

**Mathematics (MEI)**

Advanced GCE 4768

Statistics 3

**Mark Scheme for June 2010**

---

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.

© OCR 2010

Any enquiries about publications should be addressed to:

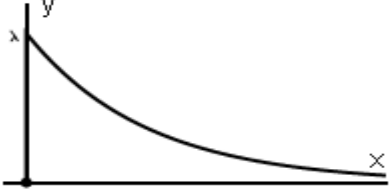
OCR Publications  
PO Box 5050  
Annesley  
NOTTINGHAM  
NG15 0DL

Telephone: 0870 770 6622  
Facsimile: 01223 552610  
E-mail: [publications@ocr.org.uk](mailto:publications@ocr.org.uk)

Q1	$D \sim N(2018, \sigma = 96)$		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)	$P(D > 2100) = P\left(Z > \frac{2100 - 2018}{96} = 0.8542\right)$ $= 1 - 0.8034 = 0.1966$	M1 A1  A1	For standardising. Award once, here or elsewhere.  c.a.o.	[3]
(iii)	$D_1 + D_2 + D_3 \sim N(6054,$ $\sigma^2 = 96^2 + 96^2 + 96^2 = 27648)$ $P(\text{this} < 6000) = P\left(Z < \frac{6000 - 6054}{166.277} = -0.3248\right)$ $= 1 - 0.6273 = 0.3727$  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 $\sim N(2018 \times 0.45 + 21200 \times 0.10 = 3028.10,$ $96^2 \times 0.45^2 + 1100^2 \times 0.10^2 = 13966.24$  $P(3000 < \text{this} < 3300)$ $= P\left(\frac{3000 - 3028.1}{118.18} < Z < \frac{3300 - 3028.1}{118.18}\right)$ $= P(-0.2378 < Z < 2.3008)$ $= 0.9893 - (1 - 0.5940) = 0.5833$	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]
			<b>Total</b>	<b>[18]</b>

Q2			
(i)	<p>A <i>t</i> test might be used because</p> <ul style="list-style-type: none"> <li>• sample is small</li> <li>• population variance is unknown</li> </ul> <p>Must assume background population is Normal.</p>	<p>B1 B1 B1</p>	[3]
(ii)	<p><math>H_0: \mu = 1.040</math> <math>H_1: \mu \neq 1.040</math></p> <p>where <math>\mu</math> is the mean specific gravity of the mixture.</p> <p><math>\bar{x} = 1.0452</math>      <math>s_{n-1} = 0.007155</math></p> <p>Test statistic is <math>\frac{1.0452 - 1.040}{\frac{0.007155}{\sqrt{9}}}</math></p> <p style="text-align: center;">= 2.189(60).</p> <p>Refer to <math>t_8</math>.</p> <p>Double-tailed 10% point is 1.860. Significant. Seems mean specific gravity in the mixture does not meet the requirement.</p>	<p>B1 Both hypotheses. Hypotheses in words only must include “population”. Do NOT allow “<math>\bar{X} = \dots</math>” or similar unless <math>\bar{X}</math> is clearly and explicitly stated to be a <u>population</u> mean.</p> <p>B1 For adequate verbal definition. Allow absence of “population” if correct notation <math>\mu</math> is used.</p> <p>B1 <math>s_n = 0.006746</math> but do <u>NOT</u> allow this here or in construction of test statistic, but FT from there.</p> <p>M1 Allow c’s <math>\bar{x}</math> and/or <math>s_{n-1}</math>. Allow alternative: <math>1.040 + (c’s\ 1.860) \times \frac{0.007155}{\sqrt{9}}</math> (= 1.0444) for subsequent comparison with <math>\bar{x}</math>. (Or <math>\bar{x} - (c’s\ 860) \times \frac{0.007155}{\sqrt{9}}</math> (= 1.0407) for comparison with 1.040.)</p> <p>A1 c.a.o. but ft from here in any case if wrong. Use of <math>1.040 - \bar{x}</math> scores M1A0, but ft.</p> <p>M1 No ft from here if wrong. <math>P(t &gt; 2.1896) = 0.05996</math>.</p> <p>A1 No ft from here if wrong.</p> <p>A1 ft only c’s test statistic.</p> <p>A1 ft only c’s test statistic.</p>	[9]
(iii)	<p>CI is given by</p> $1.0452 \pm 2.306 \times \frac{0.007155}{\sqrt{9}}$ <p>= 1.0452 <math>\pm</math> 0.0055 = (1.039(7), 1.050(7))</p> <p>In repeated sampling, 95% of confidence intervals constructed in this way will contain the true population mean.</p>	<p>M1 B1</p> <p>M1</p> <p>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 <math>t_8</math> is OK.</p> <p>E2 E2, 1, 0.</p>	[6]
<b>Total</b>			<b>[18]</b>

Q3																																	
(a) (i)	Use paired data in order to eliminate differences between authorities.	B1	[1]																														
(ii)	<p><math>H_0: m = 0</math>    <math>H_1: m &gt; 0</math> where <math>m</math> is the population median difference.</p> <table border="1"> <tr> <td>Diff (After – Before)</td> <td>6</td> <td>-1</td> <td>5</td> <td>-4</td> <td>-3</td> <td>11</td> <td>8</td> <td>2</td> <td>9</td> </tr> <tr> <td>Rank of  diff </td> <td>6</td> <td>1</td> <td>5</td> <td>4</td> <td>3</td> <td>9</td> <td>7</td> <td>2</td> <td>8</td> </tr> </table> <p><math>W_- = 1 + 3 + 4 = 8</math> (or <math>= 2 + 5 + 6 + 7 + 8 + 9 = 37</math>)</p> <p>Refer to tables of Wilcoxon paired (/single sample) statistic for <math>n = 9</math>. Lower 5% point is 8 (or upper is 37 if <math>W_+</math> used). Result is significant. Evidence suggests the percentage has been raised (on the whole).</p>	Diff (After – Before)	6	-1	5	-4	-3	11	8	2	9	Rank of  diff	6	1	5	4	3	9	7	2	8	<p>B1 Both. Accept hypotheses in words. B1 Adequate definition of <math>m</math> to include “population”.</p> <p>M1 For differences. ZERO in this section if differences not used. M1 For ranks. A1 FT from here if ranks wrong B1</p> <p>M1 No ft from here if wrong.</p> <p>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.</p>	[10]										
Diff (After – Before)	6	-1	5	-4	-3	11	8	2	9																								
Rank of  diff	6	1	5	4	3	9	7	2	8																								
(b)	<p><math>H_0</math>: Stock market prices can be modelled by Benford’s Law. <math>H_1</math>: Stock market prices can not be modelled by Benford’s Law.</p> <table border="1"> <tr> <td>Prob</td> <td>0.301</td> <td>0.176</td> <td>0.125</td> <td>0.097</td> <td>0.079</td> <td>0.067</td> <td>0.058</td> <td>0.051</td> <td>0.046</td> </tr> <tr> <td>Exp f</td> <td>60.2</td> <td>35.2</td> <td>25.0</td> <td>19.4</td> <td>15.8</td> <td>13.4</td> <td>11.6</td> <td>10.2</td> <td>9.2</td> </tr> <tr> <td>Obs f</td> <td>55</td> <td>34</td> <td>27</td> <td>16</td> <td>15</td> <td>17</td> <td>12</td> <td>15</td> <td>9</td> </tr> </table> <p><math>\chi^2 = 0.44917 + 0.04091 + 0.16 + 0.59588 + 0.04051 + 0.96716 + 0.01379 + 2.25882 + 0.00435 = 4.5305(9)</math></p> <p>Refer to <math>\chi^2_8</math>.</p> <p>Upper 5% point is 13.36. Not significant. Suggests Benford’s Law provides a reasonable model in the context of share prices.</p>	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	<p>M1 Probs <math>\times</math> 200 for expected frequencies. All correct. M1 Calculation of <math>\chi^2</math>. A1 c.a.o.</p> <p>M1 Allow correct df (= cells – 1) from wrongly grouped table and ft. Otherwise, no ft if wrong. <math>P(\chi^2 &gt; 4.53059) = 0.80636</math>. A1 No ft from here if wrong. A1 ft only c’s test statistic. A1 ft only c’s test statistic.</p>	[7]
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																								
		<b>Total</b>	<b>[18]</b>																														

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)	$\int_0^{\infty} f(x) dx = \int_0^{\infty} \lambda e^{-\lambda x} dx$ $= \left[ -e^{-\lambda x} \right]_0^{\infty}$ $= (0 - (-e^0)) = 1$ 	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)	$\mu = 6 \quad \therefore \lambda = \frac{1}{6}$ $\bar{X} \sim (\text{approx}) N\left(6, \frac{6^2}{50}\right)$	B1 Obtained $\lambda$ from the mean. B1 Normal. B1 Mean. ft c's $\lambda$ . B1 Variance. ft c's $\lambda$ .	[4]
(iv)	<p><b>EITHER</b> can argue that 7.8 is more than 2 SDs from <math>\mu</math>.  <math>(6 + 2\sqrt{0.72} = 7.697;</math>  <u>must</u> refer to SD (<math>\bar{X}</math>), not SD(<math>X</math>))                  i.e. outlier.  <math>\Rightarrow</math> doubt.</p> <p><b>OR</b> formal significance test:  <math>\frac{7.8 - 6}{\sqrt{0.72}} = 2.121</math>, refer to <math>N(0,1)</math>, sig at (eg) 5%  <math>\Rightarrow</math> doubt.</p>	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]
<b>Total</b>			[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](mailto:general.qualifications@ocr.org.uk)

**[www.ocr.org.uk](http://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