4.1 Statistics Toolkit

Question Paper

Course	DP IB Maths
Section	4. Statistics & Probability
Topic	4.1 Statistics Toolkit
Difficulty	Medium

Time allowed: **90**

Score: **/68**

Percentage: **/100**

Question 1a

Every week an orangutan sanctuary measures the weight of each of its orangutans.

The weights, to the nearest kg, of **ALL** their 18 adult males are listed below:

52, 57, 63, 80, 56, 66, 101, 68, 55, 96, 70, 62, 66, 64, 99, 91, 55, 92

(a) Using a convenience sample of size six, calculate the mean weight of the male orangutans from the data set above.

[2 marks]

Question 1b

(b) Starting from the third data value, take a systematic sample of size six and re-calculate the mean weight of the male orangutans from the data set above.

[2 marks]

Question 1c

(c) Compare your results from parts (a) and (b) and state, with a reason, which sampling method is more reliable.

[2 marks]

Question 2a

A supermarket wants to gather data from its shoppers on how far they have travelled to shop there. One lunchtime an employee is stationed at the door of the shop for half an hour and instructed to ask every customer how far they have travelled.

(a) (i) State the sampling method the employee is using.

(ii) Give one advantage and one disadvantage of using this method.

[3 marks]

Question 2b

(b) State and briefly describe an alternative method of non-random sampling that the employee could use to obtain the required data for a sample of 30 customers.

[2 marks]

Question 3

A pharmacy sells face masks in a variety of sizes. Their sales over a week are recorded in the table below:

	Kids		Adults			
Size	Small	Large	S	M	L	XL
Frequency *f*	29	4	8	24	15	4

(i) Write down the mode for this data.

(ii) Explain why, in this case, the mode from part (i) would not be particularly helpful to the shop owner when reordering masks.

(iii) Given that the shop is open every day of the week, calculate the mean number of masks sold per day.

[4 marks]

Question 4

The lengths (l cm) of a sample of nine otters, measured to the nearest centimetre by a wildlife research team, are:

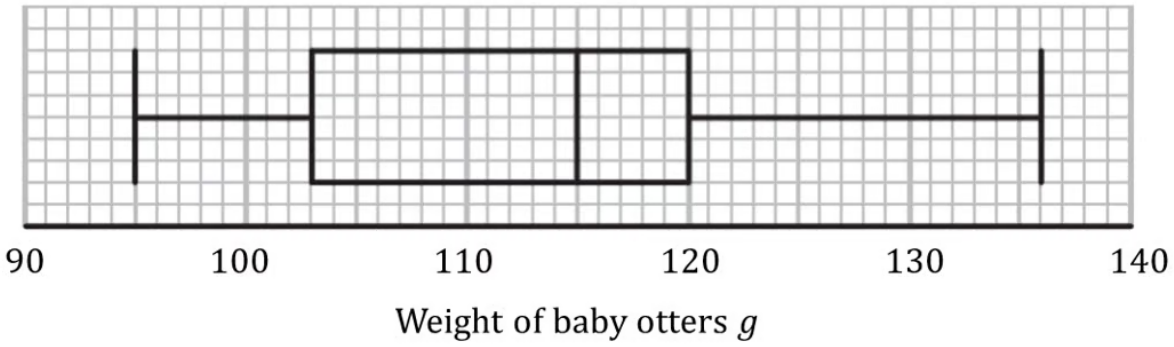
76 77 91 65 63 83 92 61 88

Calculate the mean and standard deviation of the nine recorded lengths.

[3 marks]

Question 5a

Jeanette works for a conservation charity who rescue orphaned otters. Over many years she records the weight (g) of each otter when it first arrives. The data is illustrated in the following box and whisker diagram:



(a) Using the box plot above:

(i) Write down the median weight of the otters.

(ii) Write down the lower quartile.

(iii) Find the interquartile range.

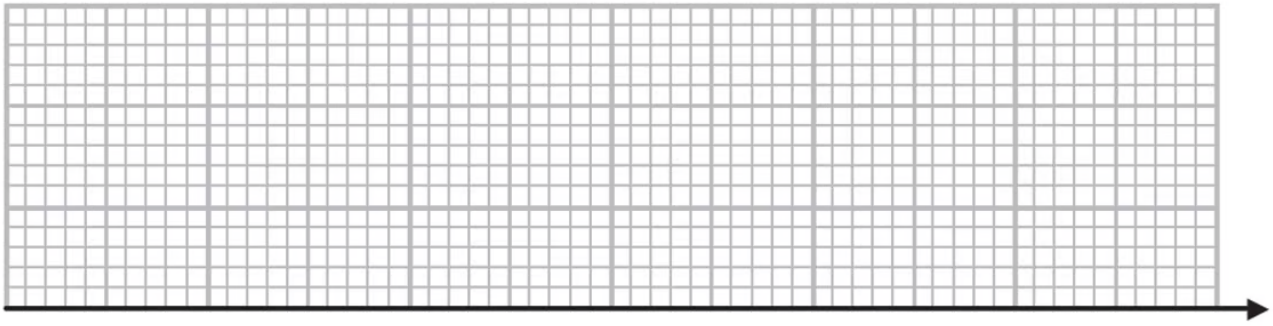
[4 marks]

Question 5b

Otters are then weighed weekly to track their growth. Summary data on the weights (g) of otters after one month is shown in the table below:

	Weight g
Smallest weight	125
Range	48
Median	152
Upper Quartile	164
Interquartile Range	33

(b) On the grid, draw a box plot for the information given above.



[3 marks]

Question 6a

The heights, in metres, of a flock of 20 flamingos are recorded and shown below:

0.4	0.9	1.0	1.0	1.2	1.2	1.2	1.2	1.2	1.2
1.3	1.3	1.3	1.4	1.4	1.4	1.4	1.5	1.5	1.6

An outlier is an observation that falls either more than $1.5 \times$ (interquartile range) above the upper quartile or less than $1.5 \times$ (interquartile range) below the lower quartile.

(a) (i) Find the values of Q_1, Q_2 and Q_3.

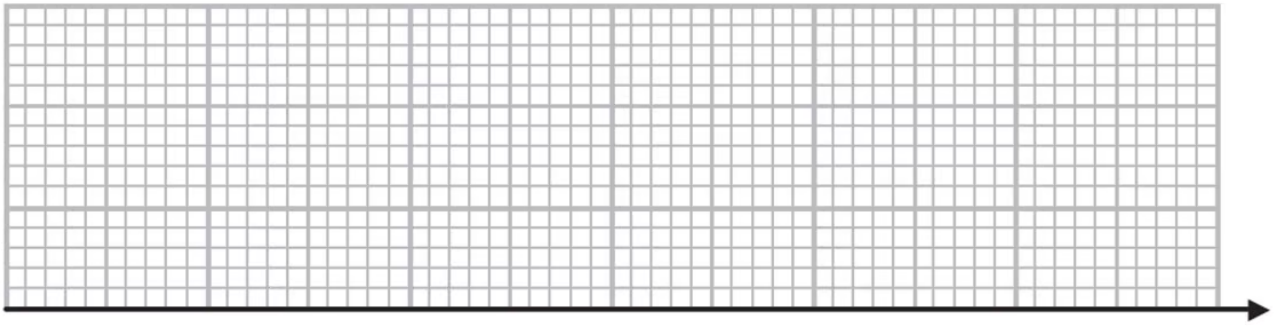
(ii) Find the interquartile range.

(iii) Identify any outliers.

[4 marks]

Question 6b

(b) Using your answers to part (a), draw a box plot for the data.



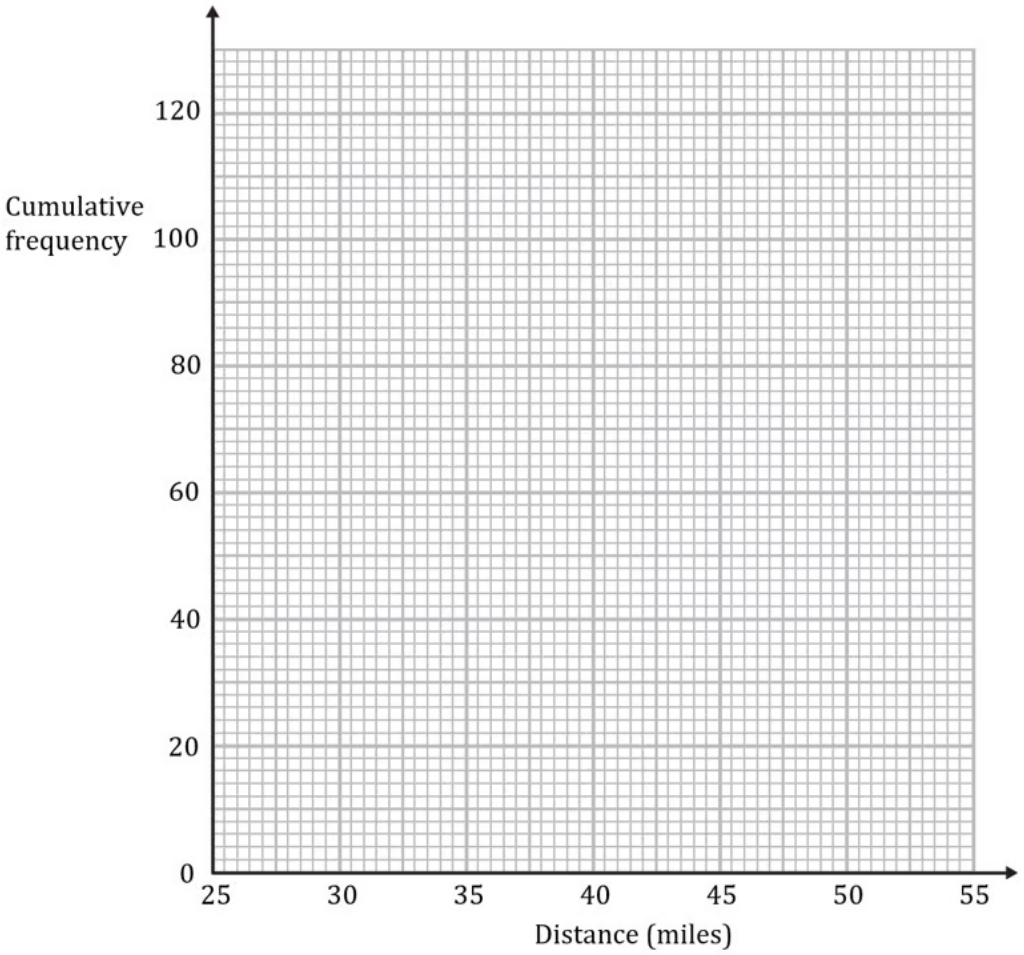
[3 marks]

Question 7a

120 competitors enter an elimination race for charity. Runners set off from the same start running as many laps of the course as possible. Their total distance is tracked and the competitor who runs the furthest over a 6-hour period is the winner. The distances runners achieved are recorded in the table below:

Distance d (miles)	Frequency f
$25 \leq d < 30$	8
$30 \leq d < 35$	10
$35 \leq d < 40$	32
$40 \leq d < 45$	54
$45 \leq d < 50$	10
$50 \leq d < 55$	6

(a) On the grid below, draw a cumulative frequency graph for the information in the table.



[3 marks]

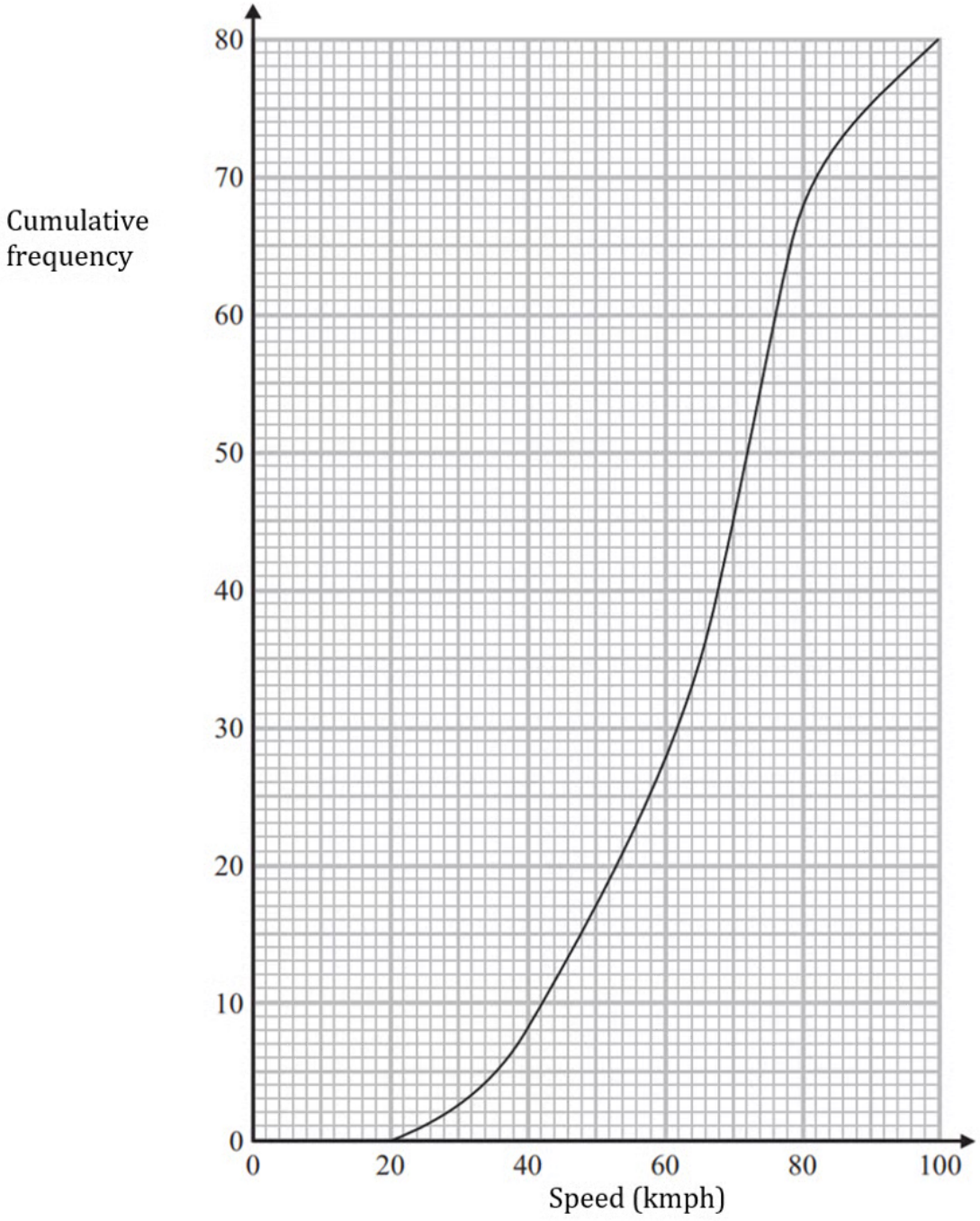
Question 7b

(b) Use your graph to find an estimate for the median and interquartile range.

[3 marks]

Question 8a

Police check the speed of vehicles travelling along a stretch of highway. The cumulative frequency curve below summarises the data for the speeds, in kmph, of 80 vehicles:



(a) Use the graph to find an estimate for the median speed.

[2 marks]

Question 8b

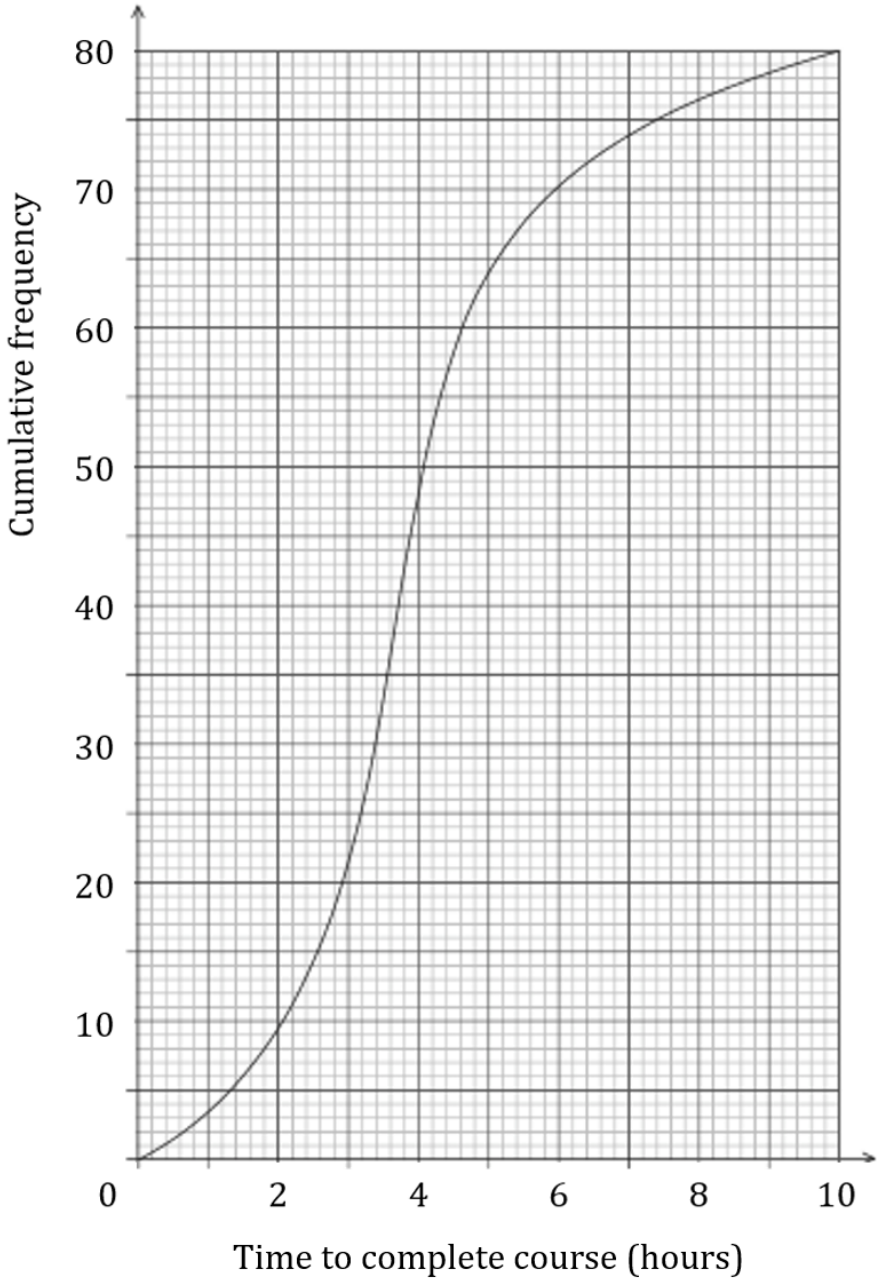
The speed limit for this section of road is 80 kmph.

(b) Vehicles travelling above the speed limit are issued with a speeding ticket. Those travelling more than 10% over the speed limit are pulled over. Use the graph to estimate the percentage of vehicles that the police pull over.

[3 marks]

Question 9a

The following cumulative frequency curve shows the number of hours, h, students took to complete their online driving course. The data is taken from 80 students, randomly selected from a large sample over a 12 month period.



(a) Find the median number of hours spent completing the online driving course.

[2 marks]

Question 9b

(b) Find the number of students whose online course time was within 1 hour of the median.

[2 marks]

Question 9c

(c) Calculate the interquartile range.

[2 marks]

Question 9d

The same information is represented by the following table.

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

(d) Find the value of p and the value of q.

[3 marks]

Question 9e

It is known that 10% of students take longer than d hours to complete the online driving course.

(e) Find the value of d.

[3 marks]

Question 9f

It is known that over a 12 month period, 4000 students in total sat the online driving course.

(f) Estimate the number of students over a 12 month period who took less than 3 hours to complete the course.

[3 marks]

Question 10a

For her IA, Mia decides to investigate the popularity of the social media platform SMEsocial amongst 16 to 18 year olds. She thinks of two possible methods for collecting data to analyse.

Method 1: Use visitor statistics from the SMEsocial website.

Method 2: Survey pupils from her school on their usage of SMEsocial.

(a) State one advantage and one disadvantage of method 2 outlined above.

[2 marks]

Question 10b

Mia decides to go ahead with the survey of pupils in her school and designs a questionnaire. She randomly asks 15 students to complete it.

(b) State the type of sampling method used.

[1 mark]

Question 10c

One of the questions that Mia asks on the survey is the following:

"How long do you spend on SMEsocial each week?"

(c) State one criticism of the question used by Mia.

[1 mark]

Question 10d

(d) Describe one way in which Mia could increase the reliability of her investigation.

[1 mark]

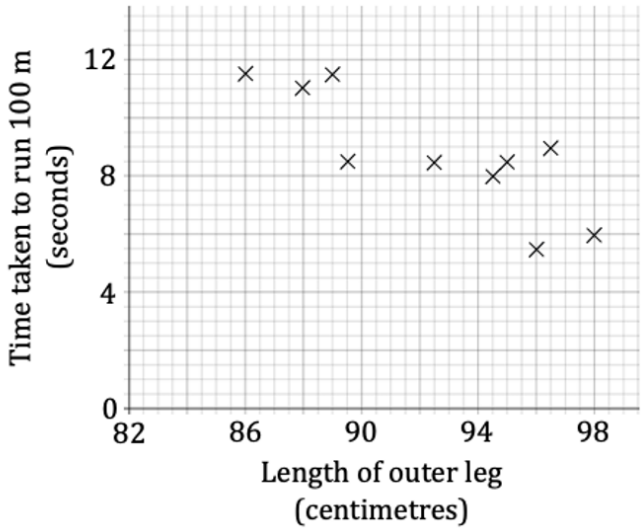
Question 10e

(e) Describe one way that Mia could improve the validity of her data.

[1 mark]

Question 11a

Nils believes that people with longer legs can run quicker. He decides to test his hypothesis by measuring the outer leg length of 10 boys aged between 10 and 12, and then using a data logger to record the time taken for each participant to run 100 metres. The results are shown on the scatter diagram below.



Nils notices that the scatter graph appears to support his hypothesis. Recalling that the world record for running 100 metres is 9.58 seconds, however, he believes that there may be a problem with the data that he has collected.

(a) Based on the information given in the question and the data, explain what the data suggests about the students.

[1 mark]

Question 11b

(b) Explain what could be wrong with the data collected and how it may have occurred.

[2 marks]

Question 11c

(c) State what Nils should do to confirm the reliability of the data.

[1 mark]