4768 Statistics 3

Q1 (a)	$P(T > t) = \frac{k}{t^2}$, $t \geq 1$,			
(i)	$F(t) = P(T < t) = 1 - P(T > t)$ $\therefore F(t) = 1 - \frac{k}{t^2}$ $F(1) = 0$ $\therefore 1 - \frac{k}{1^2} = 0$ $\therefore k = 1$	M1 M1 A1	Use of 1 – P(…). Beware: answer given.	3
(ii)	$f(t) = \frac{dF(t)}{dt}$ $= \frac{2}{t^3}$	M1 A1	Attempt to differentiate c's cdf. (For $t \geq 1$, but condone absence of this.) Ft c's cdf provided answer sensible.	2
(iii)	$\mu = \int_1^\infty t f(t) dt = \int_1^\infty \frac{2}{t^2} dt$ $= \left[\frac{-2}{t}\right]_1^\infty$ $= 0 - (-2) = 2$	M1 A1 A1	Correct form of integral for the mean, with correct limits. Ft c's pdf. Correctly integrated. Ft c's pdf. Correct use of limits leading to correct value. Ft c's pdf provided answer sensible.	3
(b)	H_0: $m = 5.4$ H_1: $m \neq 5.4$ where m is the population median time for the task.	B1 B1	Both hypotheses. Hypotheses in words only must include "population". 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	$X \sim N(260,\ \sigma = 24)$		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)	$Y \sim N(260 \times 0.6 = 156,$ $24^2 \times 0.6^2 = 207.36$ $P(Y > 175) = P(Z > \frac{175-156}{14.4} = 1.3194)$ $= 1 - 0.9063 = 0.0937$	B1 B1 A1	Mean. Variance. Accept sd (= 14.4). c.a.o.	3
(iii)	$Y_1 + Y_2 + Y_3 + Y_4 \sim N(624,$ $829.44)$ $P(\text{this} < 600) = P(Z < \frac{600-624}{28.8} = -0.8333)$ $= 1 - 0.7976 = 0.2024$	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 $0.975 = P(\text{above} > w) = P\left(Z > \frac{w-624}{28.8}\right)$ $= P(Z > -1.96)$ $\therefore w - 624 = 28.8 \times -1.96 \Rightarrow w = 567.5(52)$	M1 B1 A1	Formulation of requirement. − 1.96 Ft parameters of (iii).	3
(v)	$On \sim N(150,\ \sigma = 18)$ $X_1 + X_2 + X_3 + On_1 + On_2 \sim N(1080,$ $2376)$ $P(\text{this} > 1000) = P(Z > \frac{1000-1080}{48.744} = -1.6412)$ $= 0.9496$	 B1 B1 A1	 Mean. Variance. Accept sd (= 48.744). c.a.o.	3
(vi)	Given $\bar{x} = 252.4 \quad s_{n-1} = 24.6$ CI is given by $252.4 \pm 2.576 \times \frac{24.6}{\sqrt{100}}$ $= 252.4 \pm 6.33(6) = (246.0(63), 258.7(36))$	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: $\mu < 380$ where μ is the mean temperature in the chamber.	B1 B1	Both hypotheses. Hypotheses in words only must include "population". 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.	
	$\bar{x} = 373.825$ $\quad s_{n-1} = 9.368$	B1	s_n = 8.969 but do NOT allow this here or in construction of test statistic, but FT from there.	
	Test statistic is $\dfrac{373.825 - 380}{\frac{9.368}{\sqrt{12}}}$	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}. Single-tailed 5% point is –1.796.	M1 A1	No ft from here if wrong. Must be minus 1.796 unless absolute values are being compared. No ft from here if wrong.	
	Significant. Seems mean temperature in the chamber has fallen.	A1 A1	ft only c's test statistic. 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. Disadvantage: less precision.	E1 E1	Or equivalents.	 2 18

Q4				
(a) (i)	$\bar{x} = \frac{1125}{500} = 2.25$ For binomial E(X) = $n \times p$ $\therefore \hat{p} = \frac{2.25}{5} = 0.45$	B1 M1 A1	 Use of mean of binomial distribution. May be implicit. 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 A1	Calculation of expected frequencies. 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. Significant. Suggests binomial model does not fit.	A1 A1 A1	No ft from here if wrong. ft only c's test statistic. ft only c's test statistic.	
	The model appears to overestimate in the middle and to underestimate at the tails. The biggest discrepancy is at X = 5.	E1 E1	Accept also any other sensible comment e.g. at 2.5% significance, the result would NOT have been significant.	
	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 which would involve establishing a sampling frame and using some form of random number generator.	E1 E1 E1	 Allow sensible discussion of practical limitations of choosing a random sample. Allow other sensible suggestions. E.g systematic sample - choosing every tenth family; stratified sample - by the number of girls in a family.	3
				18