Q1	$F(t)=1-e^{-t/3}$ (t>0)			
(i)	For median m, $\frac{1}{2}=1-e^{-m/3}$ $\therefore\ e^{-m/3}=\frac{1}{2}\Rightarrow -\frac{m}{3}=\ln\frac{1}{2}=-0.6931$ $\Rightarrow m=2.079$ For 90th percentile p, $0.9=1-e^{-p/3}$ $\therefore e^{-p/3}=0.1\Rightarrow -\frac{p}{3}=\ln 0.1=-2.3026$ $\Rightarrow\ p=6.908$	M1 M1 A1 M1 A1	attempt to solve, here or for 90th percentile. Depends on previous M mark.	5
(ii)	$f(t)=\frac{d}{dt}F(t)$ $=\frac{1}{3}e^{-t/3}$ $\mu=\int_0^\infty \frac{1}{3}te^{-t/3}dt$ $=\frac{1}{3}\left\{\left[\frac{te^{-t/3}}{-1/3}\right]_0^\infty+3\int_0^\infty e^{-t/3}dt\right\}$ $=[0-0]+\left[\frac{e^{-t/3}}{-1/3}\right]_0^\infty=3$	M1 A1 M1 M1 A1	(for t>0, but condone absence of this) Quoting standard result gets 0/3 for the mean. attempt to integrate by parts	5
(iii)	$P(T>\mu)=$ [from cdf] $e^{-\mu/3}=e^{-1}$ =0.3679	M1 A1	[or via pdf] ft c's mean (>0)	2
(iv)	$\bar{T}$ ~ (approx) $N\left(3,\frac{9}{30}=0.3\right)$	B1 B1 B1	N ft c's mean (>0) 0.3	3
(v)	**EITHER** can argue that 4.2 is more than 2 SDs from μ ($3+2\sqrt{0.3}$ = 4.095; must refer to SD ($\bar{\text{T}}$), not SD(T)) i.e. outlier ⇒ doubt **OR** formal significance test: $\frac{4.2-3}{3/\sqrt{30}}=2.191$, refer to N(0,1), sig at (eg) 5% ⇒ doubt	M1 M1 A1 M1 M1 A1	Depends on first M, but could imply it.	3 18

Q2	$X \sim \mathrm{N}(180,\ \sigma = 12)$		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 (X <170) = P $(Z < \frac{170-180}{12} = -0.8333)$ = 1–0.7976 = 0.2024	M A1 A1	For standardising. Award once, here or elsewhere.	3
(ii)	$X_1 + X_2 + X_3 + X_4 + X_5 \sim \mathrm{N}(900, \sigma^2 = 720[\sigma = 26.8328]$ $\mathrm{P}\,(\text{this} < 840) = \mathrm{P}(Z < \frac{840-900}{26.8328} = -2.236)$ $= 1 - 0.9873 = 0.0127$	B1 B1 A1	Mean. Variance. Accept sd. c.a.o.	3
(iii)	$Y \sim \mathrm{N}(50, \sigma = 6)$ $X + Y \sim \mathrm{N}(230, \sigma^2 = 180[\sigma = 13.4164])$ $\mathrm{P}\,(\text{this} > 240) = \mathrm{P}(Z > \frac{240-230}{13.4164} = 0.7454)$ $= 1 - 0.7720 = 0.2280$	B1 B1 A1	Mean. Variance. Accept sd. c.a.o.	3
(iv)	$\frac{1}{4}X \sim N\left(45, \sigma^2 = \frac{1}{16} \times 144 = 9[\sigma = 3]\right)$ Require t such that $0.9 = \mathrm{P}(\text{this} < t) = \mathrm{P}\left(Z < \frac{t-45}{3}\right) = \mathrm{P}(Z < 1.282)$ $\therefore t - 45 = 3 \times 1.282 \Rightarrow t = 48.85\ (48.846)$	B1 M B1 A1	Variance. Accept sd. FT incorrect mean. Formulation of requirement. 1.282 ft only for incorrect mean	4
(v)	I = 45 + T where $T \sim \mathrm{N}\ (120,\ \sigma = 10)$ $\therefore I \sim \mathrm{N}(165, \sigma = 10)$ $\mathrm{P}(I < 150) = \mathrm{P}(Z < \frac{150-165}{10} = -1.5)$ $= 1 - 0.9332 = 0.0668$	B1 A1	for unchanged σ (candidates might work with P (T<105)) c.a.o.	2
(vi)	$J = 30 + \frac{3}{5}T$ where $T \sim \mathrm{N}\ (120,\ \sigma = 10)$		Cands might work with $\mathrm{P}(\frac{3}{5}T < 75)$. $\frac{3}{5}T \sim N(72,36)$	

	$\therefore J \sim \mathrm{N}\left(102, \sigma^2 = \frac{9}{25} \times 100 = 36\left[\sigma = 6\right]\right)$	B1 B1	Mean. Variance. Accept sd.	
	$\mathrm{P}(J < 105) = \mathrm{P}(Z < \frac{105 - 102}{6} = 0.5) = 0.6915$	A1	c.a.o.	3
				18

Q3				
(a)	H$_0$: $\mu_D = 0$ (or $\mu_A = \mu_B$)	B1	Hypotheses in words only must include "population".	
	H$_1$: $\mu_D > 0$ (or $\mu_B > \mu_A$)	B1	Or "<" for $A – B$.	
	where μ_D is "mean for B – mean for A"	B1	For adequate verbal definition. Allow absence of "population" if correct notation μ is used, but do NOT allow "$\bar{X}_A = \bar{X}_B$" or similar unless $\bar{X}$ is clearly and explicitly stated to be a population mean.	
	Normality of differences is required MUST be PAIRED COMPARISON t test.	B1		
	Differences are: 2.1 1.0 0.8 0.6 0.4 -1.0 -0.3 0.8 0.9 1.1			
	$\bar{d} = 0.64$ $s_{n-1} = 0.8316$	B1	s_n = 0.7889 but do NOT allow this here or in construction of test statistic, but FT from there.	
	Test statistic is $\dfrac{0.64 - 0}{\frac{0.8316}{\sqrt{10}}}$	M	Allow c's $\bar{d}$ and/or s_{n-1}. Allow alternative: 0 + (c's 1.833) × $\frac{0.8316}{\sqrt{10}}$ (= 0.4821) for subsequent comparison with $\bar{d}$. (Or $\bar{d}$ – (c's 1.833) × $\frac{0.8316}{\sqrt{10}}$ (= 0.1579) for comparison with 0.)	
	=2.43(37).	A1	c.a.o. but ft from here in any case if wrong. Use of 0 – $\bar{d}$ scores M1A0, but ft.	
	Refer to t_9.	M	No ft from here if wrong.	
	Single-tailed 5% point is 1.833.	A1	No ft from here if wrong.	
	Significant.	E1	ft only c's test statistic.	
	Seems mean amount delivered by B is greater that that by A	E1	ft only c's test statistic. Special case: (t_{10} and 1.812) can score 1 of these last 2 marks if either form of conclusion is given.	11
(b)	We now require Normality for the amounts delivered by machine A.	B1		

	For machine A, $\bar{x} = 250.19 \quad s_{n-1} = 3.8527$	B1	s_n = 3.6549(83) but do NOT allow this here or in construction of CI.	
	CI is given by $250.19 \pm 2.262 \dfrac{3.8527}{\sqrt{10}}$	M B1 M	ft c's $\bar{x}$ ±. 2.262 ft c's s_{n-1}.	
	$= 250.19 \pm 2.75(6) = (247.43(4),\ 252.94(6))$	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_9 is OK.	
	250 is in the CI, so would accept $H_0 : \mu = 250$, so no evidence that machine is not working correctly in this respect.	E1		7
				18

Q4				
(i)				

	$\underbrace{1 \qquad 30}_{31}$	62	70	$\underbrace{34 \qquad 3}_{37}$
e_i	$\underbrace{1.49 \qquad 37.85}_{39.34}$	55.62	58.32	$\underbrace{44.62 \qquad 2.10}_{46.72}$

		M	for grouping	
	$X^2 = 1.7681 + 0.7318 + 2.3392 + 2.0222$	M	Allow the M1 for correct method from wrongly grouped or ungrouped table.	
	$= 6.86$	A1		
	Refer to χ_1^2.	M	Allow correct df (= cells – 3) from wrongly grouped or ungrouped table, and FT. Otherwise, no FT if wrong.	
	Upper 5% point is 3.84	A1	No ft from here if wrong.	
	Significant	E1	ft only c's test statistic.	
	Suggests Normal model does not fit	E1	ft only c's test statistic.	7
(ii) (A)	*t* test unwise …	E1		
	… because underlying population appears non-Normal	E1	FT from result of candidate's work in (i)	2

<table>
<tr>
<td>(B)</td>
<td>
<table>
<tr><th>Data</th><th>Median 301</th><th>Difference</th><th>Rank of |diff|</th></tr>
<tr><td>301.3</td><td></td><td>0.3</td><td>3</td></tr>
<tr><td>301.4</td><td></td><td>0.4</td><td>4</td></tr>
<tr><td>299.6</td><td></td><td>- 1.4</td><td>8</td></tr>
<tr><td>302.2</td><td></td><td>1.2</td><td>7</td></tr>
<tr><td>300.3</td><td></td><td>- 0.7</td><td>5</td></tr>
<tr><td>303.2</td><td></td><td>2.2</td><td>10</td></tr>
<tr><td>302.6</td><td></td><td>1.6</td><td>9</td></tr>
<tr><td>301.8</td><td></td><td>0.8</td><td>6</td></tr>
<tr><td>300.9</td><td></td><td>- 0.1</td><td>1</td></tr>
<tr><td>300.8</td><td></td><td>- 0.2</td><td>2</td></tr>
</table>

T = 1 +2 + 5 + 8 = 16 (or 3+4+6+7+9+10 = 39)
Refer to tables of Wilcoxon single sample (/paired) statistic
Lower (or upper if 39 used) 5% tail is needed
Value for n = 10 is 10 (or 45 if 39 used)
Result is not significant
No evidence against median being 301
</td>
<td>M
M
A1
B1
M
M
A1
E1
E1</td>
<td>for differences.
ZERO in this section if differences not used.
for ranks.
FT if ranks wrong.</td>
<td>9</td>
</tr>
<tr><td></td><td></td><td></td><td></td><td>18</td></tr>
</table>

<table>
<tr><td>Q1</td><td>$\mathrm{f}(x) = 12x^3 - 24x^2 + 12x, \qquad 0 \le x \le 1$</td><td></td><td></td><td></td></tr>
<tr><td>(i)</td><td>$\mathrm{E}(X) = \int_0^1 x\mathrm{f}(x)\mathrm{d}x$</td><td>M1</td><td>Integral for E(X) including limits (which may appear later).</td><td></td></tr>
<tr><td></td><td>$= 12\left[\frac{x^5}{5} - 2\frac{x^4}{4} + \frac{x^3}{3}\right]_0^1$</td><td>A1</td><td>Successfully integrated.</td><td></td></tr>
<tr><td></td><td>$= 12\left[\frac{1}{5} - \frac{2}{4} + \frac{1}{3}\right] = 12 \times \frac{1}{30} = \frac{2}{5}$</td><td>A1</td><td>Correct use of limits leading to final answer. C.a.o.</td><td></td></tr>
<tr><td></td><td>For mode, $\mathrm{f}'(x) = 0$</td><td>M1</td><td></td><td></td></tr>
<tr><td></td><td>$\mathrm{f}'(x) = 12(3x^2 - 4x + 1) = 12(3x - 1)(x - 1)$
$\therefore \mathrm{f}'(x) = 0$ for $x = 1$ and $x = \frac{1}{3}$</td><td>A1</td><td></td><td></td></tr>
<tr><td></td><td>Any convincing argument (e.g. $\mathrm{f}''(x)$) that $\frac{1}{3}$ (and not 1) is the mode.</td><td>A1</td><td></td><td>6</td></tr>
<tr><td>(ii)</td><td>Cdf $\mathrm{F}(x) = \int_0^x \mathrm{f}(t)\mathrm{d}t$
$= 12\left(\frac{x^4}{4} - 2\frac{x^3}{3} + \frac{x^2}{2}\right)$</td><td>M1</td><td>Definition of cdf, including limits (or use of “+c” and attempt to evaluate it), possibly implied later. Some valid method must be seen.</td><td></td></tr>
<tr><td></td><td>$= 3x^4 - 8x^3 + 6x^2$</td><td>A1</td><td>Or equivalent expression; condone absence of domain [0,1].</td><td></td></tr>
<tr><td></td><td>$\mathrm{F}\left(\frac{1}{4}\right) = \frac{3}{256} - \frac{8}{64} + \frac{6}{16} = \frac{3-32+96}{256} = \frac{67}{256}$
$\mathrm{F}\left(\frac{1}{2}\right) = \frac{3}{16} - \frac{8}{8} + \frac{6}{4} = \frac{3-16+24}{16} = \frac{11}{16}$
$\mathrm{F}\left(\frac{3}{4}\right) = \frac{3\times81}{256} - \frac{8\times27}{64} + \frac{6\times9}{16} = \frac{243}{256}$</td><td>B1</td><td>For all three; answers given; must show convincing working (such as common denominator)! Use of decimals is not acceptable.</td><td>3</td></tr>
<tr><td>(iii)</td><td><table><tr><td>o_i</td><td>126</td><td>209</td><td>131</td><td>46</td></tr><tr><td>e_i</td><td>134</td><td>352 – 134 = 218</td><td>486 – 352 = 134</td><td>26</td></tr></table></td><td>B2</td><td>For e_i.
B1 if any 2 correct, provided Σ = 512.</td><td></td></tr>
<tr><td></td><td>$X^2 = 0{\cdot}4776 + 0{\cdot}3716 + 0{\cdot}0672 + 15{\cdot}3846 = 16{\cdot}30(1)$</td><td>M1
A1</td><td></td><td></td></tr>
<tr><td></td><td>Refer to χ^2_3.</td><td>M1</td><td>Must be some clear evidence of reference to χ^2_3, probably implicit by reference to a critical point (5% : 7·815; 1% : 11·34). No ft (to the A marks) if incorrect χ^2 used, but E marks are still available.</td><td></td></tr>
<tr><td></td><td>Very highly significant.
Very strong evidence that the model does not fit.</td><td>A1
A1</td><td>There must be at least one reference to “very …”, i.e. the extremeness of the test statistic.</td><td></td></tr>
<tr><td></td><td>The main feature is that we observe many</td><td></td><td>Or e.g. “big/small” contributions</td><td></td></tr>
</table>

	more loads at the "top end" than expected. The other observations are below expectation, but discrepancies are comparatively small.	E1 E1	to X^2 gets E1, … … and directions of discrepancies gets E1.	9
				18

Q2	A to B : $X \sim \text{N}(26,\ \sigma = 3)$ B to C : $Y \sim \text{N}(15,\ \sigma = 2)$		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(X < 24) = $\text{P}\left(Z < \frac{24-26}{3} = -0\cdot 6667\right)$ = 1 – 0·7476 = 0·2524	M1 A1 A1	For standardising. Award once, here or elsewhere. c.a.o.	3
(ii)	$X + Y \sim \text{N}(41,$ $\sigma^2 = 9 + 4 = 13\ [\sigma = 3\cdot 6056])$ P(this < 42) = $\text{P}\left(Z < \frac{42-41}{3\cdot 6056} = 0\cdot 2774\right) = 0\cdot 6093$	B1 B1 A1	Mean. Variance. Accept sd. c.a.o.	3
(iii)	$0\cdot 85X \sim \text{N}(22\cdot 1,$ $\sigma^2 = (0\cdot 85)^2 \times 9 = 6\cdot 5025\ [\sigma = 2\cdot 55])$ P(this < 24) = $\text{P}\left(Z < \frac{24-22\cdot 1}{2\cdot 55} = 0\cdot 7451\right)$ = 0·7719	B1 B1 A1	Mean. Variance. Accept sd. c.a.o.	3
(iv)	$0\cdot 9X + 0\cdot 8Y \sim N(23\cdot 4 + 12 = 35\cdot 4,$ $\sigma^2 = (0\cdot 9)^2 \times 9 + (0\cdot 8)^2 \times 4 = 9\cdot 85\ [\sigma = 3\cdot 1385])$ Require t such that 0·75 = P(this < t) $= \text{P}\left(Z < \frac{t - 35\cdot 4}{3\cdot 1385}\right) = \text{P}(Z < 0\cdot 6745)$ $\therefore t - 35\cdot 4 = 3\cdot 1385 \times 0\cdot 6745 = 2\cdot 1169$ $\Rightarrow t = 37\cdot 52$ Must therefore take scheduled time as 38	B1 B1 M1 B1 A1 M1	Mean. Variance. Accept sd. Formulation of requirement (using c's parameters). Any use of a continuity correction scores M0 (and hence A0). 0·6745 c.a.o. Round to next integer above c's value for t.	6
(v)	CI is given by $13\cdot 4 \pm 1\cdot 96 \frac{2}{\sqrt{15}}$ = 13·4 ± 1·0121 = (12·38(79), 14·41(21))	M1 B1 A1	If <u>both</u> 13·4 and $2/\sqrt{15}$ are correct. (N.B. 13·4 is given as $\bar{x}$ in the question.) (If $3/\sqrt{15}$ used, treat as mis-read and award this M1, but not the final A1.) For 1·96 c.a.o. Must be expressed as an interval.	3
				18

Q3				
(i)	Simple random sample might not be representative - e.g. it might contain only managers.	E1 E1	Or other sensible comment.	2
(ii)	Presumably there is a list of staff, so systematic sampling would be possible. List is likely to be alphabetical, in which case systematic sampling might not be representative. But if the list is in categories, systematic sampling could work well.	E1 E1 E1	Or other sensible comments.	3
(iii)	Would cover the entire population. Can get information for each category.	E1 E1		2
(iv)	5, 11, 24	B1	(4·8, 11·2, 24)	1
(v)	$\bar{x}$ = 345818, s_{n-1} = 69241 Underlying Normality H_0: μ = 300 000, H_1: μ > 300 000 Test statistic is $\frac{345818 - 300000}{\frac{69241}{\sqrt{11}}}$ =2·19(47). Refer to t_{10}. Upper 5% point is 1·812. Significant. Evidence that mean wealth is greater than 300 000. CI is given by 345818 ± 2·228 $\times \frac{69241}{\sqrt{11}}$ = 345818 ± 46513·84 = (299304(·2),	M1 A1 M1 A1 A1 A1 M1 B1 M1 A1	All given in the question. Allow alternatives: 300000 + (c's 1·812) × $\frac{69241}{\sqrt{11}}$ (= 337829) for subsequent comparison with 345818. or 345818 – (c's 1·812) × $\frac{69241}{\sqrt{11}}$ (= 307988) for comparison with 300000. c.a.o. but ft from here in any case if wrong. Use of $\mu - \bar{d}$ scores M1A0, but ft. No ft from here if wrong. No ft from here if wrong. ft only c's test statistic. ft only c's test statistic. Special case: (t_{11} and 1·796) can score 1 of these last 2 marks if either form of conclusion is given. c.a.o. Must be expressed as an	10

	392331(·8))		interval. ZERO/4 if not same distribution as test. Same wrong distribution scores maximum M1B0M1A0. Recovery to t_{10} is OK.	
				18

<table>
<tr><td>Q4</td><td></td><td></td><td></td><td></td></tr>
<tr><td>(i)</td><td><table><tr><th>Differences</th><th>Rank of |diff|</th></tr><tr><td>– 2</td><td>2</td></tr><tr><td>– 1</td><td>1</td></tr><tr><td>– 6</td><td>5</td></tr><tr><td>– 3</td><td>3</td></tr><tr><td>4</td><td>4</td></tr><tr><td>– 12</td><td>9</td></tr><tr><td>7</td><td>6</td></tr><tr><td>– 8</td><td>7</td></tr><tr><td>– 10</td><td>8</td></tr></table></td><td>M1

M1
A1</td><td>For differences.
ZERO in this section if differences not used.

For ranks.
FT from here if ranks wrong</td><td></td></tr>
<tr><td></td><td>T = 4 +6 = 10 (or 1+2+3+5+7+8+9 = 35)</td><td>B1</td><td></td><td></td></tr>
<tr><td></td><td>Refer to tables of Wilcoxon paired (/single sample) statistic.</td><td>M1</td><td>No ft from here if wrong.</td><td></td></tr>
<tr><td></td><td>Lower (or upper if 35 used) 5% tail is needed.</td><td>M1</td><td>i.e. a 1-tail test. No ft from here if wrong.</td><td></td></tr>
<tr><td></td><td>Value for n = 9 is 8 (or 37 if 35 used).</td><td>A1</td><td>No ft from here if wrong.</td><td></td></tr>
<tr><td></td><td>Result is not significant.</td><td>A1</td><td>ft only c’s test statistic.</td><td></td></tr>
<tr><td></td><td>No evidence to suggest a real change.</td><td>A1</td><td>ft only c’s test statistic.</td><td>9</td></tr>
<tr><td>(ii)</td><td>Normality of differences is required.</td><td>B1</td><td></td><td></td></tr>
<tr><td></td><td>CI MUST be based on DIFFERENCES.</td><td></td><td>ZERO/6 for the CI if differences not used.</td><td></td></tr>
<tr><td></td><td>Differences are 53, 15, 32, 13, 61, 82, 70</td><td></td><td>Accept negatives throughout.</td><td></td></tr>
<tr><td></td><td>$\bar{d} = 46\cdot5714$ $s_{n-1} = 27\cdot0485$</td><td>B1</td><td>Accept $s_{n-1}^2 = 731\cdot62$ …
[$s_n = 25\cdot0420$, but do NOT allow this here or in construction of CI.]</td><td></td></tr>
<tr><td></td><td>CI is given by
46·5714 ±</td><td>M1</td><td>Allow c’s $\bar{d}$ ± …</td><td></td></tr>
<tr><td></td><td>3·707</td><td>B1</td><td></td><td></td></tr>
<tr><td></td><td></td><td>B1</td><td>**If t_6 used.**
99% 2-tail point for c’s t distribution. (Independent of previous mark.)</td><td></td></tr>
<tr><td></td><td>$\times \frac{27\cdot0485}{\sqrt{7}}$</td><td>M1</td><td>Allow c’s s_{n-1}.</td><td></td></tr>
<tr><td></td><td>= 46·5714 ± 37·8980 = (8·67(34), 84·47)</td><td>A1</td><td>c.a.o. Must be expressed as an interval. [Upper boundary is 84·4694]</td><td></td></tr>
<tr><td></td><td>Cannot base CI on Normal distribution because</td><td>E1</td><td></td><td></td></tr>
<tr><td></td><td>sample is small
population s.d. is not known</td><td>E1</td><td>Insist on “population”, but allow “σ”.</td><td>9</td></tr>
<tr><td></td><td></td><td></td><td></td><td>18</td></tr>
</table>

Q1	$\mathrm{f}(x) = k(1-x) \qquad 0 \le x \le 1$			
(i)	$\int_0^1 k(1-x)\mathrm{d}x = 1$ $\therefore k[x - \frac{1}{2}x^2]_0^1 = 1$ $\therefore k(1-\frac{1}{2}) - 0 = 1$ $\therefore k = 2$ Labelled sketch: straight line segment from (0,2) to (1,0).	M1 E1 G1 G1	Integral of f(x), including limits (possibly implied later), equated to 1. Convincingly shown. Beware printed answer. Correct shape. Intercepts labelled.	4
(ii)	$\mathrm{E}(X) = \int_0^1 2x(1-x)\mathrm{d}x$ $= [x^2 - \frac{2}{3}x^3]_0^1 = (1-\frac{2}{3}) - 0 = \frac{1}{3}$ $\mathrm{E}(X^2) = \int_0^1 2x^2(1-x)\mathrm{d}x$ $= [\frac{2}{3}x^3 - \frac{2}{4}x^4]_0^1 = (\frac{2}{3} - \frac{1}{2}) - 0 = \frac{1}{6}$ $\mathrm{Var}(X) = \frac{1}{6} - (\frac{1}{3})^2$ $= \frac{1}{18}$	M1 A1 M1 M1 A1	Integral for E(X) including limits (which may appear later). Integral for E(X^2) including limits (which may appear later). Convincingly shown. Beware printed answer.	5
(iii)	$\mathrm{F}(x) = \int_0^x 2(1-t)\mathrm{d}t$ $= [2t - t^2]_0^x = (2x - x^2) - 0 = 2x - x^2$ $\mathrm{P}(X > \mu) = \mathrm{P}(X > \frac{1}{3}) = 1 - \mathrm{F}(\frac{1}{3})$ $= 1 - (2 \times \frac{1}{3} - (\frac{1}{3})^2) = 1 - \frac{5}{9} = \frac{4}{9}$	M1 A1 M1 A1	Definition of cdf, including limits, possibly implied later. Some valid method must be seen. [for 0 ≤ x ≤ 1; do not insist on this.] For 1 – c's F(μ). ft c's E(X) and F(x). If answer only seen in decimal expect 3 d.p. or better.	4
(iv)	$\mathrm{F}\left(1 - \frac{1}{\sqrt{2}}\right) = 2\left(1 - \frac{1}{\sqrt{2}}\right) - \left(1 - \frac{1}{\sqrt{2}}\right)^2$ $= 2 - \frac{2}{\sqrt{2}} - 1 + \frac{2}{\sqrt{2}} - \frac{1}{2} = \frac{1}{2}$ **Alternatively:** $2m - m^2 = \frac{1}{2}$ $\therefore m^2 - 2m + \frac{1}{2} = 0$ $\therefore m = 1 \pm \frac{1}{\sqrt{2}}$ so $m = 1 - \frac{1}{\sqrt{2}}$	M1 E1 M1 E1	Substitute $m = 1 - \frac{1}{\sqrt{2}}$ in c's cdf. Convincingly shown. Beware printed answer. Form a quadratic equation $\mathrm{F}(m) = \frac{1}{2}$ and attempt to solve it. ft c's cdf provided it leads to a quadratic. Convincingly shown. Beware printed answer.	2
(v)	$\bar{X} \sim \mathrm{N}(\frac{1}{3}, \frac{1}{1800})$	B1 B1 B1	Normal distribution. Mean. ft c's E(X). Correct variance.	3
				18

Q2				
(i)	$H_0 : \mu = 0{\cdot}6$ $H_1 : \mu < 0{\cdot}6$ Where μ is the (population) mean height of the saplings.	B1 B1 B1	Allow absence of "population" if correct notation μ is used, but do NOT allow "$\overline{X} = \ldots$" or similar unless $\overline{X}$ is clearly and explicitly stated to be a <u>population</u> mean. Hypotheses in words only must include "population".	
	$\bar{x} = 0\cdot5883$, $s_{n-1} = 0{\cdot}03664$ ($s_{n-1}^2 = 0{\cdot}00134$)	B1	Do not allow $s_n = 0{\cdot}03507$ ($s_n^2 = 0{\cdot}00123$).	
	Test statistic is $\dfrac{0\cdot5883 - 0\cdot6}{\left(\dfrac{0\cdot03664}{\sqrt{12}}\right)}$	M1	Allow c's $\bar{x}$ and/or s_{n-1}. Allow alternative: 0·6 ± (c's –1·796) × $\frac{0{\cdot}03664}{\sqrt{12}}$ (=0·5810, 0·6190) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ ± (c's –1·796) × $\frac{0{\cdot}03664}{\sqrt{12}}$ (=0·5693, 0·6073) for comparison with 0·6.)	
	= –1·103	A1	c.a.o. but ft from here in any case if wrong. Use of 0·6 – $\bar{x}$ scores M1A0, but ft.	
	Refer to t_{11}.	M1	No ft from here if wrong.	
	Lower 5% point is –1·796.	A1	No ft from here if wrong. Must be –1·796 unless it is clear that absolute values are being used.	
	–1·103 > –1·796, $\therefore$ Result is not significant.	E1	ft only c's test statistic.	
	Seems mean height of saplings meets the manager's requirements.	E1	ft only c's test statistic.	11
	Underlying population is Normal.	B1		
(ii)	CI is given by 0·5883 ±	M1	ft c's $\bar{x}$ ±.	
	2·201	B1		
	$\times \dfrac{0\cdot03664}{\sqrt{12}}$	M1	ft c's s_{n-1}.	
	= 0·5883 ± 0·0233 = (0·565(0), 0·611(6))	A1	c.a.o. Must be expressed as an interval. ZERO if not same distribution as test. Same wrong distribution scores maximum M1B0M1A0. Recovery to t_{11} is OK.	

	In repeated sampling, 95% of intervals constructed in this way will contain the true population mean.	E1		5
(iii)	Could use the Wilcoxon test. Null hypothesis is “Median = 0.6”.	E1 E1		2
				18

Q3	$M \sim \text{N}(44,\ 4\cdot8^2)$ $H \sim \text{N}(32,\ 2\cdot6^2)$ $P \sim \text{N}(21,\ 3\cdot7^2)$		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)	$\text{P}(M < 50) = \text{P}(Z < \frac{50-44}{4\cdot8} = 1\cdot25)$ $= 0\cdot8944$	M1 A1 A1	For standardising. Award once, here or elsewhere.	3
(ii)	$H + P \sim \text{N}(32 + 21 = 53,\ 2\cdot6^2 + 3\cdot7^2 = 20\cdot45)$ $\text{P}(H + P < 50) = \text{P}(Z < \frac{50-53}{\sqrt{20\cdot45}} = -0\cdot6634)$ $= 1 - 0\cdot7465 = 0\cdot2535$	B1 B1 A1	Mean. Variance. Accept sd = √20·45 = 4·522... c.a.o.	3
(iii)	Want $\text{P}(M > H + P)$ i.e. $\text{P}(M - (H + P) > 0)$ $M - (H + P) \sim \text{N}(44 - (32 + 21) = -9,\ 4\cdot8^2 + 2\cdot6^2 + 3\cdot7^2 = 43\cdot49)$ $\text{P(this} > 0) = \text{P}(Z > \frac{0-(-9)}{\sqrt{43\cdot49}} = 1\cdot365)$ $= 1 - 0\cdot9139 = 0\cdot0861$	M1 B1 B1 A1	Allow $H + P - M$ provided subsequent work is consistent. Mean. Variance. Accept sd = √43·49 = 6·594... c.a.o.	4
(iv)	Mean = 44 + 44 + 32 + 32 + 21 + 21 = 194 Variance = $4\cdot8^2 + 4\cdot8^2 + 2\cdot6^2 + 2\cdot6^2 + 3\cdot7^2 + 3\cdot7^2$ = 86·98	B1 B1	(sd = 9·3263…)	2
(v)	$C \sim \text{N}(194 \times 0\cdot15 + 10 = 39\cdot10,$ $86\cdot98 \times 0\cdot15^2 = 1\cdot957)$ $\text{P}(C \le 40) = \text{P}(Z \le \frac{40-39\cdot10}{\sqrt{1\cdot957}} = 0\cdot6433)$ $= 0\cdot7400$ **Alternatively:** $\text{P}(C \le 40) = \text{P(total time} \le \frac{40-10}{0.15} = 200$ minutes) $= \text{P}(Z \le \frac{200-194}{\sqrt{86\cdot98}} = 0\cdot6433)$	M1 M1 A1 M1 A1 A1 M1 M1 A1 M1 A1	c's mean in (iv) × 0·15 + 10 (or subtract 10 from 40 below) ft c's mean in (iv). c's variance in (iv) × $0\cdot15^2$ ft c's variance in (iv). c.a.o. – 10 ÷ 0.15 c.a.o. Correct use of c's variance in (iv). ft c's mean and variance in (iv).	6

	= 0·7400	A1	c.a.o.	
				18

Q4				
(a)				

Obs	Exp
10	6·68

		M1	Combine first two rows.	
	$\therefore X^2 = \frac{(10-6\cdot68)^2}{6\cdot68} + \text{etc}$	M1		
	= 1·6501 + 1·7740 + 3·3203 + 4·5018 + 0·4015 + 0·8135 = 12·46(12)	A1		
	d.o.f. = 6 – 3 = 3		Require d.o.f. = No. cells used – 3.	
	Refer to χ_3^2.	M1	No ft from here if wrong.	
	Upper 5% point is 7·815	A1	No ft from here if wrong.	
	12·46 > 7·815 ∴ Result is significant.	E1	ft only c's test statistic.	
	Seems the Normal model does not fit the data at the 5% level.	E1	ft only c's test statistic.	
	E.g. • The biggest discrepancy is in the class 1·01 < *a* ≤ 1·02	E1		
	• The model overestimates in classes …, but underestimates in classes …	E1	Any two suitable comments.	9
(b)				

Old – New:	0·007	0·002	–0·001	–0·003	0·004	–0·008	–0·010	0·009	–0·005	–0·016
Rank of |diff|	6	2	1	3	4	7	9	8	5	10

		M1	For differences. ZERO in this section if differences not used.	
		M1	For ranks of |difference|.	
		A1	All correct. ft from here if ranks wrong.	
	$W_+ = 6 + 2 + 4 + 8 = 20$	B1	Or $W_- = 1 + 3 + 7 + 9 + 5 + 10 = 35$	
	Refer to Wilcoxon single sample (/paired) tables for *n* = 10.	M1	No ft from here if wrong.	
	Lower two-tail 10% point is …	M1	Or, if 35 used, upper point is 45.	
	… 10.	A1	No ft from here if wrong.	
	20 > 10 ∴ Result is not significant.	E1	Or 35 < 45. ft only c's test statistic.	
	Seems there is no reason to suppose the barometers differ.	E1	ft only c's test statistic.	9
				18

Q1	$\text{f}(t) = kt^3(2-t) \qquad 0 < t \le 2$			
(i)	$\int_0^2 kt^3(2-t)\text{d}t = 1$ $\therefore \left[k\left(\frac{2t^4}{4} - \frac{t^5}{5}\right)\right]_0^2 = 1$ $\therefore k\left(8 - \frac{32}{5}\right) - 0 = 1$ $\therefore k \times \frac{8}{5} = 1 \qquad \therefore k = \frac{5}{8}$	M1 E1	Integral of f(t), including limits (possibly implied later), equated to 1. Convincingly shown. Beware printed answer.	2
(ii)	$\frac{\text{d}f}{\text{d}t} = \frac{5}{8}\left(6t^2 - 4t^3\right) = 0$ $\therefore 6t^2 - 4t^3 = 0$ $\therefore 2t^2(3-2t) = 0$ $\therefore t = (0 \text{ or}) \ \frac{3}{2}$	M1 A1	Differentiate and set equal to zero. c.a.o.	2
(iii)	$\text{E}(T) = \int_0^2 \frac{5}{8}t^4(2-t)\text{d}t$ $= \left[\frac{5}{8}\left(\frac{2t^5}{5} - \frac{t^6}{6}\right)\right]_0^2 = \frac{5}{8} \times \left(\frac{64}{5} - \frac{64}{6}\right) = \frac{4}{3}$ $\text{E}(T^2) = \int_0^2 \frac{5}{8}t^5(2-t)\text{d}t$ $= \left[\frac{5}{8}\left(\frac{2t^6}{6} - \frac{t^7}{7}\right)\right]_0^2 = \frac{5}{8} \times \left(\frac{128}{6} - \frac{128}{7}\right) = \frac{40}{21}$ $\text{Var}(T) = \frac{40}{21} - \left(\frac{4}{3}\right)^2 = \frac{120-112}{63} = \frac{8}{63}$	M1 A1 M1 M1 A1	Integral for E(T) including limits (which may appear later). Integral for E(T^2) including limits (which may appear later). Convincingly shown. Beware printed answer.	5
(iv)	$\overline{T} \sim \text{N}\left(\frac{4}{3}, \ \frac{8}{63n}\right)$	B1 B1 B1	Normal distribution. Mean. ft c's E(T). Correct variance.	3

(v)	$n = 100, \quad \bar{t} = \frac{145 \cdot 2}{100} = 1 \cdot 452,$ $s_{n-1}^2 = \frac{223 \cdot 41 - 100 \times 1 \cdot 452^2}{99} = 0 \cdot 12707$	B1	Both mean and variance. Accept sd = 0·3565	
	CI is given by 1·452 ±	M1	ft c's $\bar{t}$ ± .	
	1·96	B1		
	$\times \frac{0 \cdot 3565}{\sqrt{100}}$	M1	ft c's s_{n-1}.	
	= 1·452 ± 0·0698 = (1·382, 1·522)	A1	c.a.o. Must be expressed as an interval.	
	Since E(*T*) (= 4/3) lies outside this interval it seems the model may not be appropriate.	E1		6
				18

Q2	$Ca \sim \mathrm{N}(60{\cdot}2,\ 5{\cdot}2^2)$ $Co \sim \mathrm{N}(33{\cdot}9,\ 6{\cdot}3^2)$ $L \sim \mathrm{N}(52{\cdot}4,\ 4{\cdot}9^2)$		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)	$\mathrm{P}(Co < 40) = \mathrm{P}(Z < \frac{40-33\cdot9}{6\cdot3} = 0{\cdot}9683)$ $= 0{\cdot}8336$	M1 A1 A1	For standardising. Award once, here or elsewhere. c.a.o.	3
(ii)	Want $\mathrm{P}(L > Ca)$ i.e. $\mathrm{P}(L - Ca > 0)$ $L - Ca \sim \mathrm{N}(52{\cdot}4 - 60{\cdot}2 = -7{\cdot}8,$ $4{\cdot}9^2 + 5{\cdot}2^2 = 51{\cdot}05)$ $\mathrm{P}(\text{this} > 0) = \mathrm{P}(Z > \frac{0-(-7\cdot8)}{\sqrt{51\cdot05}} = 1{\cdot}0917)$ $= 1 - 0{\cdot}8625 = 0{\cdot}1375$	M1 B1 B1 A1	Allow $Ca - L$ provided subsequent work is consistent. Mean. Variance. Accept sd = $\sqrt{51{\cdot}05}$= 7·1449... c.a.o.	4
(iii)	Want $\mathrm{P}(Ca_1 + Ca_2 + Ca_3 + Ca_4 > 225)$ $Ca_1 + \ldots \sim \mathrm{N}(60{\cdot}2 + 60{\cdot}2 + 60{\cdot}2 + 60{\cdot}2 = 240{\cdot}8,$ $5{\cdot}2^2 + 5{\cdot}2^2 + 5{\cdot}2^2 + 5{\cdot}2^2 = 108{\cdot}16)$ $\mathrm{P}(\text{this} > 225) = \mathrm{P}(Z > \frac{225-240\cdot8}{\sqrt{108\cdot16}} = -1{\cdot}519)$ $= 0{\cdot}9356$ Must assume that the weeks are independent of each other.	M1 B1 B1 A1 B1	Mean. Variance. Accept sd=$\sqrt{108{\cdot}16}$=10·4. c.a.o.	5
(iv)	$R \sim \mathrm{N}(0{\cdot}05{\times}60{\cdot}2 + 0{\cdot}1{\times}33{\cdot}9 + 0{\cdot}2{\times}52{\cdot}4 = 16{\cdot}88,$ $0{\cdot}05^2{\times}5{\cdot}2^2 + 0{\cdot}1^2{\times}6{\cdot}3^2 + 0{\cdot}2^2{\times}4{\cdot}9^2 = 1{\cdot}4249)$ $\mathrm{P}(R > 20) = \mathrm{P}(Z > \frac{20-16\cdot88}{\sqrt{1\cdot4249}} = 2{\cdot}613)$ $= 1 - 0{\cdot}9955 = 0{\cdot}0045$	M1 A1 M1 M1 A1 A1	Mean. For $0{\cdot}05^2$ etc. For $\times 5{\cdot}2^2$ etc. Accept sd = $\sqrt{1{\cdot}4249}$ = 1·1937. c.a.o.	6
				18

Q3				
(a) (i)	$H_0 : \mu_D = 0$ $H_1 : \mu_D > 0$	B1	Both. Accept alternatives e.g. $\mu_D < 0$ for H_1, or $\mu_A - \mu_B$ etc provided adequately defined.	
	Where μ_D is the (population) mean reduction in absenteeism.	B1	Allow absence of "population" if correct notation μ is used, but do NOT allow " $\bar{X} = \dots$ " or similar unless $\bar{X}$ is clearly and explicitly stated to be a <u>population</u> mean. Hypotheses in words only must include "population".	
	Must assume Normality …	B1		4
	… of differences.	B1		
(ii)	Differences (reductions) (before – after) 1·7, 0·7, 0·6, –1·3, 0·1, –0·9, 0·6, –0·7, 0·4, 2·7, 0·9		Allow "after – before" if consistent with alternatives above.	
	$\bar{x} = 0 \cdot 4364$, $s_{n-1} = 1 \cdot 1518$ ($s_{n-1}{}^2 = 1 \cdot 3265$)	B1	Do not allow $s_n = 1 \cdot 098$ ($s_n{}^2 = 1 \cdot 205$).	
	Test statistic is $\dfrac{0 \cdot 4364 - 0}{\left(\dfrac{1 \cdot 1518}{\sqrt{11}}\right)}$	M1	Allow c's $\bar{x}$ and/or s_{n-1}. Allow alternative: 0 ± (c's 1·812) × $\frac{1 \cdot 1518}{\sqrt{11}}$ (= –0·6293, 0·6293) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ ± (c's 1·812) × $\frac{1 \cdot 1518}{\sqrt{11}}$ (= –0·1929, 1·0657) for comparison with 0.)	
	= 1·256(56…)	A1	c.a.o. but ft from here in any case if wrong. Use of $0 - \bar{x}$ scores M1A0, but ft.	
	Refer to t_{10}.	M1	No ft from here if wrong.	
	Upper 5% point is 1·812.	A1	No ft from here if wrong. For alternative H_1 expect –1·812 unless it is clear that absolute values are being used.	
	1·256 < 1·812, ∴ Result is not significant.	E1	ft only c's test statistic.	
	Seems there has been no reduction in mean absenteeism.	E1	ft only c's test statistic. Special case: (t_{11} and 1·796) can score 1 of these last 2 marks if either form of conclusion is given.	7

(b)	For "days lost after" $\bar{x} = 4\cdot6182$, $s_{n-1} = 1\cdot4851$ ($s_{n-1}^2 = 2\cdot2056$)	B1	Do not allow $s_n = 1\cdot4160$ ($s_n^2 = 2\cdot0051$).	
	CI is given by 4·6182 ±	M1	ft c's $\bar{x}$ ±.	
	2·228	B1		
	$\times \frac{1\cdot4851}{\sqrt{11}}$	M1	ft c's s_{n-1}.	
	= 4·6182 ± 0·9976 = (3·620(6), 5·615(8))	A1	c.a.o. Must be expressed as an interval. ZERO if not same distribution as test. Same wrong distribution scores maximum M1B0M1A0. Recovery to t_{10} is OK.	
	Assume Normality of population of "days lost after".	E1		
	Since 3·5 lies outside the interval it seems that the target has not been achieved.	E1		7
				18

Q4				
(i)				

Obs	21	24	12	15	13	9	6
Exp	26·53	17·22	20·25	11·00	10·94	8·74	5·32

		M1	Probabilities × 100.	
		A1	All Expected frequencies correct.	
	$\therefore X^2 = \frac{(21-26\cdot53)^2}{26\cdot53} + \text{etc}$	M1		
	= 1·1527 + 2·6695 + 3·3611 + 1·4545 + 0·3879 + 0·0077+ 0·0869	A1	At least 4 values correct.	
	= 9·1203	A1		
	d.o.f. = 7 – 1 = 6			
	Refer to χ_6^2.	M1	No ft from here if wrong.	
	Upper 5% point is 12·59	A1	No ft from here if wrong.	
	9·1203 < 12·59 ∴ Result is not significant.	E1	ft only c's test statistic.	
	Evidence suggests the model fits the data at the 5% level.	E1	ft only c's test statistic.	9
(ii)				

Data	Diff = data –124	Rank of |diff|
239	115	9
77	–47	3
179	55	4
221	97	7
100	–24	2
312	188	10
52	–72	5
129	5	1
236	112	8
42	–82	6

		M1	For differences.	
		M1	For ranks of |difference|.	
		A1	All correct. ft from here if ranks wrong.	
	$W_- = 3 + 2 + 5 + 6 = 16$	B1	Or $W_+ = 9 + 4 + 7 + 10 + 1 + 8 = 39$	
	Refer to Wilcoxon single sample (/paired) tables for $n = 10$.	M1	No ft from here if wrong.	
	Lower two-tail 10% point is … … 10.	M1A1	Or, if 39 used, upper point is 45. No ft from here if wrong.	
	16 > 10 ∴ Result is not significant.	E1	Or 39 < 45. ft only c's test statistic.	
	Seems there is no evidence against the median length being 124.	E1	ft only c's test statistic.	9
				18

4768 Statistics 3

Q1 (a)	$\mathrm{P}(T>t)=\frac{k}{t^2}$, $t \ge 1$,			
(i)	$\mathrm{F}(t) = \mathrm{P}(T < t) = 1 - \mathrm{P}(T > t)$ $\therefore \mathrm{F}(t) = 1 - \frac{k}{t^2}$	M1	Use of 1 – P(…).	
	$\mathrm{F}(1) = 0$ $\therefore 1 - \frac{k}{1^2} = 0$	M1		
	$\therefore k = 1$	A1	Beware: answer given.	3
(ii)	$\mathrm{f}(t) = \frac{\mathrm{d}\,\mathrm{F}(t)}{\mathrm{d}t}$	M1	Attempt to differentiate c's cdf.	
	$= \frac{2}{t^3}$	A1	(For $t \ge 1$, but condone absence of this.) Ft c's cdf provided answer sensible.	2
(iii)	$\mu = \int_1^\infty t\mathrm{f}(t)\mathrm{d}t = \int_1^\infty \frac{2}{t^2}\mathrm{d}t$	M1	Correct form of integral for the mean, with correct limits. Ft c's pdf.	
	$= \left[\frac{-2}{t}\right]_1^\infty$	A1	Correctly integrated. Ft c's pdf.	
	$= 0 - (-2) = 2$	A1	Correct use of limits leading to correct value. Ft c's pdf provided answer sensible.	3
(b)	H_0: $m = 5.4$ H_1: $m \neq 5.4$	B1	Both hypotheses. Hypotheses in words only must include "population".	
	where m is the population median time for the task.	B1	For adequate verbal definition.	

Times	− 5.4	Rank of |diff|
6.4	1.0	8
5.9	0.5	5
5.0	−0.4	4
6.2	0.8	7
6.8	1.4	10
6.0	0.6	6
5.2	−0.2	2
6.5	1.1	9
5.7	0.3	3
5.3	−0.1	1

		M1	for subtracting 5.4.	
		M1	for ranks.	
		A1	FT if ranks wrong.	
	$W_- = 1 + 2 + 4 = 7$ (or $W_+ = 3+5+6+7+8+9+10 = 48$)	B1		
	Refer to tables of Wilcoxon single sample (/paired) statistic for $n = 10$.	M1	No ft from here if wrong.	
	Lower (or upper if 48 used) double-tailed 5% point is 8 (or 47 if 48 used).	A1	i.e. a 2-tail test. No ft from here if wrong.	
	Result is significant.	A1	ft only c's test statistic.	
	Seems that the median time is no longer as previously thought.	A1	ft only c's test statistic.	10

Q2	X ~ N(260, σ = 24)		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 (X <300) = P$(Z < \frac{300-260}{24} = 1.6667)$ = 0.9522	M1 A1 A1	For standardising. Award once, here or elsewhere.	3
(ii)	Y ~ N(260 ×0.6 = 156, $24^2 \times 0.6^2$ = 207.36 $\text{P}(Y > 175) = \text{P}(Z > \frac{175-156}{14.4} = 1.3194)$ $= 1 - 0.9063 = 0.0937$	B1 B1 A1	Mean. Variance. Accept sd (= 14.4). c.a.o.	3
(iii)	$Y_1 + Y_2 + Y_3 + Y_4$ ~ N(624, 829.44) $\text{P}(\text{this} < 600) = \text{P}(Z < \frac{600-624}{28.8} = -0.8333)$ $= 1 - 0.7976 = 0.2024$	B1 B1 A1	Mean. Ft mean of (ii). Variance. Accept sd (= 28.8). Ft variance of (ii). c.a.o.	3
(iv)	Require w such that $0.975 = \text{P}(\text{above} > w) = \text{P}\left(Z > \frac{w-624}{28.8}\right)$ $= \text{P}(Z > -1.96)$ $\therefore w - 624 = 28.8 \times -1.96 \Rightarrow w = 567.5(52)$	M1 B1 A1	Formulation of requirement. − 1.96 Ft parameters of (iii).	3
(v)	On ~ N(150, σ = 18) $X_1 + X_2 + X_3 + On_1 + On_2$ ~ N(1080, 2376) $\text{P}(\text{this} > 1000) = \text{P}(Z > \frac{1000-1080}{48.744} = -1.6412)$ $= 0.9496$	B1 B1 A1	Mean. Variance. Accept sd (= 48.744). c.a.o.	3
(vi)	Given $\bar{x} = 252.4$ $s_{n-1} = 24.6$ CI is given by $252.4 \pm 2.576 \times \frac{24.6}{\sqrt{100}}$ = 252.4 ± 6.33(6) = (246.0(63), 258.7(36))	M1 B1 A1	Correct use of 252.4 and $24.6/\sqrt{100}$. For 2.576. c.a.o. Must be expressed as an interval.	3
				18

Q3				
(i)	A t test should be used because the sample is small, the population variance is unknown, the background population is Normal	 E1 E1 E1		3
(ii)	H$_0$: $\mu = 380$ H$_1$: $\mu < 380$	B1	Both hypotheses. Hypotheses in words only must include "population".	
	where μ is the mean temperature in the chamber.	B1	For adequate verbal definition. Allow absence of "population" if correct notation μ is used, but do NOT allow "$\bar{X} = \ldots$" or similar unless $\bar{X}$ is clearly and explicitly stated to be a population mean.	
	$\bar{x} = 373.825$ $s_{n-1} = 9.368$	B1	s_n = 8.969 but do NOT allow this here or in construction of test statistic, but FT from there.	
	Test statistic is $\frac{373.825 - 380}{\frac{9.368}{\sqrt{12}}}$	M1	Allow c's $\bar{x}$ and/or s_{n-1}. Allow alternative: 380 + (c's –1.796) × $\frac{9.368}{\sqrt{12}}$ (= 375.143) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ – (c's –1.796) × $\frac{9.368}{\sqrt{12}}$ (= 378.681) for comparison with 380.)	
	= –2.283(359).	A1	c.a.o. but ft from here in any case if wrong. Use of 380 – $\bar{x}$ scores M1A0, but ft.	
	Refer to t_{11}.	M1	No ft from here if wrong.	
	Single-tailed 5% point is –1.796.	A1	Must be minus 1.796 unless absolute values are being compared. No ft from here if wrong.	
	Significant.	A1	ft only c's test statistic.	
	Seems mean temperature in the chamber has fallen.	A1	ft only c's test statistic.	9
(iii)	CI is given by 373.825 ± 2·201 $\times \frac{9.368}{\sqrt{12}}$	 M1 B1 M1		
	= 373.825 ± 5.952= (367.87(3), 379.77(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_{11} is OK.	4
(iv)	Advantage: greater certainty.	E1	Or equivalents.	
	Disadvantage: less precision.	E1		2 18

Q4				
(a) **(i)**	$\bar{x} = \frac{1125}{500} = 2.25$	B1		
	For binomial E(X) = $n \times p$	M1	Use of mean of binomial distribution. May be implicit.	
	$\therefore \hat{p} = \frac{2.25}{5} = 0.45$	A1	Beware: answer given.	3
(ii)				

f_o	32	110	154	125	63	16
f_e (calc)	25.164	102.944	168.455	137.827	56.384	9.226
f_e (tables)	25.15	102.95	168.45	137.85	56.35	9.25

		M1	Calculation of expected frequencies.	
		A1	All correct.	
	X^2 = 1.8571 + 0.4836 + 1.2404 + 1.1938 + 0.7763 + 4.9737	M1	Or using tables: 1.8657 + 0.4828 + 1.2396 + 1.1978 + 0.7848 + 4.9257	
	= 10.52(49)	A1	c.a.o. Or using tables: 10.49(64)	
	Refer to χ_4^2.	M1	Allow correct df (= cells – 2) from wrongly grouped or ungrouped table, and FT. Otherwise, no FT if wrong.	
	Upper 5% point is 9.488.	A1	No ft from here if wrong.	
	Significant.	A1	ft only c's test statistic.	
	Suggests binomial model does not fit.	A1	ft only c's test statistic.	
	The model appears to overestimate in the middle and to underestimate at the tails.	E1	Accept also any other sensible comment e.g. at 2.5% significance, the result would NOT have been significant.	
	The biggest discrepancy is at X = 5.	E1		
	A binomial model assumes all trials are independent with a constant probability of "success". It seems unlikely that there will be independence within families and/or that p will be the same for all families.	E2	(E2, 1, 0) Any sensible comment which addresses independence and constant p.	12
(b)	She should try to choose a simple random sample	E1		
	which would involve establishing a sampling frame	E1	Allow sensible discussion of practical limitations of choosing a random sample.	
	and using some form of random number generator.	E1	Allow other sensible suggestions. E.g systematic sample - choosing every tenth family; stratified sample - by the number of girls in a family.	3
				18

4768 Statistics 3

Q1	$\mathrm{f}(x) = k(20 - x) \qquad 0 \le x \le 20$			
(a) (i)	$\int_0^{20} k(20-x)\mathrm{d}x = \left[k\left(20x - \frac{x^2}{2}\right)\right]_0^{20} = k \times 200 = 1$	M1	Integral of f(x), including limits (which may appear later), set equal to 1. Accept a geometrical approach using the area of a triangle.	
	$\therefore k = \frac{1}{200}$	A1	C.a.o.	
	Straight line graph with negative gradient, in the first quadrant.	G1		
	Intercept correctly labelled (20, 0), with nothing extending beyond these points.	G1		
	Sarah is more likely to have only a short time to wait for the bus.	E1		5
(ii)	Cdf $\mathrm{F}(x) = \int_0^x \mathrm{f}(t)\mathrm{d}t$ $= \frac{1}{200}\left(20x - \frac{x^2}{2}\right)$	M1	Definition of cdf, including limits (or use of "+c" and attempt to evaluate it), possibly implied later. Some valid method must be seen.	
	$= \frac{x}{10} - \frac{x^2}{400}$	A1	Or equivalent expression; condone absence of domain [0, 20].	
	P(X > 10) = 1 – F(10)	M1	Correct use of c's cdf.	
	= 1 – (1 – ¼) = ¼	A1	f.t. c's cdf. Accept geometrical method, e.g area = ½(20 – 10)f(10), or similarity.	4
(iii)	Median time, m, is given by F(m) = ½. $\therefore \frac{m}{10} - \frac{m^2}{400} = \frac{1}{2}$	M1	Definition of median used, leading to the formation of a quadratic equation.	
	$\therefore m^2 - 40m + 200 = 0$	M1	Rearrange and attempt to solve the quadratic equation.	
	$\therefore m = 5.86$	A1	Other solution is 34.14; no explicit reference to/rejection of it is required.	3

(b) (i)	A simple random sample is one where every sample of the required size has an equal chance of being chosen.	E2	S.C. Allow E1 for "Every member of the population has an equal chance of being chosen independently of every other member".	2
(ii)	Identify clusters which are capable of representing the population as a whole. Choose a random sample of clusters. Randomly sample or enumerate within the chosen clusters.	E1 E1 E1		3
(iii)	A random sample of the school population might involve having to interview single or small numbers of pupils from a large number of schools across the entire country. Therefore it would be more practical to use a cluster sample.	E1 E1	For "practical" accept e.g. convenient / efficient / economical.	2
				19

Q2	$A \sim \text{N}(100,\ \sigma = 1.9)$ $B \sim \text{N}(50,\ \sigma = 1.3)$		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(*A* < 103) = $\text{P}\left(Z < \frac{103-100}{1.9} = 1.5789\right)$ = 0·9429	M1 A1 A1	For standardising. Award once, here or elsewhere. c.a.o.	3
(ii)	$A_1 + A_2 + A_3 \sim \text{N}(300,$ $\sigma^2 = 1.9^2 + 1.9^2 + 1.9^2 = 10.83\ )$ P(this > 306) = $\text{P}\left(Z > \frac{306-300}{3\cdot 291} = 1\cdot 823\right) = 1 - 0\cdot 9658 = 0.0342$	B1 B1 A1	Mean. Variance. Accept sd (= 3.291). c.a.o.	3
(iii)	$A + B \sim \text{N}(150,$ $\sigma^2 = 1.9^2 + 1.3^2 = 5.3\ )$ P(this > 147) = $\text{P}\left(Z > \frac{147-150}{2\cdot 302} = -1.303\right)$ = 0·9037	B1 B1 A1	Mean. Variance. Accept sd (= 2.302). c.a.o.	3
(iv)	$B_1 + B_2 - A \sim N(0,$ $1\cdot 3^2 + 1\cdot 3^2 + 1\cdot 9^2 = 6\cdot 99)$ P(−3< this < 3) $= \text{P}\left(\frac{-3-0}{2.644} < Z < \frac{3-0}{2.644}\right) = \text{P}(-1\cdot 135 < Z < 1\cdot 135)$ = 2 × 0.8718 – 1 = 0.7436	B1 B1 M1 A1 A1	Mean. Or *A* – (B_1 + B_2). Variance. Accept sd (= 2.644). Formulation of requirement ... … two sided. c.a.o.	5
(v)	Given $\bar{x} = 302.3 \quad s_{n-1} = 3.7$ CI is given by $302.3 \pm 1\cdot 96 \times \frac{3.7}{\sqrt{100}}$ = 302·3 ± 0·7252 = (301·57(48), 303·02(52)) The batch appears not to be as specified since 300 is outside the confidence interval.	M1 B1 A1 E1	Correct use of 302.3 and $3.7/\sqrt{100}$. For 1·96 c.a.o. Must be expressed as an interval.	4
				18

Q3				
(a) (i)	H_0: $\mu_D = 0$ (or $\mu_I = \mu_{II}$) H_1: $\mu_D \neq 0$ (or $\mu_{II} \neq \mu_I$)	B1	Both. Hypotheses in words only must include "population".	
	where μ_D is "mean for II – mean for I"	B1	For adequate verbal definition. Allow absence of "population" if correct notation μ is used, but do NOT allow "$\overline{X}_I = \overline{X}_{II}$" or similar unless $\overline{X}$ is clearly and explicitly stated to be a population mean.	
	Normality of differences is required.	B1		3
(ii)	MUST be PAIRED COMPARISON t test. Differences are:			

10.0	26.8	42.7	2.4	–14.9	–2.0	16.3	11.5

	$\bar{d} = 11.6$ $\quad s_{n-1} = 17.707$	B1	s_n = 16.563 but do NOT allow this here or in construction of test statistic, but FT from there.	
	Test statistic is $\dfrac{11.6-0}{\frac{17.707}{\sqrt{8}}}$	M1	Allow c's $\bar{d}$ and/or s_{n-1}. Allow alternative: 0 + (c's 2.365) × $\frac{17.707}{\sqrt{8}}$ (= 14.806) for subsequent comparison with $\bar{d}$. (Or $\bar{d}$ – (c's 2.365) × $\frac{17.707}{\sqrt{8}}$ (=–3.206) for comparison with 0.)	
	= 1.852(92).	A1	c.a.o. but ft from here in any case if wrong. Use of 0 – $\bar{d}$ scores M1A0, but ft.	
	Refer to t_7.	M1	No ft from here if wrong.	
	Double-tailed 5% point is 2.365.	A1	No ft from here if wrong.	
	Not significant.	A1	ft only c's test statistic.	
	Seems there is no difference between the mean yields of the two types of plant.	A1	ft only c's test statistic. Special case: (t_8 and 2.306) can score 1 of these last 2 marks if either form of conclusion is given.	7

(b)

Diff	−5	4	−14	−3	6	1	−11	−8	−7	−9
Rank of |diff|	4	3	10	2	5	1	9	7	6	8

(b)		M1	For differences. ZERO in this section if differences not used.	
		M1	For ranks.	
		A1	FT from here if ranks wrong	
	$W_+ = 1 + 3 + 5 = 9$ (or $W_- = 2+4+6+7+8+9+10 = 46$)	B1		
	Refer to tables of Wilcoxon paired (/single sample) statistic for $n = 10$.	M1	No ft from here if wrong.	
	Lower (or upper if 46 used) double-tailed 5% point is 8 (or 47 if 46 used).	A1	i.e. a 2-tail test. No ft from here if wrong.	
	Result is not significant.	A1	ft only c's test statistic.	
	No evidence to suggest the tasters differ on the whole.	A1	ft only c's test statistic.	8
				18

Q4				
(a) (i)	$\bar{x} = \frac{310}{100} = 3.1$ $s^2 = \frac{1288 - 100 \times 3.1^2}{99} = \frac{327}{99} = 3.303$ Evidence could support Poisson since the variance is fairly close to the mean.	B1 B1 E1		3
(ii)				

f_o	6	16	19	18	17	14	6	4	0
f_e	4.50	13.97	21.65	22.37	17.33	10.75	5.55	2.46	1.42
Merged	22 18.47						10 9.43		

		M1 A1 A1	Calculation of expected frequencies. Last cell correct. All others correct, but ft if wrong.	
		M1	Combining cells. (Condone if not combined as fully as shown above, but require top two cells combined as a minimum.)	
	X^2 = 0.6747 + 0.3244 + 0.8537 + 0.0063 + 0.9826 + 0.0345	M1	Calculation of X^2.	
	= 2.876(2)	A1	(Condone wrong last cell.) Depends on both of the preceding M marks.	
	Refer to χ_4^2. e.g. Upper 10% point is 7.779.	M1	Allow correct df (= cells – 2) from wrongly grouped or ungrouped table, and FT. Otherwise, no FT if wrong.	
	Not significant.	A1	ft only c's test statistic.	
	Suggests Poisson model does fit …	A1	ft only c's test statistic.	
	… at any reasonable level of significance.	A1	Or other sensible comment.	10
(b)	CI is given by 1.465 ±	M1	If <u>both</u> 1.465 and $0.3288/\sqrt{10}$ are correct.	
	2·262	B1 B1	**If t_9 used.** 95% 2-tail point for c's *t* distribution (Independent of previous mark).	
	$\times \frac{0.3288}{\sqrt{10}}$			
	= 1.465 ± 0.2352= (1.2298, 1.7002)	A1	c.a.o. Must be expressed as an interval.	4
				17

4768 Statistics 3

<table>
<tr><td>Q1
(a)</td><td>$\mathrm{f}(x) = \lambda x^c, \ 0 \le x \le 1, \ \lambda > 1$</td><td></td><td></td><td></td></tr>
<tr><td>(i)</td><td>$\int_0^1 \lambda x^c \mathrm{d}x = 1$
$\therefore \left[\frac{\lambda x^{c+1}}{c+1}\right]_0^1 = 1$
$\therefore \frac{\lambda}{c+1} = 1 \qquad \therefore c = \lambda - 1$</td><td>M1
M1
A1</td><td>Correct integral, with limits (possibly appearing later), set equal to 1.
Integration correct and limits used.
c.a.o.</td><td>3</td></tr>
<tr><td>(ii)</td><td>$\mathrm{E}(X) = \int_0^1 \lambda x^{\lambda} \mathrm{d}x$
$= \left[\frac{\lambda x^{\lambda+1}}{\lambda+1}\right]_0^1 \quad = \frac{\lambda}{\lambda+1}.$</td><td>M1
M1
A1</td><td>Correct form of integral for E(<i>X</i>). Allow c’s expression for <i>c</i>.
Integration correct and limits used. ft c’s <i>c</i>.</td><td>3</td></tr>
<tr><td>(iii)</td><td>$\mathrm{E}(X^2) = \int_0^1 \lambda x^{\lambda+1} \mathrm{d}x$
$= \left[\frac{\lambda x^{\lambda+2}}{\lambda+2}\right]_0^1 = \frac{\lambda}{\lambda+2}.$
$\mathrm{Var}(X) = \frac{\lambda}{\lambda+2} - \left(\frac{\lambda}{\lambda+1}\right)^2 = \frac{\lambda(\lambda+1)^2 - \lambda^2(\lambda+2)}{(\lambda+2)(\lambda+1)^2}$
$= \frac{\lambda^3 + 2\lambda^2 + \lambda - \lambda^3 - 2\lambda^2}{(\lambda+2)(\lambda+1)^2} = \frac{\lambda}{(\lambda+2)(\lambda+1)^2}.$</td><td>M1
A1
M1
A1</td><td>Correct form of integral for E(X^2). Allow c’s expression for <i>c</i>.
Use of Var(<i>X</i>) = E(X^2) – E(X)^2. Allow c’s E(X^2) and E(<i>X</i>).
Algebra shown convincingly. Beware printed answer.</td><td>4</td></tr>
<tr><td>(b)</td><td>
<table>
<tr><th>Times</th><th>− 32</th><th>Rank of |diff|</th></tr>
<tr><td>40</td><td>8</td><td>4</td></tr>
<tr><td>20</td><td>−12</td><td>7</td></tr>
<tr><td>18</td><td>−14</td><td>8</td></tr>
<tr><td>11</td><td>−21</td><td>12</td></tr>
<tr><td>47</td><td>15</td><td>9</td></tr>
<tr><td>36</td><td>4</td><td>2</td></tr>
<tr><td>38</td><td>6</td><td>3</td></tr>
<tr><td>35</td><td>3</td><td>1</td></tr>
<tr><td>22</td><td>−10</td><td>5</td></tr>
<tr><td>14</td><td>−18</td><td>10</td></tr>
<tr><td>12</td><td>−20</td><td>11</td></tr>
<tr><td>21</td><td>−11</td><td>6</td></tr>
</table>
$W_+ = 1 + 2 + 3 + 4 + 9 = 19$
Refer to Wilcoxon single sample tables for <i>n</i> = 12.
Lower (or upper if 59 used) 5% tail is 17 (or 61 if 59 used).
Result is not significant.
Seems that there is no evidence that Godfrey’s times have decreased.</td><td>M1
M1
A1
B1
M1
A1
A1
A1</td><td>H_0: $m = 32$, H_1: $m < 32$, where <i>m</i> is the population median time.
for subtracting 32.
for ranks. ft if ranks wrong.
(or $W_- = 5 + 6 + 7 + 8 + 10 + 11 + 12 = 59$)
No ft from here if wrong.
i.e. a 1-tail test. No ft from here if wrong.
ft only c’s test statistic.
ft only c’s test statistic.</td><td>8</td></tr>
<tr><td></td><td></td><td></td><td></td><td>18</td></tr>
</table>

Q2	$V_G \sim \mathrm{N}(56.5,\ 2.9^2)$ $V_W \sim \mathrm{N}(38.4,\ 1.1^2)$		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)	$\mathrm{P}(V_G < 60) = \mathrm{P}(Z < \frac{60-56.5}{2.9} = 1.2069)$ $= 0.8862$	M1 A1 A1	For standardising. Award once, here or elsewhere.	3
(ii)	$V_T \sim \mathrm{N}(56.5 + 38.4 = 94.9,$ $2.9^2 + 1.1^2 = 9.62)$ $\mathrm{P}(\text{this} > 100) = \mathrm{P}(Z > \frac{100-94.9}{3.1016} = 1.6443)$ $= 1 - 0.9499 = 0.0501$	B1 B1 A1	Mean. Variance. Accept sd (= 3.1016). c.a.o.	3
(iii)	$W_T \sim \mathrm{N}(3.1 \times 56.5 + 0.8 \times 38.4 = 205.87,$ $3.1^2 \times 2.9^2 + 0.8^2 \times 1.1^2 = 81.5945)$ $\mathrm{P}(200 < \text{this} < 220)$ $= \mathrm{P}(\frac{200-205.87}{9.0330} < Z < \frac{220-205.87}{9.0330})$ $= \mathrm{P}(-0.6498 < Z < 1.5643)$ $= 0.9411 - (1 - 0.7422) = 0.6833$	M1 A1 M1 A1 M1 A1	Use of "mass = density × volume" Mean. Variance. Accept sd (= 9.0330). Formulation of requirement. c.a.o.	6
(iv)	Given $\bar{x} = 205.6 \quad s_{n-1} = 8.51$ H_0: $\mu = 200$, H_1: $\mu > 200$ Test statistic is $\frac{205.6 - 200}{\frac{8.51}{\sqrt{10}}}$ $= 2.081.$ Refer to t_9. Single-tailed 5% point is 1.833. Significant. Seems that the required reduction of the mean weight has not been achieved.	M1 A1 M1 A1 A1 A1	Allow alternative: 200 + (c's 1.833) × $\frac{8.51}{\sqrt{10}}$ (= 204.933) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ – (c's 1.833) × $\frac{8.51}{\sqrt{10}}$ (= 200.667) for comparison with 200.) c.a.o. but ft from here in any case if wrong. Use of 200 – $\bar{x}$ scores M1A0, but ft. No ft from here if wrong. $\mathrm{P}(t > 2.081) = 0.0336$. No ft from here if wrong. ft only c's test statistic. ft only c's test statistic.	6
				18

Q3				
(i)	In this situation a paired test is appropriate because there are clearly differences between specimens … … which the pairing eliminates.	E1 E1		2
(ii)	H$_0$: $\mu_D = 0$ H$_1$: $\mu_D > 0$	B1	Both. Accept alternatives e.g. $\mu_D < 0$ for H$_1$, or $\mu_A - \mu_B$ etc provided adequately defined. Hypotheses in words only must include "population".	
	Where μ_D is the (population) mean reduction in hormone concentration.	B1	For adequate verbal definition. Allow absence of "population" if correct notation μ is used, but do NOT allow "$\overline{X} = ...$" or similar unless $\overline{X}$ is clearly and explicitly stated to be a <u>population</u> mean.	
	Must assume • Sample is random • Normality of differences	 B1 B1		4
(iii)	<u>MUST</u> be PAIRED COMPARISON t test. Differences (reductions) (before – after) are		Allow "after – before" if consistent with alternatives above.	
	–0.75 2.71 2.59 6.07 0.71 –1.85 –0.98 3.56 1.77 2.95 1.59 4.17 0.38 0.88 0.95			
	$\bar{x} = 1.65$ $s_{n-1} = 2.100(3)$ $(s_{n-1}{}^2 = 4.4112)$	B1	Do not allow $s_n = 2.0291$ ($s_n{}^2 = 4.1171$)	
	Test statistic is $\dfrac{1.65-0}{\frac{2.100}{\sqrt{15}}}$	M1	Allow c's $\bar{x}$ and/or s_{n-1}. Allow alternative: 0 + (c's 2.624) × $\frac{2.100}{\sqrt{15}}$ (= 1.423) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ – (c's 2.624) × $\frac{2.100}{\sqrt{15}}$ (= 0.227) for comparison with 0.)	
	= 3.043.	A1	c.a.o. but ft from here in any case if wrong. Use of $0 - \bar{x}$ scores M1A0, but ft.	
	Refer to t_{14}.	M1	No ft from here if wrong. P(t > 3.043) = 0.00438.	
	Single-tailed 1% point is 2.624.	A1	No ft from here if wrong.	
	Significant.	A1	ft only c's test statistic.	
	Seems mean concentration of hormone has fallen.	A1	ft only c's test statistic.	7
(iv)	CI is 1.65 ±	M1	ft c's $\bar{x}$ ±.	
	$k \times \dfrac{2.100}{\sqrt{15}}$	M1	ft c's s_{n-1}.	
	= (0.4869, 2.8131)	A1	A correct equation in k using either end of the interval or the width of the interval.	
	∴ $k = 2.145$	A1	Allow ft c's $\bar{x}$ and s_{n-1}.	
	By reference to t_{14} tables this is a 95% CI.	A1	c.a.o.	5
				18

Q4				
(i)	Sampling which selects from those that are (easily) available. Circumstances may mean that it is the only economically viable method available. Likely to be neither random nor representative.	E1 E1 E1		3
(ii)	$p + pq + pq^2 + pq^3 + pq^4 + pq^5 + q^6$ $= \frac{p(1-q^6)}{1-q} + q^6 = \frac{p(1-q^6)}{p} + q^6$ $= 1 - q^6 + q^6 = 1$	M1 A1	Use of GP formula to sum probabilities, or expand in terms of p or in terms of q. Algebra shown convincingly. Beware answer given.	2
(iii)	With p = 0.25			

Probability	0.25	0.1875	0.140625	0.105469	0.079102	0.059326	0.177979
Expected fr	25.00	18.75	14.0625	10.5469	7.9102	5.9326	17.7979

		M1 M1 A1	Probabilities correct to 3 dp or better. × 100 for expected frequencies. All correct and sum to 100.	
	X^2 = 0.04 + 0.0033 + 0.6136 + 0.5706 + 1.2069 + 0.7204 + 7.8206 = 10.97(54)	M1 A1	c.a.o.	
	(If e.g. only 2dp used for expected f's then X^2 = 0.04 + 0.0033 + 0.6148 + 0.5690 + 1.2071 + 0.7226 + 7.8225 = 10.97(93))			
	Refer to χ_6^2.	M1	Allow correct df (= cells – 1) from wrongly grouped table and ft. Otherwise, no ft if wrong. P(X^2 > 10.975) = 0.0891.	
	Upper 10% point is 10.64. Significant. Suggests model with p = 0.25 does not fit.	A1 A1 A1	No ft from here if wrong. ft only c's test statistic. ft only c's test statistic.	9
(iv)	Now with X^2 = 9.124 Refer to χ_5^2.	M1	Allow correct df (= cells – 2) from wrongly grouped table and ft. Otherwise, no ft if wrong. P(X^2 > 9.124) = 0.1042.	
	Upper 10% point is 9.236. Not significant. (Suggests new model does fit.) Improvement to the model is due to estimation of p from the data.	A1 A1 E1	No ft from here if wrong. Correct conclusion. Comment about the effect of estimated p, consistent with conclusion in part (iii).	4
				18

4768 Statistics 3

Q1	W ~ N(14, 0.552) $G \sim \text{N}(144,\ 0.9^2)$		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(G < 145) = $\text{P}\left(Z < \frac{145-144}{0.9} = 1.1111\right)$	M1 A1	For standardising. Award once, here or elsewhere.	
	= 0·8667	A1	c.a.o.	3
(ii)	$W + G \sim \text{N}(14 + 144 = 158,$	B1	Mean.	
	$\sigma^2 = 0.55^2 + 0.9^2 = 1.1125\ )$	B1	Variance. Accept sd (= 1.0547…).	
	P(this > 160) = $\text{P}\left(Z > \frac{160-158}{1.0547} = 1\cdot 896\right) = 1 - 0\cdot 9710 = 0.0290$	A1	c.a.o.	3
(iii)	$H = W_1 + \ldots + W_7 + G_1 + \ldots + G_6 \sim \text{N}(962,$	B1	Mean.	
	$\sigma^2 = 0.55^2 + \ldots + 0.55^2 + 0.9^2 + \ldots + 0.9^2 = 6.9775)$	B1	Variance. Accept sd (= 2.6415).	
	P(960 < this < 965) = $\text{P}\left(\frac{960-962}{2\cdot 6415} = -0.7571 < Z < \frac{965-962}{2\cdot 6415} = 1.1357\right)$	M1	Two-sided requirement.	
	= 0·8720 – (1 – 0.7755) = 0.6475	A1	c.a.o.	
	Now want P(B(4, 0.6475) ≥ 3)	M1	Evidence of attempt to use binomial. ft c's p value.	
	= 4 × 0.6475^3 × 0.3525 + 0.6475^4	M1	Correct terms attempted. ft c's p value. Accept 1 – P(… ≤ 2)	
	= 0.38277 + 0.17577 = 0.5585	A1	c.a.o.	7
(iv)	$D = H_1 - H_2 \sim \text{N}(0,$	B1	Mean. (May be implied.)	
	$6.9775 + 6.9775 = 13.955)$	B1	Variance. Accept sd (= 3.7356). Ft 2 × c's 6.9775 from (iii).	
	Want h s.t. P(−h < D < h) = 0.95	M1	Formulation of requirement as 2-sided.	
	i.e. P(D < h) = 0 975 $\therefore h = \sqrt{13.955} \times 1.96 \quad = 7.32$	B1 A1	For 1.96. c.a.o.	5
				18

Q2				
(i)	H_0: $\mu = 1$ H_1: $\mu < 1$	B1	Both hypotheses. Hypotheses in words only must include "population".	
	where μ is the mean weight of the cakes.	B1	For adequate verbal definition. Allow absence of "population" if correct notation μ is used, but do NOT allow " $\bar{X} = \ldots$ " or similar unless $\bar{X}$ is clearly and explicitly stated to be a population mean.	
	$\bar{x} = 0.957375 \quad s_{n-1} = 0.07314(55)$	B1	s_n = 0.06842 but do NOT allow this here or in construction of test statistic, but FT from there.	
	Test statistic is $\dfrac{0.957375 \ -1}{\frac{0.07314}{\sqrt{8}}}$	M1	Allow c's $\bar{x}$ and/or s_{n-1}. Allow alternative: 1 + (c's – 1.895) × $\frac{0.07314}{\sqrt{8}}$ (= 0.950997) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ – (c's –1.895) × $\frac{0.07314}{\sqrt{8}}$ (= 1.006377) for comparison with 1.)	
	= –1.648(24).	A1	c.a.o. but ft from here in any case if wrong. Use of 1 – $\bar{x}$ scores M1A0, but ft.	
	Refer to t_7.	M1	No ft from here if wrong. P(t < –1.648(24)) = 0.0716.	
	Single-tailed 5% point is –1.895.	A1	Must be minus 1.895 unless absolute values are being compared. No ft from here if wrong.	
	Not significant.	A1	ft only c's test statistic.	
	Insufficient evidence to suggest that the cakes are underweight on average.	A1	ft only c's test statistic.	9
(ii)	CI is given by 0.957375 ± 2·365 $\times \frac{0.07314}{\sqrt{8}}$	M1 B1 M1		
	= 0.957375 ± 0.061156= (0.896(2), 1.018(5))	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_7 is OK.	4
(iii)	$\bar{x} \pm 1.96 \times \sqrt{\frac{0.006}{n}}$	M1 B1 A1	Structure correct, incl. use of Normal. 1.96.	 3

			All correct.	
(iv)	$2 \times 1.96 \times \sqrt{\frac{0.006}{n}} < 0.025$ $n > \left(\frac{2 \times 1.96}{0.025}\right)^2 \times 0.006 = 147.517$ So take n = 148	M1 M1 A1	Set up appropriate inequation. Condone an equation. Attempt to rearrange and solve. c.a.o. (expressed as an integer). S.C. Allow max M1A1(c.a.o.) when the factor “2” is missing. (n > 36.879)	3
				19

Q3				
(i)	For a systematic sample			
	• she needs a list of all staff	E1		
	• with no cycles in the list.	E1		
	All staff equally likely to be chosen if she			
	• chooses a random start between 1 and 10	E1		
	• then chooses every 10th.	E1		
	Not simple random sampling since not all samples are possible.	E1		5
(ii)	Nothing is known about the background population ..	E1	Any reference to unknown distribution or "non-parametric" situation.	
	... of differences between the scores.	E1	Any reference to pairing/differences.	
	H_0: $m = 0$ H_1: $m \neq 0$	B1	Both hypotheses. Hypotheses in words only must include "population".	
	where m is the population median difference for the scores.	B1	For adequate verbal definition.	4
(iii)				

Diff	−0.8	−2.6	8.6	6.2	6.0	−3.6	−2.4	−0.4	−4.0	5.6	6.6	2.2
Rank	2	5	12	10	9	6	4	1	7	8	11	3

		M1	For differences. ZERO in this section if differences not used.	
		M1	For ranks.	
		A1	ft from here if ranks wrong.	
	$W_- = 1 + 2 + 4 + 5 + 6 + 7 = 25$	B1	(or $W_+ = 3 + 8 + 9 + 10 + 11 + 12 = 53$)	
	Refer to tables of Wilcoxon paired (/single sample) statistic for $n = 12$.	M1	No ft from here if wrong.	
	Lower (or upper if 53 used) 2½% tail is 13 (or 65 if 53 used).	A1	i.e. a 2-tail test. No ft from here if wrong.	
	Result is not significant.	A1	ft only c's test statistic.	
	No evidence to suggest a preference for one of the uniforms.	A1	ft only c's test statistic.	8
				17

<table>
<tr><td>Q4</td><td>$\mathrm{f}(x) = \frac{2x}{\lambda^2}$ for $0 < x < \lambda$, $\lambda > 0$</td><td></td><td></td><td></td></tr>
<tr><td>(i)</td><td>f(x) > 0 for all x in the domain.
$\int_0^\lambda \frac{2x}{\lambda^2}\mathrm{d}x = \left[\frac{x^2}{\lambda^2}\right]_0^\lambda = \frac{\lambda^2}{\lambda^2} = 1$</td><td>E1
M1
A1</td><td>
Correct integral with limits.
Shown equal to 1.</td><td>3</td></tr>
<tr><td>(ii)</td><td>$\mu = \int_0^\lambda \frac{2x^2}{\lambda^2}\mathrm{d}x = \left[\frac{2x^3/3}{\lambda^2}\right]_0^\lambda = \frac{2\lambda}{3}$
$\mathrm{P}(X < \mu) = \int_0^\mu \frac{2x}{\lambda^2}\mathrm{d}x = \left[\frac{x^2}{\lambda^2}\right]_0^\mu$
$= \frac{\mu^2}{\lambda^2} = \frac{4\lambda^2/9}{\lambda^2} = \frac{4}{9}$
which is independent of λ.</td><td>M1
A1
M1
A1</td><td>Correct integral with limits.
c.a.o.
Correct integral with limits.
Answer plus comment. ft c's μ provided the answer does not involve λ.</td><td>4</td></tr>
<tr><td>(iii)</td><td>Given $\mathrm{E}(X^2) = \frac{\lambda^2}{2}$
$\sigma^2 = \frac{\lambda^2}{2} - \frac{4\lambda^2}{9} = \frac{\lambda^2}{18}$</td><td>M1
A1</td><td>Use of Var(X) = E(X^2) – E(X)^2.
c.a.o.</td><td>2</td></tr>
<tr><td>(iv)</td><td colspan="3">
<table>
<tr><td>Probability</td><td>0.18573</td><td>0.25871</td><td>0.36983</td><td>0.18573</td></tr>
<tr><td>Expected f</td><td>9.2865</td><td>12.9355</td><td>18.4915</td><td>9.2865</td></tr>
</table>
</td><td></td></tr>
<tr><td></td><td>

X^2 = 3.0094 + 0.2896 + 0.1231 + 3.5152
= 6.937(3)</td><td>M1
A1

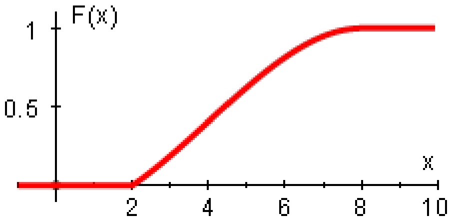
M1
A1</td><td>Probs × 50 for expected frequencies.
All correct.
Calculation of X^2.
c.a.o.</td><td></td></tr>
<tr><td></td><td>Refer to χ_3^2.</td><td>M1</td><td>Allow correct df (= cells – 1) from wrongly grouped table and ft. Otherwise, no ft if wrong.
P(X^2 > 6.937) = 0.0739.</td><td></td></tr>
<tr><td></td><td>Upper 5% point is 7.815.
Not significant.
Suggests model fits the data for these jars.
But with a 10% significance level (cv = 6.251) a different conclusion would be reached.</td><td>A1
A1
A1
E1</td><td>No ft from here if wrong.
ft only c's test statistic.
ft only c's test statistic.
Any valid comment which recognises that the test statistic is close to the critical values.</td><td>9</td></tr>
<tr><td></td><td></td><td></td><td></td><td>18</td></tr>
</table>

4768 Statistics 3

1 (i)	H_0: The number of eggs hatched can be modelled by B(3, ½)	B1		
	H_1: The number of eggs hatched cannot be modelled by B(3, ½)	B1		
	With $p = ½$			

Probability	0.125	0.375	0.375	0.125
Exp'd frequency	10	30	30	10
Obs'd frequency	7	23	29	21

		M1	Probs × 80 for expected frequencies.	
		A1	All correct.	
	$X^2 = 0.9 + 1.6333 + 0.0333 + 12.1$	M1	Calculation of X^2.	
	$= 14.666(7)$	A1	c.a.o.	
	Refer to χ_3^2.	M1	Allow correct df (= cells – 1) from wrongly grouped table and ft. Otherwise, no ft if wrong. $P(X^2 > 14.667) = 0.00212$.	
	Upper 5% point is 7.815.	A1	No ft from here if wrong.	
	Significant.	A1	ft only c's test statistic.	
	Suggests it is reasonable to suppose model with $p = ½$ does not apply.	A1	ft only c's test statistic.	**[10]**
(ii)	$\bar{x} = \frac{144}{80} = 1.8$	B1	C.a.o.	
	$\therefore \hat{p} = \frac{1.8}{3} = 0.6$	B1	Use of $E(X) = np$. ft c's mean, provided $0 < \hat{p} < 1$.	**[2]**
(iii)	Refer to χ_2^2.	M1	Allow df 1 less than in part (i). No ft if wrong.	
	Upper 5% point is 5.991.	A1	No ft if wrong.	
	Suggests it is reasonable to suppose model with estimated p does apply.	A1	ft provided previous A mark awarded.	**[3]**
(iv)	For example: Estimating p leads to an improved fit … … at the expense of the loss of 1 degree of freedom. The model in (i) fails due to a large underestimate for $X = 3$.	E2	Reward any two sensible points for E1 each.	**[2]**
			Total	**[17]**

2 (a)	$\mathrm{f}(x)=\frac{1}{72}\left(8x-x^2\right)\ ,\ 2\le x\le 8$			
(i)	$\mathrm{F}(x)=\int_2^x \frac{1}{72}\left(8t-t^2\right)\mathrm{d}t$	M1	Correct integral with limits (which may be implied subsequently).	
	$=\frac{1}{72}\left[4t^2-\frac{t^3}{3}\right]_2^x$	A1	Correctly integrated	
	$=\frac{1}{72}\left(4x^2-\frac{x^3}{3}-16+\frac{8}{3}\right)=\frac{12x^2-x^3-40}{216}$	A1	Limits used. Accept unsimplified form.	**[3]**
(ii)		G1	Correct shape; nothing below y = 0; non-negative gradient.	
		G1	Labels at (2, 0) and (8, 1).	
		G1	Curve (horizontal lines) shown for x < 2 and x > 8.	**[3]**
(iii)	F(m) = ½ $\quad\therefore \frac{12m^2-m^3-40}{216}=\frac{1}{2}$	M1	Use of definition of median. Allow use of c's F(x).	
	$\therefore 12m^2-m^3-40=108$ $\therefore m^3-12m^2+148=0$	A1	Convincingly rearranged. Beware: answer given.	
	Either F(4.42) = 0.5003(977) ≈ 0.5 Or $4.42^3-12\times 4.42^2+148=-0.0859(12)\approx 0$ $\therefore m\approx 4.42$	E1	Convincingly shown, e.g. 4.418 or better seen.	**[3]**

<table>
<tr><td>2 (b)</td><td>H_0: $m = 4.42$ H_1: $m \neq 4.42$
where m is the population median</td><td>B1
B1</td><td>Both. Accept hypotheses in words.
Adequate definition of m to include “population”.</td><td></td></tr>
<tr><td></td><td>
<table>
<tr><th>Weights</th><th>− 4.42</th><th>Rank of |diff|</th></tr>
<tr><td>3.16</td><td>−1.26</td><td>7</td></tr>
<tr><td>3.62</td><td>−0.80</td><td>6</td></tr>
<tr><td>3.80</td><td>−0.62</td><td>4</td></tr>
<tr><td>3.90</td><td>−0.52</td><td>3</td></tr>
<tr><td>4.02</td><td>−0.40</td><td>2</td></tr>
<tr><td>4.72</td><td>0.30</td><td>1</td></tr>
<tr><td>5.14</td><td>0.72</td><td>5</td></tr>
<tr><td>6.36</td><td>1.94</td><td>8</td></tr>
<tr><td>6.50</td><td>2.08</td><td>9</td></tr>
<tr><td>6.58</td><td>2.16</td><td>10</td></tr>
<tr><td>6.68</td><td>2.26</td><td>11</td></tr>
<tr><td>6.78</td><td>2.36</td><td>12</td></tr>
</table>
</td><td>M1
M1
A1</td><td>for subtracting 4.42.
for ranks.
ft if ranks wrong.</td><td></td></tr>
<tr><td></td><td>$W_- = 2 + 3 + 4 + 6 + 7 = 22$</td><td>B1</td><td>($W_+ = 1 + 5 + 8 + 9 + 10 + 11 + 12 = 56$)</td><td></td></tr>
<tr><td></td><td>Refer to Wilcoxon single sample tables for $n = 12$.</td><td>M1</td><td>No ft from here if wrong.</td><td></td></tr>
<tr><td></td><td>Lower 2½% point is 13 (or upper is 65 if 56 used).</td><td>A1</td><td>i.e. a 2-tail test. No ft from here if wrong.</td><td></td></tr>
<tr><td></td><td>Result is not significant.</td><td>A1</td><td>ft only c’s test statistic.</td><td></td></tr>
<tr><td></td><td>Evidence suggests that a median of 4.42 is consistent with these data.</td><td>A1</td><td>ft only c’s test statistic.</td><td>[10]</td></tr>
<tr><td></td><td></td><td></td><td>Total</td><td>[19]</td></tr>
</table>

3 (i)	Must assume			
	• Normality of population …	B1		
	• … of <u>differences</u>.	B1		
	H_0: $\mu_D = 0$ H_1: $\mu_D > 0$	B1	Both. Accept alternatives e.g. $\mu_D < 0$ for H_1, or $\mu_B - \mu_A$ etc provided adequately defined. 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_D is the (population) mean reduction/difference in cholesterol level.	B1	For adequate verbal definition. Allow absence of "population" if correct notation μ is used.	
	<u>MUST</u> be PAIRED COMPARISON *t* test. Differences (reductions) (before – after) are: –0.1 1.7 –1.2 1.1 1.4 0.5 0.9 2.2 –0.1 2.0 0.7 0.3		Allow "after – before" if consistent with alternatives above.	
	$\bar{x} = 0.7833$ $s_{n-1} = 0.9833(46)$ $(s_{n-1}^{2} = 0.966969)$	B1	Do not allow $s_n = 0.9415$ ($s_n^2 = 0.8864$)	
	Test statistic is $\dfrac{0.7833 - 0}{\frac{0.9833}{\sqrt{12}}}$	M1	Allow c's $\bar{x}$ and/or s_{n-1}. Allow alternative: 0 + (c's 2.718) × $\frac{0.9833}{\sqrt{12}}$ (= 0.7715) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ – (c's 2.718) × $\frac{0.9833}{\sqrt{12}}$ (= 0.0118) for comparison with 0.)	
	= 2.7595.	A1	c.a.o. but ft from here in any case if wrong. Use of 0 – $\bar{x}$ scores M1A0, but ft.	
	Refer to t_{11}.	M1	No ft from here if wrong. P(t > 2.7595) = 0.009286.	
	Single-tailed 1% point is 2.718.	A1	No ft from here if wrong.	
	Significant.	A1	ft only c's test statistic.	
	Seems mean cholesterol level has fallen.	A1	ft only c's test statistic.	**[11]**
(ii)	CI is $\bar{x} \pm$	M1	Overall structure, seen or implied.	
	2.201	B1	From t_{11}, seen or implied.	
	$\times \dfrac{s}{\sqrt{12}} = (-0.5380, 1.4046)$	A1	Fully correct pair of equations using the given interval, seen or implied.	
	$\bar{x} = ½(1.4046 - 0.5380) = 0.4333$	B1		
	$s = (1.4046 - 0.4333) \times \dfrac{\sqrt{12}}{2.201} = 1.5287$	M1 A1	Substitute $\bar{x}$ and rearrange to find s. c.a.o.	
	Using this interval the doctor might conclude that the mean cholesterol level did not seem to have been reduced.	E1	Accept any sensible comment or interpretation of <u>this</u> interval.	**[7]**
			Total	**[18]**

4	$A \sim N(80,\ \sigma = 11)$ $B \sim N(70,\ \sigma = v)$		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)	$P(A < 90) = P\left(Z < \frac{90-80}{11} = 0.9091\right)$ $= 0.8182$	M1 A1 A1	For standardising. Award once, here or elsewhere. c.a.o.	**[3]**
(ii)	$W_B = B_1 + B_2 + \ldots + B_6 + 15 \sim N(435,$ $\sigma^2 = v^2 + v^2 + \ldots + v^2 = 6v^2)$ $P(\text{this} < 450) = P\left(Z < \frac{450-435}{v\sqrt{6}}\right) = 0.8463$ $\therefore \frac{450-435}{v\sqrt{6}} = \Phi^{-1}(0.8463) = 1.021$ $\therefore v = \frac{15}{1.021 \times \sqrt{6}} = 5.9977 = 6$ grams (nearest gram)	B1 B1 M1 B1 A1	Mean. Expression for variance. Formulation of the problem. Inverse Normal. Convincingly shown, beware A.G.	**[5]**
(iii)	$W_A = A_1 + A_2 + \ldots + A_5 + 25 \sim N(425,$ $\sigma^2 = 11^2 + 11^2 + \ldots + 11^2 = 605\ \ )$ $D = W_A - W_B \sim N(-10,$ $605 + 216 = 821)$ Want P($W_A > W_B$) = P($W_A - W_B > 0$) $= P\left(Z > \frac{0-(-10)}{\sqrt{821}} = 0.3490\right) = 1 - 0.6365 = 0.3635$	B1 M1 A1 M1 A1	Mean. Accept "$B - A$". Variance. Accept sd (= 28.65). c.a.o.	**[5]**
(iv)	$\bar{x} = \frac{3126.0}{60} = 52.1,$ $s = \sqrt{\frac{164223.96 - 60 \times 52.1^2}{59}} = 4.8$ CI is given by 52.1 ± 1.96 $\times \frac{4.8}{\sqrt{60}}$ = 52.1 ± 1.2146= (50.885(4), 53.314(6))	B1 M1 B1 M1 A1	Both correct. c.a.o. Must be expressed as an interval.	**[5]**
			Total	**[18]**

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

Q2				
(i)	A t test might be used because			
	• sample is small	B1		
	• population variance is unknown	B1		
	Must assume background population is Normal.	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 “ $\bar{X} = \dots$ ” or similar unless $\bar{X}$ 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.	
	$\bar{x} = 1.0452$ $s_{n-1} = 0.007155$	B1	s_n = 0.006746 but do NOT allow this here or in construction of test statistic, but FT from there.	
	Test statistic is $\dfrac{1.0452 - 1.040}{\dfrac{0.007155}{\sqrt{9}}}$	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 ±	M1		
	2·306	B1		
	$\times \dfrac{0.007155}{\sqrt{9}}$	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]**

<table>
<tr><td>Q3</td><td></td><td></td><td></td></tr>
<tr><td>(a)
(i)</td><td>Use paired data in order to eliminate differences between authorities.</td><td>B1</td><td>[1]</td></tr>
<tr><td>(ii)</td><td>H_0: $m = 0$ H_1: $m > 0$
where m is the population median difference.
<table>
<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>
$W_- = 1 + 3 + 4 = 8$ (or = 2+5+6+7+8+9 = 37)
Refer to tables of Wilcoxon paired (/single sample) statistic for $n = 9$.
Lower 5% point is 8 (or upper is 37 if W_+ used).
Result is significant.
Evidence suggests the percentage has been raised (on the whole).</td>
<td>B1 Both. Accept hypotheses in words.
B1 Adequate definition of m to include “population”.
M1 For differences. ZERO in this section if differences not used.
M1 For ranks.
A1 FT from here if ranks wrong
B1
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.</td><td>[10]</td></tr>
<tr><td>(b)</td><td>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.
<table>
<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>
$X^2 = 0.44917 + 0.04091 + 0.16 + 0.59588 + 0.04051 + 0.96716 + 0.01379 + 2.25882 + 0.00435 = 4.5305(9)$
Refer to χ_8^2.
Upper 5% point is 13.36.
Not significant.
Suggests Benford’s Law provides a reasonable model in the context of share prices.</td>
<td>M1 Probs × 200 for expected frequencies. All correct.
M1 Calculation of X^2.
A1 c.a.o.
M1 Allow correct df (= cells – 1) from wrongly grouped table and ft. Otherwise, no ft if wrong. $P(X^2 > 4.53059) = 0.80636$.
A1 No ft from here if wrong.
A1 ft only c’s test statistic.
A1 ft only c’s test statistic.</td><td>[7]</td></tr>
<tr><td></td><td></td><td>Total</td><td>[18]</td></tr>
</table>

Q4	$f(x) = \lambda e^{-\lambda x}$ for $x \ge 0$, where $\lambda > 0$.	Given $\int_0^\infty x^r e^{-\lambda x} \mathrm{d}x = \dfrac{r!}{\lambda^{r+1}}$	
(i)	$\int_0^\infty f(x)\mathrm{d}x = \int_0^\infty \lambda e^{-\lambda x}\mathrm{d}x$ $= \left[-e^{-\lambda x}\right]_0^\infty$ $= \left(0 - (-e^0)\right) = 1$ 	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, λ) labelled; asymptotic to x-axis.	**[5]**
(ii)	$E(X) = \int_0^\infty \lambda x e^{-\lambda x}\mathrm{d}x$ $= \lambda \dfrac{1}{\lambda^2} = \dfrac{1}{\lambda}$ $E(X^2) = \int_0^\infty \lambda x^2 e^{-\lambda x}\mathrm{d}x$ $= \lambda \dfrac{2}{\lambda^3} = \dfrac{2}{\lambda^2}$ Var(X) = E(X^2) − E(X)^2 = $\dfrac{2}{\lambda^2} - \left(\dfrac{1}{\lambda}\right)^2 = \dfrac{1}{\lambda^2}$	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)	$\mu = 6 \quad \therefore \lambda = \dfrac{1}{6}$ $\bar{X}$ ~ (approx) N$\left(6, \dfrac{6^2}{50}\right)$	B1 Obtained λ from the mean. B1 Normal. B1 Mean. ft c's λ. B1 Variance. ft c's λ.	**[4]**
(iv)	**EITHER** 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 ($\bar{\mathrm{X}}$), not SD(X)) i.e. outlier. ⇒ doubt. **OR** formal significance test: $\dfrac{7.8-6}{\sqrt{0.72}}$ = 2.121, refer to N(0,1), sig at (eg) 5% ⇒ 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]**

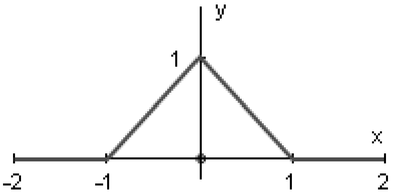
Q1	$E \sim N(406,\ 12^2)$ When a candidate's answers suggest that (s)he appears to have neglected to use the difference columns of the Normal distribution tables penalise the first occurrence only.			
(i)	$P(E < 420) = P\left(Z < \frac{420-406}{12} = 1.1666\right)$	M1 A1	For standardising. Award once, here or elsewhere.	
	$= 0\cdot8783/4$	A1	c.a.o.	3
(ii)	$C \sim N(406\times14.6 = 5927.6,$ $\sigma^2 = 12^2\times14.6^2 = 30695.04)$	 B1 B1	Accept equivalent in £. Mean. Variance. Accept sd (= 175.2).	
	P(this > 6000) = $P\left(Z > \frac{6000-5927.6}{175.2} = 0.4132\right) = 1-0\cdot6602 = 0.3398$	A1	Accept P(E > 6000/14.6) o.e. c.a.o.	3
(iii)	$B = C_1 + C_2 + C_3 \sim N(17782.8,$ $\sigma^2 = 175.2^2 + 175.2^2 + 175.2^2 = 92085.12)$	 B1 B1	Accept equivalent in £, or $E_1 + E_2 + E_3$. Mean. ft from (ii). Variance. Accept sd (= 303.455…). ft from (ii).	
	Require b s.t. $P(B < 100b) = 0.99$ $\therefore \frac{100b-17782.8}{303.455} = 2.326$	B1	Accept $P(E_1 + E_2 + E_3 < 100b/14.6)$ o.e. 2.326 seen.	
	$\therefore 100b = 17782.8 + 2.326\times303.455 = 18488.6...$ (p) b = £184.89	A1	c.a.o. (Minimum 4 s.f. required in final answer.)	4
(iv)	H_0: $\mu = 432$ H_1: $\mu < 432$	B1	Both hypotheses. Hypotheses in words only must include "population".	
	where μ is the mean amount of electricity used.	B1	For adequate verbal definition. Allow absence of "population" if correct notation μ is used, but do NOT allow "$\bar{X} = ...$" or similar unless $\bar{X}$ is clearly and explicitly stated to be a population mean.	
	$\bar{x} = 422.16...$ $s_{n-1} = 13.075(4)$	B1	s_n = 11.936 but do NOT allow this here or in construction of test statistic, but FT from there.	
	Test statistic is $\frac{422.16-432}{\frac{13.075}{\sqrt{6}}}$	M1	Allow c's $\bar{x}$ and/or s_{n-1}. Allow alternative: 432 + (c's –2.015) × $13.075/\sqrt{6}$ (= 421.24) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ – (c's –2.015) × $13.075/\sqrt{6}$ (= 432.92) for comparison with 432.)	
	= –1.842(13).	A1	c.a.o. but ft from here in any case if wrong. Use of $\mu - \bar{x}$ scores M1A0.	
	Refer to t_5.	M1	No ft from here if wrong. $P(t < -1.842(13)) = 0.0624$.	
	Single-tailed 5% point is –2.015.	A1	Must be minus 2.015 unless absolute values are being compared. No ft from here if wrong.	
	Not significant.	A1	ft only c's test statistic.	
	Insufficient evidence to suggest that the amount of electricity used has decreased on average.	A1	ft only c's test statistic. Conclusion in context to include "on average" o.e.	9
				19

Q2				
(a) (i)	There are identifiable subgroups or strata that might exhibit different characteristics. Each stratum is randomly sampled. Use it to obtain a representative sample. Can get information on the individual strata.	E1 E1 E1 E1		4
(ii)	For each stratum $\ldots \times \frac{2000}{79368}$ giving 813.9, 836.9, 245.4, 103.8 so 814, 837, 245, 104	M1 A1	All correct.	2
(b) (i)	The population (or underlying distribution) is assumed to be symmetrical about its median.	E2	E2, 1, 0. Award E1 for 2 out of 3 of the key features.	2
(ii)	H_0: $m = 0$ H_1: $m \neq 0$ where m is the population median difference for the percentages.	B1 B1	Both hypotheses. Hypotheses in words only must include "population". For adequate verbal definition.	

Diff	−0.66	0.02	−0.80	−0.91	0.28	0.76	0.40	1.68	−0.07	1.12
Rank	5	1	7	8	3	6	4	10	2	9

		M1 M1 A1	For differences. ZERO (out of 8) in this section if paired differences not used. For ranks. ft from here if ranks wrong.	
	$W_- = 2 + 5 + 7 + 8 = 22$	B1	(or $W_+ = 1 + 3 + 4 + 6 + 9 + 10 = 33$)	
	Refer to tables of Wilcoxon paired (/single sample) statistic for $n = 10$.	M1	No ft from here if wrong.	
	Lower (or upper if 33 used) 5% tail is 10 (or 45 if 33 used).	A1	i.e. a 2-tail test. No ft from here if wrong.	
	Result is not significant.	A1	ft only c's test statistic.	
	No evidence to suggest a change in spending on average.	A1	ft only c's test statistic. Conclusion in context to include "on average" o.e.	10
				18

Q3				
(i)	Using mid- intervals 1.5, 1.7, etc $\bar{x} = \frac{205}{100} = 2.05$ $s = \sqrt{\frac{425.16 - 100 \times 2.05^2}{99}} = 0.2227(01...)$	M1 A1 E1	 Mean. s.d. Answer given; must show convincingly.	3
(ii)	$f = 100 \times \mathrm{P}(1.8 \le M < 2.0)$ $= 100 \times \mathrm{P}(-1.1226 \le z < -0.2245)$ $= 100 \times ((1 - 0.5888) - (1 - 0.8691))$ $= 100 \times (0.4112 - 0.1309) = 28.03$	M1 A1 A1	Probability × 100. Correct Normal probabilities. ft c's mean. Must show convincingly using Normal distribution. ft c's mean.	3
(iii)	H_0: The Normal model fits the data. H_1: The Normal model does not fit the data. $X^2 = 0.7294 + 0.1384 + 1.9623 + 3.5155 + 0.2437$ $= 6.589(3)$ Refer to χ_2^2. Upper 5% point is 5.991. Significant. Evidence suggests that the model does not fit the data.	B1 B1 M1 M1 A1 M1 A1 A1 A1	Ignore any reference to parameters. Merge first 2 and last 2 cells. Calculation of X^2. c.a.o. Allow correct df (= cells – 3) from wrongly grouped table and ft. Otherwise, no ft if wrong. $\mathrm{P}(X^2 > 6.589) = 0.0371$. No ft from here if wrong. ft only c's test statistic. ft only c's test statistic. Conclusion in context.	9
(iv)	The model • overestimates in the 2.2 – 2.4 class, • underestimates in the 2 – 2.2 class. At lower significance levels the test would not have been significant.	 E1 E1 E1		3
				18

Q4				
(i)		G1 G1 G1	One (straight) line segment correct. Second (straight) line segment correct. Fully labelled intercepts + no spurious other lines.	3
(ii)	E(*X*) = 0 (By symmetry.) $\text{E}(X^2) = \int_{-1}^{0} x^2(1+x)\text{d}x + \int_{0}^{1} x^2(1-x)\text{d}x$ $= \left[\frac{x^3}{3}+\frac{x^4}{4}\right]_{-1}^{0} + \left[\frac{x^3}{3}-\frac{x^4}{4}\right]_{0}^{1}$ $= 0 - \left(\frac{-1}{3}+\frac{1}{4}\right) + \left(\frac{1}{3}-\frac{1}{4}\right) - 0$ $= \frac{1}{6}$ $\therefore \text{Var}(X) = \frac{1}{6}\left(-0^2\right) = \frac{1}{6}$	B1 M1 M1 M1 A1	One correct integral with limits (which may be implied subsequently). Second integral correct (with limits) or allow use of symmetry. Correctly integrated and attempt to use limits. c.a.o. Condone absence of explicit evidence of use of Var(*X*) = E(X^2) – E(X)^2.	5
(iii)	$\bar{L} \sim N\left(k, \frac{1}{300}\right)$ Normal distribution because of the Central Limit Theorem.	B1 B1 B1 E1	Normal. Mean. Variance. ft c's variance in (ii) (> 0) / 50. Any reference to the CLT.	4
(iv)	CI is given by 90.06 ± 1.96 $\times \frac{1}{\sqrt{300}}$ = 90.06 ± 0.11316= (89.947, 90.173)	M1 B1 M1 A1	ft c's variance in (ii) (> 0) / 50. Must be expressed as an interval.	4
(v)	It is reasonable, because 90 lies within the interval found in (iv).	E1	Or equivalent.	1
				17

GCE

Mathematics (MEI)

Advanced GCE

Unit **4768:** Statistics 3

Mark Scheme for June 2011

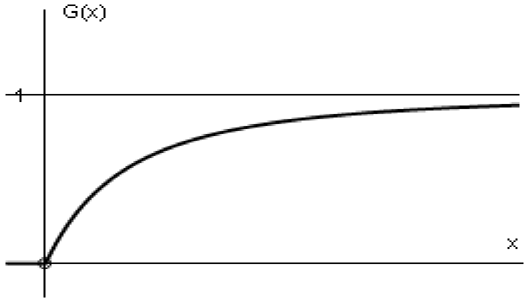
Oxford Cambridge and RSA Examinations

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 “ $\bar{X} = ...$ ” or similar unless $\bar{X}$ is clearly and explicitly stated to be a population 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.	
	$\bar{x} = 14.987$ $s_{n-1} = 0.4567(5)$	B1	$s_n = 0.4333$ but do NOT allow this here or in construction of test statistic, but FT from there.	
	Test statistic is $\frac{14.987 - 15.3}{\frac{0.45675}{\sqrt{10}}}$	M1	Allow c’s $\bar{x}$ and/or s_{n-1}. Allow alternative: 15.3 + (c’s –1.833) $\times\frac{0.45675}{\sqrt{10}}$ (= 15.035) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ – (c’s –1.833)$\times\frac{0.45675}{\sqrt{10}}$ (= 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 $\mu - \bar{x}$ scores M1A0, but ft.	
	Refer to t_9.	M1	No ft from here if wrong.	
	Single-tailed 5% point is –1.833.	A1	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.	A1	ft only c’s test statistic.	
	Seems that Gerry’s times have been reduced on average.	A1	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.	E1		
	Decreasing the significance level would make it less likely that a true H_0 would be rejected.	E1		
	Evidence for rejecting H_0 would need to be stronger.	E1	Or equivalent. Allow answers that relate to the context of the question.	3
(iv)	CI is given by 14.987 ±	M1	ZERO/4 if not same distribution as test. Same wrong distribution scores maximum M1B0M1A0. Recovery to t_9 is OK.	
	2·262	B1		
	$\times\frac{0.45675}{\sqrt{10}}$	M1		
	= 14.987 ± 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	
Expfr	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					

		M1 M1 A1	Probs correct to 3d.p. or better. × 100 for expected frequencies. All correct.	
	$X^2 = 1.3128 + 0.7843 + 0.1183 + 0.3063 + 0.1083 + 0.1813 + 0.6676 + 0.4056$ $= 3.884(5)$	M1 M1 A1	Merge first 2 cells. Calculation of X^2. c.a.o. (For ungrouped cells $X^2 = 6.816$.)	
	H_0: The Poisson model fits the data. H_1: The Poisson model does not fit the data.	B1 B1	Ignore any reference to the parameter. Do not accept "data fit model" oe.	
	Refer to χ_6^2 .	M1	Allow correct df (= cells – 2) from wrongly grouped table and ft. Otherwise, no ft if wrong.	
	Upper 10% point is 10.64.	A1	No ft from here if wrong. ($\chi_7^2 = 12.02$) $P(X^2 > 3.884) = 0.6924$.	
	Not significant. Evidence suggests that the model fits the data.	A1 A1	ft only c's test statistic. 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).	B1 B1	Both. Accept hypotheses in words. Adequate definition of m to include "population".	
	Given $W_- = 53$ ($\therefore$ $W_+ = 157$)			
	Refer to tables of Wilcoxon paired (/single sample) statistic for $n = 20$.	M1	No ft from here if wrong.	
	Lower 5% point is 60 (or upper is 150 if W_+ used).	A1	i.e. a 1-tail test. No ft from here if wrong.	
	Result is significant. Evidence suggests that the median diameter appears to be more than 15 μm.	A1 A1	ft only c's test statistic. 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 $\therefore \left(1+\frac{u}{200}\right)^{-2} = \frac{1}{4} \quad \therefore u = 200$ For the LQ G(l) = 0.25 $\therefore \left(1+\frac{l}{200}\right)^{-2} = \frac{3}{4} \quad \therefore l = 200\left(\frac{2}{\sqrt{3}}-1\right) = 30.94...$ $\therefore$IQR = 200 – 30.94 = 169(.06) For an outlier x > UQ +1.5 × IQR = 200 + 1.5 × 169 =453(.58) ≈ 454 (nearest hour)	M1 A1 A1 M1 M1 E1	Use of G(x) for either quartile. c.a.o. c.a.o. UQ – LQ UQ +1.5 × IQR. Answer given; must be obtained genuinely.	**6**
(ii) (A)	$F(x) = \int_0^x \frac{1}{200} e^{\frac{-t}{200}} dt$ $= \left[-e^{\frac{-t}{200}}\right]_0^x = \left(-e^{\frac{-x}{200}}\right) - \left(-e^{\frac{-0}{200}}\right) = 1 - e^{\frac{-x}{200}}$	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)	$P(X > 400) = e^{\frac{-400}{200}} = 0.1353(35)$ $P(X > 450) = e^{\frac{-450}{200}} = 0.1053(99)$ $P(X > 450 \mid X > 400) = \frac{P(X > 450)}{P(X > 400)}$ $= \frac{e^{\frac{-450}{200}}}{e^{\frac{-400}{200}}} = e^{\frac{-50}{200}} = e^{-0.25} \left(= 0.7788\right)$	B1 B1 M1 A1	Accept any form. Accept any form. Conditional probability. Not P(X > 50) × 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	$C \sim \mathrm{N}(10,\ 0.4^2)$, $D \sim \mathrm{N}(35,\ 3.5^2)$ 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)	$\mathrm{P}(C < 9.5) = \mathrm{P}\left(Z < \frac{9.5-10}{0.4} = -1.25\right)$	M1 A1	For standardising. Award once, here or elsewhere.	
	$= 1 - 0{\cdot}8944 = 0.1056$	A1	c.a.o.	**3**
(ii)	$D - S = D - (C_1 + C_2 + C_3 + C_4) \sim \mathrm{N}(-5,$	B1	Mean. Accept +5 for $S - D$.	
	$\sigma^2 = 3.5^2 + (0.4^2 + 0.4^2 + 0.4^2 + 0.4^2) = 12.89)$	B1	Variance. Accept sd (= 3.590…).	
	Want $\mathrm{P}(D > S) = \mathrm{P}(D - S > 0)$	M1	Formulation of requirement. Accept S – D < 0. This mark could be awarded in (iii) if not earned here.	
	$= 1 - \Phi\left(\frac{0-(-5)}{3.59} = 1.39(27)\right)$			
	$= 1 - 0.9182 = 0.0818$	A1	c.a.o.	**4**
(iii)	*New* $(D - S) = (D \times 1.3) - (C_1 + \ldots + C_5) \sim \mathrm{N}(-4.5,$	B1	Mean. Accept +4.5 for $S - D$.	
	$\sigma^2 = (3.5^2 \times 1.3^2) + (0.4^2 + \ldots + 0.4^2) = 21.5025)$	M1 A1	Correct use of $\times 1.3^2$ for variance. c.a.o. Accept sd (= 4.637…)	
	Again want $\mathrm{P}(D > S) = \mathrm{P}(D - S > 0)$		Or S – D < 0. M1 for formulation in (ii) available here.	
	$= 1 - \Phi\left(\frac{0-(-4.5)}{4.637} = 0.9704\right)$			
	$= 1 - 0.8341 = 0.1659$	A1	c.a.o.	**4**
(iv)	CI is given by $9.73 \pm$	M1		
	1.96	B1	1.96 seen.	
	$\times \frac{0.4}{\sqrt{12}}$	M1		
	$= 9.73 \pm 0.2263 = (9.50(37),\ 9.95(63))$	A1	c.a.o. Must be expressed as an interval.	
	Since 10 lies above this interval, it seems that the cheeses are underweight.	E1	Ft c's interval.	
	In repeated sampling, 95% of all confidence intervals	E1		
	constructed in this way will contain the true mean.	E1		**7**
				18

Question			Answer	Marks	Guidance
1	(i)		A paired sample is used in this context in order to eliminate any effects due to the surfaces used.	E1 [1]	Must refer to (differences between) surfaces.
1	(ii)		A t test might be used since … … the sample is small and … the population variance is not known (it must be estimated from the data). Must assume: Normality of population … … of differences.	E1 E1 B1 B1 [4]	Allow use of "σ", otherwise insist on "population". Allow "underlying" or "distribution" to imply "population".
1	(iii)		H$_0$: $\mu_D = 0$ H$_1$: $\mu_D > 0$	B1	Both. Accept alternatives e.g. $\mu_D < 0$ for H$_1$, or $\mu_B - \mu_A$ etc provided adequately defined. Hypotheses in words only must include "population". Do NOT allow "$\bar{X} = ...$" or similar. unless $\bar{X}$ is clearly and explicitly stated to be a population mean.
			Where μ_D is the (population) mean reduction/difference in drying time.	B1	For adequate verbal definition. Allow absence of "population" if correct notation μ is used.
			MUST be PAIRED COMPARISON t test. Differences (reductions) (before – after) are: 0.7 0.7 0.2 –0.3 0.8 –0.1 0.3 –0.1 0.1 0.5		Allow "after – before" if consistent with alternatives above.
			$\bar{x} = 0.28$ $s_{n-1} = 0.3852(84)$ $(s_{n-1}^2 = 0.1484(44))$	B1	Do not allow $s_n = 0.3655$ $(s_n^2 = 0.1336)$
			Test statistic is $\dfrac{0.28 - 0}{\frac{0.3853}{\sqrt{10}}}$	M1	Allow c's $\bar{x}$ and/or s_{n-1}. Allow alternative: 0 + (c's 1.833) × $\frac{0.3853}{\sqrt{10}}$ (= 0.2233) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ – (c's 1.833) × $\frac{0.3853}{\sqrt{10}}$ (= 0.0566) for comparison with 0.)
			= 2.298.	A1	c.a.o. but ft from here in any case if wrong. Require 3/4 sf; condone up to 6. Use of $0 - \bar{x}$ scores M1A0, but ft.
			Refer to t_9.	M1	No ft from here if wrong. P(t > 2.298) = 0.02357.
			Single-tailed 5% point is 1.833.	A1	No ft from here if wrong.
			Significant.	A1	ft only c's test statistic.
			Seems mean drying time has fallen.	A1 [9]	ft only c's test statistic. "Non-assertive" conclusion in context to include "on average" oe.

Question			Answer	Marks	Guidance
1	(iv)		CI is given by 0.28 ± 2.262 $\times \frac{0.3853}{\sqrt{10}}$ = 0.28 ± 0.2756 = (0.0044, 0.5556)	M1 B1 M1 A1 **[4]**	Allow c's $\overline{x}$. Allow c's s_{n-1}. c.a.o. Must be expressed as an interval. Require 3/4 dp; condone 5. If the final answer is centred on a negative sample mean then do not award the final A mark. ZERO/4 if not same distribution as test. Same wrong distribution scores maximum M1 B0 M1 A0. Recovery to t_9 is OK.
2	(a)	(i)	For example, need to take a sample because the population might be too large for it to be sensible to take a complete census. Because the sampling process might be destructive.	E1 E1 **[2]**	Reward 1 mark each for any two distinct, sensible points.
2	(a)	(ii)	For example Sample should be unbiased. Sample should be representative (of the population).	E1 E1 **[2]**	Reward 1 mark each for any two distinct, sensible points that the sample/data should be fit for purpose. Further examples include: data should not be distorted by the act of sampling; data should be relevant.
2	(a)	(iii)	A random sample … enables proper statistical inference to be undertaken …… because we know the probability basis on which it has been selected	E2 **[2]**	Award E2, 1, 0 depending on the quality of response.
2	(b)	(i)	A Wilcoxon signed rank test might be used when nothing is known about the distribution of the background population. Must assume symmetry (about the median).	E1 E1 **[2]**	Do not allow "sample", or "data" unless it clearly refers to the population. Do not allow if "Normality" forms part of the assumption.

Question			Answer	Marks	Guidance
2	(b)	(ii)	H_0: $m = 28.7$ H_1: $m > 28.7$	B1	Both. Accept hypotheses in words.
			where m is the population median	B1	Adequate definition of m to include "population".

Speeds	−28.7	Rank of |diff|
32.0	3.3	8
29.1	0.4	3
26.1	−2.6	6
35.2	6.5	12
34.4	5.7	11
28.6	−0.1	1
32.3	3.6	9
28.5	−0.2	2
27.0	−1.7	5
33.3	4.6	10
28.2	−0.5	4
31.9	3.2	7

				M1	for subtracting 28.7.
				M1	for ranks.
				A1	ft if ranks wrong. If candidate has tied ranks then penalise A0 here but ft from here.
			$W_- = 1 + 2 + 4 + 5 + 6 = 18$	B1	($W_+ = 3 + 7 + 8 + 9 + 10 + 11 + 12 = 60$)
			Refer to Wilcoxon single sample tables for $n = 12$.	M1	No ft from here if wrong.
			Lower 5% point is 17 (or upper is 61 if 60 used).	A1	ie a 1-tail test. No ft from here if wrong.
			Result is not significant.	A1	ft only c's test statistic.
			No evidence to suggest that the median speed has increased.	A1	ft only c's test statistic. "Non-assertive" conclusion in context to include "on average" oe.
				[10]	
3			$S \sim \text{N}(11.07,\ 2.36^2)$ $C \sim \text{N}(57.33,\ 8.76^2)$ $R \sim \text{N}(24.23,\ 3.75^2)$		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)		$\text{P}(10 < S < 13)$		
			$= \text{P}\left(\frac{10 - 11.07}{2.36} < Z < \frac{13 - 11.07}{2.36}\right)$	M1	For standardising. Award once, here or elsewhere.
			$= \text{P}(-0.4534 < Z < 0.8178)$	A1	
			$= 0.7931 - (1 - 0.6748)$		
			$= 0.4679$	A1	Cao Accept 0.468(0), 0.4681, 0.4682, but not 0.4683.
				[3]	

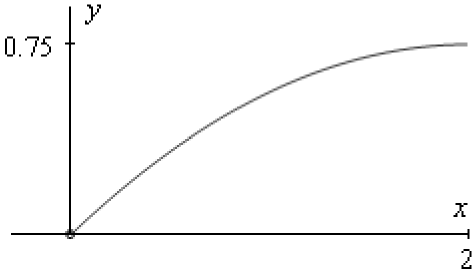
Question			Answer	Marks	Guidance
3	(ii)		Want $P(R > S + 10)$ i.e. $P(R - S > 10)$ R – S ~ N(24.23 – 11.07 = 13.16, $3.75^2 + 2.36^2 = 19.6321$) P(this > 10) = P($Z > \frac{10 - 13.16}{\sqrt{19.6321}} = -0.7132$) = 0.7621	M1 B1 B1 A1 **[4]**	Allow $S - R$ provided subsequent work is consistent. Mean. Variance. Accept sd = √19.6321 = 4.4308... cao
3	(iii)		Want $P(S + R > \tfrac{2}{3}C)$ i.e. $P(S + R - \tfrac{2}{3}C > 0)$ $S + R - \tfrac{2}{3}C$ ~ N(11.07 + 24.23 – ⅔ × 57.33 = –2.92, $2.36^2 + 3.75^2 + (\tfrac{2}{3} \times 8.76)^2 = 53.7377$) P(this > 0) = P($Z > \frac{0 - (-2.92)}{\sqrt{53.7377}} = 0.3983$) = 1 – 0.6548 = 0.3452	M1 B1 B1 A1 **[4]**	Allow $\tfrac{2}{3}L - (S + R)$ provided subsequent work is consistent. Mean Variance. Accept sd = √53.7377 = 7.3306... cao
3	(iv)		$\bar{x} = 98.484$, $s_{n-1} = 10.1594$ CI is given by 98.484 ± 2.201 $\times \frac{10.1594}{\sqrt{12}}$ = 98.484 ± 6.455 = (92.03, 104.94)	B1 M1 B1 M1 A1 **[5]**	Do not allow s_n = 9.7269. ft c's $\bar{x}$ ±. From t_{11}. ft c's s_{n-1}. cao Must be expressed as an interval. Require 1 or 2 dp; condone 3dp.
3	(v)		Normality is unlikely to be reasonable – times could well be (positively) skewed. Independence is unlikely to be reasonable – e.g. a competitor who is fast in one stage may well be fast in all three.	E1 E1 **[2]**	Discussion required. Accept any reasonable point. Accept "reasonable" provided an adequate explanation is given. Discussion required. Accept any reasonable point. This is independence between stages for a particular competitor, not between competitors.

Question			Answer	Marks	Guidance
4	(i)		H_0: The model for the number of callouts fits the data	B1	Do not allow "Data fit the model" o.e for either hypothesis.
			H_1: The model for the number of callouts does not fit the data.	B1	

Obs'd frequency	145	79	22	6	3	0
Exp'd frequency	139.947	83.968	25.190	5.038	0.756	0.101

Question			Answer	Marks	Guidance
			Merge last 3 cells. Obs 9 Exp 5.895	M1	
			$X^2 = 0.1824 + 0.2939 + 0.4040 + 1.6355$	M1	Calculation of X^2.
			$= 2.515(8)$	A1	Cao Require 3/4 sf; condone up to 6.
			Refer to χ^2_2.	M1	Allow correct df (= cells – 2) from wrongly grouped table and ft. Otherwise, no ft if wrong. $P(X^2 > 2.5158) = 0.2842$.
			Upper 5% point is 5.991.	A1	No ft from here if wrong.
			Not significant.	A1	ft only c's test statistic.
			Suggests it is reasonable to suppose that the model fits the data.	A1	ft only c's test statistic. "Non-assertive" conclusion in words (+context). Do not allow "Data fit model" o.e.
				[9]	
4	(ii)		Mean = 5/3 $\therefore\ \lambda = 0.6$	B1 **[1]**	
4	(iii)		$F(t) = \int_0^t 0.6\mathrm{e}^{-0.6x}\,\mathrm{d}x$	M1	Correct integral with limits (which may be implied subsequently). Allow use of "+ c" accompanied by a valid attempt to evaluate it.
			$= \left[-\mathrm{e}^{-0.6x}\right]_0^t$	A1	Correctly integrated.
			$= \left(-\mathrm{e}^{-0.6t} - (-\mathrm{e}^0)\right) = 1 - \mathrm{e}^{-0.6t}$	A1	Limits used or c evaluated correctly. Accept unsimplified form. If final answer is given in terms of λ then allow max M1A1A0.
				[3]	
4	(iv)		$P(T > 1) = 1 - F(1)$	M1	ft c's F(t).
			$= 1 - \left(1 - \mathrm{e}^{-0.6}\right) = 0.5488$	A1	cao Allow any exact form of the correct answer.
				[2]	
4	(v)		$F(m) = \frac{1}{2} \qquad \therefore 1 - \mathrm{e}^{-0.6m} = \frac{1}{2}$	M1	Use of definition of median. Allow use of c's F(t).
			$\therefore \mathrm{e}^{-0.6m} = \frac{1}{2} \qquad \therefore -0.6m = -\ln 2 \qquad \therefore m = \frac{\ln 2}{0.6}$	M1	Convincing attempt to rearrange to "m = …", to include use of logs.
			$m = 1.155$ (days)	A1	Cao obtained only from the correct F(t). Must be evaluated. Require 2 to 4 sf; condone 5.
				[3]	

Question			Answer	Marks	Guidance
1	**(i)**		A Normal test is not appropriate since …		
			… the sample is small and	E1	
			… the population variance is not known (it must be estimated from the data).	E1 **[2]**	Allow use of "σ", otherwise insist on "population".
1	**(ii)**		The sample is taken from a Normal population.	B1 **[1]**	
1	**(iii)**		H_0: $\mu = 7.8$ H_1: $\mu \neq 7.8$	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 <u>population</u> mean.
			where μ is the mean water pressure.	B1	For adequate verbal definition. Allow absence of "population" if correct notation μ is used.
			$\bar{x} = 7.631$ $\quad s = 0.1547$	B1	$s_n = 0.1459$ but do <u>NOT</u> allow this here or in construction of test statistic, but ft from there.
			Test statistic is $\dfrac{7.631 - 7.8}{\left(\dfrac{0.1547}{\sqrt{9}}\right)}$	M1	Allow c's $\bar{x}$ and/or s_{n-1}. Allow alternative: 7.8 + (c's –2.896) × $0.1547/\sqrt{9}$ (= 7.65…) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ – (c's –2.896) × $0.1547/\sqrt{9}$ (= 7.78…) for comparison with 7.8.)
			$= -3.27(7)$.	A1	c.a.o. but ft from here in any case if wrong. Use of $\mu - \bar{x}$ scores M1A0.
			Refer to t_8.	M1	No ft from here if wrong.
			Double-tailed 2% point is ±2.896.	A1	Must compare test statistic with <u>minus</u> 2.896 unless absolute values are being compared. No ft from here if wrong. Allow $P(t < -3.27(7) \text{ or } t > 3.27(7)) = 0.0113$ for M1A1.
			Significant.	A1	ft only c's test statistic if both M's scored.
			Sufficient evidence to suggest that the mean water pressure has changed.	A1 **[9]**	ft only c's test statistic if both M's scored. Conclusion in context to include "average" o.e.

Question			Answer	Marks	Guidance
1	**(iv)**		In repeated sampling, 95% of all confidence intervals constructed in this way will contain the true mean.	E1 E1 **[2]**	
1	**(v)**		CI is given by 7.631 ± 2·306 $\times \frac{0.1547}{\sqrt{9}}$ = 7.631 ± 0.118(9) = (7.512, 7.750)	M1 B1 M1 A1 **[4]**	ZERO/4 if not same distribution as test. Same wrong distribution scores maximum M1B0M1A0. Recovery to t_8 is OK. Allow c's $\overline{x}$. 2.306 seen. Allow c's s_{n-1}. c.a.o. Must be expressed as an interval.
2	**(i)**			G1 G1 G1 **[3]**	Curve with positive gradient, through the origin and in the first quadrant only. Correct shape for an inverted parabola ending at maximum point. End point (2, 3/4) labelled.

Question			Answer	Marks	Guidance
2	(ii)		$\mathrm{E}(X)=\frac{3}{16}\int_0^2(4x^2-x^3)\mathrm{d}x$	M1	Correct integral for E(X) with limits (which may appear later).
			$=\frac{3}{16}\left[\frac{4x^3}{3}-\frac{x^4}{4}\right]_0^2$	M1	Correctly integrated. Dep on previous M1.
			$=\frac{3}{16}\left\{\left(\frac{32}{3}-\frac{16}{4}\right)-0\right\}$		
			$=\frac{5}{4}$	A1	Limits used correctly to obtain PRINTED ANSWER (BEWARE) convincingly. Condone absence of "–0".
			$\mathrm{E}(X^2)=\frac{3}{16}\int_0^2(4x^3-x^4)\mathrm{d}x$	M1	Correct integral for E(X) with limits (which may appear later).
			$=\frac{3}{16}\left[x^4-\frac{x^5}{5}\right]_0^2$	M1	Correctly integrated. Dep on previous M1.
			$=\frac{3}{16}\left\{\left(16-\frac{32}{5}\right)-0\right\}$		
			$=\frac{9}{5}$	A1	Limits used correctly to obtain result. Condone absence of "–0".
			$\mathrm{Var}(X)=\frac{9}{5}-\left(\frac{5}{4}\right)^2=\frac{19}{80}$	M1	Use of Var(X) = E(X^2) – E(X)^2.
			$\mathrm{sd}=\sqrt{\frac{19}{80}}=0.487(3)$	A1	cao
				[8]	
2	(iii)		$\mathrm{SE}(\bar{X})=\frac{0.487}{\sqrt{100}}$	M1	
			$=0.0487$	A1 **[2]**	ft c's $\sigma/10$.

Question			Answer	Marks	Guidance
2	(iv)		$\mathrm{P}(X<1)=\frac{3}{16}\int_0^1(4x-x^2)\mathrm{d}x$ $=\frac{3}{16}\left[2x^2-\frac{x^3}{3}\right]_0^1$ $=\frac{3}{16}\left\{\left(2-\frac{1}{3}\right)-0\right\}$ $=\frac{5}{16}$	M1 A1 **[2]**	Correct integral for P(X < 1) with limits (which may appear later). cao. Condone absence of "–0" when limits applied.
2	(v)		Regard the reed beds as clusters. Select a few clusters (maybe only one) at random. Take a (simple random) sample of reeds (or maybe all of them) from the selected cluster(s).	E1 E1 E1 **[3]**	NB "Clusters of reeds" scores 0 unless clearly and correctly explained.
3			$P1 \sim \mathrm{N}(2025, 44.6^2)$ $P2 \sim \mathrm{N}(1565, 21.8^2)$ $I \sim \mathrm{N}(1410, 33.8^2)$		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.
3	(i)		$\mathrm{P}(P1 < 2100) =$ $\mathrm{P}\left(Z<\frac{2100-2025}{44.6}=1.681(6)\right)$ $= 0.9536/7$	M1 A1 A1 **[3]**	For standardising. Award once, here or elsewhere. c.a.o.

Question			Answer	Marks	Guidance
3	**(ii)**		Require P($P1 - P2 > 400$)	M1	
			$P1 - P2 \sim (2025 - 1565 = 460,$	B1	Mean.
			$44.6^2 + 21.8^2 = 2464.4)$	B1	Variance. Accept sd (= 49.64).
			P(this > 400) =		
			$\mathrm{P}\left(Z > \frac{400-460}{\sqrt{2464.4}} = -1.208(6)\right) = 0\cdot 8864/5$	A1	cao
				[4]	
3	**(iii)**		$T = P1 + P2 + I \sim \mathrm{N}(5000,$	B1	Mean.
			$\sigma^2 = 44.6^2 + 21.8^2 + 33.8^2 = 3606.84)$	B1	Variance. Accept sd (= 60.056…).
			Require b s.t. P($T > b$) = 0.95		
			$\therefore \frac{b - 5000}{\sqrt{3606.84}} = -1.645$	B1	–1.645 seen.
			$\therefore b = 5000 - 1.645 \times \sqrt{3606.84} = 4901.2..$	A1	c.a.o.
				[4]	
3	**(iv)**		Mean = (1.2 × 2025) + (1.3 × 1565) + (0.8 × 1410) = £5592.50	B1	Condone absence of £.
			Var = $(1.2^2 \times 44.6^2) + (1.3^2 \times 21.8^2) +$	M1	Use of at least one of $(1.2^2 \times 44.6^2)$ etc…
			$(0.8^2 \times 33.8^2) = 4398.7076 \approx £^2 4399$	A1	Condone absence of $£^2$.
				[3]	
3	**(v)**		Mean = (123.72 + 127.38)/2 = 125.55	B1	Cao
			$s = \frac{127.38 - 125.55}{2.576/\sqrt{50}} = 5.02(3)$	B1	Sight of 2.576.
				M1	Or equivalent.
				A1	cao
				[4]	

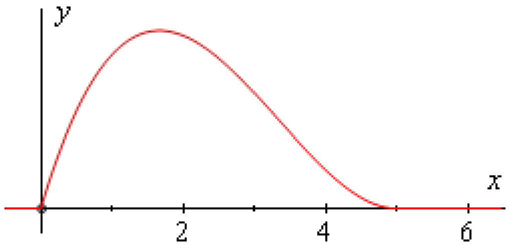
<table>
<tr><th colspan="3">Question</th><th>Answer</th><th>Marks</th><th>Guidance</th></tr>
<tr><td rowspan="2">**4**</td><td rowspan="2">**(a)**</td><td rowspan="2">**(i)**</td><td>Number all the projects to be marked. (Sampling frame.)</td><td>E1</td><td>Do not award if candidate subsequently describes a different method of sampling (eg systematic sampling).</td></tr>
<tr><td>Use a form of random number generator to select the projects in the sample until 12 projects have been selected.</td><td>E1
[2]</td><td>Condone absence of 12.</td></tr>
<tr><td rowspan="7">**4**</td><td rowspan="7">**(a)**</td><td rowspan="7">**(ii)**</td><td>H_0: $m = 0$ H_1: $m \neq 0$
where m is the population median difference between the examiners’ marks.</td><td></td><td>This is given in the question.</td></tr>
<tr><td><table><tr><td>Diff</td><td>15</td><td>10</td><td>2</td><td>–7</td><td>11</td><td>19</td><td>–8</td><td>–14</td><td>17</td><td>13</td><td>–5</td><td>–4</td></tr><tr><td>Rank</td><td>10</td><td>6</td><td>1</td><td>4</td><td>7</td><td>12</td><td>5</td><td>9</td><td>11</td><td>8</td><td>3</td><td>2</td></tr></table></td><td>M1
M1
A1</td><td>For differences. ZERO (out of 8) in this section if differences not used.
For ranks.
ft from here if ranks wrong.</td></tr>
<tr><td>$W_{-} = 2 + 3 + 4 + 5 + 9 = 23$</td><td>B1</td><td>(or $W_{+} = 1 + 6 + 7 + 8 + 10 + 11 + 12 = 55$)</td></tr>
<tr><td>Refer to tables of Wilcoxon paired (/single sample) statistic for $n = 12$.</td><td>M1</td><td>No ft from here if wrong.</td></tr>
<tr><td>Lower (or upper if 55 used) 5% tail is 17 (or 61 if 55 used).</td><td>A1</td><td>i.e. a 2-tail test. No ft from here if wrong.</td></tr>
<tr><td>Result is not significant.</td><td>A1</td><td>ft only c’s test statistic.</td></tr>
<tr><td>Insufficient evidence to suggest a difference in the marks awarded, on average.</td><td>A1
[8]</td><td>ft only c’s test statistic. Conclusion in context to include “average” o.e.</td></tr>
</table>

Question			Answer	Marks	Guidance
4	(b)		H_0: The random number function is performing as it should. H_1: The random number function is not performing as it should.	B1	Both hypotheses. Must be the right way round. Allow use of the uniform distribution/model. Do not accept "data fit model" oe.
			All expected frequencies are 10	B1	
			$X^2 = 1.6 + 0.4 + 0.1 + 1.6 + 0.4 + 0.1 + 2.5 + 2.5 + 1.6 + 1.6$	M1	Calculation of X^2.
			$= 12.4$	A1	c.a.o.
			Refer to χ_9^2.	M1	Allow correct df (= cells – 1) from wrongly grouped table and ft. Otherwise, no ft if wrong. $P(X^2 > 12.4) = 0.1916$.
			Upper 10% point is 14.68.	A1	No ft from here if wrong.
			Not significant.	A1	ft only c's test statistic.
			Insufficient evidence to suggest that the random number function is not performing as it should.	A1	ft only c's test statistic. Conclusion in context. Allow in terms of the uniform distribution/model. Do not accept "data fit model" oe.
				[8]	

<table>
<tr><th colspan="3">Question</th><th>Answer</th><th>Marks</th><th>Guidance</th></tr>
<tr><td>1</td><td>(i)</td><td></td><td>H_0: $m = 7.4$ H_1: $m < 7.4$
where m is the population median time.</td><td>B1</td><td>Both. Accept hypotheses in words, but must include “population”. Do NOT allow symbols other than m unless clearly and explicitly stated to be a population median.</td></tr>
<tr><td></td><td></td><td></td><td></td><td>B1</td><td>Adequate definition of m to include “population”.</td></tr>
<tr><td></td><td></td><td></td><td>
<table>
<tr><th>Times</th><th>−7.4</th><th>Rank of |diff|</th></tr>
<tr><td>6.90</td><td>−0.50</td><td>8</td></tr>
<tr><td>7.23</td><td>−0.17</td><td>3</td></tr>
<tr><td>6.54</td><td>−0.86</td><td>10</td></tr>
<tr><td>7.62</td><td>0.22</td><td>4</td></tr>
<tr><td>7.04</td><td>−0.36</td><td>6</td></tr>
<tr><td>7.33</td><td>−0.07</td><td>1</td></tr>
<tr><td>6.74</td><td>−0.66</td><td>9</td></tr>
<tr><td>6.45</td><td>−0.95</td><td>11</td></tr>
<tr><td>7.81</td><td>0.41</td><td>7</td></tr>
<tr><td>7.71</td><td>0.31</td><td>5</td></tr>
<tr><td>7.50</td><td>0.10</td><td>2</td></tr>
<tr><td>6.32</td><td>−1.08</td><td>12</td></tr>
</table>
</td><td>M1
M1
A1</td><td>for subtracting 7.4.
for ranking.
All correct. ft if ranks wrong.</td></tr>
<tr><td></td><td></td><td></td><td>$W_+ = 2 + 4 + 5 + 7 = 18$</td><td>B1</td><td>($W_- = 1 + 3 + 6 + 8 + 9 + 10 + 11 + 12 = 60$)</td></tr>
<tr><td></td><td></td><td></td><td>Refer to Wilcoxon single sample tables for $n = 12$.</td><td>M1</td><td>No ft from here if wrong.</td></tr>
<tr><td></td><td></td><td></td><td>Lower 5% point is 17 (or upper is 61 if 60 used).</td><td>A1</td><td>i.e. a 1-tail test. No ft from here if wrong.</td></tr>
<tr><td></td><td></td><td></td><td>Result is not significant.</td><td>A1</td><td>ft only c’s test statistic.</td></tr>
<tr><td></td><td></td><td></td><td>Insufficient evidence to suggest that the median time has been reduced.</td><td>A1</td><td>ft only c’s test statistic. Conclusion in context.</td></tr>
<tr><td></td><td></td><td></td><td></td><td>[10]</td><td></td></tr>
</table>

Question			Answer	Marks	Guidance
1	**(ii)**		$\bar{x} = 6.94 \quad s = 0.37$	B1	Accept $s^2 = 0.1369$. Beware use of msd (0.13518875) or rmsd (0.3676(8)). Do not allow here or below.
			CI is given by 6.94 ±	M1	ft c's $\bar{x}$ ±.
			1.96	B1	1.96 seen.
			$\times \frac{0.37}{\sqrt{80}}$	M1	ft c's s but not rmsd.
			= 6.94 ± 0.0811 = (6.859, 7.021)	A1	c.a.o. Must be expressed as an interval. [rmsd gives 6.94 ± 0.0805(7) = (6.8594(2), 7.0205(7))]
			Normal distribution can be used because the sample size is large enough for the Central Limit Theorem to apply.	E1 **[6]**	CLT essential
1	**(iii)**		Advantage: A 99% confidence interval is more likely to contain the true mean.	E1	O.e.
			Disadvantage: A 99% confidence interval is less precise/wider.	E1 **[2]**	O.e.
2	**(i)**		A paired test would eliminate any differences between individual cattle.	E1 **[1]**	
2	**(ii)**		Must assume: Normality of population … … of <u>differences</u>.	B1 B1 **[2]**	

Question			Answer	Marks	Guidance
2	(iii)		H_0: $\mu_D = 10$ H_1: $\mu_D < 10$	B1	Both. Accept alternatives e.g. $\mu_D > -10$ for H_1, or $\mu_A - \mu_B$ etc provided adequately defined. 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 <u>population</u> mean.
			Where μ_D is the (population) mean increase/difference in milk yield. <u>MUST</u> be PAIRED COMPARISON *t* test.	B1	For adequate verbal definition. Allow absence of "population" if correct notation μ is used.
			Differences (increases) (after – before) are: 4 9 6 13 1 8 6 7 9 12	M1	Allow "before – after" if consistent with alternatives for hypotheses above.
			$\bar{x} = 7.5 \quad s_{n-1} = 3.566(8) \quad (s_{n-1}{}^2 = 12.722(2))$	A1	Do not allow $s_n = 3.3837$ ($s_n{}^2 = 11.45$).
			Test statistic is $\dfrac{7.5 - 10}{\frac{3.5668}{\sqrt{10}}}$	M1	Allow c's $\bar{x}$ and/or s_{n-1}. Allow reversed numerator compared with 2.2164 Allow alternative: 10 – (c's 1.833) × $\dfrac{3.5668}{\sqrt{10}}$ (= 7.933) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ + (c's 1.833) × $\dfrac{3.5668}{\sqrt{10}}$ (= 9.567) for comparison with 10.)
			= –2.2164.	A1	c.a.o. but ft from here in any case if wrong. Use of 10 – $\bar{x}$ scores M1A0, but ft.
			Refer to t_9.	M1	No ft from here if wrong.
			Single-tailed 5% point is –1.833.	A1	Must be minus 1.833 unless absolute values are being compared. No ft from here if wrong. $P(t < -2.2164) = 0.0269$.
			Significant.	A1	ft only c's test statistic.
			Sufficient evidence to suggest that the mean milk yield has not increased by 10 litres (per cow per week).	A1	ft only c's test statistic. Conclusion in context to include "on average" o.e. Accept "Sufficient evidence to suggest that the **company's** claim is not justified." o.e.
				[10]	

Question			Answer	Marks	Guidance
2	(iv)		CI is given by $7.5 \pm$ $2{\cdot}262$ $\times \dfrac{3.5668}{\sqrt{10}}$ $= 7.5 \pm 2.5514 = (4.948, 10.052)$	M1 B1 M1 A1 **[4]**	ZERO/4 if not same distribution as test. Same wrong distribution scores maximum M1B0M1A0. Recovery to t_9 is OK. Allow c's $\overline{x}$. 2.262 seen. Allow c's s_{n-1}. c.a.o. Must be expressed as an interval.
3	(i)			G1 G1 G1 **[3]**	Curve, through the origin and in the first quadrant only. A single maximum; curve returns to $y = 0$; nothing to the right of $x = 5$. No t.pt at $x = 0$; t.pt. at $x = 5$; (5, 0) labelled (p.i. by an indicated scale).
3	(ii)		$\mathrm{F}(x) = k\int_0^x t(t-5)^2\,\mathrm{d}t$ $= k\left[\dfrac{t^4}{4} - \dfrac{10t^3}{3} + \dfrac{25t^2}{2}\right]_0^x$ $= k\left(\dfrac{x^4}{4} - \dfrac{10x^3}{3} + \dfrac{25x^2}{2}\right)$	M1 M1 A1 **[3]**	Correct integral for F(x) with limits (which may appear later). Correctly integrated. Limits used correctly to obtain expression. Condone absence of "–0". Do not require complete definition of F(x). Dependent on both M1's

Question			Answer	Marks	Guidance
3	**(iii)**		F(5) = 1 $\therefore k\left(\frac{5^4}{4}-\frac{10\times5^3}{3}+\frac{25\times5^2}{2}\right)=1$ $\therefore k\left(\frac{1875-5000+3750}{12}\right)=1$ $\therefore k\times\frac{625}{12}=1$ $\therefore k=\frac{12}{625}$	M1 A1 **[2]**	Substitute $x = 5$ and equate to 1. Expect to see evidence of at least this line of working (oe) for A1. Convincingly shown. Beware printed answer.
3	**(iv)**		For $0 \le x < 1$, Expected f = 60 × F(1) $= 60\times\frac{12}{625}\left(\frac{1^4}{4}-\frac{10\times1^3}{3}+\frac{25\times1^2}{2}\right)=10.848$ For $1 \le x < 2$, Expected f = 60 – Σ(the rest) = 20.64	M1 A1 B1 **[3]**	Use of 60 × F(x) with correct k. Allow also 31.488 – frequency for $1 \le x < 2$ provided that one found using F(x). Allow either frequency found by integration. FT 31.488 – previous answer. Or allow 60 × (F(2) – F(1))

Question			Answer	Marks	Guidance
3	(v)		H$_0$: The model is suitable / fits the data. H$_1$: The model is not suitable / does not fit the data.	B1	Both hypotheses. Must be the right way round. Do not accept "data fit model" oe.
			Merge last 2 cells: Obs f = 17, Exp f = 10.752	M1	
			$X^2 = 3.1525 + 1.5411 + 1.5460 + 3.6307$	M1	Calculation of X^2.
			$= 9.870$	A1	c.a.o.
			Refer to χ_3^2.	M1	Allow correct df (= cells – 1) from wrongly grouped table and ft. Otherwise, no ft if wrong.
			Upper 2.5% point is 9.348.	A1	No ft from here if wrong. $P(X^2 > 9.870) = 0.0197$.
			Significant.	A1	ft only c's test statistic.
			Sufficient evidence to suggest that the model is not suitable in this context.	A1	ft only c's test statistic. Conclusion in context. Do not accept "data do not fit model" oe.
				[8]	
4			$C \sim N(96, 21)$ $M \sim N(57, 14)$		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.
4	(i)		$P(90 < C < 100)$		
			$= P\left(\frac{90-96}{\sqrt{21}} < Z < \frac{100-96}{\sqrt{21}}\right)$	M1	For standardising. Award once, here or elsewhere. SC – candidates with consistent variances of 21² and 14² can be awarded all M and B marks
			$= P(-1.3093 < Z < 0.8729)$	A1	Either side correct. SC – 0.2857, 0.1905
			$= 0.8086 - (1 - 0.9047)$	A1	Both table values correct. Or 0.8086 – 0.0953 SC 0.5755 – (1 – 0.6125)
			$= 0.7133$	A1	c.a.o.
				[4]	
4	(ii)		Total weight $T \sim N(153, 35)$	B1	Mean.
				B1	Variance. Accept sd = 5.916… SC 637 sd = 25.239
			$P(T < 145) = P\left(Z < \frac{145-153}{\sqrt{35}} = -1.3522\right)$		
			$= 1 - 0.9118 = 0.0882$	A1	c.a.o.
				[3]	

Question			Answer	Marks	Guidance
4	**(iii)**		$T_1 + T_2 + T_3 + T_4 \sim \text{N}(612, 140)$	B1	Mean.
				B1	Variance. Accept sd = 11.832… SC = 2548 sd = 50.478
			Require w such that P(this > w) = 0.95	M1	
			$\therefore w = 612 - 1.645 \times \sqrt{140}$	B1	1.645 seen.
			= 592.5(3)	A1	c.a.o.
				[5]	
4	**(iv)**		Require $M \geq 0.35(M + C)$	M1	Formulate requirement.
			$\therefore 0.65M \geq 0.35C$		
			$\therefore 0.65M - 0.35C \geq 0$	A1	Convincingly shown. Beware printed answer.
			$0.65M - 0.35C \sim$	B1	Mean.
			N((0.65 × 57) – (0.35 × 96) = 3.45,	M1	For use of at least one of $0.65^2 \times \ldots$ or $0.35^2 \times \ldots$
			$(0.65^2 \times 14) + (0.35^2 \times 21) = 8.4875$)	A1	Variance. Accept sd = 2.913… SC variance = 136.83 sd = 11.70
			P(This ≥ 0) $= \text{P}\left(Z \geq \dfrac{0 - 3.45}{\sqrt{8.4875}} = -1.1842\right)$		
			= 0.8818	A1	c.a.o.
				[6]	