Mark Scheme (Results)

January 2021

Pearson Edexcel International Advanced Level
In Statistics 1 (WST03/01)

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications come from Pearson, the world's leading learning company. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information, please visit our website at www.edexcel.com.

Our website subject pages hold useful resources, support material and live feeds from our subject advisors giving you access to a portal of information. If you have any subject specific questions about this specification that require the help of a subject specialist, you may find our Ask The Expert email service helpful.

www.edexcel.com/contactus

Pearson: helping people progress, everywhere

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 2021
Publications Code WST03_01_2101_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.

Special notes for marking Statistics exams (for AAs only)

- If a method leads to "probabilities" which are greater than 1 or less than 0 then M0 should be awarded unless the mark scheme specifies otherwise.
- 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.
- If a candidate is "hedging their bets" e.g. give Attempt 1…Attempt 2…etc then please send to review.

Question Number	Scheme	Marks
1. (a)	**[In QP:** *33, 15, 23* **]** 29, 34, 39, 06, 31, 13, 42	M1A1 (2)
(b)	This will give 4 girls with numbers 15, 23, 06, 13 This will give 6 boys with numbers 33, 29, 34, 39, 31, 42	B1 B1 (2)
(c)	Since the highest number is 42 …therefore may miss <u>older players</u>	M1 A1 (2) **[6 marks]**
Notes		
(a)	M1 for 7 numbers (at least 4 correct in any order) (Condone repeats but only count once towards the "4") e.g. <u>29</u>, 33, <u>34</u>, <u>39</u>, 15, 29, <u>31</u> The 33 and 15 are repeats of those in QP and 29 is a repeat but all will count for the "7" This will score M1 as there are 4 of the correct numbers listed: 29, 34, 39 and 31 A1 for all 7 correct with no repeats	
(b)	1st B1 for showing the 4 girls in sample (No ft for incorrect random numbers) 2nd B1 for showing the 6 boys in the sample (No ft for incorrect random numbers)	
(c)	M1 for mention of highest number of 42 (or ft their highest number as long as < 60) A1 for stating that this means older players may be missing from the sample This can be awarded if their highest number is stated for M1 and is < 42	

Question Number	Scheme	Marks
2. (a)	<table><tr><td>**Student**</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><td>*I*</td><td>*J*</td><td>*K*</td></tr><tr><td>**Objects rank**</td><td>9</td><td>6</td><td>8</td><td>2</td><td>1</td><td>10</td><td>7</td><td>3</td><td>5</td><td>4</td><td>11</td></tr><tr><td>**Maths rank**</td><td>11</td><td>4</td><td>5</td><td>1</td><td>2</td><td>9</td><td>3</td><td>7</td><td>8</td><td>6</td><td>10</td></tr></table>	M1
		M1
	$\sum d^2 = 4+4+9+1+1+1+16+16+9+4+1 = 66$	M1
	$r_s = 1 - \frac{6\times"66"}{11\left(11^2-1\right)}$; = **<u>0.7</u>**	dM1; A1
		(5)
(b)	$H_0 : \rho = 0$ $H_1 : \rho > 0$	B1
	Critical value (n = 11 5% one-tail) is 0.5364	B1
	(Significant result so) there is evidence to support the teacher's belief <u>or</u> there is evidence of a positive correlation between short term memory and mathematical ability (o.e.) <u>or</u> evidence that students with strong maths ability also have good short term memory (o.e.)	B1
		(3)
(c)	Data shows positive correlation but does not necessarily imply that enhanced short term memory <u>causes</u> increase in mathematical ability.	B1
		(1)
		[9 marks]

Notes

(a)

1st M1 for attempt to rank one row with at least 5 correct (could be reversed)
2nd M1 for both rows ranked with at least 5 correct in each row (one or both reversed)
3rd M1 for an attempt at $\sum d^2$ ft their values and at least 5 correct
4th dM1 (dep on at least one M1) for use of their $\sum d^2$ in a correct formula
A1 for 0.7 or exact equivalent

(b)

1st B1 for both hypotheses in terms of ρ or ρ_s [If $r_s < 0$ in (a) allow $H_1 : \rho < 0$]
2nd B1 for critical value of 0.5364 (sign compatible with r_s) [If $r_s < 0$ in (a) need -0.5364]
Allow 0.6182 if 1st B0 for $H_1 : \rho \neq 0$
3rd B1 for correct conclusion in context. Penalise contradictory comments e.g. "not significant so supports teacher's belief" [No ft]

(c)

B1 for a comment that states that correlation does <u>not</u> imply <u>causation</u>
Need to see "cause" or "causation" clearly mentioned.

Question Number	Scheme	Marks
3. (a)	All expected frequencies are $(88 \div 4)$ = **22**	B1
	Degrees of freedom = 3, so critical value $\chi_3^2(5\%) = 7.815$	B1, B1ft
	(Not significant so) insufficient evidence to suggest not uniformly distributed	B1
		(4)
(b)	e.g. H_0: School is independent of club chosen H_1: Club chosen depends on which school a student is from	B1
		(1)
(c)	$\frac{28 \times 17}{88} = 5.409...$ awrt **5.41**	B1
		(1)
(d)	Expected frequency for Music and School C = 4.77 < 5 (Allow $\frac{105}{22}$ for 4.77)	B1
	So combine Music column with another column giving 3x3 table so 4 df	B1
		(2)
(e)	Critical value $\chi_4^2(5\%) = 9.488$	B1
	[Not significant so] insufficient evidence of an association between school and choice of club	B1
		(2)
		[10 marks]

Notes

Ignore values of any test statistics calculated in (a) or (e)

(a)
1st B1 for 22
2nd B1 for degrees of freedom = 3 (can be implied by sight of 7.815 as cv)
3rd B1ft for 7.815 (or better - cal: 7.814727910… or correct 5% cv for their d.f.)
4th B1 for comment suggesting uniform distribution is a suitable model.
Must follow from comparing 6.09 with their cv.
Do not allow contradictory statements e.g. "significant" so uniform dist' is suitable

(b)
B1 for both hypotheses with some context ("club" and "school" mentioned at least once)
Use of "independence" or "association"

(c)
B1 for a correct expression or awrt 5.41 (allow $\frac{119}{22}$)

(d)
1st B1 for identifying that Music & School C has E_i that is < 5
(a value to 2 sf should be seen, may be in (c), but must state this E_i < 5 as well)
2nd B1 for pooling music with another column leading to 3x3 table and 4 degrees of freedom
Must clearly state the pooling and evidence for 4 df e.g. allow $(3-1)\times(4-1-1)$

[NB pooling with Art gives 4.3987.., with Sports 4.3247..., with Computers 7.2879...]

(e)
1st B1 for 9.488 (or awrt 9.488)
2nd B1 for a correct, not significant, conclusion mentioning school and clubs

Question Number	Scheme	Marks
4. (a)	Use of $\bar{x} \pm z \times \frac{18}{\sqrt{25}}$; z = 2.3263 (or better)	M1;B1
	= (44.0253…, 60.7746…) awrt **(44.0, 60.8)**	A1, A1
		(4)
(b)	$H_0 : \mu_A = \mu_B$ $H_1 : \mu_B > \mu_A$	B1
	$z = (\pm)\frac{57.8-52.4}{18\sqrt{\frac{1}{25}+\frac{1}{30}}}$	M1dM1
	= (±) 1.1078… awrt (±) **1.11**	A1
	5% one-tail critical value is 1.6449 (or p-value = 0.13396… i.e. awrt 0.134)	B1
	(not sig') so insufficient evidence (in these data) to support newspaper's claim	A1
		(6)
(c)	Require $\frac{\bar{x}-\mu}{\frac{18}{\sqrt{n}}} > z$ where z = – 1.6449 (o.e.)	M1
	$\mu < 52.4+1.64(49)\times\frac{18}{5}$ or $\mu < 57.8+1.64(49)\times\frac{18}{\sqrt{30}}$	A1
	i.e. $\mu < 58.3216...$ and $\mu < 63.2056...$	M1
	So μ = **58.3**	A1
		(4)
		[14 marks]

Notes

(a)

M1 for use of correct expression with 18, 25 and $1 < z < 3$ (Ignore $\bar{x}$ for this mark)

B1 for z = 2.3263 or better (calc: 2.32634787..)

1st A1 for awrt 44.0 (ans only of 44.02…or awrt 44.03 scores M1B1 implied)

2nd A1 for awrt 60.8 (ans only of 60.77… or awrt 60.77 scores M1B1 implied)

(b)

1st B1 for both hypotheses in terms of μs (If using μ_1 etc they must define which is which)

1st M1 for a correct denominator (18 needn't be outside square root) [4.87(44…)]

2nd dM1 for a correct expression for test statistic

1st A1 for awrt (±) 1.11

2nd B1 for critical value of 1.6449 or better (If B0 in (a) for 2.33 allow 1.64 or 1.645 here)
[Allow p-value of awrt 0.134 and condone awrt 0.866 if compared with 0.95]

2nd A1 Correct contextual conclusion, ft comparing their "1.11" with 1.64 (or their cv) **but** must be not significant and mention "claim" **or** "score in town *A*" and "score in town *B*"

(c)

1st M1 for a correct starting inequality with any z such that $|z| > 1$ (Allow $\geqslant$)

1st A1 for either correct inequality for μ, allow z = 1.64 or better

2nd M1 for both cases of $\bar{x} + z\frac{18}{\sqrt{n}}$ ($z > 1$) can allow "=" or inequality, may be in CI

2nd A1 (dep on both Ms) for sight of both awrt 58.3 and awrt 63.2 and selecting awrt 58.3

Question Number	Scheme	Marks
5. (a)	H_0 : $N(6, 0.75^2)$ is a suitable model for the length of fallen pine cones H_1 : $N(6, 0.75^2)$ is NOT a suitable model for the lengths of the pine cones	B1
	e.g. E_i : $5 \leqslant x < 5.5 = 80 \times P(5 \leqslant X < 5.5) = 80 \times P(-\frac{4}{3} \leqslant Z < -\frac{2}{3})$ [= **12.77~12.90]** or E_i : $6 \leqslant x < 6.5 = 80 \times P(0 \leqslant Z < \frac{2}{3})$ [= **19.80~19.89**]	M1 A1
	E_i : $5.5 \leqslant x < 6$ = **19.80~19.89** or $x \geqslant 6.5$ = 40 – "19.80" = **20.11~20.20**	M1
	$x < 5$ | $5 \leqslant x < 5.5$ | $5.5 \leqslant x < 6$ | $6 \leqslant x < 6.5$ | $x \geqslant 6.5$ E_i: **7.30~7.43** | 12.77~12.90 | 19.80~19.89 | 19.80~19.89 | 20.11~20.20 $\frac{(O-E)^2}{E}$: 0.23~0.28 | 0.093~0.12 | 0.84~0.90 | 1.87~1.95 | 5.08~5.16	A1
	$\sum \frac{(O_i - E_i)^2}{E_i}$ or $\sum \frac{O_i^2}{E_i} - 80$ = 8.308… ; answer in [**8.15 ~ 8.4]**	dM1; A1
	$\nu = 5 - 1 = 4 \implies$; $\chi_4^2(10\%) = 7.779$	B1; B1ft
	(significant result so) the data do not support Chrystal's belief	A1ft (10)
(b)	$\hat{\mu} = \frac{464}{80}$ = **<u>5.8</u>** (cm) ; $s^2 = \frac{2722.59 - 80 \times "5.8^2"}{79}$	B1; M1
	$s^2 = 0.39734\ldots$ awrt **<u>0.397</u>** (cm^2)	A1 (3)
(c)	$\nu = 5 - 3 = 2$; so $\chi_2^2(10\%) = 4.605$	B1; B1ft
	(Not sig') so a normal distribution is a plausible model for length of pine cones	B1ft (3)
(d)	P($X > 7$ | $\mu = 5.8$ and $s = \sigma = 0.63035\ldots$) = $P\left(Z > \frac{7 - "5.8"}{\sqrt{0.397..}}\right)$ = P($Z > 1.90..$)	M1
	= **<u>0.028~0.029</u>**	A1 (2) **[18m'ks]**

Notes

(a)
1st B1 for both hypotheses. Must include the model and mention "length(s)" and "cones"
1st M1 for correct use of normal to find E_i for one cell
1st A1 for a middle value e.g. awrt 12.77~12.90 inclusive (12.77 is from tables, 12.90 calc)
2nd M1 for use of symmetry to get E_i for $5.5 \leqslant x < 6$ (same as $6 \leqslant x < 6.5$) or $x \geqslant 6.5$ (40 – ...)
2nd A1 for a correct set of expected frequencies (all awrt in given ranges)
3rd dM1 (dep on 1st M1) for a correct attempt to find test statistic…at least one correct term
3rd A1 for answer in the range 8.15-8.4 (inclusive)
2nd B1 for degrees of freedom = 4
3rd B1ft for a correct 10% critical value using their degrees of freedom
4th A1ft dep on M3 and cv = awrt 7.78 for contextual conclusion: length, cones, N (μ, σ not needed) or Chrystal's belief

(b)
B1 for 5.8
M1 for a correct expression (ft their mean)
A1 for awrt 0.397 (Condone $\frac{3139}{7900}$)

(c)
1st B1 for degrees of freedom = 2
2nd B1ft for a correct cv (different from their part (a)) ft their df
3rd B1ft for a correct conclusion in context ft cv ("length" and "cones") Ignore any μ or σ

(d)
M1 for standardising with 7, their 5.8 ($\neq 6$) and their s.d. from (b). Ignore any $\times 80$
A1 for a correct proportion of 0.028 or 0.029. (ISW if correct ans followed by $\times 80$)

Question Number	Scheme	Marks
6. (a)	Let $D = Y - R$ then $E(D) = -3$; $Var(D) = 0.8^2 + 1.5^2$ or 1.7^2 or 2.89	B1, M1
	$P(D > 0) = P\left(Z > \frac{0 - -3}{1.7}\right)$ <u>or</u> $P(Z > 1.7647\ldots)$	M1
	$= 0.03880655\ldots$ <u>or</u> $1 - 0.9608 = 0.0392$ awrt **<u>0.039</u>**	A1
		(4)
(b)	$(R_1 + R_2 + R_3) \sim N\left(45, \sqrt{3\times1.5^2}^2\right)$; $4Y \sim N\left(48, \sqrt{4^2\times0.8^2}^2\right)$	M1A1A1
	$L = 4Y - (R_1 + R_2 + R_3) \Rightarrow L \sim N\left(3, \sqrt{16.99}^2\right)$	M1A1
	$P(L > 0) = P\left(Z > \frac{0-3}{\sqrt{16.99}}\right)$ <u>or</u> $P(Z > 0 - 0.7278\ldots)$ [use 0 – 0.73 in tables]	dM1
	= awrt **<u>0.767</u>**	A1
		(7)
(c)	$E(X) = 780$ gives $15a + 12b = 780$	M1A1
	[Var(X) =] $1.5^2 \times a^2 + 0.8^2 \times b^2$	M1
	Sub for a: $Var(X) = 2.25(52 - 0.8b)^2 + 0.64 \times b^2$ <u>or</u> $2.08b^2 - 187.2b + 6084$	M1
	$\frac{d}{db}[Var(X)] = 0 \Rightarrow 4.16b - 187.2 = 0$	M1
	<u>$b = 45$</u>	A1
	So $a = 52 - 0.8\times45 = 52 - 36$ **<u>$a = 16$</u>**	A1
		(7)
		[18 marks]

Notes

(a)

B1 for $E(D) = -3$ (or +3 if using $R - Y$) and 1st M1 for $Var(D) = 0.8^2 + 1.5^2$ o.e.
2nd M1 for attempt at $P(D > 0)$ must standardise with their -3 and their 1.7 and inequality
A1 for awrt 0.039

(b)

1st M1 for correct mean or variance for either $R_1 + R_2 + R_3$ or $4Y$

1st A1 for $(R_1 + R_2 + R_3) \sim N\left(45, \sqrt{6.75}^2\right)$ 2nd A1 for $4Y \sim N\left(48, \sqrt{10.24}^2\right)$

2nd M1 for attempting a suitable L (condone $3R - 4L$ etc)
Must have L with mean of ± 3 and $\sigma_L{}^2$ = "6.75" + "10.24" = $(4.1218\ldots)^2$

3rd A1 for a correct mean and variance. **Sight of N($\pm$ 3, 16.99) scores 1st 5 marks**

3rd dM1 (dep on 2nd M1) for attempting a prob ($\rightarrow$ ans > 0.5) using $\mu_L = \pm 3$ and their σ_L

4th A1 for awrt 0.767 (Calc: 0.7666384… or tables 0.7673)

(c)

1st M1 for an attempt to use $E(X) = 780$ must see a linear equation in a and b using 780
1st A1 for $15a + 12b = 780$ o.e. e.g. $5a + 4b = 260$ or $a + 0.8b = 52$ etc
2nd M1 for an attempt to find an expression for $Var(X)$ (condone a and b wrong way around)
3rd M1 for forming a quadratic expression for $Var(X)$ in terms of a or b only (M0 for = k, $k \neq 0$)
4th M1 suitable method for finding min (e.g. differentiation, or completing square or calc)
e.g. $\frac{13}{4}(a^2 - 32a + 832)$ [3rd M1] then $k\left[(a-16)^2 + m\right]$ would score 4th M1

2nd A1 for $b = 45$ <u>or</u> $a = 16$
3rd A1 for **both** $b = 45$ <u>and</u> $a = 16$

Correct answers should be accompanied by evidence for 1st 4 marks

Pearson Education Limited. Registered company number 872828
with its registered office at 80 Strand, London, WC2R 0RL, United Kingdom