Cambridge Assessment International Education

Cambridge O Level

CANDIDATE NAME									
CENTRE NUMBER						CANDIDATE NUMBER				

8484042231

STATISTICS **4040/23**

Paper 2 **October/November 2023**

2 hours 15 minutes

You must answer on the question paper.

You will need: Calculator
Pair of compasses
Protractor

INSTRUCTIONS

- Answer **all** questions.
- Use a black or dark blue pen. You may use an HB pencil for any diagrams or graphs.
- Write your name, centre number and candidate number in the boxes at the top of the page.
- Write your answer to each question in the space provided.
- Do **not** use an erasable pen or correction fluid.
- Do **not** write on any bar codes.
- You should use a calculator where appropriate.
- You must show all necessary working clearly.

INFORMATION

- The total mark for this paper is 100.
- The number of marks for each question or part question is shown in brackets [ ].

This document has **16** pages. Any blank pages are indicated.

DC (AP) 332444

1 The students in a class take a Chemistry assessment consisting of a written test and a practical test.
The table gives some information about the marks of the class overall and the marks of one of the students, Yazan.

	Class mean	*Class standard deviation*	*Yazan*
Written test	60.5	7.8	41
Practical test	37.2	y	40

The marks in both parts of the assessment are to be scaled to a mean of 50 and a standard deviation of 12.

(a) Find Yazan's scaled mark in the written test.

..... [2]

In the practical test, Yazan's scaled mark is 54.

(b) Find the class standard deviation, y, for the practical test.

..... [2]

In the written test, Lorato's mark is unchanged by the scaling process.

(c) Find Lorato's mark in the written test.

..... [2]

2 Two independent events A and B are such that

$$\mathrm{P}(A) = \frac{3}{5} \quad \text{and} \quad \mathrm{P}(B) = \frac{1}{2}.$$

Find

(a) $\mathrm{P}(A')$,

..... [1]

(b) $\mathrm{P}(A \cap B)$,

..... [2]

(c) $\mathrm{P}(A \cup B)$.

..... [2]

(d) State which pair of events from A, B and A' are mutually exclusive.
Give a reason for your answer.

.....

..... [1]

3 A quality control manager at a clothing factory tests a sample of the 1200 items of clothing made in one day. He numbers each item from 0001 to 1200 and selects a systematic sample.
Two of the items in the sample are those numbered 0127 and 0187.

(a) Find the smallest possible size of this systematic sample.

..... [2]

He discovers a fault in one of the sampled items. As the item is a shirt, he decides to take a further sample from the 84 shirts made that day.
Each shirt was made using one of the machines A, B or C. He numbers each shirt according to the machine used to make it.

Machine	*Shirt number*
A	01–36
B	37–72
C	73–84

He decides to take a sample of **size 7**, stratified by machine.

(b) (i) Find the number of shirts that he should sample from each machine.

Machine A .....

Machine B .....

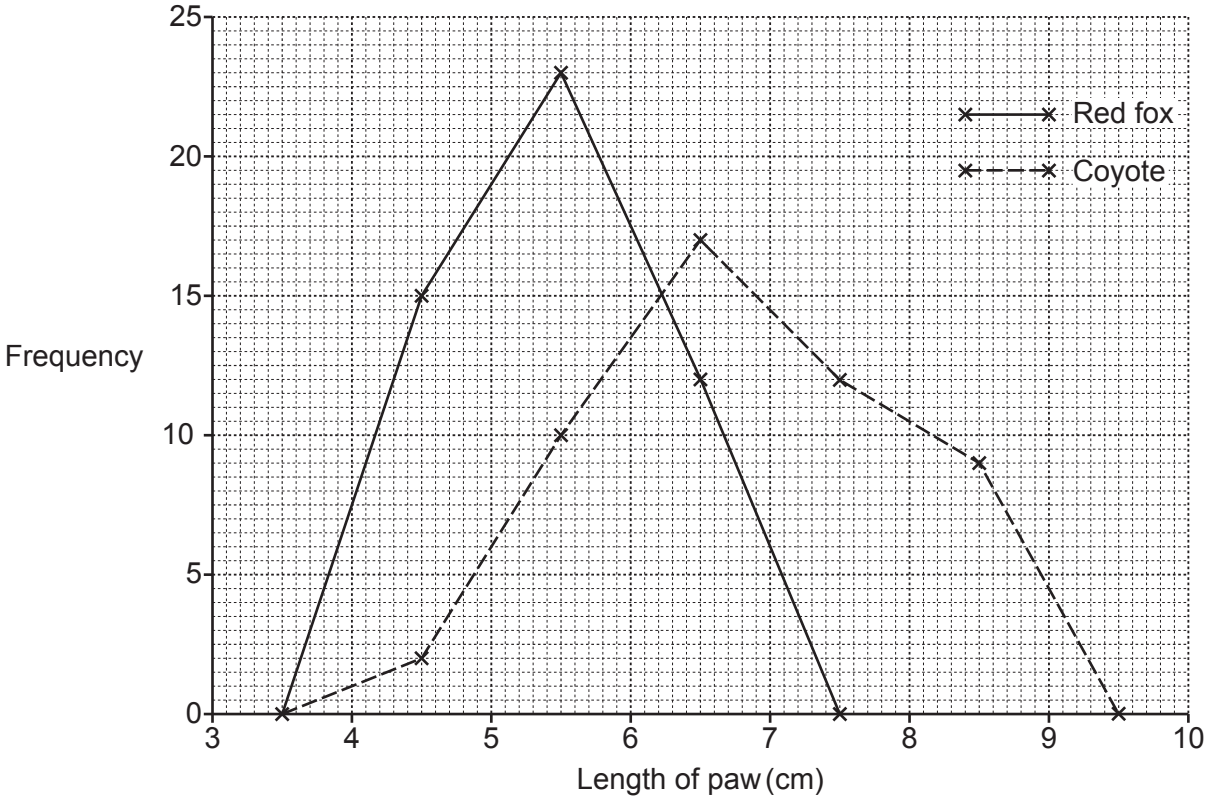
Machine C .....
[2]

(ii) Use the random number table below to select the sample.
Start on the left of the table, and use every number if the machine to which it relates has not yet been fully sampled.

79 04 41 97 23 82 04 35 17 58 60 26 66 71 09

..... [3]

4 The paw lengths of 50 red foxes and 50 coyotes are measured and the results displayed in the frequency polygons.



(a) Describe **two** differences between the paw lengths of the red foxes and the paw lengths of the coyotes.

.....

.....

.....

..... [2]

Alternatively the data could have been displayed using histograms.

(b) Give **one** advantage that the frequency polygons have over histograms.

.....

..... [1]

Paw prints of red foxes and coyotes look very similar, with the main difference being their size. A paw print from one of the animals that has been measured is found. It has a length of 7.1 cm. John says that this must have been from one of the coyotes.

(c) Explain whether or not you think he is correct.

.....

..... [1]

5 A company director produces the following diagrams for his annual report.

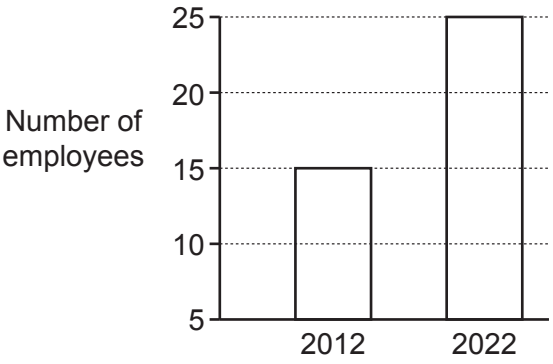


Fig. 1

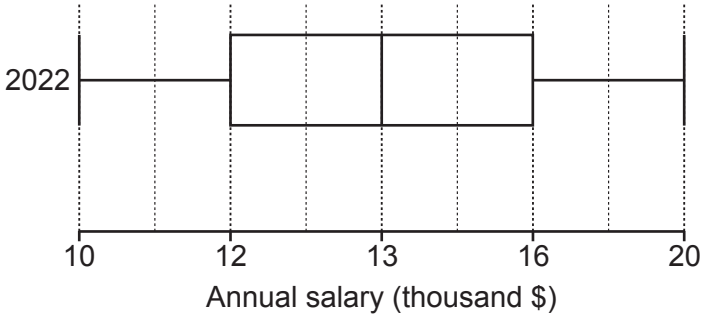


Fig. 2

(a) Explain what is misleading about each of these diagrams.

Fig. 1 .....

.....

Fig. 2 .....

.....

[2]

The stem-and-leaf diagram below shows the annual salaries of the 15 employees in 2012.

7	1	2	5	8	8	
8	0	2	3	6	7	9
9	3	3	7			
10						
11	9					

Key: 7|1 represents an annual salary of $7100

(b) Find the median salary of the employees in 2012.

..... [1]

(c) Draw a box-and-whisker diagram for the salaries in 2012.

2012

[4]

6 A scientist is studying the water quality in naturally occurring springs in a national park. She divides the national park into 30 equally sized areas and counts the number of springs in each area.

Number of springs	0	1	2	3
Number of areas	6	9	10	5

(a) Find the total number of springs in the national park.

..... [2]

She decides to test the water quality in two different springs, selected at random.

(b) Find the probability that the two springs are from the same area.

..... [5]

7 Some children at a school take a Mathematics challenge. The data collected is the time it takes each of them to complete the challenge.

(a) Use statistical language to describe fully the type of data that is being collected.

..... [2]

Information about the times taken, in minutes, is shown in the table.

Time to complete the challenge, t (mins)	$20 \leqslant t < 40$	$40 \leqslant t < 50$	$50 \leqslant t < 55$	$55 \leqslant t < 60$	$60 \leqslant t < 70$
Number of children	13	15	38	42	12

Estimates, obtained by linear interpolation, of the lower quartile and the median are 50.26 minutes and 54.21 minutes respectively.

(b) Use linear interpolation to find an estimate for the upper quartile, and hence find an estimate for the interquartile range of the times to complete the challenge.

..... [5]

(c) State, with a reason, whether you think that the mean time taken is likely to be greater than, less than or equal to the median time taken. You do **not** need to calculate the mean.

.....

..... [1]

Anyone who completes the challenge in less than 48 minutes qualifies to enter a national challenge.

(d) Use linear interpolation to estimate the number of these children who qualify for the national challenge.

..... [3]

It is discovered later that the children should have been allowed to read the challenge instructions before the timing started. It is estimated that it took each of them $2\frac{1}{2}$ minutes to read the instructions, so this amount of time is subtracted from each child's time.

(e) Find estimates for the new median and interquartile range of the times spent completing the challenge.

Median .....

Interquartile range .....
[1]

(f) Use linear interpolation to estimate how many more of these children now qualify for the national challenge.

..... [4]

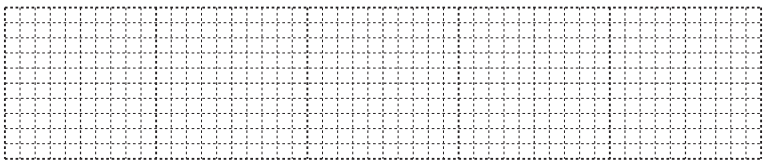
8 A driving instructor wants to calculate a weighted aggregate cost of driving index. She divides her expenditure into three categories: Tax and insurance, Maintenance, and Fuel.

In 2018, she travelled a distance of 24 000 km,
her mean fuel consumption was 7.2 litres per 100 km,
the cost of her tax and insurance was $720,
her maintenance costs were $40 per 1000 km,
the cost of her fuel was $1.25 per litre.

(a) Show that the driving instructor should assign weights to the three categories Tax and insurance, Maintenance, and Fuel in the ratio 3 : 4 : 9.

[3]

(b) Display these weights using a percentage sectional bar chart on the grid below.



[3]

The price relatives for the three categories Tax and insurance, Maintenance, and Fuel in 2019 were 107, 102 and 92 respectively, taking 2018 as the base year.

(c) Using part **(b)** and these price relatives, explain, without further calculation, whether you think that the overall costs for the driving instructor increased, decreased or stayed the same between 2018 and 2019.

.....

.....

..... [2]

In 2022 the driving instructor bought a new car.

Compared to 2018, her tax and insurance costs reduced by 16%,
her maintenance costs reduced to $36 per 1000 km,
the cost of her fuel increased to $1.45 per litre.

(d) Taking 2018 as the base year, complete the table below to show price relatives for 2018 and 2022.

	Price relative	
	2018	2022
Tax and insurance		
Maintenance		
Fuel		

[4]

(e) Calculate a weighted aggregate cost of driving index for 2022, taking 2018 as the base year.

..... [3]

The weighted aggregate cost of driving index may prove to be incorrect if the weights have changed.

(f) Give one reason why the weights may have changed.

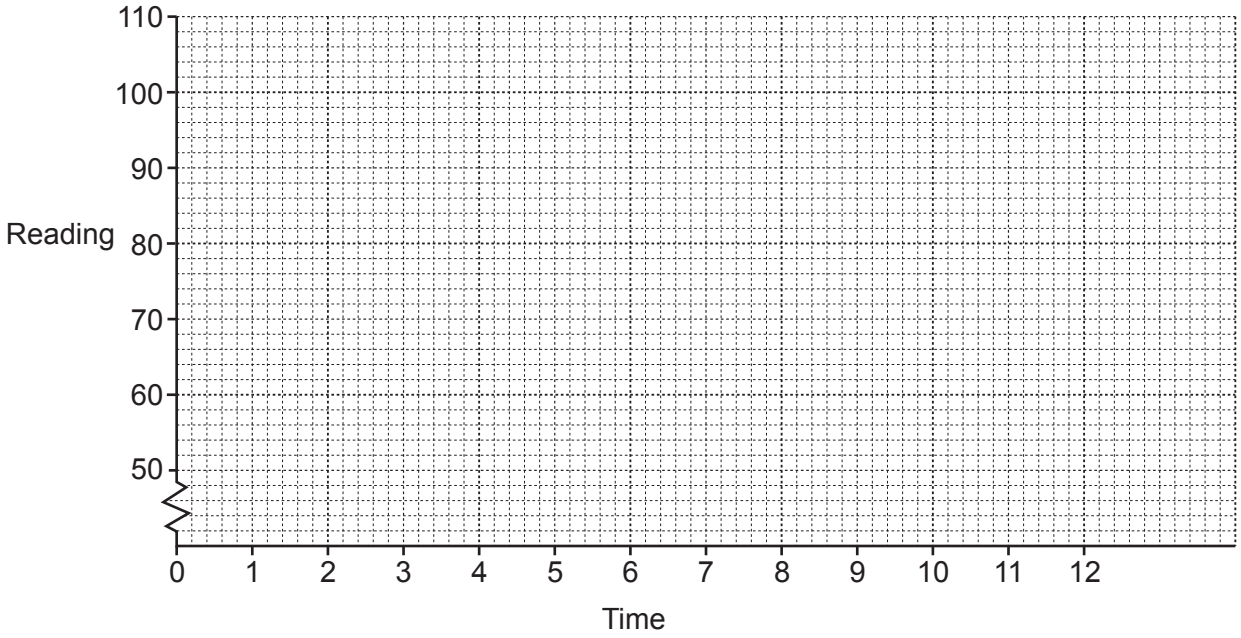
.....

..... [1]

9 Readings from a scientific instrument were taken at equal time intervals.

Time	1	2	3	4	5	6	7	8	9	10	11	12
Reading	103	70	65	98	68	62	98	65	59	95	62	56

(a) Plot a time series graph for the readings.



[2]

The trend of the readings is downward.
Elsie says that this is because the last two readings, 62 and 56, are each lower than the previous reading.

(b) Explain what is wrong with Elsie's statement.

.....

..... [1]

(c) Show, correctly, that the trend of the readings is downward.
Add appropriate points and a trend line to the graph in part **(a)**.

Time	*Reading*	
1	103	
2	70	
3	65	
4	98	
5	68	
6	62	
7	98	
8	65	
9	59	
10	95	
11	62	
12	56	

[7]

(d) By using your work in part **(c)** and by calculating an appropriate seasonal component, find an estimate for the reading at time 13.

..... [5]

10 A game involves rolling a fair 4-sided die, with faces numbered 1, 2, 3 and 4, three times. Once a 4 is obtained, the player puts a counter on the start square.

Start	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**

Once on the start square the player moves the counter forward according to the numbers on any remaining rolls of the die.
For example, if [4, 2, 4] is obtained, the counter will end up on square 6.
If [2, 4, 1] is obtained, the counter will end up on square 1.

(a) Find the probability that the counter

(i) does not get onto the start square,

..... [2]

(ii) gets onto the start square, but moves no further on.

..... [2]

If the counter does not get onto the start square, the prize is \$0. Otherwise the prizes are shown in the table below.

Final square	**Start**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Prize	\$0	\$1	\$1	\$1	\$1	\$1	\$1	\$2	\$3

(b) Find the probability that the prize is

(i) \$0,

..... [1]

(ii) \$3,

..... [1]

(iii) $2.

..... [2]

(c) Complete a probability distribution table for all the prize values.

[3]

(d) Calculate the amount that should be charged to play if this is to be a fair game.

..... [2]

It is later decided to double the amount charged to play a game. All the prizes remain the same, except the prize for square 8.

(e) Calculate the prize that should be given for ending up on square 8 if this is to remain a fair game.

..... [2]

(f) Explain why the difference between the original prize for ending up on square 8 and the answer to part **(e)** is so large.

.....

..... [1]

BLANK PAGE