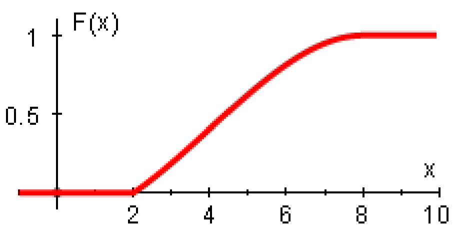
4768 Statistics 3

1 (i)	H_0: The number of eggs hatched can be modelled by B(3, ½) H_1: The number of eggs hatched cannot be modelled by B(3, ½)	B1 B1		
	With p = ½ (see table below)	M1 A1	Probs × 80 for expected frequencies. All correct.	
	$X^2 = 0.9 + 1.6333 + 0.0333 + 12.1$ $= 14.666(7)$	M1 A1	Calculation of X^2. 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. Significant. Suggests it is reasonable to suppose model with p = ½ does not apply.	A1 A1 A1	No ft from here if wrong. ft only c's test statistic. ft only c's test statistic.	**[10]**
(ii)	$\bar{x} = \frac{144}{80} = 1.8$ $\therefore \hat{p} = \frac{1.8}{3} = 0.6$	B1 B1	C.a.o. Use of $E(X) = np$. ft c's mean, provided $0 < \hat{p} < 1$.	**[2]**
(iii)	Refer to χ_2^2. Upper 5% point is 5.991. Suggests it is reasonable to suppose model with estimated p does apply.	M1 A1 A1	Allow df 1 less than in part (i). No ft if wrong. No ft if wrong. 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. **Total**	**[2]** **[17]**

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

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

Question	Answer	Mark	Notes	Total
2 (b)	H_0: $m = 4.42$ H_1: $m \neq 4.42$ where m is the population median	B1 B1	Both. Accept hypotheses in words. Adequate definition of m to include "population".	

Weights	− 4.42	Rank of |diff|
3.16	−1.26	7
3.62	−0.80	6
3.80	−0.62	4
3.90	−0.52	3
4.02	−0.40	2
4.72	0.30	1
5.14	0.72	5
6.36	1.94	8
6.50	2.08	9
6.58	2.16	10
6.68	2.26	11
6.78	2.36	12

M1 for subtracting 4.42.

M1 for ranks.
A1 ft if ranks wrong.

Answer	Mark	Notes	Total
$W_- = 2 + 3 + 4 + 6 + 7 = 22$	B1	($W_+ = 1 + 5 + 8 + 9 + 10 + 11 + 12 = 56$)	
Refer to Wilcoxon single sample tables for $n = 12$.	M1	No ft from here if wrong.	
Lower 2½% point is 13 (or upper is 65 if 56 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.	
Evidence suggests that a median of 4.42 is consistent with these data.	A1	ft only c's test statistic.	[10]
		Total	**[19]**

3 (i)	Must assume • Normality of population … • … of <u>differences</u>.	B1 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 $\bar{x}$ ±	M1	Overall structure, seen or implied.	
	2.201	B1	From t_{11}, seen or implied.	
	$\times \frac{s}{\sqrt{12}} = (-0.5380, 1.4046)$	A1	Fully correct pair of equations using the given interval, seen or implied.	
	$\bar{x}$ = ½(1.4046 − 0.5380) = 0.4333	B1		
	$s = (1.4046 - 0.4333) \times \frac{\sqrt{12}}{2.201}$ = 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 $52.1 \pm$ 1.96 $\times \frac{4.8}{\sqrt{60}}$ $= 52.1 \pm 1.2146 = (50.885(4), 53.314(6))$	B1 M1 B1 M1 A1	Both correct. c.a.o. Must be expressed as an interval.	**[5]**
			Total	**[18]**