RECOGNISING ACHIEVEMENT

ADVANCED GCE UNIT   **4767/01**

MATHEMATICS (MEI)

Statistics 2

FRIDAY 12 JANUARY 2007   Morning

Time: 1 hour 30 minutes

Additional Materials:
            Answer booklet (8 pages)
            Graph paper
            MEI Examination Formulae and Tables (MF2)

INSTRUCTIONS TO CANDIDATES

- Write your name, centre number and candidate number in the spaces provided on the answer booklet.
- Answer **all** the questions.
- You are permitted to use a graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

- 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 72.

ADVICE TO CANDIDATES

- Read each question carefully and make sure you know what you have to do before starting your answer.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.

This document consists of **6** printed pages and **2** blank pages.

                  OCR is an exempt Charity   **[Turn over**

1   In a science investigation into energy conservation in the home, a student is collecting data on the time taken for an electric kettle to boil as the volume of water in the kettle is varied. The student's data are shown in the table below, where v litres is the volume of water in the kettle and t seconds is the time taken for the kettle to boil (starting with the water at room temperature in each case). Also shown are summary statistics and a scatter diagram on which the regression line of t on v is drawn.

v	0.2	0.4	0.6	0.8	1.0
t	44	78	114	156	172

$n = 5, \ \Sigma v = 3.0, \ \Sigma t = 564, \ \Sigma v^2 = 2.20, \ \Sigma vt = 405.2.$
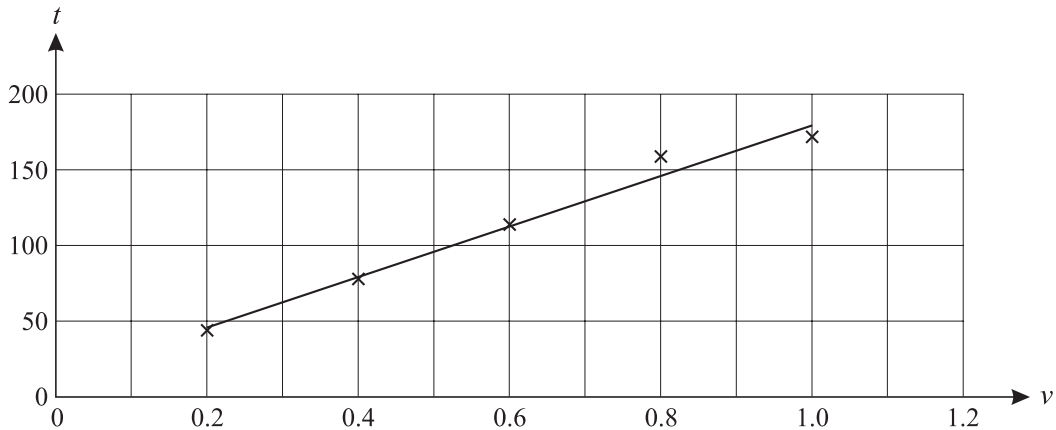
(i)   Calculate the equation of the regression line of t on v, giving your answer in the form $t = a + bv$. [5]

(ii)   Use this equation to predict the time taken for the kettle to boil when the amount of water which it contains is

(A)   0.5 litres,

(B)   1.5 litres.

Comment on the reliability of each of these predictions. [4]

(iii)   In the equation of the regression line found in part (i), explain the role of the coefficient of v in the relationship between time taken and volume of water. [2]

(iv)   Calculate the values of the residuals for $v = 0.8$ and $v = 1.0$. [4]

(v)   Explain how, on a scatter diagram with the regression line drawn accurately on it, a residual could be measured and its sign determined. [3]

2 **(a)** A farmer grows Brussels sprouts. The diameter of sprouts in a particular batch, measured in mm, is Normally distributed with mean 28 and variance 16. Sprouts that are between 24 mm and 33 mm in diameter are sold to a supermarket.

    (i) Find the probability that the diameter of a randomly selected sprout will be within this range. [4]

    (ii) The farmer sells the sprouts in this range to the supermarket for 10 pence per kilogram. The farmer sells sprouts under 24 mm in diameter to a frozen food factory for 5 pence per kilogram. Sprouts over 33 mm in diameter are thrown away. Estimate the total income received by the farmer for the batch, which weighs 25 000 kg. [3]

    (iii) By harvesting sprouts earlier, the mean diameter for another batch can be reduced to k mm. Find the value of k for which only 5% of the sprouts will be above 33 mm in diameter. You may assume that the variance is still 16. [3]

(b) The farmer also grows onions. The weight in kilograms of the onions is Normally distributed with mean 0.155 and variance 0.005. He is trying out a new variety, which he hopes will yield a higher mean weight. In order to test this, he takes a random sample of 25 onions of the new variety and finds that their total weight is 4.77 kg. You should assume that the weight in kilograms of the new variety is Normally distributed with variance 0.005.

    (i) Write down suitable null and alternative hypotheses for the test in terms of μ. State the meaning of μ in this case. [2]

    (ii) Carry out the test at the 1% level. [6]

3 An electrical retailer gives customers extended guarantees on washing machines. Under this guarantee all repairs in the first 3 years are free. The retailer records the numbers of free repairs made to 80 machines.

Number of repairs	0	1	2	3	>3
Frequency	53	20	6	1	0

   (i) Show that the sample mean is 0.4375. [1]

  (ii) The sample standard deviation s is 0.6907. Explain why this supports a suggestion that a Poisson distribution may be a suitable model for the distribution of the number of free repairs required by a randomly chosen washing machine. [2]

The random variable X denotes the number of free repairs required by a randomly chosen washing machine. For the remainder of this question you should assume that X may be modelled by a Poisson distribution with mean 0.4375.

 (iii) Find $P(X = 1)$. Comment on your answer in relation to the data in the table. [4]

 (iv) The manager decides to monitor 8 washing machines sold on one day. Find the probability that there are at least 12 free repairs in total on these 8 machines. You may assume that the 8 machines form an independent random sample. [3]

  (v) A launderette with 8 washing machines has needed 12 free repairs. Why does your answer to part **(iv)** suggest that the Poisson model with mean 0.4375 is unlikely to be a suitable model for free repairs on the machines in the launderette? Give a reason why the model may not be appropriate for the launderette. [3]

The retailer also sells tumble driers with the same guarantee. The number of free repairs on a tumble drier in three years can be modelled by a Poisson distribution with mean 0.15. A customer buys a tumble drier and a washing machine.

 (vi) Assuming that free repairs are required independently, find the probability that

     (*A*) the two appliances need a total of 3 free repairs between them,

     (*B*) each appliance needs exactly one free repair. [5]

4 Two educational researchers are investigating the relationship between personal ambitions and home location of students. The researchers classify students into those whose main personal ambition is good academic results and those who have some other ambition. A random sample of 480 students is selected.

(i) One researcher summarises the data as follows.

Observed		Home location	
		City	Non-city
Ambition	Good results	102	147
	Other	75	156

Carry out a test at the 5% significance level to examine whether there is any association between home location and ambition. State carefully your null and alternative hypotheses. Your working should include a table showing the contributions of each cell to the test statistic. [9]

(ii) The other researcher summarises the same data in a different way as follows.

Observed		Home location		
		City	Town	Country
Ambition	Good results	102	83	64
	Other	75	64	92

(*A*) Calculate the expected frequencies for both 'Country' cells. [2]

(*B*) The test statistic for these data is 10.94. Carry out a test at the 5% level based on this table, using the same hypotheses as in part **(i)**. [3]

(*C*) The table below gives the contribution of each cell to the test statistic. Discuss briefly how personal ambitions are related to home location. [2]

Contribution to the test statistic		Home location		
		City	Town	Country
Ambition	Good results	1.129	0.596	3.540
	Other	1.217	0.643	3.816

(iii) Comment briefly on whether the analysis in part **(ii)** means that the conclusion in part **(i)** is invalid. [2]

BLANK PAGE

BLANK PAGE