4768 Statistics 3

Q1	W ~ N(14, 0.552) G ~ 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) = $\mathrm{P}\left(Z < \frac{145-144}{0.9} = 1.1111\right)$	M1 A1	For standardising. Award once, here or elsewhere.	
	= 0·8667	A1	c.a.o.	3
(ii)	$W+G \sim \mathrm{N}(14+144=158,$	B1	Mean.	
	$\sigma^2 = 0.55^2+0.9^2 = 1.1125\ )$	B1	Variance. Accept sd (= 1.0547…).	
	P(this > 160) = $\mathrm{P}\left(Z > \frac{160-158}{1.0547} = 1\cdot 896\right) = 1-0\cdot 9710 = 0.0290$	A1	c.a.o.	3
(iii)	$H = W_1+...+W_7+G_1+...+G_6 \sim \mathrm{N}(962,$	B1	Mean.	
	$\sigma^2 = 0.55^2+...+0.55^2+0.9^2+...+0.9^2 = 6.9775)$	B1	Variance. Accept sd (= 2.6415).	
	P(960 < this < 965) = $\mathrm{P}\left(\frac{960-962}{2\cdot 6415} = -0.7571 < Z < \frac{965-962}{2\cdot 6415} = 1.1357\right)$	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)	$D = H_1 - H_2 \sim \mathrm{N}(0,$	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			
	$\therefore h = \sqrt{13.955} \times 1.96 \qquad = 7.32$	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 <u>population</u> mean.	
	$\bar{x} = 0.957375 \qquad s_{n-1} = 0.07314(55)$	B1	s_n = 0.06842 but do <u>NOT</u> allow this here or in construction of test statistic, but FT from there.	
	Test statistic is $\dfrac{0.957375 \ -1}{\dfrac{0.07314}{\sqrt{8}}}$	M1	Allow c’s $\bar{x}$ and/or s_{n-1}. Allow alternative: 1 + (c’s – 1.895) × $\dfrac{0.07314}{\sqrt{8}}$ (= 0.950997) for subsequent comparison with $\bar{x}$. (Or $\bar{x}$ – (c’s –1.895) × $\dfrac{0.07314}{\sqrt{8}}$ (= 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 ±	M1		
	2·365	B1		
	$\times \dfrac{0.07314}{\sqrt{8}}$	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)	$\bar{x} \pm 1.96 \times \sqrt{\dfrac{0.006}{n}}$	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 E1		
	• then chooses every 10th.			
	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)
Refer to χ_3^2.
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>M1
A1
M1
A1
M1
A1
A1
A1
E1</td><td>Probs × 50 for expected frequencies.
All correct.
Calculation of X^2.
c.a.o.
Allow correct df (= cells – 1) from wrongly grouped table and ft. Otherwise, no ft if wrong.
P(X^2 > 6.937) = 0.0739.
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>