OCR
RECOGNISING ACHIEVEMENT

GCE

Mathematics (MEI)

Advanced GCE **4768**

Statistics 3

Mark Scheme for June 2010

Oxford Cambridge and RSA Examinations

Any enquiries about publications should be addressed to:

OCR Publications
PO Box 5050
Annesley
NOTTINGHAM
NG15 0DL

Telephone: 0870 770 6622
Facsimile: 01223 552610
E-mail: publications@ocr.org.uk

Q1	D ~ N(2018, σ = 96)		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)	Systematic Sampling. It lacks any element of randomness. Choose a random starting point in the range 1 – 10.	B1 E1 E1	 May be implied by the next mark. Allow reasonable alternatives e.g. "the list may contain cycles." Beware proposals for a different sampling method.	[3]
(ii)	$\text{P}(D > 2100) = \text{P}\left(Z > \frac{2100 - 2018}{96} = 0.8542\right)$ $= 1 - 0.8034 = 0.1966$	M1 A1 A1	For standardising. Award once, here or elsewhere. c.a.o.	[3]
(iii)	$D_1 + D_2 + D_3 \sim \text{N}(6054,$ $\sigma^2 = 96^2 + 96^2 + 96^2 = 27648\ )$ $\text{P(this} < 6000) = \text{P}\left(Z < \frac{6000 - 6054}{166.277} = -0.3248\right)$ $= 1 - 0.6273 = 0.3727$ Must assume that the months are independent. This is unlikely to be realistic since e.g. consecutive months may not be independent.	B1 B1 A1 E1 E1	Mean. Variance. Accept sd (= 166.277). c.a.o. Reference to independence of months. Any sensible comment.	[5]
(iv)	Claim ~ N(2018 × 0.45 + 21200 × 0.10 = 3028.10, $96^2 \times 0.45^2 + 1100^2 \times 0.10^2 = 13966.24$ P(3000 < this < 3300) $= \text{P}\left(\frac{3000 - 3028.1}{118.18} < Z < \frac{3300 - 3028.1}{118.18}\right)$ $= \text{P}(-0 \cdot 2378 < Z < 2.3008)$ $= 0.9893 - (1 - 0.5940) = 0.5833$	M1 A1 M1 A1 M1 A1 A1	Mean. c.a.o. Variance. Accept sd (= 118.18). c.a.o. Formulation of requirement: a two-sided inequality. Ft c's parameters. c.a.o.	[7]
			Total	**[18]**

Q2				
(i)	A t test might be used because • sample is small • population variance is unknown Must assume background population is Normal.	 B1 B1 B1		[3]
(ii)	H$_0$: $\mu = 1.040$ H$_1$: $\mu \neq 1.040$	B1	Both hypotheses. Hypotheses in words only must include "population". Do NOT allow " $\overline{X} = \ldots$ " or similar unless $\overline{X}$ is clearly and explicitly stated to be a population mean.	
	where μ is the mean specific gravity of the mixture.	B1	For adequate verbal definition. Allow absence of "population" if correct notation μ is used.	
	$\bar{x} = 1.0452$ $\quad s_{n-1} = 0.007155$	B1	$s_n = 0.006746$ but do NOT allow this here or in construction of test statistic, but FT from there.	
	Test statistic is $\dfrac{1.0452 - 1.040}{\frac{0.007155}{\sqrt{9}}}$	M1	Allow c's $\bar{x}$ and/or s_{n-1}. Allow alternative: 1.040 + (c's 1.860) × $\frac{0.007155}{\sqrt{9}}$ (= 1.0444) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ – (c's 860) × $\frac{0.007155}{\sqrt{9}}$ (= 1.0407) for comparison with 1.040.)	
	= 2.189(60).	A1	c.a.o. but ft from here in any case if wrong. Use of 1.040 – $\bar{x}$ scores M1A0, but ft.	
	Refer to t_8.	M1	No ft from here if wrong. P(t > 2.1896) = 0.05996.	
	Double-tailed 10% point is 1.860.	A1	No ft from here if wrong.	
	Significant.	A1	ft only c's test statistic.	
	Seems mean specific gravity in the mixture does not meet the requirement.	A1	ft only c's test statistic.	[9]
(iii)	CI is given by 1.0452 ± 2·306 × $\frac{0.007155}{\sqrt{9}}$	 M1 B1 M1		
	= 1.0452 ± 0.0055= (1.039(7), 1.050(7))	A1	c.a.o. Must be expressed as an interval. ZERO/4 if not same distribution as test. Same wrong distribution scores maximum M1B0M1A0. Recovery to t_8 is OK.	
	In repeated sampling, 95% of confidence intervals constructed in this way will contain the true population mean.	E2	E2, 1, 0.	[6]
			Total	**[18]**

Q3

(a)(i) Use paired data in order to eliminate differences between authorities. B1 **[1]**

(ii) $H_0: m = 0 \qquad H_1: m > 0$
where m is the population median difference.

B1 Both. Accept hypotheses in words.
B1 Adequate definition of m to include "population".

Diff (After − Before)	6	−1	5	−4	−3	11	8	2	9
Rank of |diff|	6	1	5	4	3	9	7	2	8

M1 For differences. ZERO in this section if differences not used.
M1 For ranks.
A1 FT from here if ranks wrong

$W_- = 1 + 3 + 4 = 8$ (or $= 2+5+6+7+8+9 = 37$) B1

Refer to tables of Wilcoxon paired (/single sample) statistic for $n = 9$. M1 No ft from here if wrong.
Lower 5% point is 8 (or upper is 37 if W_+ used). A1 i.e. a 1-tail test. No ft from here if wrong.
Result is significant. A1 ft only c's test statistic.
Evidence suggests the percentage has been raised (on the whole). A1 ft only c's test statistic. **[10]**

(b) H_0: Stock market prices can be modelled by Benford's Law.
H_1: Stock market prices can not be modelled by Benford's Law.

Prob	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046
Exp f	60.2	35.2	25.0	19.4	15.8	13.4	11.6	10.2	9.2
Obs f	55	34	27	16	15	17	12	15	9

M1 Probs × 200 for expected frequencies. All correct.

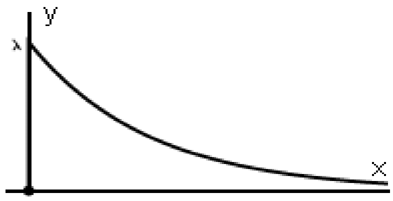
$X^2 = 0.44917 + 0.04091 + 0.16 + 0.59588 + 0.04051 + 0.96716 + 0.01379 + 2.25882 + 0.00435 = 4.5305(9)$

M1 Calculation of X^2.
A1 c.a.o.

Refer to χ^2_8. M1 Allow correct df (= cells – 1) from wrongly grouped table and ft. Otherwise, no ft if wrong.
$P(X^2 > 4.53059) = 0.80636$.

Upper 5% point is 13.36. A1 No ft from here if wrong.
Not significant. A1 ft only c's test statistic.
Suggests Benford's Law provides a reasonable model in the context of share prices. A1 ft only c's test statistic. **[7]**

Total [18]

Q4	$f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$, where $\lambda > 0$.	Given $\int_0^{\infty} x^r e^{-\lambda x} dx = \frac{r!}{\lambda^{r+1}}$	
(i)	$\int_0^{\infty} f(x)dx = \int_0^{\infty} \lambda e^{-\lambda x} dx$ $= \left[-e^{-\lambda x}\right]_0^{\infty}$ $= \left(0 - (-e^0)\right) = 1$ 	M1 Integration of $f(x)$. M1 Use of limits or the given result. A1 Convincingly obtained (Answer given.) G1 Curve, with negative gradient, in the first quadrant only. Must intersect the y-axis. G1 $(0, \lambda)$ labelled; asymptotic to x-axis.	[5]
(ii)	$E(X) = \int_0^{\infty} \lambda x e^{-\lambda x} dx$ $= \lambda \frac{1}{\lambda^2} = \frac{1}{\lambda}$ $E(X^2) = \int_0^{\infty} \lambda x^2 e^{-\lambda x} dx$ $= \lambda \frac{2}{\lambda^3} = \frac{2}{\lambda^2}$ $\text{Var}(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$	M1 Correct integral. A1 c.a.o. (using given result) M1 Correct integral. A1 c.a.o. (using given result) M1 Use of $E(X^2) - E(X)^2$ A1	[6]
(iii)	$\mu = 6 \quad \therefore \lambda = \frac{1}{6}$ $\bar{X} \sim$ (approx) $N\left(6, \frac{6^2}{50}\right)$	B1 Obtained λ from the mean. B1 Normal. B1 Mean. ft c's λ. B1 Variance. ft c's λ.	[4]
(iv)	**<u>EITHER</u>** can argue that 7.8 is more than 2 SDs from μ. ($6 + 2\sqrt{0.72}$ = 7.697; <u>must</u> refer to SD $(\bar{X})$, not SD(X)) i.e. outlier. $\Rightarrow$ doubt. **<u>OR</u>** formal significance test: $\frac{7.8 - 6}{\sqrt{0.72}}$ = 2.121, refer to N(0,1), sig at (eg) 5% $\Rightarrow$ doubt.	M1 A 95% C.I would be (6.1369, 9.4631). M1 A1 M1 M1 Depends on first M, but could imply it. $P(\|Z\| > 2.121)$= 0.0339 A1	[3]
		Total	**[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