OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MEI STRUCTURED MATHEMATICS   **4767**

Statistics 2

Wednesday   **25 JANUARY 2006**   Morning   1 hour 30 minutes

Additional materials:
 8 page answer booklet
 Graph paper
 MEI Examination Formulae and Tables (MF2)

TIME 1 hour 30 minutes

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.

- 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.

- The total number of marks for this paper is 72.

This question paper consists of 5 printed pages and 3 blank pages.

2

1   A roller-coaster ride has a safety system to detect faults on the track.

   (i) State conditions for a Poisson distribution to be a suitable model for the number of faults occurring on a randomly selected day. [2]

Faults are detected at an average rate of 0.15 per day. You may assume that a Poisson distribution is a suitable model.

   (ii) Find the probability that on a randomly chosen day there are

   (*A*)   no faults,

   (*B*)   at least 2 faults. [4]

   (iii) Find the probability that, in a randomly chosen period of 30 days, there are at most 3 faults. [3]

There is also a separate safety system to detect faults on the roller-coaster train itself. Faults are detected by this system at an average rate of 0.05 per day, independently of the faults detected on the track. You may assume that a Poisson distribution is also suitable for modelling the number of faults detected on the train.

   (iv) State the distribution of the total number of faults detected by the two systems in a period of 10 days. Find the probability that a total of 5 faults is detected in a period of 10 days. [4]

   (v) The roller-coaster is operational for 200 days each year. Use a suitable approximating distribution to find the probability that a total of at least 50 faults is detected in 200 days. [5]

3

2 The drug EPO (erythropoetin) is taken by some athletes to improve their performance. This drug is in fact banned and blood samples taken from athletes are tested to measure their 'hematocrit level'. If the level is over 50 it is considered that the athlete is likely to have taken EPO and the result is described as 'positive'. The measured hematocrit level of each athlete varies over time, even if EPO has not been taken.

　(i) For each athlete in a large population of innocent athletes, the variation in measured hematocrit level is described by the Normal distribution with mean 42.0 and standard deviation 3.0.

　　(*A*) Show that the probability that such an athlete tests positive for EPO in a randomly chosen test is 0.0038. [3]

　　(*B*) Find the probability that such an athlete tests positive on at least 1 of the 7 occasions during the year when hematocrit level is measured. (These occasions are spread at random through the year and all test results are assumed to be independent.) [3]

　　(*C*) It is standard policy to apply a penalty after testing positive. Comment briefly on this policy in the light of your answer to part **(i)(*B*)**. [2]

　(ii) Suppose that 1000 tests are carried out on innocent athletes whose variation in measured hematocrit level is as described in part **(i)**. It may be assumed that the probability of a positive result in each test is 0.0038, independently of all other test results.

　　(*A*) State the exact distribution of the number of positive tests. [2]

　　(*B*) Use a suitable approximating distribution to find the probability that at least 10 tests are positive. [4]

　(iii) Because of genetic factors, a particular innocent athlete has an abnormally high natural hematocrit level. This athlete's measured level is Normally distributed with mean 48.0 and standard deviation 2.0. The usual limit of 50 for a positive test is to be altered for this athlete to a higher value h. Find the value of h for which this athlete would test positive on average just once in 200 occasions. [4]

[Turn over

3 A researcher is investigating the relationship between temperature and levels of the air pollutant nitrous oxide at a particular site. The researcher believes that there will be a positive correlation between the daily maximum temperature, x, and nitrous oxide level, y. Data are collected for 10 randomly selected days. The data, measured in suitable units, are given in the table and illustrated on the scatter diagram.

x	13.3	17.2	16.9	18.7	18.4	19.3	23.1	15.0	20.6	14.4
y	9	11	14	26	43	25	52	15	10	7
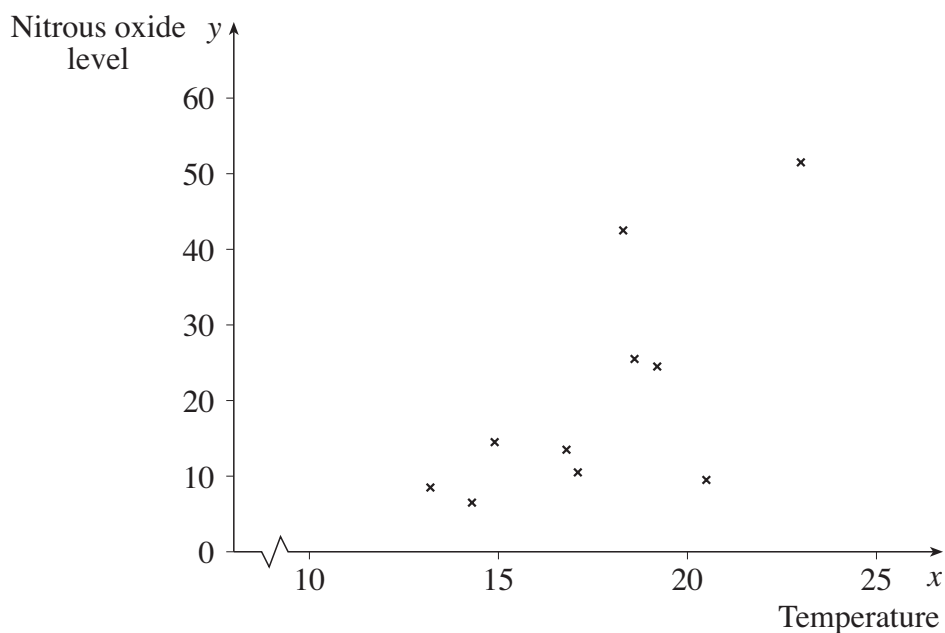
(i) Calculate the value of Spearman's rank correlation coefficient for these data. [5]

(ii) Perform a hypothesis test at the 5% level to check the researcher's belief, stating your hypotheses clearly. [5]

(iii) It is suggested that it would be preferable to carry out a test based on the product moment correlation coefficient. State the distributional assumption required for such a test to be valid. Explain how a scatter diagram can be used to check whether the distributional assumption is likely to be valid and comment on the validity in this case. [3]

(iv) A statistician investigates data over a much longer period and finds that the assumptions for the use of the product moment correlation coefficient are in fact valid. Give the critical region for the test at the 1% level, based on a sample of 60 days. [2]

(v) In a different research project, into the correlation between daily temperature and ozone pollution levels, a positive correlation is found. It is argued that this shows that high temperatures cause increased ozone levels. Comment on this claim. [3]

4 The table summarises the usual method of travelling to school for 200 randomly selected pupils from primary and secondary schools in a city.

		Primary	Secondary
Method of travel	Bus	21	49
	Car	65	15
	Cycle or Walk	34	16

   (i) Write down null and alternative hypotheses for a test to examine whether there is any association between method of travel and type of school. [1]

  (ii) Calculate the expected frequency for primary school bus users. Calculate also the corresponding contribution to the test statistic for the usual χ^2 test. [4]

 (iii) Given that the value of the test statistic for the usual χ^2 test is 42.64, carry out the test at the 5% level of significance, stating your conclusion clearly. [4]

The mean travel time for pupils who travel by bus is known to be 18.3 minutes. A survey is carried out to determine whether the mean travel time to school by car is different from 18.3 minutes. In the survey, 20 pupils who travel by car are selected at random. Their mean travel time is found to be 22.4 minutes.

 (iv) Assuming that car travel times are Normally distributed with standard deviation 8.0 minutes, carry out a test at the 10% level, stating your hypotheses and conclusion clearly. [7]

  (v) Comment on the suggestion that pupils should use a bus if they want to get to school quickly. [2]