OCR
RECOGNISING ACHIEVEMENT

Friday 22 June 2012 – Afternoon

A2 GCE MATHEMATICS (MEI)

4768 Statistics 3

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4768
- 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.
- Answer **all** the questions.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- 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 Technologists at a company that manufactures paint are trying to develop a new type of gloss paint with a shorter drying time than the current product. In order to test whether the drying time has been reduced, the technologists paint a square metre of each of the new and old paints on each of 10 different surfaces. The lengths of time, in hours, that each square metre takes to dry are as follows.

Surface	A	B	C	D	E	F	G	H	I	J
Old paint	16.6	17.0	16.5	15.6	16.3	16.5	16.4	15.9	16.3	16.1
New paint	15.9	16.3	16.3	15.9	15.5	16.6	16.1	16.0	16.2	15.6

(i) Explain why a paired sample is used in this context. **[1]**

(ii) The mean reduction in drying time is to be investigated. Why might a t test be appropriate in this context and what assumption needs to be made? **[4]**

(iii) Using a significance level of 5%, carry out a test to see if there appears to be any reduction in mean drying time. **[9]**

(iv) Find a 95% confidence interval for the true mean reduction in drying time. **[4]**

2 **(a)** **(i)** Give two reasons why an investigator might need to take a sample in order to obtain information about a population. **[2]**

(ii) State two requirements of a sample. **[2]**

(iii) Discuss briefly the advantage of the sampling being random. **[2]**

(b) **(i)** Under what circumstances might one use a Wilcoxon single sample test in order to test a hypothesis about the median of a population? What distributional assumption is needed for the test? **[2]**

(ii) On a stretch of road leading out of the centre of a town, highways officials have been monitoring the speed of the traffic in case it has increased. Previously it was known that the median speed on this stretch was 28.7 miles per hour. For a random sample of 12 vehicles on the stretch, the following speeds were recorded.

32.0 29.1 26.1 35.2 34.4 28.6 32.3 28.5 27.0 33.3 28.2 31.9

Carry out a test, with a 5% significance level, to see whether the speed of the traffic on this stretch of road seems to have increased on the whole. **[10]**

3 The triathlon is a sports event in which competitors take part in three stages, swimming, cycling and running, one straight after the other. The winner is the competitor with the shortest overall time. In this question the times for the separate stages are assumed to be Normally distributed and independent of each other.

For a particular triathlon event in which there was a very large number of competitors, the mean and standard deviation of the times, measured in minutes, for each stage were as follows.

	Mean	Standard deviation
Swimming	11.07	2.36
Cycling	57.33	8.76
Running	24.23	3.75

(i) For a randomly chosen competitor, find the probability that the swimming time is between 10 and 13 minutes. **[3]**

(ii) For a randomly chosen competitor, find the probability that the running time exceeds the swimming time by more than 10 minutes. **[4]**

(iii) For a randomly chosen competitor, find the probability that the swimming and running times combined exceed $\frac{2}{3}$ of the cycling time. **[4]**

(iv) In a different triathlon event the total times, in minutes, for a random sample of 12 competitors were as follows.

103.59   99.04   85.03   81.34   106.79   89.14   98.55   98.22   108.87   116.29   102.51   92.44

Find a 95% confidence interval for the mean time of all competitors in this event. **[5]**

(v) Discuss briefly whether the assumptions of Normality and independence for the stages of triathlon events are reasonable. **[2]**

[Question 4 is printed overleaf.]

4 The numbers of call-outs per day received by a fire station for a random sample of 255 weekdays were recorded as follows.

Number of call-outs	0	1	2	3	4	5 or more
Frequency (days)	145	79	22	6	3	0

The mean number of call-outs per day for these data is 0.6. A Poisson model, using this sample mean of 0.6, is fitted to the data, and gives the following expected frequencies (correct to 3 decimal places).

Number of call-outs	0	1	2	3	4	5 or more
Expected frequency	139.947	83.968	25.190	5.038	0.756	0.101

(i) Using a 5% significance level, carry out a test to examine the goodness of fit of the model to the data. **[9]**

The time T, measured in days, that elapses between successive call-outs can be modelled using the exponential distribution for which f(t), the probability density function, is

$$f(t) = \begin{cases} 0 & t < 0, \\ \lambda e^{-\lambda t} & t \geqslant 0, \end{cases}$$

where λ is a positive constant.

(ii) For the distribution above, it can be shown that $E(T) = \frac{1}{\lambda}$. Given that the mean time between successive call-outs is $\frac{5}{3}$ days, write down the value of λ. **[1]**

(iii) Find F(t), the cumulative distribution function. **[3]**

(iv) Find the probability that the time between successive call-outs is more than 1 day. **[2]**

(v) Find the median time that elapses between successive call-outs. **[3]**