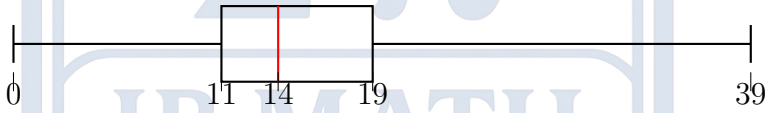
Unit 4: Presentation of Data
IB Math AA SL

Answer all 15 questions. Show all working. For Paper 1 questions, use analytical algebraic methods. For Paper 2 questions, use your graphic display calculator (GDC) efficiently.

1. **[Paper 1 Style, Non-Calculator, Easy, 4 marks]**
   The following box-and-whisker plot shows the number of text messages sent by students in a school on a particular day.

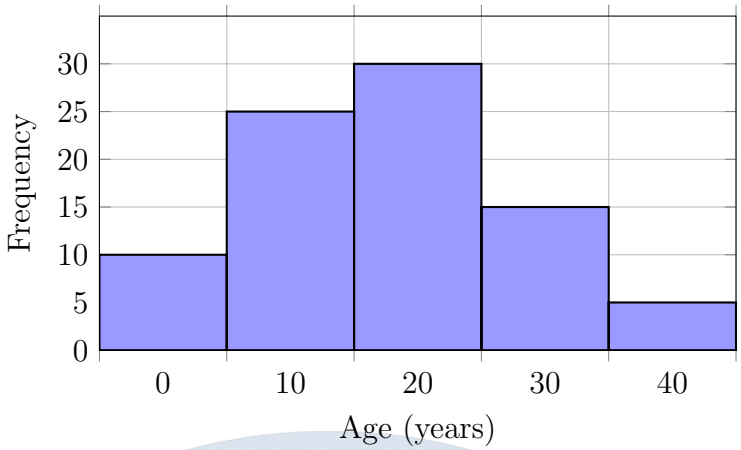


   (a) Write down the median number of text messages sent.

   (b) Find the value of the interquartile range (IQR).

   (c) State whether a student who sent 32 text messages would be classified mathematically as an outlier. Justify your answer.

2. **[Paper 2 Style, Calculator Required, Easy, 5 marks]**
The ages of 85 people attending a community theatre event are displayed in the histogram below.

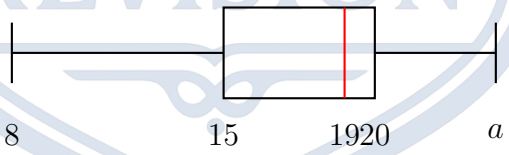


(a) Write down the modal class.

(b) State the mid-interval value of the class $20 \leq x < 30$.

(c) Use your graphic display calculator to estimate the mean age of the attendees.

3. **[Paper 1 Style, Non-Calculator, Easy, 4 marks]**
A group of 10 girls recorded the number of hours they spent watching television during a particular week. Their results are summarized in the box-and-whisker plot below.



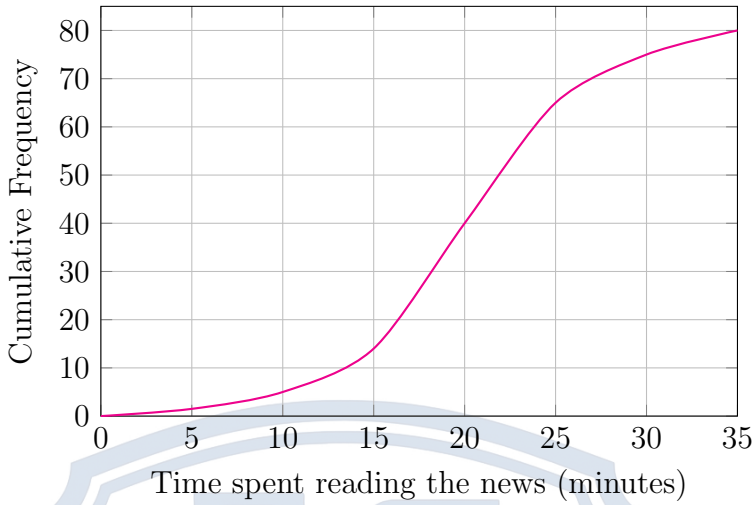
Given that the range of the data is exactly 16 hours:

(a) Find the value of a.

(b) Find the value of the interquartile range.

(c) State whether the distribution of hours spent watching television is positively skewed, negatively skewed, or symmetrical.

4. **[Paper 2 Style, Calculator Required, Easy, 5 marks]**
A school surveyed 80 of its final year students to find out how much time they spent reading the news on a given day. The results are shown in the cumulative frequency curve below.



(a) Use the graph to find the median number of minutes spent reading the news.

(b) Estimate the number of students who spent less than 15 minutes reading the news.

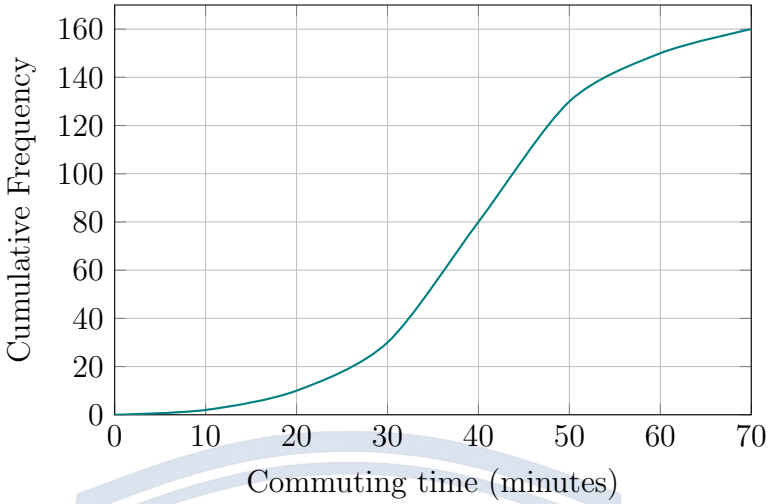
(c) Find the interquartile range.

5. **[Paper 1 Style, Non-Calculator, Medium, 5 marks]**
A frequency distribution for the continuous variable X is given in the table below.

Class Interval	Frequency
$0 \leq x < 10$	2
$10 \leq x < 20$	5
$20 \leq x < 30$	k
$30 \leq x < 40$	3

An estimate of the mean of X, calculated using the mid-interval values, is exactly 20. By forming an algebraic equation, find the exact value of the unknown frequency k.

6. **[Paper 2 Style, Calculator Required, Medium, 5 marks]**
A company records the commuting times, in minutes, of 160 employees. The data is represented in the cumulative frequency graph below.



(a) Find the median commuting time.

(b) Find the number of employees whose commuting time is within 10 minutes of the median.

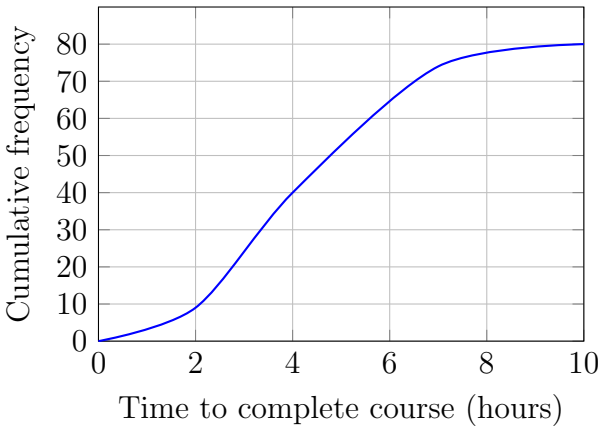
7. **[Paper 1 Style, Non-Calculator, Medium, 5 marks]**
A research student weighed lizard eggs in grams and recorded the results. The following box-and-whisker diagram shows a summary of the results, where L and U are the lower and upper quartiles respectively.

The interquartile range is 20 grams and it is known that there are absolutely **no outliers** in the results.

(a) By using the mathematical definition of an upper outlier, find the minimum possible value of U.

(b) Hence, find the minimum possible value of L.

8. **[Paper 2 Style, Calculator Required, Medium, 5 marks]**
The cumulative frequency graph below shows the number of hours, h, that 80 students took to complete an online driving course.



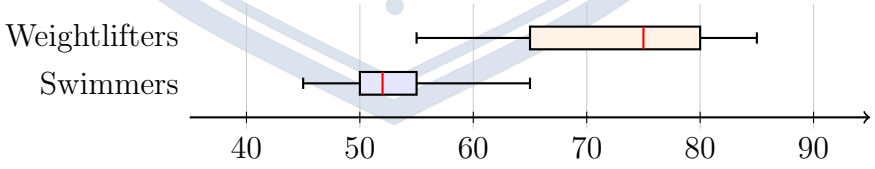
The same information is represented by the following grouped frequency table:

Hours, h	$0 \leq h < 2$	$2 \leq h < 4$	$4 \leq h < 7$	$7 \leq h < 10$
Frequency	9	p	q	6

(a) Use the cumulative frequency graph to find the value of p and the value of q.

(b) Use your graphic display calculator to estimate the standard deviation of the times taken to complete the online driving course.

9. **[Paper 1 Style, Non-Calculator, Hard, 6 marks]**
The resting heart rates (in beats per minute) of two groups of athletes, Swimmers and Weightlifters, are displayed in the parallel box-and-whisker plots below.



(a) Compare the two distributions, making two distinct statistical observations regarding their central tendency and dispersion.

(b) For the Weightlifters, calculate the upper outlier boundary and determine if a weightlifter with a heart rate of 98 bpm would be considered an outlier.

(c) Which of the two distributions exhibits a strong negative skew? Give a geometric reason for your answer.

10. **[Paper 2 Style, Calculator Required, Hard, 5 marks]**
Only 15% of the final year students from the survey in **Question 4** spent more than k minutes reading the news.

(a) Determine the number of students who represent this top 15%.

(b) By utilizing the cumulative frequency graph from Question 4, find the value of k.

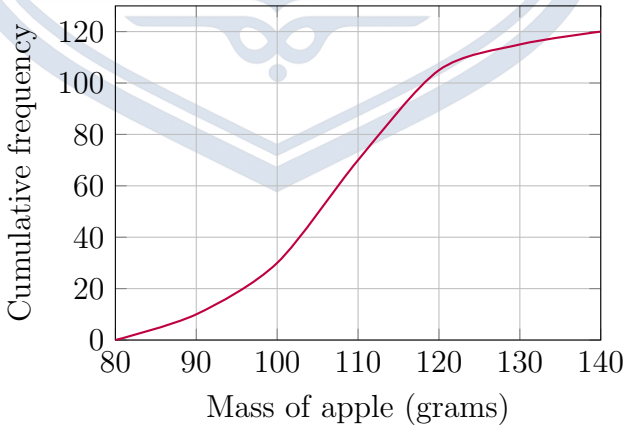
(c) The school introduces an initiative to encourage reading. If every student increases their reading time by exactly 5 minutes, state what effect this will have on the interquartile range.

11. **[Paper 1 Style, Non-Calculator, Hard, 5 marks]**
A histogram is drawn to represent a continuous data set. The data is divided into three equal class intervals: $0 \leq x < 10$, $10 \leq x < 20$, and $20 \leq x < 30$. The frequencies for these intervals are p, $2p$, and p respectively, where p is a positive integer.

(a) By using the mid-interval values and summing algebraically, prove that the estimated mean of this dataset is exactly 15, regardless of the value of p.

(b) Explain why the standard deviation of this dataset will also be independent of the value of p.

12. **[Paper 2 Style, Calculator Required, Hard, 5 marks]**
The cumulative frequency graph below shows the masses, m in grams, of 120 apples harvested from an orchard.



(a) Find the probability that a randomly chosen apple has a mass greater than 120 g.

(b) Given that an apple has a mass of less than 110 g, find the conditional probability that its mass is less than 100 g.

13. **[Paper 1 Style, Non-Calculator, Very Hard, 6 marks]**
A dataset of n observations has a box-and-whisker plot with the following parameters:

$$\text{Minimum} = 2, \quad Q_1 = q, \quad \text{Median} = 15, \quad Q_3 = 3q, \quad \text{Maximum} = 35$$

where q is a positive constant. If it is known that the dataset contains absolutely **no outliers**, find the full range of possible values for q. Show all algebraic reasoning clearly.

14. **[Paper 2 Style, Calculator Required, Very Hard, 6 marks]**
A researcher generates a cumulative frequency curve for a set of 200 data points. From the curve, she records the following percentiles:

- 10^{th} percentile = 12
- 25^{th} percentile = 22
- 50^{th} percentile = 38
- 75^{th} percentile = 56
- 90^{th} percentile = 82

The absolute minimum value in the dataset is 8, and the absolute maximum is 110.

(a) Using the appropriate percentiles, calculate the interquartile range (IQR).

(b) By creating a box-and-whisker plot profile, determine if the maximum value (110) and the 90^{th} percentile (82) are outliers.

(c) The researcher removes all outliers from the dataset. Assuming only the absolute maximum was an outlier, state how the median will shift (if at all).

15. **[Paper 1 Style, Non-Calculator, Very Hard, 7 marks]**
A continuous random variable X is organized into a grouped frequency table with three intervals of equal width c, where $c > 0$.

Class Interval	Frequency
$0 \leq x < c$	1
$c \leq x < 2c$	2
$2c \leq x < 3c$	1

(a) Using the mid-interval values, prove algebraically that the estimated mean of X is $\frac{3}{2}c$.

(b) Using the variance formula $\text{Var}(X) = \frac{\sum fx^2}{n} - \mu^2$, prove algebraically that the estimated variance of this grouped data is exactly $0.5c^2$.