Mark Scheme (Results)

January 2023

Pearson Edexcel International Advanced Level In Statistics S3 (WST03) Paper 01

Question	Scheme		Marks
1 (a)(i)	Method 1	Method 2	
	$[\bar{y} =]\frac{847}{100}[=8.47]$	$847+100\times1000$ [=100847]	M1
	So $\bar{x}=1000+\frac{847}{100}=1008.47$ *	$\bar{x}=\frac{847+1000\times100}{100}$ = 1008.47*	A1*
(ii)	$\left[s_x{}^2=s_y{}^2=\right]\frac{13510.09-100\times"8.47"^2}{99}$	$\left[s_x{}^2=\right]\frac{101707510.1-\frac{"100847"^2}{100}}{99}$	M1
	= 64		A1
			(4)
(b)	$H_0:\mu_x=1010$ $H_1:\mu_x\neq1010$		B1
			(1)
(c)	$\frac{\bar{X}-1010}{"8"/\sqrt{100}}=-1.96$ oe $\frac{\bar{X}-1010}{"8"/\sqrt{100}}=1.96$ oe		M1 B1
	$\bar{X}=1008.432$ $\bar{X}=1011.568$ awrt 1008 and 1012(or 1011)		A1
	$\bar{X}\leqslant"1008.432"$ $\bar{X}\geqslant"1011.568"$		A1ft
			(4)
(d)	1008.47 is not in the critical region		M1
	The machine does not need to be stopped /reset		A1ft
			(2)
(e)	It is reasonable since the sample size is (reasonably) large		B1
			(1)

	Notes		**Total 12**
(a)(i)	**M1**	For 8.47 or $847/100$ or $847+100\times1000$ or $847=\sum x-100\times1000$ or 100847 seen	
	A1*	cso correct solution including $\bar{x}$ = .. and …=1008.47 allow alt notation for $\bar{x}$ but must refer to *x* not *y* and must not be just *x* eg E(*X*), μ_x, mean of *x*	
(ii)	**M1**	For a correct expression ft their 100847 Allow for answer of 1064	
	A1	Cao do not ISW Allow 64.00…	
(b)	**B1**	Both hypotheses correct. Must be in terms of μ. (Allow $H_0:\mu_y=10$ $H_1:\mu_y\neq10$)	
		Mark (c) and (d) together	
(c)	**M1**	For $\pm$ standardisation with 1010 and their sd Allow equivalent eg $1010\pm n\times"8"/\sqrt{100}$ SC condone use of 1008.47 for 1010 or for $\bar{X}$	
	B1	For c.v. = ±1.96 or better seen (Calculator gives 1.95996…) Condone 1.6449 or better if they have a one tail hypotheses in (b)	
	A1	For **both** limits 1008 or better **and** 1012 or better seen. (condone 1011from correct working)	
	A1	For selecting the correct region ft their figures(not *z* value). Allow use of < and > also allow other letters(condone μ) Allow other notation eg [1012, ∞], (∞, 1008] allow [1012, ∞] , [∞, 1008]	
(d)	**M1**	ft their CR if the final A mark in part (c) is awarded. For a correct comment compatible with their CR. Must refer to 1008.47 (allow mean of *x*) is in or out of their CR Allow writing in the form "1008.432" < 1008.47 < "1011.568" etc but if in middle it must have both ends. If no clear CR it is M0A0	
	A1ft	dep on M1 awarded. Correct conclusion consistent with comparing 1008.47 with their CR(allow interval/ range etc). If it is in the CR they must say it needs to be reset/stopped. If it is not in the CR it must say it does not need to be stopped/reset. (allow equivalent wording)	

	SC	If the CR in (c) is of the form "1008.432" < $\bar{X}$ < "1011.568" oe (not z values) then award M0A1 for concluding the machine does not need to be stopped/reset.
(e)	**B1**	Any suitable comment about the sample being large eg n is large

Question	Scheme	Marks
2 (a)	<table><tr><td>**Athlete**</td><td>*A*</td><td>*B*</td><td>*C*</td><td>*D*</td><td>*E*</td><td>*F*</td><td>*G*</td><td>*H*</td></tr><tr><td>**Rank SBT**</td><td>4</td><td>2</td><td>1</td><td>3</td><td>5</td><td>6</td><td>8</td><td>7</td></tr><tr><td>**FP**</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td></tr></table>	M1
	$\sum d^2 = 9+0+4+1+0+0+1+1\ [=16]$	M1
	$r_s = 1 - \frac{6("16")}{8(63)} = 0.8095....$ awrt 0.81	dM1 A1
		(4)
(b)	$H_0 : \rho = 0$, $H_1 : \rho > 0$	B1
	Critical Value $r_s = 0.8333$ or CR: $r_s \geqslant 0.8333$	B1
	Do not reject H_0 or not significant or does not lie in the critical region or there is no evidence of a positive correlation	M1
	There is no evidence of a **positive correlation** between **season's best time** and finishing **position** for these athletes	A1ft
		(4)
(c)	$r = \frac{0.225175}{\sqrt{0.1286875 \times 0.55275}}$	M1
	$= 0.84428...$ awrt 0.844	A1
		(2)
(d)	Critical Value $r = 0.7887$ or CR: $r \geqslant 0.7887$	M1
	so there is evidence of a **positive correlation** between **season's best time** and **finishing time** for these athletes	A1 ft
		(2)

		Notes	**Total 12**
(a)	**M1**	attempt to rank seasonal best time (at least four correct), May be implied by $\sum d^2 = 16$	
	M1	Attempt to find the difference between each of the **ranks** (at least 3 correct) and evaluating $\sum d^2$ May be implied by awrt 0.81 NB if no ranks for SBT it is M0	
	dM1	dependent on 1st M1. Using $1 - \frac{6\sum d^2}{8(63)}$ with their $\sum d^2$	
	A1	$\frac{17}{21}$ or awrt 0.81(0)	
	SC	for reverse rankings May score M1M1dM1A0 order 5 7 8 6 4 3 1 2 $\sum d^2 = 158$	
(b)	**B1**	both hypotheses correct. Must be in terms of ρ (allow something that looks like rho eg *p*). Must be attached to H_0 and H_1	
	B1	critical value of 0.8333 Sign should match there H_1 or r_s	
	M1	correct statement comparing their CV with their r_s - no context needed but do not allow contradicting non contextual comments. If no CV or test statistic given or the |test value| or |CV| > 1 then it is M0	
	A1ft	correct conclusion in context for their value of r_s from (a) and their stated CV. Conclusion must refer to **positive correlation**, **seasonal best** or **time** and **position**.	
	SC	For use of two-tailed test: May score B0B1M1A0 CV allow 0.881…)	
(c)	**M1**	correct method used	
	A1	awrt 0.844	
(d)	**M1**	Critical value of 0.7887 Allow 0.8343 if hypotheses are two tailed in (b)	

A1ft	M1 must be awarded. A correct conclusion for their value of *r* from (c) Conclusion must refer to **positive correlation**, **seasonal best** or **time** and **finishing time**. Do not allow contradicting comments. if the |test value| or |CV| > 1 then it is M0

Question	Scheme	Marks
3 (a)	$\frac{86\times300}{1200}$ or $\frac{1114\times300}{1200}$	M1
	21.5 and 278.5	A1
		(2)
(b)	H_0: Making a claim and age are independent (not associated) H_1: Making a claim and age are not independent (associated)	B1
	Observed / Expected / $\frac{(O-E)^2}{E}$ 14 / "21.5" / $\frac{(14-"21.5")^2}{"21.5"} = 2.6162\ldots$ 286 / "278.5" / $\frac{(286-"278.5")^2}{"278.5"} = 0.20197\ldots$	M1
	$\sum\frac{(O-E)^2}{E} = 7.123+"2.616..."+"0.2019..."$	M1
	$= 9.941\ldots$ awrt 9.94	A1
	$\nu = (2-1)(3-1) = 2$	B1
	$\chi^2_2(0.01) = 9.210 \Rightarrow$ CR: $X^2 \geqslant 9.21[0]$	B1ft
	[in the CR/significant/Reject H_0] There is sufficient evidence to suggest that making a **claim** is not independent of **age**.	dA1ft
		(7)

	Notes	**Total 9**
(a)	**M1**	A correct method for finding one expected value. Implied by one correct value.
	A1	Correct answer for both 21.5 and 278.5
(b)	**B1**	For both hypotheses correct. Must mention claim and age at least once. Use of "relationship" or "correlation" or "connection" is B0
	M1	A correct method for finding both contributions to the χ^2 value or awrt 2.62 or awrt 0.202 Allow truncated answers of 2.61 and 0.201 May be implied by awrt 9.94
	M1	Adding their two values to 7.123 (may be implied by a full χ^2 calculation, with at least 3 correct expressions or values. Do not ISW)
	A1	awrt 9.94
	B1	$\nu = 2$ This mark can be implied by a correct critical value of 9.21 or better
	B1ft	9.21[0] or better ft their Degrees of freedom common ones $\nu = 3$ is 11.345
	dA1ft	Independent of hypotheses but dependent on both M marks being awarded. We will ft their test statistic and CV only. A correct contextual conclusion compatible with their values, which has the words claim and age. eg if they have 11.345 and 9.94 they should say it is independent/ not associated. Do not allow contradicting statements.

Full calculations for(b)

eg $\frac{(24-14.33)^2}{14.33}+\frac{(176-185.67)^2}{185.67}+\frac{(48-50.17)^2}{50.17}+\frac{(652-649.83)^2}{649.83}+\frac{(14-"21.5")^2}{21.5}+\frac{(286-"278.5")^2}{278.5}$

or awrt $6.52+$ awrt $0.5+$ awrt $0.09+$ awrt $0.01+$ awrt $2.62+0.20$

or $\frac{24^2}{14.33}+\frac{176^2}{185.67}+\frac{48^2}{50.17}+\frac{652^2}{649.83}+\frac{14^2}{"21.50"}+\frac{286^2}{"278.5"}-1200$

or awrt 40.19 + awrt 166.83 + awrt 45.92 + awrt 654.17 + awrt 9.116 + awrt 293.702 – 1200

Question	Scheme	Marks
4 (a)	H_0 : B(4, 0.5) is a suitable model H_1 : B(4, 0.5) is not a suitable model	B1
	Expected frequencies 12.5, 50, 75, 50, 12.5	M1 A1
	$\sum \frac{(O-E)^2}{E} = \frac{(15-"12.5")^2}{"12.5"} + ... + \frac{(10-"12.5")^2}{"12.5"}$ or $\sum \frac{O^2}{E} - N = \frac{15^2}{"12.5"} + ... + \frac{10^2}{"12.5"} - 200$	M1
	$= 10.84$ (or 10.8)	A1
	$\nu = 4$	B1
	$\chi_4^2(0.05) = 9.488 \quad \Rightarrow CR \geqslant 9.488$	B1
	Sufficient evidence to say that the research students claim is not supported	A1ft
		(8)
(b)	$[0\times15+]1\times68+2\times69+3\times38+4\times10[=360]$	M1
	$\frac{360}{200\times4} = 0.45$ *	A1*
		(2)
(c)	H_0 : Binomial is a suitable model H_1 : Binomial is not a suitable model	B1
	$\nu = 3$	B1
	$\chi_3^2(0.05) = 7.815 \quad \Rightarrow CR \geqslant 7.815$	B1ft
	No significant evidence to say that the binomial is not a reasonable model	B1ft
		(4)

		Notes	**Total 14**
(a)	**B1**	Both hypotheses correct. Must mention B(4 ,0.5) at least once. (may be in words need Binomial, probability (p) = 0.5 and a reference to 4 children or n = 4) Condone B(0.5, 4)	
	M1	For a correct method to find at least one expected frequency e.g. $0.5^4\times200[=12.5]$ or $4\times0.5^4\times200[=50]$ or $6\times0.5^4\times200[=75]$ May be implied by correct answer 10.84 or 10.8	
	A1	For all 5 expected frequencies correct. These must be seen and cannot be implied.	
	M1	For an attempt at the test statistic, at least 2 correct expressions/ values seen (include – 200 if needed) $\sum \frac{(O-E)^2}{E} = 0.5+6.48+0.48+2.88+0.5$ **or** $\sum \frac{O^2}{E} - N = 18+92.48+63.48+28.88+8-200$ May be implied by correct answer 10.84 or 10.8	
	A1	10.84 Allow 10.8	
	B1	$\nu = 4$ This mark can be implied by a correct critical value of 9.488	
	B1	9.488 ft their degrees of freedom if given. For $\nu = 3$ it is 7.815	
	A1ft	Dep on the 2nd M1. independent of hypotheses. Need claim or student or binomial. ft their CV and their test statistic only. A correct conclusion based on their test statistic value and their χ^2 critical value (Allow in terns of Binomial eg does not follow a binomial distribution) If their Test statistic > their CV then must say not supported (not binomial) . If their Test statistic < their CV then must say supported (is binomial)	
(b)	**M1**	A correct method for finding the total number of girls. At least 3 non zero terms correct. useful figures [0] + 68+138 + 114 + 40. Implied by 360 or 1.8	
	A1*	cso allow for 360/800 or 1.8/4 or 1.8 = 4p	
(c)	**B1**	Both hypotheses correct. Must mention binomial at least once. Condone inclusion of B(4,0.45)/B(0.45,4)	
	B1	$\nu = 3$ This mark can be implied by a correct critical value of 7.815 Condone (their ν in part(a) – 1)	
	B1ft	7.815 ft their degrees of freedom if they have (their ν in part(a) – 1)	
	B1ft	Ft their CV only. Independent of hypotheses. A correct conclusion based on awrt 2.47 and their χ^2 critical value only. Ignore any parameter given. Do not allow contradicting statements.	

Question	Scheme	Marks
5 (a)	$H_0 : \mu_A = \mu_B$ $H_1 : \mu_A > \mu_B$ oe	B1
	se = $\sqrt{\frac{17.8^2}{50}+\frac{18.4^2}{40}}$	M1
	$z = \pm\frac{1377-1368}{\sqrt{\frac{17.8^2}{50}+\frac{18.4^2}{40}}}$	M1
	$=\pm 2.339...$ awrt ± 2.34	A1
	One tailed c.v. $\lvert Z\rvert = 2.3263$ **or** CR: $Z \leqslant -2.3263$ or $Z \geqslant 2.3263$	B1
	In CR/Significant/Reject H_0	dM1
	Sufficient evidence to support that the mean **yield** from plants using fertiliser ***A*** is greater than the mean **yield** from plants using fertiliser ***B***	A1ft
		(7)
ALT	**finding the CI can get B1M1M1A0B1M1A1 unless test statistic given**	
	award M1 for $z = \pm\frac{D}{\sqrt{\frac{17.8^2}{50}+\frac{18.4^2}{40}}}$ dep on first M1 where $2.3 \leqslant z \leqslant 2.4$ May be implied by $\lvert D\rvert = 8.949$	
(b)	Expected profit per plant	
	A: $3\times1.377-\frac{75}{50}$ *B*: $3\times1.368-\frac{50}{40}$	M1
	A: £2.63(1) *B*: £2.85(4)	A1
	Claire should use fertiliser *B*	dA1 (3)
	Notes	**Total 10**

(a)	**B1**	Both hypotheses correct. Allow equivalent hypotheses. Must be in terms of μ If *A* and *B* not used the letter must be defined
	M1	For a correct attempt to find the se or se^2 Condone slip in sample sizes May be implied by se = awrt 3.85 or se^2 = awrt 14.8. Allow for a *p*-value of 0.0096 or awrt 0.0097
	M1	For an attempt to find *z* value. Allow slip in sample sizes and/or use of 17.8 and 18. 4 rather than 17.8^2 and 18.4^2 Allow for a *p*-value of 0.0096 or awrt 0.0097
	A1	awrt $=\pm 2.34$ Allow for a *p*-value of 0.0096 or awrt 0.0097
	B1	$\pm$ 2.3263 or better seen (Calculator gives 2.3263479…) must be compatible with their test statistic
	dM1	dep on previous dM1 awarded, ft their test statistic and CV only. A correct statement compatible with their test statistic and CV only – need not be contextual but do not allow contradicting non contextual comments.
	A1ft	ft their *z* value and CR only. A correct contextual statement compatible with their test statistic and CV with context of yield (at least once) and *A* and *B*
		NB id they give a *p*-value of awrt 0.0096/7 they could get B1M1dM1A1B0dM1A1
(b)	**M1**	A correct method to find the profit per *n* plants or *m* kg for either fertiliser *A* or fertiliser *B* $n\left(3\times1.377-{}^{75}\!/_{50}\right)$ or $n\left(3\times1.368-{}^{50}\!/_{40}\right)$ or $m\left(3-{}^{75}\!/_{50\times1.377}\right)$ or $m\left(3-{}^{50}\!/_{(40\times1.368)}\right)$ where n and $m\neq 0$ Implied by one correct value for *A* or *B*
	A1	must have 2 values which can be compared. ie using same *n* or *m*. Profit per *n* plant £2.63(1) *n* and £2.85(4) *n* or profit per *m* kg awrt £1.91 *m* and awrt £2.09 *m* (2dp) or cost per *m* kg awrt £1.09 *m* and awrt £0.91 *m* or number plants per £*y* awrt 0.38…*y* and awrt 0.35… *y* Useful numbers (*n* = 50 gives profit 131.55, 142.7) or (*n* = 40 gives profits105.24 and 114.16) gain M1A1
	dA1	dependent on 1st A1 being awarded. For a correct statement.

Question	Scheme	Marks
6 (a)	$\left[\bar{x} = \frac{806.4}{36} = \right] 22.4$	B1
	$"22.4" \pm 2.3263 \times \frac{0.4}{\sqrt{36}}$	M1 B1
	(22.24…, 22.55…) awrt (22.2, 22.6)	A1
	NB answers which are awrt (22.2, 22.6) gain full marks	
		(4)
(b)	[The Central Limit Theorem is not required as] the original population is **normally distributed**	B1
		(1)
(c)	22.5 is within the confidence interval	B1 ft
	So no reason to doubt the manufacturers claim	dB1 ft
		(2)
(d)	$\bar{Y} \sim \mathrm{N}\left(850, \left(\frac{5}{\sqrt{10}}\right)^2\right)$	B1
	$\mathrm{P}\left(\bar{Y} < 848\right) = \mathrm{P}\left(Z < \frac{848-850}{\frac{5}{\sqrt{10}}}\right) = [\mathrm{P}(Z < -1.26)]$	M1
	$= 0.1038$ (Calculator gives 0.10295…) awrt 0.103 / 0.104	A1
		(3)
ALT	N(8500, 250)	B1
	$\mathrm{P}\left(\bar{Y} < 848\right) = \mathrm{P}\left(Z < \frac{8480-8500}{\sqrt{250}}\right) = [\mathrm{P}(Z < -1.26)]$	M1
	$= 0.1038$	A1
	Notes	**Total 10**

(a)	**B1**	For 22.4
	M1	For use of $\bar{x} \pm z$ value $\times \frac{\sigma}{\sqrt{n}}$ with $1.2 < z < 2.6$
	B1	For z value = 2.3263 or better seen (Calculator gives 2.3263479…)
	A1	awrt (22.2, 22.6) This does not imply the B1
(b)	**B1**	For reference to the data is modelled by **normal distribution**
(c)	**B1 ft**	ft their CI For a comment on whether 22.5 (or it) is or is not in their CI allow eg range for CI Allow "22.24" < 22.5 < " 22.6" Answer must be compatible with their CI
	dB1 ft	Dependent on B1 ft. For a correct comment ft their CI eg claim is correct oe
(d)	**B1**	for $\bar{Y} \sim \mathrm{N}(850, ...)$ or $\bar{Y} < \frac{848-850}{5}$ Must have $\bar{Y}$ **or** $\mathrm{N}\left(850, \left(\frac{5}{\sqrt{10}}\right)^2\right)$ or $\mathrm{N}(850, 2.5)$ seen or used or N(8500, 250) seen or used. Both implied by a correct standardisation.
	M1	For $\pm$(a correct standardisation) implied by a correct answer
	A1	awrt 0.103 to 0.104

Question	Scheme	Marks
7 (a)	Let P = time to serve a customer at a standard checkout	
	$Q = P_1 + P_2 + P_3 \quad [Q \sim] \mathbf{N(720, 1200)}$	B1
	$\mathrm{P}(Q < 660) = \mathrm{P}\left(Z < \pm \frac{660 - "720"}{"\sqrt{1200}"}\right) \left[= \mathrm{P}(Z < -1.732...)\right]$	M1
	$= 0.0418$ (Calculator gives 0.04163…) **awrt 0.041 / 0.042**	A1
		(3)
ALT	for the B1 M1 B1 for $[Q \sim] \mathrm{N}\left(12, \frac{1}{3}\right)$ M1 for $\mathrm{P}(Q < 11) = \mathrm{P}\left(Z < \pm \frac{11 - "12"}{\sqrt{"1/3"}}\right) \left[= \mathrm{P}(Z < -1.732...)\right]$	
(b)	Assume the time taken to serve customers is independent	B1
		(1)
(c)	R = time to serve a customer at an express checkout	
	$S = (P_1 + P_2 + P_3) - (R_1 + ... + R_7) \quad [S \sim] \mathbf{N(20, 1648)}$	M1 A1
	$\mathrm{P}(S > 0) = P\left(Z > \pm \frac{0 - 20}{"\sqrt{1648}"}\right) \left[= P(Z > -0.492...)\right]$	M1
	$= 0.6879$ (Calculator gives 0.6888…) **awrt 0.688 / 0.689**	A1
ALT	For the M1A1M1 M1 for $\mathrm{N}\left(\frac{1}{3}, ...\right)$ A1 for $\mathrm{N}\left(\frac{1}{3}, \frac{103}{225}\right)$ M1 for $\pm \frac{0 - 1/3}{"\sqrt{103/225}"}$	
		(4)

	Notes	**Total 8**
(a)	**B1**	For $\mathrm{N}(720, 1200)$ or $\mathrm{N}\left(12, \frac{1}{3}\right)$ Maybe awarded if used in standardisation
	M1	For standardising using 660, their mean $\neq$ 240 or 4 and their standard deviation $\neq$ 20 or $\frac{1}{3}$. If no distribution given the mean and sd must be correct in the standardisation. Allow $\pm$ stand
	A1	awrt 0.041 or awrt 0.042
(b)	**B1**	A correct assumption. Must have context of customers or time and independence(allow random)
(c)	**M1**	For N(±20, ….) or $\mathrm{N}\left(\frac{1}{3}, ...\right)$ maybe awarded in standardisation
	A1	For N(±20, 1648) or $\mathrm{N}\left(\frac{1}{3}, \frac{103}{225}\right)$ maybe awarded if used in standardisation
	M1	For standardising using 0 and mean of $\pm$ 20 or $\pm$ 1/3 and their standard deviation. The 0 may be implied by having just the mean on the numerator Allow $\pm$ stand
	A1	awrt 0.688 to 0.689