GCE

Statistics (MEI)

Advanced Subsidiary GCE **AS H132**

Mark Scheme for the Units

June 2007

H132/MS/R/07

Oxford Cambridge and RSA Examinations

Any enquiries about publications should be addressed to:

OCR Publications
PO Box 5050
Annesley
NOTTINGHAM
NG15 0DL

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

CONTENTS

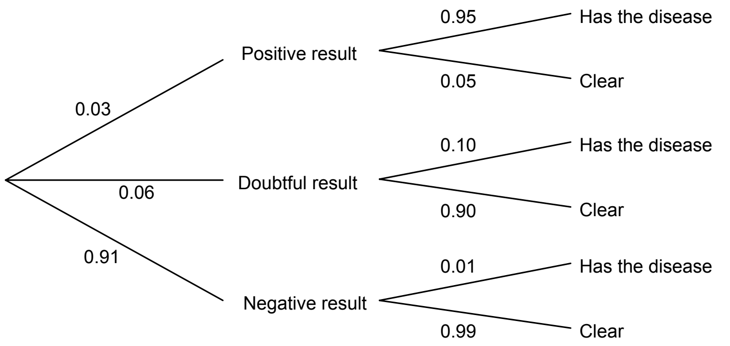
Advanced Subsidiary GCE Statistics (H132)

MARK SCHEMES FOR THE UNITS

Mark Scheme G241
June 2007

<table>
<tr><td>Q1 (i)</td><td>$\binom{8}{4}$ ways to select = 70</td><td>M1 for $\binom{8}{4}$
A1 CAO</td><td>2</td></tr>
<tr><td>(ii)</td><td>4! = 24</td><td>B1 CAO</td><td>1</td></tr>
<tr><td></td><td></td><td>TOTAL</td><td>3</td></tr>
<tr><td>Q2 (i)</td><td><table><tr><td>Amount</td><td>0- <20</td><td>20- <50</td><td>50- <100</td><td>100- <200</td></tr><tr><td>Frequency</td><td>800</td><td>480</td><td>400</td><td>200</td></tr></table></td><td>B1 for amounts
B1 for frequencies</td><td>2</td></tr>
<tr><td>(ii)</td><td>Total ≈
$10\times800+35\times480+75\times400+150\times200=£84800$</td><td>M1 for their midpoints × their frequencies
A1 CAO</td><td>2</td></tr>
<tr><td></td><td></td><td>TOTAL</td><td>4</td></tr>
<tr><td>Q3 (i)</td><td>Mean = $\frac{3026}{56}$ = 54.0
S_{xx} = $178890-\frac{3026^2}{56}$ = 15378
$s=\sqrt{\frac{15378}{55}}$ = 16.7</td><td>B1 for mean
M1 for attempt at S_{xx}
A1 CAO</td><td>3</td></tr>
<tr><td>(ii)</td><td>$\bar{x}$ + 2s = 54.0 + 2×16.7 = 87.4
So 93 is an outlier</td><td>M1 for their $\bar{x}$ + 2×their s
A1 FT for 87.4 and comment</td><td>2</td></tr>
<tr><td>(iii)</td><td>New mean = 1.2 ×54.0 – 10 = 54.8
New s = 1.2 ×16.7 = 20.1</td><td>B1 FT
M1A1 FT</td><td>3</td></tr>
<tr><td></td><td></td><td>TOTAL</td><td>8</td></tr>
<tr><td>Q4 (i)</td><td>(A) P(at least one) = $\frac{36}{50}=\frac{18}{25}$ = 0.72
(B) P(exactly one) = $\frac{9+6+5}{50}=\frac{20}{50}=\frac{2}{5}$ = 0.4</td><td>B1 aef
M1 for (9+6+5)/50
A1 aef</td><td>3</td></tr>
<tr><td>(ii)</td><td>P(not paper | aluminium) = $\frac{13}{24}$</td><td>M1 for denominator 24 or 24/50 or 0.48
A1 CAO</td><td>2</td></tr>
<tr><td>(iii)</td><td>P(one kitchen waste) = $2\times\frac{18}{50}\times\frac{32}{49}=\frac{576}{1225}=0.470$</td><td>M1 for both fractions
M1 for 2 × product of both, or sum of 2 pairs
A1</td><td>3</td></tr>
<tr><td></td><td></td><td>TOTAL</td><td>8</td></tr>
</table>

Q5 (i)	11^{th} value is 4, 12^{th} value is 4 so median is 4 Interquartile range = 5 – 2 = 3	B1 M1 for either quartile A1 CAO	**3**
(ii)	No, not valid any two valid reasons such as : • the sample is only for two years, which may not be representative • the data only refer to the local area, not the whole of Britain • even if decreasing it may have nothing to do with global warming • more days with rain does not imply more total rainfall • a five year timescale may not be enough to show a long term trend	B1 E1 E1	**3**
		TOTAL	**6**
Q6 (i)	Either P(all 4 correct) $= \frac{4}{7}\times\frac{3}{6}\times\frac{2}{5}\times\frac{1}{4} = \frac{1}{35}$ or P(all 4 correct) $= \frac{1}{{}^7C_4} = \frac{1}{35}$	M1 for fractions, or ${}^7C_4$ seen A1 **NB answer given**	**2**
(ii)	E(X) = $1\times\frac{4}{35}+2\times\frac{18}{35}+3\times\frac{12}{35}+4\times\frac{1}{35}=\frac{80}{35}=2\frac{2}{7}=2.29$ E(X^2) = $1\times\frac{4}{35}+4\times\frac{18}{35}+9\times\frac{12}{35}+16\times\frac{1}{35}=\frac{200}{35}=5.714$ Var(X) = $\frac{200}{35}-\left(\frac{80}{35}\right)^2=\frac{24}{49}$ = 0.490 (to 3 s.f.)	M1 for Σrp (at least 3 terms correct) A1 CAO M1 for Σx^2p (at least 3 terms correct) M1*dep* for – their E(X)² A1 FT their E(X) provided Var(X) > 0	**5**
		TOTAL	**7**

	Section B		
Q7 (i)		G1 probabilities of result G1 probabilities of disease G1 probabilities of clear G1 labels	**4**
(ii)	P(negative and clear) = 0.91×0.99 $= 0.9009$	M1 for their 0.91×0.99 A1 CAO	**2**
(iii)	P(has disease) $= 0.03 \times 0.95 + 0.06 \times 0.10 + 0.91 \times 0.01$ $= 0.0285 + 0.006 + 0.0091$ $= 0.0436$	M1 three products M1*dep* sum of three products A1 FT their tree	**3**
(iv)	P(negative | has disease) $= \dfrac{\text{P(negative } \textit{and} \text{ has disease)}}{\text{P(has disease)}} = \dfrac{0.0091}{0.0436} = 0.2087$	M1 for their 0.01×0.91 or 0.0091 on its own or as numerator M1 *indep* for their 0.0436 as denominator A1 FT their tree	**3**
(v)	Thus the test result is not very reliable. A relatively large proportion of people who have the disease will test negative.	E1 FT for idea of 'not reliable' or 'could be improved', etc E1 FT	**2**
(vi)	P(negative or doubtful and declared clear) $= 0.91 + 0.06 \times 0.10 \times 0.02 + 0.06 \times 0.90 \times 1$ $= 0.91 + 0.00012 + 0.054 = 0.96412$	M1 for their 0.91 + M1 for either triplet M1 for second triplet A1 CAO	**4**
		TOTAL	**18**

Q8 (i)	$X \sim B(17, 0.2)$ $P(X \geq 4) = 1 - P(X \leq 3)$ $= 1 - 0.5489 = 0.4511$	B1 for 0.5489 M1 for 1 – their 0.5489 A1 CAO	**3**
(ii)	$E(X) = np = 17 \times 0.2 = 3.4$	M1 for product A1 CAO	**2**
(iii)	$P(X = 2) = 0.3096 - 0.1182 = 0.1914$ $P(X = 3) = 0.5489 - 0.3096 = 0.2393$ $P(X = 4) = 0.7582 - 0.5489 = 0.2093$ So 3 applicants is most likely	B1 for 0.2393 B1 for 0.2093 A1 CAO *dep* on both B1s	**3**
(iv)	(*A*) Let *p* = probability of a randomly selected maths graduate applicant being successful (for population) H_0: $p = 0.2$ H_1: $p > 0.2$ (*B*) H_1 has this form as the suggestion is that mathematics graduates are <u>more</u> likely to be successful.	B1 for definition of *p* in context B1 for H_0 B1 for H_1 E1	**4**
(v)	Let $X \sim B(17, 0.2)$ $P(X \geq 6) = 1 - P(X \leq 5) = 1 - 0.8943 = 0.1057 > 5\%$ $P(X \geq 7) = 1 - P(X \leq 6) = 1 - 0.9623 = 0.0377 < 5\%$ So critical region is {7,8,9,10,11,12,13,14,15,16,17}	B1 for 0.1057 B1 for 0.0377 M1 for at least one comparison with 5% A1 CAO for critical region *dep* on M1 and at least one B1	**4**
(vi)	Because $P(X \geq 6) = 0.1057 > 10\%$ Either: comment that 6 is still outside the critical region Or comparison $P(X \geq 7) = 0.0377 < 10\%$	E1 E1	**2**
		TOTAL	**18**

Mark Scheme G242
June 2007

1(i)	P(X<30) = P($Z < \frac{30-26}{2.4}$) = P(Z<1.666..) = 0.952 (3sf)	M1 A1 A1	**3**
1(ii)	Φ^{-1} (0.99) = 2.326 $\frac{x-26}{2.4} = 2.326$ (Equation in positive z) x = 31.58 (CAO) must leave at 06 28 (dependent on M1)	B1 M1 A1 A1 FT	**4**
1(iii)	C.I. centred on 67.4 67.4 ± 1.96×2.45÷√30 (B1 for 1.96) (66.5, 68.3)	M1 B1 M1 A1 (FT sensible z)	**4**
1(iv)	The CI does not contain the scheduled journey time of 65 minutes. This suggests that the scheduled time is not accurate. The mean journey time could be greater than 65 minutes.	E1(65 not in) E1(suggests) E1(mean greater)	**3**
1(v)	Valid comment on sample size. Valid comment on randomness.	E1 E1	**2**
2(i)	H_0: population median = 280 H_1: population median > 280 Actual differences -89 -56 -2 20 1 -63 132 210 130 153 80 96 85 10 -102 Associated ranks 9 5 2 4 1 6 13 15 12 14 7 10 8 3 11 T^- = 9 + 5 + 2 + 6 + 11 = 33 T^+ = 4 + 1 + 13 + 15 + 12 + 14 + 7 + 10 + 8 + 3 = 87 ∴ T = 33 From tables – at the 5% level of significance in a one-tailed Wilcoxon signed rank test, the critical value of T is 30 33 > 30 ∴the result is not significant The evidence does not suggest an increase in the numbers of Manx Shearwaters flying past the observatory at this time of year.	B1 B1 M1 M1 A1 B1 B1 B1FT B1 M1 A1 E1	**12**
2(ii)	Variable - symmetry about median Sample - random	E1 E1	**2**

3(i)	mean = 45.9 standard deviation = 4.65	B1 B1	**2**
3(ii)	$H_0 : \mu = 50$ & $H_1 : \mu < 50$ Where μ represents the mean fungicide level in the underlying population. $t = \frac{45.9 - 50}{SD/\sqrt{10}}$ = -2.79 (3s.f.) 9 degrees of freedom At 5% level, critical value of *t* = 1.833 -2.79 < -1.833 so the result is significant Evidence suggests the mean level of fungicide in the crop is below the agreed safe level.	B1 E1 M1 A1 B1 B1 M1 A1 A1	**9**
3(iii)	The amount of fungicide in the underlying population follows a Normal distribution	E1(Normal) E1dep(pop$^{\underline{n}}$)	**2**
4(i)	Bacteria occur randomly with a uniform mean rate of occurrence	E1(random) E1(uniform mean rate)	**2**
4(ii)A	$\Sigma fx \div \Sigma f$ = 580÷200 = 2.9 (A.G.)	M1 A1	**2**
4(ii)B	Variance = 1.762^2 = 3.1046... mean ≈ variance so does not give reason to doubt suitability of the Poisson model	B1 E1dep	**2**
4(iii)	P(*X* = 0) = 0.055 P(*X* = 1) = 0.1596 P(*X* ≥ 8) 0.0099 11 (*X* = 0), 31.92 (*X* = 1), 1.98 (*X* ≥ 8)	B1 B1 B1 M1 (×200) A1	**5**
4(iv)	6 degrees of freedom (8 – 1 – 1) Critical value at 5% level is X^2 = 12.59 9.032 < 12.59 so not significant The Poisson model seems a good fit.	M1 (for 8-1-1) A1 M1 A1 (F.T. sensible c.v.)	**4**

5(i)	H_0: there is no association between sex and viewing preference H_1: there is an association between sex and viewing preference	B1	
	Expected frequencies (see table A below)	M1 A1	
	Contribution to X^2 (see table B below)	M1 M1	
	X^2 = 17.88715 5 degrees of freedom Critical value at 5% level is 11.07 As 17.88715 > 11.07 the result is significant There is evidence of an association between sex and viewing preference.	A1 B1 B1 M1A1 A1	**11**
5(ii)	The largest contributions to the test statistic come from news and sport programmes. The value 3.43310 shows a positive association between females and news programmes whereas the value 3.84042 shows a negative association between males and news programmes. With sport, the preferences are reversed. The smallest contributions to the test statistic were for film programmes showing similarities between male and female.	E1 E1 E1	**3**

Table A – Expected frequencies

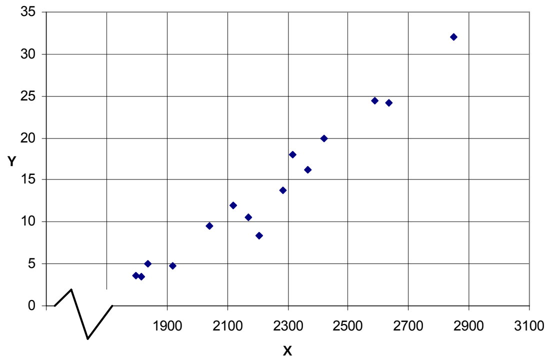
		Female	**Male**
Programme Type	**Film**	24.288	21.712
	Drama	19.008	16.992
	News	20.592	18.408
	Sport	33.264	29.736
	Music	**16.368**	**14.632**
	Wildlife	**18.48**	**16.52**

Table B – Contribution to X^2

		Female	**Male**
Programme Type	**Film**	0.12067	0.13499
	Drama	1.31103	1.46658
	News	3.43310	3.84042
	Sport	2.58002	2.88612
	Music	**0.34258**	**0.38323**
	Wildlife	**0.65532**	**0.73308**

Mark Scheme G243
June 2007

Q1 (i)	There may be differences in 'fertility' in different parts of the field – any such differences should affect the pairs equally.	**E2**	(2, 1, 0)	**2**
(ii)	Population of <u>differences</u> in ratings is Normally distributed. HH_0: $\mu_D = 0$ (or $\mu_A = \mu_B$ etc) H_1: $\mu_D \neq 0$ (or $\mu_A \neq \mu_B$ etc) Where μ_D is the population mean for differences. MUST be paired comparison t test. Use of differences. Differences are: 3.7, 1.4, –2.6, –0.6, 0.4, –1.0, 0.2, 2.1, –0.1, 0.7 $\bar{d} = 0.42$ $s_{n-1} = 1.736$ Test statistic is $\frac{0.42 - 0}{1.736/\sqrt{10}} = 0.765$ Refer to t_9 Double-tailed 10% critical value is 1.833 Not significant Seems mean ratings may be assumed equal. Wilcoxon signed rank test for paired samples.	**B1** **B1** **B1** **B1** **M1** **A1** **M1** **A1** **M1** **A1** **E1** **E1** **B1**	Do NOT allow $\bar{D}$ or similar unless it is clearly and explicitly stated to be a <u>population</u> mean. Hypotheses in words only must include "population". For adequate verbal definition. Allow absence of "population" here if correct notation *μ* has been used. For both. [s_n = 1.647 <u>NOT</u> allowed.] A1 is 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 M1 A0, but FT. Allow c's $\bar{d}$ and/or s_{n-1}. Allow alternatives: 0 + (c's 1·833) × $\frac{1.736}{\sqrt{10}}$ (= 1.006) for subsequent comparison with $\bar{d}$, <u>OR</u> $\bar{d}$ – (c's 1.833) × $\frac{1.736}{\sqrt{10}}$ (= $\bar{d}$ – 1.006) for subsequent comparison with 0. No FT if wrong. No FT if wrong.	**13**

<table>
<tr><td>Q2
(i)</td><td>
Looks strongly linear.</td><td>G1

G1

G1

B1</td><td>Axes, including labels. “x” and “y” suffice as they are defined in the question.

Clear “break” in the x-axis.

All points correct (allow one error).

Comments re bivariate Normality may be rewarded in part (iii).</td><td>4</td></tr>
<tr><td>(ii)</td><td>r = 0.9774

If used:
$\sum x = 33361$ $\sum x^2 = 75604869$ $S_{xx} = 1407780.9$
$\sum y = 205.9$ $\sum y^2 = 3870.63$ $S_{yy} = 1044.3093$
$\sum xy = 495410.1$ $S_{xy} = 37474.773$</td><td>M1
A1</td><td>Regard as implicit from correct answer.</td><td>2</td></tr>
<tr><td>(iii)</td><td>Bivariate Normality.
Yes – (long thin !) “cigar” shape.</td><td>M1
E1</td><td></td><td>2</td></tr>
<tr><td>(iv)</td><td>H_0 is ρ=0
where ρ is the correlation coefficient for the underlying bivariate population.
From tables, upper 1% point for n=15 is 0.5923
Significant.
Seems there is a positive correlation between traffic flow and air pollution.</td><td>B1

B1
A1

E1

E1</td><td>

No FT if wrong.</td><td>5</td></tr>
<tr><td>(v)</td><td>For any sensible comments, <u>such as</u>:
– correlation does not necessarily imply causation.
– there could be another confounding factor
– this might be a “false positive”.</td><td>E1,
E1</td><td></td><td>2</td></tr>
</table>

Q3 (i)	H_0: the medians of the two populations are the same. H_1: the medians of the two populations are different. [Or more formal statements.] Wilcoxon rank sum test (or Mann-Whitney form thereof). Ranks are: A 5 15 6 3 18 10 7 1 4 11 B 8 2 22 13 9 20 14 19 17 16 21 12 Rank sum for smaller sample is 80. Refer to (10, 12) table. Two-tail 5% critical value is 84 [or 29 for M-W]. Significant. Seems median heights are different.	**B1** **B1** **M1** **M1** **A1** **B1** **M1** **A1** **E1** **E1**	Allow 1 for "medians", but require "population" for second mark. Combined ranking. All ranks correct (FT if M1 earned). [Or M-W statistic = 1+6+1+1+8+3+1+0+1+3=25] No FT from here if wrong. No FT from here if wrong.	**10**
(ii)	A No, seems to need "pairing". B Not clear-cut, perhaps no strong reason for "pairing", but large-scale weather systems affecting whole country might be important.	**E1** **E1** **E1** **E1** **E1**	 Or other sensible comments.	**5**

Q4 (i)	Need to define the region geographically. Age and/or sex sub-classifications? May be problems of people who go out of/come into the region to buy goods. Frequency of purchase sub-classifications?	**E1** **E1** **E1** **E1**	Or other sensible comments	**4**
(ii)	Elect a (simple) random sample of schools/colleges. Select sample (or complete enumeration) in each selected school/college.	**E1** **E1**		**2**
(iii)	Comments to effect of administrative convenience, less cost.	**E1** **E1**		**2**
(iv)	Each cluster should be representative of entire population. The chosen clusters might all represent only a sub-population. Stratified sampling.	 **E1** **E1** **B1**		**3**
(v)	Pooled s^2 is $\frac{(6\times 9.28)+(9\times 12.16)}{15}=\frac{165.12}{15}=11.008$ Test statistic is $\frac{68.6-64.2\ (-0)}{\sqrt{11.008}\sqrt{\frac{1}{7}+\frac{1}{10}}}$ $=\frac{4.4}{1.635}=2.691$ Refer to t_{15}. Double-tailed 5% point is 2.131. Significant. Seems that population mean ratings are different. Must assume <u>population</u> variances are the same.	**M1** **A1** **M1** **M1** **M1** **A1** **M1** **A1** **E1** **E1** **B1**	For any reasonable attempt at pooling If correct For numerator For $\sqrt{11.008}$ (or candidate's value) For $\sqrt{\frac{1}{7}+\frac{1}{10}}$ FT from here if all M marks earned. Allow alternatives as in question 1. No FT if wrong No FT if wrong	**11**
(vi)	(From e.g. college roll) select one of the first 60 at random. and then every 60^{th}. Equivalent to simple random sampling, with some reason consistent within candidate's answer. Discussion	 **M1** **M1** **M1** **E2**	E0,1 or 2. Allow E1 for comment re each student has equal chance of selection. Allow E2 for comment re no cycles in list of names.	**5**

Advanced Subsidiary GCE (MEI Statistics) (H132)
June 2007 Assessment Session

Unit Threshold Marks

Unit		**Maximum Mark**	**a**	**b**	**c**	**d**	**e**	**u**
G241	Raw	72	55	48	41	35	29	0
	UMS	100	80	70	60	50	40	0
G242	Raw	72	58	50	43	36	29	0
	UMS	100	80	70	60	50	40	0
G243	Raw	72	58	50	43	36	29	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	14.3	42.9	71.4	85.7	100	100	7

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 Customer Contact Centre

(General Qualifications)
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

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