Mark Scheme (Results)

January 2024

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

January 2024
Question Paper Log Number P75603A
Publications Code WST03_01_2401_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.
- When examiners are in doubt regarding the application of the mark scheme to a candidate's response, the team leader must be consulted.
- 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 $\sqrt{}$ 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.

<u>Special notes for marking Statistics exams (for AAs only)</u>

- Any correct method should gain credit. If you cannot see how to apply the mark scheme but believe the method to be correct then please send to review.
- For method marks, we generally allow or condone a slip or transcription error if these are seen in an expression. We do not, however, condone or allow these errors in accuracy marks.

Question Number	Scheme	Marks
1.	$\text{H}_0:$ There is no association between treatment and presence of fungus $\text{H}_1:$ There is association between treatment and presence of fungus	B1

Expected	No treatment	Sulphur	Copper sulphate
No Fungus	$\frac{123\times 30}{150}$ [=24.6]	$\frac{123\times 63}{150}$ [=51.66]	$\frac{123\times 57}{150}$ [=46.74]
Fungus	$\frac{27\times 30}{150}$ [=5.4]	$\frac{27\times 63}{150}$ [=11.34]	$\frac{27\times 57}{150}$ [=10.26]

M1

Observed	Expected	$\frac{(O-E)^2}{E}$	$\frac{O^2}{E}$
20	24.6	0.86016…	16.2601...
55	51.66	0.21594…	58.5559…
48	46.74	0.03396…	49.2939…
10	5.4	3.91851…	18.5185…
8	11.34	0.98373…	5.6437…
9	10.26	0.15473…	7.8947…
	Totals:	6.167…	156.167…

dM1

$\text{X}^2 = \sum\frac{(O-E)^2}{E}$ or $\sum\frac{O^2}{E} - 150$ dM1

= awrt 6.17 A1

$\nu = (3-1)(2-1) = 2$ B1

$\chi_2^2(0.05) = 5.991$ B1ft

[Reject H_0/significant/in the CR] There is sufficient evidence to suggest there is an association between treatment and presence of fungus. A1ft

[8]

Notes

1st B1 both hypotheses correct with treatment (oe) and fungus (oe) (treatment and fungus need to only appear in either H_0 or H_1). May be written in terms of independence.

1st M1 attempt at $\frac{\text{row total} \times \text{column total}}{\text{total}}$ (can be implied by at least one correct E_i to 1dp)

2nd M1 (dep on 1st M1) at least 2 correct terms for $\frac{(O-E)^2}{E}$ or $\frac{O^2}{E}$ or correct expressions with their E_i (allow 2sf accuracy) (May be implied by awrt 6.17 or awrt 156.17)

3rd M1 (dep on 2nd M1) for using $\sum\frac{(O-E)^2}{E}$ or $\sum\frac{O^2}{E} - 150$ (May be implied by awrt 6.17)

1st A1 awrt 6.17

2nd B1 DoF/$\nu = 2$ (May be implied by 5.991)

3rd B1ft 5.991 (or better) allow ft from their stated degrees of freedom)

2nd A1ft (dep on 3rd M1 and 3rd B1) for a correct ft contextualised conclusion. Must include “treatment” and “fungus”. Ignore any non-contextual statements. If hypotheses are the wrong way round then A0.

Question Number	Scheme	Marks
2. (a)	**Number** all employees [1-800] Use a **random** number to select the **first employee** oe Then select every **10^{th}** employee from the list of employees	B1 B1 B1 (3)
(b)	**Number** all employees **by city/for each city** Use **random** numbers to select **54** employees from London, **31** employees from Edinburgh and **15** employees from Cardiff	B1 B1 B1 (3)
(c)	e.g Stratified sample reflects the population structure	B1 (1) **[7]**

	Notes
(a)	1st B1 idea of **numbering** all employees 2nd B1 idea of **randomly** selecting a starting point 3rd B1 selecting every $\boldsymbol{k}^{\text{th}}$ employee
(b)	1st B1 idea of **numbering** employees for **each city** 2nd B1 use of **random** numbers (oe) 3rd B1 **54** from London, **31** from Edinburgh, **15** from Cardiff cao
(c)	Any correct advantage e.g. Allows calculations [of statistics] for each city/group

Question Number	Scheme	Marks
3. (a)	$H_0 : \rho = 0$ $H_1 : \rho \neq 0$	B1
	$[r=]\dfrac{83.634}{\sqrt{2.486 \times 3026.234}} = 0.9642.....$ awrt 0.964	M1 A1
	CV = 0.7545	B1ft
	[Reject H_0/Significant] There is evidence of correlation between annual tea consumption and population.	A1
		(5)
(b)	**Country** A B C D E F G ***T* Rank** 5 6 4 7 1 2 3 ***P* Rank** 7 6 4 3 1 2 5 or **Country** A B C D E F G ***T* Rank** 3 2 4 1 7 6 5 ***P* Rank** 1 2 4 5 7 6 3	M1
	$\sum d^2 = 4+0+0+16+0+4[=24]$	M1
	$[r_s =]1-\dfrac{6(24)}{7(48)} = 0.571428...$ awrt 0.571	dM1A1
		(4)
(c)	$H_0 : \rho_s = 0$ $H_1 : \rho_s > 0$	B1
	CV = 0.7143	M1
	[Do not reject H_0 / not significant] There is not enough evidence to suggest a positive correlation between annual tea consumption and population.	A1ft
		(3)
		[12]

	Notes
(a)	1st B1 both hypotheses correct in terms of ρ (must be two-tailed). Condone use of *p* M1 use of formula for *r* (May be implied by awrt 0.964) 1st A1 awrt 0.964 2nd B1ft 0.7545 (or better) or ft 1-tailed alternative hypothesis (0.6694) 2nd A1 correct contextual conclusion including tea consumption/*t* and population/*p*. Must be consistent with their *r* and their CV. (Ignore any non-contextual conclusion) Allow positive correlation
(b)	1st M1 attempt to rank each country for tea and population (at least 4 correct in each) 2nd M1 for $\sum d^2$ for their ranks (implied by $\sum d^2 = 24$) 2nd M1 (dep on 1st M1) use of $1-\dfrac{6('24')}{7(48)}$ A1 awrt 0.571 (or $\dfrac{4}{7}$)
(c)	B1 both hypotheses correct in terms of ρ or ρ_s. Condone use of *p* M1 0.7143 (or better) A1ft correct contextual conclusion including positive, tea consumption/*t* and population/*p*. (Ignore any non-contextual conclusion) ft their part (b)

Question Number	Scheme	Marks
4. (a)	$\dfrac{[0\times 24]+1\times 34+2\times 28+3\times 21+4\times 8+5\times 5}{120}\left[=1.75\right]*$	B1*cso (1)
(b)	$\left[s=\right]120\times\dfrac{e^{-1.75}1.75^4}{4!}$ [= 8.15] * or $\left[s=\right]120-\left(20.85+36.49+31.93+120\times\dfrac{e^{-1.75}1.75^3}{3!}+3.95\right)$[= 8.15] *	B1*cso (1)
(c)	$\left[r=\right]18.63$	B1 (1)
(d)	$H_0:$ Poisson distribution is a good fit. $H_1:$ Poisson distribution is not a good fit	B1
	$\sum\dfrac{(O_i-E_i)^2}{E_i}=1.43+\dfrac{(8+5-(8.15+3.95))^2}{8.15+3.95}$	M1 M1
	= 1.49694… awrt 1.5(0)	A1
	$\nu=5-1-1=3$	B1
	$\chi_3^2(0.05)=7.815$	B1ft
	[Do not reject H_0/not significant] There is insufficient evidence to reject the office manager's belief or the number of jobs sent to the printer are consistent with a Poisson distribution.	A1 (7)
		[10]

	Notes
(a)	B1cso correct calculation, minimum working $\dfrac{34+56+63+32+25}{120}=1.75*$
(b)	B1cso fully correct calculation (may be seen in stages) leading to 8.15*
(c)	For 18.63 (This may be seen in part (b) if labelled as *r*)
(d)	1st B1 both hypotheses correct (mention of 1.75 is B0) 1st M1 evidence of combining last 2 cells e.g. 8 + 5 and 8.15 + 3.95 2nd M1 use of 1.43 + $\sum\dfrac{(O_i-E_i)^2}{E_i}$ for remaining cells (Condone cells not combined. May be implied by 1.43 + 0.00276…+ 0.279… or awrt 1.71) 1st A1 awrt 1.50 (allow 1.5 from correct working) 2nd B1 Dof/ $\nu=3$ implied by a correct critical value of 7.815 3rd B1ft 7.815 (allow ft on the ν so may see 9.488 or 11.070 etc) 2nd A1 (dep on 2nd M1) a correct conclusion which states that the office manager's belief is correct/the data are consistent with a Poisson distribution which must be consistent with the test statistic and CV. Condone Po(1.75) is a suitable model. This mark is independent of the hypotheses

Question Number	Scheme	Marks
5. (a)	$H_0: \mu_H - \mu_M = 15$ $\quad$ $H_1: \mu_H - \mu_M > 15$	B1
	$z = \dfrac{56.3 - 39.8 - 15}{\sqrt{\dfrac{27.2}{38} + \dfrac{18.5}{45}}}$	M1 M1
	= 1.4130…. $\quad$ awrt 1.41	A1
	CV = 1.6449 (or better) or p = awrt 0.0788	B1
	Do not reject H_0/Not significant	M1
	There is not sufficient evidence to support the professor's claim/there is not sufficient evidence to suggest that undergraduates studying History type more than 15 words/minute faster than undergraduates studying Maths.	A1 (7)
(b)	$s^2 \approx \sigma^2$ for **both** History and Maths	B1
	Assume sample sizes are large enough so that CLT applies or $\bar{X}$ is normally distributed for **both**	B1 (2) **[9]**

	Notes
(a)	1st B1 both hypotheses correct in terms of μ_H and μ_M Allow equivalent rearrangements. Allow other letters as long it is clear which is History and which is Maths Must be attached to H_0 and H_1 1st M1 for $z = \dfrac{a - b - 15}{\sqrt{\dfrac{c}{38} + \dfrac{d}{45}}}$ with at least 2 of a, b, c or d correct (allow ±) 2nd M1 for $z = \dfrac{56.3 - 39.8 - 15}{\sqrt{\dfrac{27.2}{38} + \dfrac{18.5}{45}}}$ (allow ±) 1st A1 awrt 1.41 2nd B1 for CV = ±1.6449 and compatible sign with their test statistic (allow p = awrt 0.0788) 3rd M1 correct statement consistent with their test statistic and CV (no contradictory non-contextual comments) May be implied by correct contextual comment. 2nd A1 contextual conclusion that is consistent with their test statistic and their CV. Must mention professor's claim or History, Maths and typing (oe).
(b)	1st B1 must mention both. Allow $s \approx \sigma$ for both History and Maths 2nd B1 either correct assumption

Question Number	Scheme	Marks
6. (a)	$[\bar{x} = 49.8]$	
	$2\times1.96\left(\frac{\sigma}{\sqrt{8}}\right) = 53.88 - 45.72 = 8.16$ | $49.8 + 1.96\left(\frac{\sigma}{\sqrt{8}}\right) = 53.88$ or $49.8 - 1.96\left(\frac{\sigma}{\sqrt{8}}\right) = 45.72$	M1
	$2\times2.5758\left(\frac{\sigma}{\sqrt{8}}\right) = \frac{8.16\times2.5758}{1.96} = 10.7238...$ | $2.5758\left(\frac{\sigma}{\sqrt{8}}\right) = \frac{4.08\times2.5758}{1.96} = 5.3618...$	B1 M1
	99%CI = $49.8 \pm \frac{10.7238}{2}$ | 99%CI $= 49.8 \pm 5.3618$	M1
	$= (44.438..., 55.1619...)$ (awrt 44.4, awrt 55.2)	A1 (5)
(b)	$\hat{\mu} = \bar{x} = \frac{91.2}{8} = 11.4$	B1
	$\hat{\sigma}^2 = s^2 = \frac{1145.16 - 8\times"11.4^2"}{7} = 15.06857...$ awrt 15.1	M1 A1 (3)
(c)	Combined $\Sigma x = 10.8\times24 + 91.2 = 350.4$	M1
	Combined $\Sigma x^2 = 1145.16 + 23\times17.64 + 24\times10.8^2 = 4350.24$	M1A1
	Combined $s^2 = \frac{"4350.24" - 32\times\left(\frac{"350.4"}{32}\right)^2}{31} = 16.56$	M1 A1
	$\frac{s}{\sqrt{n}} = \frac{\sqrt{16.56}}{\sqrt{32}} = 0.719374...$ awrt 0.719	M1 A1 (7) **[15]**

	Notes
(a)	1st M1 use of $2z\frac{\sigma}{\sqrt{n}}$ or $z\frac{\sigma}{\sqrt{n}}$ with 1.5 < |z| < 2. Allow σ_m for $\frac{\sigma}{\sqrt{n}}$
	B1 1.96 (or better) and 2.5758 (or better)
	2nd M1 attempt to find width or semi-width of 99% CI with |z| > 2 Allow $\sigma = \frac{4.08\times\sqrt{8}}{1.96}[= 5.887...]$
	3rd M1 Use of 49.8 ± awrt 5.36 or $49.8 \pm 2.5758\left(\frac{"5.887..."}{\sqrt{8}}\right)$ If σ is incorrect then working must be shown.
	A1 correct interval with (awrt 44.4, awrt 55.2)
	Correct answer from less accurate *z* –values scores M1B0M1M1A1
(b)	B1 11.4 cao
	M1 full attempt at s^2 ft their $\bar{x}$
	A1 awrt 15.1
(c)	M1 for correct combined sum (may be implied by combined mean of 10.95)
	2nd M1 for attempt at combined sum of squares 1145.16 + $(n - 1)$ ×17.64 + n ×10.8^2 (allow 1 error)
	1st A1 fully correct expression or awrt 4350
	3rd M1 using their values in a complete expression for combined s^2 oe
	2nd A1 s^2 =16.56 or s = awrt 4.07 (either of these implies M1M1A1M1A1)
	4th M1 use of $\frac{s}{\sqrt{n}}$ with combined values
	3rd A1 awrt 0.719

Question Number	Scheme	Marks
7. (a)	$a = 2\times180 - 330 = 30$	B1
	$b = 4.5^2\times2 + 6.7^2 = 85.39$	M1 A1
		(3)
(b)	$X = L - 1.8S$	
	$\text{E}(X) = 330 - 1.8\times180 = 6$	M1
	$\text{Var}(X) = 6.7^2 + 1.8^2\times4.5^2 = 110.5$	M1 A1
	$\text{P}(X > 0) = \text{P}\left(Z > \frac{0-6}{\sqrt{110.5}}\right)$	M1
	$\text{P}(Z > -0.57) = 0.7157$	A1
		(5)
(c)	$T = S_1 - \frac{S_1+S_2+S_3}{3} = \frac{2S_1 - S_2 - S_3}{3}$	M1 A1
	E(*T*) = 0	M1
	$\text{Var}(T) = \frac{1}{9}\left(2^2\times4.5^2 + 4.5^2 + 4.5^2\right) = \frac{6}{9}\left(4.5^2\right) = 13.5$	M1
	$\text{P}(T > 5) = \text{P}\left(Z > \frac{5-0}{\sqrt{13.5}}\right)$	M1
	P(*Z* > 1.36) = 1 – 0.9131 = 0.0869	A1
		(6)
		[14]
	Notes	
(a)	B1 30 cao	
	M1 $2\times\text{Var}(S) + \text{Var}(L)$	
	A1 85.39 (allow 85.4)	
(b)	1st M1 Seeing or using E(*X*) = 6 or correct expression for mean	
	2nd M1 $\text{Var}(L) + 1.8^2\,\text{Var}(S)$ (condone mixing variances for M1)	
	1st A1 for 110.5 (allow 65.61 + 6.7^2)	
	3rd M1 standardising with their mean and s.d. leading to a probability $p > 0.5$	
	2nd A1 awrt 0.716 [calc: 0.7159262…]	
(c)	1st M1 realising the need to write as a single distribution using $\bar{S} = \frac{S_1+S_2+S_3}{3}$	
	1st A1 for $\frac{2S_1 - S_2 - S_3}{3}$	
	2nd M1 Using mean = 0	
	3rd M1 using Var(*aS*) = a^2Var(*S*)	
	4th M1 standardising with their mean and sd	
	2nd A1 awrt 0.0868 to awrt 0.0869 [calc: 0.08678…]	
	Note: Assuming S_1 and $\bar{S}$ are independent, leads to E(*T*) = 0, Var (*T*) = 27, P(*T* > 5) = 0.167… scores M0A0M1M0M1A0	