Mark Scheme (Results)

October 2020

Pearson Edexcel IAL In Statistics 1
Paper WST03/01

Question Number	Scheme	Marks
1.		
(a)	$\mathrm{E}(2\bar{X}) = 2\mathrm{E}(\bar{X}) = \dfrac{2(1+\alpha)}{2}$	M1
	$= 1+\alpha \neq \alpha$ (therefore $2\bar{X}$ is a biased estimator of α)	A1
		(2)
(b)	$\bar{x} = 6 \qquad 2\bar{x} - 1 =$	M1
	<u>11</u>	A1
		(2)
		Total 4
	Notes	
(a)	M1 for use of $2 \times \dfrac{1+\alpha}{2}$ A1cso correct comparison with α and no incorrect working seen. $\dfrac{1+\alpha}{2} \neq \alpha$ is M0A0	
(b)	M1 for attempt at $\bar{x}$ and use of $2\bar{x} - 1$ or ft use of their $\mathrm{E}(\bar{X})$ from part (a) to find an estimate for α A1 11 cao	

Question Number	Scheme	Marks
2.		
(a)	Expected values: 13.75, 19.25, 17	B1
	$\frac{(18-13.75)^2}{13.75}$ $\frac{(15-19.25)^2}{19.25}$ $\frac{(17-17)^2}{17}$	M1
	Pass | 1.31 | 0.94 | 0 | 2.25	
	9.43	A2
		(4)
(b)	H_0 : Degree/category is independent of department (no association) H_1 : Degree/category is not independent of department (association)	B1
	$\nu = (3-1)(3-1) = 4$	B1
	$\chi^2_4(5\%) = 9.488$	B1ft
	'9.43' < 9.488, so do not reject H_0 / not significant	M1
	Not enough evidence to show that the **Degree/category** is not independent of **department** (no association) (at the 5% level of significance)	A1ft
		(5)
		Total 9
	Notes	
(a)	B1 All correct expected values (may be implied by A2) M1 Attempting $\frac{(O-E)^2}{E}$ for their E (may be implied by one correct value) A2 all 5 values correct (allow awrt) (A1 any 2 values correct)	
(b)	B1 Both hypotheses required must mention degree/category and department at least once. Use of "relationship" or "correlation" or "connection" or "link" award B0. B1 4 can be implied by 9.488 seen B1ft 9.488 or better. Follow through their ν so may see $\chi^2_{3,0.05} = 7.815$ $\chi^2_{2,0.05} = 5.991$ M1 Correct non-contextual conclusion for their (a) and c.v. Can be implied by correct conclusion in context ft their (a) and c.v. A1ft A correct comment in context. Condone "relationship" or "connection" here but **not** "correlation". Follow through from their test statistic and cv, but **hypotheses must be correct.**	

Question Number	Scheme	Marks
3.		
(a)	<table><tr><td></td><td>200m</td><td>400m</td><td>d</td><td>d^2</td></tr><tr><td>A</td><td>3</td><td>1</td><td>2</td><td>4</td></tr><tr><td>B</td><td>2</td><td>2</td><td>0</td><td>0</td></tr><tr><td>C</td><td>1</td><td>4</td><td>–3</td><td>9</td></tr><tr><td>D</td><td>4</td><td>5</td><td>–1</td><td>1</td></tr><tr><td>E</td><td>7</td><td>7</td><td>0</td><td>0</td></tr><tr><td>F</td><td>6</td><td>6</td><td>0</td><td>0</td></tr><tr><td>G</td><td>5</td><td>3</td><td>2</td><td>4</td></tr></table>	M1 A1
	$\sum d^2 = 18$	A1
	$r_s = 1 - \frac{6\times 18}{7\times 48} = \frac{19}{28} =$ awrt **<u>0.679</u>**	dM1 A1
		(5)
(b)	$H_0 : \rho = 0$ $H_1 : \rho > 0$	B1
	Critical value 0.7143	B1
	$(0.679 < 0.7143)$ so insufficient evidence to reject H_0	M1
	There is insufficient evidence to suggest a **positive** correlation between the finishing order in the 200 metre race and the finishing order in the 400 metre race.	A1ft
		(4)
(c)	There are tied lengths so use average ranks	B1
	Then use $r_s = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$	dB1
		(2)
		Total 11
	Notes	
(a)	M1 for an attempt to rank athletes lists (at least 3 correct for each) A1 for correct rankings for both (may be reversed). Can be implied by correct $\sum d^2$ A1 $\sum d^2 = 18$ dM1 (dep upon previous M1) for use of the correct formula, follow through their $\sum d^2$ A1 awrt 0.679	
(b)	B1 both hypotheses correct ρ or ρ_s B1 for correct c.v. (sign should match H_1 or r_s) M1 ft their r_s and their c.v. A1 ft a correct contextualised comment including positive correlation and races/finishing orders (oe) **with all previous marks in (b) scored**	
(c)	B1 for use of average ranks or use of 1.5 or 6.5(for C and D)/4.5 or 3.5 (for B and G) dB1 (dep on 1st B1) for use of pmcc (with the average ranks)	

Question Number	Scheme	Marks
4. (a)	Randomly select a student from 1 to 7 oe	B1
	Take every 7th student	B1
		(2)
(b)	$\nu = 10-1=9$ $\quad \chi^2_9(10\%) = 14.684$	B1 B1ft
	No evidence (at 10% level of significance) that the **digits** generated do not follow a **uniform distribution**.	dB1
		(3)
(c)	H_0 : Uniform distribution is a good fit (for the two-digit numbers generated) H_1 : Uniform distribution is not a good fit (for the two-digit numbers generated)	B1
	O | E | $\frac{(O-E)^2}{E}$ | $\frac{O^2}{E}$ 31 | 40 | 2.025 | 24.025 49 | 40 | 2.025 | 60.025 30 | 40 | 2.5 | 22.5 42 | 40 | 0.1 | 44.1 48 | 40 | 1.6 | 57.6 200 | 200 | 8.25 | 208.25	B1 M1 A1
	$\sum \frac{(O-E)^2}{E} = 8.25$ or $\sum \frac{O^2}{E} - 200 = 8.25$	A1
	$\chi^2_4(10\%) = 7.779$	B1
	[8.25 > 7.779] Reject H_0 / Significant	M1
	There is evidence to suggest the two-digit **numbers** generated may not follow a **uniform distribution**.	A1
		(8)
(d)	To generate a simple random sample, Luka would need to generate **two-digit** numbers (from 00 to 69)	B1
	But Luka's table would not be suitable for generating **random** two-digit numbers	B1ft
		(2)
		Total 15
	Notes	
(a)	B1 for the idea of generating a random number to determine first student selected B1 for every 7th student (not just select 40 students)	
(b)	B1 for correct degrees of freedom (may be implied by 2nd B1) B1ft for correct critical value 14.684 or better ft their d.f. so may see $\chi^2_{10}(10\%) = 15.987$ B1 (dependent upon 2nd B1) for correct conclusion in context with digits (oe) and uniform distr.	
(c)	B1 Both hypotheses correct B1 All expected values = 40 M1 Attempting $\frac{(O-E)^2}{E}$ or $\frac{O^2}{E}$ with at least 3 correct (1 dp truncated or rounded) or f.t. their E A1 all $\frac{(O-E)^2}{E}$ or $\frac{O^2}{E}$ correct (1 dp truncated or rounded). May be implied by 8.25 or 208.25 A1 8.25 oe B1 correct critical value 7.779 or better M1 for a correct non-contextual conclusion (ignore any contradictory contextual comments for this mark) based on their cv and their test statistic A1 correct conclusion in context **with all previous marks scored**	
(d)	B1 for the idea of generating **two-digit** numbers (for a simple random sample) B1ft Luka's table not suitable for generating two-digit random numbers for a simple random sample, ft parts (b) and (c)	
SC:	If B0B0 scored, 'table only suitable for generating one-digit **random** numbers' scores B0B1	

Question Number	Scheme	Marks
5.		
(a)	$s_A^2 = \frac{1}{39}\left(790258 - 140.4^2 \times 40\right)$ $s_B^2 = \frac{1}{31}\left(581430 - 134.7^2 \times 32\right)$	M1
	= 45.4256…. **awrt 45.4** = 26.4232… **awrt 26.4**	A1 A1
		(3)
(b)	$H_0 : \mu_A - \mu_B = 5$	B1
	$H_1 : \mu_A - \mu_B > 5$	B1
	s.e. = $\sqrt{\frac{45.4256...}{40} + \frac{26.4232...}{32}}$ (= awrt 1.4)	M1
	$z = \frac{\pm(140.4 - 134.7 - 5)}{s.e.}$ = awrt 0.50	dM1 A1
	c.v. = 1.6449	B1
	(Do not reject H_0) Insufficient evidence to support the **greengrocer's belief**. (Insufficient evidence that the **difference** in **weight** between type *A* oranges and type *B* oranges is over 5 grams).	A1ft
		(7)
(c)	Large sample sizes so…	
	Sample means are normally distributed (CLT)	B1
	$s_A^2 = \sigma_A^2$ **and** $s_B^2 = \sigma_B^2$	B1
		(2)
		Total 12
	Notes	
(a)	M1 one correct expression A1 either awrt 45.4 or awrt 26.4 A1 both awrt 45.4 and awrt 26.4	
(b)	B1 Allow equivalent rearrangements. Must use μ B1 Allow equivalent rearrangements. Must use μ. For both hypotheses do not allow e.g. μ_1 and μ_2 unless each has been clearly defined M1 attempt at standard error (allow one slip) ft their (a) dM1 standardising with (140.4 – 134.7 – 5) and their s.e. (dep on previous M1) A1 for 0.5 or awrt 0.50 B1 correct c.v. 1.6449 or better. Allow –1.6449 or better with use of $\mu_B - \mu_A$ Allow p = awrt 0.309 A1ft correct ft conclusion in context with either greengrocer's belief (oe) or difference in weights (oe) **dependent on all B and M marks scored.**	
(c)	B1 must comment on both sample means, $\bar{A}$ and $\bar{B}$ B1 must comment on both variances/standard deviations	

Question Number	Scheme	Marks
6.		
(a)	$\overline{T} \sim \mathrm{N}(4, \frac{4}{35})$	M1 A1
		(2)
(b)	$\overline{K} \sim \mathrm{N}(\lambda, \frac{\lambda}{40})$	M1
	$2 \times 2.5758 \times \sqrt{\frac{\lambda}{40}} = 2.6$	B1 M1
	λ = awrt **10.2**	A1
		(4)
(c)	$2 \times 0.99 \times 0.01$	M1
	= **0.0198**	A1
		(2)
		Total 8

	Notes
(a)	M1 for Normal distribution A1 for correct mean and variance (allow N(4, awrt 0.114))
(b)	M1 for use of $\frac{\lambda}{40}$ if $\lambda = \sigma$ is used, then M0 B1 for ±2.5758 (may be implied by sight of 10.188…) M1 for use of $2 \times z \times \frac{'\sigma'}{\sqrt{40}} = 2.6$ with $\lvert z \rvert > 2$ A1 awrt 10.2 (an answer of 10.15…or awrt 10.2 on its own scores M1B0M1A1) **SC:** Use of $\sqrt{\lambda}$ instead of λ leading to an answer of awrt 3.19… scores M0B1M1A0
(c)	M1 for $2p(1-p)$ for any p $0 < p < 1$ A1 0.0198

Question Number	Scheme	Marks
7.		
(a)	$C_1 + C_2 + C_3 \sim \mathrm{N}(480,\ 3\times 1.25^2)$	M1 A1
	$\mathrm{P}(C_1 + C_2 + C_3 > 475.8) = \mathrm{P}\left(Z > \dfrac{475.8-480}{\sqrt{3\times 1.25^2}} (= -1.94)\right)$	M1
	= awrt **0.974**	A1
		(4)
(b)	$W = T_1 + T_2 + T_3 + T_4 + T_5 + C_1 + C_2 \sim \mathrm{N}(5\times 60 + 2\times 160,\ 5\times 2^2 + 2\times 1.25^2)$	M1 A1
	$\mathrm{P}(W > 625) = \mathrm{P}\left(Z > \dfrac{625-620}{\sqrt{23.125}} (= 1.03975...)\right)$	M1
	= awrt **0.149**	A1
		(4)
(c)	$Y = (n-1)T_1 - \sum_{r=2}^{n} T_r$	
	$Y \sim \mathrm{N}(\mu\ ,\ \sigma^2)$	
	$\mu = (n-1)\times 60 - (n-1)\times 60\ [= 0]$	M1 A1
	$\sigma^2 = (n-1)^2 \times 4 + (n-1)\times 4\ [= 4n^2 - 4n]$	M1 A1
	$\dfrac{40-0}{\sqrt{4n^2-4n}} = 1.38$	M1 B1
	$4n^2 - 4n - 840(.159...) = 0$	dM1
	$n = 15$	A1
		(8)
		Total 16
	Notes	
(a)	M1 for setting up Normal distribution with mean 480 A1 for correct expression for variance (= 4.6875) or for standard deviation (= 2.165…) M1 standardising with 475.8, 480 and their standard deviation (their standard deviation ≠ 3.75) A1 awrt 0.974	
(b)	M1 for setting up Normal distribution with mean 620 A1 for correct expression for variance (= 23.125) or for standard deviation (= 4.8088…) M1 standardising with 625, 620 and their standard deviation A1 awrt 0.149	
(c)	M1 for a single combined normal distribution (may be implied by a single standardisation) A1 correct expression for μ M1 for use of $a^2 \times 4 + a \times 4$ A1 correct expression for σ^2 M1 standardising with their mean and their sd = z where $1 < \lvert z \rvert < 1.5$ B1 awrt 1.38 dM1 solving their 3TQ (working must be shown if answer is incorrect) (dependent upon 2nd M1) A1 15 cao (must reject –14 if found). Must come from compatible signs in standardisation.	