Mark Scheme (Results)

Autumn 2020

Pearson Edexcel GCE In A Level Statistics

(9ST0/01)

Paper 1: Data and Probability

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

Autumn 2020
Publications Code 9ST0_01_2010_MS

General Marking Guidance

Total marks

The total number of marks for the paper is 80.

Mark types

The Edexcel Statistics mark schemes use the following types of marks:

- **M** **Method** marks, awarded for 'knowing a method and attempting to apply it',
  unless otherwise indicated.
- **A** **Accuracy** marks can only be awarded if the relevant method (M) marks have been earned.
- **B** **Unconditional accuracy** marks are independent of M marks
- **E** **Explanation** marks

NOTE: Marks should not be subdivided.

Abbreviations

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

- ft follow through
- PI possibly implied
- cao correct answer only
- cso correct solution only
  (There must be no errors in this part of the question)
- awrt answers which round to
- awfw answers which fall within (a given range)
- SC special case
- nms no method shown
- oe or equivalent
- dep dependent (on a given mark or objective)
- dp decimal places
- sf significant figures
- ∗ The answer is printed on the paper

Further notes

- 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.
- All A marks are 'correct answer only' (cao.), unless shown, for example, as A1ft to indicate that previous wrong working is to be followed through.
- After a **misread**, the subsequent A marks affected are treated as A1ft, but manifestly absurd answers should never be awarded A marks.
- **Crossed out** work should be marked UNLESS the candidate has replaced it with an alternative response.
- If **two solutions** are given, each should be marked, and the resultant mark should be the mean of the two marks, rounded down to the nearest integer if needed.

Question	Scheme	Marks	AO	Notes
1(a)(i)	**Possible advantages** **(Not exhaustive)**			
	• Level of response will be very high.			
	• Very quick turnaround.			
	• Easy to administer.			
		E1	3.1a	oe Any of above
1(a)(ii)	**Possible sources of bias** **(Not exhaustive)**			
	• People who are ill often are more likely to be sampled.			
	• People who work in the morning are less likely to respond.			**or** on that specific day
	• People who are ill may not give full responses (or may answer badly/untruthfully).			
	• Some patients in the waiting room may be from the same household.			
	• Some patients may be too young to complete a questionnaire.			or certain disabilities
	• Patients may discuss the questionnaire in the waiting room, potentially leading to peer-pressure responses.			oe Do not accept: 'They may discuss answers' alone. Needs to be clear why this may lead to bias
		E1, E1	2.1b, 2.1b	E1 for each relevant source of bias (max E2)

Question	Scheme	Marks	AO	Notes
1(b)(i)	**Possible advantages** **(Not exhaustive)**			
	• The sample is random which will reduce bias.			Accept: Everyone is equally likely to be selected.
	• The responses will be electronically stored, making analysis easier.			
	• Patients are all equally likely to be contacted.			
	• Using a sample without replacement will ensure no duplicates.			
	• The survey will not be rushed, so patients can take their time to answer the questions carefully.			
		E1	3.1a	oe

Question	Scheme	Marks	AO	Notes
1(b)(ii)	**Possible sources of bias** **(Not exhaustive)**			
	• The results will be biased towards the type of people who regularly fill out surveys.			**or** towards those who have strong opinions.
	• Only people who regularly use the internet are likely to respond.			Accept: Non-response bias oe
	• Certain email providers may consider the email spam/junk.			
	• If a patient has changed his/her email address, then they will not be able to complete the survey.			
	• If a patient's email address has been entered into the system incorrectly, then they will not be able to complete the survey.			
	• Some patients may not have an email address.			
	• Some patients may have given a fake email address			
		E1, E1	2.1b, 2.1b	E1 for each relevant source of bias (max E2)
	Total	**6**		

Question	Scheme	Marks	AO	Notes
2(a)	**'Unsuitable' solution**			
	The Poisson distribution is an **unsuitable** model…	E1dep	2.1b	Dep on attempt at explanation
	…as the buses are timetabled…	E1	2.1a	**or** if traffic is bad, it is likely to affect multiple buses (oe) Must be in context
	…so the events are not independent.	E1	2.1b	**Accept:** The rate is not constant **Condone:** The events are not random
	'Dependent' solution			
	The suitability is **dependent**…	(E1dep)		Dep on next E1
	…on whether the buses are **timetabled**. **or** on whether there is a (timed) traffic light junction nearby.	(E1)		Implication (or direct statement) of considerations of constant rate and/or independence. Condone consideration of randomness No context needed Note: This mark may be scored in (b) if not here
		(E1)		Fully correct interpretation in context

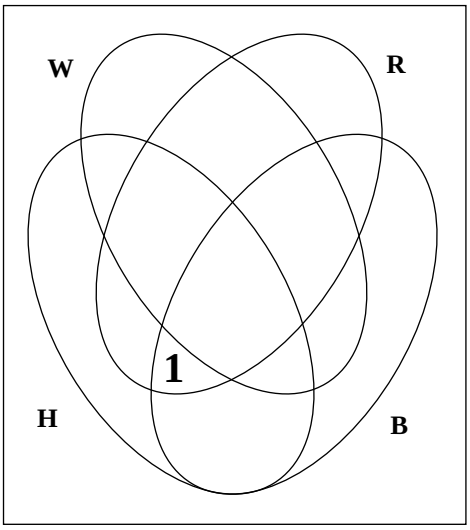
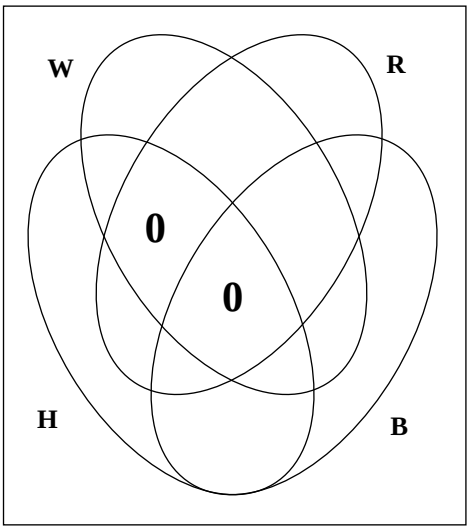
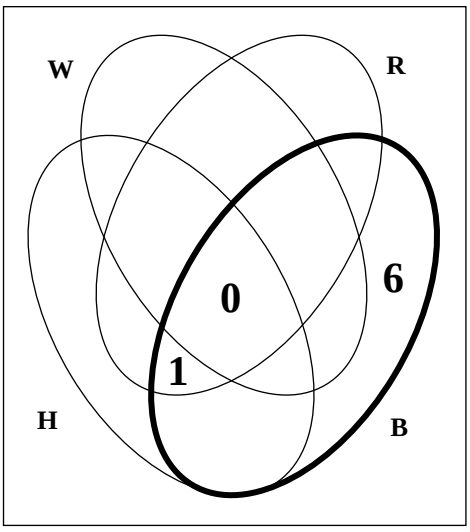
Question	Scheme	Marks	AO	Notes
2(b)	**'Suitable' solution**			
	The Poisson distribution is a **suitable** model.	E1dep	2.1b	Dep on attempt at finding parameter
	$\mu_Y = 7.73\times10^{-5}\times5.1\times10^{8}\times10$	M1	1.2	Attempt to find mean **Condone** single year (39 423)
	$\lambda = \mu_Y = 394\,000$ (3 s.f.)	A1	1.2	awrt Actual: 394 230
	'Unsuitable' solution			
	Meteorites often fall in groups as showers…	(E1)		Must be in context
	…so the rate is not constant…	(M1)		**Accept:** The events are not independent
	…so the Poisson distribution is an **unsuitable** model.	(A1dep)		Dep on either E1 or M1 scored
	'Dependent' solution			
	The suitability is **dependent** on whether meteorites fall: • at a constant rate • independently	(E1)		Scores E1 max Must be in context May score second E1 in (a) if not awarded there
	Total	**6**		

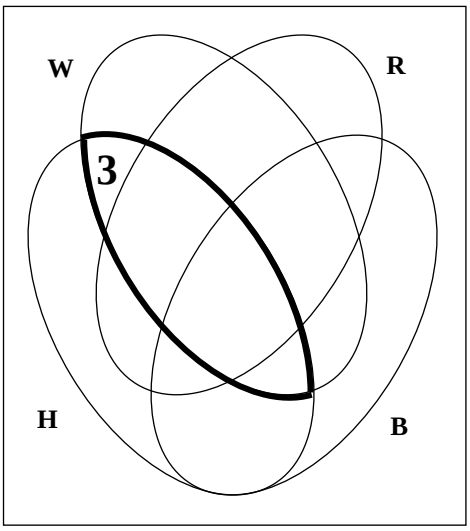
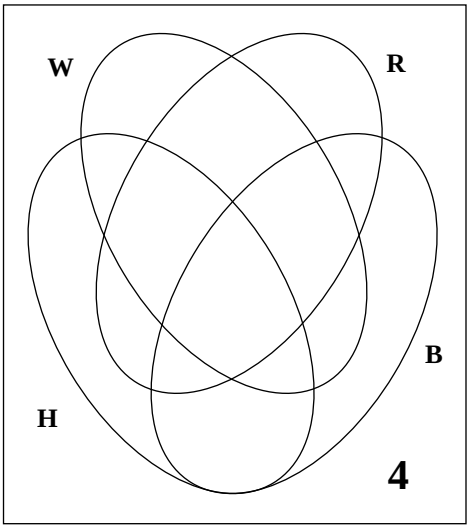
Question	Scheme	Marks	AO	Notes
3(a)(i)	**Possible advantages (not exhaustive)**			
	Probabilities are easily found (using a calculator/spreadsheet).			
	The distribution of the total score will also be known.			**or** total score also normally distributed
		E1	3.1a	
3(a)(ii)	**Possible disadvantages (not exhaustive)**			
	The model may predict a score above the maximum mark.			**or** below 0
	A normal distribution is a continuous distribution (whereas this data is discrete).	E1	3.1a	SC: E1 E0 for comments solely on shape (bell-shaped/ symmetrical etc)
3(b)	[Let C = Coursework score]			
	$C \sim N\left(76.75, 4.76^2\right)$			
	$\mathrm{P}\left(C < c\right) = 0.1$	M1	1.2	PI Clear attempt at cumulative dist = 0.1 (accept correct sketch) **or** use of $z = -1.2816$ or $z = -1.28$ **Note:** This mark may be scored in (c) if it is not scored here.

Question	Scheme	Marks	AO	Notes
3(b) (cont)	c = 70.6498…	M1	1.2	PI Correct use of inverse normal in calc. **or** use of tables and correct standardisation: $\frac{c-76.75}{4.76}=\pm 1.2816$ **Note:** This mark may be scored in (c) if it is not scored here.
	71 marks	A1	1.2	cao
3(c)	[Let X = Examination score]			
	$X \sim N\left(27.39, 8.19^2\right)$			
	$\mathrm{P}\left(X < x\right)=0.9$			May be awarded first M1 in (b) if not scored above
	c = 37.8859…			May be awarded second M1 in (b) if not scored above
	38 marks	B1	1.2	cao SC may be awarded if neither (b) or (c) is rounded

Question	Scheme	Marks	AO	Notes
3(d)	Student A scores 71 + 38			
	= 109 marks	B1ft	1.2	ft their (c) and (d)
	Student A scores **more** than student B (in total).	E1ft	2.1a	or 109 > 100 Clear comparison of **total** scores. PI
	However, Student A scored **less** than Student B in the coursework	E1ft	2.1a	oe Accept correct comparison of examination scores: 71 < 83 **or** ft their (c) < 83
	So Sylvia's weighting has not been successful (as it has advantaged those who did better in the examination rather than the coursework).	E1	3.1a	oe
3(e)	**Possible suggestions**			
	Make the coursework marks more spread out…	E1	3.1a	oe Accept greater SD/Var
	…by changing the coursework rubric/mark scheme.	E1	3.1a	Practical solution, must be in context
	Alternative			
	Increase the marks available in the coursework. **or** Reduce the marks available in the examination.	(E1)		Either scores E1 max Condone: use coursework scores only
	Total	**12**		

Question	Scheme	Marks	AO	Notes
4	I would use a **filter**…			Must see ‘filter’.
	…on ‘IncidentGroup’ to find ‘Fire’ incidents…			
	…and on ‘DateOfCall’ between 1/1/18 and 31/12/18.			
		E1, E1	1.1, 1.1	E1 for any two of these. E2 for all three.
	I would **sort** by ‘FirstPumpArriving_AttendanceTime” (in ascending order).	E1	1.1	Must see ‘sort’ **Condone** truncated field name.
	Then I would select the top/bottom 10 records (to find the fastest times)	E1	1.1	
	Total	**4**		

Question	Scheme	Marks	AO	Notes
5(a)		B1	1.1	cao
		B1	1.1	cao
		B1	1.1	If numbers incorrect, these three numbers adding to 7 scores B1

Question	Scheme	Marks	AO	Notes
5(a) (cont)		B1	1.1	If number incorrect, numbers adding to 3 inside region scores B1
		B1	1.1	If number incorrect, all numbers adding to 24 scores B1
5(b)(i)	$\mathrm{P}\left(H \cap W' \cap R \cap B\right)$			
	$= \frac{1}{24} = 0.0417$ (3 s.f.)	B1	1.1	awrt
5(b)(ii)	$\mathrm{P}\left(H' \cap W' \cap R' \cap B'\right)$			
	$= \frac{4}{24} = \frac{1}{6} = 0.167$ (3 dp)	B1ft	1.1	awrt 0.167 ft their diagram/24
5(b)(iii)	$\mathrm{P}\left(W \cap B\right)$			
	= 0	B1ft	1.1	ft their diagram/24

Question	Scheme	Marks	AO	Notes
5(b)(iv)	$\mathrm{P}(R \mid H) = \dfrac{\mathrm{P}(R \cap H)}{\mathrm{P}(H)}$	M1	1.2	Evidence of correct use of multiplication rule
	$= \dfrac{2}{8} = \dfrac{1}{4} = 0.25$	A1ft	1.2	oe cao
	Alternative			
	2 out of 8 people with hats have red hair…	(M1ft)		ft their diagram
	…so $\mathrm{P}(R \mid H) = \dfrac{2}{8} = \dfrac{1}{4} = 0.25$	(A1)		oe cao
5(c)	A **man**			**or** ‘not a woman’
	with **no hat**,			
	no beard,			
	and **without red hair**.			**Accept** ‘hair a colour other than red’ oe
		E1	2.1a	All four needed
5(d)	Does the person in the picture have a beard?	E1	2.1b	

Question	Scheme	Marks	AO	Notes
5(e)	**Likely questions**			
	Does the person in the picture:			
	• have either red hair or a beard?			
	Is the person in the picture:			
	• either a woman, or a man/person with red hair? • either a woman, or a man/person with a hat?			
		E1	2.1b	Any valid compound question.
		E1ft	2.1b	One which divides the 24 cards 10/14 at worst.
	Total	**14**		

Question	Scheme	Marks	AO	Notes
6(a)(i)	[X = volume of concrete in a single lorry (m^3)]			
	$P(X > 5.75) = 0.252$ (3 s.f.)	B1	1.2	awrt
6(a)(ii)	$X_1 + X_2 \sim N(11.34, 0.0288)$	B1	2.1b	Normal distribution stated or used **Note:** Can be scored in (b) if not here
		M1	1.2	PI Correct mean **or** variance (awrt) **or** SD = 0.170 (3 sf.)
	$P(X_1 + X_2 < 11) = 0.0226$ (3 s.f.)	A1	1.2	awrt
6(b)	$X_1 + ... + X_k \sim N(5.67k, 0.0144k)$			Normal distribution
		B1	1.2	Correct mean oe
		B1	1.2	Correct variance oe Condone SD $= \sqrt{0.0144k} = 0.12\sqrt{k}$

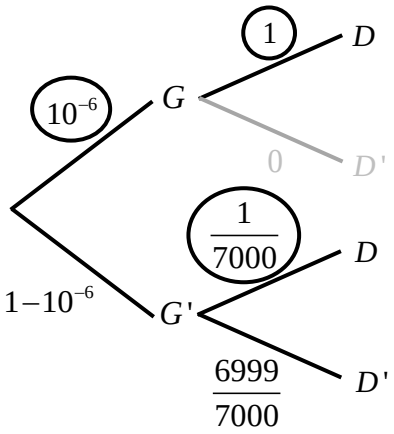
Question	Scheme	Marks	AO	Notes
6(c)	Box 1: 0.99			Top box
	Box 2: $[\mu =] 5.67k$			Middle box ft (b)
	Box 3: $[\sigma =] \sqrt{0.0144k} = 0.12\sqrt{k}$			Bottom-left box ft (b) Do not accept variance
	Box 4: -2.33 (3 s.f.)			Bottom-right box awrt Actual: -2.326347… Negative sign needed
		B1, B1, B1	1.2, 1.2, 1.2	B1 for one box correct B2 for two boxes correct B3 for all four boxes correct
6(d)	$k = 4\,797\,285.967$	M1	1.2	PI Clear attempt at squaring awrt 4 797 286
	She should plan to purchase (at least) 4 800 000 lorry loads.	A1	2.1a	Sensible rounding Condone: 4 797 000
	Total	**11**		

<table>
<tr><th>Question</th><th>Scheme</th><th>Marks</th><th>AO</th><th>Notes</th></tr>
<tr><td rowspan="2">7(a)</td><td>$\mathrm{P}(Swindon) = \frac{1}{1000} \times \frac{182\,441}{24.6^2}$</td><td>M1</td><td>1.2</td><td>PI
Correct use of formula for any town/city
or
At least one probability correct awrt:
Swindon 0.30
Oxford 0.23</td></tr>
<tr><td><table><tr><th>T</th><th>P(T)</th></tr><tr><td>Faringdon</td><td>0.2738</td></tr><tr><td>Wantage</td><td>0.0426</td></tr><tr><td>Witney</td><td>0.0899</td></tr><tr><td>Swindon</td><td>0.3015</td></tr><tr><td>Oxford</td><td>0.2274</td></tr></table></td><td>A1</td><td>1.2</td><td>Both probabilities correct, each to 4 d.p.</td></tr>
<tr><td rowspan="2">7(b)</td><td>$0.2738 + 0.0426 + 0.0899$</td><td></td><td></td><td></td></tr>
<tr><td>= 0.406 (3 s.f.)</td><td>B1</td><td>1.2</td><td>awrt</td></tr>
<tr><td>7(c)</td><td>There are bound to be other towns/cities/villages where families may choose to shop.</td><td>E1</td><td>2.1a</td><td></td></tr>
<tr><td rowspan="3">7(d)</td><td>300×2.4
= 720 people in new village

720×1.9
= 1368 shopping trips per week</td><td>M1</td><td>1.2</td><td>PI
Attempt to calculate (or estimate) number of weekly shoppers from the new village
estimate:
awfw 570~1368</td></tr>
<tr><td>1368×0.2738</td><td>M1</td><td>1.2</td><td>PI
Multiplying by 0.2738</td></tr>
<tr><td>= 375 (3 s.f.) Faringdon shoppers per week</td><td>A1</td><td>1.2</td><td>awfw 155~375</td></tr>
</table>

Question	Scheme	Marks	AO	Notes
7(e)	[Using model,]			
	$\mathrm{P}(London) = 0.685$ (3 s.f.)	B1	1.2	awrt Must be seen for B1
	Possible reasons (not exhaustive)			
	The model predicts a total probability greater than 1.			
	The model predicts that households would shop in London more than anywhere else, which is not realistic as it is nearly 120 km away.			Condone: London is too far away to go shopping oe
	If London were included, then other large cities would need to be included too.			
	The model may be too complicated if more towns/cities are added.			
		E1, E1	3.1a, 3.1a	E1 for each sensible reason (max E2)
	Total	**10**		

Question	Scheme	Marks	AO	Notes
8(a)	**Bayes' theorem method**			
	$\mathrm{P}(D \mid G) = 1$	B1, B1	1.2, 1.2	PI B1 for one correct B2 for both
	$\mathrm{P}(D \mid G') = \frac{1}{7000} = 0.00014286...$			
	$\mathrm{P}(G \mid D)$	M1	2.1a	Clear intent to find this probability
	$= \frac{\mathrm{P}(G) \times \mathrm{P}(D \mid G)}{\mathrm{P}(G) \times \mathrm{P}(D \mid G) + \mathrm{P}(G') \times \mathrm{P}(D \mid G')}$ $= \frac{10^{-6} \times 1}{10^{-6} \times 1 + \left(1 - 10^{-6}\right) \times \frac{1}{7000}}$	M1	1.2	PI Correct use of Bayes' theorem using assumptions
	$= 0.00695 = 0.695\%$	A1	1.2	$= 6.95 \times 10^{-3}$ awrt 0.007
	$\neq 99.987\%$ so the expert witness's statement is incorrect	E1dep	3.1b	oe Either scores E1 Dep on previous A1

Question	Scheme	Marks	AO	Notes
8(a) (cont)	**Tree diagram method**			
		(B1)		Correct diagram structure with G, G′, D and D′ correctly placed. [Top D′ branch optional] Condone D/D′ as first branch.
		(B1)		Correct circled probabilities.
	$\mathrm{P}(G \cap D) = 10^{-6}$ $\mathrm{P}(G \cap D') = 0$ $\mathrm{P}(G' \cap D) = 0.000143$ (3 s.f.) $\mathrm{P}(G' \cap D') = 0.999856...$	(M1)		Clear attempt at multiplying probabilities along branches PI
	$\mathrm{P}(G \mid D)$	(M1)		Clear intent to find this probability
	$= \dfrac{10^{-6}}{10^{-6} + 0.000143}$			
	$= 0.00695 = 0.695\%$	(A1)		$= 6.95 \times 10^{-3}$ awrt 0.007
	$\neq 99.987\%$ so the expert witness's statement is incorrect	(E1dep)		oe Either scores E1 Dep on previous A1

Question	Scheme	Marks	AO	Notes
8(b)	DNA evidence is not sufficiently rigorous for a prosecution…	E1	3.1b	DNA evidence insufficient
	…without other evidence.	E1	3.1a	Other evidence is also needed
	Alternative			
	Expert witnesses should be vetted before being allowed to give evidence.	(E1)		Scores E1 max
	Total	**8**		

<table>
<tr><th>Question</th><th>Scheme</th><th>Marks</th><th>AO</th><th>Notes</th></tr>
<tr><td rowspan="13">9</td><td>Proportion</td><td></td><td></td><td></td></tr>
<tr><td>Total = $2\times78+39+22$</td><td>M1</td><td>1.2</td><td>PI
Attempt to find total score</td></tr>
<tr><td>= 217</td><td>A1</td><td>1.2</td><td>cao</td></tr>
<tr><td>Proportion scored
$=\frac{217}{2\times151}=0.719$</td><td>B1ft</td><td>1.2</td><td>ft their total/302
Condone total/151</td></tr>
<tr><td>$0.719>70\%$</td><td>B1ft</td><td>2.1b</td><td>Correct comparison with 70%/0.7
ft their proportion</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
<tr><td>Independence (conditional probabilities method)</td><td></td><td></td><td></td></tr>
<tr><td><table><tr><td></td><td>2 S</td><td>2 S’</td><td>Total</td></tr><tr><td>1 S</td><td>78</td><td>39</td><td>117</td></tr><tr><td>1 S’</td><td>22</td><td>12</td><td>34</td></tr><tr><td>Total</td><td>100</td><td>51</td><td>151</td></tr></table></td><td>M1</td><td>1.2</td><td>PI
Attempt to find row and column totals</td></tr>
<tr><td>[A = Score on first throw
B = Score on second throw]</td><td></td><td></td><td></td></tr>
<tr><td>$\mathrm{P}(A)=\frac{117}{151}=0.775$ (3 s.f.)</td><td>M1</td><td>1.2</td><td>Clear attempt at one of these probabilities</td></tr>
<tr><td>$\mathrm{P}(A\mid B)=\frac{78}{100}=0.780$ (3 s.f.)
$\mathrm{P}(A\mid B')=\frac{39}{51}=0.765$ (3 s.f.)</td><td>A1</td><td>1.2</td><td>Two correct probabilities
awrt</td></tr>
<tr><td>(These probabilities are very close, so) the first and second throws are approximately independent.</td><td>E1dep</td><td>2.1a</td><td>Dep on previous A1</td></tr>
<tr><td>9 (cont)</td><td>Independence (multiplying probabilities method)</td><td></td><td></td><td></td></tr>
</table>

Question	Scheme	Marks	AO	Notes
	<table><tr><td></td><td>2 S (B)</td><td>2 S' (B')</td><td>**Total**</td></tr><tr><td>1 S (A)</td><td>78</td><td>39</td><td>**117**</td></tr><tr><td>1 S' (A')</td><td>22</td><td>12</td><td>**34**</td></tr><tr><td>**Total**</td><td>**100**</td><td>**50**</td><td>**151**</td></tr></table>	(M1)		PI Attempt to find row and column totals
	[A = Score on first throw B = Score on second throw]			
	$\mathrm{P}(A)=\frac{117}{151}=0.775$ (3 s.f.) $\mathrm{P}(B)=\frac{100}{151}=0.662$ (3 s.f.) $\mathrm{P}(A\cap B)=\frac{78}{151}=0.517$ (3 s.f.)	(M1)		Clear attempt at one of these probabilities
	$\mathrm{P}(A)\times P(B)=\frac{11\,700}{22\,801}=0.513$ (3 s.f.) $\mathrm{P}(A\cap B)=\frac{78}{151}=0.517$ (3 s.f.)	(A1)		Both probabilities correct awrt
	(These probabilities are very close, so) the first and second throws are approximately independent.	(E1dep)		Dep on previous A1
	Conclusion			
	[The evidence suggests that] Aoife is free throwing at a professional level.	E1dep	2.1a	Dep on either: • 2nd B1 in proportion, or • E1 in independence.
	Total	**9**		