OCR

Thursday 12 June 2014 – Afternoon

AS GCE MEI STATISTICS

G243/01 Statistics 3 (Z3)

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book G243/01
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

12412833391

INSTRUCTIONS TO CANDIDATES

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

- The Question Paper will be found inside 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.

Section A (48 marks)

1 A hospital consultant is checking the waiting times for patients at hospital clinics. The waiting times, in minutes, for random samples of morning and afternoon appointments are as follows.

Morning	6	9	14	15	18	23	24	29	33	37	52	61
Afternoon	2	5	7	12	13	19	20	21	27	47		

The consultant suspects that the waiting times are shorter, on average, at afternoon appointments than at morning appointments.

(i) Name two tests which could be used to analyse these data. Under what circumstances would you choose one test rather than the other? **[4]**

(ii) Given that it is not known whether or not the populations are Normally distributed, carry out a test at the 5% significance level to investigate the consultant's suspicion. **[10]**

(iii) Another consultant suggests that if there had been equal numbers of morning and afternoon appointments, a paired design could have been used. Comment briefly on this suggestion. **[2]**

2 An internet retailer employs workers to pick goods for packing at its warehouse. The management introduces an incentive bonus scheme in order to try to reduce the average time which it takes a picker to pick an item. The picking times in seconds for a random sample of 11 workers, before the introduction of the incentive scheme and one week after the introduction of the scheme, are as follows.

Worker	A	B	C	D	E	F	G	H	I	J	K
Before incentive	33.5	31.7	34.2	36.5	37.1	29.6	30.2	33.0	32.9	28.7	30.3
After incentive	32.6	30.3	33.7	36.8	36.9	30.1	29.6	32.0	31.6	29.2	29.4

A test is to be carried out to investigate whether it appears that the average picking time is reduced after the incentive scheme is introduced.

(i) Briefly explain why it is better to carry out a paired sample test rather than a two-sample test for these data. **[2]**

(ii) Name a paired sample test which can be used if the distribution of the population of differences is unknown. **[1]**

(iii) State any distributional assumptions necessary for the use of a paired sample *t* test. **[2]**

(iv) Use a *t* test to examine, at the 5% significance level, whether it appears that the average picking time is less after the incentive scheme is introduced. **[11]**

3 In order to investigate whether there is positive correlation between rainfall and crop yields, the total rainfall, x mm, and the weights per square metre, y kg, of a particular crop were recorded in a number of fields. These fields were chosen randomly from a large number of fields. The data are shown below.

x	36	50	44	72	44	74	64	50	39	30	61
y	2.2	5.0	6.2	8.4	1.8	7.4	4.3	2.2	7.5	3.6	7.6

(i) Draw a scatter diagram to illustrate these data. **[3]**

(ii) Calculate the value of the product moment correlation coefficient. **[2]**

(iii) State an assumption about the underlying population which is required to carry out a test based on the product moment correlation coefficient. Explain why, in the light of the scatter diagram, it is reasonable to suppose that this assumption may be valid. **[2]**

(iv) Carry out a hypothesis test at the 1% significance level to determine whether there appears to be positive correlation between x and y. **[6]**

(v) Explain why it is important that the 11 fields were chosen randomly. **[2]**

(vi) Name an alternative test which could have been carried out if the assumption in part **(iii)** was not valid. **[1]**

Question 4 begins on page 4.

Section B (24 marks)

4 A researcher is investigating whether the average lengths of fish of a particular species in two African lakes P and Q are equal. Each day fishermen catch large numbers of these fish from each lake. The researcher is able to measure the lengths of any of the fish which have been caught. For each lake, she considers a number of ways of selecting a sample of 50 fish of this species.

A: Randomly select 50 fish which have been caught by a particular fisherman on a particular day.
B: Choose 5 fishermen and randomly select 10 fish caught by each of them on a particular day.
C: Choose 5 fishermen and randomly select 2 fish caught by each of them for a period of 5 days.

(i) Which of these methods is likely to result in the most representative sample from the population of fish in the lake? Briefly explain your answer. **[2]**

(ii) The researcher decides to use Method B. Given that a particular fisherman on one of the lakes catches 120 fish, explain how she can select a simple random sample of size 10 from these 120 fish. **[3]**

(iii) Explain why she should sample the fish from the two lakes on the same day if possible. **[2]**

On a later occasion the researcher selects simple random samples of 50 fish from each lake. The lengths in mm of the 50 fish from Lake P are summarised by $\Sigma x = 7683.5, \Sigma x^2 = 1191300$. For the 50 fish from Lake Q, the sample mean is 151.7 mm and the sample standard deviation is 15.6 mm.

(iv) Calculate the sample mean and sample standard deviation for the fish from Lake P. **[3]**

The researcher wishes to investigate whether there appears to be any difference in the average lengths of the fish from the two lakes.

(v) Explain why it is appropriate to carry out a hypothesis test based on the Normal distribution. **[1]**

(vi) Carry out the hypothesis test at the 10% significance level. **[11]**

(vii) Explain why, even if the result of a hypothesis test is 'reject H_0', the null hypothesis may still be true. **[2]**

END OF QUESTION PAPER