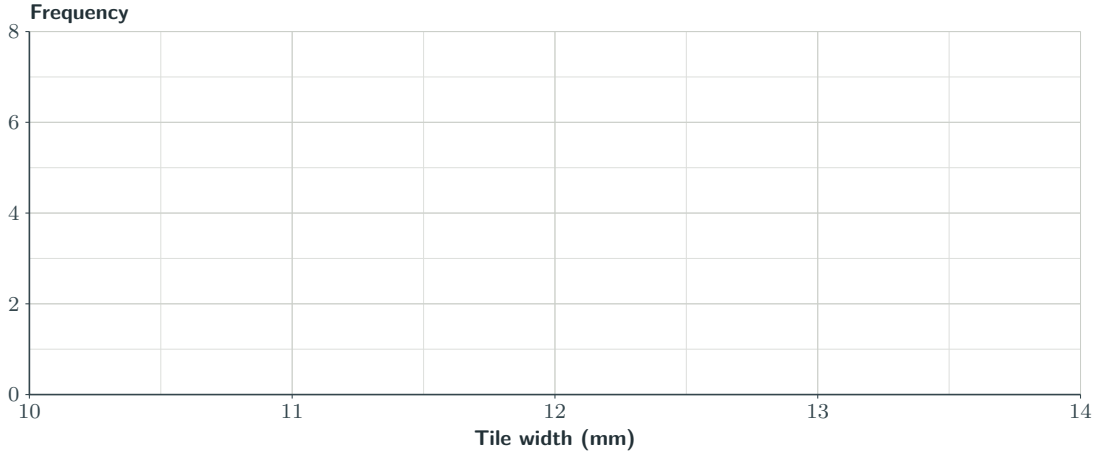
SECTION 2 | Extended response Q16–19

16. A technician measures the widths, w mm, of 16 tiles.

11.8, 10.0, 13.2, 12.0, 11.0, 10.6, 13.4, 12.1, 11.5, 12.7, 10.9, 11.2, 12.0, 11.9, 13.0, 11.3

(a) Construct a frequency table with four equal-width classes, starting at 10 mm and ending at 14 mm. Use inequalities.

(b) Draw the corresponding frequency histogram on the grid.



17. A chart groups 24 online orders by delivery time. Its four one-day intervals, in order, contain 8, 0, 10 and 6 orders. A designer removes the empty interval and slides the last two bars to the left.

(a) Explain how this changes what a reader would infer about delivery times.

(b) Describe how the interval with zero frequency should appear.

18. A theatre counts tickets bought per booking over one morning.

Tickets per booking	1	2	3	4
Number of bookings	6	9	3	2

(a) Find the median number of tickets per booking.

(b) Find the fraction of all tickets that were sold in bookings of four.

19. A table groups 48 journey times into ten-minute intervals. The first occupied class is $10 \le t < 20$ and the last is $40 \le t < 50$. A student reports an exact range of 40 minutes. Explain the error and state what additional information would give the exact range.

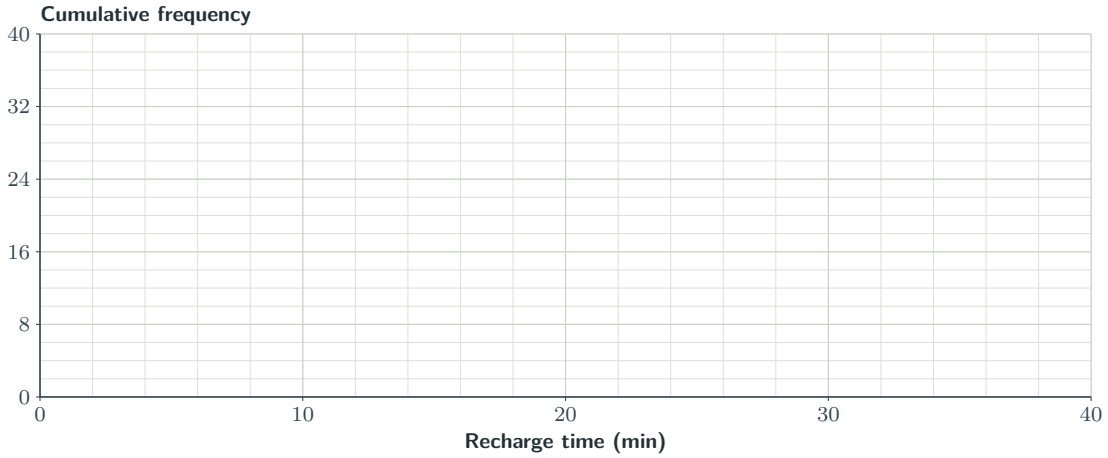
SECTION 2 | Extended response Q20–23

20. A team records the recharge times of 40 batteries.

Time t (min)	$0 \le t < 10$	$10 \le t < 20$	$20 \le t < 30$	$30 \le t < 40$
Frequency	4	12	16	8

(a) Make a cumulative frequency table at the upper class boundaries.

(b) Plot the cumulative frequencies on the grid. Include the starting point and join consecutive points with straight line segments.



21. Use your graph from Question 20.

(a) Estimate the percentage of batteries that finish charging in under 25 minutes.

(b) A technician wants a time by which about 90% finish charging. Estimate this time and explain how you located it.

22. A changing room has 12 lockers. A student counts how many are occupied every hour. She needs to report both the most frequent count and the number of observations with at least eight lockers occupied. Would an exact frequency table or a box plot better support both tasks? Explain how she could obtain each answer.

23. A cumulative frequency graph has a horizontal segment from 20 g to 30 g at a cumulative frequency of 18. The vertical axis counts items with mass strictly below each stated boundary.

(a) How many items have masses in $20 \le m < 30$? Explain.

(b) A student says that 18 items must each weigh 20 g. Use the graph information to assess this statement.

SECTION 2 | Extended response Q24–27

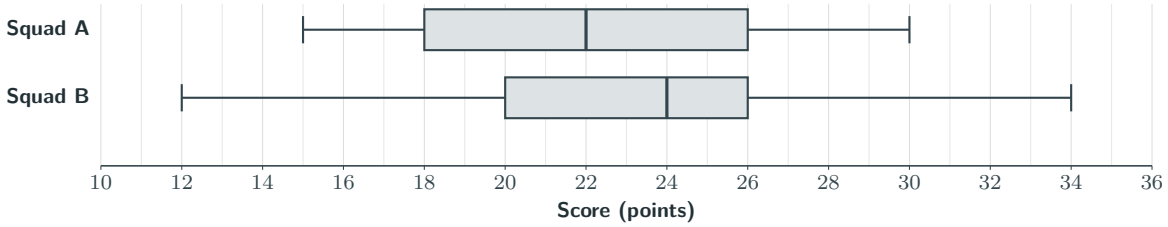
24. A student counts the people waiting at a lunch counter on 13 days. The ordered counts are $8, 9, 10, 10, 12, 13, 14, 15, 15, 16, 16, 17, 29$.

(a) Find the quartiles and use the $1.5 \times \text{IQR}$ rule to identify any outliers.

(b) Draw a box-and-whisker diagram on the scale. Use a cross for each outlier.



25. Two squads complete the same archery test.



(a) Which squad has the higher median score, and by how many points?

(b) A coach says Squad B is more consistent. Compare the IQRs and the ranges, then explain why the choice of measure matters.

26. Six pupils take two tests. Their first scores are $4, 5, 6, 7, 8, 9$. Their second scores are $5, 6, 7, 8, 9, 10$. The names on the second list are missing.

(a) Pair the scores so that every pupil improved.

(b) Find another pairing in which only one pupil improved. What does this tell you about using the change in the median to describe individual pupils?

27. Two sets of seven integer scores have the same five-number summary $(0, 2, 5, 8, 10)$. Construct one set with seven distinct scores and another with a repeated score. Explain why their box plots are identical.

SECTION 2 | Extended response Q28–30

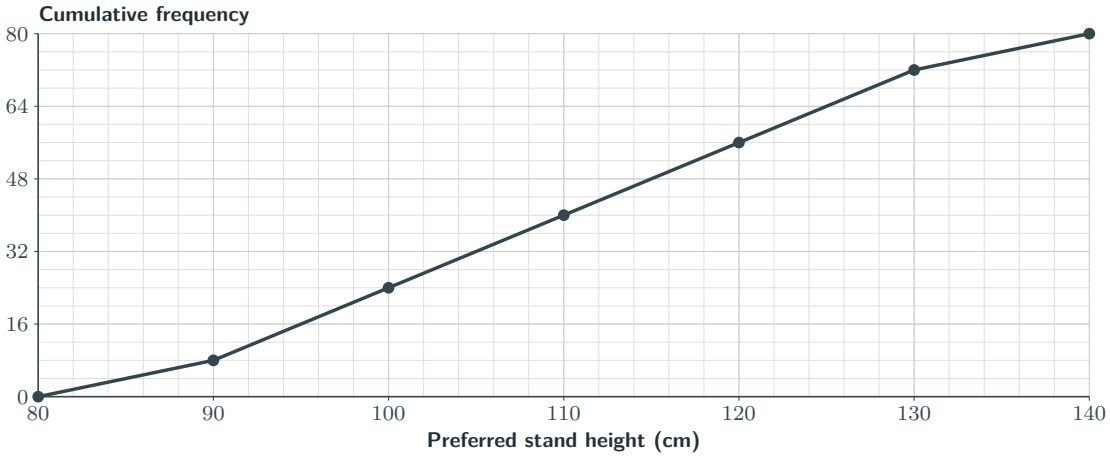
28. All recorded journeys on two bus routes lasted at most 25 minutes. Each entry is the number of journeys lasting at most b minutes.

Time b (min)	5	10	15	20	25
East route	6	18	30	48	60
West route	4	16	30	36	40

(a) Compare the percentages of journeys lasting at most 10 minutes.

(b) Which percentile does 15 minutes represent on each route's cumulative frequency graph?

29. An orchestra measures the preferred music-stand heights of 80 players. A new stand should accommodate the middle 80% of preferred heights.



(a) Identify the two percentile cutoffs and use the graph to estimate the minimum and maximum heights the stand should reach.

(b) A cheaper stand has the same minimum but reaches only 120 cm. Estimate the percentage of players whose preferred height it accommodates.

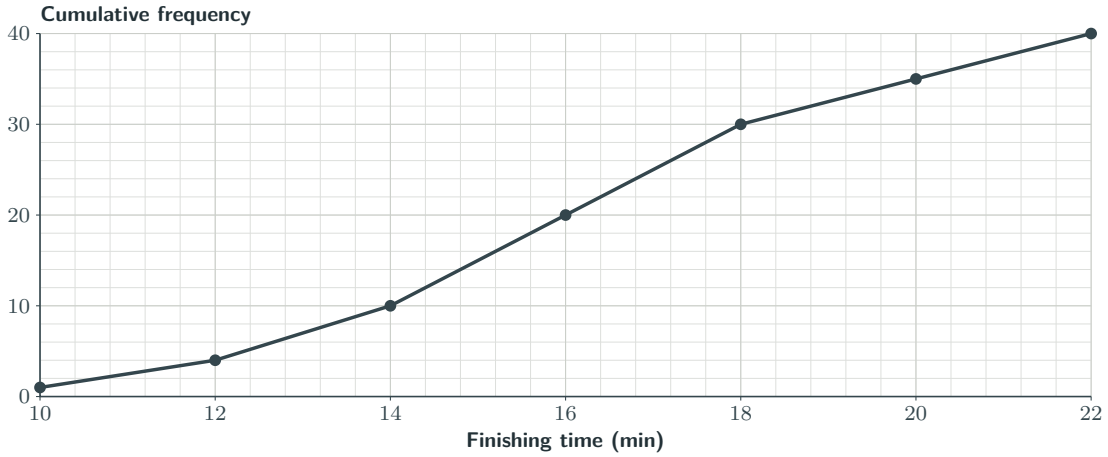
30. For 20 visits to a gallery, 12 lasted less than 10 minutes and all 20 lasted less than 30 minutes. A student joins $(10, 12)$ directly to $(30, 20)$ on a cumulative frequency graph. In fact, no visits lasted from 10 minutes up to, but excluding, 20 minutes.

(a) Give the missing boundary point and describe the graph from 10 to 20 minutes.

(b) Using straight lines between successive boundary points, estimate the 80th percentile on the student's graph and on the corrected graph.

SECTION 3 | Paper 1 practice Q31–33

31. A coach records the finishing times of 40 runners. The graph joins selected counts of runners finishing at or before each time. Its first and last points correspond to the fastest and slowest observed times.



(a) Use the graph to find the range of the finishing times. **[2]**

(b) Estimate the interquartile range. **[2]**

(c) Estimate the percentage of runners who took more than 20 minutes. **[2]**

32. Forty students record how many times they borrowed a school bicycle in one week.

Number of times	0	1	2	3	4
Frequency	4	8	12	13	3

(a) Find the number of students who borrowed a bicycle at least once. **[1]**

(b) The school office logged 80 bicycle loans that week. Determine whether this matches the students' total and give the discrepancy, if any. **[3]**

33. Fourteen bus delays are recorded to the nearest whole minute. Exactly two delays are 6 minutes and one delay is 10 minutes.

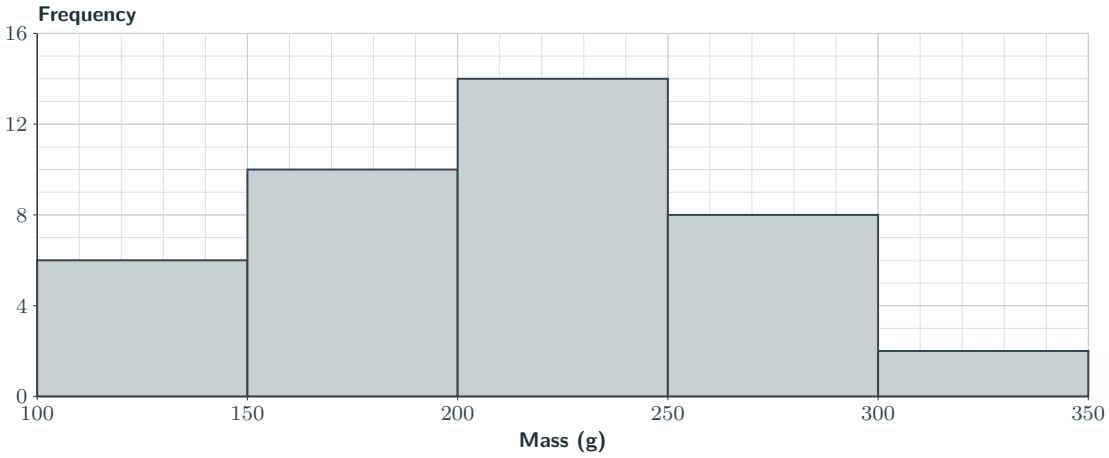
Delay d (min)	$2 \le d < 6$	$6 \le d < 10$	$10 \le d < 14$
Frequency	4	7	3

(a) Find the frequencies for revised classes $2 \le d \le 6$, $6 < d \le 10$ and $10 < d < 14$. **[3]**

(b) Using the revised table, a student adds its first two frequencies to count delays strictly below 10 minutes. Explain why this overcounts, and by how much. **[2]**

SECTION 3 | Paper 1 practice Q34–36

34. A grower weighs a sample of apples. Each bar includes its lower mass limit and excludes its upper limit.



(a) Find the percentage of apples with masses from 150 g up to, but excluding, 250 g. **[3]**

(b) Estimate how many apples have a mass of at least 225 g. State the assumption you use about the apples in the relevant class. **[3]**

35. For a sample of journey times, $Q_1 = 18$ minutes and $Q_3 = 22$ minutes. In the ordered list, the two smallest times are 11 and 12 minutes, and the two largest are 27 and 29 minutes.

(a) Use the $1.5 \times$ IQR rule to identify the outliers. **[3]**

(b) State the two times at which the whiskers end on a box-and-whisker diagram. **[2]**

36. A coach groups athletes' completion times. The table gives the percentage below 60 seconds for each group.

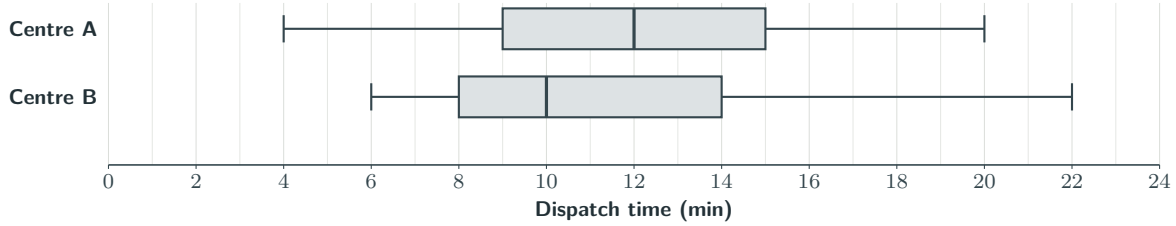
Group	Juniors	Seniors
Number of athletes	60	20
Percentage below 60 s	75%	25%

(a) Find the percentage of all 80 athletes who completed the exercise in under a minute. **[3]**

(b) A student averages 75% and 25% to obtain 50%. Explain why this method gives the wrong percentage here. **[1]**

SECTION 3 | Paper 1 practice Q37–39

37. Two dispatch centres record the time between receiving an order and sending it out.



(a) Compare the typical dispatch times using the medians. **[2]**

(b) Use the interquartile ranges to compare the spread of the middle 50%. **[2]**

(c) Which centre's diagram is more consistent with a normal distribution? Refer to the positions of the median and the ends of the whiskers. **[2]**

38. Forty-eight readers report how long they read yesterday. Of these, 35 read for at least 10 minutes, 18 for at least 20, and 6 for at least 30. All read for less than 40 minutes.

(a) Find the frequencies for the classes $0 \le t < 10$, $10 \le t < 20$, $20 \le t < 30$ and $30 \le t < 40$. **[3]**

(b) Determine the class containing the median and justify your choice. **[2]**

39. A cumulative frequency graph for 60 cycle journeys has a straight segment from 10 to 15 minutes. Exactly 15 journeys lasted more than 10 and at most 15 minutes. The graph gives a 40th percentile of 13 minutes.

(a) Find the cumulative frequency at 13 minutes. **[1]**

(b) Determine how many journeys lasted at most 10 minutes. **[3]**

(c) Hence estimate the median journey time. **[2]**

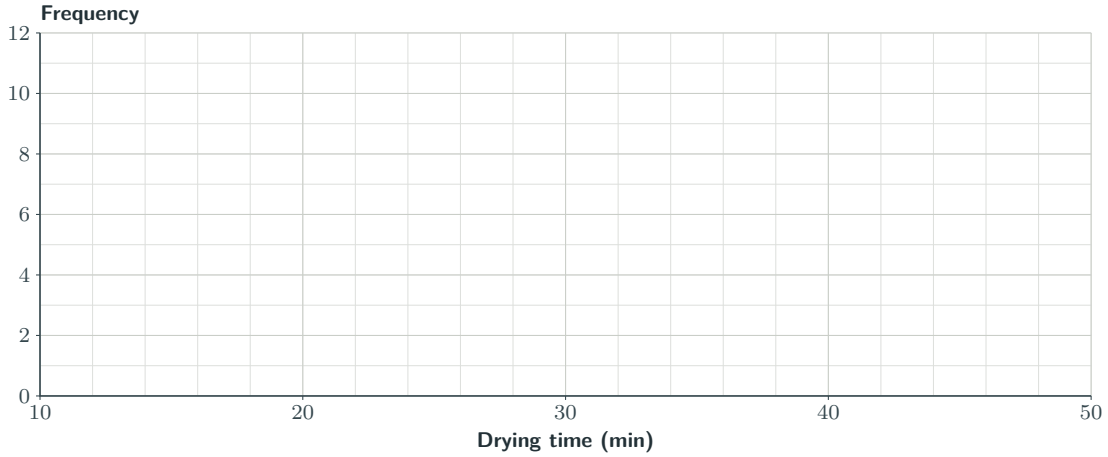
SECTION 3 | Paper 1 practice Q40–42

40. A technician records drying times, t, in minutes, for 32 paint samples. The table gives the number with $t < b$ for each boundary b.

Boundary b (min)	10	20	30	40	50
Number with $t < b$	0	6	16	28	32

(a) Find the frequencies for $20 \le t < 30$ and $30 \le t < 40$. **[2]**

(b) On the grid, draw a frequency histogram for all 32 samples. **[3]**



41. Nine archers' scores, in non-decreasing order, are $2, 4, 4, a, 7, b, 10, 10, 12$. The median is excluded when forming the lower and upper halves.

(a) Find the five-number summary of the scores. **[2]**

(b) Only three of the nine scores exceed 7. Determine b and give a reason. **[2]**

42. Fifty parcels are weighed. The table shows the cumulative totals before a recording error is corrected. Two readings recorded as 28 kg should each have been 38 kg.

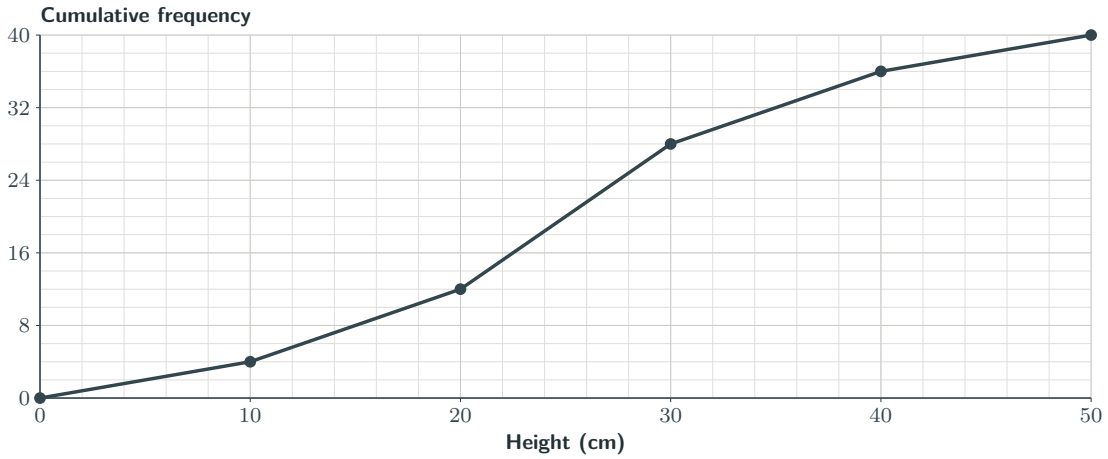
Boundary b (kg)	10	20	30	40	50
Number with mass $< b$	4	15	30	43	50

(a) Write down the five corrected cumulative totals, in order. **[2]**

(b) Find the corrected frequency for $30 \le m < 40$ and explain how the correction changed that frequency. **[2]**

SECTION 3 | Paper 1 practice Q43–45

43. A gardener measures the heights of 40 seedlings. The graph shows the cumulative frequencies, with successive points joined by straight line segments.



(a) Find the percentage of seedlings shorter than 20 cm. **[2]**

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

(c) Estimate the interquartile range of the heights. **[3]**

44. A museum records how long 40 visitors spend in one gallery.

Time t (min)	$0 \le t < 5$	$5 \le t < 10$	$10 \le t < 15$	$15 \le t < 20$
Frequency	6	14	11	9

(a) The museum redraws its frequency histogram using intervals of width 10 minutes, starting at 0. Find the heights of the two bars. **[2]**

(b) A curator says the two equal bar heights show that visits were evenly spread across all four original intervals. Use the data to evaluate this claim. **[2]**

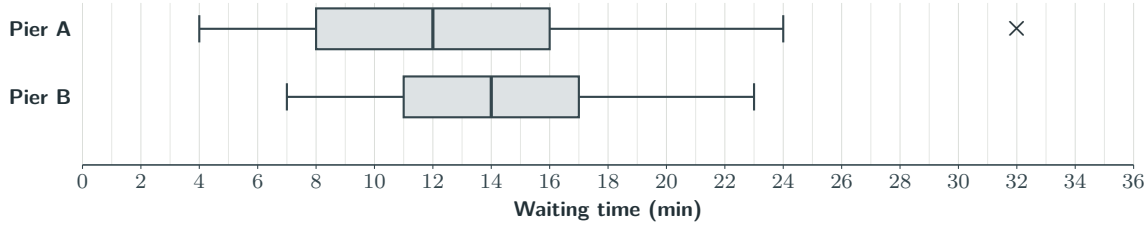
45. The middle 50% of the lengths of a batch of steel pins lie between the quartiles 24 mm and 28 mm. The smallest and largest lengths are 18 mm and 34 mm.

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

(b) An outlier lies strictly more than 1.5 interquartile ranges beyond either quartile. Determine whether either extreme value should be shown with a cross. **[3]**

SECTION 3 | Paper 1 practice Q46–48

46. The diagrams show waiting times for passengers at two ferry piers.



(a) Compare the typical waiting times and the spread of the middle 50% at the two piers. Use numerical evidence. **[4]**

(b) A passenger insists that every wait recorded at Pier A was shorter than the longest wait at Pier B. Use the diagram to evaluate this claim. **[2]**

47. In a cycling event, times for 60 finishers are grouped into four intervals of equal width. The cumulative frequencies at the successive upper boundaries are 9, 27, 51 and 60.

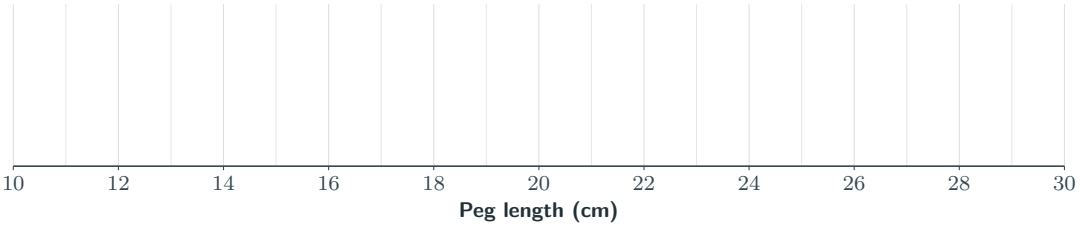
(a) Find the frequency of the third interval. **[2]**

(b) The fastest 20% of finishers qualify. In which interval does the qualifying cutoff lie? Justify your answer. **[3]**

48. A scout measures the lengths of a set of wooden pegs. The shortest is 12 cm, the lower quartile is 16 cm and the median is 20 cm. The interquartile range is 6 cm and the range is 15 cm. There are no outliers.

(a) Determine the upper quartile and the length of the longest peg. **[2]**

(b) Draw a box-and-whisker diagram on the scale below. **[3]**

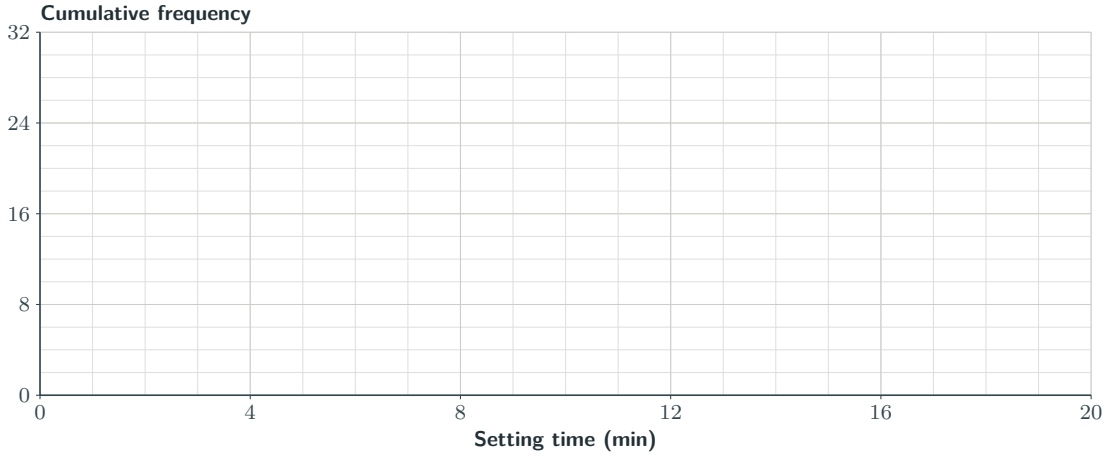


SECTION 3 | Paper 1 practice Q49–50

49. A conservator times how long 32 samples of glue take to set.

Time t (min)	$0 \le t < 4$	$4 \le t < 8$	$8 \le t < 12$	$12 \le t < 16$	$16 \le t < 20$
Frequency	4	8	8	8	4

(a) On the grid, draw a cumulative frequency graph. Join successive points with straight line segments. **[4]**



(b) Use your graph to estimate the times marking the lower and upper ends of the middle 50% of setting times. **[3]**

(c) Estimate how many samples took at least 9 minutes but less than 13 minutes to set. **[2]**

50. A factory records the masses of 20 tins to the nearest gram. All parts refer to the recorded whole-number masses.

Mass m (g)	$96 \le m < 98$	$98 \le m < 100$	$100 \le m < 102$	$102 \le m < 104$
Frequency	3	7	8	2

(a) Find the percentage of recorded masses that are at least 100 g. **[2]**

(b) Find the smallest and largest possible ranges of the recorded masses. **[3]**

(c) A tin passes inspection if its recorded mass is from 97 g to 102 g inclusive. Find the smallest and largest possible numbers that pass. Explain both bounds. **[3]**

SECTION 4 | AA HL exclusive challenge Q51–54

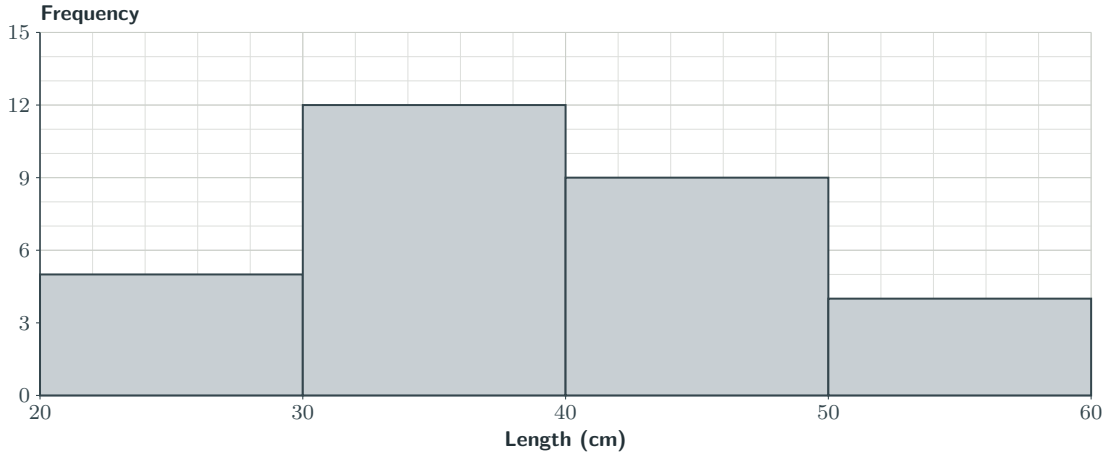
51. Forty loading times are grouped into $0 \le t < 10$, $10 \le t < 20$ and $20 \le t < 30$ minutes. On the cumulative frequency graph, successive points are joined by straight segments. The estimated median is 15 minutes and the estimated upper quartile is 20 minutes. Determine all three class frequencies.

52. Seven distinct whole-number scores have minimum 0, lower quartile 2, median 6, upper quartile 10 and maximum 12. The quartiles are the second and sixth ordered scores.

(a) Find the smallest and largest possible numbers of scores in the interval $4 \le x \le 8$. **[3]**

(b) Give one ordered list achieving each bound. **[2]**

53. A conveyor sorts 30 pieces of timber by length. The histogram records counts in classes that include their lower boundary and exclude their upper boundary.



(a) Find the tightest possible lower and upper bounds on the number of pieces with $25 \le x < 55$ cm. **[3]**

(b) Can the supplier guarantee that at least two thirds of the pieces meet this specification? Justify your answer. **[2]**

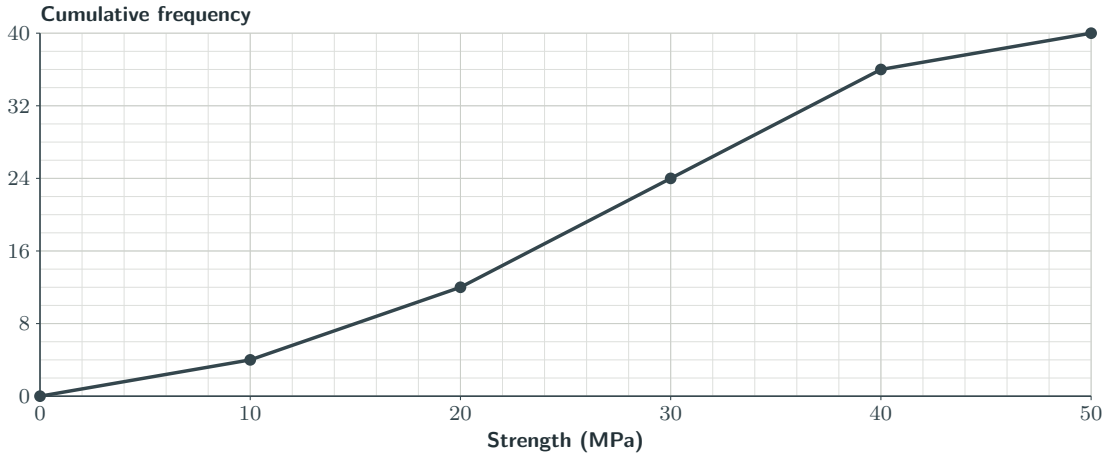
54. Seven ordered journey times, in minutes, are $4, 6, 7, 8, 10, 12, t$, where $t \ge 12$. Use the second and sixth observations as the quartiles. An outlier lies strictly more than 1.5 IQRs beyond a quartile.

(a) Find all values of t for which the upper whisker ends at 12 minutes. **[3]**

(b) Of those values, which give a box plot with no outlier? **[1]**

SECTION 4 | AA HL exclusive challenge Q55–57

55. Two laboratories test batches of the same material. The graph shows the strength measurements for the 40 samples at Laboratory A. At Laboratory B, 45% of measurements are below 30 MPa and 75% are below 40 MPa. At both laboratories all measurements are below 50 MPa.



(a) Exactly half the combined measurements are below 30 MPa. Find the number of samples at B. **[3]**

(b) Hence estimate the 90th percentile of the combined measurements, using a straight segment between the cumulative frequencies at 40 and 50 MPa. **[3]**

56. A box plot for a set of scores has lower quartile 10. Its upper whisker ends at 25 and a cross marks an outlier at 27. The upper quartile is a whole number. An upper outlier is strictly greater than $Q_3 + 1.5\,\text{IQR}$. Determine Q_3 and justify why it is unique.

57. A technician prepares a cumulative frequency graph from the table below. She then discovers that two readings of 18 seconds had each been entered as 28 seconds. Use straight segments between successive plotted points.

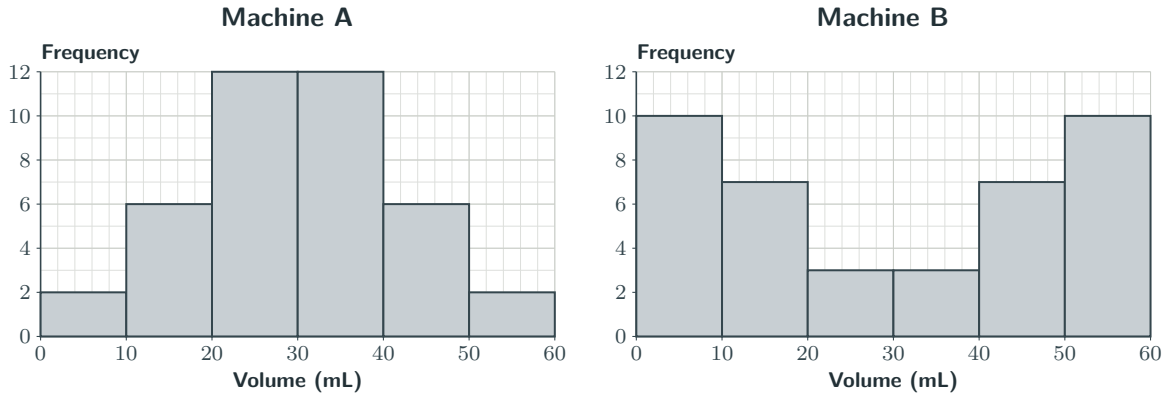
Time t (s)	$0 \le t < 10$	$10 \le t < 20$	$20 \le t < 30$	$30 \le t < 40$
Original frequency	4	8	20	8

(a) Calculate the estimated median before and after the correction. **[4]**

(b) Explain why the estimate of the 90th percentile is unchanged. **[2]**

SECTION 4 | AA HL exclusive challenge Q58–60

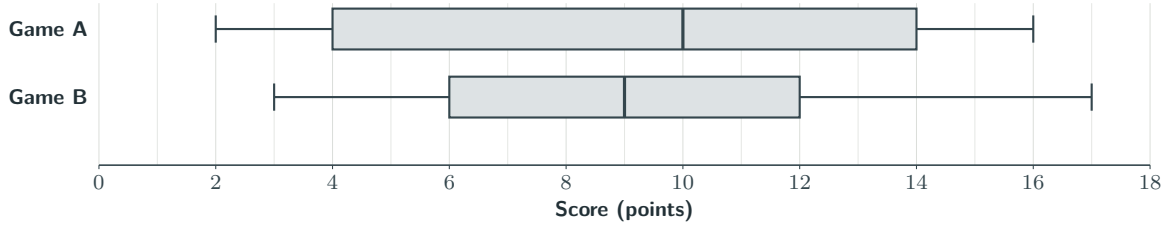
58. Two machines dispense liquid. The volumes from each machine are grouped in the same six equal-width classes.



(a) Both frequency histograms are symmetric. For which machine is an approximately normal distribution more plausible? Justify your choice using the shape of each histogram. **[3]**

(b) For one machine both quartiles on its grouped cumulative frequency graph lie at class boundaries. Identify this machine and find its estimated IQR. **[3]**

59. Seven distinct whole-number scores were recorded in each of two games. The diagrams summarise the results. A score can occur in both games.



(a) Determine how many of the 14 scores were at most 9. **[3]**

(b) Hence find the median of the combined scores. Explain why it is determined exactly. **[2]**

60. Two reports group the same 40 waiting times t, measured in minutes. One report gives frequencies 10, 18 and 12 for the classes $0 \le t < 4$, $4 \le t < 8$ and $8 \le t < 12$, respectively. The other gives 15 for $2 \le t < 6$ and 16 for $6 \le t < 10$.

(a) Let a be the frequency for $0 \le t < 2$. Express the six frequencies for the consecutive classes of width 2, starting at 0, in terms of a. **[4]**

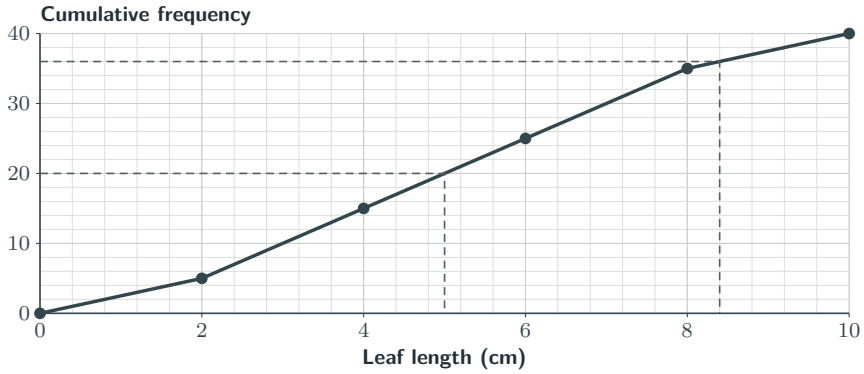
(b) Exactly half of the waiting times are below 6 minutes. Hence recover all six frequencies. **[2]**

Answer key and solution outlines

Graph readings are estimates. Allow for the precision of the scale when comparing your answers.

Section 1 | Short response

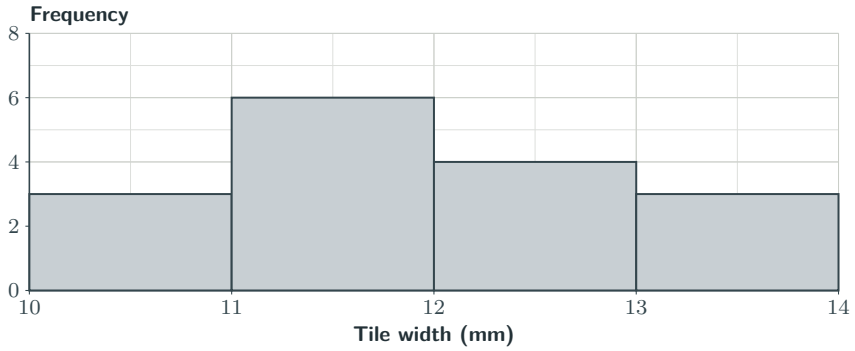
1. The numbers of trays with 0, 1, 2 and 3 germinated seeds are 2, 3, 5 and 2 respectively. The frequencies sum to 12.
2. The classes are $0 \le x < 5$, $5 \le x < 10$ and $10 \le x < 15$, respectively.
3. $7 + 4 + 1 = 12$ journeys.
4. $10 \le t < 20$, $20 \le t < 30$ and $30 \le t < 40$. A time on a shared boundary belongs to the class on its right.
5. The third and fourth bars give $10 + 7 = 17$ people.
6. 5, $5 + 7 = 12$, $12 + 0 = 12$, $12 + 8 = 20$.
7. $0.65(80) = 52$. Read horizontally from 52 on the cumulative frequency axis.
8. Median $= 8$. The lower half gives $Q_1 = (4 + 6)/2 = 5$ and the upper half gives $Q_3 = (12 + 14)/2 = 13$. Hence IQR $= 8$.
9. The largest observation is the outlier at 25 min. Range $= 25 - 6 = 19$ min.
10. $(5, 0)$ and $(10, 3)$. The second point gives the count below 10 s.
11. The median is about 5 cm at cumulative frequency 20. The 90th percentile is about 8.4 cm at cumulative frequency 36 (one fifth of the way along the final segment).



12. $25 - 18 = 7$ observations.
13. The count falls from 17 to 15. Successive cumulative totals can stay the same or increase, since each new class contributes a non-negative frequency.
14. $Q_1 = 6$ min. The middle 50% lies between 6 and 12 min.
15. The five-number summary is symmetric. This is compatible with normally distributed data, although the box plot alone does not reveal the shape within each quarter.

Section 2 | Extended response

16. **(a)** The frequencies for $10 \le w < 11$, $11 \le w < 12$, $12 \le w < 13$ and $13 \le w < 14$ are $3, 6, 4, 3$. **(b)** Each bar spans 1 mm. Its height gives the frequency.

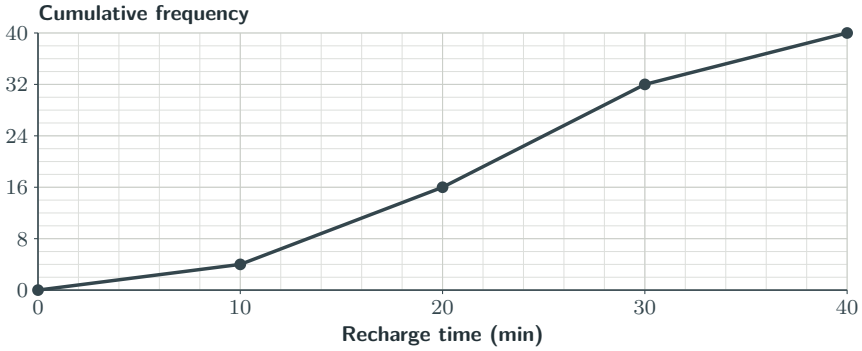


17. **(a)** Moving the last two bars left makes those deliveries look one day earlier. **(b)** Retain the one-day interval at its correct position on the scale, with zero bar height.

18. **(a)** There are 20 bookings. Bookings 10 and 11 are both for 2 tickets, so the median is 2. **(b)** There are $6 + 18 + 9 + 8 = 41$ tickets. The two four-ticket bookings account for $8/41$ of all tickets.

19. The endpoints 10 and 50 are class boundaries. The actual minimum is at least 10 min and the maximum is less than 50 min, so the range is strictly less than 40 min. The exact minimum and maximum are needed.

20. **(a)** At upper boundaries 10, 20, 30 and 40 min the cumulative frequencies are 4, 16, 32 and 40. **(b)** Plot these points and start at $(0, 0)$.



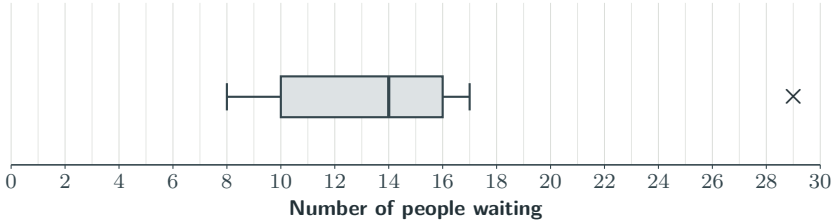
21. **(a)** At 25 min read about 24 batteries, or $24/40 = 60\%$. **(b)** At $0.9(40) = 36$, read across to the graph and down to the time axis to obtain approximately 35 min.

22. Use the exact frequency table. The count with the highest frequency is the most frequent occupancy (there can be a tie). Sum the frequencies for 8 through 12 to find how often at least eight lockers were occupied.

23. **(a)** $18 - 18 = 0$ items, since the cumulative count is unchanged. **(b)** All 18 items counted at the first boundary weigh less than 20 g. None weighs exactly 20 g because that would contribute to the 20–30 g interval.

Section 2 | Extended response CONTINUED

24. **(a)** $Q_1 = 10$, median $= 14$, $Q_3 = 16$. The IQR is 6 and the fences are $10 - 9 = 1$ and $16 + 9 = 25$. Only 29 is an outlier. **(b)** The whiskers end at 8 and 17.



25. **(a)** Squad B's median is $24 - 22 = 2$ points higher. **(b)** The IQRs are $26 - 18 = 8$ (A) and $26 - 20 = 6$ (B). B's middle half is more tightly grouped. The ranges are $30 - 15 = 15$ (A) and $34 - 12 = 22$ (B), so B has greater spread overall.

26. **(a)** In the order of the first scores, assign the second scores $5, 6, 7, 8, 9, 10$. Every score rises by one. **(b)** Assign $10, 5, 6, 7, 8, 9$. Only the pupil originally scoring 4 improves. The median rises from 6.5 to 7.5 in both pairings, so the change in the group median does not determine how many individuals improved.

27. For example, $0, 2, 3, 5, 7, 8, 10$ and $0, 2, 2, 5, 7, 8, 10$. Both have the same extremes, median and quartiles, which determine the box and whiskers here.

28. **(a)** East has $18/60 = 30\%$, compared with $16/40 = 40\%$ on West. **(b)** At 15 min the ranks are $30/60 = 50\%$ (East, 50th percentile) and $30/40 = 75\%$ (West, 75th percentile).

29. **(a)** Use the 10th and 90th percentiles. At cumulative frequencies 8 and 72, read approximately 90 cm and 130 cm. **(b)** The cumulative frequency at 120 cm is 56, giving approximately $(56 - 8)/80 = 60\%$.

30. **(a)** $(20, 12)$. The segment from 10 to 20 min is horizontal because the intervening class is empty. **(b)** $0.8(20) = 16$ is halfway from 12 to 20 on the cumulative scale. The student's graph gives 20 min, whereas the corrected graph gives 25 min, halfway between 20 and 30 min.

Section 3 | Paper 1 practice

31. **(a)** The extreme observations are 10 and 22 minutes, so the range is $22 - 10 = 12$ min. **(b)** At cumulative counts 10 and 30, read $Q_1 \approx 14$, $Q_3 \approx 18$, so IQR ≈ 4 min. **(c)** $(40 - 35)/40 = 12.5\%$.

32. **(a)** $40 - 4 = 36$ students. **(b)** The total number of loans is $1(8) + 2(12) + 3(13) + 4(3) = 83$. The office logged three fewer loans than the students reported.

33. **(a)** The frequencies are $4 + 2 = 6$, $7 - 2 + 1 = 6$, $3 - 1 = 2$. **(b)** The first two revised classes include the delay of exactly 10 min. The count is too high by one (12 instead of 11).

34. **(a)** The total is $6 + 10 + 14 + 8 + 2 = 40$. The required percentage is $100(10 + 14)/40 = 60\%$. **(b)** Half of the 200–250 g class and all heavier classes give $14/2 + 8 + 2 = 17$ apples, assuming masses are uniformly distributed within the 200–250 g class.

35. **(a)** IQR $= 4$, so the fences are $18 - 6 = 12$ and $22 + 6 = 28$ min. Mark 11 and 29 min with crosses. **(b)** Whiskers end at 12 and 27 min. The observation at the lower fence is included.

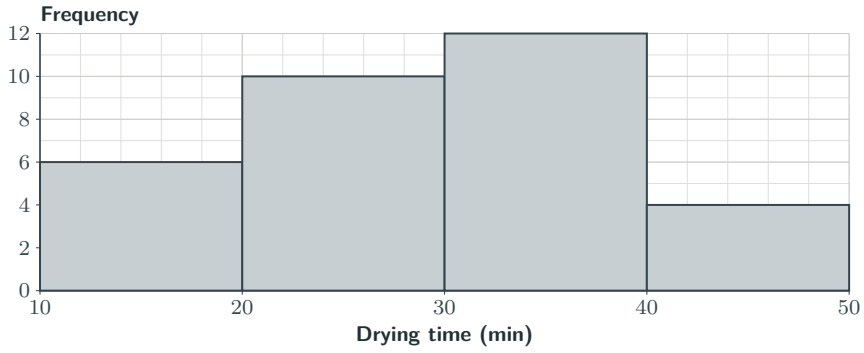
36. **(a)** $45 + 5 = 50$ finish in under a minute, giving $100(50/80) = 62.5\%$. **(b)** There are three times as many juniors as seniors, so their percentages must carry different weights.

37. **(a)** Centre B typically dispatches sooner. Its median is 10 min compared with 12 min at A. **(b)** At A, $15 - 9 = 6$ min. At B, $14 - 8 = 6$ min. The middle halves have equal spread. **(c)** A is more consistent with normality because its median is central in the box and its whiskers have equal lengths. B has a longer upper whisker.

38. **(a)** The frequencies are $48 - 35 = 13$, $35 - 18 = 17$, $18 - 6 = 12$, $6 - 0 = 6$. **(b)** $10 \le t < 20$ min. The cumulative frequencies at 10 and 20 are 13 and 30 (strictly below the boundary), so the 24th and 25th observations fall in this class.

39. **(a)** $0.4(60) = 24$. **(b)** From 10 to 13 min the graph rises by $(3/5)(15) = 9$. The cumulative frequency at 10 min is $24 - 9 = 15$. **(c)** At 15 min the cumulative count is $15 + 15 = 30 = N/2$, so the estimated median is 15 min.

40. **(a)** $16 - 6 = 10$, $28 - 16 = 12$. **(b)** Draw touching bars on $[10, 20)$, $[20, 30)$, $[30, 40)$, $[40, 50)$ with respective heights $6, 10, 12, 4$.



Section 3 | Paper 1 practice CONTINUED

41. **(a)** $(2, 4, 7, 10, 12)$. The middle pair of each half is $(4, 4)$ or $(10, 10)$. **(b)** $b = 7$. The three scores $10, 10, 12$ already exceed 7, and the order requires $b \geq 7$.

42. **(a)** $4, 15, 28, 43, 50$. Only the total below 30 kg loses the two readings. **(b)** $43 - 28 = 15$ parcels. The two corrected masses move into this interval, increasing its frequency from 13 to 15.

43. **(a)** $12/40 = 30\%$. **(b)** At cumulative frequency 20, the estimated median is 25 cm. **(c)** Reading at cumulative frequencies 10 and 30 gives about 17.5 cm and 32.5 cm. The estimated IQR is 15 cm.

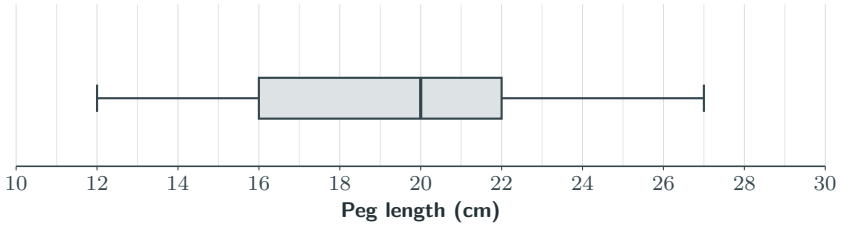
44. **(a)** $6 + 14 = 20$ and $11 + 9 = 20$ visitors. **(b)** The claim is false. For instance, 14 visitors spent 5 to under 10 min, while only 6 spent under 5 min. Combining the intervals hides that difference.

45. **(a)** IQR $= 28 - 24 = 4$ mm. Range $= 34 - 18 = 16$ mm. **(b)** The fences are $24 - 6 = 18$ mm and $28 + 6 = 34$ mm. Both extremes lie exactly on a fence, so neither is marked with a cross.

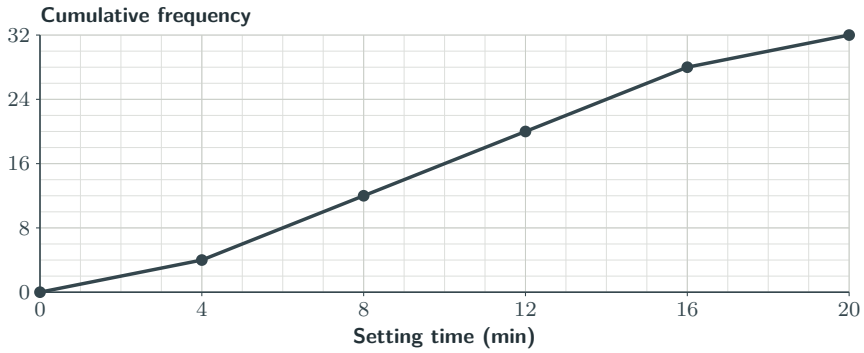
46. **(a)** At A the median is lower (12 versus 14 min), but the middle half has a larger spread (IQR $16 - 8 = 8$ versus $17 - 11 = 6$ min). **(b)** False. The 32-minute wait at A exceeds the longest wait at B, 23 min.

47. **(a)** $51 - 27 = 24$ finishers. **(b)** 20% of 60 is 12. The 12th finisher lies in the second interval because $9 < 12 \leq 27$.

48. **(a)** $Q_3 = 16 + 6 = 22$ cm. The longest peg measures $12 + 15 = 27$ cm. **(b)** Draw the box from 16 to 22, with its median at 20. Draw whiskers from 12 to 16 and from 22 to 27, with end caps at 12 and 27 (all lengths in cm).



49. **(a)** Plot and join $(0, 0), (4, 4), (8, 12), (12, 20), (16, 28), (20, 32)$. **(b)** Read at cumulative frequencies 8 and 24. The middle half runs from approximately 6 to 14 min. **(c)** The cumulative frequencies at 9 and 13 min are about 14 and 22. Their difference gives about 8 samples.



50. **(a)** $(8 + 2)/20 = 50\%$. **(b)** Smallest range $102 - 97 = 5$ g. Largest range $103 - 96 = 7$ g. **(c)** All $7 + 8 = 15$ tins in the two central classes pass. The other five can all fail (recorded as 96 or 103 g), or all pass (97 or 102 g), so the bounds are 15 and 20.

Section 4 | AA HL exclusive challenge

51. At 20 min the cumulative frequency is $0.75(40) = 30$. At 15 min it is 20. Since 15 is halfway from 10 to 20, the cumulative frequency at 10 min is $2(20) - 30 = 10$. The frequencies are $10, 20, 10$.

52. **(a)** The median 6 must count. Only the third and fifth scores can also count, giving bounds 1 and 3. **(b)** Lists attaining them are $(0, 2, 3, 6, 9, 10, 12)$ and $(0, 2, 4, 6, 8, 10, 12)$, respectively.

53. **(a)** The two inner bars guarantee $12 + 9 = 21$. All nine pieces in the outer classes could lie outside the specified interval, or all nine could lie inside. The bounds are 21 and 30. **(b)** Yes. Even the minimum 21 exceeds $\frac{2}{3}(30) = 20$.

54. **(a)** $Q_1 = 6$, $Q_3 = 12$, so the upper fence is $12 + 1.5(6) = 21$. For $12 \le t \le 21$ the upper whisker ends at t. For $t > 21$ it ends at 12. Thus $t = 12$ or $t > 21$. **(b)** Only $t = 12$ gives no outlier among those possibilities.

55. **(a)** At A, 24 are below 30 MPa. If B has n samples, $24 + 0.45n = 0.5(40 + n)$, giving $n = 80$. **(b)** At 40 MPa the combined count is $36 + 0.75(80) = 96$. At 50 MPa it is 120. The target $0.9(120) = 108$ is halfway between, so the estimate is 45 MPa.

56. Write $q = Q_3$. The upper fence is $q + 1.5(q - 10) = 2.5q - 15$. Since 25 is an inlier and 27 is an outlier, $25 \le 2.5q - 15 < 27$. Thus $16 \le q < 16.8$, whose only integer is 16.

57. **(a)** The original cumulative counts are $4, 12, 32, 40$. At count 20, the median estimate is $20 + 10(8/20) = 24$ s. Corrected counts are $4, 14, 32, 40$, giving $20 + 10(6/18) = 23\frac{1}{3}$ s. **(b)** The 90th percentile is at count 36. The endpoints $(30, 32)$ and $(40, 40)$ of that segment are unchanged, so the estimate stays at 35 s.

58. **(a)** A is more plausible. Its two central bars are highest and heights taper symmetrically. B has its lowest bars in the centre and its highest at the extremes. **(b)** Each sample has size 40. For B the cumulative counts reach 10 at 10 mL and 30 at 50 mL. Thus the estimated IQR is $50 - 10 = 40$ mL.

59. **(a)** A's median 10 has exactly three smaller scores. Since all scores are whole numbers, those three are at most 9. B's median 9 is its fourth score, giving $3 + 4 = 7$ altogether. **(b)** The seventh combined score is 9 (observed at B) and the eighth is 10 (observed at A). The median is $(9 + 10)/2 = 9.5$ points.

60. **(a)** Successively subtract the known preceding frequency from each overlapping total. The six frequencies are $a,\ 10 - a,\ 5 + a,\ 13 - a,\ 3 + a,\ 9 - a$. **(b)** The first three sum to $15 + a = 20$, so $a = 5$. The frequencies are $5, 5, 10, 8, 8, 4$.