Cambridge O Level

Cambridge Assessment International Education
Cambridge Ordinary Level

CANDIDATE NAME			
CENTRE NUMBER		CANDIDATE NUMBER	

2592749566

STATISTICS **4040/22**

Paper 2 **October/November 2019**

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.

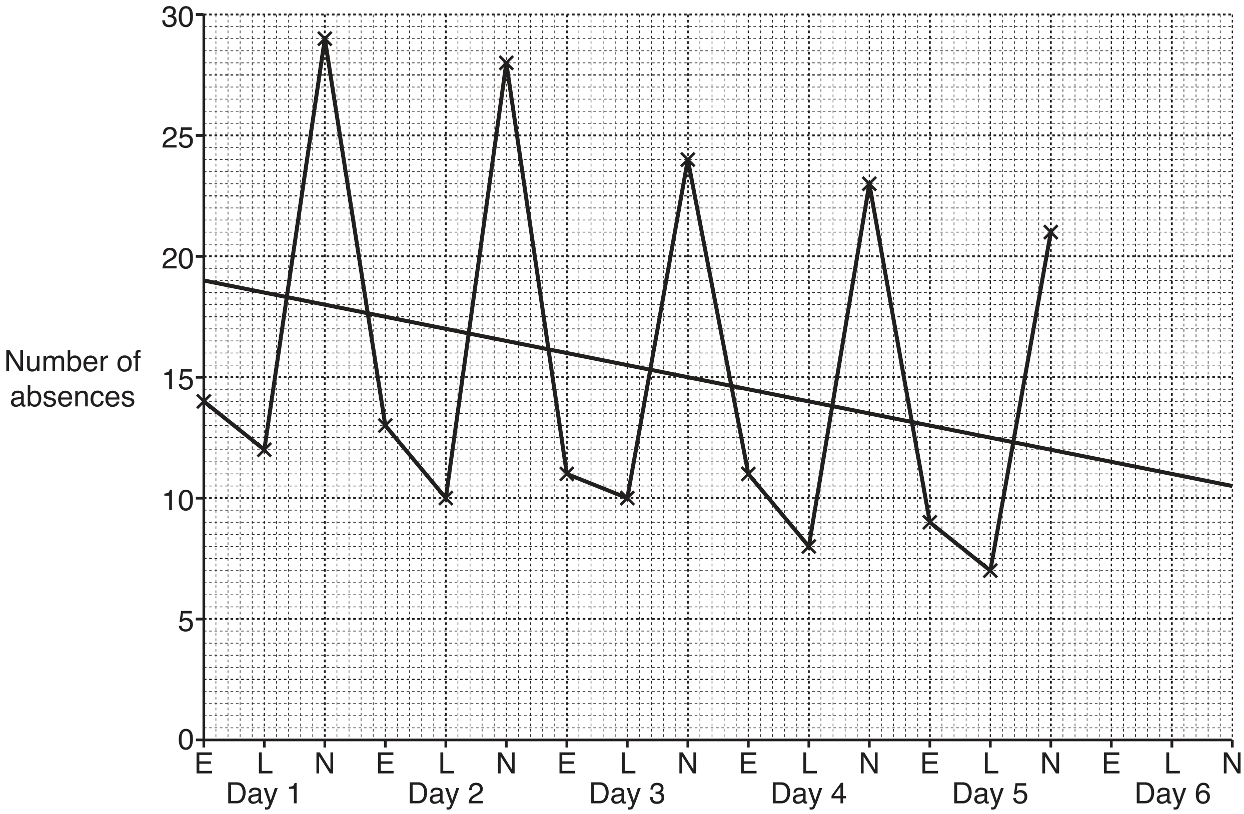
This document consists of **16** printed pages.

DC (JM/AR) 170616/2

Cambridge Assessment International Education

[Turn over

1 Each employee at a factory works one of three possible shifts, either an Early (E) or a Late (L) or a Night (N) shift each day. The time series graph and trend line give information about the number of people absent for each shift over a five-day period.



The factory owner has been trying to reduce the number of people who are absent.

(a) Use the graph above to decide whether or not you think he has been successful over these five days. Give a reason for your answer.

....

.... [1]

(b) Use the time series graph and trend line to find an estimate for the seasonal component of the Late (L) shift.

.... [3]

(c) Use the graph and your answer to part **(b)** to estimate the number of people who will be absent on the Late (L) shift of day 6.

.... [2]

2 An inspector at a bus company wishes to conduct mechanical tests on a sample of the 40 buses parked at the bus station overnight. He numbers the buses from 00 to 39.

(a) Use the random number table below, starting at the beginning of the table, to select a simple random sample of **size 8**.

52 31 01 24 00 12 79 27 31 11 36 83 68 08 35 98 16 27

.... [2]

18 of the buses (numbered 00 to 17) are small local buses and 22 of the buses (numbered 18 to 39) are large long-distance buses.

(b) Comment on whether or not your sample is representative in terms of the two different types of bus.

....

....

.... [3]

3 The students at a school compete in the high jump and the long jump, recording their best result for each event. The table below gives information about the recorded results and scaled results for two of the students, Makena and Ebele, together with the means and standard deviations for the whole school.

	Recorded results (m)		Scaled results (m)	
	High jump	Long jump	High jump	Long jump
Makena	1.72	5.16	68	58
Ebele	1.65	4.95	z	44
Mean	x	5.04	50	50
Standard deviation	0.06	y	12	12

(a) Find the values of x, y and z.

x = ...

y = ...

z = ... [6]

(b) Explain why it is useful to calculate scaled results for these events.

....

.... [1]

4 A speed camera on a rural road records, to the **nearest km/h**, the speeds of the vehicles that pass it. Shown below are the recorded speeds of the first 16 vehicles, in order, as they pass it one day.

54 60 63 26 26 29 28 32 35 34 53 54 59 61 63 58

(a) Display the data using a stem-and-leaf diagram.

[4]

(b) State the maximum possible range of the speeds of these vehicles.

.... [1]

(c) Suggest a possible explanation for the shape of the distribution of the data.

....

.... [1]

(d) Alternatively the data could have been displayed using a histogram with class widths of 10 km/h.

State an advantage that the stem-and-leaf diagram has over a histogram.

....

.... [1]

5 A and B are two events such that

$$P(A) = 0.75 \quad \text{and} \quad P(B) = 0.6\ .$$

(a) Show that A and B cannot be mutually exclusive events.

[1]

It is also known that $P(A \cup B) = 0.85$.

(b) Find $P(A \cap B)$.

.... [2]

(c) Show whether or not A and B are independent events.

[2]

(d) Find the probability that either event A or event B, but not both, occurs.

.... [1]

(e) Find the probability that neither event A nor event B occurs.

.... [1]

6 On 1 January 2016 the ages, measured in completed years, of the 60 workers at a company were collected and are summarised in the table below.

Age	*Frequency*	
18–24	12	
25–29	23	
30–39	18	
40–59	7	

(a) Show that the mid-point of the 18–24 class is 21.5 .

[1]

(b) Calculate estimates for the mean and the standard deviation of the ages of the workers on 1 January 2016.

Mean = ...

Standard deviation = ... [6]

On 1 January 2018 the same 60 workers were still employed at the company.

(c) Find estimates for the mean and standard deviation of the ages of the workers on 1 January 2018.

Mean = ...

Standard deviation = ... [1]

7 An article from the World News section of a newspaper is analysed by counting the number of letters in each word of the article. Abbreviations are considered to be single words.

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

.... [2]

The article contains 102 words and has a mean of 4.5 letters per word.
The article must be edited so that it contains exactly 100 words. In order to do this, the editor replaces the 3 words '**World Health Organisation**' with the abbreviation '**WHO**'.

(b) Find the new mean number of letters per word.

.... [5]

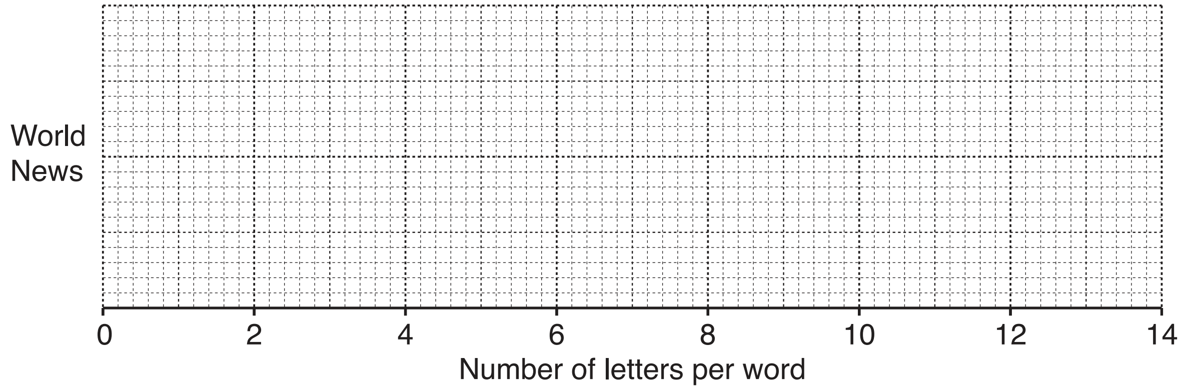
The table below gives some information about the number of letters per word in the final edited article.

(c) Complete the table.

	Number of letters per word
Smallest	1
Largest	
Mean	
Median	4
Lower quartile	
Upper quartile	6.5
Interquartile range	3.5
Standard deviation	2.3
Range	12

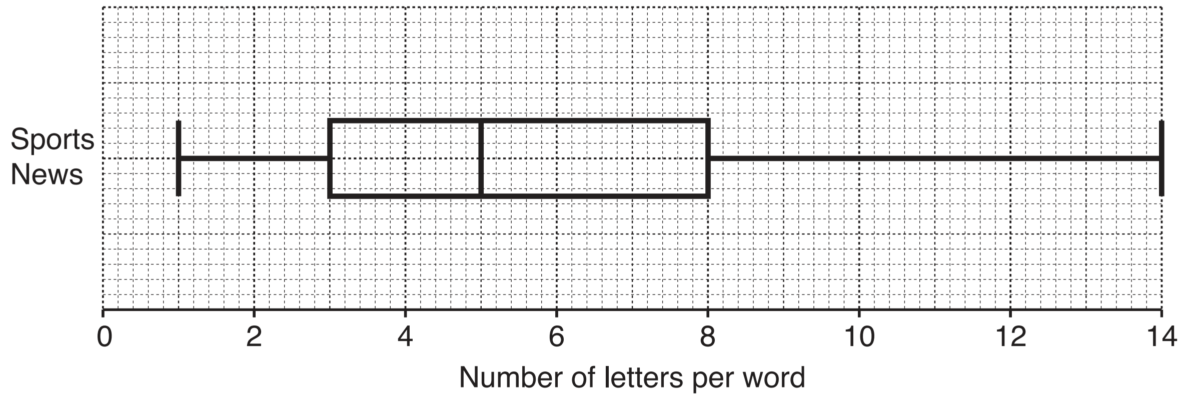
[2]

(d) Use appropriate information from the table to draw a box-and-whisker diagram for the number of letters per word in the World News article.



[3]

A 100-word article from the Sports News section of the newspaper is also analysed and the box-and-whisker diagram is shown below.



(e) Make two comparisons between the number of letters per word in the World News article and the number of letters per word in the Sports News article.

....

....

.... [2]

8 Amare wishes to calculate a weighted aggregate cost of living index for students at the university he attends.
He calculates weights based on the spending patterns of a stratified sample of the students.

(a) Suggest one appropriate factor that Amare might use to divide the students into strata. Give a reason for your suggestion.

....

....

.... [2]

Amare divides a student's cost of living into five categories. The base year for the price relatives in the table below is 2016.

Category	*Weight*	*Price relative*	
		2017	2018
Housing	35	103	105
Food	20	100	102
Transport	3		111
Clothing	2		98
Entertainment	12		115

(b) Find the price relative of Housing in 2018, taking 2017 as the base year. Show your working.

.... [2]

(c) Explain what the value of 100 in the table tells you.

....

....[1]

(d) Use the following information to complete the table.

Transport prices increased by 9% between 2016 and 2017.
Clothing prices decreased by 5% between 2016 and 2017.
Entertainment prices increased by 8% between 2017 and 2018.

[3]

(e) **(i)** Calculate, correct to one decimal place, a weighted aggregate cost of living index for 2018, taking 2016 as the base year.

.... [3]

(ii) Explain what this figure suggests.

....

....

.... [2]

(iii) Explain why it would be inappropriate to use this figure for all the people in the country where Amare lives.

....

.... [1]

Amare spent a total of $6300 in 2016.

(f) Estimate, correct to 3 significant figures, his expenditure in 2018.

.... [2]

9 A ranger in a nature reserve counts the geese that she can see and displays her results in a two-way table.

	With tag	Without tag
Barnacle geese	25	14
Pink-footed geese	17	19
Canada geese	10	5

A goose is to be selected at random.

Find the probability that

(a) it is a pink-footed goose without a tag,

.... [2]

(b) it has a tag,

.... [1]

(c) it is a barnacle goose, given that it has a tag.

.... [1]

Two geese are to be selected without replacement.

(d) Find the probability that one is a barnacle goose and the other is a pink-footed goose.

.... [3]

Barnacle geese lay their eggs in nests high up in cliffs away from predators. Young barnacle geese (goslings), unable to fly, jump from these nests to the beach below when they are a few days old. An island of Greenland, with east cliffs and west cliffs, has many such nests. The probabilities of goslings surviving the jump are shown in the table.

Location of nest	*Probability of surviving*
East cliffs	0.86
West cliffs	0.72

The goslings then try to reach the sea without being caught by arctic foxes.
For those that survive the jump there is a probability of 0.1 that they will **not** reach the sea.

(e) Show that the probability of a gosling from a nest in the east cliffs reaching the sea is 0.774 .

[2]

In one particular year there are 83 goslings in nests in the east cliffs and 57 goslings in nests in the west cliffs.

(f) Find the expected number of goslings to reach the sea from these cliffs in that year.

.... [3]

In another year there were twice as many goslings in nests in the east cliffs as in nests in the west cliffs and 79 goslings reached the sea.

(g) Calculate an estimate for the total number of goslings in nests that year.

.... [4]

10 Hazel travels to work by train. She takes a train that departs at the same time each day. On each of 50 mornings she records how late her train arrives at her destination station using the tally chart below.

Number of minutes late, x	*Tally*	*Frequency*
$0 \leqslant x < 8$	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~	20
$8 \leqslant x < 16$	~~IIII~~ ~~IIII~~ ~~IIII~~	15
$16 \leqslant x < 24$	~~IIII~~ ~~IIII~~	10
$24 \leqslant x < 32$	IIII	4
$32 \leqslant x < 40$	I	1

From the station she has a 5-minute walk to work and must arrive by 8 am.
She is late for work on 11 of these days.

(a) Use linear interpolation to calculate an estimate, to the nearest minute, for the time that the train arrives at her destination station when it is on time.

.... [5]

Hazel complains to the train company about the delays to her train. She decides to quote the largest measure of central tendency in her complaint.

(b) State which measure of central tendency she should quote. Justify your answer, with reference to the data in the table.

....

....

.... [2]

The train company offers to refund her money, as shown in the table below.

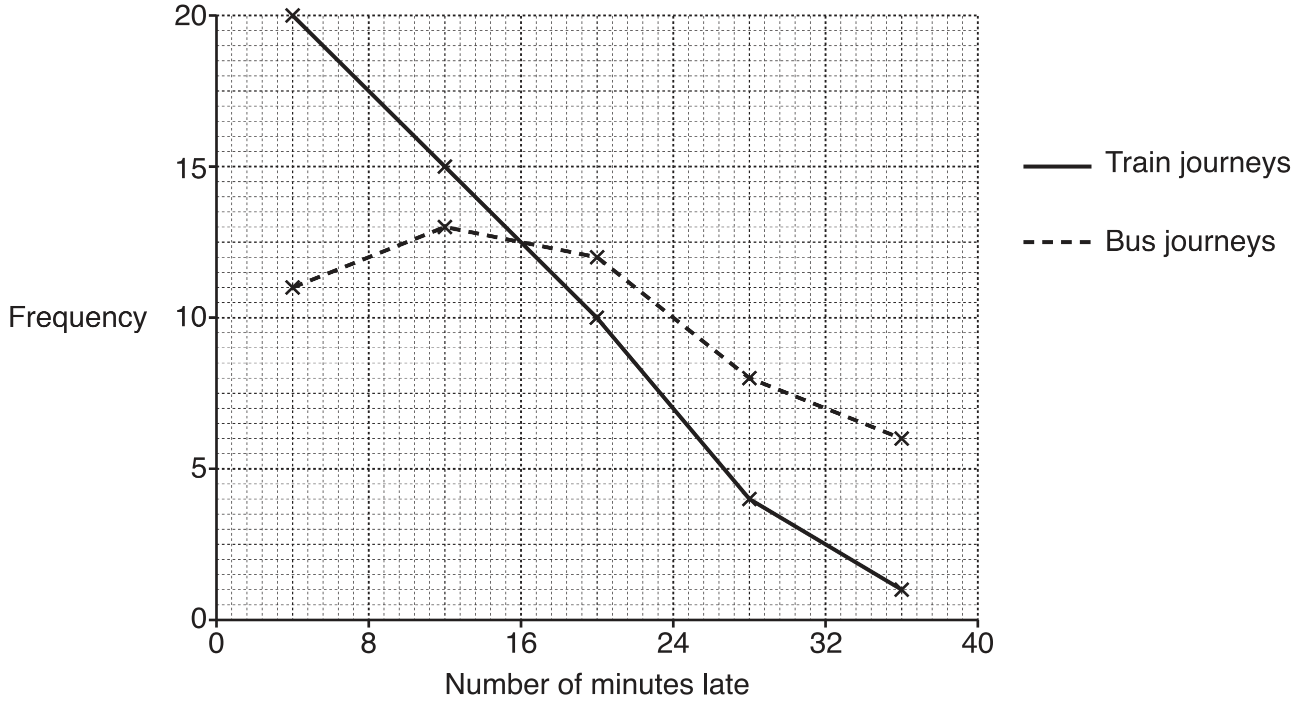
Number of minutes late	*Amount of money refunded*
Less than 15	Nothing
From 15 to under 30	50% of the cost of the journey
30 or more	100% of the cost of the journey

(c) Hazel paid $14.60 for each journey. Use linear interpolation to calculate an estimate for the total amount of money she can get back in refunds due to delays to these 50 journeys.

.... [5]

[Question 10 continues on the next page]

For the next 50 days she travels to work by bus and records how late her bus arrives at its destination. Her results for the train journeys and the bus journeys are shown in the frequency polygons below.



(d) State the number of days that her bus was less than 16 minutes late.

.... [1]

(e) Use the frequency polygons to compare the number of minutes late for the train journeys with the number of minutes late for the bus journeys.

....

.... [1]