GCE

Statistics (MEI)

Advanced Subsidiary GCE **AS H132**

Mark Scheme for the Unit

June 2006

H132/MS/R/06

Oxford Cambridge and RSA Examinations

OCR (Oxford, Cambridge and RSA Examinations) is a unitary awarding body, established by the University of Cambridge Local Examinations Syndicate and the RSA Examinations Board in January 1998. OCR provides a full range of GCSE, A level, GNVQ, Key Skills and other qualifications for schools and colleges in the United Kingdom, including those previously provided by MEG and OCEAC. It is also responsible for developing new syllabuses to meet national requirements and the needs of students and teachers.

This mark scheme is published as an aid to teachers and students, to indicate the requirements of the examination. It shows the basis on which marks were awarded by Examiners. It does not indicate the details of the discussions which took place at an Examiners' meeting before marking commenced.

All Examiners are instructed that alternative correct answers and unexpected approaches in candidates' scripts must be given marks that fairly reflect the relevant knowledge and skills demonstrated.

Mark schemes should be read in conjunction with the published question papers and the Report on the Examination.

OCR will not enter into any discussion or correspondence in connection with this mark scheme.

© OCR 2006

Any enquiries about publications should be addressed to:

OCR Publications
PO Box 5050
Annersley
NOTTINGHAM
NG15 0DL

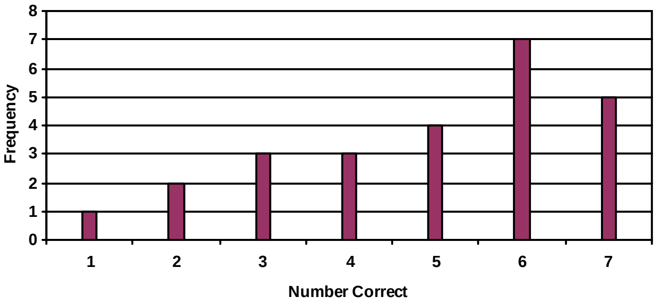
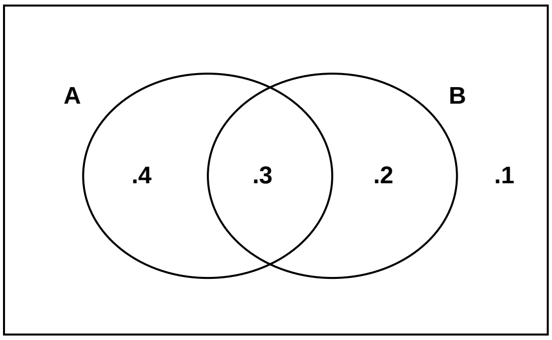
Telephone: 0870 870 6622
Facsimile: 0870 870 6621
E-mail: publications@ocr.org.uk

CONTENTS

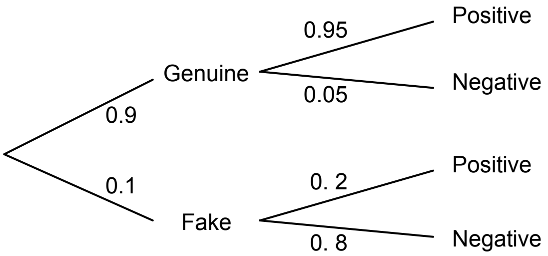
Advanced Subsidiary GCE Statistics (H132)

MARK SCHEMES FOR THE UNITS

Mark Scheme G241
June 2006

Q1 **(i)**		G1 Labelled linear scales G1 Height of lines	**2**
(ii)	Negative (skewness)	B1	**1**
(iii)	Σ*fx* = 123 so mean = 123/25 = 4.92 o.e. $S_{xx} = 681 - \frac{123^2}{25} = 75.84$ M.s.d = $\frac{75.84}{25} = 3.034$	B1 M1 for S_{xx} attempted A1 FT their 4.92	**3**
(iv)	Total for 25 days is 123 and totals for 31 days is 155. Hence total for next 6 days is 32 and so mean = 5.33	M1 31 x 5 – 25xtheir 4.92 A1 FT their 123	**2**
		TOTAL	**8**
Q2 **(i)**	$\mathrm{P}(A\cap B)=\mathrm{P}(A)\mathrm{P}(B\mid A)=\frac{7}{10}\times\frac{3}{7}$ $\rightarrow \mathrm{P}(A\cap B)=0.3$ o.e.	M1 Product of these fractions A1	**2**
(ii)		B1FT either 0.4 or 0.2 in correct place B1FT all correct and labelled	**2**
(iii)	P(*B*|*A*) ≠ P(*B*), 3/7 ≠ 0.5 Unequal so not independent	E1 Correct comparison E1*dep* for 'not independent'	**2**
(iv)	3/7 < 0.5 so Isobel is less likely to score when her parents attend	E1 for comparison E1*dep*	**2**
		TOTAL	**8**

Q3 **(i)**	P(X = 1) = 7k, P(X = 2) = 12k, P(X = 3) = 15k, P(X = 4) = 16k 50k = 1 so k = 1/50	M1 for addition of four multiples of k A1 **ANSWER GIVEN**	**2**
(ii)	E(X) = 1 x 7k + 2 x 12k + 3 x 15k + 4 x 16k = 140k = 2.8 OR E(X) = 1 x ${}^{7}/_{50}$ + 2 x ${}^{12}/_{50}$ + 3 x ${}^{15}/_{50}$ + 4 x ${}^{16}/_{50}$ = ${}^{140}/_{50}$ = 2.8 oe Var(X) = 1 x 7k + 4 x 12k + 9 x 15k + 16 x 16k -7.84 = 1.08 OR Var(X) = 1 x ${}^{7}/_{50}$ + 4 x ${}^{12}/_{50}$ + 9 x ${}^{15}/_{50}$ + 16 x ${}^{16}/_{50}$ -7.84 = 8.92 – 7.84 = 1.08	M1 for Σxp (at least 3 terms correct) A1 CAO M1 Σx^2p (at least 3 terms correct) M1*dep* for – their E(X)² NB provided Var(X) > 0 A1 FT their E(X)	**5**
		TOTAL	**7**
Q4 **(i)**	4 x 5 x 3 = 60	M1 for 4 x 5 x 3 A1 CAO	**2**
(ii)	**(*A*)** $\binom{4}{2}$ = 6 **(*B*)** $\binom{4}{2}\binom{5}{2}\binom{3}{2}$ = 180	B1 **ANSWER GIVEN** B1 CAO	**2**
(iii)	**(*A*)** 1/5 **(*B*)** $\frac{3}{4}\times\frac{4}{5}\times\frac{2}{3}=\frac{2}{5}$	B1 CAO M1 for $\frac{3}{4}\times\frac{4}{5}\times\frac{2}{3}$ A1	**3**
		TOTAL	**7**
Q5 **(i)**	$\text{P}(X=2)=\binom{3}{2}\times 0.87^2\times 0.13=0.2952$	M1 0.87^2 x 0.13 M1 $\binom{3}{2}$ x p^2q **with p+q=1** A1 CAO	**3**
(ii)	In 50 throws expect 50 (0.2952) = 14.76 times	B1 FT	**1**
(iii)	P (two 20's twice) = $\binom{4}{2}\times 0.2952^2\times 0.7048^2=0.2597$	M1 $0.2952^2\times 0.7048^2$ A1 FT their 0.2952	**2**
		TOTAL	**6**

Q6 **(i)**		G1 for left hand set of branches fully correct including labels and probabilities G1 for right hand set of branches fully correct	**2**
(ii)	P (test is positive) = (0.9)(0.95) + (0.1)(0.2) = 0.875	M1 Two correct pairs added A1 CAO	**2**
(iii)	P (test is correct) = (0.9)(0.95) + (0.1)(0.8) = 0.935	M1 Two correct pairs added A1 CAO	**2**
(iv)	P (Genuine|Positive) = 0.855/0.875 = 0.977	M1 Numerator M1 Denominator A1 CAO	**3**
(v)	P (Fake|Negative) = 0.08/0.125 = 0.64	M1 Numerator M1 Denominator A1 CAO	**3**
(vi)	EITHER: A positive test means that the painting is almost certain to be genuine so no need for a further test. However, more than a third of those paintings with a negative result are genuine so a further test is needed. NOTE: Allow sensible alternative answers	E1FT E1FT	**2**
(vii)	P (all 3 genuine) = (0.9 x 0.05 x 0.96)^3 = (0.045 x 0.96)^3 = (0.0432)^3 = 0.0000806	M1 for 0.9 x 0.05 (=0.045) M1 for complete correct triple product M1*indep* for cubing A1 CAO	**4**
		TOTAL	**18**

Q7 **(i)**	$X \sim B(20, 0.1)$ (**A**) $P(X = 1) = \binom{20}{1} \times 0.1 \times 0.9^{19} = 0.2702$ OR from tables $0.3917 - 0.1216 = 0.2701$	M1 0.1 x 0.9^{19} M1 $\binom{20}{1}$ x pq^{19} A1 CAO OR: M2 for 0.3917 – 0.1216 A1 CAO	**3**
	(**B**) $P(X \geq 1) = 1 - 0.1216 = 0.8784$	M1 P(X=0) *provided that P(X≥1)=1–P(X≤1) not seen* M1 1-P(X=0) A1 CAO	**3**
(ii)	EITHER: $1 - 0.9^n \geq 0.8$ $0.9^n \leq 0.2$ Minimum $n = 16$ OR (using trial and improvement): Trial with 0.9^{15} or 0.9^{16} or 0.9^{17} $1 - 0.9^{15} = 0.7941 < 0.8$ and $1 - 0.9^{16} = 0.8147 > 0.8$ Minimum $n = 16$ NOTE: $n = 16$ unsupported scores SC1 only	M1 for 0.9^n M1 for inequality A1 CAO M1 M1 A1 CAO	**3**
(iii)	(**A**) Let p = probability of a randomly selected rock containing a fossil (for population) H_0: $p = 0.1$ H_1: $p < 0.1$	B1 for definition of p B1 for H_0 B1 for H_1	**3**
	(**B**) Let $X \sim B(30, 0.1)$ $P(X \leq 0) = 0.0424 < 5\%$ $P(X \leq 1) = 0.0424 + 0.1413 = 0.1837 > 5\%$ So critical region consists only of 0.	M1 for attempt to find $P(X \leq 0)$ or $P(X \leq 1)$ using binomial M1 for both attempted M1 for comparison of either of the above with 5% A1 for critical region dep on both comparisons **(NB Answer given)**	**4**
	(**C**) 2 does not lie in the critical region. So there is insufficient evidence to reject the null hypothesis and we conclude that it seems that 10% of rocks in this area contain fossils.	M1 for comparison A1 for conclusion **in context**	**2**
		TOTAL	**18**

Mark Scheme G242
June 2006

Q1			
(a)(i)	P(time < 94) = P(Z < $\frac{94-93}{0.9}$) P(Z < 1.111) 0.8667	M1 A1 A1	3
(a)(ii)	For 4 games, the probability that all four last less than 94 minutes is $(0.8667)^4$ = 0.5642… 1 - $(0.8667)^4$ 0.4357…	M1 M1 A1	3
(b)(i)	$\dfrac{1989670-\dfrac{10920^2}{60}}{59}$ = 37.7966…. = 37.80 (2d.p.)	M1 A1	2
(b)(ii)	Sample mean = 182 182 ± 1.96×√37.80÷√60 (B1 for 1.96) (180.4,183.6)	B1 B1M1 A1	4
(b)(iii)	The C.I. does not contain 188 suggesting the mean height of goalkeepers could be greater than that of outfield players. Relevant comment based on probability – in context - e.g. A wider CI based on the outfield players data – 99%, say - would still not contain 188 cm.	E1(188 not in) E1(suggesting) E1(mean greater) E1(prob^y link)	4

Q2			
(i)	$\Sigma fx \div \Sigma f = 340 \div 100$ (=3.4) $(1486 - 340^2/100) \div 99$ Sample variance = $3\frac{1}{3}$ No reason to doubt manager as mean ≈ variance	B1 M1 A1 E1	4
(ii)	0.0334, 0.1134, 0.1929, 0.2187, 0.1858, 0.1263, 0.0716, 0.0348, 0.0231 3.34, 11.34, 19.29, 21.87, 18.58, 12.63, 7.16, 3.48, 2.31	M1A2 M1A1	5
(iii)	(*A*) Low expected frequencies have a disproportionate influence on the value of X^2 and may make the procedure a poor approx$^{\underline{n}}$ (*B*) 5 degrees of freedom (7 – 1 – 1) Critical value at 5% level is χ^2 = 11.07 5.127 < 11.07 so not significant The Poisson model seems a good fit.	E1 B1 B1 M1 E1	5

Q3			
(i)	Mean = $3\frac{13}{30}$ & SD = 1.49989…	B1, B1	**2**
(ii)	$H_0 : \mu = 4.9$ & $H_1 : \mu < 4.9$ Where μ represents the population mean pollution level *t* distribution needed $t = \dfrac{3\frac{13}{30} - 4.9}{SD/\sqrt{12}}$ = -3.39 (3s.f.) 11 degrees of freedom At 5% level, critical value of *t* = 1.796 -3.394 < -1.796 so the result is significant Evidence suggests there is a reduction in mean pollution level	B1 B1 B1 M1A1 (FT) B1 B1 M1 A1 E1	**10**
(iii)	Sample is random	B1	**1**

<table>
<tr><td>Q4</td><td></td><td></td><td></td></tr>
<tr><td>(i)</td><td>
H₀: there is no association between personality and colour preference
H₁: there is an association between personality and colour preference

Expected frequencies
<table>
<tr><td></td><td></td><td>Introvert</td><td>Extrovert</td></tr>
<tr><td>Preferred colour</td><td>Red</td><td>32.4</td><td>47.6</td></tr>
<tr><td></td><td>Yellow</td><td>9.72</td><td>14.28</td></tr>
<tr><td></td><td>Green</td><td>20.25</td><td>29.75</td></tr>
<tr><td></td><td>Blue</td><td>18.63</td><td>27.37</td></tr>
</table>
Contribution to X^2
<table>
<tr><td></td><td></td><td>Introvert</td><td>Extrovert</td></tr>
<tr><td>Preferred colour</td><td>Red</td><td>2.18</td><td>1.48</td></tr>
<tr><td></td><td>Yellow</td><td>0.76</td><td>0.52</td></tr>
<tr><td></td><td>Green</td><td>0.15</td><td>0.10</td></tr>
<tr><td></td><td>Blue</td><td>4.71</td><td>3.21</td></tr>
</table>
X^2= 13.11 (13.11399… without rounding)

3 degrees of freedom

Critical value for 5% significance level is 7.815
As 13.11 > 7.815 the result is significant

There is evidence of an association between personality and colour preference.
</td><td>B1
B1

M1A1

M1A1

A1

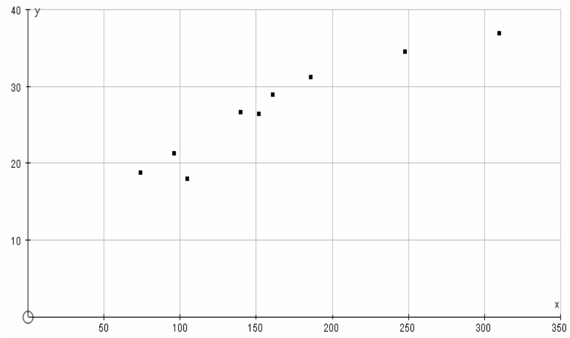
B1

B1
M1A1

E1</td><td>12</td></tr>
<tr><td>(ii)</td><td>People classed as extrovert tend to prefer red.
People classed as introvert tend to prefer blue.
Third relevant comment e.g. referring to specific contribution to X^2</td><td>E1
E1
E1</td><td>3</td></tr>
</table>

Q5			
(i)	H_0: population median = 12 H_1: population median < 12 Actual differences +8 -1 -3 +5 +7 -11 -6 -2 -9 -10 Associated ranks 7 1 3 4 6 10 5 2 8 9 $T^- = 1 + 3 + 10 + 5 + 2 + 8 + 9 = 38$ $T^+ = 7 + 4 + 6 = 17$ $\therefore\ T = 17$ From tables – at the 5% level of significance in a one-tailed Wilcoxon signed rank test, the critical value of T is 10 $17 > 10$ $\therefore$ the result is not significant The evidence does not suggest the drug is effective.	B1 B1 B1 M1A1 B1 B1 B1 B1 M1A1 E1	**12**
(ii)	Sample too small *t* distribution	B1 B1	**2**

Mark Scheme G243
June 2006

Q1				
(i)	 Looks reasonably linear, … … maybe flattens a bit at the top.	G1 G1 G1 E1 E1	Axes, including labels. "x" and "y" suffice as they are defined in the question. Correct zero, or clear "breaks". All points correct. Allow one error.	5
(ii)	x ranks 5 1 2 9 8 3 4 7 6 y ranks 4 2 3 9 8 1 5 7 6 | d | 1 1 1 0 0 2 1 0 0 Σd^2 = 8 $r_s = 1 - \frac{6 \times 8}{9 \times 80} = 0 \cdot 9333$	B1 B1 M1 A1	c.a.o.	4
(iii)	Critical value for n = 9 at two-sided 5% level is 0·7000. Significant. Seems there is an association between rainfall and yield.	B1 E1 E1	No ft if wrong. S.C. Use of the 1-tail point (0·6000) can get either, but not both, of these E marks.	3
(iv)	No real suggestion of bivariate Normality. H_0 is $\rho = 0$, where ρ is the correlation coefficient for the underlying bivariate population.	E1 B1 B1	Or equivalent statement.	3
				15

Q2				
(a) (i)	H_0 : the medians of the two populations are the same. H_1 : the medians of the two populations are different. Wilcoxon rank sum test (or Mann-Whitney form thereof). Ranks are A 1 2 4 5 6 10 11 13 B 3 7 8 9 12 14 15 16 17 18 Rank sum for smaller sample is 52. Refer to (8,10) table. 2-tail 5% critical value is 53. Significant. Seems areas are not equivalent in this regard.	B1 B1 M1 B1 M1 A1 E1 E1	Or more formal statements. (Or M-W statistic = 16) No ft from here if wrong. (Or 17 for M-W). No ft if wrong. ft only c's test statistic. ft only c's test statistic.	8
(ii)	There is nothing obvious that can be measured. Depends on subjective judgement of analyst.	E1 E1	Or other sensible comments. Or other sensible comments.	2
(b) (i)	The population standard deviation.	B1 B1		2
(ii)	The critical point is 1·96. Not significant. Appears that the average air pollution in towns in the two regions may be assumed to be the same.	M1 E1 E1	No ft if wrong.	3
				15

Q3				
(i)	There is a list of the population.	E1		
	(As far as is known) there are no cycles or patterns in the list that might be related to the aim of the sampling.	E1		2
(ii)	Pick one of first 8 in alphabetical list at random …	M1		
	… then every 8th.	M1		2
(iii)	H$_0$: $\mu_D = 0$ (or $\mu_1 = \mu_1$ etc)	B1	Do NOT allow $\overline{D}$ or similar unless it is clearly and explicitly stated to be a population mean. Hypotheses in words only must include "population".	
	H$_1$: $\mu_D > 0$ (or $\mu_2 > \mu_1$ etc)	B1	Or "<" for "first" – "second".	
	where μ_D is the population mean for the differences.	B1	For adequate verbal definition. Allow absence of "population" here if correct notation μ has been used.	
	MUST be PAIRED COMPARISON *t* test.			
	Use of differences.	M1		
	Differences are: 38 18 72 29 –12 99 23 41 –2 46 52 62			
	$\bar{d} = 38\cdot8\dot{3}$ $s_{n-1} = 30\cdot924(64)$	A1	For both. [s_n = 29·6081 NOT allowed.]	
	Test statistic is $\dfrac{38\cdot8\dot{3}-0}{\dfrac{30\cdot924(64)}{\sqrt{12}}}$	M1	Allow c's $\bar{d}$ and/or s_{n-1}. Allow alternative: 0 + (c's 1·796) × $\dfrac{30\cdot924(64)}{\sqrt{12}}$ (= 16·03) for subsequent comparison with $\bar{d}$. (Or $\bar{d}$ – (c's 1·796) × $\dfrac{30\cdot924(64)}{\sqrt{12}}$ (= 22·80) for comparison with 0.)	
	=4·35.	A1	c.a.o. but ft from here if M1 awarded, but no marks from here on if not paired *t* test. Use of 0 – $\bar{d}$ scores M1A0, but ft.	
	Refer to t_{11}.	M1	No ft from here if wrong.	
	Upper 5% point is 1·796.	A1	No ft from here if wrong.	
	Significant.	E1	ft only c's test statistic.	
	Seems mean score for second test is greater than for first.	E1	ft only c's test statistic.	11
				15

Q4				
(i)	Both the pressure and the temperature are tried at settings above and below those currently used.	E2	(E0, E1, E2)	2
	There is no replication so variability cannot be assessed.	E2	(E0, E1, E2)	2
(ii)	Pooled $s^2 = \frac{(4\times10\cdot243)+(4\times14\cdot647)}{8} = 12\cdot445$	M1 A1	For any reasonable attempt at pooling. If correct.	
	Test statistic is $\frac{26\cdot96-24\cdot92\ \ (-0)}{\sqrt{12\cdot445}\sqrt{\frac{1}{5}+\frac{1}{5}}}$ $= \frac{2\cdot04}{2\cdot2311} = 0\cdot9143$	M1 M1 M1 A1	For numerator. $\sqrt{12\cdot445}$. Use of c's pooled variance. $\sqrt{\frac{1}{5}+\frac{1}{5}}$. ft from here if all M marks earned.	
	Refer to t_8.	M1	No ft from here if wrong.	
	Double-tailed 5% point is 2·306.	A1	No ft from here if wrong.	
	Not significant.	E1	ft only c's test statistic.	
	No evidence that population mean yields are different.	E1	ft only c's test statistic.	10
(iii)	All runs in first chamber are on different days from all runs in second, …	E1		
	… so any variability in the raw materials may affect the chambers in different ways.	E1		
	Do a run in each chamber on Monday, then a run in each chamber on Tuesday, and so on.	E1	Or equivalent.	
	Any variability would affect each chamber equally.	E1		4
(iv)	Differences are 0·4 –0·5 1·2 –0·3 0·7 2·0 –0·1 0·6 1·5 1·0	B1		
	Ranks of | *d* | 3 4 8 2 6 10 1 5 9 7	M1 A1		
	Test statistic is 4 + 2 + 1 = 7 (or 3 + 8 + 6 + 10 + 5 + 9 + 7 = 48)	M1 A1		
	Refer to paired Wilcoxon table with *n* = 10	M1	No ft from here if wrong.	
	Need lower 2½ % point which is 8 (or, if 48 used, upper 2½ % point which is 47).	A1	No ft from here if wrong.	
	Result is significant.	E1	ft only c's test statistic.	
	Seems underlying medians are not the same.	E1	ft only c's test statistic.	9
				27

Advanced Subsidiary GCE (MEI Statistics) (H132)
June 2006 Assessment Series

Unit Threshold Marks

Unit		Maximum Mark	a	b	c	d	e	u
G241	Raw	72	54	47	40	33	27	0
	UMS	100	80	70	60	50	40	0
G242	Raw	72	56	49	42	35	28	0
	UMS	100	80	70	60	50	40	0
G243	Raw	72	56	48	40	32	24	0
	UMS	100	80	70	60	50	40	0

Specification Aggregation Results

Overall threshold marks in UMS (i.e. after conversion of raw marks to uniform marks)

	Maximum Mark	A	B	C	D	E	U
H132	300	240	210	180	150	120	0

The cumulative percentage of candidates awarded each grade was as follows:

	A	B	C	D	E	U	Total Number of Candidates
H132	0	0	0	50.0	50.0	100	6

For a description of how UMS marks are calculated see;
www.ocr.org.uk/OCR/WebSite/docroot/understand/ums.jsp

Statistics are correct at the time of publication

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

OCR Information Bureau

(General Qualifications)
Telephone: 01223 553998
Facsimile: 01223 552627
Email: helpdesk@ocr.org.uk

www.ocr.org.uk

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