Cambridge
O Level

**Cambridge International Examinations**
Cambridge Ordinary Level

| | |
|---|---|
| CANDIDATE NAME | |
| CENTRE NUMBER | |
| CANDIDATE NUMBER | |

*1608516764*

**STATISTICS** **4040/22**

Paper 2 **October/November 2017**

**2 hours 15 minutes**

Candidates answer on the Question Paper.

Additional Materials: Pair of compasses
Protractor

**READ THESE INSTRUCTIONS FIRST**

Write your Centre number, candidate number and name on all the work you hand in.
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 in Section A and not more than **four** questions from Section B.
If working is needed for any question it must be shown below that question.
The use of an electronic calculator is expected in this paper.

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.

This document consists of **18** printed pages and **2** blank pages.



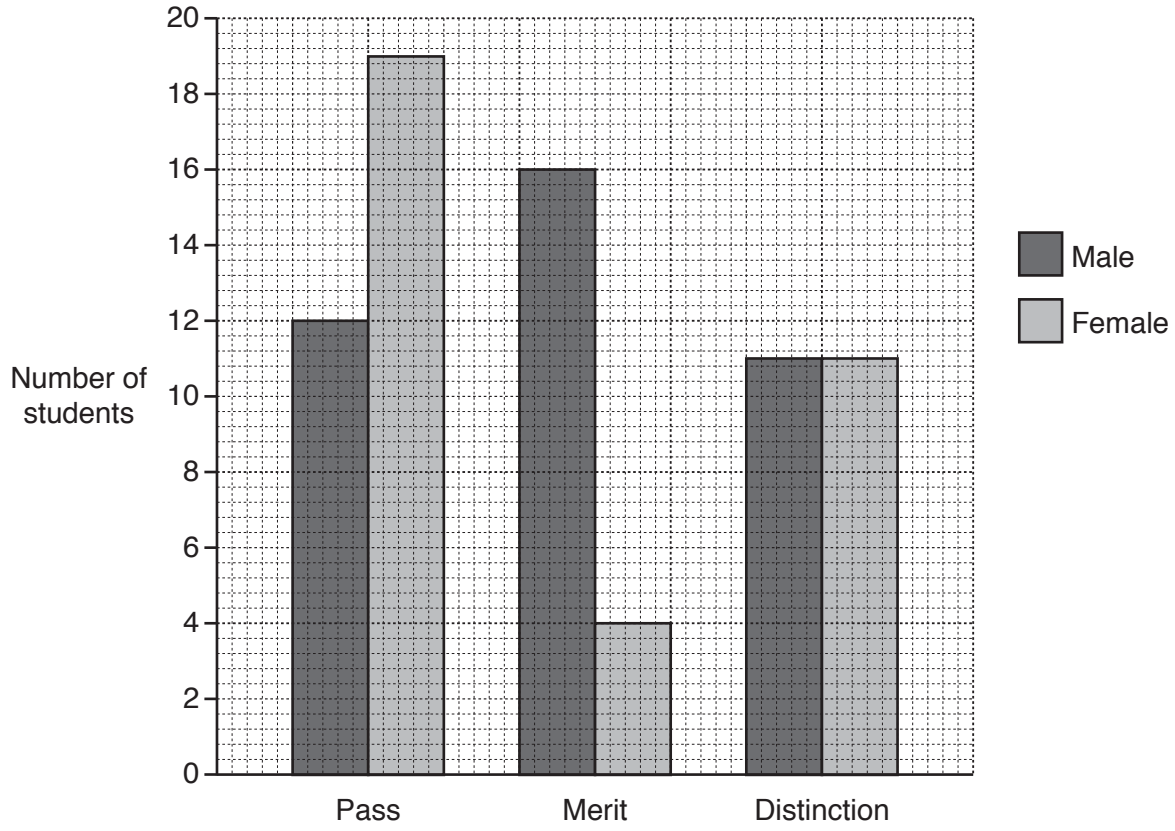

CAMBRIDGE
International Examinations

**Section A** [36 marks]

Answer **all** of the questions 1 to 6.

**1** The students who completed a catering course at a college each achieved one of a pass, a merit or a distinction.

The dual bar chart shows the number of students of each gender achieving each grade.

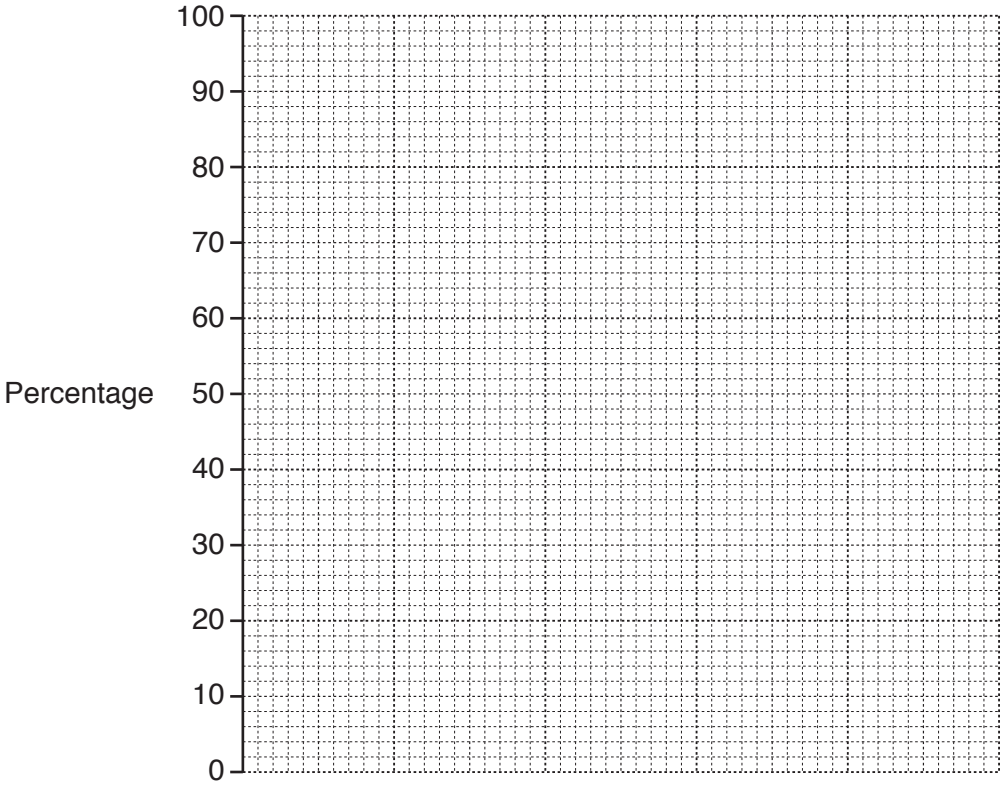



**(i)** Find the total number of male students and the total number of female students who completed the course.

Male students .....

Female students .....[1]

**(ii)** On the grid below, display the data as a percentage sectional bar chart with one bar for each gender.




[4]

**2** Data has been collected from the last Olympic Games.

**(i)** In each case below, put a tick in the column with the correct description of the type of data.

| *Data* | *Qualitative variable* | *Discrete quantitative variable* | *Continuous quantitative variable* | *Not a variable* |
|---|---|---|---|---|
| The height of each competitor | | | | |
| The country of origin of each competitor | | | | |
| The number of events entered by each country | | | | |
| The name of each competitor | | | | |
| The winning time for the men's 110 m hurdles | | | | |

[4]

The ages, in completed years, of the archery competitors are grouped into classes labelled 16 – 18 , 19 – 21 , 22 – 24 etc.

**(ii)** State the lower and upper class boundaries of the 19 – 21 class.

Lower class boundary .....

Upper class boundary .....[1]

The masses, measured to the nearest kg, of the archery competitors are grouped into classes labelled 45 – 49 , 50 – 54 , 55 – 59 etc.

**(iii)** State the lower and upper class boundaries of the 50 – 54 class.

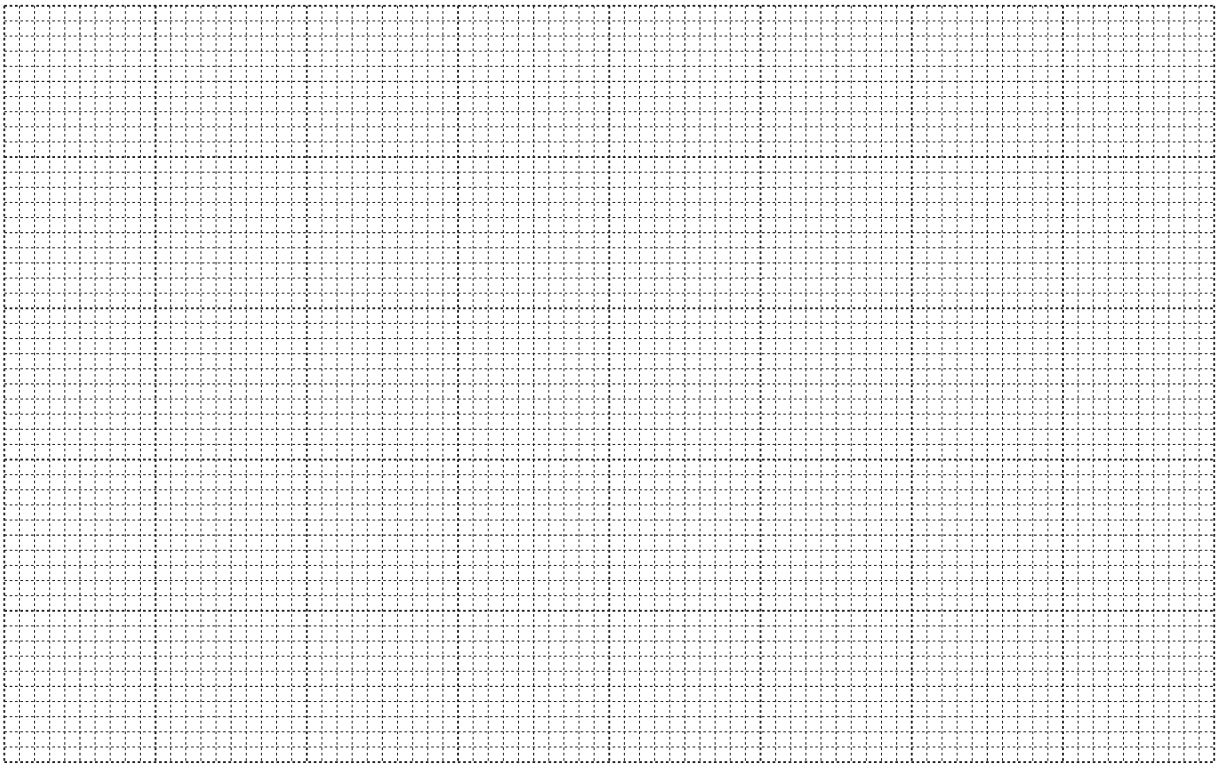
Lower class boundary .....

Upper class boundary .....[1]

**3** The shoulder heights, in centimetres, of 50 male and 50 female African elephants in a herd were measured.

| *Shoulder height, h* (cm) | *Number of male elephants* | *Number of female elephants* |
|---|---|---|
| $0 \leqslant h < 50$ | 0 | 0 |
| $50 \leqslant h < 100$ | 1 | 2 |
| $100 \leqslant h < 150$ | 3 | 4 |
| $150 \leqslant h < 200$ | 4 | 24 |
| $200 \leqslant h < 250$ | 21 | 15 |
| $250 \leqslant h < 300$ | 17 | 4 |
| $300 \leqslant h < 350$ | 4 | 1 |
| $350 \leqslant h < 400$ | 0 | 0 |

**(i)** On the grid below, draw two frequency polygons so that the data for the male and female elephants can be compared.



[4]

**(ii)** Using your frequency polygons, compare the shoulder heights of the male and female elephants.

.....

.....[1]

**4** **(a)** $A$ and $B$ are independent events such that

$$P(A) = 0.5 \quad \text{and} \quad P(A \cap B) = 0.25 .$$

**(i)** Find $P(B)$.

.....[2]

**(ii)** If event $A$ is 'obtaining a head when a fair coin is thrown', give a possible example for event $B$.

.....

.....[1]

**(b)** $C$ and $D$ are mutually exclusive events such that

$$P(C) = 0.62 \quad \text{and} \quad P(D) = 0.21 .$$

Find $P(C \cup D)$ and $P(C \cap D)$.

$P(C \cup D)$ = .....

$P(C \cap D)$ = .....[3]

**5** The houses in a village are numbered 1 to 40.
A systematic sample of the houses is to be selected.

**(i)** Use the random number table below to help you select a systematic sample of size five from the houses numbered 1 to 40.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 23 | 02 | 28 | 14 | 79 | 26 | 09 | 97 | 42 | 38 | 64 | 75 | 12 | 21 | 66 | 04 |
| 22 | 47 | 51 | 27 | 31 | 84 | 52 | 17 | 36 | 65 | 40 | 79 | 94 | 08 | 92 | 16 |

.....[3]

The odd-numbered houses are on one side, and the even-numbered houses are on the other side of the road through the village.
A new railway is proposed that will pass close to the houses on one side of the road.
All the adults living in each of the selected houses are to be asked for their opinions on the proposal.

**(ii)** **(a)** Give two reasons why the sample selected in part **(i)** may not provide opinions which represent those of all the adults living in the village.

Reason 1 .....

.....

Reason 2 .....

.....[2]

**(b)** Suggest an alternative sampling method that might produce a more representative sample.

.....

.....[2]



**[Turn over**

**6** A shop selling shirts has 25 large shirts which are either long- or short-sleeved and either white, blue or cream, as shown in the table.

| | White | Blue | Cream | TOTAL |
|---|---|---|---|---|
| Long-sleeved | 7 | 2 | 6 | 15 |
| Short-sleeved | 3 | 4 | 3 | 10 |
| TOTAL | 10 | 6 | 9 | 25 |

**(i)** If a man selects a large shirt at random from the shop, find the probability that it is

**(a)** a blue, short-sleeved shirt,

.....[1]

**(b)** either a white shirt or a cream shirt,

.....[1]

**(c)** blue, given that it is long-sleeved,

.....[1]

**(d)** either a white shirt or a short-sleeved shirt, or both.

.....[1]

**(ii)** If the man selects two large shirts at random from the shop, find the probability that they are the same colour.

.....[3]

**BLANK PAGE**

**[Turn over for Section B]**

**Section B** [64 marks]

Answer not more than **four** of the questions 7 to 11.

Each question in this section carries 16 marks.

**7** A garage owner noted the number of kilometres, $k$, that each of 120 cars brought in for a service had completed.

| *Number of km*, $k$ | *Number of cars* |
|---|---|
| $10\,000 \leqslant k < 15\,000$ | 5 |
| $15\,000 \leqslant k < 20\,000$ | 20 |
| $20\,000 \leqslant k < 25\,000$ | 33 |
| $25\,000 \leqslant k < 30\,000$ | 29 |
| $30\,000 \leqslant k < 40\,000$ | 24 |
| $40\,000 \leqslant k < 50\,000$ | 6 |
| $50\,000 \leqslant k < 60\,000$ | 0 |
| $60\,000 \leqslant k < 80\,000$ | 3 |

**(i)** Find the maximum possible value of the range of these data.

.....[1]

There are some extreme values in these data.

**(ii)** **(a)** State the most appropriate measure of central tendency to represent these data.

.....[1]

**(b)** Give two possible extreme values which might have been in the original raw data.

.....

.....[2]

**(iii)** Use linear interpolation to calculate an estimate of the interquartile range of these data.

.....[7]

A basic service at this garage costs $45.
Cars which have completed less than 22 000 km are given a 20% reduction on the cost of a basic service.

**(iv)** Use linear interpolation to find an estimate of the total amount collected by the garage for the basic services on these 120 cars.

.....[5]

**8** A charity fundraiser divides her costs into three categories: Leaflets, Phone Calls and Petrol. A summary, for last year, of the quantities used and the costs are shown in the table below.

| | | |
|---|---|---|
| *Leaflets* | 1200 were delivered | The cost of each was \$0.12 |
| *Phone Calls* | 600 minutes of calls were made | The cost per minute was \$0.40 |
| *Petrol* | 20 litres were used | The cost per litre was \$1.20 |

A weighted aggregate cost index for this year is to be calculated.

**(i)** Use the information in the table to find appropriate weights, based on expenditure last year, and express them as a ratio in its lowest terms.

.....[3]

This year,
the cost of the leaflets has increased by 3%,
the cost per minute of phone calls has decreased by \$0.02,
and the cost per litre of petrol is \$1.26 .

**(ii)** Use the information above to find price relatives for each category for this year, using last year as base.

Leaflets .....

Phone Calls .....

Petrol .....[5]

**(iii)** **(a)** Use your answers to parts **(i)** and **(ii)** to find a weighted aggregate cost index for this year, taking last year as base. Give your answer correct to one decimal place.

.....[3]

**(b)** Explain clearly what your answer suggests about the charity fundraiser's costs for this year.

.....

.....

.....[3]

A forecast for expenditure this year calculated using the index found in part **(iii)(a)** may not be accurate, if the weights have changed.

**(iv)** Give two possible reasons why the weights may have changed.

Reason 1 .....

.....

Reason 2 .....

.....[2]



**[Turn over**

**9** The probability that Baruti's alarm clock will wake him is 0.8 .

If Baruti's alarm clock wakes him the probability that he will be late for work is 0.1 .
If Baruti's alarm clock does not wake him the probability that he will be late for work is 0.7 .

**(i)** Show that on any one day the probability that Baruti will be late for work is 0.22 .

.....[4]

Baruti is paid $16.50 per day, but he must pay back $2.00 per day for each day that he is late for work.

**(ii)** Find his expected earnings per day.

.....[3]

He works for 50 days at the above rate of pay.

**(iii)** Find the number of these days that you would expect him to be late for work and find his total expected earnings for this period.

Expected number of days late .....

Total expected earnings for the 50 days .....[2]

His employer then decides to increase his pay, but also to increase to $3.00 the amount that he must pay back for being late.

**(iv)** Find his new rate of pay if his expected earnings are to remain the same.

.....[4]

Baruti works with Jake and they arrive for work independently.
The probability that Jake is late for work is 0.16 .

**(v)** Find the probability that, on any one day, at least one of these men is late for work.

.....[3]



**[Turn over**

**10** 50 people in a running race each completed four circuits of a track. The table below summarises the times taken, in seconds (s), for the first and second circuits.

| *Circuit* | *Mean* (s) | *Standard deviation* (s) |
|---|---|---|
| First circuit | 52.3 | 3.2 |
| Second circuit | 57.6 | 4.8 |

**(i)** Use the table above to state two differences between the times taken for the first circuit and the times taken for the second circuit.

1 .....

.....

2 .....

.....[2]

Zara took 57.1s to complete the first circuit and 63.6s to complete the second circuit.

**(ii)** Scale Zara's times to a distribution with a mean of 0s and a standard deviation of 1s.
Hence state, with a reason, in which of the two circuits you would consider Zara to have performed better, in relation to the rest of the people.

Scaled time for first circuit .....

Scaled time for second circuit .....

.....

.....[3]

The 50 people completed all four circuits of the track and their times, $t$, in seconds, for the whole race are summarised in the table below.

| *Time to complete race, t* (s) | *Number of people* | |
|---|---|---|
| $200 \leqslant t < 240$ | 5 | |
| $240 \leqslant t < 280$ | 33 | |
| $280 \leqslant t < 320$ | 12 | |

**(iii)** **(a)** Use an assumed mean of 260 s to find estimates for the mean and the standard deviation of the times to complete the race.

Mean .....

Standard deviation .....[9]

**(b)** Explain why these answers are only estimates.

.....

.....[1]

**(c)** State a change that could have been made to the grouped frequency table above, so that better estimates could be obtained.

.....

.....[1]



**[Turn over**

**11** The number of meals consumed in a school canteen is recorded each day for two weeks.

| *Week* | *Day* | *Number of meals* | |
|---|---|---|---|
| 1 | Monday | 252 | |
| | Tuesday | 265 | |
| | Wednesday | 281 | |
| | Thursday | 242 | |
| | Friday | 229 | |
| 2 | Monday | 258 | |
| | Tuesday | 270 | |
| | Wednesday | 289 | |
| | Thursday | 251 | |
| | Friday | 237 | |

*n*-point moving average values are to be calculated.

**(i)** **(a)** Explain why a suitable value for *n* is 5.

.....

.....[1]

**(b)** Explain clearly why centring will not be necessary.

.....

.....[2]

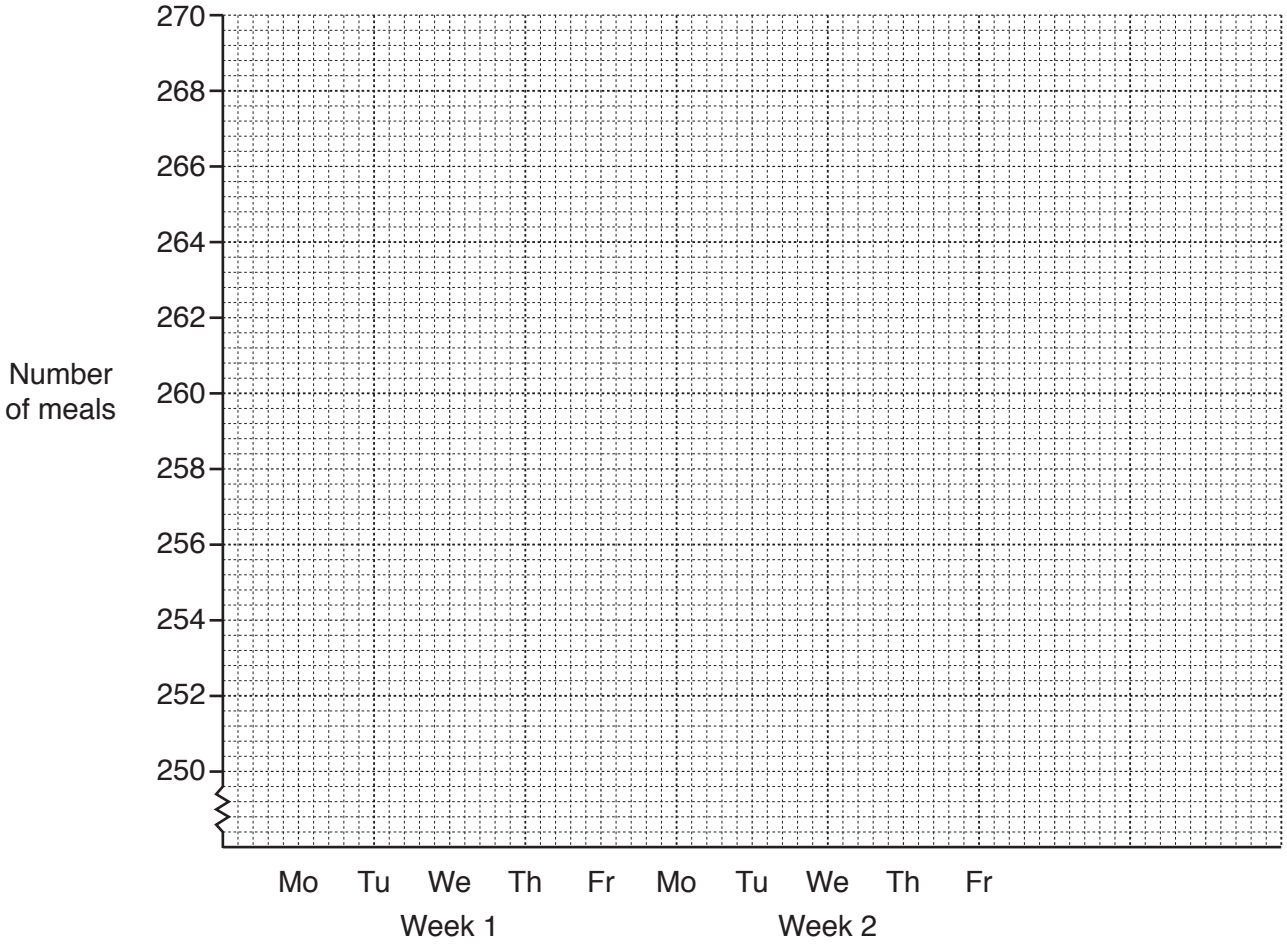
**(c)** Calculate all the 5-point moving average values and insert them in appropriate places in the table above.

[3]

**(ii)** Use values from your table to find an estimate for the seasonal component for Wednesday.

.....[3]

**(iii)** Plot your moving average values on the grid below and draw the trend line.




[3]

**(iv)** Use your trend line and answer to part **(ii)** to estimate the number of meals that will be consumed at the school on the Wednesday of week 3.

.....[2]

**(v)** **(a)** Describe the trend in the number of meals consumed at the school.

.....[1]

**(b)** Explain whether or not you think it is reasonable to assume that this trend will continue over the long term.

.....

.....[1]

**BLANK PAGE**