OCR
RECOGNISING ACHIEVEMENT

GCE

Mathematics (MEI)

Advanced GCE

Unit **4768:** Statistics 3

Mark Scheme for June 2011

OCR (Oxford Cambridge and RSA) is a leading UK awarding body, providing a wide range of qualifications to meet the needs of pupils of all ages and abilities. OCR qualifications include AS/A Levels, Diplomas, GCSEs, OCR Nationals, Functional Skills, Key Skills, Entry Level qualifications, NVQs and vocational qualifications in areas such as IT, business, languages, teaching/training, administration and secretarial skills.

It is also responsible for developing new specifications to meet national requirements and the needs of students and teachers. OCR is a not-for-profit organisation; any surplus made is invested back into the establishment to help towards the development of qualifications and support which keep pace with the changing needs of today's society.

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.

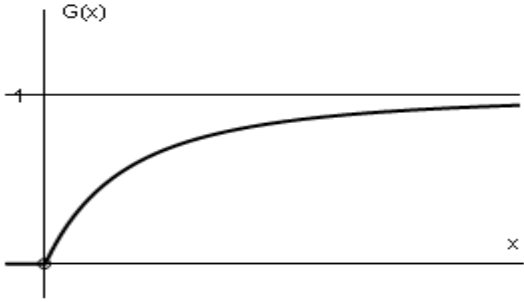| Q1				
(i)	t test might be used because • population variance is unknown • background population is Normal	E1 E1	Allow "sample is small" as an alternative.	**2**
(ii)	H$_0$: $\mu = 15.3$ H$_1$: $\mu < 15.3$	B1	Both hypotheses. Hypotheses in words only must include "population". Do NOT allow "$\overline{X} = ...$" or similar unless $\overline{X}$ is clearly and explicitly stated to be a <u>population</u> mean.	
	where μ is the mean of Gerry's times.	B1	For adequate verbal definition. Allow absence of "population" if correct notation μ is used.	
	$\overline{x} = 14.987$ $s_{n-1} = 0.4567(5)$	B1	$s_n = 0.4333$ but do <u>NOT</u> allow this here or in construction of test statistic, but FT from there.	
	Test statistic is $\dfrac{14.987 - 15.3}{\dfrac{0.45675}{\sqrt{10}}}$	M1	Allow c's $\overline{x}$ and/or s_{n-1}. Allow alternative: $15.3 + ($c's $-1.833)$ $\times \dfrac{0.45675}{\sqrt{10}}$ $(= 15.035)$ for subsequent comparison with $\overline{x}$. (Or $\overline{x} - ($c's $-1.833) \times \dfrac{0.45675}{\sqrt{10}}$ $(= 15.252)$ for comparison with 15.3.)	
	$= -2.167(0)$.	A1	c.a.o. but ft from here in any case if wrong. Use of $\mu - \overline{x}$ scores M1A0, but ft.	
	Refer to t_9. Single-tailed 5% point is -1.833.	M1 A1	No ft from here if wrong. Must be minus 1.833 unless absolute values are being compared. No ft from here if wrong. $P(t < -2.167(0)) = 0.0292$.	
	Significant. Seems that Gerry's times have been reduced on average.	A1 A1	ft only c's test statistic. ft only c's test statistic. Conclusion in context to include "average" o.e.	**9**
(iii)	A 5% significance level means that the probability of rejecting H$_0$ given that it is true is 0.05. Decreasing the significance level would make it less likely that a true H$_0$ would be rejected. Evidence for rejecting H$_0$ would need to be stronger.	E1 E1 E1	Or equivalent. Allow answers that relate to the context of the question.	**3**
(iv)	CI is given by $14.987 \pm$	M1	ZERO/4 if not same distribution as test. Same wrong distribution scores maximum M1B0M1A0. Recovery to t_9 is OK.	
	$2 \cdot 262$ $\times \dfrac{0.45675}{\sqrt{10}}$	B1 M1		
	$= 14.987 \pm 0.3267 = (14.66(0),\ 15.31(3))$	A1	c.a.o. Must be expressed as an interval.	**4**
				18

Q2

(i)

No. particles	0	1	2	3	4	5
Obs fr	4	7	10	20	17	
Prob'y	0.0150	0.0630	0.1322	0.1852	0.1944	
Exp fr	1.50	6.30	13.22	18.52	19.44	
Contrib to X^2	(4.1667)	(0.0778)	0.7843	0.1183	0.3063	

Combined	
	11
	7.80
	1.3128

$X^2 = 1.3128 + 0.7843 + 0.1183 + 0.3063 + 0.1083 +$
     $0.1813 + 0.6676 + 0.4056$
     $= 3.884(5)$

H_0: The Poisson model fits the data.
H_1: The Poisson model does not fit the data.

Refer to χ_6^2.

Upper 10% point is 10.64.

Not significant.
Evidence suggests that the model fits the data.

M1	Probs correct to 3d.p. or better.
M1	$\times 100$ for expected frequencies.
A1	All correct.
M1	Merge first 2 cells.
M1	Calculation of X^2.
A1	c.a.o. (For ungrouped cells $X^2 = 6.816$.)
B1	Ignore any reference to the parameter.
B1	Do not accept "data fit model" oe.
M1	Allow correct df (= cells – 2) from wrongly grouped table and ft. Otherwise, no ft if wrong.
A1	No ft from here if wrong. ($\chi_7^2 = 12.02$) $P(X^2 > 3.884) = 0.6924$.
A1	ft only c's test statistic.
A1	ft only c's test statistic. Do not accept "data fit model" oe.

12

(ii)

H_0: $m = 15$ H_1: $m > 15$
where m is the population median diameter (in μm).

Given $W_- = 53$ ($\therefore W_+ = 157$)

Refer to tables of Wilcoxon paired (/single sample) statistic for $n = 20$.
Lower 5% point is 60 (or upper is 150 if W_+ used).

Result is significant.
Evidence suggests that the median diameter appears to be more than 15 μm.

B1	Both. Accept hypotheses in words.
B1	Adequate definition of m to include "population".
M1	No ft from here if wrong.
A1	i.e. a 1-tail test. No ft from here if wrong.
A1	ft only c's test statistic.
A1	ft only c's test statistic. Conclusion in context to include "average" o.e.

6

18

Q3				
(i) (A)		M1 A1 A1	Increasing curve, through (0, 0), in first quadrant only. Asymptotic behaviour. Asymptote labelled; condone absence of axis labels.	3
(B)	For the UQ $G(u) = 0.75$ $\therefore \left(1 + \dfrac{u}{200}\right)^{-2} = \dfrac{1}{4}$ $\therefore u = 200$ For the LQ $G(l) = 0.25$ $\therefore \left(1 + \dfrac{l}{200}\right)^{-2} = \dfrac{3}{4}$ $\therefore l = 200\left(\dfrac{2}{\sqrt{3}} - 1\right) = 30.94...$ $\therefore \text{IQR} = 200 - 30.94 = 169(.06)$ For an outlier $x > \text{UQ} + 1.5 \times \text{IQR} = 200 + 1.5 \times 169$ $= 453(.58) \approx 454$ (nearest hour)	M1 A1 A1 M1 M1 E1	Use of $G(x)$ for either quartile. c.a.o. c.a.o. $\text{UQ} - \text{LQ}$ $\text{UQ} + 1.5 \times \text{IQR}$. Answer given; must be obtained genuinely.	6
(ii) (A)	$F(x) = \displaystyle\int_0^x \dfrac{1}{200} e^{\frac{-t}{200}} \, dt$ $= \left[-e^{\frac{-t}{200}} \right]_0^x = \left(-e^{\frac{-x}{200}} \right) - \left(-e^{\frac{-0}{200}} \right) = 1 - e^{\frac{-x}{200}}$	M1 A1 E1	Correct integral, including limits (which may be implied subsequently). Correctly integrated. Limits used. Answer given; must be shown convincingly. Condone the omission of $x < 0$ part. Allow use of "$+ c$" with $F(0) = 0$.	3
(B)	$P(X > 50) = 1 - F(50)$ $= e^{\frac{-50}{200}} = e^{-0.25}$	M1 E1	Use of $1 - F(x)$ Answer given: must be convincing. $(= 0.7788(0))$	2
(C)	$P(X > 400) = e^{\frac{-400}{200}} = 0.1353(35)$ $P(X > 450) = e^{\frac{-450}{200}} = 0.1053(99)$ $P(X > 450 \mid X > 400) = \dfrac{P(X > 450)}{P(X > 400)}$ $= \dfrac{e^{\frac{-450}{200}}}{e^{\frac{-400}{200}}} = e^{\frac{-50}{200}} = e^{-0.25} \; (= 0.7788)$	B1 B1 M1 A1	Accept any form. Accept any form. Conditional probability. Not $P(X > 50)$ $\times P(X > 400)$ unless <u>clearly</u> justified. Accept division of decimals, 3dp or better. Accept a.w.r.t. 0.778 or 0.779.	4
				18

Q4	$C \sim N(10,\ 0.4^2), \qquad D \sim N(35,\ 3.5^2)$ When a candidate's answers suggest that (s)he appears to have neglected to use the difference columns of the Normal distribution tables penalise the first occurrence only.			
(i)	$P(C < 9.5) = P\left(Z < \dfrac{9.5 - 10}{0.4} = -1.25 \right)$ $= 1 - 0{\cdot}8944 = 0.1056$	M1 A1 A1	For standardising. Award once, here or elsewhere. c.a.o.	**3**
(ii)	$D - S = D - (C_1 + C_2 + C_3 + C_4) \sim N(-5,$ $\sigma^2 = 3.5^2 + (0.4^2 + 0.4^2 + 0.4^2 + 0.4^2) = 12.89)$ Want $P(D > S) = P(D - S > 0)$ $= 1 - \Phi\left(\dfrac{0 - (-5)}{3.59} = 1.39(27) \right)$ $= 1 - 0.9182 = 0.0818$	B1 B1 M1 A1	Mean. Accept +5 for $S - D$. Variance. Accept sd (= 3.590…). Formulation of requirement. Accept $S - D < 0$. This mark could be awarded in (iii) if not earned here. c.a.o.	**4**
(iii)	$New\ (D - S) = (D \times 1.3) - (C_1 + \ldots + C_5) \sim N(-4.5,$ $\sigma^2 = (3.5^2 \times 1.3^2) + (0.4^2 + \ldots + 0.4^2) = 21.5025)$ Again want $P(D > S) = P(D - S > 0)$ $= 1 - \Phi\left(\dfrac{0 - (-4.5)}{4.637} = 0.9704 \right)$ $= 1 - 0.8341 = 0.1659$	B1 M1 A1 A1	Mean. Accept +4.5 for $S - D$. Correct use of $\times 1.3^2$ for variance. c.a.o. Accept sd (= 4.637…) Or $S - D < 0$. M1 for formulation in (ii) available here. c.a.o.	**4**
(iv)	CI is given by $9.73 \pm$ 1.96 $\times \dfrac{0.4}{\sqrt{12}}$ $= 9.73 \pm 0.2263 = (9.50(37),\ 9.95(63))$ Since 10 lies above this interval, it seems that the cheeses are underweight. In repeated sampling, 95% of all confidence intervals constructed in this way will contain the true mean.	M1 B1 M1 A1 E1 E1 E1	 1.96 seen. c.a.o. Must be expressed as an interval. Ft c's interval.	**7**
				18

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

OCR Customer Contact Centre

14 – 19 Qualifications (General)
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