Cambridge O Level

CANDIDATE NAME

CENTRE NUMBER

CANDIDATE NUMBER

STATISTICS **4040/22**

Paper 2 **October/November 2018**

2 hours 15 minutes

Candidates answer on the Question Paper.

Additional Materials: Pair of compasses
Protractor
Electronic calculator

READ THESE INSTRUCTIONS FIRST

Write your Centre number, candidate number and name in the spaces at the top of this page.
Write in dark blue or black pen.
You may use an HB pencil for any diagrams or graphs.
Do not use staples, paper clips, glue or correction fluid.
DO **NOT** WRITE IN ANY BARCODES.

Answer **all** questions.
If working is needed for any question it must be shown below that question.
Essential working must be shown for full marks to be awarded.
Electronic calculators should be used.

At the end of the examination, fasten all your work securely together.
The number of marks is given in brackets [ ] at the end of each question or part question.
The total number of marks for this paper is 100.

This document consists of **15** printed pages and **1** blank page.

CAMBRIDGE International Examinations

[Turn over

1   Gita is collecting data from the pupils in her school.
    Three of the questions that she asks in her survey are as follows:

Question A
How old are you in completed years?

Less than 10 years ☐   10 – 13 years ☐   14 – 16 years ☐   17 years or more ☐

Question B
How many siblings (brothers and sisters) do you have?

Question C
How do you travel to school?

Walk ☐   Cycle ☐   Bus ☐   Other ☐

(a) Give the upper and lower class boundaries for the 10 – 13 years class in Question A.

Lower class boundary...

Upper class boundary ....[1]

(b) Use statistical language to describe fully the type of data that Gita will collect from Question B and Question C.

Question B ....

Question C ....[3]

2 *X* and *Y* are two outcomes of an experiment such that:

$$P(X) = q \qquad P(Y) = 3/5 \qquad P(X \cup Y) = 7/10$$

Find the value of *q* if *X* and *Y* are

(a) mutually exclusive events,

....[2]

(b) independent events.

....[3]

3 A choir consists of 7 women and 3 men.
Three different members of the choir are to be chosen at random to sing solo parts.

Find the probability that more men than women are chosen to sing solo parts.

....[5]

 [Turn over

4 Pedro is conducting a survey of the 90 patients at his local hospital. He would like to know if they are satisfied with the care that they have received.

He is going to take a sample of the patients, and gives each a number from 00 to 89, allocated according to their age group and the department they are staying in.

The table below shows how the allocated numbers are divided between the two age groups.

50 years old or under	Over 50 years old
00 – 29	30 – 89

He decides to take a sample of **size 6** stratified by age group.

(a) Use the random number table below, starting at the beginning of the table, to select the stratified sample.

78   06   92   47   13   03   78   09   64   98   51   25   80   33   91   52

....[4]

Each patient is staying in one of three departments, A, B or C. The table below shows how the allocated numbers are divided between the departments and the age groups.

	50 years old or under	Over 50 years old
Department A	00 – 14	
Department B		30 – 59
Department C	15 – 29	60 – 89

(b) Ignoring the ages of the patients, comment on how well your sample represents the departments they are staying in.

....

....

....[3]

(c) State whether you would consider a sample stratified by age group or by department to be more useful in this case. Give a reason for your answer.

....

....[1]

5 David and the rest of his class take tests in Singing, Acting and Dancing.
The scores they obtain in each test are to be standardised to a mean of 40 and a standard deviation of 12.
The table below gives some information about his scores and those of the whole class.

Subject	David's raw score	David's standardised score	Class mean	Class standard deviation
Singing	46	....	58	9
Acting	....	....	37.5	10
Dancing	38	31	....	8

   (a) Find David's standardised score in Singing and enter it into the table above.

[2]

   (b) Find David's scores in Acting, given that his raw score and standardised score are the same, and enter them into the table above.

[2]

   (c) Find the class mean for Dancing, and enter it into the table above.

[2]

   (d) State the subject in which you would consider David to have performed best in relation to the rest of the class. Give a reason for your answer.

   ....

   ....[1]

   Later the class teachers decide to increase each standardised score by 10 marks.

   (e) Find the new mean and new standard deviation for the standardised scores.

New mean ....

New standard deviation ....[1]

6 Two types of tomato are Golden Yellow and Ruby Red.

The probability that each Golden Yellow seed will produce fruit is 0.8 and the probability that each Ruby Red seed will produce fruit is 0.75.

A packet of 50 tomato seeds contains 30 Golden Yellow seeds and 20 Ruby Red seeds.

(a) State the probability that a seed, randomly selected from the packet, is a Golden Yellow seed.

....[1]

(b) Find the probability that a seed, randomly selected from the packet, will produce fruit.

....[3]

(c) Find how many seeds from the packet you would expect to produce fruit.

....[2]

The manufacturer of the packets of seeds wishes to be able to state on each packet that the number of seeds expected to produce fruit is 40.

They put 35 Golden Yellow seeds into a packet.

(d) Find how many Ruby Red seeds must be in this packet.

....[2]

7 **(a)** The following time series graphs are for variable *V* and variable *W*.

In each case, state a suitable value for *n* if *n*-point moving average values are to be calculated. Also, in each case, state, with a reason, whether or not it would be necessary to centre the moving average values.
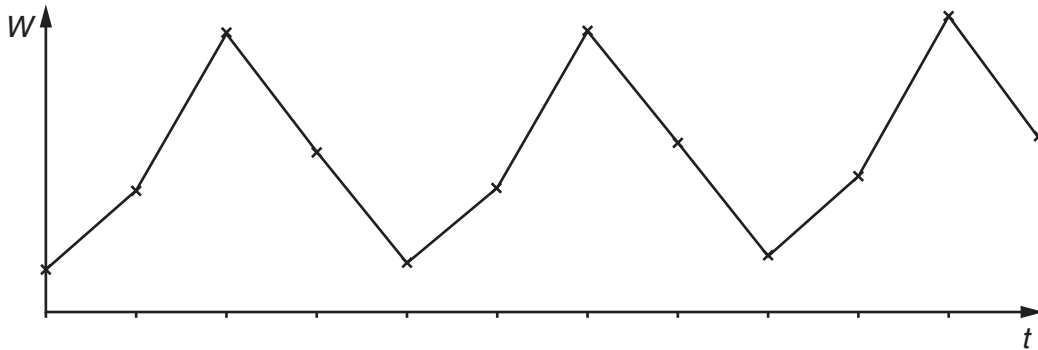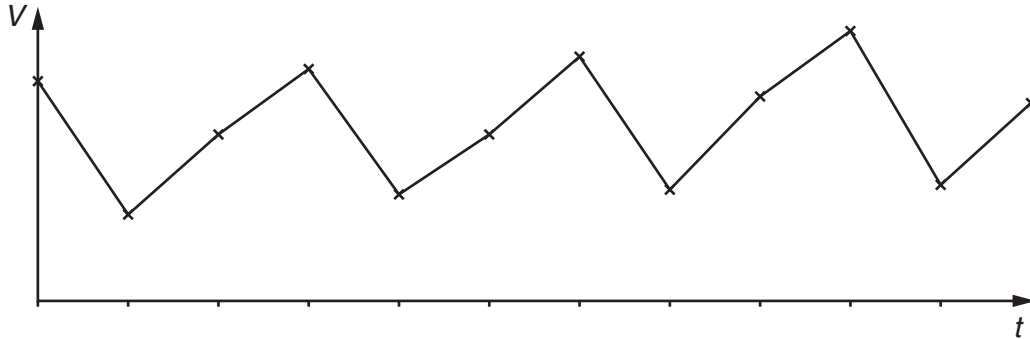
Variable *V*: *n* = ...

....

....

Variable *W*: *n* = ...

....

....[4]

A baker wishes to predict the profit he will make in Jan – Apr 2019.
He collects data on the profits he made every four months over the last 3 years.

| Four month period	Profit ($)	3-point moving total ($)	3-point moving average ($)
Sep – Dec 2015	9200		
Jan – Apr 2016	7800	27100	9033
May – Aug 2016	10100	27900	9300
Sep – Dec 2016	10000	28400	9467
Jan – Apr 2017	8300	29800	9933
May – Aug 2017	11500	31700	10570
Sep – Dec 2017	11900	33200	11070
Jan – Apr 2018	9800		
May – Aug 2018	14400		

(b) Complete the final two entries in the table.

[2]

(c) Use values from the table to make an estimate for the seasonal component for Jan – Apr.

....[3]

(d) Plot the 3-point moving average values on the grid below and draw a straight trend line.
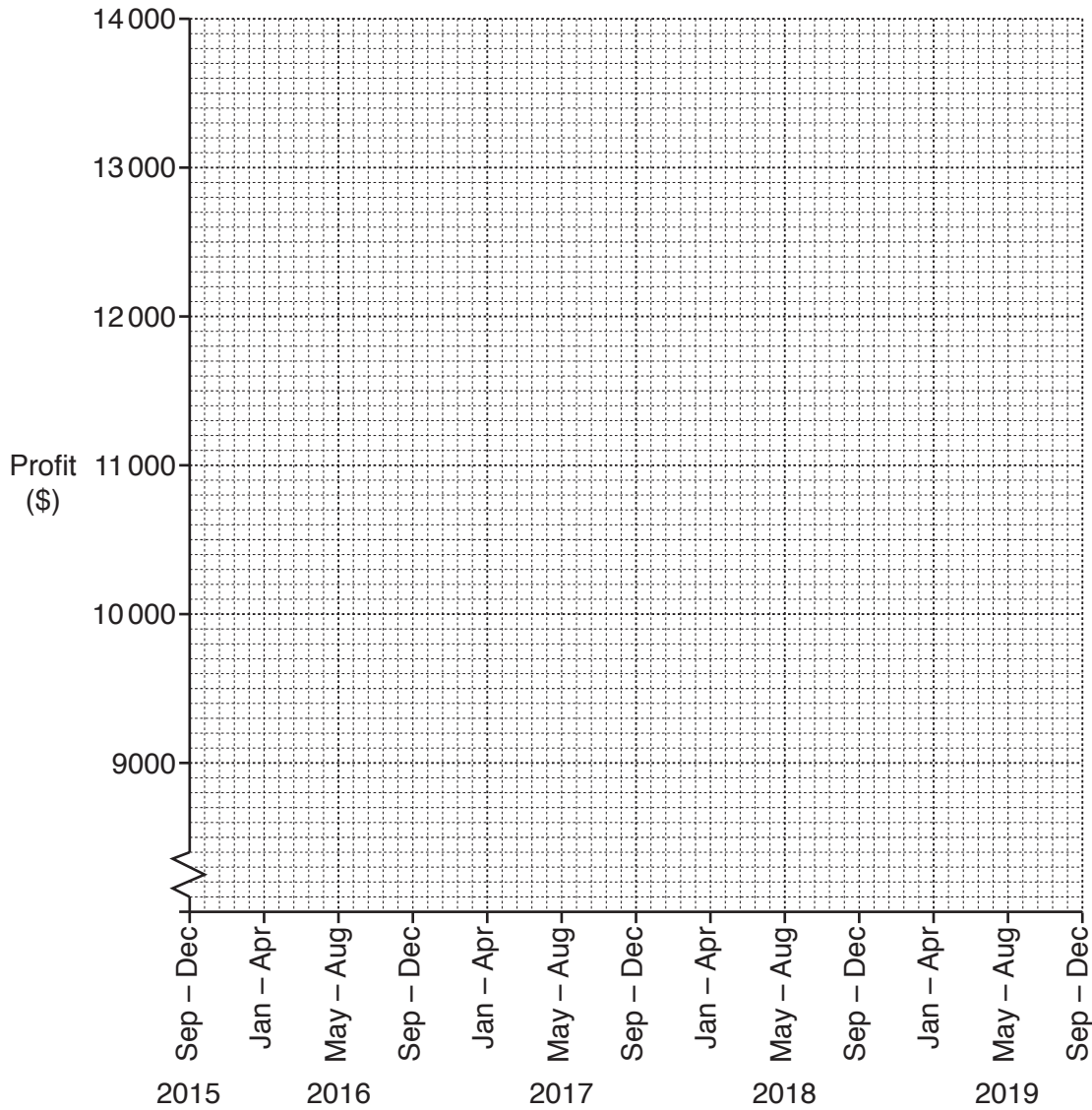
[3]

(e) Explain what the trend line you have drawn tells you.
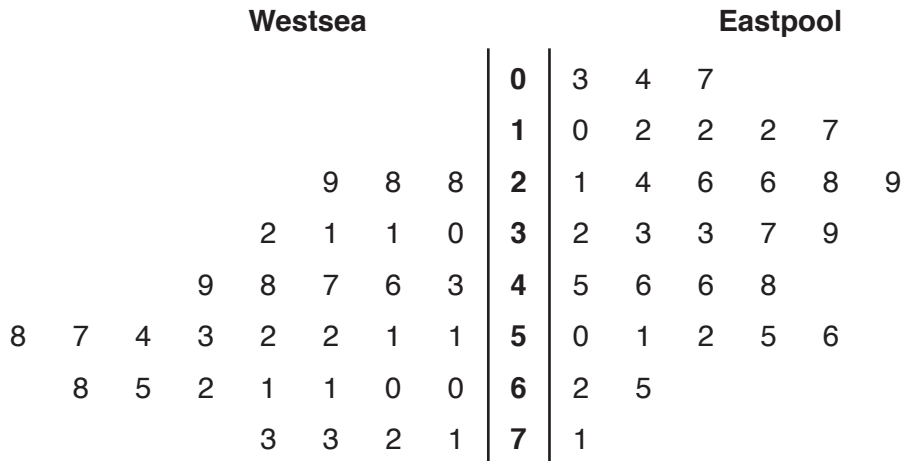
....
....[1]

(f) Use your answers to parts **(c)** and **(d)** to predict the baker's profit in Jan – Apr 2019.

....[2]

(g) Looking again at the points you have plotted in part **(d)**, suggest a reason why the prediction you have made may prove to be inaccurate.

....
....[1]

8 The back-to-back stem-and-leaf diagram below shows the number of hours of sunshine, measured to the nearest 0.1 hours, in the towns of Westsea and Eastpool, for each of the 31 days in May.

	Westsea											**Eastpool**					
											0	3	4	7			
											1	0	2	2	2	7	
						9	8	8		**2**	1	4	6	6	8	9
					2	1	1	0		**3**	2	3	3	7	9	
				9	8	7	6	3		**4**	5	6	6	8		
8	7	4	3	2	2	1	1		**5**	0	1	2	5	6		
	8	5	2	1	1	0	0		**6**	2	5				
			3	3	2	1		**7**	1						

Key: 8 | **2** | 1 represents 2.8 hours of sunshine in Westsea and 2.1 hours of sunshine in Eastpool

(a) Give one advantage that a stem-and-leaf diagram has over a box-and-whisker diagram.
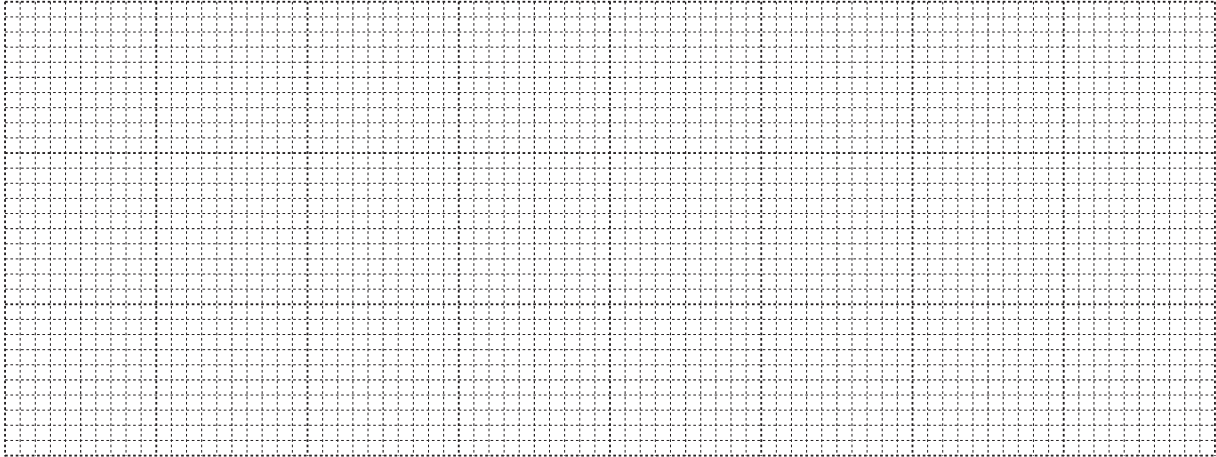
....

....[1]

(b) Complete the table below for the data in the stem-and-leaf diagram.

	Least value	Lower quartile	Median	Upper quartile	Greatest value
Eastpool					
Westsea					

[5]

(c) On the grid below, draw two box-and-whisker diagrams, one for the daily hours of sunshine in May in Eastpool and one for the daily hours of sunshine in May in Westsea, using a single axis.

[4]

(d) Use the box-and-whisker diagrams to compare the daily hours of sunshine in Eastpool and the daily hours of sunshine in Westsea in that May.

....

....

....[2]

(e) Name the statistical measure represented in a box-and-whisker diagram by

   (i) the overall length of the diagram,

   ....[1]

   (ii) the length of the box in the diagram.

   ....[1]

9 A gardener divides her costs into four categories: Seeds, Equipment, Fertilizer and Other Costs. The table below is to show the price relatives for each of these categories, taking 2015 as base year.

	Price relative		
	2015	2016	2017
Seeds	100	107	121
Equipment	100	*a*	109
Fertilizer	100	*b*	111
Other Costs	100	*c*	105

(a) Find the price relative for Seeds in 2017, taking 2016 as base year.

....[2]

Equipment costs increased by 4% between 2015 and 2016.
Fertilizer costs, per kg, increased from $1.23 in 2015 to $1.33 in 2016.
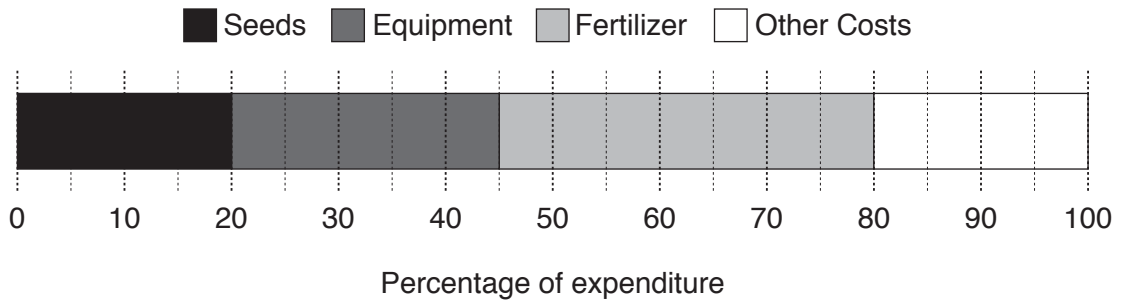Other Costs decreased by 3% between 2015 and 2016.

(b) Find the values of *a*, *b* and *c*.

a = ...

b = ...

c = ...[4]

The total expenditure in 2015 was divided between the four categories as shown in the percentage sectional bar chart.

■ Seeds ■ Equipment □ Fertilizer □ Other Costs

| 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |

Percentage of expenditure

(c) Using expenditure for weights, calculate, correct to 1 decimal place, a weighted aggregate cost index for 2017, taking 2015 as base year.

....[4]

(d) Explain what your answer to part **(c)** tells you.

....

....[2]

The gardener spent a total of $352 in 2015.

(e) Use your answer to part **(c)** to find an estimate for the total expenditure in 2017.

....[2]

Later it was discovered that the total expenditure in 2017 was actually greater than this estimate.

The gardener's children, Gouta and Suma, each gave a possible explanation for the low estimate.
Gouta said, 'It is because the cost of seeds went up by 21% between 2015 and 2017.'
Suma said, 'It is because more seeds were used in 2017 than in 2015.'

One of her children gave a correct explanation.

(f) For each child state, with a reason, whether the explanation given for the low estimate is correct or not.

Gouta ....

....

Suma ....

....[2]

10 Amy sells fish in a market. On one particular day she has 14 fish to sell, with a total mass of 7.49 kg. In the morning she sells nine of the fish, which have a mean mass of 0.72 kg. She sells the rest of the fish in the afternoon.

(a) By finding the mean mass of the fish she sells in the afternoon, make a comparison between the fish she sells in the morning and the fish she sells in the afternoon.

....

....[5]

The table below gives some information about all the fish that she sold that week, with masses measured in kg.

Number of fish	Sum of the masses	Sum of the squares of the masses
84	39.96	28.29

(b) Calculate the standard deviation of the masses of the fish she sold in that week.

....[2]

(c) State an advantage that the interquartile range has, as a measure of dispersion, over the standard deviation.

....

....[1]

The cumulative frequency distribution below gives further details about the masses of the fish sold that week.

Mass, m (kg)	Frequency	Cumulative frequency
$0.05 \leqslant m < 0.1$	6	6
$0.1 \leqslant m < 0.25$	21	27
$0.25 \leqslant m < 0.5$	26	53
$0.5 \leqslant m < 0.75$	17	70
$0.75 \leqslant m < 1$	9	79
$1 \leqslant m < 1.5$	4	83
$1.5 \leqslant m < 2$	1	84

An estimate, using linear interpolation, for the lower quartile mass of the fish is 0.207 kg.

(d) Use linear interpolation to calculate an estimate for the upper quartile mass and hence find an estimate for the interquartile range of the masses of the fish.

Upper quartile ....

Interquartile range ....[5]

Amy finds it difficult to sell fish that are under 0.2 kg, and often needs to reduce the price of these fish.

(e) Use linear interpolation to calculate an estimate for the number of fish that were under 0.2 kg that week.

....[3]

16

BLANK PAGE