OCR
RECOGNISING ACHIEVEMENT

Friday 1 June 2012 – Morning

AS GCE MEI STATISTICS

G243 Statistics 3 (Z3)

QUESTION PAPER

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

G243601012

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]**

OCR
RECOGNISING ACHIEVEMENT

Friday 1 June 2012 – Morning

AS GCE MEI STATISTICS

G243 Statistics 3 (Z3)

PRINTED ANSWER BOOK

Candidates answer on this Printed Answer Book.

OCR supplied materials:
- Question Paper G243 (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 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.

Section A (48 marks)

1 (i)

1 (ii)

1 (ii)	(continued)
1 (iii)	

2 (i)

2 (ii)

2 (ii)	(continued)

2 (iii)

2 (iv)

3 (i)

3 (ii)

3 (iii)	
3 (iv)	
3 (v)	

Section B (24 marks)

4 (i)	
4 (ii)	
4 (iii)	

4 (iv)	
4 (v)	
4 (vi)	
4 (vii)	(answer space continued overleaf)

4 (vii)	(continued)

OCR
RECOGNISING ACHIEVEMENT

GCE

Mathematics (MEI)

Advanced GCE

Unit **G243:** Statistics 3

Mark Scheme for June 2012

Oxford Cambridge and RSA Examinations

Annotations and abbreviations

Annotation in scoris	**Meaning**
✓and ✖	
BOD	Benefit of doubt
FT	Follow through
ISW	Ignore subsequent working
M0, M1	Method mark awarded 0, 1
A0, A1	Accuracy mark awarded 0, 1
B0, B1	Independent mark awarded 0, 1
SC	Special case
^	Omission sign
MR	Misread
Highlighting	
Other abbreviations in mark scheme	**Meaning**
E1	Mark for explaining
U1	Mark for correct units
G1	Mark for a correct feature on a graph
M1 dep*	Method mark dependent on a previous mark, indicated by *
cao	Correct answer only
oe	Or equivalent
rot	Rounded or truncated
soi	Seen or implied
www	Without wrong working

Subject-specific Marking Instructions for GCE Mathematics (MEI) Pure strand

a. Annotations should be used whenever appropriate during your marking.

The A, M and B annotations must be used on your standardisation scripts for responses that are not awarded either 0 or full marks. It is vital that you annotate standardisation scripts fully to show how the marks have been awarded.

For subsequent marking you must make it clear how you have arrived at the mark you have awarded.

b. An element of professional judgement is required in the marking of any written paper. Remember that the mark scheme is designed to assist in marking incorrect solutions. Correct *solutions* leading to correct answers are awarded full marks but work must not be judged on the answer alone, and answers that are given in the question, especially, must be validly obtained; key steps in the working must always be looked at and anything unfamiliar must be investigated thoroughly.

Correct but unfamiliar or unexpected methods are often signalled by a correct result following an *apparently* incorrect method. Such work must be carefully assessed. When a candidate adopts a method which does not correspond to the mark scheme, award marks according to the spirit of the basic scheme; if you are in any doubt whatsoever (especially if several marks or candidates are involved) you should contact your Team Leader.

c. The following types of marks are available.

M
A suitable method has been selected and *applied* in a manner which shows that the method is essentially understood. Method marks are not usually lost for numerical errors, algebraic slips or errors in units. However, it is not usually sufficient for a candidate just to indicate an intention of using some method or just to quote a formula; the formula or idea must be applied to the specific problem in hand, eg by substituting the relevant quantities into the formula. In some cases the nature of the errors allowed for the award of an M mark may be specified.

A
Accuracy mark, awarded for a correct answer or intermediate step correctly obtained. Accuracy marks cannot be given unless the associated Method mark is earned (or implied). Therefore M0 A1 cannot ever be awarded.

B
Mark for a correct result or statement independent of Method marks.

E

A given result is to be established or a result has to be explained. This usually requires more working or explanation than the establishment of an unknown result.

Unless otherwise indicated, marks once gained cannot subsequently be lost, eg wrong working following a correct form of answer is ignored. Sometimes this is reinforced in the mark scheme by the abbreviation isw. However, this would not apply to a case where a candidate passes through the correct answer as part of a wrong argument.

d. When a part of a question has two or more 'method' steps, the M marks are in principle independent unless the scheme specifically says otherwise; and similarly where there are several B marks allocated. (The notation 'dep *' is used to indicate that a particular mark is dependent on an earlier, asterisked, mark in the scheme.) Of course, in practice it may happen that when a candidate has once gone wrong in a part of a question, the work from there on is worthless so that no more marks can sensibly be given. On the other hand, when two or more steps are successfully run together by the candidate, the earlier marks are implied and full credit must be given.

e. The abbreviation ft implies that the A or B mark indicated is allowed for work correctly following on from previously incorrect results. Otherwise, A and B marks are given for correct work only — differences in notation are of course permitted. A (accuracy) marks are not given for answers obtained from incorrect working. When A or B marks are awarded for work at an intermediate stage of a solution, there may be various alternatives that are equally acceptable. In such cases, exactly what is acceptable will be detailed in the mark scheme rationale. If this is not the case please consult your Team Leader.

Sometimes the answer to one part of a question is used in a later part of the same question. In this case, A marks will often be 'follow through'. In such cases you must ensure that you refer back to the answer of the previous part question even if this is not shown within the image zone. You may find it easier to mark follow through questions candidate-by-candidate rather than question-by-question.

f. Wrong or missing units in an answer should not lead to the loss of a mark unless the scheme specifically indicates otherwise. Candidates are expected to give numerical answers to an appropriate degree of accuracy, with 3 significant figures often being the norm. Small variations in the degree of accuracy to which an answer is given (e.g. 2 or 4 significant figures where 3 is expected) should not normally be penalised, while answers which are grossly over- or under-specified should normally result in the loss of a mark. The situation regarding any particular cases where the accuracy of the answer may be a marking issue should be detailed in the mark scheme rationale. If in doubt, contact your Team Leader.

g. Rules for replaced work

If a candidate attempts a question more than once, and indicates which attempt he/she wishes to be marked, then examiners should do as the candidate requests.

If there are two or more attempts at a question which have not been crossed out, examiners should mark what appears to be the last (complete) attempt and ignore the others.

NB Follow these maths-specific instructions rather than those in the assessor handbook.

h. For a *genuine* misreading (of numbers or symbols) which is such that the object and the difficulty of the question remain unaltered, mark according to the scheme but following through from the candidate's data. A penalty is then applied; 1 mark is generally appropriate, though this may differ for some units. This is achieved by withholding one A mark in the question.

Note that a miscopy of the candidate's own working is not a misread but an accuracy error.

Question			Answer	Marks	Guidance
1	(i)		Make a list of all of the lorries. Choose one of the first 15 lorries at random, then choose every 15th lorry	E1 E1 E1 **[3]**	Or number all of the lorries or line up the lorries
1	(ii)		H_0: $\mu_D = 0$ H_1: $\mu_D \neq 0$ Where μ_D denotes the population mean for differences. Differences (biodiesel – conventional) are –0.5 –0.4 0.6 0.3 0.0 0.7 0.3 0.6 $\overline{d} = 0.2$ $s_{n-1} = 0.460$ Test statistic is $\frac{0.2-0}{0.460/\sqrt{8}} = 1.23$ Refer to t_7 Two tailed 5% critical value is 2.365 Not significant There is insufficient evidence to suggest that the mean fuel consumption for lorries using biodiesel is different from that for lorries using conventional diesel	B1 B1 B1 M1 A1 M1 A1 M1 A1 E1 E1 **[11]**	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. Condone μ in place of μ_D Hypotheses explained in words only must include "population" No further marks unless paired comparison *t* test For differences – allow one error For both CAO FT their $\overline{d}$ and s_{n-1} CAO but FT from here if M1 awarded No FT if wrong No FT if wrong
1	(iii)		Population of differences is Normally distributed Wilcoxon signed rank test	B1 B1 **[2]**	Ignore comments about sample size, etc Allow just 'Normally distributed' but not 'Data is Normally distributed . Not just 'Wilcoxon', but allow 'Wilcoxon paired'
2	(i)		No Since the samples are not paired	B1 E1 **[2]**	Do not allow if the reason is wrong, but do allow if reason not relevant. Zero if answer yes.
2	(ii)		H_0: the medians of the two populations are the same H_1: the medians of the two populations are different	B1 B1	For medians Need population for second mark (1) Explicit statement re shapes of distributions (eg that they are the same shape) is not required. (2) More formal

Question			Answer	Marks	Guidance
					statements of hypotheses gain both marks (eg cdfs are $F(x)$ and $F(x-\Delta)$, H_0 is $\Delta = 0$ etc).
			Wilcoxon rank sum test (or Mann-Whitney form thereof)	M1	Soi . NB: Nofurther marks if not Wilcoxon rank sum test
			Ranks are Rank A 3 13 18 14 20 17 16 7 6 8 Rank B 15 11 4 9 10 2 1 19 12 5	M1	Combined ranking
			Smaller rank sum (for B) is 88	A1	FT (M–W stat = 6 + 4 + 1 + 4 + 4 + 0 + 0 + 9 + 4 + 1 = 33)
			Refer to (10,10) table	B1	Soi (can be implied by correct cv)
			2–tail 10% critical value is 82 [or 27 for M–W]	M1 A1	No FT from here if cv wrong
			Not significant	E1	
			There is insufficient evidence to suggest that the breaking stresses are different	E1 **[10]**	
2	(iii)		No	B1 **[1]**	Ignore reason (if given)
2	(iv)		Two sample test based on the Normal distribution	B1	Allow 'normal test' but NOT 'normal test for paired samples'
			They would not need to be known because the samples are large so they can be estimated from the data	E1 E1 **[3]**	
3	(i)		$S_{xy} = \Sigma xy - \frac{1}{n}\Sigma x \Sigma y = 147.6 - \frac{1}{50} \times 26.43 \times 265.4$ $= 7.310$	M1	For method for S_{xy} NB $\bar{x} = 0.5286 \quad \bar{y} = 5.308$
			$S_{xx} = \Sigma x^2 - \frac{1}{n}(\Sigma x)^2 = 16.62 - \frac{1}{50} \times 26.43^2 = 2.649$ $S_{yy} = \Sigma y^2 - \frac{1}{n}(\Sigma y)^2 = 1576.9 - \frac{1}{50} \times 265.4^2 = 168.2$	M1	For method for at least one of S_{xx} or S_{yy}
			$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{7.310}{\sqrt{2.649 \times 168.2}} = 0.346$	M1 M1 A1 **[5]**	For structure of r For √ CAO (0.34 to 0.35)

Question			Answer	Marks	Guidance
3	(ii)		H_0: $\rho = 0$ H_1: $\rho > 0$ (one-tailed test) where ρ is the correlation coefficient for the underlying bivariate population. For $n = 50$, one tailed 5% critical value = 0.2353 Significant There is evidence to suggest that there is positive correlation between attendance and academic performance.	B1 B1 B1 B1 E1 E1 **[6]**	Hypotheses explained in words only must include "population" But allow 'in its schools' (oe) as implying population. Allow even if H_0 and H_1 wrong but not if H_1 two tailed No ft if cv wrong Do not allow unless H_0 and H_1 essentially correct Only allow final two marks if $0 < r < 1$. Do not allow 'between x and y'
3	(iii)		The population should have a bivariate Normal distribution. The scatter diagram should have a roughly elliptical shape.	B1 B1 **[2]**	Must include 'bivariate'
3	(iv)		Because the whole population has been used.	E1 **[1]**	Or 'Hypothesis tests are based on a sample'
3	(v)		Correlation does not imply causation. There may be a third variable which causes correlation between attendance and academic performance. An example of a possible third variable is parental encouragement	E1 E1 **[2]**	For one of these statements. For another of these statements. Allow any sensible alternatives. If neither of first two statements (oe), max B1
4	(i)		Treatment D is a control. It is needed to check whether any of the other treatments have an effect.	E1 E1 **[2]**	 Or 'To allow a comparison to be made' oe
4	(ii)		The experiment is not replicated. With only one observation for each treatment, it is not possible to gain any knowledge of variability. Thus it is not possible to compare the 4 treatments. There need to be many observations for each treatment.	 E3,2,1, 0 **[3]**	 Award marks dependent on quality of answer Max 2 if nothing about variability Allow E1 for sample too small (oe) Allow E1 for 'the people may have different types of cold' (oe)

Question			Answer	Marks	Guidance
4	(iii)		Examples of suitable comments: Infecting people with the cold virus is not ethical. People might suffer from complications. There may be more than one type of common cold virus. The treatments need to be allocated at random. This is a distinct improvement on the suggestion in part (ii).	E1 E1 **[2]**	Award E2 for two suitable comments Award E1 for one suitable comment
4	(iv)		If people were not allocated randomly then perhaps younger people might be given for example treatment A, and perhaps younger people recover more quickly than older people.	E1 E1 **[2]**	Allow other suitable answers Allow E1 for 'to avoid any bias' oe Allow E2 for 'random sample required to perform a hypothesis test'
4	(v)		Because it would not include children or elderly people.	E1 E1 **[2]**	Allow other suitable answers eg The company may not be located in the UK
4	(vi)		They are given placebo tablets so that they appear to be having some treatment.	E2 **[2]**	
4	(vii)		H_0: $\mu_A = \mu_D$ H_1: $\mu_A < \mu_D$ Where μ_A μ_D denote the population mean recovery times under treatments A and D respectively 2–sample test based on N(0,1) Test statistic is $\frac{4.770 - 4.926}{\sqrt{\frac{1.747}{43} + \frac{1.594}{49}}} = \frac{-0.156}{0.2705} = -0.577$ 1–tailed 5% point of N(0,1) is –1.645 –0.577 > –1.645 Not significant There is insufficient evidence to suggest that the average recovery time with the new drug is less than that with no drug.	B1 B1 B1 E1 M1 M1 A1 B1 M1 A1 E1 **[11]**	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. soi. No further marks for other tests Numerator Denominator CAO (–0.57 to –0.58) FT from here if at least one M1 earned For ± 1.645 No FT if wrong

OCR (Oxford Cambridge and RSA Examinations)
1 Hills Road
Cambridge
CB1 2EU

OCR Customer Contact Centre

Education and Learning
Telephone: 01223 553998
Facsimile: 01223 552627
Email: general.qualifications@ocr.org.uk

www.ocr.org.uk

For staff training purposes and as part of our quality assurance programme your call may be recorded or monitored

Oxford Cambridge and RSA Examinations
is a Company Limited by Guarantee
Registered in England
Registered Office; 1 Hills Road, Cambridge, CB1 2EU
Registered Company Number: 3484466
OCR is an exempt Charity

OCR (Oxford Cambridge and RSA Examinations)
Head office
Telephone: 01223 552552
Facsimile: 01223 552553