



UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS
General Certificate of Education Ordinary Level

CANDIDATE
NAME

CENTRE
NUMBER

--	--	--	--	--

CANDIDATE
NUMBER

--	--	--	--



STATISTICS

4040/02

Paper 2

October/November 2009

2 hours 15 minutes

Candidates answer on the question paper.

- Additional Materials:
- Mathematical tables
 - 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 a soft pencil for any diagrams or graphs.
 Do not use staples, paper clips, highlighters, 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.



Section A [36 marks]Answer **all** of the questions 1 to 6.

- 1 The masses of all 79 apples in a batch were measured to the nearest gram. The masses are summarised in the following table.

<i>Mass (to nearest gram)</i>	<i>Number of apples</i>
80–99	2
100–109	13
110–119	20
120–129	22
130–149	17
150–179	5
Total	79

- (i) Explain why the class width of the 100–109 gram class is 10 grams.

.....
[1]

- (ii) Obtain the cumulative frequencies of the data.

.....[1]

- (iii) Without drawing a graph, calculate, to 1 decimal place, an estimate of the upper quartile mass of this batch of apples.

.....g [4]

2 A and B are two outcomes of a particular statistical experiment.

The probabilities of each of these outcomes occurring are as follows.

$$P(A) = 0.55$$

$$P(B) = 0.7$$

(i) Explain why A and B cannot be mutually exclusive; that is, why $P(A \cap B)$ cannot equal 0.

.....
..... [1]

(ii) If $P(A \cap B) = 0.4$, calculate the value of $P(A \cup B)$.

.....[2]

(iii) If the experiment is carried out once, calculate the probability that

(a) neither A nor B occurs,

.....[1]

(b) either A or B but not both occurs.

.....[2]

3 On one day an insurance company received 42 claims for storm damage. There were only sufficient staff available to investigate six of these claims. In order to select the six claims to be investigated, they were randomly numbered 00 to 41, then various methods of sampling were suggested for selecting the six, using a table of two-digit random numbers.

In selecting the random numbers, all repeats are ignored.

The suggested methods are

- A Select six random numbers, ignoring any greater than 41.
- B Select six random numbers. Divide each one by 42, and choose the claims whose numbers correspond to the remainders, (for example, if the selected random number is 45, claim 03 would be chosen.)
- C As in B, but ignore random numbers of 84 and over.
- D Select a number at random from the range 00 to 06. Choose the claim corresponding to that number, and every seventh number thereafter, (for example, if 05 is selected, choose the claims numbered 05 12 19 26 33 and 40).
- E Do not use random numbers, just select the six largest claims.

(i) Give the name of the method of sampling described in A, and the name of that described in D.

Method A

Method D [2]

(ii) State which of the five suggested **methods** of sampling are unbiased, and which are biased.

Unbiased

Biased [2]

(iii) For each of the methods which is biased, give a reason why it is biased.

.....

.....

.....

..... [2]

- 4 Twelve O-level Statistics students attended a revision course shortly before taking their first test. They all took a test before the start of the course, and another at the end of the course. The improvement during the course was measured as the difference between their two test marks (end mark minus start mark). A negative value means that the end mark was lower than the start mark. The differences were as follows.

13 15 6 -23 9 0 7 3 8 11 -1 14

For each of the following state whether it is true or false.

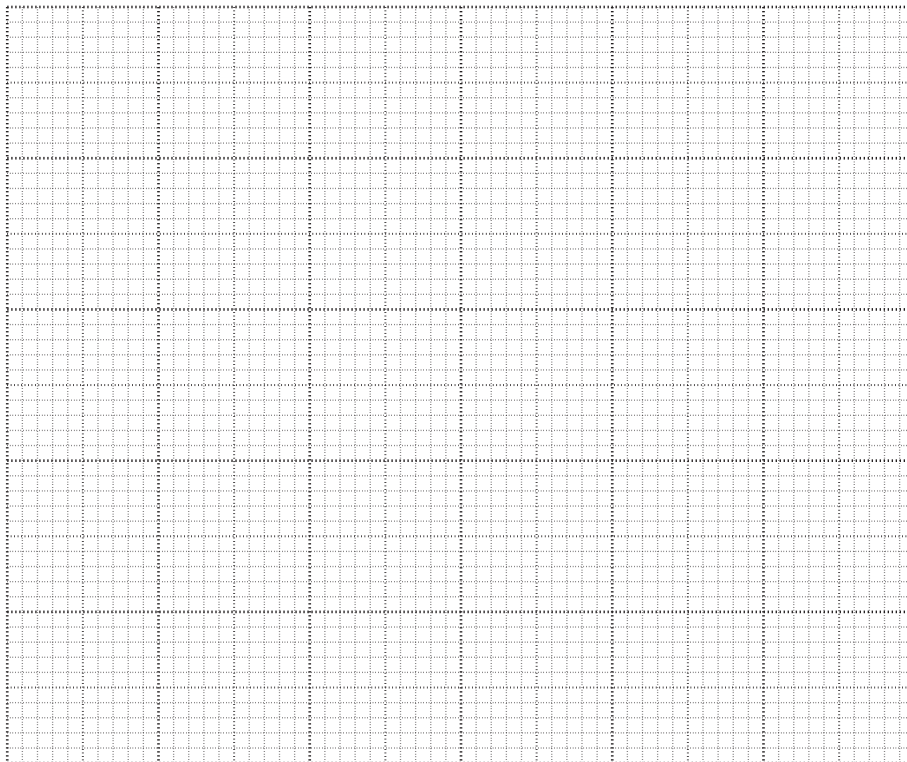
- (i) The median difference is 3.5.
[1]
- (ii) The modal difference is -23.
[1]
- (iii) The standard deviation of these differences cannot be calculated because some of the values are negative.
[1]
- (iv) The range is not an appropriate measure of dispersion for the data because of the presence of one extreme value.
[1]
- (v) The mean of these values is calculated by summing them and then dividing the total by 12.
[1]
- (vi) The range of the values is 1.
[1]

- 5 In a city there are three hospitals, and any person requiring treatment at an Accident and Emergency department is equally likely to attend any of the three.

The number of people attending each hospital, in thousands, to the nearest thousand, in each of the years 2007 and 2008 is given in the following table.

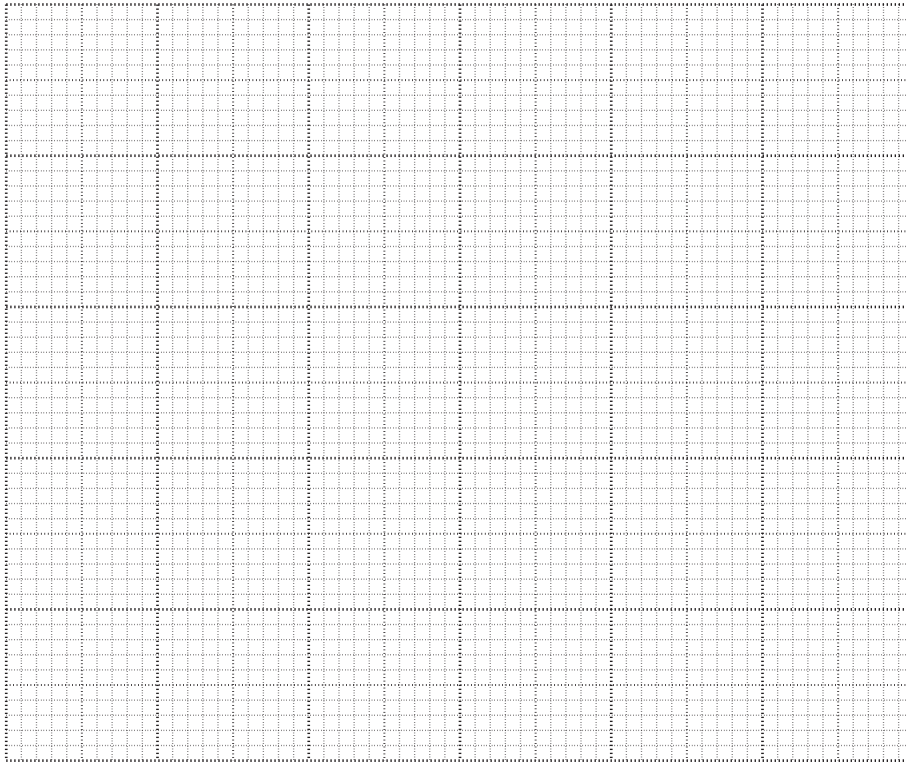
Year	Hospital		
	A	B	C
2007	10	14	16
2008	12	18	10

- (i) Draw, on the grid below, a dual bar chart for each hospital to illustrate the data.



[2]

(ii) Draw, on the grid below, a sectional bar chart for each year to illustrate the data.



[2]

(iii) For each chart state one feature of the data which it illustrates.

.....

.....

.....

.....

[2]

- 6 In a series of ten experiments in which the speed of light, in thousands of km per second, was measured, the following results were obtained.

299.3 299.6 299.9 299.4 299.1 299.3 300.1 300.6 299.6 299.3

- (i) Values of a variable, X , were obtained by applying an assumed mean of 299 to the results. Calculate, correct to 2 decimal places, the mean and the standard deviation of X .

Mean

Standard Deviation[4]

- (ii) **Use your answers to (i)** to state the mean and the standard deviation of the experimental results, (in units of thousands of km per second).

Mean

Standard Deviation[2]



Section B starts on page 10

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) A school wished to estimate its examination costs per candidate for the coming year using the previous year's fees. For each candidate there was a basic fee of \$20 plus an additional fee of \$30 for each subject taken by that candidate. Using these figures, the school estimated that the cost per candidate would have a mean of \$230 and a standard deviation of \$90. However, before the entries were made, the fees had been increased to \$25 per candidate and \$35 per subject.

Calculate the actual mean and standard deviation of the cost per candidate which the school will have to pay for its examination entries in the coming year.

Mean \$

Standard deviation \$ [7]

- (b) An examination consists of three papers, I, II and III, all of which are marked out of a maximum marks. The marks are then weighted, and a candidate's final mark is the sum of his/her three weighted marks. The following table gives the raw and weighted maximum marks, together with the raw marks achieved by one candidate, Victor.

<i>Paper</i>	<i>Raw maximum</i>	<i>Standardised maximum</i>	<i>Victor's marks</i>
I	25	50	16
II	50	50	24
III	80	100	48

- (i) Calculate Victor's final mark out of a total of 200.

.....[3]

An alternative scheme has been proposed for obtaining the final marks. This involves adjusting the mean and standard deviation for each paper to a new scaled mean and standard deviation.

The raw and scaled values are given in the following table.

<i>Paper</i>	<i>Raw mean</i>	<i>Raw standard deviation</i>	<i>Scaled mean</i>	<i>Scaled standard deviation</i>
I	12	4	25	10
II	20	8	25	10
III	48	14	50	20

- (ii) Calculate Victor's final mark if this alternative scheme is used.



(ii) State the seven possible sequences of outcomes when a player plays the game

(iii) Calculate the probability of each of the sequences you have listed.

[2]

(iv) State the amount received for each of the sequences you have listed.

[2]

(v) Find the probability that a player will make neither a profit nor a loss in one game.

.....[2]

(vi) Find, to the nearest cent, a player's expected profit or loss when he plays the game once.

.....[4]



9 In order to calculate a weighted aggregate cost of motoring index, a man divided his costs into three categories: Insurance and Tax, Maintenance, and Fuel. He collected the following data for his motoring in the year 2004.

- Insurance and Tax cost \$440.
- Maintenance cost \$55 per 1000 km travelled.
- Fuel cost \$0.90 per litre.
- Total distance travelled was 19000 km.
- Fuel consumption was 7.8 litres per 100 km.

(i) Calculate the total cost of each of the categories in the year 2004, giving each result to the nearest \$50.

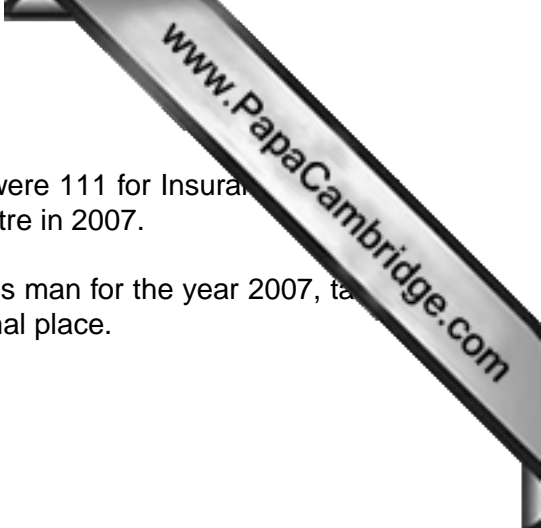
Insurance and Tax \$

Maintenance \$

Fuel \$[6]

(ii) Use your results in (i) to estimate suitable weights for the three categories, expressing them in their lowest terms.

.....[2]



The price relatives for the year 2007, taking 2004 as base year, were 111 for Insurance and 108 for Maintenance. The cost of fuel had risen to \$1.08 per litre in 2007.

- (iii) Calculate a weighted aggregate cost of motoring index for this man for the year 2007, taking 2004 as base year, and giving your answer correct to 1 decimal place.

.....[4]

- (iv) The man's total expenditure on motoring in 2004, to the nearest \$10, was \$2820. Estimate, to the nearest \$10, his expenditure for 2007.

\$[2]

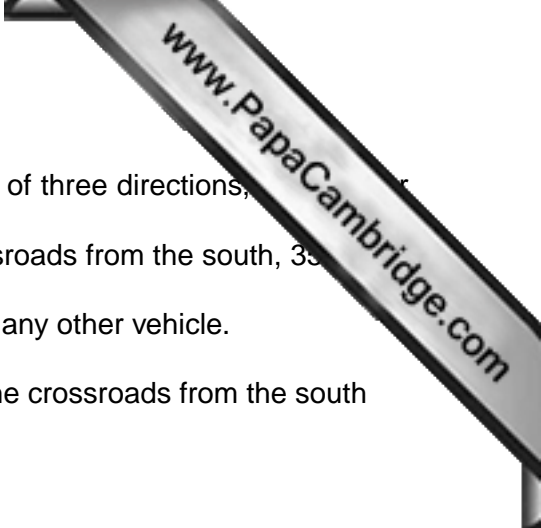
- (v) Give **two** reasons why the estimate in (iv) may be very inaccurate.

Reason 1

.....

Reason 2

.....[2]



10 Vehicles approaching a crossroads from the south may go in one of three directions, straight ahead.

A traffic census has shown that, of vehicles approaching the crossroads from the south, 35% go straight ahead, 35% turn left, 45% turn right and 20% go straight ahead.

The direction taken by any vehicle is independent of that taken by any other vehicle.

Calculate the probability that for any three vehicles approaching the crossroads from the south

(i) all go straight ahead,

.....[2]

(ii) all go in the same direction,

.....[3]

(iii) two turn left and one turns right,

.....[3]



(iv) all go in different directions,

.....[3]

(v) exactly two turn left.

.....[3]

Three particular vehicles all went in the same direction.

(vi) Calculate the probability that they all went straight ahead.

.....[2]

- 11 The table below shows the quarterly sales of petrol, measured in thousands of litres at a petrol station over a three-year period, together with appropriate totals and values of a four-point moving average.

Year	Quarter	Sales (thousand litres)	Four-quarter totals	Centred totals	Centred moving average values
2005	I	57			
	II	64			
			248		
	III	68		494	61.75
			w		
	IV	59		491	61.375
			245		
2006	I	55		490	61.25
			245		
	II	63		487	60.875
			242		
	III	68		x	60.375
			241		
	IV	56		481	60.125
			240		
2007	I	54		479	59.875
			239		
	II	62		475	y
			236		
	III	67			
			z		
	IV	53			

- (i) Explain why it is necessary to centre the moving average values in this table.

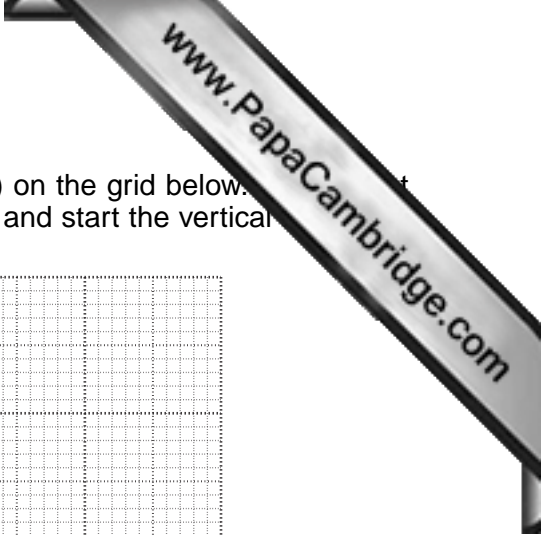
.....
[1]

- (ii) Calculate the values of w, x and y in the table.

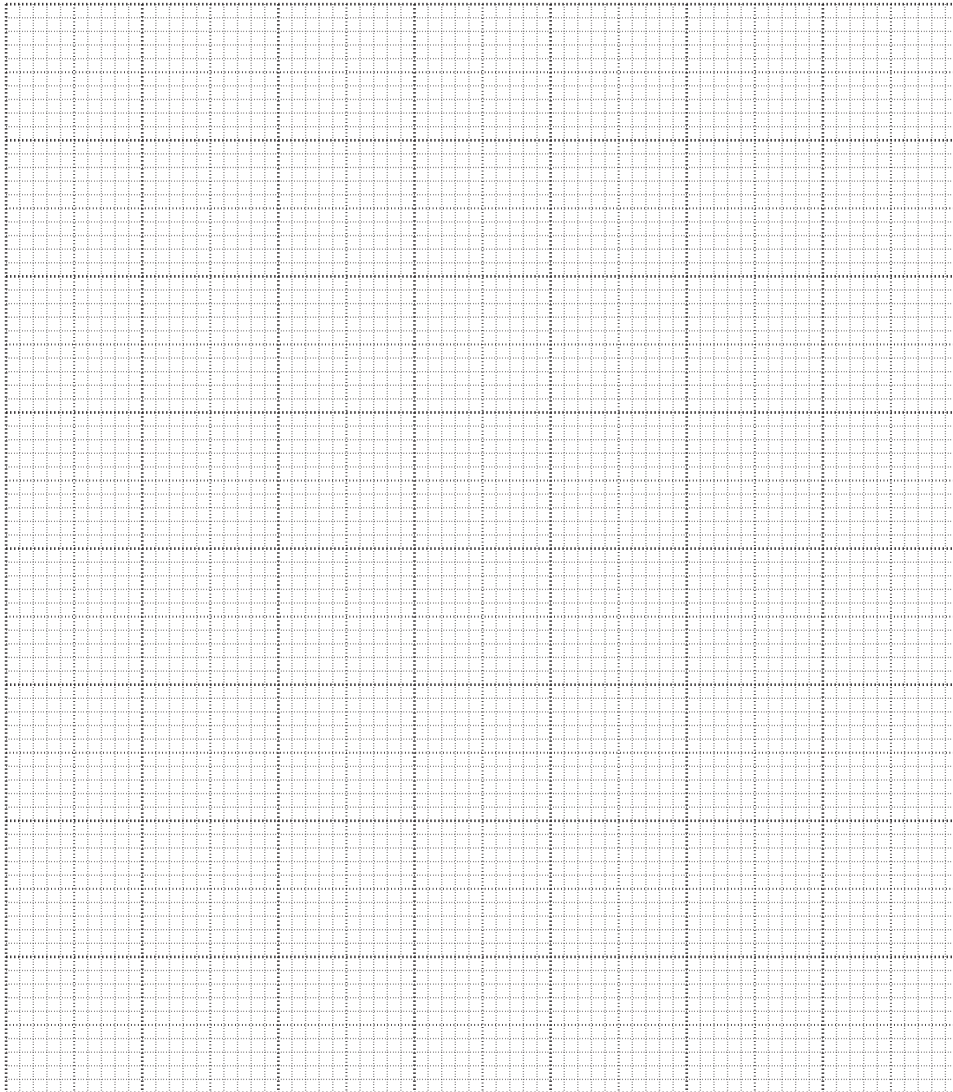
w = x = y = [3]

- (iii) Explain why it is not possible to calculate the value of z in the table.

.....



- (iv) Plot the values of the moving average (**not** the original data) on the grid below. The horizontal axis extends to cover the first quarter of 2008, and start the vertical axis (thousand litres).



[5]

- (v) Draw a single straight line on your graph to represent the trend, and comment on what it tells you about sales of petrol at this filling station.

.....
 [2]

The quarterly components for the data are summarised in the following table.

Quarter	I	II	III	IV
Quarterly component	-5.9	2.1	q	-3.1

- (vi) Calculate the value of q .
 [2]

- (vii) Use your trend line and the appropriate quarterly component to estimate the sales, in thousands of litres, in the first quarter of 2008.

