Statistics: centre and spread

IB Mathematics AA SL & HL | Syllabus 4.3

60 questions ▪ Averages, variation and changing data

Learning goal	Calculate, compare and interpret measures of centre and spread.
Syllabus link	SL 4.3, common to AA SL and AA HL; challenges extend the same content.
Structure	Q1–15 short response; Q16–30 extended response; Q31–50 IB exam-style; Q51–60 HL exclusive; answer key at the back.
Technology	Sections 1, 2 and 4: GDC permitted. Section 3: Paper 1 / no calculator. Use technology for raw standard deviation, variance and discrete quartiles.

Instructions. Show your method and give units. Work in your exercise book. Give exact answers where practical; otherwise use 3 significant figures, unless stated differently. For descriptive statistics use σ_x (division by n), not s_x.
Quartiles. The answers use the medians of the lower and upper halves; for odd n, exclude the overall median. Different technology conventions may give different quartiles; identify your convention if it differs.

1. Short response

Questions 1–15 | Build fluency; GDC permitted.

1. Find the mean of 4, 7, 9, 12, 8.

2. Find the median of each data set. Show the ordered list you use.

(a) 3, 10, 6, 8, 4.

(b) 14, 8, 11, 7, 12, 10.

3. State the mode or modes of each data set, or state that there is no mode.

(a) 2, 4, 4, 6, 6, 7.

(b) 3, 5, 8, 10.

4. The table shows the scores obtained by 20 players in a puzzle.

Score, x	0	1	2	3	4
Frequency, f	2	5	7	4	2

Find the mean score and state the modal score.

5. Six people earn the following amounts, in euros, from a one-day event:

$$60, \quad 65, \quad 70, \quad 70, \quad 75, \quad 260.$$

Find the mean and median. Explain which better represents a typical person's earnings here.

6. The mean of 18 measurements is 12.5 cm. Find the sum of the measurements.

7. The four numbers 4, 7, 9, k have mean 8.5. Find k.

8. The masses of 20 small objects are grouped below.

Mass, m (g)	$0 \leq m < 5$	$5 \leq m < 10$	$10 \leq m < 15$	$15 \leq m < 20$
Frequency	2	8	6	4

Write down the class midpoints, estimate the mean mass and state the modal class.

9. The lower quartile, median and upper quartile of some journey times are 14, 19 and 27 minutes respectively. Find the interquartile range and state what it measures in this context.

10. Use technology to find Q_1, the median, Q_3 and the interquartile range of

$$3,\ 4,\ 6,\ 6,\ 7,\ 9,\ 11,\ 12,\ 13,\ 15,\ 16,\ 18.$$

Use the quartile convention stated on page 1.

11. Eight waiting times, in minutes, are

$$2,\quad 2,\quad 4,\quad 4,\quad 4,\quad 6,\quad 8,\quad 10.$$

Use your GDC to find the mean, standard deviation σ_x and variance. Include units.

12. Use the relationship between variance and standard deviation.

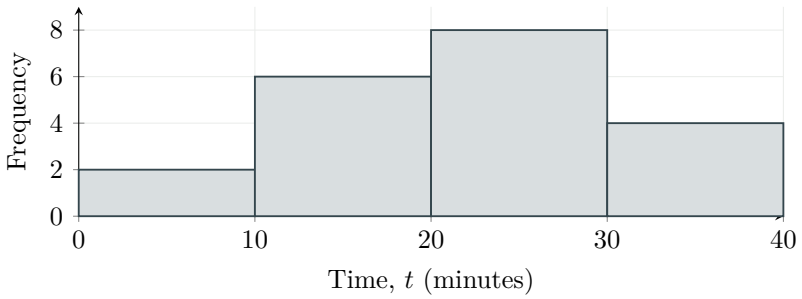
(a) Some lengths have standard deviation 2.4 cm. Find their variance.

(b) Some times have variance $18\,\text{s}^2$. Give their standard deviation in exact form.

13. A data set has mean 18, median 17, interquartile range 6 and standard deviation 3. Each data value is decreased by 5. Find the new mean, median, interquartile range and standard deviation.

14. Some distances, measured in kilometres, have mean 2.4 and variance 0.09. Find the mean, standard deviation and variance when the same distances are recorded in metres.

15. The histogram shows the time spent on a task by 20 students. All class intervals are of the form $a \leq t < b$.



State the modal class and use class midpoints to estimate the mean time.

2. Extended response

Questions 16–30 | Apply, compare and interpret; GDC permitted.

16. The recorded journey times, in minutes, for six days are

$$18, \quad 21, \quad 20, \quad 16, \quad 17, \quad 22.$$

(a) Find the mean and median of the recorded times.

(b) The entry 16 should have been 26. Find the corrected mean by adjusting the original total.

(c) Find the corrected median. Explain why the correction does not increase the mean by 10 minutes.

17. A librarian records how many books each student borrowed in one month.

Number of books, x	0	1	2	3	4
Frequency	3	6	8	2	1

(a) Find the number of students and the total number of books borrowed.

(b) Find the mean, median and mode.

(c) Use technology to find the variance and standard deviation. Give the standard deviation to three significant figures.

18. The heights h, in centimetres, of 20 seedlings are grouped below.

Height, h (cm)	$10 \le h < 14$	$14 \le h < 18$	$18 \le h < 22$	$22 \le h < 26$
Frequency	2	5	9	4

(a) Use the class midpoints to estimate the mean height.

(b) State the modal class.

(c) Explain why the mean in part (a) is an estimate. Does this estimated mean imply that exactly half the seedlings are shorter than that height? Give a reason.

19. Two counters record the waiting times, in minutes, of eight customers each.

Counter A	12	14	14	15	15	16	16	18
Counter B	8	11	13	15	15	17	19	22

(a) Find the mean waiting time at each counter.

(b) Use technology to find each standard deviation, to three significant figures. Find each interquartile range.

(c) Which counter had more consistent waiting times in these observations? Justify your answer using a measure of dispersion.

20. The number of exercise sessions completed by a group of students in one week is shown below. The mean number of sessions is $\frac{5}{3}$.

Number of sessions	0	1	2	3
Frequency	2	4	k	3

(a) Find k.
(b) Find the median and mode.
(c) One additional student reports no exercise sessions. Find the new mean and median.

21. Two groups complete the same puzzle.

Group	Number of participants	Mean completion time (minutes)
A	12	14
B	18	19

(a) Find the mean completion time for all 30 participants.
(b) A third group of 10 participants is included. The mean for all 40 participants is now 19 minutes. Find the mean completion time of the third group.
(c) Explain why simply averaging the three group means is not a valid method here.

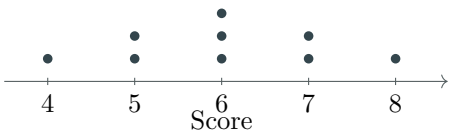
22. A temperature sensor gives the following summary of its readings. Every displayed reading is 0.8 °C too high.

Statistic	Mean	Median	Q_1	Q_3	σ_x
Value (°C)	18.6	18.4	17.9	19.2	1.7

(a) Find the corrected mean, median and quartiles.
(b) Find the corrected interquartile range, standard deviation and variance.
(c) An assistant suggests subtracting 0.8 from the standard deviation as well. Explain the error.

23. An engraving service charges $C = 2.5t + 4$ euros for a job taking t minutes. For a set of jobs, the times have mean 12, median 11, quartiles 9 and 15, and standard deviation 3 minutes.
(a) Find the mean and median charge.
(b) Find the standard deviation and variance of the charges, including their units.
(c) Find the interquartile range of the charges. Explain why the fixed charge of 4 euros does not affect this measure of spread.

24. The dot plot shows the scores in a game. Each dot represents one student.



(a) Find the mean, median and mode.
(b) Find Q_1, Q_3 and the interquartile range.
(c) The student with the highest score leaves the group. Find the mean and median of the remaining scores, and compare them with the original values.

25. A shop records the European shoe sizes sold in one afternoon.

Shoe size	36	37	38	39	40
Frequency	2	7	4	7	2

(a) Find the mean, median and all modes.

(b) A manager says, "The mean tells us which size was sold most often." Explain why this is incorrect for these data.

(c) Two further pairs of size 40 are sold. Find the new mean and state whether the median or modes change.

26. The ordered times, in minutes, spent on a task are

$$4,\quad 6,\quad 7,\quad 9,\quad 10,\quad 12,\quad 14,\quad 15,\quad 18.$$

(a) Find the median, quartiles and interquartile range. Exclude the overall median when forming the lower and upper halves.

(b) Another app reports $Q_1 = 7$ and $Q_3 = 14$ for the same observations. Explain why this alone does not show that the app is faulty. State what should be kept consistent when comparing the quartiles of different data sets.

27. Only the following grouped information about four journey times is available.

Time, t (minutes)	$0 \le t < 10$	$10 \le t < 20$
Frequency	2	2

(a) Estimate the mean using class midpoints.

(b) Write down one possible set of four observations consistent with the table that has an actual mean of 12 minutes.

(c) Can you decide from the table alone whether the midpoint estimate is greater than or less than the actual mean? Justify your answer.

28. Five players score 4, 5, 5, 6 and 10 points. A sixth player's score is added, making the mean 7 points.

(a) Find the sixth player's score.

(b) Find the median of the six scores.

(c) Use technology to find the standard deviation before and after adding the sixth score, to three significant figures. State how the spread changes.

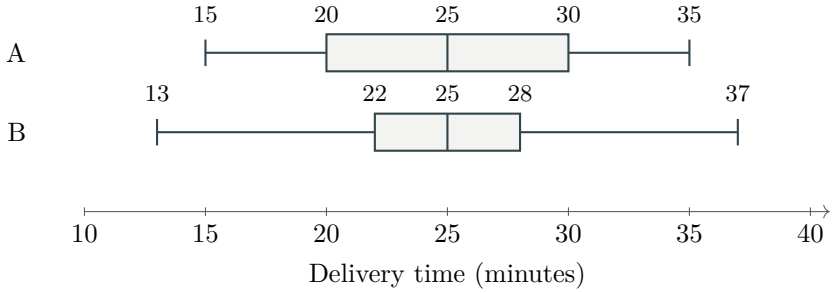
29. A set of instrument readings has mean 40 and standard deviation 5. Each reading x is transformed to $y = ax + b$, where $a > 0$. The transformed readings must have mean 50 and standard deviation 8.

(a) Find a and b.

(b) The original median is 39 and the original interquartile range is 6. Find the transformed median and interquartile range.

(c) State the variance of the transformed readings.

30. The box plots summarise delivery times for services A and B. The five displayed values on each plot are its minimum, quartiles, median and maximum.



(a) Find each service's median, interquartile range and range.

(b) Compare the spread of the central 50% of times and the overall spread.

(c) Can you determine which service has the larger standard deviation from these box plots alone? Explain.

3. IB exam-style / Paper 1

Questions 31–50 | No calculator. Original questions with mark allocations.

31. Six quiz scores are 5, 8, 8, 9, 10 and k. Their mean is 9.

(a) Find k. [2 marks]

(b) Write down the median and the mode. [2 marks]

(c) Only the score k is increased by 6. Find the new mean and explain why the median is unchanged. [2 marks]

32. The table records the number of emails received during several equal time intervals. The mean number of emails per interval is $\frac{3}{2}$.

Number of emails	0	1	2	3
Frequency	3	4	k	1

(a) Find k. [3 marks]

(b) Find the median and the mode. [2 marks]

(c) One additional interval, with no emails, is recorded. Find the new mean. [2 marks]

33. The time t minutes spent reading by each of 20 students is summarized below.

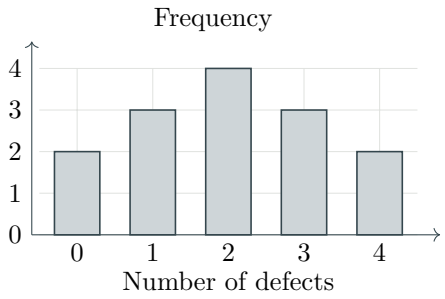
Time, t (min)	$0 \le t < 10$	$10 \le t < 20$	$20 \le t < 30$	$30 \le t < 40$
Frequency	3	8	7	2

(a) Use class midpoints to estimate the mean reading time. [3 marks]

(b) Write down the modal class. [1 mark]

(c) Explain why the exact mean cannot be obtained from this table alone. [2 marks]

34. The bar chart shows the numbers of printing defects found in a collection of notebooks.



(a) Find the number of notebooks in the collection. [1 mark]

(b) Find the mean, median and mode of the numbers of defects. [4 marks]

(c) One notebook recorded as having 4 defects actually has none. Find the corrected mean. [2 marks]

35. The recorded mean delivery time for 25 deliveries is 18 minutes. One time was entered as 13 minutes instead of 43 minutes.

(a) Find the recorded total delivery time. [1 mark]

(b) Find the corrected mean delivery time. [3 marks]

(c) A student claims that the standard deviation is unchanged because the correction only changes the mean. Explain why this conclusion is not justified. [2 marks]

36. Two classes sit the same test.

	Class A	Class B
Number of students	16	24
Mean score	12	17

(a) Find the mean score of all 40 students. [3 marks]

(b) Explain why $\frac{12+17}{2}$ does not give the combined mean. [2 marks]

(c) Every Class B score is increased by 5; Class A scores stay the same. Find the new combined mean. [2 marks]

37. Two packing lines produce bags with the following mass summaries. All entries are in grams.

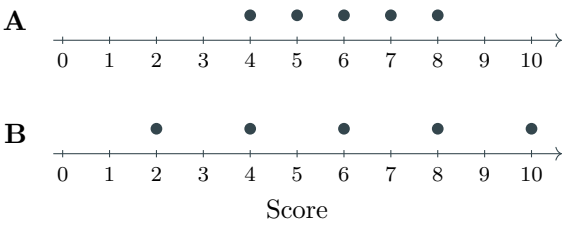
Statistic	Line A	Line B
Mean	250	250
Median	249	250
Q_1	246	248
Q_3	254	252
Standard deviation	6	3

(a) Find the interquartile range for each line. [2 marks]

(b) State which line has more consistent masses, supporting your answer with two measures of spread. [2 marks]

(c) Does this guarantee that every bag from Line B is closer to 250 g than every bag from Line A? Justify your answer. [2 marks]

38. The dot plots show two sets of five scores. Each dot represents one score.



(a) Find the mean and median of each set. [2 marks]

(b) Each score a in Set A is transformed into a score $b = 2a + c$ in Set B. Find c. [2 marks]

(c) The variance of Set A is 2. Find the exact standard deviation of Set B. [2 marks]

39. The table summarizes completion times t minutes. The estimated mean, calculated using class midpoints, is $\frac{13}{2}$ minutes.

Time, t (min)	$0 \le t < 4$	$4 \le t < 8$	$8 \le t < 12$
Frequency	2	k	3

(a) Find k. [3 marks]

(b) Write down both modal classes. [2 marks]

(c) Every completion time is doubled. Find the new estimated mean. [1 mark]

40. The annual salaries of five employees, in thousands of euros, are

$$21, \quad 22, \quad 22, \quad 25, \quad 110.$$

(a) Find the mean and median salary. [3 marks]

(b) State which of these measures better represents the salary of a typical employee in this group. Give a reason. [2 marks]

(c) The employees earning 21 and 25 thousand euros each receive a raise of 5 thousand euros. Find the new mean and median salary. [3 marks]

41. A data set has mean 10, standard deviation 2, median 9 and interquartile range 3. Each value x is replaced by $y = ax + b$, where $a > 0$. The transformed data have mean 25 and standard deviation 6.

(a) Find a. [2 marks]

(b) Find b. [2 marks]

(c) Find the median and interquartile range of the transformed data. [2 marks]

42. A set of n values has mean 12. When one further value, 18, is added, the mean becomes 13.

(a) Find n. [3 marks]

(b) Find the sum of the original values. [1 mark]

(c) The original set contains the value 5. Starting again with the original set, remove this value instead of adding 18. Find the resulting mean. [2 marks]

43. Ten real data values are grouped as follows.

Value, x	$0 \le x < 10$	$10 \le x < 20$	$20 \le x < 30$
Frequency	2	5	3

(a) Use class midpoints to estimate the mean. [3 marks]
(b) Find the smallest possible value of the true mean. [2 marks]
(c) Determine whether the true mean could equal 21, giving a reason. [2 marks]

44. The five values

$$3, \quad a, \quad 8, \quad b, \quad 17$$

are listed in non-decreasing order. Their mean is 9 and their only mode is 8.

(a) Show that $a + b = 17$. [2 marks]
(b) Find a and b. [3 marks]
(c) The largest value is removed. Find the mean and median of the remaining values. [2 marks]

45. A data set has the following summary statistics.

Statistic	Mean	Variance	Median	Q_1	Q_3
Value	12	9	11	9	15

Each data value x is replaced by $y = 30 - 2x$.

(a) Find the mean and variance of the y values. [3 marks]
(b) Find the median and interquartile range of the y values. [2 marks]
(c) Explain why the lower quartile of the y values is 0, rather than 12. [2 marks]

46. The table compares journey times on two routes. All entries are in minutes.

Statistic	Route P	Route Q
Mean	20	22
Median	18	22
Q_1	16	18
Q_3	20	26
Standard deviation	8	5

(a) Find the interquartile range for each route. [2 marks]
(b) Compare the spread of the middle 50% of journey times, and then compare overall spread using standard deviation. [2 marks]
(c) A traveller concludes that Route P is quicker on every journey. Explain why this conclusion is not justified. [2 marks]

47. Group A contains n measurements with mean 12. Group B contains $n + 4$ measurements with mean 18. The combined mean is 16.

(a) Find n. [3 marks]
(b) Four additional measurements, with mean 13, are added to the combined data. Find the new combined mean. [2 marks]
(c) Every value in the enlarged set is then multiplied by 2 and decreased by 1. Find the resulting mean. [1 mark]

48. A set of n measurements has mean 7 and standard deviation 0.

(a) Explain why every measurement must equal 7. [2 marks]

(b) One measurement of 19 is added. The new mean is 8. Find n. [3 marks]

(c) Explain why the new standard deviation is greater than 0. [1 mark]

49. A set of 10 values has mean 6 and standard deviation 2. Two new data sets are formed independently.

Set A	Multiply each original value by 3.
Set B	Make three copies of each original observation, giving 30 observations.

(a) Find the mean and standard deviation of Set A. [2 marks]

(b) Find the mean and standard deviation of Set B. [2 marks]

(c) Show that the sums of the values in the two new sets are equal, and explain why their means are different. [2 marks]

50. Fourteen students report how many films they watched in a month. The mean is $\frac{5}{2}$ films.

Number of films	1	2	3	4
Frequency	p	4	4	q

(a) Form two equations in p and q. [2 marks]

(b) Find p and q. [2 marks]

(c) Find the median and state both modes. [2 marks]

(d) One value recorded as 4 should be 1. Find the corrected mean and median. [3 marks]

4. AA HL exclusive challenge

Questions 51–60 | Optional for SL. Deeper reasoning within SL 4.3; GDC permitted.

These challenges extend the same section 4.3 ideas. Technology may be used. For calculations from summaries, you may use $\sum x_i^2 = n(\sigma^2 + \bar{x}^2)$.

51. Six observations are $2, 4, 8, 10, p, q$, where $p \leq q$. Their mean is 8 and their variance is 15.

(a) Determine p and q.

(b) Find the median and interquartile range of the six observations.

52. The following frequency distribution is symmetric about $x = 4$, where a and b are non-negative integers.

Value x	0	2	4	6	8
Frequency	a	b	6	b	a

(a) Show that the mean is 4, regardless of the values of a and b.

(b) There are 30 observations and the variance is $\frac{28}{5}$. Determine a and b.

(c) Find the median and state all modes.

53. A data set has the following summary statistics.

Mean	Median	Q_1	Q_3	Standard deviation
14	13	9	18	7

Each value x is replaced by $y = a - bx$, where $b > 0$. The new median is 22 and the new interquartile range is 18.

(a) Determine a and b.

(b) Find the mean, variance and quartiles of the new data set.

(c) Explain why applying the transformation to the original Q_1 gives the new Q_3.

54. The times, in minutes, taken by two groups to complete a task have the following summaries.

Group	Number of people	Mean time	Standard deviation
A	20	12	3
B	30	17	4

(a) Find the mean time for all 50 people.

(b) Find the variance of all 50 times, and their standard deviation to 3 significant figures.

(c) A student calculates $\frac{20(3) + 30(4)}{50} = 3.6$ as the combined standard deviation. Explain why this method fails.

55. Twelve recorded values have mean 10 and variance 16. Two entries, recorded as 4 and 16, should have been 7 and 13, respectively.

(a) Show that the corrected mean is unchanged.

(b) Determine the corrected variance and standard deviation. Give the standard deviation to 3 significant figures.

(c) Explain why the standard deviation decreases, although the mean does not change.

56. The durations t, in minutes, of 20 tasks are grouped as follows.

Duration t	$0 \le t < 10$	$10 \le t < 20$	$20 \le t < 30$	$30 \le t < 40$
Frequency	3	7	6	4

(a) Use class midpoints to estimate the mean duration.

(b) Determine the tightest interval that must contain the actual mean. Pay attention to the interval endpoints.

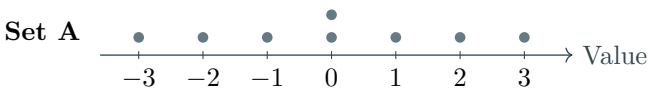
(c) Find the largest possible absolute error in the midpoint estimate. Explain whether this error can be attained.

57. Ten observations have mean 6 and variance 4. Two further observations, p and q, where $p \le q$, are added. The mean remains 6, but the variance becomes 9.

(a) Determine the exact values of p and q.

(b) Explain how the mean can stay unchanged while the variance increases.

58. The dot plot shows the eight observations in set A. Each dot represents one observation.



Set B consists of $-t, -2, -1, 0, 0, 1, 2, t$, where $t > 3$.

(a) Find the mean and variance of set A.

(b) Determine t if the standard deviation of B is twice the standard deviation of A.

(c) Find the interquartile range of each set. Explain why the IQR does not distinguish the two spreads here.

59. Seven integer scores, in non-decreasing order, are shown below.

Score	0	a	4	4	7	b	10

The mean is 5 and 4 is the unique mode.

(a) Find all possible ordered pairs (a, b).

(b) The interquartile range is 6. Determine the data set and find its standard deviation to 3 significant figures.

60. Twelve observations $x_1, x_2, \ldots, x_{12}$ have mean 8 and variance 9. For a real number c, define

$$E(c) = \sum_{i=1}^{12} (x_i - c)^2.$$

(a) Show that $E(c) = 108 + 12(c - 8)^2$.

(b) Hence find the value of c that minimizes $E(c)$, and state the minimum value. Interpret your result in terms of the mean.

(c) Each observation is replaced by $y_i = ax_i + b$, where $a > 0$. Find a and b so that the new data have mean 0 and variance 1.

Answer key and solution outlines

Use the working to check your reasoning as well as your final answer. Decimal answers are given to 3 significant figures where appropriate. Quartiles follow the convention on page 1; σ_x uses division by n.

Section 1: Short response (1–15)

1. $\bar{x} = \dfrac{4+7+9+12+8}{5} = \boxed{8}$.

2.

(a) Ordered list: $3, 4, 6, 8, 10$. Median $= \boxed{6}$.

(b) Ordered list: $7, 8, 10, 11, 12, 14$. Median $= \dfrac{10+11}{2} = \boxed{10.5}$.

3.

(a) Modes: $\boxed{4 \text{ and } 6}$; both occur twice.

(b) No mode: every value occurs once.

4. $\sum f = 20$, $\sum fx = 39$, so $\bar{x} = \dfrac{39}{20} = \boxed{1.95}$. The modal score is $\boxed{2}$, with frequency 7.

5. Mean $= \dfrac{600}{6} = \boxed{100 \text{ euros}}$; median $= \boxed{70 \text{ euros}}$. The median better represents a typical person's earnings: the unusually high 260 euros pulls the mean upwards.

6. Sum $= 18(12.5) = \boxed{225\,\text{cm}}$.

7. $4 + 7 + 9 + k = 4(8.5) = 34$, so $\boxed{k = 14}$.

8. Midpoints: $2.5, 7.5, 12.5, 17.5$ g.

$$\bar{m} \approx \frac{2(2.5) + 8(7.5) + 6(12.5) + 4(17.5)}{20} = \frac{210}{20} = \boxed{10.5\,\text{g}}.$$

Modal class: $\boxed{5 \le m < 10}$.

9. IQR $= 27 - 14 = \boxed{13 \text{ minutes}}$. It measures the spread of the middle 50% of the journey times, between the lower and upper quartiles.

10. $\boxed{Q_1 = 6}$, median $= \boxed{10}$, $\boxed{Q_3 = 14}$ and IQR $= \boxed{8}$.

11. GDC output gives $\boxed{\bar{x} = 5 \text{ min}}$, $\boxed{\sigma_x = \sqrt{7} \text{ min} \approx 2.65 \text{ min}}$ and $\boxed{\sigma_x^2 = 7 \text{ min}^2}$. Use σ_x, not s_x.

12.

(a) Variance $= (2.4)^2 = \boxed{5.76\,\text{cm}^2}$.

(b) Standard deviation $= \sqrt{18} = \boxed{3\sqrt{2}\,\text{s}}$.

13. New mean $= \boxed{13}$; median $= \boxed{12}$; IQR $= \boxed{6}$; standard deviation $= \boxed{3}$. Subtracting the same constant shifts the centre and leaves the spread unchanged.

14. Original standard deviation $= \sqrt{0.09} = 0.3$ km. Multiply each distance by 1000: new mean $= \boxed{2400\,\text{m}}$, standard deviation $= \boxed{300\,\text{m}}$ and variance $= \boxed{90000\,\text{m}^2}$.

15. Modal class: $\boxed{20 \le t < 30}$.

$$\bar{t} \approx \frac{2(5) + 6(15) + 8(25) + 4(35)}{20} = \frac{440}{20} = \boxed{22 \text{ minutes}}.$$

Section 2: Extended response (16–30)

16.

(a) Total $= 114$; mean $= 114/6 = 19$ minutes. Ordered times: $16, 17, 18, 20, 21, 22$; median $= (18+20)/2 = 19$ minutes.

(b) Replacing 16 by 26 increases the total by $26 - 16 = 10$ minutes. Corrected mean $= (114+10)/6 = \dfrac{62}{3}$ minutes.

(c) Corrected order: $17, 18, 20, 21, 22, 26$; median $= 20.5$ minutes. The increase of 10 is shared across six observations, so the mean increases by $10/6$ minutes.

17.

(a) $n = 20$ students; total books $= 0(3) + 1(6) + 2(8) + 3(2) + 4(1) = 32$.

(b) Mean $= 32/20 = 1.6$ books; median $= 2$ books; mode $= 2$ books. The 10th and 11th observations are both 2.

(c) From one-variable statistics, $\sigma_x^2 = 1.04$ books^2 and $\sigma_x = 1.02$ books (3 s.f.).

18.

(a) Midpoints: $12, 16, 20, 24$. Estimated mean $= \dfrac{2(12) + 5(16) + 9(20) + 4(24)}{20} = 19$ cm.

(b) $18 \leq h < 22$ cm.

(c) Every height in a class is represented by its midpoint; the individual heights are unknown. No: the mean is an average and need not divide the data into two equal groups.

19.

(a) Each total is 120 minutes, so both means are $120/8 = 15$ minutes.

(b) A: $\sigma_x = 1.66$ minutes, $Q_1 = 14$, $Q_3 = 16$, IQR $= 2$ minutes.
B: $\sigma_x = 4.15$ minutes, $Q_1 = 12$, $Q_3 = 18$, IQR $= 6$ minutes.

(c) Counter A: its smaller standard deviation (or smaller IQR) shows less variation in these observations, despite the equal means.

20.

(a) $\dfrac{13+2k}{9+k} = \dfrac{5}{3}$, so $39 + 6k = 45 + 5k$ and $k = 6$.

(b) There are 15 observations; the 8th is 2. Median $= 2$ sessions and mode $= 2$ sessions.

(c) The total remains 25 sessions but the count becomes 16: mean $= \dfrac{25}{16}$ sessions. The 8th and 9th observations are both 2, so the median remains 2 sessions.

21.

(a) Combined mean $= \dfrac{12(14) + 18(19)}{30} = 17$ minutes.

(b) The third group's total is $40(19) - 510 = 250$ minutes; its mean is $250/10 = 25$ minutes.

(c) The group sizes differ. Their means must be weighted by the numbers of participants; the unweighted mean $\dfrac{14+19+25}{3} = \dfrac{58}{3}$ minutes gives each group equal influence instead of each participant.

22.

(a) Subtract 0.8 from each location measure: mean $17.8\,^\circ$C, median $17.6\,^\circ$C, $Q_1 = 17.1\,^\circ$C, $Q_3 = 18.4\,^\circ$C.

(b) IQR $= 1.3\,^\circ$C; standard deviation $= 1.7\,^\circ$C; variance $= 1.7^2 = 2.89\,(^\circ\text{C})^2$.

(c) Subtracting the same constant moves all readings together. Differences from the mean, and hence the standard deviation, are unchanged.

23.

(a) Mean charge $= 2.5(12) + 4 = 34$ euros; median charge $= 2.5(11) + 4 = 31.5$ euros.

(b) Standard deviation $= 2.5(3) = 7.5$ euros; variance $= 7.5^2 = 56.25$ euros^2.

(c) IQR $= 2.5(15 - 9) = 15$ euros. Adding 4 to both quartiles leaves their difference unchanged.

Section 2: Extended response (16–30) (continued)

24.

(a) Scores: $4, 5, 5, 6, 6, 6, 7, 7, 8$; total 54. Mean $= 54/9 = 6$, median $= 6$, mode $= 6$.

(b) $Q_1 = 5$, $Q_3 = 7$; IQR $= 2$ points.

(c) Removing 8 gives mean $= (54 - 8)/8 = 5.75$ points. The 4th and 5th remaining scores are 6, so the median remains 6 points. The mean decreases; the median is unchanged.

25.

(a) Total frequency $= 22$; weighted total $= 836$. Mean $= 38$; median $= 38$; modes $= 37$ and 39.

(b) Size 38 occurred four times, whereas sizes 37 and 39 each occurred seven times. The modes identify the most frequent sizes; the mean does not.

(c) New mean $= \dfrac{836 + 2(40)}{24} = \dfrac{229}{6}$. The 12th and 13th values are both 38, so the median stays 38. The modes remain 37 and 39.

26.

(a) Median $= 10$ minutes. Excluding it, $Q_1 = (6 + 7)/2 = 6.5$ and $Q_3 = (14 + 15)/2 = 14.5$ minutes; IQR $= 8$ minutes.

(b) Different quartile conventions can give different values for the same discrete data. Use the same stated quartile convention when comparing data sets; the app's different output alone is not evidence of an error.

27.

(a) Midpoints 5 and 15 give estimated mean $= \dfrac{2(5) + 2(15)}{4} = 10$ minutes.

(b) For example, $7, 9, 15, 17$ minutes have total 48 and mean 12 minutes, with two values in each class.

(c) No. Part (b) gives an actual mean greater than 10. For example, $1, 3, 11, 13$ also fits the table but has mean 7, below 10. The midpoint estimate can therefore be either too low or too high.

28.

(a) The original total is 30; $\dfrac{30 + x}{6} = 7$ gives $x = 12$ points.

(b) The ordered scores are $4, 5, 5, 6, 10, 12$, so the median is $(5 + 6)/2 = 5.5$ points.

(c) Before: $\sigma_x = 2.10$ points. After: $\sigma_x = 2.94$ points (both 3 s.f.). The standard deviation increases, indicating greater spread about the mean.

29.

(a) Since $a > 0$, $5a = 8$, so $a = \dfrac{8}{5} = 1.6$. Then $40a + b = 50$ gives $b = -14$.

(b) Transformed median $= 1.6(39) - 14 = 48.4$; transformed IQR $= 1.6(6) = 9.6$.

(c) Transformed variance $= 8^2 = 64$.

30.

(a) A: median 25 minutes, IQR $30 - 20 = 10$ minutes, range $35 - 15 = 20$ minutes.
B: median 25 minutes, IQR $28 - 22 = 6$ minutes, range $37 - 13 = 24$ minutes.

(b) B has a less spread-out central 50% of times, since its IQR is smaller. A has the smaller overall range. Their medians are equal.

(c) No. Box plots give only five summary values, not the full distribution of observations about the mean, so the standard deviations cannot be compared conclusively.

Section 3: IB exam-style / Paper 1 (31–50)

31.

(a) $5 + 8 + 8 + 9 + 10 + k = 6(9)$, so $k = 14$.

(b) Median $= \dfrac{8 + 9}{2} = \dfrac{17}{2}$; mode $= 8$.

(c) New mean $= \dfrac{54 + 6}{6} = 10$. The middle two scores remain 8 and 9, so the median remains $\dfrac{17}{2}$.

Section 3: IB exam-style / Paper 1 (31–50) (continued)

32.

(a) $\dfrac{4+2k+3}{3+4+k+1}=\dfrac{3}{2}$, so $14+4k=24+3k$ and $k=10$.

(b) There are 18 observations; the 9th and 10th are both 2. Median $=2$; mode $=2$.

(c) The total stays 27 while the number of observations becomes 19. New mean $=\dfrac{27}{19}$ emails per interval.

33.

(a) Midpoints: 5, 15, 25, 35. Estimated mean $=\dfrac{3(5)+8(15)+7(25)+2(35)}{20}=19$ min.

(b) $10\le t<20$ minutes.

(c) The individual times within each class are unknown. The calculation replaces every time in a class by its midpoint, which need not equal the actual class mean.

34.

(a) $2+3+4+3+2=14$ notebooks.

(b) Mean $=\dfrac{0(2)+1(3)+2(4)+3(3)+4(2)}{14}=2$ defects. The 7th and 8th values are 2, so the median is 2. The highest frequency occurs at 2, so the mode is 2.

(c) The total decreases from 28 to 24. Corrected mean $=\dfrac{24}{14}=\dfrac{12}{7}$ defects.

35.

(a) Recorded total $=25(18)=450$ min.

(b) Replace 13 by 43: the total increases by $43-13=30$ min. Corrected mean $=\dfrac{450+30}{25}=\dfrac{96}{5}$ min.

(c) Only one observation changes. The rule that standard deviation is unchanged applies when the same constant is added to *every* observation; it cannot be applied to this correction.

36.

(a) Combined mean $=\dfrac{16(12)+24(17)}{40}=\dfrac{600}{40}=15$.

(b) The classes have different sizes. Their means must be weighted by their respective numbers of students; a simple average gives both classes equal weight.

(c) The total increases by $24(5)=120$. New combined mean $=\dfrac{600+120}{40}=18$.

37.

(a) Line A: $254-246=8$ g. Line B: $252-248=4$ g.

(b) Line B: its IQR is $4<8$ g and its standard deviation is $3<6$ g.

(c) No. These statistics describe distributions as a whole; they do not give the mass of every bag. The distributions may overlap, and individual exceptions are possible.

38.

(a) Set A and Set B both have mean 6 and median 6.

(b) Comparing means, $6=2(6)+c$, so $c=-6$.

(c) $\sigma_A=\sqrt{2}$; multiplying by 2 doubles standard deviation. Thus $\sigma_B=2\sqrt{2}$ (and the variance is 8).

39.

(a) Midpoints are 2, 6, 10. Thus $\dfrac{2(2)+6k+3(10)}{5+k}=\dfrac{13}{2}$, giving $68+12k=65+13k$, so $k=3$.

(b) $4\le t<8$ and $8\le t<12$ minutes, both with frequency 3.

(c) New estimated mean $=2\left(\dfrac{13}{2}\right)=13$ min.

40.

(a) Mean $=\dfrac{21+22+22+25+110}{5}=40$ thousand euros; median $=22$ thousand euros.

(b) The median better represents a typical employee here. The single salary of 110 thousand euros pulls the mean well above the salaries of the other four employees.

(c) The new total is 210 thousand euros, so the mean is 42 thousand euros. Reordering gives $22, 22, 26, 30, 110$; the new median is 26 thousand euros.

Section 3: IB exam-style / Paper 1 (31–50) (continued)

41.

(a) $6 = a(2)$ with $a > 0$, so $a = 3$.

(b) $25 = 3(10) + b$, so $b = -5$.

(c) Median $= 3(9) - 5 = 22$; IQR $= 3(3) = 9$.

42.

(a) $\dfrac{12n + 18}{n + 1} = 13$, so $12n + 18 = 13n + 13$ and $n = 5$.

(b) Original sum $= 12(5) = 60$.

(c) Removing 5 leaves sum 55 and 4 values. Mean $= \dfrac{55}{4}$.

43.

(a) Estimated mean $= \dfrac{2(5) + 5(15) + 3(25)}{10} = 16$.

(b) Use the included lower boundaries: smallest true mean $= \dfrac{2(0) + 5(10) + 3(20)}{10} = 11$.

(c) No. Every value is strictly below its class's upper boundary, so the total is strictly less than $2(10) + 5(20) + 3(30) = 210$. The mean is therefore strictly less than 21.

44.

(a) $3 + a + 8 + b + 17 = 5(9)$, so $a + b = 45 - 28 = 17$.

(b) For 8 to be the mode, $a = 8$ or $b = 8$. If $b = 8$, then $a = 9$, contradicting the stated order. Hence $a = 8$ and $b = 9$.

(c) The remaining values are $3, 8, 8, 9$. Mean $= \dfrac{28}{4} = 7$; median $= 8$.

45.

(a) Mean $= 30 - 2(12) = 6$. Variance $= (-2)^2(9) = 36$.

(b) Median $= 30 - 2(11) = 8$. IQR $= |-2|(15 - 9) = 12$.

(c) The negative multiplier reverses the order. The original upper quartile becomes the new lower quartile: $Q_{1,y} = 30 - 2(15) = 0$. Similarly, $Q_{3,y} = 30 - 2(9) = 12$.

46.

(a) Route P: $20 - 16 = 4$ min. Route Q: $26 - 18 = 8$ min.

(b) The middle 50% is less spread out on Route P, since $4 < 8$. Overall spread measured by standard deviation is smaller on Route Q, since $5 < 8$.

(c) The lower mean and median for Route P describe typical times, not every journey. The summaries do not rule out journeys on Route P taking longer than journeys on Route Q.

47.

(a) $\dfrac{12n + 18(n + 4)}{2n + 4} = 16$, so $30n + 72 = 32n + 64$ and $n = 4$.

(b) The original combined set has 12 values and sum $12(16) = 192$. Adding sum $4(13) = 52$ gives mean $\dfrac{192 + 52}{16} = \dfrac{61}{4}$.

(c) Resulting mean $= 2\left(\dfrac{61}{4}\right) - 1 = \dfrac{59}{2}$.

48.

(a) Zero standard deviation means there is no spread: every observation equals the mean, 7.

(b) $\dfrac{7n + 19}{n + 1} = 8$, so $7n + 19 = 8n + 8$ and $n = 11$.

(c) The enlarged set contains both 7 and 19, so its values are not all equal and its standard deviation is positive.

Section 3: IB exam-style / Paper 1 (31–50) (continued)

49.

(a) Set A: mean $= 3(6) = 18$; standard deviation $= 3(2) = 6$.

(b) Set B: mean $= 6$; standard deviation $= 2$. Repeating every observation the same number of times leaves the relative frequencies, mean and population standard deviation unchanged.

(c) Set A has sum $10(18) = 180$. Set B has sum $30(6) = 180$. The means differ because the equal totals are divided by different numbers of observations, 10 and 30.

50.

(a) From the frequencies, $p + q = 6$. From the mean, $p + 2(4) + 3(4) + 4q = 14\left(\frac{5}{2}\right)$, so $p + 4q = 15$.

(b) Subtracting gives $3q = 9$, so $q = 3$ and $p = 3$.

(c) The 7th and 8th observations are 2 and 3. Median $= \frac{5}{2}$; the modes are 2 and 3.

(d) The corrected total is $35 - 4 + 1 = 32$, so the mean is $\frac{16}{7}$. The corrected frequencies are $4, 4, 4, 2$; the 7th and 8th observations are both 2, so the median is 2.

Section 4: AA HL exclusive challenge (51–60)

51.

(a) The total is $6(8) = 48$, so $p + q = 24$. Also $\sum x_i^2 = 6(15 + 8^2) = 474$, giving $p^2 + q^2 = 474 - 184 = 290$. Thus $2pq = 24^2 - 290 = 286$, so p, q are the roots of $z^2 - 24z + 143 = (z - 11)(z - 13) = 0$. Hence $p = 11,\ q = 13$.

(b) The ordered data are $2, 4, 8, 10, 11, 13$. Median $= 9$; $Q_1 = 4,\ Q_3 = 11$, so IQR $= 7$.

52.

(a) $\bar{x} = \dfrac{8a + 8b + 24}{2a + 2b + 6} = 4$.

(b) The total gives $a + b = 12$. Using deviations from 4, $\dfrac{32a + 8b}{30} = \dfrac{28}{5}$, so $4a + b = 21$. Subtracting gives $a = 3$, and then $b = 9$.

(c) The 15th and 16th observations are both 4, so the median is 4. The modes are 2 and 6, each occurring 9 times.

53.

(a) The original IQR is $18 - 9 = 9$. Hence $9b = 18$, so $b = 2$. The median gives $a - 2(13) = 22$, so $a = 48$.

(b) New mean $= 48 - 2(14) = 20$; new variance $= 2^2(7^2) = 196$. The transformation reverses order, so $Q_1' = 48 - 2(18) = 12$ and $Q_3' = 48 - 2(9) = 30$.

(c) Smaller original values become larger transformed values because the multiplier is negative. The lower and upper quarters therefore exchange roles.

54.

(a) $\bar{x} = \dfrac{20(12) + 30(17)}{50} = 15$ minutes.

(b) The combined sum of squares is $20(3^2 + 12^2) + 30(4^2 + 17^2) = 12210$. Hence $\sigma^2 = \dfrac{12210}{50} - 15^2 = \dfrac{96}{5}$ min^2, and $\sigma = \sqrt{96/5} = 4.38$ minutes (3 s.f.).

(c) Standard deviations do not combine by taking a weighted average. The groups have different means, so the combined spread includes variation between those means as well as variation within each group.

55.

(a) The original and corrected pairs both sum to 20. The total stays 120, so the mean remains 10.

(b) The original sum of squares is $12(16 + 10^2) = 1392$. The corrected sum is $1392 - 4^2 - 16^2 + 7^2 + 13^2 = 1338$. Thus $\sigma^2 = \dfrac{1338}{12} - 100 = \dfrac{23}{2}$, and $\sigma = \sqrt{23/2} = 3.39$ (3 s.f.).

(c) Both corrected values are closer to the unchanged mean: deviations $-6, 6$ become $-3, 3$. The sum of squared deviations decreases.

Section 4: AA HL exclusive challenge (51–60) (continued)

56.

(a) $\bar{t} \approx \dfrac{3(5)+7(15)+6(25)+4(35)}{20} = 20.5$ minutes.

(b) Using all lower endpoints gives a mean of 15.5 minutes. The upper endpoints would give 25.5 minutes. The lower endpoints are allowed and the upper endpoints are excluded, so $15.5 \le \bar{t} < 25.5$ minutes.

(c) The largest possible absolute error is 5 minutes. It is attained when every duration equals its class's lower endpoint, giving a true mean of 15.5 minutes. Values can approach the upper endpoints, but cannot attain them.

57.

(a) The unchanged mean gives $p+q=12$. The original sum of squares is $10(4+6^2)=400$; the new sum is $12(9+6^2)=540$. Hence $p^2+q^2=140$, so $pq=(12^2-140)/2=2$. Solving $z^2-12z+2=0$ gives $p=6-\sqrt{34}$ and $q=6+\sqrt{34}$.

(b) The new values lie equally far below and above 6, so their average is 6. Their squared deviations, each 34, increase the average squared deviation and therefore the variance.

58.

(a) The mean is 0, and $\sigma_A^2 = \dfrac{2(3^2+2^2+1^2)}{8} = \dfrac{7}{2}$.

(b) Set B also has mean 0, with $\sigma_B^2 = \dfrac{2t^2+10}{8} = \dfrac{t^2+5}{4}$. Doubling the standard deviation multiplies the variance by 4, so $\frac{t^2+5}{4} = 4(\frac{7}{2}) = 14$. Thus $t=\sqrt{51}$, using $t>3$.

(c) Both sets have $Q_1=-\frac{3}{2}$, $Q_3=\frac{3}{2}$, and IQR 3. Only the smallest and largest values changed, leaving the middle half unchanged. Standard deviation uses every observation, so it detects the larger tail spread.

59.

(a) The mean gives $a+b=10$. The ordering implies $0 \le a \le 3$ and $b=10-a$. The candidates are $(0,10),(1,9),(2,8),(3,7)$. The first and last make other values occur twice, so 4 would not be the unique mode. Hence $(a,b)=(1,9)$ or $(2,8)$.

(b) With the overall median excluded, $Q_1=a$, $Q_3=b$, so $b-a=6$. Together with $a+b=10$, this gives $a=2, b=8$. The data are $0,2,4,4,7,8,10$. Their variance is $\dfrac{249}{7}-25=\dfrac{74}{7}$, so $\sigma=\sqrt{74/7}=3.25$ (3 s.f.).

60.

(a) $\sum x_i = 96$ and $\sum x_i^2 = 12(9+8^2)=876$. Therefore

$$E(c) = 876 - 192c + 12c^2 = 108 + 12(c-8)^2.$$

(b) A square is non-negative, so the minimum is 108, attained at $c=8$. The sum of squared distances from a common value is smallest when that value is the mean.

(c) The new variance is $9a^2=1$. Since $a>0$, $a=\frac{1}{3}$. The new mean is $8a+b=0$, so $b=-\frac{8}{3}$.