4768 Statistics 3

Q1	$f(x) = k(20 - x) \quad 0 \le x \le 20$			
(a) (i)	$\int_0^{20} k(20-x)\,dx = \left[k\left(20x - \frac{x^2}{2}\right)\right]_0^{20} = k \times 200 = 1$	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 $F(x) = \int_0^x f(t)\,dt$ $= \frac{1}{200}\left(20x - \frac{x^2}{2}\right)$	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 $\mu_{II} \neq \mu_I$) where μ_D is "mean for II – mean for I"	B1 B1	Both. 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 "$\bar{X}_I = \bar{X}_{II}$" or similar unless $\bar{X}$ 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			
	$\bar{d} = 11.6$ $s_{n-1} = 17.707$	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) $\times \frac{17.707}{\sqrt{8}}$ (= 14.806) for subsequent comparison with $\bar{d}$. (Or $\bar{d}$ – (c's 2.365) $\times \frac{17.707}{\sqrt{8}}$ (=–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

		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

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

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		

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

A1</td><td>If <u>both</u> 1.465 and $0.3288/\sqrt{10}$ are correct.
If t_9 used.
95% 2-tail point for c's t distribution (Independent of previous mark).
c.a.o. Must be expressed as an interval.</td><td>4</td></tr>
<tr><td></td><td></td><td></td><td></td><td>17</td></tr>
</table>