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

1241283391

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

OCR

Thursday 12 June 2014 – Afternoon

AS GCE MEI STATISTICS

G243/01 Statistics 3 (Z3)

PRINTED ANSWER BOOK

Candidates answer on this Printed Answer Book.

OCR supplied materials:
- Question Paper G243/01 (inserted)
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

Candidate forename		Candidate surname	

Centre number						Candidate number				

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.

Section A (48 marks)

1 (i)	
1 (ii)	
	(answer space continued on next page)

1 (ii)	**(continued)**
1 (iii)	

2 (i)	
2 (ii)	
2 (iii)	
2 (iv)	
	(answer space continued on next page)

2 (iv)	(continued)

3 (i)	
3 (ii)	

3 (iii)	
3 (iv)	

3 (v)	
3 (vi)	

Section B (24 marks)

4 (i)	

4 (ii)

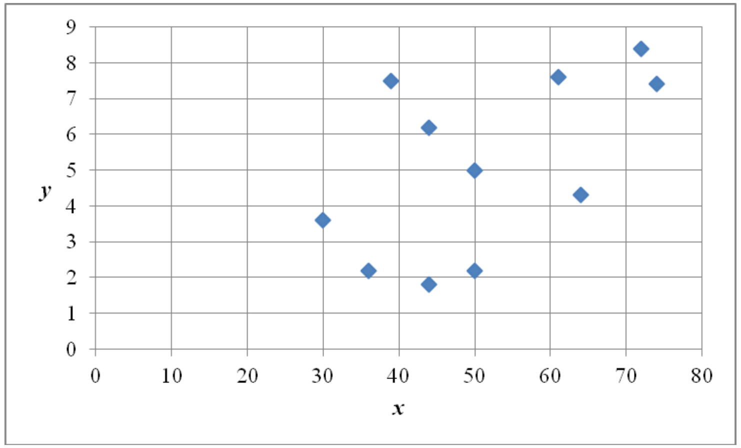
4 (iii)

4 (iv)	
4 (v)	
4 (vi)	
	(answer space continued on next page)

4 (vi)	**(continued)**
4 (vii)	

Question			Answer	Marks	Guidance
1	(i)		Unpaired *t* test –allow 2-sample *t* test	E1	
			Wilcoxon rank sum test (or Mann-Whitney 2-sample test)	E1	
			If the two populations were Normally distributed with equal variances then the unpaired *t* test would be preferable. Otherwise the Wilcoxon rank sum test would have to be used.	E1	For either Normally distributed or equal variances associated with unpaired *t* test (or for both associated with *t* test)
				E1	For full answer. Condone 'symmetrically distributed about the median'
				[4]	
1	(ii)		H_0: the medians of the two populations are the same	B1	For medians
			H_1: the median of the afternoon population is lower	B1	Need population for second mark Note: Explicit statement re shapes of distributions (eg that they are the same shape) is not required.
			Wilcoxon rank sum test (or Mann-Whitney form thereof)	M1	
			Ranks are		
			Mor Rank 3 5 8 9 10 14 15 17 18 19 21 22 Aft Rank 1 2 4 6 7 11 12 13 16 20	M1	Combined ranking
			Rank sum for smaller sample (for Afternoon) is 92	A1	FT (M-W stat = 0 + 0 + 1 + 2 + 2 + 5 + 5 + 5 + 7 + 10 = 37)
			Refer to (10,12) table	B1	
			1–tail 5% critical value is 89 [or 34 for M–W]	B1	
			92 > 89	M1	For sensible comparison leading to a correct conclusion
			Not significant	A1	
			There is insufficient evidence to suggest that, on average, the afternoon waiting times are shorter than the morning waiting times	E1	
				[10]	
1	(iii)		A paired design can only be used if there is some natural data pairing, such as the same people at both a morning and an afternoon appointment.	E1	
			This is very unlikely and so a paired design is not appropriate.	E1	For not appropriate (with justification)
				[2]	

Question			Answer	Marks	Guidance
2	(i)		The pairing will eliminate any differences in individual subjects' abilities and so will compare the times with and without the incentive scheme	E1 E1 **[2]**	
2	(ii)		The Wilcoxon signed rank test	E1 **[1]**	
2	(iii)		The population of differences of times must be Normally distributed	E1 E1 **[2]**	For population of differences For Normally distributed
2	(iv)		H_0: $\mu_D = 0$	B1	
			H_1: $\mu_D < 0$	B1	
			Where μ_D denotes the population mean for differences.	B1	Condone absence of "population" if correct notation "μ" has been used, but do NOT accept $\overline{D}$ or similar unless explicitly stated to be population means.
			No further marks unless paired comparison t test		Hypotheses explained in words only must include "population"
			Differences (after incentive – before incentive) are		
			–0.9 –1.4 –0.5 0.3 –0.2 0.5 –0.6 –1.0 –1.3 0.5 –0.9	M1	For differences
			$\bar{d} = -0.50$ $s_{n-1} = 0.6899$	A1	For both
			Test statistic is $\dfrac{-0.50 - 0}{0.6899/\sqrt{11}} = -2.404$	M1 A1	FT their $\bar{d}$ and s_{n-1} CAO but FT from here if M1 awarded
			Refer to t_{10}	M1	For t_{10}
			One tailed 5% critical value is –1.812	A1	Must be minus 1.812 unless absolute values or (before – after) are being compared. No FT if wrong. M1 can be implied by correct cv
			So significant	E1	
			There is sufficient evidence to suggest that the mean picking time after the incentive scheme is introduced is less than it was before	E1 **[11]**	

Question	Answer	Marks	Guidance
3 (i)		G1 G1 G1 **[3]**	Linear axes, including labels Correct zero or clear broken scale All points correct (allow 2 errors)
3 (ii)	r = 0.5609	M1 A1 **[2]**	
3 (iii)	The population should have a bivariate Normal distribution. In this case the points appear to lie in an elliptical pattern which suggests that the population may have a bivariate Normal distribution.	E1 E1 **[2]**	For bivariate Normal For roughly elliptical
3 (iv)	H_0: $\rho = 0$ H_1: $\rho > 0$ (one-tailed test) where ρ is the correlation coefficient for the underlying bivariate population. For $n = 11$, one tailed 1% critical value = 0.6851 Not significant There is insufficient evidence to suggest that there is positive correlation between total rainfall and weight of crop harvested per square metre.	B1 B1 B1 B1 E1 E1 **[6]**	FT their cv and test statistic from part (ii) (provided | test statistic | ≤ 1) Condone '…positive correlation between x and y'

Question			Answer	Marks	Guidance
3	(v)		If the locations were not chosen randomly, then making inferences from the sample would not be valid.	E1 E1 **[2]**	For 'making inferences' For 'not be valid' Allow other reasonable answers
3	(vi)		A test based on Spearman's rank correlation coefficient.	E1 **[1]**	
4	(i)		Method C Because with method A) a particular fisherman might just take fish from one area and with method B) if all fish are caught on a particular day, the larger or the smaller fish might not be available to catch	E1 E1 **[2]**	 Allow other sensible comments
4	(ii)		Allocate numbers 1 to 120 to the fish. Use random numbers to choose 10 random numbers. If any repeats appear, choose further random numbers to replace them.	E1 E1 E1 **[3]**	
4	(iii)		Because if she samples on different days, it might be that weather conditions vary and this may result in longer or shorter fish being caught.	E1 E1 **[2]**	Allow other sensible comments Max 1 mark unless reference to lengths
4	(iv)		$\bar{x} = \frac{7683.5}{50} = 153.67$ $s^2 = \frac{1191300 - \frac{7683.5^2}{50}}{49} = \frac{10576.55}{49} = 215.85$ $s = \sqrt{215.85} = 14.69$	B1 M1 A1 **[3]**	 For S_{xx}

Question			Answer	Marks	Guidance
4	(v)		Because both samples are large	E1 **[1]**	Condone 'large sample(s)'
4	(vi)		H_0: $\mu_P = \mu_Q$ H_1: $\mu_P \neq \mu_Q$ Where μ_P, μ_Q denote the population mean lengths of fish in Lakes P and Q respectively	B1 B1 B1	Condone absence of "population" if correct notation "μ" has been used, but do NOT accept $\overline{X}$ and $\overline{Y}$ or similar unless explicitly stated to be population means. Accept hypothesis explained in words, provided "population" appears. Condone μ = population mean
			2–sample test based on N(0,1)	E1	
			Test statistic is	M1	Numerator
			$\dfrac{153.67 - 151.7}{\sqrt{\dfrac{215.85}{50} + \dfrac{15.6^2}{50}}} = \dfrac{1.97}{3.03} = 0.650$	M1 A1	Denominator CAO
			2–tailed 10% point of N(0,1) is 1.645	B1	
			0.650 < 1.645	M1	FT their cv and test statistic
			Not significant	A1	
			There is insufficient evidence to suggest that on average the lengths of fish in the two lakes are different.	E1 **[11]**	
4	(vii)		The result of a hypothesis test always has an element of doubt. For example, at the 5% level, 5 times in every 100, the null hypothesis will be rejected when it is in fact true.	E1 E1 **[2]**	Allow E1 for 'sample might not be representative'

Unit level raw mark and UMS grade boundaries June 2014 series

AS GCE / Advanced GCE / AS GCE Double Award / Advanced GCE Double Award

GCE Mathematics (MEI)

		Max Mark	a	b	c	d	e	u
4751/01 (C1) MEI Introduction to Advanced Mathematics	Raw	72	61	56	51	46	42	0
	UMS	100	80	70	60	50	40	0
4752/01 (C2) MEI Concepts for Advanced Mathematics	Raw	72	57	51	45	39	33	0
	UMS	100	80	70	60	50	40	0
4753/01 (C3) MEI Methods for Advanced Mathematics with Coursework: Written Paper	Raw	72	58	52	47	42	36	0
4753/02 (C3) MEI Methods for Advanced Mathematics with Coursework: Coursework	Raw	18	15	13	11	9	8	0
4753/82 (C3) MEI Methods for Advanced Mathematics with Coursework: Carried Forward Coursework Mark	Raw	18	15	13	11	9	8	0
4753 (C3) MEI Methods for Advanced Mathematics with Coursework	UMS	100	80	70	60	50	40	0
4754/01 (C4) MEI Applications of Advanced Mathematics	Raw	90	68	61	54	47	41	0
	UMS	100	80	70	60	50	40	0
4755/01 (FP1) MEI Further Concepts for Advanced Mathematics	Raw	72	63	57	51	45	40	0
	UMS	100	80	70	60	50	40	0
4756/01 (FP2) MEI Further Methods for Advanced Mathematics	Raw	72	60	54	48	42	36	0
	UMS	100	80	70	60	50	40	0
4757/01 (FP3) MEI Further Applications of Advanced Mathematics	Raw	72	57	51	45	39	34	0
	UMS	100	80	70	60	50	40	0
4758/01 (DE) MEI Differential Equations with Coursework: Written Paper	Raw	72	63	56	50	44	37	0
4758/02 (DE) MEI Differential Equations with Coursework: Coursework	Raw	18	15	13	11	9	8	0
4758/82 (DE) MEI Differential Equations with Coursework: Carried Forward Coursework Mark	Raw	18	15	13	11	9	8	0
4758 (DE) MEI Differential Equations with Coursework	UMS	100	80	70	60	50	40	0
4761/01 (M1) MEI Mechanics 1	Raw	72	57	49	41	34	27	0
	UMS	100	80	70	60	50	40	0
4762/01 (M2) MEI Mechanics 2	Raw	72	57	49	41	34	27	0
	UMS	100	80	70	60	50	40	0
4763/01 (M3) MEI Mechanics 3	Raw	72	55	48	42	36	30	0
	UMS	100	80	70	60	50	40	0
4764/01 (M4) MEI Mechanics 4	Raw	72	48	41	34	28	22	0
	UMS	100	80	70	60	50	40	0
4766/01 (S1) MEI Statistics 1	Raw	72	61	53	46	39	32	0
	UMS	100	80	70	60	50	40	0
4767/01 (S2) MEI Statistics 2	Raw	72	60	53	46	40	34	0
	UMS	100	80	70	60	50	40	0
4768/01 (S3) MEI Statistics 3	Raw	72	61	54	47	41	35	0
	UMS	100	80	70	60	50	40	0
4769/01 (S4) MEI Statistics 4	Raw	72	56	49	42	35	28	0
	UMS	100	80	70	60	50	40	0
4771/01 (D1) MEI Decision Mathematics 1	Raw	72	51	46	41	36	31	0
	UMS	100	80	70	60	50	40	0
4772/01 (D2) MEI Decision Mathematics 2	Raw	72	46	41	36	31	26	0
	UMS	100	80	70	60	50	40	0
4773/01 (DC) MEI Decision Mathematics Computation	Raw	72	46	40	34	29	24	0
	UMS	100	80	70	60	50	40	0
4776/01 (NM) MEI Numerical Methods with Coursework: Written Paper	Raw	72	54	48	43	38	32	0
4776/02 (NM) MEI Numerical Methods with Coursework: Coursework	Raw	18	14	12	10	8	7	0
4776/82 (NM) MEI Numerical Methods with Coursework: Carried Forward Coursework Mark	Raw	18	14	12	10	8	7	0
4776 (NM) MEI Numerical Methods with Coursework	UMS	100	80	70	60	50	40	0
4777/01 (NC) MEI Numerical Computation	Raw	72	55	47	39	32	25	0
	UMS	100	80	70	60	50	40	0
4798/01 (FPT) Further Pure Mathematics with Technology	Raw	72	57	49	41	33	26	0
	UMS	100	80	70	60	50	40	0

GCE Statistics (MEI)

		Max Mark	a	b	c	d	e	u
G241/01 (Z1) Statistics 1	Raw	72	61	53	46	39	32	0
	UMS	100	80	70	60	50	40	0
G242/01 (Z2) Statistics 2	Raw	72	55	48	41	34	27	0
	UMS	100	80	70	60	50	40	0
G243/01 (Z3) Statistics 3	Raw	72	56	48	41	34	27	0
	UMS	100	80	70	60	50	40	0