4768 Statistics 3

Q1 (a)	$f(x) = \lambda x^c,\ 0 \le x \le 1,\ \lambda > 1$			
(i)	$\int_0^1 \lambda x^c \mathrm{d}x = 1$	M1	Correct integral, with limits (possibly appearing later), set equal to 1.	
	$\therefore \left[\frac{\lambda x^{c+1}}{c+1}\right]_0^1 = 1$	M1	Integration correct and limits used.	
	$\therefore \frac{\lambda}{c+1} = 1 \qquad \therefore c = \lambda - 1$	A1	c.a.o.	3
(ii)	$\mathrm{E}(X) = \int_0^1 \lambda x^\lambda \mathrm{d}x$	M1	Correct form of integral for E(X). Allow c's expression for c.	
	$= \left[\frac{\lambda x^{\lambda+1}}{\lambda+1}\right]_0^1 = \frac{\lambda}{\lambda+1}.$	M1 A1	Integration correct and limits used. ft c's c.	3
(iii)	$\mathrm{E}(X^2) = \int_0^1 \lambda x^{\lambda+1} \mathrm{d}x$	M1	Correct form of integral for E(X^2). Allow c's expression for c.	
	$= \left[\frac{\lambda x^{\lambda+2}}{\lambda+2}\right]_0^1 = \frac{\lambda}{\lambda+2}.$	A1		
	$\mathrm{Var}(X) = \frac{\lambda}{\lambda+2} - \left(\frac{\lambda}{\lambda+1}\right)^2 = \frac{\lambda(\lambda+1)^2 - \lambda^2(\lambda+2)}{(\lambda+2)(\lambda+1)^2}$	M1	Use of Var(X) = E(X^2) – E(X)^2. Allow c's E(X^2) and E(X).	
	$= \frac{\lambda^3 + 2\lambda^2 + \lambda - \lambda^3 - 2\lambda^2}{(\lambda+2)(\lambda+1)^2} = \frac{\lambda}{(\lambda+2)(\lambda+1)^2}.$	A1	Algebra shown convincingly. Beware printed answer.	4
(b)	(see table below)		H$_0$: m = 32, H$_1$: m < 32, where m is the population median time.	
		M1	for subtracting 32.	
		M1 A1	for ranks. ft if ranks wrong.	
	$W_+ = 1 + 2 + 3 + 4 + 9 = 19$	B1	(or W_- =5 + 6 + 7 + 8 + 10 + 11 + 12 = 59)	
	Refer to Wilcoxon single sample tables for n = 12.	M1	No ft from here if wrong.	
	Lower (or upper if 59 used) 5% tail is 17 (or 61 if 59 used).	A1	i.e. a 1-tail test. No ft from here if wrong.	
	Result is not significant.	A1	ft only c's test statistic.	
	Seems that there is no evidence that Godfrey's times have decreased.	A1	ft only c's test statistic.	8
				18

Times	– 32	Rank of |diff|
40	8	4
20	−12	7
18	−14	8
11	−21	12
47	15	9
36	4	2
38	6	3
35	3	1
22	−10	5
14	−18	10
12	−20	11
21	−11	6

Q2	$V_G \sim \text{N}(56.5,\ 2.9^2)$ $V_W \sim \text{N}(38.4,\ 1.1^2)$		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)	$\text{P}(V_G < 60) = \text{P}(Z < \frac{60 - 56.5}{2.9} = 1.2069)$ $= 0.8862$	M1 A1 A1	For standardising. Award once, here or elsewhere.	3
(ii)	$V_T \sim \text{N}(56.5 + 38.4 = 94.9,$ $2.9^2 + 1.1^2 = 9.62)$ $\text{P}(\text{this} > 100) = \text{P}(Z > \frac{100 - 94.9}{3.1016} = 1.6443)$ $= 1 - 0.9499 = 0.0501$	B1 B1 A1	Mean. Variance. Accept sd (= 3.1016). c.a.o.	3
(iii)	$W_T \sim \text{N}(3.1 \times 56.5 + 0.8 \times 38.4 = 205.87,$ $3.1^2 \times 2.9^2 + 0.8^2 \times 1.1^2 = 81.5945)$ $\text{P}(200 < \text{this} < 220)$ $= \text{P}(\frac{200 - 205.87}{9.0330} < Z < \frac{220 - 205.87}{9.0330})$ $= \text{P}(-0.6498 < Z < 1.5643)$ $= 0.9411 - (1 - 0.7422) = 0.6833$	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 $\bar{x} = 205.6$ $s_{n-1} = 8.51$ 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. 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