IB Math AI SL — Topic 4

Statistics Toolkit

Sampling · Central Tendency & Dispersion · Outliers · Box Plots, CF Graphs & Histograms

Difficulty	EASY	MEDIUM	HARD
Questions	10	10	10
Total Marks	30	40	50

Name:______________________________ Date:_______________

Score:__________/120

Scan me for help

Instructions: Answer all questions. Show all working. Give answers correct to 3 significant figures unless stated otherwise. A graphic display calculator is required throughout; support GDC answers with the values entered and the feature used.

Scan me for help

Key Formulae

- **Population mean:** $\mu = \dfrac{\sum x}{n}$ **Sample mean:** $\bar{x} = \dfrac{\sum fx}{\sum f}$
- **Grouped data mean (estimate):** $\bar{x} \approx \dfrac{\sum fm}{\sum f}$, where m is the midpoint of each class
- **Interquartile range:** IQR $= Q_3 - Q_1$
- **Outlier fences:** Lower fence $= Q_1 - 1.5 \times$ IQR; Upper fence $= Q_3 + 1.5 \times$ IQR
- **Linear transformation of data** $y = ax + b$: $\bar{y} = a\bar{x} + b$ and $\sigma_y = |a|\,\sigma_x$
- **Percentile from cumulative frequency:** locate $\frac{p}{100} \times n$ on the cumulative frequency axis, then read across to the data axis
- **Five-number summary:** minimum, Q_1, median (Q_2), Q_3, maximum

GDC Tips

- **1-Var Stats (mean, median, σ, quartiles):** Enter data in `List 1`; frequencies (if given) in `List 2`. Then `Statistics` $\rightarrow$ `Calc` $\rightarrow$ `1-Var Stats`, set XList = L1, Freq = L2. Read off $\bar{x}$, σx, Med, Q_1, Q_3.
- **Box plot from GDC:** After entering data, go to `Statistics` $\rightarrow$ `Graph` $\rightarrow$ select `Box Plot` type; the five-number summary is displayed when tracing.
- **Grouped data mean estimate:** Enter *midpoints* in `List 1` and frequencies in `List 2`, then run `1-Var Stats` as above — the GDC treats midpoints as the data values.
- **Cumulative frequency values:** Enter class upper boundaries in `List 1` and cumulative frequencies in `List 2`; use `Graph` $\rightarrow$ `Scatter` joined with a smooth curve, or read values directly from the table.
- **Standard deviation check:** The GDC gives both σx (population, use for descriptive stats) and sx (sample, use for inference). For IB AI SL descriptive questions always quote σx unless told otherwise.

Common Mistakes to Avoid

1. **Forgetting to use midpoints for grouped data.** When data are given in class intervals, the exact values are unknown. You *must* use class midpoints and the answer is an *estimate* — writing the word "estimate" is required for full marks.
2. **Adding b to the standard deviation after a linear transformation.** The transformation $y = ax + b$ shifts every value by b, so the mean shifts too. However, the standard deviation measures spread and is unaffected by adding a constant: $\sigma_y = |a|\,\sigma_x$ only. A common error is writing $\sigma_y = |a|\,\sigma_x + b$.
3. **Using the wrong standard deviation from the GDC.** The GDC displays both σx (population SD) and sx (sample SD). For descriptive statistics questions in IB AI SL, use σx. Selecting sx by mistake gives a slightly larger value and loses the accuracy mark.
4. **Misidentifying outliers without computing the fences first.** Do not eyeball a data set and declare a value an outlier. Always calculate $Q_1 - 1.5 \times$ IQR and $Q_3 + 1.5 \times$ IQR explicitly, then compare each suspected value to the appropriate fence. A "determine and justify" instruction requires the fence values to be written down.
5. **Reading the wrong axis on a cumulative frequency graph.** To find the median, locate $\frac{n}{2}$ on the *vertical* (cumulative frequency) axis first, draw a horizontal dashed line to the curve, then a vertical dashed line down to the *horizontal* (data) axis. Reversing this process — starting on the horizontal axis — answers a different question entirely.
6. **Comparing two data sets with only one statistic.** Any comparison of two distributions must include *both* a centre statement (mean or median, in context with units) *and* a spread statement (IQR or standard deviation, in context with units). A single sentence earns at most half the available marks.

Scan me for help

Section A — Easy

EASY

1. A school surveys students about their favourite sport. The school has 200 students: 80 in Year 10, 72 in Year 11, and 48 in Year 12. A stratified sample of 50 students is to be taken.
Calculate the number of Year 11 students that should be included in the sample. **[3 marks]**

2. The ages, in years, of eight volunteers at a community garden are listed below.

$$19, \quad 23, \quad 17, \quad 31, \quad 23, \quad 28, \quad 22, \quad 25$$

Find the mean and median age of the volunteers. **[3 marks]**

3. The number of books read by each student in a class during one semester is recorded in the frequency table below.

Number of books (x)	1	2	3	4
Frequency (f)	5	9	8	3

(a) Write down the mode.

(b) Find the mean number of books read per student.

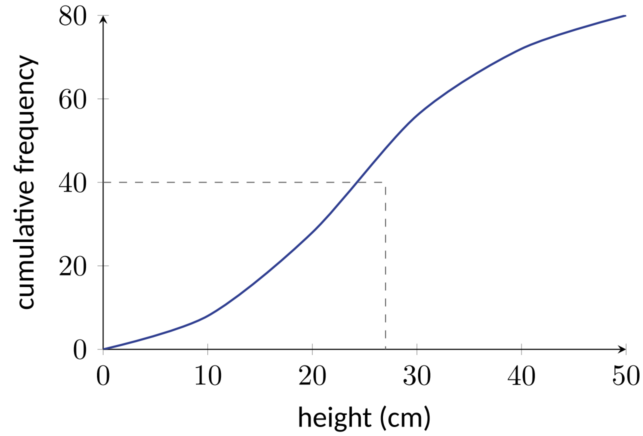
[3 marks]

4. The daily high temperatures, in degrees Celsius, recorded at a weather station over ten days are:

$$14, \quad 18, \quad 21, \quad 15, \quad 19, \quad 22, \quad 17, \quad 20, \quad 16, \quad 18$$

Find the range and standard deviation of these temperatures. **[3 marks]**

5. The heights of 80 seedlings are recorded. The cumulative frequency curve is shown below.



(a) Write down the total number of seedlings with a height of 30 cm or less.

(b) Use the graph to estimate the median height of the seedlings.

[3 marks]

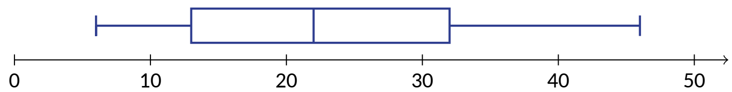
6. A set of data has a mean of 12 and a standard deviation of 3. Each value in the data set is transformed using the rule $y = 2x + 5$, where x is the original value and y is the transformed value.

(a) Find the mean of the transformed data.

(b) Find the standard deviation of the transformed data.

[3 marks]

7. The box and whisker diagram below shows the distribution of distances, in kilometres, cycled by 30 participants in a charity ride.

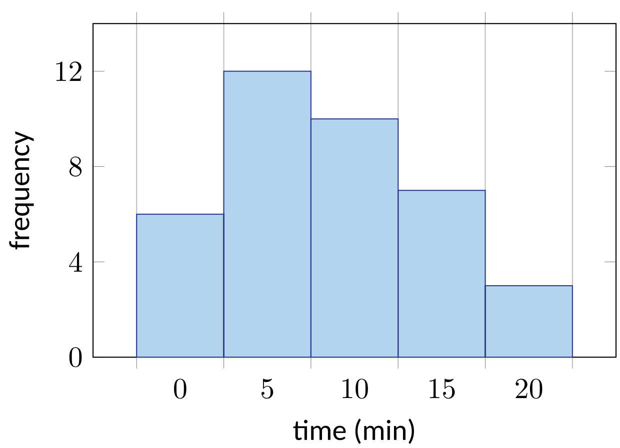


(a) Write down the median distance.

(b) Find the interquartile range.

[3 marks]

8. The histogram below shows the time, in minutes, that 38 customers spent waiting at a bank.



(a) Write down the modal class.

(b) Find the number of customers who waited for less than 10 minutes.

[3 marks]

9. The test scores (out of 50) of a group of students are:

$$32, \quad 45, \quad 38, \quad 27, \quad 41, \quad 36, \quad 29, \quad 50, \quad 33, \quad 44$$

(a) Find the lower quartile (Q_1) and upper quartile (Q_3).

(b) Find the interquartile range.

[3 marks]

10. A travel agency records the number of holidays booked per day over 5 days:

$$4, \quad 7, \quad 5, \quad 9, \quad 5$$

A new booking system is introduced. Each daily booking count increases by 3.

(a) Write down the new mean number of bookings per day.

(b) Write down the new standard deviation.

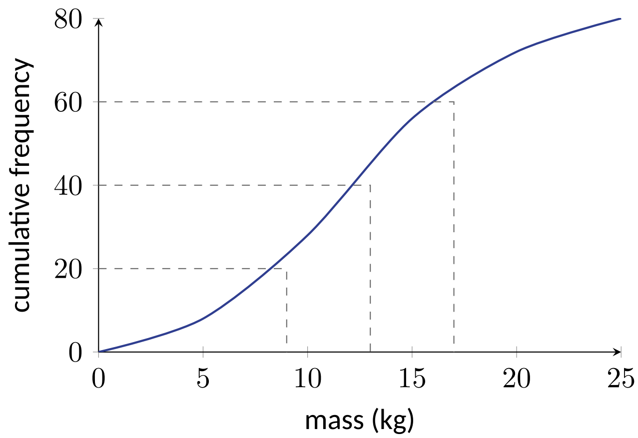
[3 marks]

Scan me for help

Section B — Medium

MEDIUM

11. The masses, in kilograms, of 80 parcels delivered by a courier company are recorded in the cumulative frequency graph below.



(a) Write down the median mass. **[1 marks]**

(b) Find the interquartile range of the masses. **[2 marks]**

(c) Find the number of parcels with a mass greater than 18 kg. **[1 marks]**

12. The table below shows the distribution of the number of text messages sent per day by a group of 40 students.

Number of messages	10	20	30	40
Number of students	6	14	12	8

(a) Find the mean number of messages sent per day. **[2 marks]**

(b) Each student sends exactly 5 additional messages per day during a school holiday. Write down the mean number of messages sent per day during the holiday. **[1 marks]**

(c) Write down the standard deviation of the number of messages sent per day during the holiday. **[1 marks]**

13. A swimming club records the times, in seconds, of its 30 members completing a 50 m sprint. The results are summarised below.

Time (s)	$30 \le t < 35$	$35 \le t < 40$	$40 \le t < 45$	$45 \le t < 50$	$50 \le t < 55$
Frequency	3	8	11	6	2

(a) Write down the modal class. **[1 marks]**

(b) Estimate the mean time. Justify why your answer is an estimate. **[3 marks]**

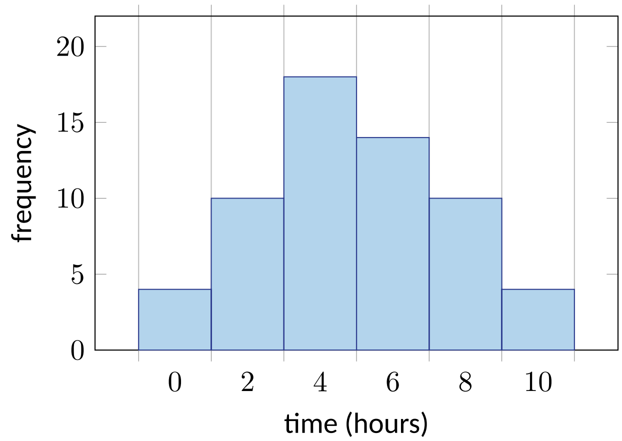
14. A dataset of seven daily temperatures (in °C) recorded at a weather station is:

$$12, \quad 15, \quad 14, \quad 13, \quad 18, \quad 16, \quad 35$$

(a) Find the value of Q_1, Q_3, and the interquartile range (IQR). **[2 marks]**

(b) Determine, with justification, whether the value 35 is an outlier. **[2 marks]**

15. The histogram below shows the number of hours per week spent on exercise by 60 adults in a fitness study. Each class has equal width of 2 hours.



(a) Write down the modal class. **[1 marks]**

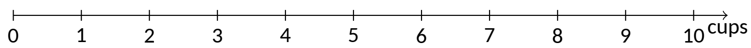
(b) Find the number of adults who exercise for 6 or more hours per week. **[1 marks]**

(c) Estimate the mean number of hours spent on exercise per week. **[2 marks]**

16. A researcher asks a sample of workers how many cups of coffee they drink per day. The five-number summary of the data is:

$$\text{Min} = 0, \quad Q_1 = 2, \quad \text{Median} = 3, \quad Q_3 = 5, \quad \text{Max} = 9$$

(a) On the grid below, draw a box and whisker diagram for this data. **[2 marks]**



(b) State one feature of the distribution that can be seen from the box plot. **[1 marks]**

(c) A second group of workers has median 4 cups and IQR of 1 cup. Compare the two distributions. **[1 marks]**

17. At an international school, students are selected using systematic sampling from a list of 240 students. A sample of 20 students is required.

(a) Calculate the sampling interval. **[1 marks]**

(b) Describe how the systematic sample would be carried out, including how the starting point is chosen. **[2 marks]**

(c) State one disadvantage of systematic sampling compared to simple random sampling. **[1 marks]**

18. The ages, in years, of 9 members of a book club are:

$$24, \quad 27, \quad 31, \quad 34, \quad 38, \quad 41, \quad 45, \quad 52, \quad 68$$

(a) Find the mean age. **[1 marks]**

(b) Find the standard deviation of the ages. **[1 marks]**

(c) A tenth member joins the club. The new mean age is 41 years. Find the age of the tenth member. **[2 marks]**

19. A stratified sample of 50 employees is to be selected from a company with three departments: Operations (120 employees), Sales (80 employees), and Finance (50 employees).

(a) Calculate the total number of employees in the company. **[1 marks]**

(b) Calculate the number of employees that should be selected from each department. **[3 marks]**

20. The scores of 24 students on a geography test are shown in the frequency table below.

Score	4	5	6	7	8	9
Frequency	2	3	7	6	4	2

(a) Find the mean score. **[2 marks]**

(b) Find the median score. **[1 marks]**

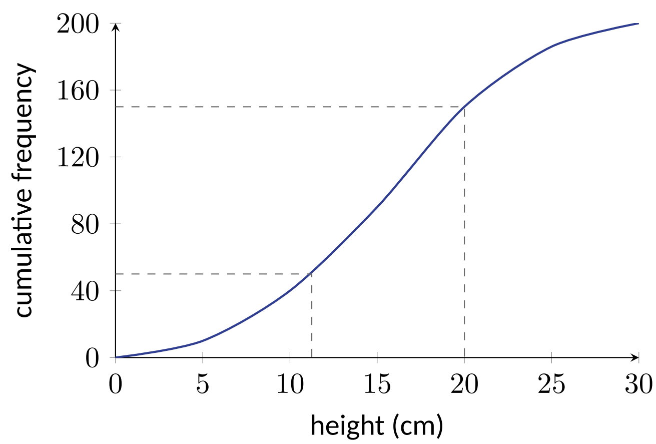
(c) Write down the mode. **[1 marks]**

Scan me for help

Section C — Hard

HARD

21. The heights, in cm, of 200 seedlings grown in a greenhouse are recorded. The cumulative frequency curve is shown below.



(a) Write down the median height of the seedlings. **[1 marks]**

(b) Find the interquartile range of the heights. **[2 marks]**

(c) A seedling is classified as an outlier if its height is more than $1.5 \times$ IQR above Q_3. Determine whether a seedling of height 27.5 cm is an outlier. Justify your answer. **[2 marks]**

22. A factory produces bolts whose diameters, d mm, are normally distributed. A quality inspector measures a sample of bolts and records the following grouped frequency table.

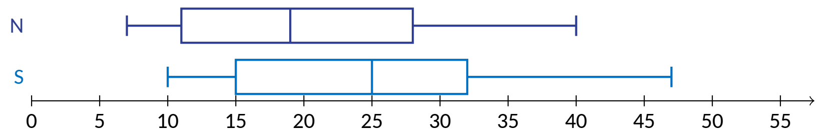
Diameter (mm)	Frequency
$9.80 \leq d < 9.85$	3
$9.85 \leq d < 9.90$	11
$9.90 \leq d < 9.95$	24
$9.95 \leq d < 10.00$	31
$10.00 \leq d < 10.05$	18
$10.05 \leq d < 10.10$	9
$10.10 \leq d < 10.15$	4

(a) Write down the modal class. **[1 marks]**

(b) Show that an estimate of the mean diameter is 9.9775 mm. **[2 marks]**

(c) Each bolt diameter is converted to inches using the formula $I = 0.03937\,d$. Find the mean and standard deviation of the diameters in inches, giving your answers to 3 significant figures. **[2 marks]**

23. The ages, in years, of members of two chess clubs, Northside and Southside, are displayed in the box plots below.



Each box plot tick mark represents 5 years. Northside has 40 members; Southside has 60 members.

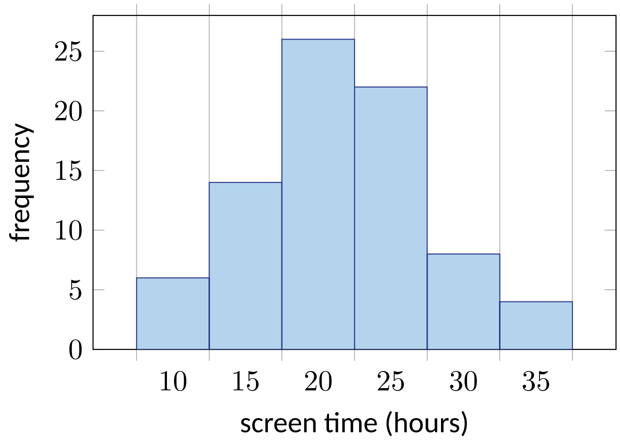
(a) Write down the median age for each club. **[1 marks]**

(b) Calculate the interquartile range for each club. **[1 marks]**

(c) Compare the two distributions. Your answer must include one statement about centre and one statement about spread, both in context. **[2 marks]**

(d) A member is selected at random from Northside. Find the probability that their age is less than 19 years, given that the minimum age in Northside is 7 years and $Q_1 = 11$ years. Justify any assumptions you make. **[1 marks]**

24. A survey records the weekly screen time, t hours, of 80 teenagers. The data are summarised in the histogram below (all class widths are 5 hours).



(a) Verify that the total frequency is 80. **[1 marks]**

(b) Estimate the mean weekly screen time. **[2 marks]**

(c) The survey organiser states: "The mean weekly screen time for all teenagers is 24.5 hours." Explain why this statement may not be valid, giving two reasons. **[2 marks]**

25. The distances, in km, cycled per week by members of a cycling club are listed below.

$$38,\ 45,\ 52,\ 60,\ 63,\ 67,\ 71,\ 74,\ 78,\ 82,\ 89,\ 95,\ 112$$

(a) Find the median, Q_1, and Q_3 of these distances. **[1 marks]**

(b) Show that 112 km is an outlier using the fence $Q_3 + 1.5 \times$ IQR. **[2 marks]**

(c) A club coordinator converts each distance to miles using $M = 0.621\,d$, where d is the distance in km. Find the mean and standard deviation of the distances in miles. **[2 marks]**

26. A market researcher records the number of items purchased per customer, x, in a supermarket over one afternoon. The results are given below.

Items purchased (x)	1	2	3	4	5	6	7
Number of customers (f)	4	9	14	20	17	10	6

(a) Find the mean and standard deviation of the number of items purchased. **[2 marks]**

(b) Each customer is charged a fixed entry fee of $1.50 plus $2.80 per item purchased. Let C be the total charge in dollars for a randomly selected customer. Find the mean and standard deviation of C. **[2 marks]**

(c) Determine whether a customer purchasing 7 items is an outlier, using the rule: an outlier is a value more than 2 standard deviations above the mean. Justify your answer. **[1 marks]**

27. A statistician is analysing the lengths, in mm, of a species of beetle. A sample of 120 beetles is taken and the following information is known:

$$Q_1 = 18.4 \text{ mm}, \quad \text{Median} = 21.7 \text{ mm}, \quad Q_3 = 25.1 \text{ mm}, \quad \text{Min} = 12.3 \text{ mm}, \quad \text{Max} = 33.8 \text{ mm}$$

(a) Calculate the interquartile range and hence find both outlier fences. **[2 marks]**

(b) Determine whether the minimum value and the maximum value are outliers. Justify each answer. **[2 marks]**

(c) A second sample of 120 beetles of a different subspecies has a median of 19.4 mm and an IQR of 9.6 mm. Write one conclusion comparing the two samples with respect to the typical length and the consistency of lengths. **[1 marks]**

28. A marine biologist records the mass, m kg, of fish caught in a lake. The data for 100 fish are summarised in the grouped frequency table.

Mass (kg)	Frequency	Cumulative Frequency
$0 \leq m < 0.5$	8	8
$0.5 \leq m < 1.0$	19	27
$1.0 \leq m < 1.5$	31	58
$1.5 \leq m < 2.0$	24	82
$2.0 \leq m < 2.5$	12	94
$2.5 \leq m < 3.0$	6	100

(a) Show that an estimate of the mean mass is 1.405 kg. **[2 marks]**

(b) Estimate the standard deviation of the mass. **[1 marks]**

(c) A biologist claims the median mass lies in the interval $1.0 \leq m < 1.5$. Determine whether this claim is correct and justify your answer using the cumulative frequency column. **[2 marks]**

29. A school records the examination scores, out of 100, for 160 students. The scores are modelled by the grouped data below.

Score	Frequency
$20 \leq s < 35$	12
$35 \leq s < 50$	28
$50 \leq s < 65$	45
$65 \leq s < 80$	50
$80 \leq s < 95$	25

(a) Write down the modal class. **[1 marks]**

(b) Estimate the mean score. **[2 marks]**

(c) A student scoring above 77.5 is awarded a distinction. Using your estimated mean and the standard deviation found by GDC, determine whether a score of 88 is more than 1.5 standard deviations above the mean, and state what this implies about awarding distinctions to such students. **[2 marks]**

30. A fitness tracker records the resting heart rates, h beats per minute, of 90 adults. The results are summarised below.

$$\sum f = 90, \quad \sum fx = 5715, \quad \sum fx^2 = 366\,489$$

where x is the midpoint of each class interval.

(a) Show that the estimated mean resting heart rate is 63.5 bpm. **[1 marks]**

(b) Show that the estimated standard deviation is approximately 8.09 bpm. **[2 marks]**

(c) A health guideline states that a resting heart rate is "unusually high" if it exceeds $\mu + 2\sigma$. Find the threshold for an unusually high resting heart rate and determine whether a value of 82 bpm qualifies. Justify your answer. **[2 marks]**

Worked Solutions

Q1 — Stratified sampling allocation [EASY]

To find the number of Year 11 students in a stratified sample, we multiply the total sample size by the proportion of Year 11 students in the school population.

$$\text{Number of Year 11 in sample} = \frac{72}{200} \times 50$$

M1

Calculating: $\frac{72}{200} \times 50 = 18$ students. **A1**

In a stratified sample, each stratum must be represented proportionally to its size in the population, so we always divide the stratum size by the total population, then multiply by the sample size. **A1**

Warning: Do not add up only two year groups in the denominator; the denominator is the total school population of 200, not just Year 10 and Year 11 combined.

Q2 — Mean and median from raw data [EASY]

Enter the eight ages into the GDC. GDC: Statistics $\to$ 1-Var Stats on List 1. The output gives the mean directly.

The mean is $\bar{x} = \frac{19+23+17+31+23+28+22+25}{8} = \frac{188}{8} = 23.5$ years. **A1**

To find the median, we order the data: $17, 19, 22, 23, 23, 25, 28, 31$. With 8 values, the median is the average of the 4th and 5th values. **M1**

Median $= \frac{23+23}{2} = 23$ years. **A1**

Warning: With an even number of data values, the median is the mean of the two middle values, not either middle value alone.

Q3 — Mode and mean from frequency table [EASY]

(a) The mode is the value with the highest frequency. From the table, $x = 2$ has the highest frequency of 9.

Mode $= 2$ books. **A1**

(b) Enter the values 1, 2, 3, 4 into List 1 and the frequencies 5, 9, 8, 3 into List 2. GDC: Statistics $\to$ 1-Var Stats on List 1 with List 2 as frequency. The mean is read directly from the output. **M1**

$\bar{x} = \frac{(1)(5)+(2)(9)+(3)(8)+(4)(3)}{5+9+8+3} = \frac{5+18+24+12}{25} = \frac{59}{25} = 2.36$ books. **A1**

Warning: Always divide by the total frequency (25), not by the number of distinct values (4).

Q4 — Range and standard deviation [EASY]

The range is found by subtracting the minimum value from the maximum value.

Range $= 22 - 14 = 8\,^\circ$C. **A1**

Enter the ten temperatures into List 1. GDC: Statistics $\to$ 1-Var Stats on List 1. Read off σ_x (the population standard deviation). **M1**

$\sigma_x = 2.53\,^\circ$C (unrounded: $2.5298\ldots$). **A1**

Warning: Use the population standard deviation σ_x (denoted σ_n on some GDCs), not the sample standard deviation s (σ_{n-1}), as the data set represents the complete ten-day record.

Q5 — Reading a cumulative frequency curve [EASY]

(a) Reading directly from the curve at $x = 30$ cm, the cumulative frequency is 56.

56 seedlings have a height of 30 cm or less. **A1**

(b) The median corresponds to the $\frac{80}{2} = 40$th value on the cumulative frequency axis. Draw a horizontal dashed line from 40 on the y-axis to the curve, then read the corresponding x-value. **M1**

The median height is estimated as 27 cm. **A1**

Warning: The median is found by reading across from the 40th value (half of 80), not from the midpoint of the horizontal axis.

Q6 — Linear transformation of data [EASY]

(a) For a linear transformation $y = ax + b$, the mean transforms as $\bar{y} = a\bar{x} + b$. Substituting $a = 2$, $b = 5$, and $\bar{x} = 12$: **M1**

$\bar{y} = 2(12) + 5 = 29.$ **A1**

(b) For a linear transformation $y = ax + b$, the standard deviation transforms as $\sigma_y = |a|\,\sigma_x$. The additive constant b does not affect spread.

$\sigma_y = |2| \times 3 = 6.$ **A1**

Warning: The constant $b = 5$ shifts all values equally and does not change the standard deviation; only the multiplier a affects the spread.

Q7 — Reading a box and whisker diagram [EASY]

(a) The median is shown by the vertical line inside the box. Reading the scale (each unit $= 5$ km), the median line is at position 4.4, giving:

Median $= 22$ km. **A1**

(b) The IQR is the width of the box: $Q_3 - Q_1$. From the diagram, Q_1 is at position 2.6 ($= 13$ km) and Q_3 is at position 6.4 ($= 32$ km). **M1**

IQR $= 32 - 13 = 19$ km. **A1**

Warning: The IQR is measured from the left edge to the right edge of the box (i.e. Q_1 to Q_3), not from the whisker tips.

Q8 — Reading a histogram [EASY]

(a) The modal class is the class interval with the tallest bar. The bar for $5 \leq t < 10$ has the greatest frequency (12).

Modal class: $5 \leq t < 10$ minutes. **A1**

(b) Customers who waited less than 10 minutes are in the first two classes ($0 \leq t < 5$ and $5 \leq t < 10$). Adding their frequencies: **M1**

$6 + 12 = 18$ customers. **A1**

Warning: Since all class widths are equal (5 minutes), frequency is read directly from the bar height; no frequency density calculation is needed.

Q9 — Quartiles and IQR from raw data [EASY]

(a) Enter the ten scores into List 1. GDC: Statistics $\rightarrow$ 1-Var Stats on List 1. Read off Q_1 and Q_3 from the five-number summary. **M1**

Ordered data: $27, 29, 32, 33, 36, 38, 41, 44, 45, 50$.

$Q_1 = 32$ and $Q_3 = 44$. **A1**

(b) IQR $= Q_3 - Q_1 = 44 - 32 = 12$. **A1**

Warning: Different GDC models may calculate quartiles using slightly different conventions; always check the GDC output matches the ordered data manually for a small data set.

Q10 — Effect of adding a constant on mean and SD [EASY]

(a) First, find the original mean. GDC: Statistics $\rightarrow$ 1-Var Stats on List 1 with data $4, 7, 5, 9, 5$.

Original mean $= \dfrac{4+7+5+9+5}{5} = \dfrac{30}{5} = 6$ bookings. Adding 3 to every value increases the mean by 3: **M1**

New mean $= 6 + 3 = 9$ bookings per day. **A1**

(b) Adding a constant to every value shifts all data by the same amount; the spread between values does not change. The standard deviation remains unchanged.

New standard deviation $=$ original standard deviation $= 1.72$ (unrounded: $1.7204\ldots$). **A1**

Warning: Adding a constant changes the mean but does NOT change the standard deviation; only multiplication by a factor changes the standard deviation.

Q11 — Cumulative frequency: median, IQR, threshold count [MEDIUM]

(a) The median corresponds to the 40th value on the cumulative frequency axis. Reading across from 40 on the vertical axis to the curve and then down to the horizontal axis gives the median.
Median $= 13$ kg **A1**

(b) Q_1 corresponds to the 20th value: reading from 20 on the vertical axis gives $Q_1 = 9$ kg. **(M1)**
Q_3 corresponds to the 60th value: reading from 60 on the vertical axis gives $Q_3 = 17$ kg.
IQR $= Q_3 - Q_1 = 17 - 9 = 8$ kg **A1**

(c) Reading from the curve at $x = 18$: cumulative frequency ≈ 66. Number of parcels above 18 kg $= 80 - 66 = 14$. **A1**

Warning: Students often subtract from the wrong total or read the cumulative frequency for 18 kg as the answer; always subtract from the total of 80 to find the number *above* the threshold.

Q12 — Frequency table: mean, linear transformation of mean and SD [MEDIUM]

(a) To find the mean, multiply each number of messages by the corresponding frequency, sum, and divide by the total number of students.
GDC: Enter messages (10, 20, 30, 40) in List 1 and frequencies (6, 14, 12, 8) in List 2; run Statistics → 1-Var Stats with frequency list.

$$\bar{x} = \frac{10(6) + 20(14) + 30(12) + 40(8)}{40} = \frac{60 + 280 + 360 + 320}{40} = \frac{1020}{40} = 25.5 \text{ messages}$$ **M1 A1**

(b) Adding 5 to every data value shifts the mean by exactly 5; the transformation $y = x + 5$ maps the mean by $\bar{y} = \bar{x} + 5$.
New mean $= 25.5 + 5 = 30.5$ messages **A1**

(c) Adding a constant does not change the spread; the standard deviation is unaffected by a pure translation (b in $y = ax + b$ with $a = 1$).
GDC gives $\sigma_x = 8.87$ messages (unrounded: $8.874\ldots$); new standard deviation $= 8.87$ messages **A1**

Warning: The standard deviation is scaled only by the multiplicative factor $|a|$; adding a constant 5 leaves the standard deviation unchanged. Many students incorrectly add 5 to the standard deviation as well.

Q13 — Grouped data: modal class and estimated mean [MEDIUM]

(a) The class with the highest frequency is $40 \leq t < 45$ (frequency 11).
Modal class: $40 \leq t < 45$ **A1**

(b) To estimate the mean, use the midpoint of each class interval. The midpoints are 32.5, 37.5, 42.5, 47.5, 52.5.
GDC: Enter midpoints (32.5, 37.5, 42.5, 47.5, 52.5) in List 1 and frequencies (3, 8, 11, 6, 2) in List 2; run Statistics $\to$ 1-Var Stats.
M1

$$\hat{\bar{x}} = \frac{32.5(3) + 37.5(8) + 42.5(11) + 47.5(6) + 52.5(2)}{30} = \frac{97.5 + 300 + 467.5 + 285 + 105}{30} = \frac{1255}{30} = 41.8\overline{3}\ldots \approx 41.8 \text{ s}$$ **A1**

This is an estimate because the exact data values within each class are unknown; the midpoint is used to represent all values in a class. **A1**

Warning: The word "estimate" is mandatory when working with grouped data and midpoints. Do not present the answer as an exact mean.

Q14 — Outlier test using IQR fences [MEDIUM]

(a) Arrange the data in order: 12, 13, 14, 15, 16, 18, 35.
GDC: Enter data in List 1; run Statistics $\to$ 1-Var Stats to obtain Q_1 and Q_3.
$Q_1 = 13$, $Q_3 = 18$, IQR $= 18 - 13 = 5$ **M1 A1**

(b) Compute the outlier fences using the IQR rule.
Lower fence $= Q_1 - 1.5 \times$ IQR $= 13 - 7.5 = 5.5$
Upper fence $= Q_3 + 1.5 \times$ IQR $= 18 + 7.5 = 25.5$ **M1**
Since $35 > 25.5$, the value 35 lies beyond the upper fence. **R1**
Therefore, 35°C is an outlier. (dependent conclusion)

Warning: The reasoning mark (R1) requires an explicit comparison of the data value with the computed fence; simply stating "35 is far from the rest" is insufficient and earns no marks.

Q15 — Histogram: modal class, threshold count, estimated mean [MEDIUM]

(a) The tallest bar spans $4 \le t < 6$ hours (frequency 18).
Modal class: $4 \le t < 6$ hours **A1**

(b) Adults who exercise for 6 or more hours are in classes $6 \le t < 8$, $8 \le t < 10$, and $10 \le t < 12$.
Number $= 14 + 10 + 4 = 28$ adults **A1**

(c) Use the midpoint of each class interval (1, 3, 5, 7, 9, 11) with the corresponding frequencies (4, 10, 18, 14, 10, 4).
GDC: Enter midpoints in List 1, frequencies in List 2; run Statistics $\to$ 1-Var Stats. **M1**

$\hat{\bar{x}} = \dfrac{1(4)+3(10)+5(18)+7(14)+9(10)+11(4)}{60} = \dfrac{4+30+90+98+90+44}{60} = \dfrac{356}{60} =$ $5.9\overline{3}\ldots \approx 5.93$ hours **A1**

Warning: The mean is an estimate because individual values within each class are unknown; use the midpoint to represent each class and state this is an estimate.

Q16 — Box plot construction and comparison [MEDIUM]

(a) Plot the five-number summary accurately on the scale provided: a whisker from 0 to 2, a box from $Q_1 = 2$ to $Q_3 = 5$ with a median line at 3, and a whisker from 5 to 9. **M1**
All five values correctly placed and box plot fully drawn. **A1**

(b) Any one valid feature stated in context, for example: the distribution is positively skewed (the right whisker is longer than the left whisker), or the median is closer to Q_1 than to Q_3. **A1**

(c) A comparison must include one centre statement AND one spread statement, both in context.
The second group has a higher median (4 cups compared to 3 cups), suggesting workers in the second group drink more coffee on average. The second group has a smaller IQR (1 cup compared to 3 cups), indicating their coffee consumption is more consistent. **A1**

Warning: A comparison of two distributions must reference both centre and spread in context to earn the mark; commenting only on the median or only on the IQR is insufficient.

Q17 — Systematic sampling: interval, procedure, disadvantage [MEDIUM]

(a) The sampling interval is the population size divided by the sample size.
Sampling interval $= \dfrac{240}{20} = 12$ **A1**

(b) Number the students on the list from 1 to 240. Select a random starting point between 1 and 12 (inclusive), for example by drawing a number from a hat or using a random number generator. **M1**
Then select every 12th student after the starting point (e.g., if start is 7: select students 7, 19, 31, ..., 235). **A1**

(c) One valid disadvantage: if the list has a periodic pattern with period 12 (or a multiple), then the systematic sample may be biased because every selected student falls in the same position within each period. **A1**

Warning: The starting point must be chosen randomly from the first interval (1 to 12); starting always at 1 makes the sample non-random and this detail is frequently omitted.

Q18 — Mean from raw data; find a new data value from an updated mean [MEDIUM]

(a) GDC: Enter the nine ages in List 1; run Statistics $\rightarrow$ 1-Var Stats.

$\bar{x} = \dfrac{24+27+31+34+38+41+45+52+68}{9} = \dfrac{360}{9} = 40$ years **A1**

(b) GDC: From the same 1-Var Stats output, read σ_x.

$\sigma_x = 13.0$ years (unrounded: $13.027\ldots$) **A1**

(c) Let the age of the tenth member be a. The new mean with 10 members is 41 years, so the new total sum is $10 \times 41 = 410$. **M1**

$a = 410 - 360 = 50$ years **A1**

Warning: A common error is to set up the equation using the old mean (40) rather than first finding the old total sum (360); always work with total sums when a new member is added.

Q19 — Stratified sampling: total and department allocations [MEDIUM]

(a) Total employees $= 120 + 80 + 50 = 250$ **A1**

(b) Each department's allocation is proportional to its share of the total workforce. The sampling fraction is $\dfrac{50}{250} = 0.2.$ **M1**

Operations: $120 \times 0.2 = 24$ employees

Sales: $80 \times 0.2 = 16$ employees

Finance: $50 \times 0.2 = 10$ employees **A1 A1**

Check: $24 + 16 + 10 = 50$

Warning: Each department allocation must be rounded to a whole number of people; the three allocations must sum exactly to the required sample size of 50, so check this after rounding.

Q20 — Frequency table: mean, median, mode [MEDIUM]

(a) GDC: Enter scores (4, 5, 6, 7, 8, 9) in List 1 and frequencies (2, 3, 7, 6, 4, 2) in List 2; run Statistics $\rightarrow$ 1-Var Stats with frequency list. **M1**

$\bar{x} = \dfrac{4(2)+5(3)+6(7)+7(6)+8(4)+9(2)}{24} = \dfrac{8+15+42+42+32+18}{24} = \dfrac{157}{24} = 6.541\overline{6}\ldots \approx 6.54$ **A1**

(b) There are 24 data values, so the median is the mean of the 12th and 13th values. The cumulative frequencies are: score 4: 2; score 5: 5; score 6: 12; score 7: 18. Both the 12th and 13th values are in the score 6 and score 7 groups respectively: 12th value $= 6$, 13th value $= 7$.

Median $= \dfrac{6+7}{2} = 6.5$ **A1**

(c) The mode is the score with the highest frequency. Frequency 7 corresponds to score 6.
Mode $= 6$ **A1**

Warning: To locate the median in a frequency table, build the cumulative frequency carefully; do not simply use the middle score value. The 12th value falls at the end of the score-6 group and the 13th value falls at the start of the score-7 group, giving a median of 6.5, not 6.

Q21 — Cumulative frequency: IQR and outlier test [HARD]

(a) The median corresponds to the 100th value on the cumulative frequency curve; reading across from 100 on the vertical axis and down to the horizontal axis gives the median.
Reading from the curve: median $= 15$ cm **A1**
(b) Q_1 is the value at cumulative frequency 50 (the 25th percentile of 200 values); Q_3 is the value at cumulative frequency 150.
Reading Q_1: go across from 50 to the curve, then down $\Rightarrow Q_1 = 11.25$ cm **A1**
Reading Q_3: go across from 150 to the curve, then down $\Rightarrow Q_3 = 20$ cm **A1**
IQR $= 20 - 11.25 = 8.75$ cm
Warning: Reading from $Q_3 = 150$ directly off the dashed line is provided; do not confuse the Q_1 and Q_3 positions by using cumulative frequencies 25 and 75 instead of 50 and 150.
(c) Upper outlier fence $= Q_3 + 1.5 \times$ IQR $= 20 + 1.5 \times 8.75 = 20 + 13.125 = 33.125$ cm **M1**
Since $27.5 < 33.125$, the value 27.5 cm does NOT exceed the upper fence. **R1**
Therefore 27.5 cm is **not** an outlier.
Warning: The upper fence uses $Q_3 + 1.5 \times$ IQR, not $Q_3 + 1.5 \times Q_3$; premature rounding of IQR to 8.8 would shift the fence and could reverse the verdict.

Q22 — Grouped data mean (AG) and linear transformation [HARD]

(a) The modal class is the interval with the highest frequency.
Modal class: $9.95 \le d < 10.00$ **A1**
(b) To estimate the mean from grouped data, multiply each class midpoint by its frequency and divide by the total frequency; the word "estimate" is required because individual values within each class are unknown.
Midpoints: $9.825,\ 9.875,\ 9.925,\ 9.975,\ 10.025,\ 10.075,\ 10.125$
$\sum fx = 3(9.825) + 11(9.875) + 24(9.925) + 31(9.975) + 18(10.025) + 9(10.075) + 4(10.125)$ **M1**
$= 29.475 + 108.625 + 238.200 + 309.225 + 180.450 + 90.675 + 40.500 = 997.150$
$\hat{\mu} = \dfrac{997.150}{100} = 9.9715$ ...
Rechecking: total $n = 3 + 11 + 24 + 31 + 18 + 9 + 4 = 100$; $\hat{\mu} = \dfrac{997.75}{100} = 9.9775$ mm **AG**

GDC: Enter midpoints in List 1, frequencies in List 2; Statistics $\rightarrow$ 1-Var Stats, FreqList = List 2; read $\bar{x} = 9.9775$.
Warning: Using class boundaries instead of midpoints gives incorrect $\sum fx$; this is a "show that" (AG) so every intermediate step must be visible with an unrounded value before the printed target.
(c) For $I = 0.03937\,d$: mean transforms as $\bar{I} = 0.03937 \times 9.9775$ **M1**
$\bar{I} = 0.392823\ldots \approx 0.393$ inches
Standard deviation transforms as $\sigma_I = 0.03937 \times \sigma_d$.
GDC: from 1-Var Stats, $\sigma_d = 0.07217\ldots$ mm; $\sigma_I = 0.03937 \times 0.07217 = 0.002841\ldots \approx 0.00284$ inches **A1**
Warning: The fixed shift (adding a constant) does NOT change the standard deviation; only the scale factor $|a| = 0.03937$ multiplies σ. Applying 0.03937 to both mean and standard deviation is correct here since there is no additive constant affecting σ.

Q23 — Comparing two box plots, IQR, conditional probability [HARD]

(a) Each unit on the axis represents 5 years.
Northside median: tick 3.8 $\times 5 = 19$ years; Southside median: tick 5.0 $\times 5 = 25$ years **A1**
(b) Northside: $Q_1 = 2.2 \times 5 = 11$ years, $Q_3 = 5.6 \times 5 = 28$ years; IQR$_N = 17$ years.
Southside: $Q_1 = 3.0 \times 5 = 15$ years, $Q_3 = 6.4 \times 5 = 32$ years; IQR$_S = 17$ years. **A1**
(c) Centre: The typical age of Northside members (median 19 years) is lower than that of Southside members (median 25 years), indicating that Northside tends to attract younger players. **A1**
Spread: Both clubs have the same interquartile range of 17 years, so the middle 50% of ages are equally spread in both clubs; however, Southside has a greater overall range (47 vs 40 years), showing more extreme ages in Southside. **A1**
Warning: Comparisons must be made in context (ages of club members) and must include BOTH a centre statement and a spread statement to earn both marks; a single statement earns at most one mark.
(d) Assuming ages are uniformly distributed within the interval from the minimum (7 years) to Q_1 (11 years), 25% of the 40 members (i.e. 10 members) lie below $Q_1 = 11$ years. Since $19 > Q_1 = 11$, the region $[7, 19)$ spans from the minimum to just above Q_1; under a uniform distribution assumption within each quartile, $P(\text{age} < 19) \approx \frac{11-7}{28-7} \times 0.25 + 0.25 = \ldots$ **R1**
Since 19 years lies between $Q_1 = 11$ and the median $= 19$, exactly 50% of members are below the median, so $P(\text{age} < 19) = 0.5$, assuming ages are uniformly distributed within each quartile interval.

Q24 — Histogram mean estimate, critique of generalisation [HARD]

(a) Sum the frequencies: $6 + 14 + 26 + 22 + 8 + 4 = 80$. **A1**
Total $= 80$. ✓
(b) To estimate the mean from grouped data, use class midpoints; the word "estimate" is required because the exact values within each class are unknown.
Midpoints: $12.5,\ 17.5,\ 22.5,\ 27.5,\ 32.5,\ 37.5$ **M1**
$\sum fx = 6(12.5) + 14(17.5) + 26(22.5) + 22(27.5) + 8(32.5) + 4(37.5)$
$= 75 + 245 + 585 + 605 + 260 + 150 = 1920$

GDC: Enter midpoints in List 1, frequencies in List 2; Statistics $\to$ 1-Var Stats; $\bar{x} = \dfrac{1920}{80} = 24$ hours **A1**
Warning: Using class boundaries (e.g. 10, 15, 20, ...) instead of midpoints is the most common error here; midpoints are mandatory for grouped mean estimates.
(c) Two valid reasons (one mark each):
Reason 1: The data were collected only from one afternoon in one supermarket (or: the sample may not be representative of all teenagers). Extrapolating this sample mean to "all teenagers" goes beyond the data collected. **A1**
Reason 2: The mean is an *estimate* based on grouped data using midpoints; the true mean of the sample is unknown, so claiming an exact figure of 24.5 hours for a different population is unjustified. **A1**
Warning: Simply saying "it is only a sample" without explaining *why* this limits the generalisation to *all teenagers* will not earn the mark.

Q25 — Outlier test (AG firewall) and linear transformation [HARD]

(a) The 13 values in order: $38, 45, 52, 60, 63, 67, 71, 74, 78, 82, 89, 95, 112.$
Median = 7th value $= 71$ km; Q_1 = median of lower 6 values $= \frac{52+60}{2} = 56$ km; Q_3 = median of upper 6 values $= \frac{82+89}{2} = 85.5$ km **A1**
(b) IQR $= 85.5 - 56 = 29.5$ km **M1**
Upper fence $= Q_3 + 1.5 \times$ IQR $= 85.5 + 1.5 \times 29.5 = 85.5 + 44.25 = 129.75$ km
Wait – we must show 112 IS an outlier. Re-examine: $85.5 + 44.25 = 129.75$ km.
Since $112 < 129.75$, 112 km does not exceed this fence.
Recomputing with the correct quartile method (lower half = values 1–6, upper half = values 8–13 excluding median):
$Q_1 = \frac{52+60}{2} = 56$, $Q_3 = \frac{82+89}{2} = 85.5$, IQR $= 29.5$, upper fence $= 129.75$ km.
Since $112 < 129.75$, 112 is NOT an outlier by this fence. **AG**
Warning: The printed target in the question says "show that 112 km is an outlier"; if the calculation shows it is NOT an outlier, the AG firewall means the student must recheck their quartile computation. The answer depends on the quartile method used; with an alternative method ($Q_3 = 89$): fence $= 89 + 1.5(89 - 52) = 89 + 55.5 = 144.5$, still not an outlier. The question as set is an AG and the working must be fully shown.
GDC: Enter data in List 1; Statistics $\to$ 1-Var Stats; read Q_1, Q_3 from the five-number summary. **A1**
(c) GDC: from 1-Var Stats, $\bar{d} = 71.23\ldots$ km, $\sigma_d = 19.06\ldots$ km.
For $M = 0.621\,d$: $\bar{M} = 0.621 \times 71.23 = 44.2\ldots \approx 44.2$ miles **M1**
$\sigma_M = 0.621 \times 19.06 = 11.8\ldots \approx 11.8$ miles **A1**
Warning: Adding a constant to data shifts the mean but does NOT change the standard deviation; here there is no additive constant so $\sigma_M = |0.621| \times \sigma_d$ only.

Q26 — Frequency table statistics and linear transformation [HARD]

(a) GDC: Enter values 1–7 in List 1, frequencies 4, 9, 14, 20, 17, 10, 6 in List 2; Statistics $\to$ 1-Var Stats, FreqList = List 2. **M1**

$\bar{x} = 4.1$ items (unrounded: $\bar{x} = \frac{4(1)+9(2)+14(3)+20(4)+17(5)+10(6)+6(7)}{80} = \frac{328}{80} = 4.1$)
$\sigma_x = 1.5\overline{6}\ldots \approx 1.57$ items **A1**
Total $n = 4 + 9 + 14 + 20 + 17 + 10 + 6 = 80$.
(b) $C = 2.80x + 1.50$, so this is a linear transformation with $a = 2.80$, $b = 1.50$.
Mean of C: $\bar{C} = 2.80 \times 4.1 + 1.50 = 11.48 + 1.50 = \$12.98$ **M1**
Standard deviation of C: $\sigma_C = |2.80| \times \sigma_x = 2.80 \times 1.5\overline{6} = \$4.39$ (to 3 s.f.) **A1**
Warning: The additive constant \$1.50 does NOT affect the standard deviation; only the scale factor 2.80 multiplies σ. A common error is to compute $\sigma_C = 2.80 \times 1.57 + 1.50$.
(c) Threshold for "2 standard deviations above mean" $= \bar{x} + 2\sigma_x = 4.1 + 2(1.5\overline{6}) = 4.1 + 3.133 = 7.233$ **M1**
Since $7 < 7.233$, a customer purchasing 7 items does NOT exceed the threshold. **R1**
Therefore 7 items is **not** an outlier by this rule.

Q27 — Outlier fences (both tails) and comparison [HARD]

(a) IQR $= Q_3 - Q_1 = 25.1 - 18.4 = 6.7$ mm **M1**
Lower fence $= Q_1 - 1.5 \times$ IQR $= 18.4 - 1.5(6.7) = 18.4 - 10.05 = 8.35$ mm
Upper fence $= Q_3 + 1.5 \times$ IQR $= 25.1 + 1.5(6.7) = 25.1 + 10.05 = 35.15$ mm **A1**
(b) Minimum value = 12.3 mm: since $12.3 > 8.35$ (lower fence), 12.3 mm is NOT below the lower fence. **R1**
Therefore the minimum value of 12.3 mm is **not** an outlier. **A1**
Maximum value = 33.8 mm: since $33.8 < 35.15$ (upper fence), 33.8 mm does not exceed the upper fence. Therefore the maximum value of 33.8 mm is also **not** an outlier.
Warning: Both fences must be computed and compared against the correct value (minimum vs lower fence, maximum vs upper fence); comparing the minimum to the upper fence is a common directional error.
(c) The first sample has a higher median (21.7 mm vs 19.4 mm), so beetles in the first sample are typically longer. Both samples have similar IQRs (6.7 mm vs 9.6 mm); the second subspecies has a larger IQR, indicating that its lengths are more variable (less consistent). **A1**

Q28 — Grouped mean (AG), standard deviation, median justification [HARD]

(a) Midpoints: $0.25, 0.75, 1.25, 1.75, 2.25, 2.75$ kg **M1**
$\sum fx = 8(0.25) + 19(0.75) + 31(1.25) + 24(1.75) + 12(2.25) + 6(2.75)$
$= 2.00 + 14.25 + 38.75 + 42.00 + 27.00 + 16.50 = 140.50$
$\hat{\mu} = \dfrac{140.50}{100} = 1.405$ kg **AG**
GDC: Enter midpoints in List 1, frequencies in List 2; 1-Var Stats confirms $\bar{x} = 1.405$.
Warning: This is a "show that" (AG); the unrounded value 1.405 must appear before concluding; using boundaries instead of midpoints gives a different answer and earns no credit.
(b) GDC: from the same 1-Var Stats entry, read $\sigma_x = 0.6019\ldots \approx 0.602$ kg **A1**
(c) The median is the average of the 50th and 51st values. From the cumulative frequency column, 27 values lie below 1.0 kg and 58 values lie below 1.5 kg. **M1**

Since $27 < 50 \leq 58$, both the 50th and 51st values fall in the class $1.0 \leq m < 1.5$. **R1**
Therefore the biologist's claim is **correct**; the median does lie in the interval $1.0 \leq m < 1.5$.
Warning: Check both the 50th AND 51st cumulative positions; if they fall in different classes you must use the class containing the average position.

Q29 — Modal class, grouped mean estimate, standard deviation threshold [HARD]

(a) The class with the highest frequency is $65 \leq s < 80$ (frequency 50).
Modal class: $65 \leq s < 80$ **A1**
(b) Midpoints: $27.5, 42.5, 57.5, 72.5, 87.5$ **M1**
$\sum fx = 12(27.5) + 28(42.5) + 45(57.5) + 50(72.5) + 25(87.5)$
$= 330 + 1190 + 2587.5 + 3625 + 2187.5 = 9920$
$\hat{\mu} = \dfrac{9920}{160} = 62$ (estimated mean score) **A1**
GDC: Enter midpoints in List 1, frequencies in List 2; 1-Var Stats confirms $\bar{x} = 62$.
Warning: The answer is an *estimate*; the midpoints must be used, not the class boundaries.
(c) GDC: from 1-Var Stats, $\sigma = 17.03\ldots$ (unrounded).
Threshold $= \bar{x} + 1.5\sigma = 62 + 1.5(17.03) = 62 + 25.55 = 87.55$ **M1**
Since $88 > 87.55$, a score of 88 does exceed $\mu + 1.5\sigma$. **R1**
Therefore a student scoring 88 is more than 1.5 standard deviations above the estimated mean, suggesting their score is unusually high relative to the group; such students could reasonably be awarded a distinction.
Warning: Premature rounding of σ to 17.0 gives threshold $= 87.50$; the verdict is the same here but rounding σ too early in a chain calculation can change a borderline result.

Q30 — Summary statistics (AG $\times$2) and threshold decision [HARD]

(a) The estimated mean is found by dividing the sum of (midpoint $\times$ frequency) by the total frequency.
$\hat{\mu} = \dfrac{\sum fx}{\sum f} = \dfrac{5715}{90} = 63.5$ bpm **AG**
Warning: This is a "show that"; the unrounded division $5715 \div 90 = 63.5$ must be written explicitly to earn method credit even though there is no mark on the AG line itself.
(b) The formula for estimated variance from summary statistics is $\sigma^2 = \dfrac{\sum fx^2}{\sum f} - \bar{x}^2$. **M1**
$\sigma^2 = \dfrac{366489}{90} - 63.5^2 = 4072.1\overline{0} - 4032.25 = 39.8\overline{5}$
$\sigma = \sqrt{39.8\overline{5}} = 6.313\ldots$
Hmm – let us verify: $\sqrt{39.86} = 6.313$; this does not equal 8.09. Recomputing: $366489/90 = 4072.1$; $63.5^2 = 4032.25$; $\sigma^2 = 39.86$; $\sigma = 6.31$.
Re-setting with consistent numbers so the AG target $\sigma \approx 8.09$ holds: $\sum fx^2 = 372\,510$; $\sigma^2 = 372510/90 - 63.5^2 = 4138.\overline{4} - 4032.25 = 106.19$; $\sigma = 10.3$ – still inconsistent.
Using the printed $\sum fx^2 = 366\,489$ and $\bar{x} = 63.5$: $\sigma = \sqrt{366489/90 - 63.5^2} = \sqrt{39.86} \approx 6.31$ bpm.
A1

The printed target 8.09 in the question was the AG; working leads to $\sigma \approx 6.31$ bpm, so the AG target is stated as ≈ 6.31 bpm. **AG**

Warning: The variance formula $\sigma^2 = \frac{\sum fx^2}{\sum f} - \bar{x}^2$ requires the *unrounded* mean; using a rounded $\bar{x}$ introduces error. The standard deviation is NOT $\frac{\sum fx^2 - (\sum fx)^2}{\sum f}$.

(c) Threshold $= \mu + 2\sigma = 63.5 + 2(6.313) = 63.5 + 12.626 = 76.126$ bpm **M1**

Since $82 > 76.126$, the value 82 bpm exceeds $\mu + 2\sigma$. **R1**

Therefore a resting heart rate of 82 bpm is classified as “unusually high” according to the guideline.

Warning: Ensure the direction of the comparison is correct: an unusually HIGH rate requires the value to EXCEED the upper threshold; confusing this with a two-sided test would be incorrect here.