OCR
RECOGNISING ACHIEVEMENT

Friday 1 June 2012 – Morning

AS GCE MEI STATISTICS

G243 Statistics 3 (Z3)

QUESTION PAPER

G21601061 2

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book G243
- 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**.
- This 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.

Section A (48 marks)

1 A delivery company has a fleet of 120 lorries. The company manager wishes to switch from conventional diesel fuel to a blend of biodiesel fuel. Before switching, he decides to check whether using biodiesel will affect the fuel consumption of the lorries. He selects 8 lorries and checks their fuel consumption using conventional diesel and then again using biodiesel.

(i) Describe how the manager could select this sample by systematic sampling. **[3]**

The results, measured in litres per 100 km, are as follows.

Lorry	A	B	C	D	E	F	G	H
Conventional diesel	39.2	34.4	28.6	25.0	27.8	31.9	33.6	38.7
Biodiesel	38.7	34.0	29.2	25.3	27.8	32.6	33.9	39.3

(ii) Use a *t* test to examine, at the 5% significance level, whether it appears that the mean fuel consumption for lorries using biodiesel is the same as that for lorries using conventional diesel. **[11]**

(iii) State the distributional assumption which is necessary for this test to be valid. Name an alternative test which could be performed if this assumption is not valid. **[2]**

2 A researcher is checking the breaking stress of titanium components from two different manufacturers. These components are costly and so the researcher can only test a small sample from each manufacturer. The researcher wishes to examine whether, on the whole, the breaking stresses of components from the two manufacturers can be considered to be the same. The breaking stresses of a random sample of 10 components from each manufacturer, measured in suitable units, are as follows.

Manufacturer A	70.6	75.2	77.9	75.4	79.3	77.6	77.0	73.1	73.0	73.5
Manufacturer B	76.9	74.6	72.6	73.7	74.2	70.4	70.1	78.5	75.0	72.7

(i) Explain briefly whether it is appropriate to carry out a Wilcoxon signed rank test in this situation. **[2]**

(ii) The researcher suspects that the populations are not Normally distributed. Carry out a suitable test at the 10% significance level. **[10]**

(iii) Would the outcome of the test be different if you had ranked the data in reverse order? **[1]**

(iv) Name an alternative test which would have been preferable if the researcher had been able to collect data on 100 components from each manufacturer. Discuss briefly whether the population variances would need to be known in order to carry out this test. **[3]**

3 An education authority collects data on attendance level, x, and academic performance, y, of children in its schools, both measured in suitable units. Summary statistics for 50 randomly selected children are shown below.

$$\Sigma x = 26.43 \qquad \Sigma y = 265.4 \qquad \Sigma x^2 = 16.62 \qquad \Sigma y^2 = 1576.9 \qquad \Sigma xy = 147.6 \qquad n = 50$$

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

(ii) Carry out a hypothesis test at the 5% significance level to determine whether there appears to be positive correlation between x and y. **[6]**

(iii) What distributional assumption is required for this test? Explain how a scatter diagram may be used to check whether this assumption may be valid. **[2]**

(iv) Subsequently it is decided to calculate the correlation coefficient for the whole population of students of the education authority and it is found to be 0.291. Explain briefly why it is not valid to perform a hypothesis test in this case. **[1]**

(v) Because this correlation coefficient is positive, it is suggested that boosting attendance levels will enhance academic performance. Explain briefly why this may not be the case. **[2]**

[Question 4 is printed overleaf.]

Section B (24 marks)

4 Scientists at a pharmaceutical company are developing a new drug to treat the common cold. They wish to find out whether patients get better more quickly if given this drug rather than being given other treatments. They compare the performance of the new drug with two other treatments which are already on the market, paracetamol and ibuprofen, and also with giving no drug.

They classify the treatments as follows.

Treatment A: new drug
Treatment B: paracetamol
Treatment C: ibuprofen
Treatment D: no drug

(i) Explain why the scientists include treatment D in the investigation. **[2]**

(ii) In order to test whether the new drug is better than the other treatments, it is suggested that 4 people who have symptoms of the common cold are selected. The first person is given treatment A, the second is given treatment B, the third is given treatment C and the fourth is given treatment D. If the person who is given treatment A gets better more quickly than the others, then the new drug would be regarded as being better than the other treatments. Comment critically on this suggestion. **[3]**

(iii) An alternative approach is suggested, in which a large number of volunteers are infected with the same strain of the common cold virus. Each is then given one of the four treatments. The speed of recovery under each treatment is noted. Briefly comment on this suggestion. **[2]**

In fact it is decided to test the treatments on all employees of the pharmaceutical company who are willing to take part in the trial. As soon as any of these employees has symptoms of the common cold, one of the four treatments is allocated to the employee. The time that the employee takes to recover is noted.

(iv) Explain why the treatments should be allocated randomly. **[2]**

(v) Explain why a sample of employees of the company would not be representative of the whole of the population of the UK. **[2]**

(vi) In fact, the employees given treatment D are given 'placebo' tablets which appear identical to the tablets containing the new drug. These placebo tablets are designed to have no medical effect. Explain the purpose of using placebo tablets. **[2]**

The scientists wish to compare treatments A and D. The recovery times, in days, of 43 employees under treatment A and 49 employees under treatment D are available. The means and variances of these recovery times are shown below.

Treatment A:	mean 4.770	variance 1.747
Treatment D:	mean 4.926	variance 1.594

(vii) Carry out a test at the 5% significance level to examine whether employees appear to have shorter recovery times under treatment A than under treatment D. **[11]**