4.1 Statistics Toolkit

Question Paper

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

Time allowed: 80

Score: /59

Percentage: /100

Question 1a

A group of Netflix subscribers participated in a research survey and the ages of participants were recorded in the following table.

Age, in years (a)	$15 \leq a < 25$	$25 \leq a < 35$	$35 \leq a < 45$	$45 \leq a < 55$	$55 \leq a < 65$
Number of participants	11	62	56	x	12

It is known that $56 < x < 62$.

(a) Write down

(i)   the modal class

(ii)   the mid interval value of the modal class.

[2 marks]

Question 1b

(b) Determine the class in which the upper quartile lies.

[2 marks]

Question 1c

Using the mid-interval values the mean of the data can be estimated to be 39.95.

(c) Find the value of x.

[2 marks]

Question 1d

The participants in this survey were chosen by randomly selecting people entering a supermarket. However, to be more efficient, the surveyor only selected people who were in groups of at least 3.

(d) Write down the type of sampling method used.

[1 mark]

Question 2a

25 people are invited to the premier of the movie "ICE", and they are asked to give the movie a score out of 1-10. The table below shows the distribution of the scores.

Score	1	2	3	4	5	6	7	8	9	10
Frequency	1	2	1	4	5	5	1	a	3	b

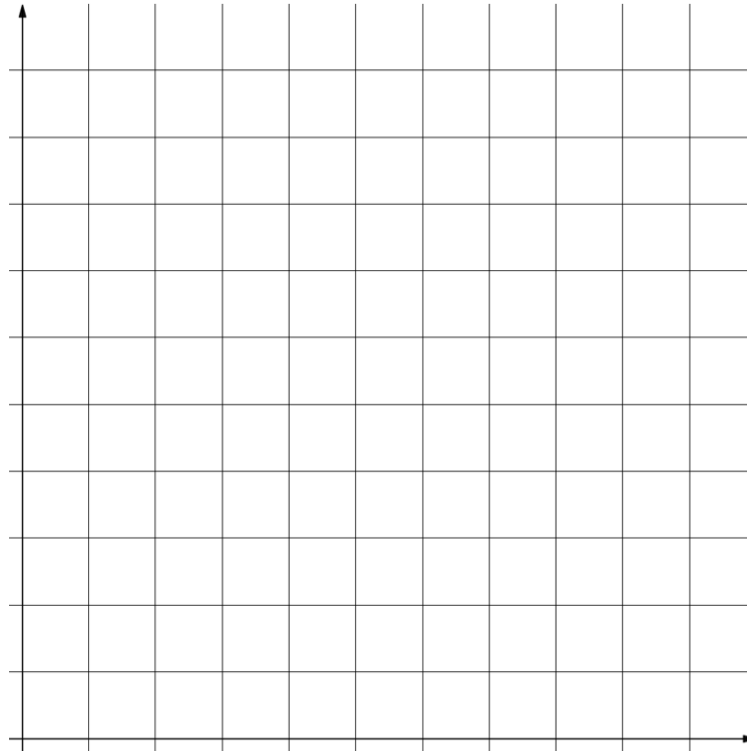It is known that $a > b$.

(a) Find the value of a and the value of b.

[2 marks]

Question 2b

(b) Draw a bar chart of the data on the grid below.

[4 marks]

Question 3a

A shoe store wants to know which shoe size is the most popular and so they record the shoe sizes of 30 customers.

Score	7	7.5	8	8.5	9	9.5	10	11	12	13
Frequency	5	1	4	p	3	3	2	1	q	1

(a) State whether the above data is continuous or discrete.

[1 mark]

Question 3b

It is known that $p = 4q$.

(b) Find the value of

    (i) p

    (ii) q.

[2 marks]

Question 3c

(c) Write down

    (i)   the mean

    (ii)   the median

    (iii) the mode.

[3 marks]

Question 3d

(d) State which statistical measure is most useful for the shoe store. Justify your answer.

[1 mark]

Question 4a

A data set has a mean of 22 and a standard deviation of 6.

Each element of the data set has 4 subtracted from it.

(a) Find the value of

    (i)   the new mean

    (ii)   the new standard deviation.

[3 marks]

Question 4b

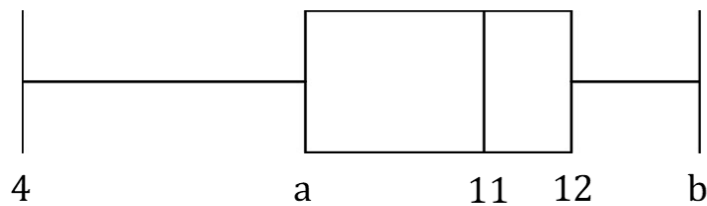After each element has 4 subtracted from it, each element is divided by $\frac{2}{3}$.

(b) Find the value of

    (i)   the new mean

    (ii)   the new variance.

[3 marks]

Question 5a

The following box and whisker diagram shows the number of social media posts made by a group of content creators over a week.

Number of social media posts

(a) State whether the data is discrete or continuous.

[1 mark]

Question 5b

It is given that $7a = 4b$ and $a + b = 22$.

(b) Calculate the value of

    (i) a

    (ii) b.

[3 marks]

Question 5c

(c) A content creator made k posts, where $k < a$. Given that k is an outlier find the maximum value of k.

[4 marks]

Question 6a

The following table displays the percentage change of an investor's portfolio over 12 months.

Month	Percentage change
1	-5.51
2	6.86
3	4.00
4	1.67
5	2.18
6	0.17
7	1.43
8	-2.31
9	3.31
10	-3.35
11	1.81
12	-3.24

(a) Calculate

    (i)   the mean

    (ii)   the standard deviation.

[2 marks]

Question 6b

(b) State which statistical measure, calculated in part (a), gives an indication of the volatility of the investor's portfolio.

[1 mark]

Question 6c

The investor's portfolio value at the beginning of the 12 months was $6000.

(c) Calculate the value of the portfolio at the end of the 12 months. Give your answer to 2 decimal places.

[3 marks]

Question 7a

At a swimming competition the mean time of the first four swimmers is 28.2 seconds. The time for the fifth and sixth swimmers are then recorded and the mean time of the first six swimmers is 29.8 seconds. The difference between the fifth and sixth swimmer's time is 0.4 seconds.

(a) Find the time achieved by

(i)   the fifth swimmer

(ii)   the sixth swimmer.

[4 marks]

Question 7b

The first swimmers time is 25.7 seconds.

(b) Calculate the range in times of the swimmers.

[1 mark]

Question 8a

The table below shows the average temperature, T °C, in a city over a normal year (not a leap year).

Temperature	$-5 \leq T < 5$	$5 \leq T < 15$	$15 \leq T < 25$	$25 \leq T < 35$
Frequency	124	p	109	q

It is given that $p = 11q$.

(a) Calculate the values of

(i) p

(ii) q.

[3 marks]

Question 8b

(b) Using your GDC, estimate the value of

    (i)   the mean

    (ii)   the standard deviation.

[3 marks]

Question 8c

It is given that the mean temperature of the city over the year is 14.2 °C.

(c) Calculate the percentage error between your estimate of the mean temperature, found in part (b) (i), and the actual mean temperature.

[2 marks]

Question 9a

Ann is the product manager of a bicycle company and she is testing a new type of tyre. Ann wants to test whether using the new tyres allows cyclists to reach higher speeds. She asks 10 volunteers to cycle 30 miles and to record their times. An extract of the data is shown below:

Volunteer	Time (hours)
A	1.05
B	1.75
C	1.50
D	1.25

It is known that the recording devices had two settings for displaying time. Ann checks and finds that volunteer C took 1 hour and 30 minutes whereas volunteer D took 1 hour and 25 minutes.

(a) Explain why Ann will need to verify the time of volunteer A.

[1 mark]

Question 9b

Ann runs the experiment again but this time she records the times herself. She asks the volunteers to repeat the process three times.

(b) State three factors which Ann should control to try to ensure that her measurements are reliable.

[3 marks]

Question 9c

(c) It is later discovered that the volunteers completed all three journeys during the same day. Explain how this would affect the reliability of the measurements.

[2 marks]