Mark Scheme (Results)

Summer 2023

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

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications are awarded by Pearson, the UK's largest awarding body. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information visit our qualifications websites at www.edexcel.com or www.btec.co.uk. Alternatively, you can get in touch with us using the details on our contact us page at www.edexcel.com/contactus.

Pearson: helping people progress, everywhere

Pearson aspires to be the world's leading learning company. Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2023
Question Paper Log Number 73489
Publications Code WST02_01_2306_MS

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.

- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.

- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.

- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.

- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.

- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.

- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

General Instructions for Marking

The total number of marks for the paper is 75.

Edexcel Mathematics mark schemes use the following types of marks:

'M' marks
These are marks given for a correct method or an attempt at a correct method. In Mechanics they are usually awarded for the application of some mechanical principle to produce an equation, e.g. resolving in a particular direction; taking moments about a point; applying a suvat equation; applying the conservation of momentum principle; etc.

The following criteria are usually applied to the equation.

To earn the M mark, the equation
(i) should have the correct number of terms
(ii) each term needs to be dimensionally correct

For example, in a moments equation, every term must be a 'force x distance' term or 'mass x distance', if we allow them to cancel 'g' s.

For a resolution, all terms that need to be resolved (multiplied by sin or cos) must be resolved to earn the M mark.

'M' marks are sometimes dependent (DM) on previous M marks having been earned, e.g. when two simultaneous equations have been set up by, for example, resolving in two directions and there is then an M mark for solving the equations to find a particular quantity – this M mark is often dependent on the two previous M marks having been earned.

'A' marks
These are dependent accuracy (or sometimes answer) marks and can only be awarded if the previous M mark has been earned. e.g. M0 A1 is impossible.

'B' marks
These are independent accuracy marks where there is no method (e.g. often given for a comment or for a graph).

A and B marks may be f.t. – follow through – marks.

General Abbreviations

These are some of the traditional marking abbreviations that will appear in the mark schemes:

- bod means benefit of doubt
- ft means follow through
    - the symbol $\surd$ will be used for correct ft
- cao means correct answer only
- cso means correct solution only, i.e. there must be no errors in this part of the question to obtain this mark
- isw means ignore subsequent working

- awrt means answers which round to
- SC means special case
- oe means or equivalent (and appropriate)
- dep means dependent
- indep means independent
- dp means decimal places
- sf means significant figures
- * means the answer is printed on the question paper
- ⌐ ⌙ means the second mark is dependent on gaining the first mark

All A marks are 'correct answer only' (cao.), unless shown, for example, as A1 ft to indicate that previous wrong working is to be followed through. After a misread however, the subsequent A marks affected are treated as A ft, but manifestly absurd answers should never be awarded A marks.

For misreading which does not alter the character of a question or materially simplify it, deduct two from any A or B marks gained, in that part of the question affected.

If a candidate makes more than one attempt at any question:

- If all but one attempt is crossed out, mark the attempt which is NOT crossed out.
- If either all attempts are crossed out or none are crossed out, mark all the attempts and score the highest single attempt.

Ignore wrong working or incorrect statements following a correct answer.

Question Number	Scheme	Marks
1 (a)	When the data is ordinal e.g. Judges' ranks	B1
	When a non-linear relationship might be expected	B1
		(2)
(b)	$H_0 : \rho = 0$, $H_1 : \rho \neq 0$	B1
	Critical value $r_s = -0.6485$ or CR: $r_s \leqslant -0.6485$ $(\text{and } r_s \geqslant 0.6485)$	B1
	Reject H_0 or significant or lies in the critical region	M1
	The Spearman's rank correlation coefficient shows there is sufficient evidence of a correlation [between the length and maximum diameter of the melons]	A1
		(4)
(c)	$H_0 : \rho = 0$, $H_1 : \rho < 0$	B1
	Critical value $r = -0.5494$ or CR: $r \leqslant -0.5494$	B1
	The product moment correlation coefficient shows there is insufficient evidence of a **negative** correlation [between the length and maximum diameter of the melons]	B1
		(3)

	Notes	**Total 9**
(a)	**B1**	For one correct condition
	B1	For a second correct condition. Condone not underlying normal
(b)	**B1**	For both hypotheses correct. Must be in terms of ρ. Must be attached to H_0 and H_1
	B1	For critical value of –0.6485 (Allow –0.5636 if a one tailed test is stated for H_1) Condone 0.6485 if compared with 0.673
	M1	A correct statement – no context needed but do not allow contradicting non contextual comments. ft their CV provided the CV is negative (May be implied by a correct conclusion) Condone a positive CV if a comparison with 0.673 seen
	A1	For a correct conclusion which is rejecting H_0 Allow negative correlation This mark is independent of the hypotheses
(c)	**B1**	For both hypotheses correct. Must be in terms of ρ. Must be attached to H_0 and H_1
	B1	For critical value of –0.5494 (Allow –0.6319 if a two tailed test is stated for H_1) Condone 0.5494 if compared with 0.525
	B1	For a correct conclusion which is not rejecting H_0

Question Number	Scheme	Marks
2 (a)	$\frac{60\times 60}{240}$ or $\frac{60\times 84}{240}$ or $\frac{60\times 96}{240}$	M1
	15 and 21 and 24	A2
		(3)
(b)	H_0 : There is no association between the payment amount and payment method used H_1 : There is an association between the payment amount and payment method used	B1
	Observed / Expected / $\frac{(O-E)^2}{E}$ 23 / 15 / $\frac{(23-'15')}{'15'}=4.2667$ 21 / 21 / $\frac{(21-'21')}{'21'}=0$ 16 / 24 / $\frac{(16-'24')}{'24'}=2.6667$	M1
	$\chi^2 = 2.4048 + '4.2667' + '0' + '2.6667'$	M1
	$= 9.3381\ldots$ awrt 9.34	A1
	$\nu = (3-1)(3-1) = 4$ $\quad \chi^2_4(0.05) = 9.488 \Rightarrow$ CR: $X^2 \geqslant 9.488$	B1 B1ft
	[Not in the CR/Not significant/Do not reject H_0] There is no evidence of an association between **the payment amount and payment method used**	dA1
		(7)

	Notes	**Total 10**
(a)	**M1**	For a correct method for finding one expected value
	A2	For all 3 answers correct (A1 for 2 correct answers or 1 correct and 3 values that sum to 60)
(b)	**B1**	Both hypotheses correct. Must mention method **and** amount with payment at least once. (may be written in terms of independence)
	M1	For a correct method for finding all three contributions to the χ^2 value ft their part a May be implied by 3 correct values If expected values are incorrect then working must be shown
	M1	For adding their values to 2.4048 (If all 9 values are calculated the 6 values not found in part (a) must have working shown or the correct values seen or awrt 9.34)
	A1	awrt 9.34
	B1	$\nu = 4$ This mark can be implied by a correct critical value of 9.488
	B1ft	9.488 or better ft their DoF
	dA1	Dependent on both M marks. A correct contextualised conclusion which is not rejecting H_0 Must mention **method** and **amount**. If no hypotheses or they are the wrong way round, then A0 here. Contradictory statements score A0. e.g. "Significant, do not reject H_0 ".Condone "relationship" or "connection" here but **not** "correlation".

Question Number	Scheme	Marks
3 (a)	It is not a statistic as it involves <u>unknown</u> [population] parameter	B1
		(1)
(b)	$\mathrm{E}(S)=\mathrm{E}\left(\frac{3}{5}X_1+\frac{5}{7}X_2\right)=\frac{3}{5}\mathrm{E}(X_1)+\frac{5}{7}\mathrm{E}(X_2)$	M1
	$=\frac{3}{5}\mu+\frac{5}{7}\mu=\frac{46}{35}\mu\neq\mu$ So S is a biased estimator for μ	A1
		(2)
(c)	$'\frac{46}{35}\mu'-\mu=\frac{11}{35}\mu$	B1ft
		(1)
(d)	$\mathrm{E}(Y)=a\mathrm{E}(X_1)+b\mathrm{E}(X_2)=\mu$ $\Rightarrow(a+b)\mu=\mu$	M1
	$a+b=1$	A1
		(2)
(e)	$\mathrm{Var}(Y)=a^2\mathrm{Var}(X_1)+b^2\mathrm{Var}(X_2)=\left(a^2+b^2\right)\sigma^2$	M1
	$\mathrm{Var}(Y)=\left(a^2+'(1-a)'^2\right)\sigma^2$	M1
	$\mathrm{Var}(Y)=\left(2a^2-2a+1\right)\sigma^2$ *	A1*
		(3)

	Notes	Total 9
(a)	**B1**	For a correct explanation Allow σ is unknown (Do not allow σ is unknown variance)
(b)	**M1**	For writing or using $\mathrm{E}(S)=a\mathrm{E}(X_1)+b\mathrm{E}(X_2)$ Condone missing subscripts
	A1	cao (Allow $1.31\mu\neq\mu$)
(c)	**B1ft**	Follow through their part (a) $-\mu$
(d)	**M1**	For writing or using $\mathrm{E}(Y)=a\mathrm{E}(X_1)+b\mathrm{E}(X_2)=\mu$ (May be implied by $a+b=1$) Condone missing subscripts
	A1	Cao
(e)	**M1**	For writing or using $\mathrm{Var}(Y)=a^2\mathrm{Var}(X_1)+b^2\mathrm{Var}(X_2)$ Condone missing subscripts
	M1	For substitution of $b=1-a$ ft their part (d) into their expression for Var(Y)
	A1*	Answer is given so no incorrect working must be seen

Question Number	Scheme	Mark
4 (a)	$\left[\int_a^{a+1} \frac{2}{25}t \ \mathrm{d}t\right] = \frac{2}{25}\left[\frac{t^2}{2}\right]_a^{a+1}$ or $\mathrm{F}(t) = \begin{cases} 0 & t<0 \\ \frac{1}{25}t^2 & 0 \leqslant t < 5 \\ 1 & t>5 \end{cases}$ or $\frac{1}{2}\left(\frac{2}{25}(a+1)+\frac{2}{25}a\right)(a+1-a)$	M1
	$\frac{1}{25}\left((a+1)^2-a^2\right)$ or $\frac{1}{25}(a+1)^2-\frac{1}{25}a^2$ or $\left(\frac{1}{25}a+\frac{1}{25}+\frac{1}{25}a\right)$	M1
	$\frac{1}{25}\left(a^2+2a+1-a^2\right)$ oe $\left[=\frac{1}{25}(2a+1)\right]$*	A1*
		(3)
(b)	H_0 : The data could be modelled by the p.d.f H_1 : The data could not be modelled by the p.d.f	B1
	Expected frequencies: 6, 18, 30, 42, 54	M1 A1
	$\sum\frac{(O-E)^2}{E} = \frac{(10-'6')^2}{'6'}+\ldots+\frac{(68-'54')^2}{'54'}$ or $\sum\frac{O^2}{E}-N = \frac{10^2}{'6'}+\ldots+\frac{68^2}{'54'}-150$ or 2.666…+ 1.388…+ 1.2 + 1.166…+ 3.629	M1
	= 10.05… awrt 10.1	A1
	$\nu = 4$	B1
	$\chi_4^2(0.05) = 9.488 \quad \Rightarrow \mathrm{CR} \geqslant 9.488$	B1ft
	[In the CR so there is sufficient evidence to reject H_0]	
	Sufficient evidence to say that data does not fit the given p.d.f	dA1
		(8)

	Notes	**Total 11**
(a)	**M1**	For correct integration, ignore limits or finding the area of a trapezium
	M1	For substitution of the limits. May be implied by $\frac{1}{25}\left(a^2+2a+1-a^2\right)$ or simplifying the expression for the area of the trapezium
	A1*	Answer is given so no incorrect working should be seen. At least one correct line of working from the method mark to the final answer should be seen
(b)	**B1**	Both hypotheses correct. Allow H_0 : The p.d.f/ $\mathrm{f}(t)$ is a suitable model H_1 : The p.d.f/ $\mathrm{f}(t)$ is not a suitable model
	M1	For a correct method to find at least one expected frequency e.g. $\frac{1}{25}\times 150$ Ignore any reference to limits
	A1	For all 5 expected frequencies correct
	M1	For an attempt at the test statistic, at least 2 correct expressions/values ft their expected frequencies
	A1	awrt 10.1
	B1	$\nu = 4$ This mark can be implied by a correct critical value of 9.488
	B1ft	9.488 or better ft their DoF
	dA1	Dependent on 2nd M1. A correct conclusion based on their χ^2 critical value If no hypotheses or they are the wrong way round, then A0 here.

Question Number	Scheme	Marks
5 (a)	$\bar{x} \pm 1.6449 \times \frac{5}{\sqrt{10}}$	M1 B1
	$\bar{x} \pm 2.60 \Rightarrow (\bar{x} - 2.60,\ \bar{x} + 2.60)$ *	A1*
		(3)
(b)	$\bar{y} \pm 1.96 \times \frac{3}{\sqrt{20}}$	M1 B1
	$\bar{y} \pm 1.31 \Rightarrow (\bar{y} - 1.31,\ \bar{y} + 1.31)$	A1
		(3)
(c)(i)	$\bar{X} - \bar{Y} \sim \mathrm{N}\left(\mu - \mu,\ \frac{5^2}{10} + \frac{3^2}{20}\right) \Rightarrow \bar{X} - \bar{Y} \sim \mathrm{N}(0,\ 2.95)$	M1 A1
(ii)	Do not overlap when either	
	$\bar{x} - 2.60 > \bar{y} + \text{'1.31'}$ or $\bar{x} + 2.60 < \bar{y} - \text{'1.31'}$	M1
	$\bar{x} - \bar{y} > 3.91$ or $\bar{x} - \bar{y} < -3.91$	A1ft
	$2 \times \mathrm{P}(\bar{X} - \bar{Y} > 3.91) = 2 \times \mathrm{P}\left(Z > \frac{\text{'3.91'} - \text{'0'}}{\text{'}\sqrt{2.95}\text{'}}\right) = [2 \times \mathrm{P}(Z > 2.276...)]$	M1 M1
	$[2 \times 0.0113] = 0.0226$ (calculator gives $[2 \times 0.0114...] = 0.0228$)	A1
		(7)
	Notes	**Total 3**

(a)	**M1**	For use of $\bar{x} \pm z$ value $\times \frac{5}{\sqrt{10}}$
	B1	For use of z = 1.6449 or better
	A1*	Answer is given so no incorrect working should be seen (condone use of 1.645)
(b)	**M1**	For use of $\bar{y} \pm z$ value $\times \frac{3}{\sqrt{20}}$
	B1	For use of z = 1.96 or better
	A1	For $(\bar{y} - \text{awrt1.31},\ \bar{y} + \text{awrt1.31})$ Allow 1.315
(c)(i)	**M1**	For a correct method to find the variance (May be seen in a standardisation expression)
	A1	For $\mathrm{N}(0,\ 2.95)$ (May be seen in a standardisation expression) Allow $\mathrm{N}\left(0,\ \frac{5^2}{10} + \frac{3^2}{20}\right)$ oe
(ii)	**M1**	For $\bar{x} - 2.60 > \bar{y} + 1.31$ oe or $\bar{x} + 2.60 > \bar{y} - 1.31$ oe ft part (b)
	A1ft	For $\bar{x} - \bar{y} > \text{'3.91'}$ or $\bar{x} - \bar{y} < -\text{'3.91'}$ ft part (b)
	M1	For multiplying by 2 (may be seen at any stage of their working)
	M1	For standardising ft their 3.91, their mean and their standard deviation (Do not allow use of 2.6 or 1.31 as their 3.91)
	A1	For answers in the range awrt 0.0226 – awrt 0.0228

Question Number	Scheme	Marks
6 (a)	$\alpha = 5.1$	B1
	$\beta = \sqrt{\dfrac{1694.65 - 65\times('5.1')^2}{64}}$	M1
	$= 0.25$	A1
		(3)
(b)	$H_0 : \mu_A = \mu_B$ $H_1 : \mu_A < \mu_B$	B1
	$z = \pm\dfrac{5.0 - '5.1'}{\sqrt{\dfrac{0.24^2}{70} + \dfrac{'0.25'^2}{65}}}$	M1 M1
	$= -2.367...$ awrt -2.37	A1
	One tailed c.v. $z = -1.6449$ **or** CR: $z \leqslant -1.6449$	B1
	In CR/Significant/Reject H_0	M1
	Sufficient evidence to support Roxane's claim	A1
		(7)
(c)	Since the sample is **large** the **CLT** applies.	M1
	No [need to assume that the fat content is normally distributed]	A1
		(2)
(d)	Assumed that $s^2 = \sigma^2$ in **both** groups	B1
		(1)
	Notes	**Total 13**

(a)	**B1**	cao
	M1	For a correct method to find β using their α
	A1	Cao
(b)	**B1**	Both hypotheses correct. Allow equivalent hypotheses. Must be in terms of μ
	M1	For correct standard error ft their *s* in part a
	M1	For an attempt to find the test statistic, ft their SE and their α
	A1	awrt –2.37 (Allow 2.37)
	B1	–1.6449 or better (seen) (Allow 1.6449 or better if comparing to their 2.37)
	M1	A correct statement – need not be contextual but do not allow contradicting non contextual comments. ft their CV and test statistic
	A1	A correct contextual statement e.g sufficient evidence to support that crisps from **brand A** have a **lower fat** content than the crisps from **brand B** (must include the words in bold)
(c)	**M1**	A suitable comment that mentions large and CLT
	A1	A correct answer, context not required.
(d)	**B1**	For the assumption that sample variance = population variance for **both** groups

Question Number	Scheme	Marks
7 (a)	$\mathrm{E}(X) = 4\times15 - 3\times10[=30]$	M1
	$\mathrm{Var}(X) = 4^2\times5^2 + 3^2\times4^2[=544]$	M1
	So $X \sim \mathrm{N}(30,\ 544)$	
	$\mathrm{P}(X<40) = \mathrm{P}\left(Z < \dfrac{40-'30'}{'\sqrt{544}'}\right)\left[=\mathrm{P}\left(Z<0.428...\right)\right]$	M1
	$= 0.6664$ (Calculator gives 0.6659…) awrt 0.666	A1
		(4)
(b)	$\mathrm{E}\left(A+B+D\right) = 15+10+3\times20 = [85]$	M1
	$\mathrm{Var}(A+B+D) = 5^2+4^2+3\times\sigma^2 = [41+3\sigma^2]$	M1
	So $A+B+D \sim \mathrm{N}(85,\ 41+3\sigma^2)$	
	$\mathrm{P}\left(A+B+D<76\right) = \mathrm{P}\left(Z < \dfrac{76-85}{\sqrt{41+3\sigma^2}}\right) = 0.242$	
	So $\dfrac{-9}{\sqrt{41+3\sigma^2}} = -0.7$ or $\dfrac{9}{\sqrt{41+3\sigma^2}} = 0.7$ (Calculator gives –0.69988…)	M1 A1
	$3\sigma^2 = \left(\dfrac{-9}{-0.7}\right)^2 - 41$	dM1
	$\sigma = 6.437...$ awrt 6.44	A1
		(6)

	Notes	**Total 10**
(a)	**M1**	For a correct method to find $\mathrm{E}(X)$. May be implied by a correct standardisation expression.
	M1	For a correct method to find $\mathrm{Var}(X)$ Allow $\sqrt{544}$ oe or 23.3^2 or better. May be implied by a correct standardisation expression.
	M1	For standardising $(\pm)$ using their mean and their variance
	A1	awrt 0.666
(b)	**M1**	For a correct method to find $\mathrm{E}(A+B+D)$
	M1	For a correct method to find $\mathrm{Var}(A+B+D)$
	M1	For standardising $(\pm)$ using their mean and their standard deviation which is in terms of σ^2 and setting equal to –0.7 or better. Allow +0.7
	A1	For the correct equation
	dM1	Dependent on the previous M mark. For squaring and rearranging leading to an equation in σ^2
	A1	awrt 6.44 (Do not award if previous A mark was not awarded)