OCR
RECOGNISING ACHIEVEMENT

Thursday 31 May 2012 – Morning

A2 GCE MATHEMATICS (MEI)

4767 Statistics 2

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4767
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the Printed Answer Book and the Question Paper.

- The Question Paper will be found in the centre of the Printed Answer Book.
- Write your name, centre number and candidate number in the spaces provided on the Printed Answer Book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the Printed Answer Book**. Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

This information is the same on the Printed Answer Book and the Question Paper.

- The number of marks is given in brackets **[ ]** at the end of each question or part question on the Question Paper.
- 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**.
- The Printed Answer Book consists of **12** pages. The Question Paper consists of **4** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER / INVIGILATOR

- Do not send this Question Paper for marking; it should be retained in the centre or recycled. Please contact OCR Copyright should you wish to re-use this document.

1 The times, in seconds, taken by ten randomly selected competitors for the first and last sections of an Olympic bobsleigh run are denoted by x and y respectively. Summary statistics for these data are as follows.

$$\Sigma x = 113.69 \quad \Sigma y = 52.81 \quad \Sigma x^2 = 1292.56 \quad \Sigma y^2 = 278.91 \quad \Sigma xy = 600.41 \quad n = 10$$

(i) Calculate the sample product moment correlation coefficient. **[5]**

(ii) Carry out a hypothesis test at the 10% significance level to investigate whether there is any correlation between times taken for the first and last sections of the bobsleigh run. **[6]**

(iii) State the distributional assumption which is necessary for this test to be valid. Explain briefly how a scatter diagram may be used to check whether this assumption is likely to be valid. **[2]**

(iv) A commentator says that in order to have a fast time on the last section, you must have a fast time on the first section. Comment briefly on this suggestion. **[2]**

(v) (*A*) Would your conclusion in part **(ii)** have been different if you had carried out the hypothesis test at the 1% level rather than the 10% level? Explain your answer. **[2]**

(*B*) State one advantage and one disadvantage of using a 1% significance level rather than a 10% significance level in a hypothesis test. **[2]**

2 A particular genetic mutation occurs in one in every 300 births on average. A random sample of 1200 births is selected.

(i) State the exact distribution of X, the number of births in the sample which have the mutation. **[2]**

(ii) Explain why X has, approximately, a Poisson distribution. **[2]**

(iii) Use a Poisson approximating distribution to find

(*A*) $P(X = 1)$,

(*B*) $P(X > 4)$. **[5]**

(iv) Twenty independent samples, each of 1200 births, are selected. State the mean and variance of a Normal approximating distribution suitable for modelling the total number of births with the mutation in the twenty samples. **[2]**

(v) Use this Normal approximating distribution to

(*A*) find the probability that there are at least 90 births which have the mutation, **[3]**

(*B*) find the least value of k such that the probability that there are at most k births with this mutation is greater than 5%. **[4]**

3 At a vineyard, the process used to fill bottles with wine is subject to variation. The contents of bottles are independently Normally distributed with mean $\mu = 751.4$ ml and standard deviation $\sigma = 2.5$ ml.

(i) Find the probability that a randomly selected bottle contains at least 750 ml. **[3]**

(ii) A case of wine consists of 6 bottles. Find the probability that all 6 bottles in a case contain at least 750 ml. **[2]**

(iii) Find the probability that, in a random sample of 25 cases, there are at least 2 cases in which all 6 bottles contain at least 750 ml. **[4]**

It is decided to increase the proportion of bottles which contain at least 750 ml to 98%.

(iv) This can be done by changing the value of μ, but retaining the original value of σ. Find the required value of μ. **[4]**

(v) An alternative is to change the value of σ, but retain the original value of μ. Find the required value of σ. **[3]**

(vi) Comment briefly on which method might be easier to implement and which might be preferable to the vineyard owners. **[2]**

[Question 4 is printed overleaf.]

4 **(a)** Mary is opening a cake shop. As part of her market research, she carries out a survey into which type of cake people like best. She offers people 4 types of cake to taste: chocolate, carrot, lemon and ginger. She selects a random sample of 150 people and she classifies the people as children and adults. The results are as follows.

		Classification of person		Row totals
		Child	Adult	
Type of cake	Chocolate	34	23	57
	Carrot	16	18	34
	Lemon	4	18	22
	Ginger	13	24	37
Column totals		67	83	150

The contributions to the test statistic for the usual χ^2 test are shown in the table below.

		Classification of person	
		Child	Adult
Type of cake	Chocolate	2.8646	2.3124
	Carrot	0.0436	0.0352
	Lemon	3.4549	2.7889
	Ginger	0.7526	0.6075

The sum of these contributions, correct to 2 decimal places, is 12.86.

(i) Calculate the expected frequency for children preferring chocolate cake. Verify the corresponding contribution, 2.8646, to the test statistic. **[3]**

(ii) Carry out the test at the 1% level of significance. **[5]**

(b) Mary buys flour in bags which are labelled as containing 5 kg. She suspects that the average contents of these bags may be less than 5 kg. In order to test this, she selects a random sample of 8 bags and weighs their contents. Assuming that weights are Normally distributed with standard deviation 0.0072 kg, carry out a test at the 5% level, given that the weights of the 8 bags in kg are as follows.

4.992 4.981 5.006 4.982 4.996 5.009 4.991 5.003 **[9]**