Cambridge O Level
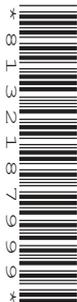
CANDIDATE NAME

CENTRE NUMBER

CANDIDATE NUMBER

STATISTICS **4040/23**

Paper 2 **October/November 2022**

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.

[Turn over

1 A researcher conducted a survey of 144 shoppers in a city centre one day.

His first question was: 'What mode of transport did you use to get to the city centre today?'
His second question was: 'What distance have you travelled to get to the city centre today?'
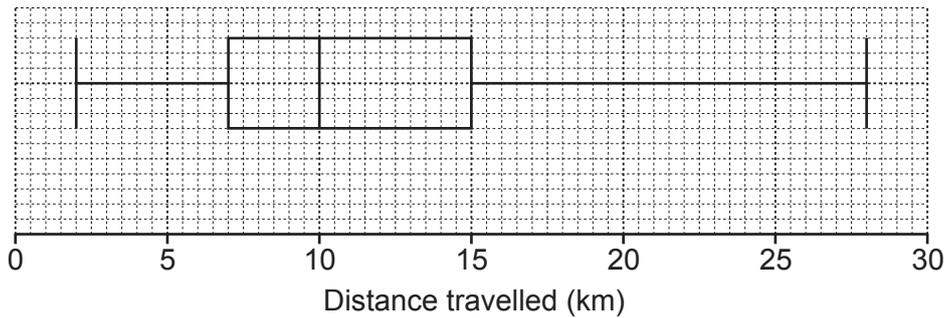
(a) For each question, use statistical language to describe fully the type of data that he collected.

Mode of transport ....

Distance travelled ....

[3]

He drew a box-and-whisker diagram of the results of his second question.

Distance travelled (km)

(b) Find the interquartile range of the distances travelled.

.... [2]

(c) Calculate the number of shoppers that had travelled more than 15 km.

.... [2]

2 Sumant wants to know which one of cricket, football or tennis is the most popular sport amongst the children in his class. He asks the boys and girls in his class to choose their favourite, and draws a dual bar chart of the results.
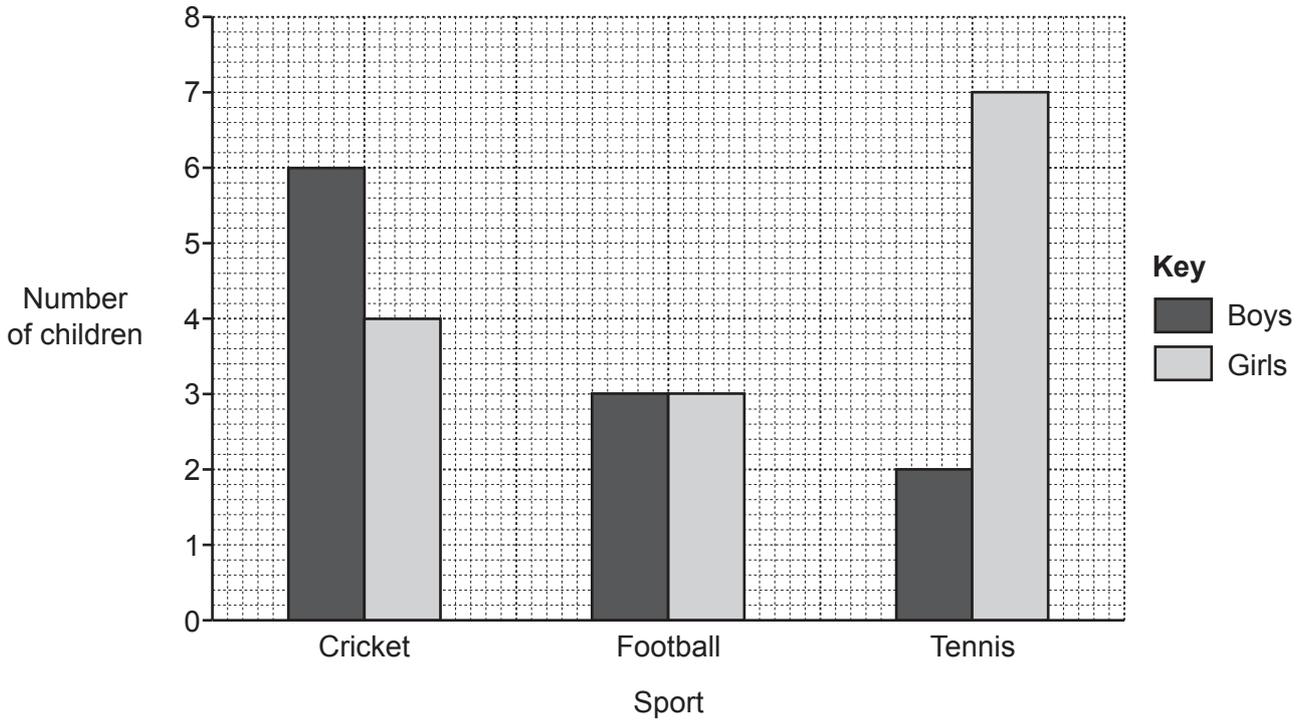
(a) Name an alternative type of bar chart that would have been more appropriate for him to use. Give a reason for your answer.

....

....

....
[2]

Use the dual bar chart to find the probability that a child chosen at random

(b) said that football was their favourite sport,

.... [2]

(c) said that football was their favourite sport, given that they were a boy,

.... [1]

(d) was a boy, given that they said that football was their favourite sport.

.... [1]

3 A factory produces three sizes of battery: D, AA and AAA. A quality control manager is to test a sample of the 600 batteries produced in one day. She gives each battery a 3-digit number as shown in the table.

Battery size	Number produced	3-digit number
D	100	000–099
AA	300	100–399
AAA	200	400–599

She uses a random number generator to obtain the following simple random sample of **size 6**:

016, 582, 409, 037, 297, 108

(a) Show whether or not this simple random sample is representative in terms of the battery sizes.

[3]

She decides instead to select a sample of **size 5**, stratified by battery size.

(b) Decide how many of each type of battery should be in the sample.
Show your reasoning.

D ....

AA ....

AAA ....

[2]

4   *A* and *B* are two independent events, such that

$$P(A) = 0.3 \quad \text{and} \quad P(B) = 0.4 \,.$$

Find

(a) P(*A* or *B*),

.... [3]

(b) P(*A* or *B* but not both),

.... [2]

(c) P(not *A* and not *B*).

.... [2]

5 Uzma wants to find a weighted aggregate cost-of-housing index.
She divides her housing costs into three categories: Rent, Electricity and Other costs.

Last year she spent:
$250 per month on rent
$0.80 per unit for 1200 units of electricity
$360 on other costs

(a) Show that weights based on expenditure last year are in the ratio $25:8:3$.

[2]

This year, her rent increased by 9%.
The cost of each unit of electricity remained the same.
Other costs decreased by 2%.

(b) Using the weights from part **(a)**, find a weighted aggregate cost-of-housing index for Uzma.

.... [5]

The index found in part **(b)** may be inaccurate if the weights have changed.

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

....

.... [1]

6 A post office recorded the masses of all the parcels that it processed last year.

28% of the parcels had a mass less than 600 g.
10% of the parcels had a mass greater than 1000 g.

(a) Use linear interpolation to find an estimate for the median mass of these parcels.
Give your answer to the nearest gram.

.... [4]

(b) State the assumption that you have made in order to give your estimate in part **(a)**.

....

.... [1]

7 The ages and genders of the workers at a company are shown in the table.

Males								Females							
50	46	61	30	50	52	56	54	47	38	39	48	33	39	62	57
55	57	61	49	62	64	65		63	64	63	41	51	42	65	

The ages of the females have been put into the incomplete back-to-back stem-and-leaf diagram below.

	3	3	8	9	9	
	4	1	2	7	8	
	5	1	7			
	6	2	3	3	4	5

(a) Complete the back-to-back stem-and-leaf diagram by adding the data for the males.
Include a key. [4]

(b) Find the lower quartile, median and upper quartile of the ages of the males and the females, and insert them into the table.

	Males	Females
Lower quartile		
Median		
Upper quartile		

[3]

Azeeb says, 'The male workers are generally younger than the female workers.'
Tebogo says, 'The ages of the male workers are less varied than those of the female workers.'

(c) For each of Azeeb and Tebogo, state whether or not they are correct and use values from your table to justify your answer.

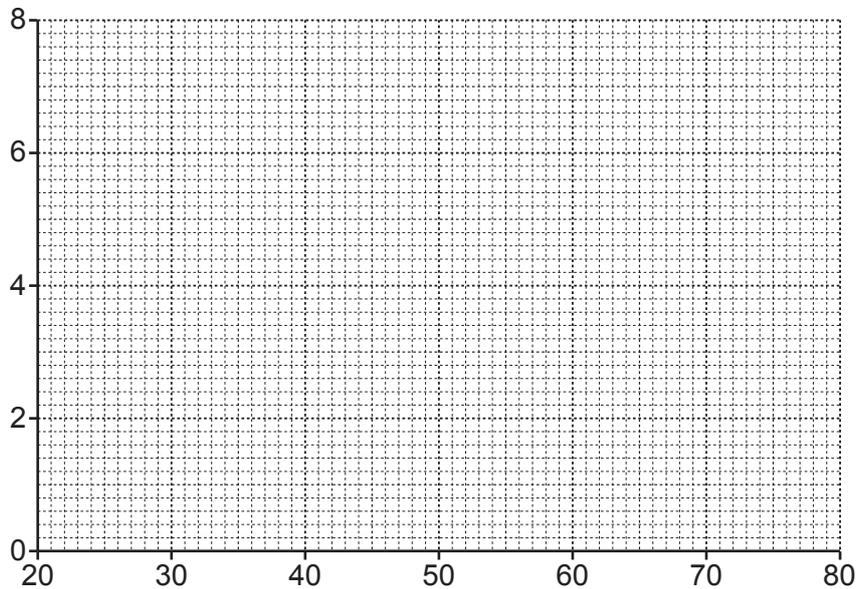Azeeb ....

....

Tebogo ....

....

[2]

(d) Complete the frequency table.

Age, x (years)	Number of males	Number of females
$20 \leqslant x < 30$		
$30 \leqslant x < 40$		
$40 \leqslant x < 50$		
$50 \leqslant x < 60$		
$60 \leqslant x < 70$		
$70 \leqslant x < 80$		

[1]

(e) On the grid, draw a pair of frequency polygons for the ages of the males and the ages of the females.
Include all necessary labels and a key.

[3]

(f) State one advantage that the stem-and-leaf diagram has over the frequency polygon.

....

.... [1]

(g) Some new female workers arrive at the company, which changes the median age of the female workers to 54.

(i) Find the smallest number of females that could have arrived.

.... [1]

(ii) What can be said about the ages of these new female workers?

....

.... [1]

8 A ranger records the number of swans that visit a nature reserve each quarter for 3 years.
Some of the swans migrate to other places at certain times of the year, causing seasonal variation.

Year and quarter		Number of swans	4-point moving average	Centred 4-point moving average
2019	Q1	2480		
2019	Q2	327		
			$x =$	
2019	Q3	418		1320
			1308.75	
2019	Q4	2100		1309.375
			1310	
2020	Q1	2390		1308.125
			$y =$	
2020	Q2	332		1306.75
			1307.25	
2020	Q3	403		$z =$ ....
			1303.75	
2020	Q4	2104		1302.25
			1300.75	
2021	Q1	2376		1301.125
			1301.5	
2021	Q2	320		1297.75
			1294	
2021	Q3	406		
2021	Q4	2074		

(a) Explain why the ranger might want to find moving average values.

....

....

.... [2]
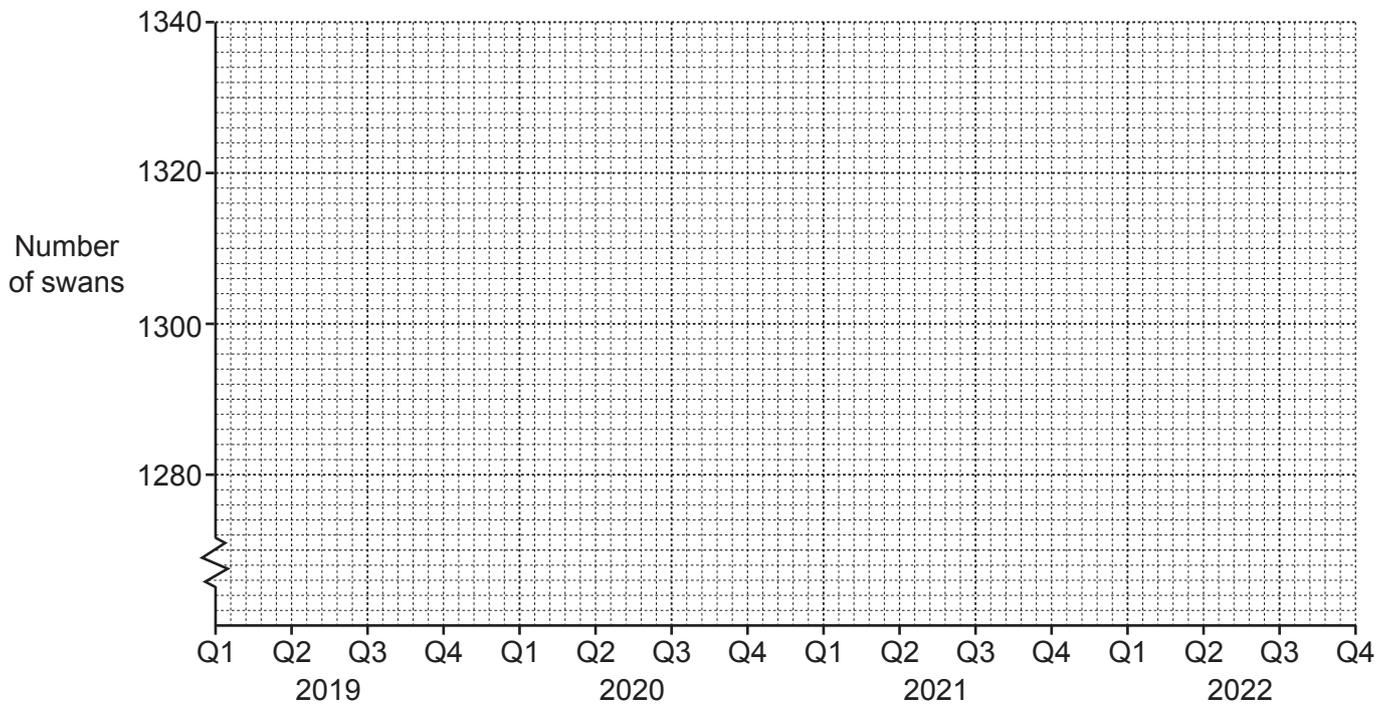
(b) Calculate the values of x, y and z and insert them in the table. [3]

(c) Use appropriate values from the table to find an estimate of the seasonal component for quarter 2.

.... [3]

(d) Plot all the centred moving average values on the grid below and draw an appropriate trend line.

[3]

The ranger's assistant says, 'The trend line shows that the number of swans is falling each quarter.'

(e) Explain whether or not you think the ranger's assistant is correct.

....

.... [1]

(f) Use your answers to parts **(c)** and **(d)** to estimate the number of swans that will visit the nature reserve in quarter 2 of 2022.

.... [2]

9 Abena and Hilda each have three cards. The cards are either a square or a triangle and are either black or white.

Abena's cards   Hilda's cards

They each choose one of their cards at random.

(a) Find the probability that the two cards chosen are

(i) both black,

.... [1]

(ii) both squares,

.... [1]

(iii) both black squares,

.... [1]

(iv) both black or both squares.

.... [2]

Abena and Hilda play a game.
If the two cards they choose are identical (the same shape and colour), Abena gives Hilda $x.
If the two cards are not identical, Hilda gives Abena $1.

(b) Find the value of x if this is a fair game.

.... [4]

Abena decides to make a new game. She puts all six cards in a bag and Hilda chooses two cards at random, without replacement.

(c) Find the probability that the two cards chosen are

(i) different shapes,

.... [3]

(ii) the same shape and colour,

.... [1]

(iii) the same shape but different colours.

.... [1]

Abena decides to give Hilda a prize for each of these outcomes as shown.

Outcome	Prize
Two cards of different shapes	$3
Two cards of the same shape and colour	$9
Two cards of the same shape and different colours	$6

(d) Find how much Abena should charge Hilda to play this game to make it a fair game.

.... [2]

10 All the students in a school are given the same History test. The table shows information about the students from Years 1 and 2 and their scores in the test.

	Number of students in year group	Mean of the scores	Standard deviation of the scores
Year 1	159	62	8
Year 2	141	68	10

Hazeema is in Year 1 and scored 52 in the test.
Kalilo is in Year 2 and scored 54 in the test.

(a) Which of these two students performed better, relative to all the students in their year group? Show your working.

.... [3]

It is decided to combine the scores of the students in Years 1 and 2.

(b) Find the mean and standard deviation for all the students in Years 1 and 2 combined.

Mean ....

Standard deviation ....

[7]

The scores of the students in Years 2 and 3 have already been combined. The combined mean for these two year groups is 73.4 . There are 149 students in Year 3.

(c) Find the mean score for Year 3.

.... [3]

[Question 10 continues on the next page]

	Number of students in year group	Mean of the scores	Standard deviation of the scores
Year 1	159	62	8
Year 2	141	68	10

It is later decided to adjust the score for each student in Year 1 by increasing each student's score by 10% of their original score.

(d) Find the mean and standard deviation for the students in Year 1 after this adjustment.

Mean ....

Standard deviation ....
[1]

A student from Year 2, who was absent on the day of the test, was included in the original data and given a score of 0.

(e) If that student's score were removed from the Year 2 data, tick to show what the effect would be on the mean and on the standard deviation for Year 2.

	It would increase	It would decrease	It would stay the same	There is not enough information to know
Mean				
Standard deviation				

[2]