Pearson
Edexcel

Pearson Edexcel
Level 3 Advanced GCE in Mathematics (9MA0)

May/June 2023 Exemplar
9MA0-31 A level Mathematics
Paper 31: Statistics
Senior Examiner's feedback on student responses

Contents

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

About this booklet

This document has been produced to support mathematics teachers delivering the GCE Mathematics specification.

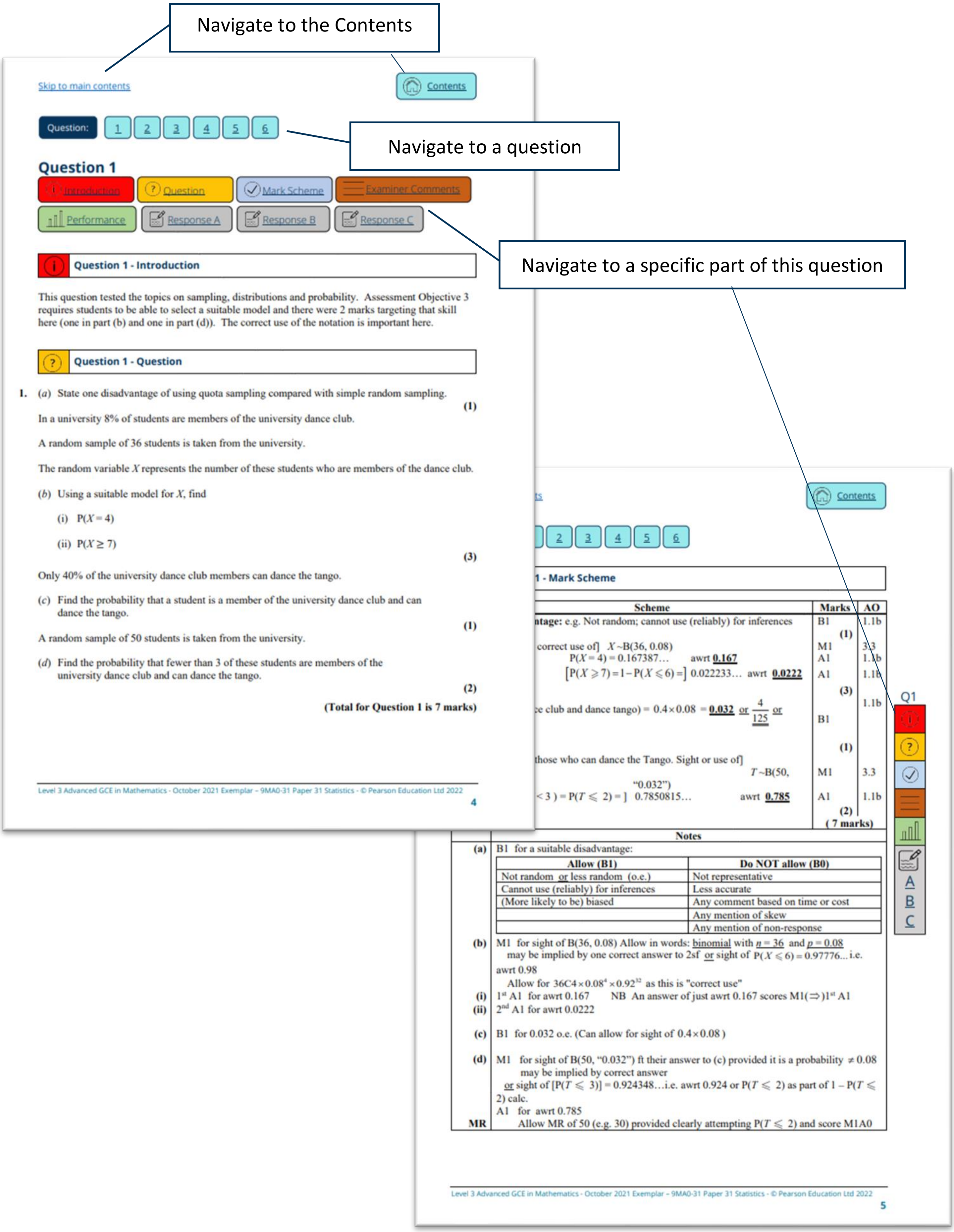
This document looks at questions from the 9MA0-31 A level Mathematics Paper 31: Statistics June 2023 examination paper. It shows real student responses to these questions, and how the examining team follow the mark schemes to demonstrate how the students are awarded the marks. For the mark scheme notes and details of alternative methods please see the full mark scheme for this question paper on our website.

For 2023, the approach all exam boards have taken to grading was to return to pre-pandemic grading, while giving students protection against any impact of disruption. Results in summer 2023 therefore will be far more in line with summer exams that were sat in 2019, but lower than in 2022, when grades awarded were based on a mid-point between 2019 and 2021 outcomes. For more information please read our 'Understanding grade boundaries 2023' document.

* The question level performance data is there to give an indication only of how students performed, on each question, in the context of sitting the entire exam paper and is not an indication of how students may perform sitting a question in isolation.

Skip to Main Contents

How to use this booklet



General Examiner Feedback

The first 4 questions on this paper were more accessible and most students made progress in each of these. In particular, question 1 parts (a) and (b), question 2 parts (b) and (d), question 3 part (b) and question 4(a) gave the grade E candidates a good source of marks. In addition, question 4(b) was now a familiar hypothesis test and those who had practised similar questions were able to make reasonable progress. The final 2 questions, and especially question 5, were more challenging and provided good differentiation for the more able students.

It was encouraging to see improved familiarity with the large data set in question 3.

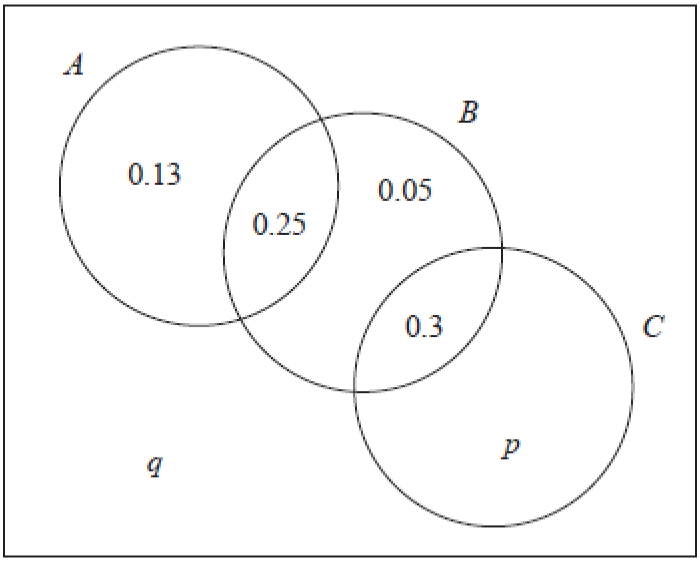
Question 1

Question | Mark Scheme | Examiner Comments

Performance | Response A | Response B | Response C

Question 1 - Question

1. The Venn diagram, where p and q are probabilities, shows the three events A, B and C and their associated probabilities.



(a) Find P(A)

(1)

The events B and C are independent.

(b) Find the value of p and the value of q

(3)

(c) Find P(A|B')

(2)

(Total for Question 1 is 6 marks)

Question 1 - Mark Scheme

Qu 1	Scheme	Marks	AO
(a)	[0.13 + 0.25 =] **0.38**	B1	1.1b
		(1)	
(b)	Independence implies:		
	e.g. $\left[\text{P}(B\cap C)=\text{P}(B)\times\text{P}(C) \quad \Rightarrow\right]$ $0.3=(0.3+0.05+0.25)\times(0.3+p)$	M1	1.1b
	So $p=\underline{\mathbf{0.2}}$	A1	1.1b
	[Sum of probabilities = 1 gives] $q=\underline{\mathbf{0.07}}$	B1ft	1.1b
		(3)	
(c)	$[\text{P}(A\mid B')=]\dfrac{\text{P}(A\cap B')}{\text{P}(B')}$ or $\dfrac{0.13}{(1-0.6) \text{ or } (0.13+\text{"0.2"}+\text{"0.07"})}$	M1	1.1b
	$=\dfrac{13}{40}$ or **0.325**	A1	1.1b
		(2)	
		(6 marks)	

	Notes
(a)	B1 for 0.38 (or exact equivalent) **If answers are given on Venn Diagram and in the script then the script takes precedence.**
(b)	M1 for a correct equation in p or P(C) only. May be implied by an answer of $p = 0.2$ **provided** this does not come from incorrect working. Condone missing brackets if they get 0.2 Other rules for independence will give simple rearrangements of this equation.
Beware	If $p = 0.2$ comes from incorrect working, we've seen $p=\frac{0.6}{0.3}=0.2$, score M0A0
	A1 for $p = 0.2$ (or exact equivalent) B1ft for $q = 0.07$ (or exact equivalent) ft their p i.e. $q = 0.27 - \text{"0.2"}$ where $0 \leqslant p \leqslant 0.27$
(c)	M1 for a correct ratio of probability expressions or a correct ratio of probabilities ft their values of p and q (provided both probabilities) or letters p and q A1 for 0.325 or exact equivalent. Correct answer only will score 2/2 **NB on epen this is labelled M1 but treat it as A1**

Skip to Main Contents

Question: 1 2 3 4 5 6

Question 1 - Examiner Comments

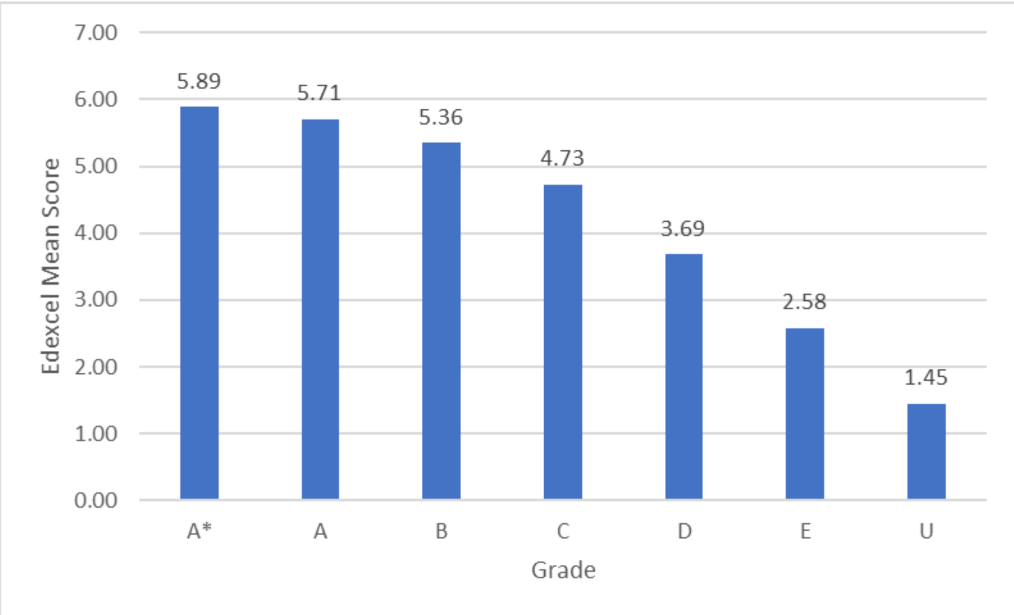
This proved to be a positive and accessible start to the paper.

Part (a) was nearly always correct and most knew how to write down a suitable expression for independence for part (b) which usually resulted in a correct equation for p. Those that didn't achieve this often thought that $\mathrm{P}(C) = 0.3$ or $0.3p$ or even just p. Solving for p was mostly completed correctly and even those who had an incorrect value for p could often score the follow through for q provided their $p + q = 0.27$

In part (c), some used the simple conditional probability formula $[\,\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(A \mid B)\,]$ in the formula booklet to get started whereas others just wrote down a correct ratio of probabilities from their Venn diagram. A fairly common error was to have $\mathrm{P}(B) = 0.6$ as the denominator and a few misinterpreted the formula and used $\frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B')}$. Those using the more complicated formula for $\mathrm{P}(A|B)$ in the formula booklet rarely managed to complete part (c) and students should be encouraged to know how to use the formulae in the formula booklet efficiently.

Question 1 - Performance

			Edexcel averages: mean scored by candidates achieving grade:						
Mean score	**Max score**	**Mean %**	**A***	**A**	**B**	**C**	**D**	**E**	**U**
4.87	6	81%	5.89	5.71	5.36	4.73	3.69	2.58	1.45



Q1
?
A
B
C

Question 1 - Response A

A) ~~0.13~~ $0.25 - 0.13 = 0.12$

$P(A) = 0.12$

B) $0.13 + 0.25 + 0.05 + 0.3 + p + q = 1$

$0.73 + p + q = 1$

$p + q = 0.27$

$0.3 - 0.27 = 0.03$

$p = 0.03$

$q = 0.24$

1/6 marks

Part (a)

B0 - This is incorrect, which was uncommon.

Part (b)

They do not use an independence property and there is no correct equation for *p* or P(*C*) (M0 and therefore A0). They have a value for *p* and they give a value for *q* which could be read as 0.14 or 0.24. Since they have previously stated that $p + q = 0.27$ it is assumed that their $q = 0.24$ and is therefore a correct follow through (B1)

Note that students should always ensure their writing is clear, and letters and number are clearly distinguishable. Whilst examiners will try and see if it could be correct, unclear handwriting can affect this.

Part (c)

There is no attempt (M0 and therefore A0)

Examiner comments

This student did not score the first mark, which was uncommon, and didn't understand or use the statement about events *B* and *C* being independent but they still scored the B1 mark in part (b).

Question 1 - Response B

(a) Find P(*A*)

(1)

The events *B* and *C* are independent.

$P(A \cap B) = P(A) \times P(B)$

(b) Find the value of *p* and the value of *q*

(3)

(c) Find $P(A|B')$

(2)

(a) $P(A) = 0.13 + 0.25 = 0.38$

(b) Independant so don't depend on each other.

$0.73 + p + q = 1$

$P(B) = \cancel{0.7}\ 0.6 \qquad P(B \cap C) = 0.3$

$P(B \cap C) = P(B) \times \cancel{P(B)}\ P(C)$

$0.3 = 0.6 \times P(C)$

(c) $P(A|B') = \dfrac{P(A \cap B')}{P(B')} =$

3/6 marks

Part (a)

This is correct (B1)

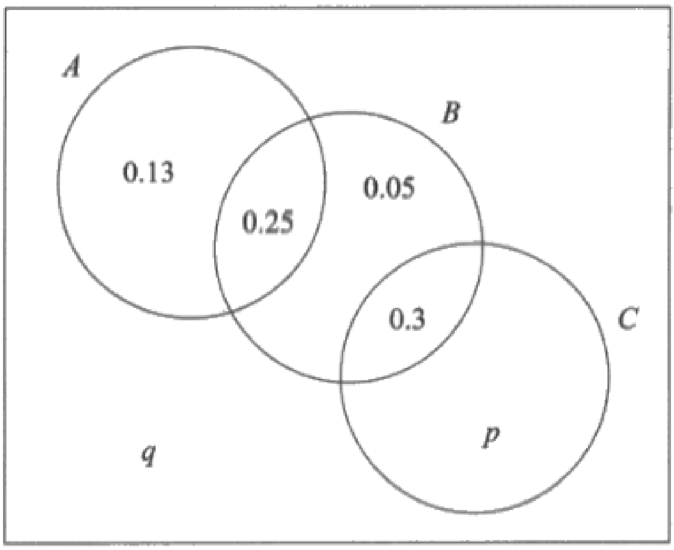
Part (b)

There is a correct equation for P(*C*) based on an independence property (M1) but it does not lead to $p = 0.2$ (A0). There is no attempt to find a value for q (B0)

Part (c)

There is a correct ratio of probability expressions (M1) but no attempt to use this to complete the question (A0)

Question 1 - Response C



(a) Find P(A) **(1)**

The events B and C are independent.

(b) Find the value of p and the value of q **(3)**

(c) Find P($A|B'$) **(2)**

a) 0.38

b) $0.3 \div (0.3+0.05+0.25) = 0.5$ $\quad P(B \cap C) = P(B) \times P(C)$

$= P(C)$ $\quad p = 0.2$

$q = 0.33$

c) $B' = 0.4$ $\quad P(A|B') = \frac{0.13}{0.4} = 0.325$

5/6 marks

Part (a)

Fully correct (B1)

Part (b)

There is a correct expression for P(C) based on independence which leads to the correct value for p and this is enough for both of the first two marks (M1A1) but their answer for q is incorrect (B0)

Part (c)

This is correct, we do not need to see the ratio of probability expressions if the ratio of probabilities is correctly given. (M1A1). Note that some "shorthand" notation such as B' = 0.4 rather than P(B') = 0.4 was condoned. Provided this didn't invalidate their solution the marks were awarded for correct expressions that lead to the correct answer.

Q1
A
B
C

Question 2

Question | Mark Scheme | Examiner Comments

Performance | Response A | Response B | Response C

Question 2 - Question

2. A machine fills packets with sweets and $\frac{1}{7}$ of the packets also contain a prize.

The packets of sweets are placed in boxes before being delivered to shops.
There are 40 packets of sweets in each box.

The random variable T represents the number of packets of sweets that contain a prize in each box.

(*a*) State a condition needed for T to be modelled by $\mathrm{B}(40, \frac{1}{7})$

(1)

A box is selected at random.

(*b*) Using $T \sim \mathrm{B}(40, \frac{1}{7})$ find

(i) the probability that the box has exactly 6 packets containing a prize,

(ii) the probability that the box has fewer than 3 packets containing a prize.

(2)

Kamil's sweet shop buys 5 boxes of these sweets.

(*c*) Find the probability that exactly 2 of these 5 boxes have fewer than 3 packets containing a prize.

(2)

Kamil claims that the proportion of packets containing a prize is less than $\frac{1}{7}$

A random sample of 110 packets is taken and 9 packets contain a prize.

(*d*) Use a suitable test to assess Kamil's claim.
You should

- state your hypotheses clearly
- use a 5% level of significance

(4)

(Total for Question 2 is 9 marks)

Question 2 - Mark Scheme

Qu 2	Scheme	Marks	AO
(a)	Comment in context about either **independence** or **random** packing e.g. "<u>prizes</u> must be placed in <u>packets</u> at <u>random/independently</u> of each other" <u>**or**</u> about **constant probability** e.g. "the <u>probability</u> of a <u>packet</u> containing a <u>prize</u> is <u>constant/ the same/fixed</u>"	B1 **(1)**	3.5b
(b)(i)	[P($T = 6$) =] 0.17273… awrt <u>**0.173**</u>	B1	1.1b
(ii)	[P($T < 3$) = P($T \leq 2$) =] 0.061587… awrt <u>**0.0616**</u>	B1 **(2)**	1.1b
(c)	[K= no. of boxes with fewer than 3 packets containing a prize] $K \sim$ B(5, "0.0616") P($K = 2$) = 0.031344… in the range **[<u>0.0313~0.0314</u>]**	M1 A1 **(2)**	1.1b 1.1b
(d)	$H_0 : p = \frac{1}{7}$ $H_1 : p < \frac{1}{7}$ [X = no of packets containing a prize] X~B(110, $\frac{1}{7}$) [P($X \leq 9$)] = 0.038292… [Significant result <u>or</u> reject H_0] E.g. there <u>is</u> evidence to <u>support</u> Kamil's <u>claim</u>	B1 M1 A1 A1 **(4)**	2.5 3.3 3.4 2.2b
		(9 marks)	

	Notes
(a)	B1 **May use idea of independent events:** a suitable reason, **in context**, covering idea of <u>random</u> packing or packets filled <u>independently.</u> Should mention key words/ideas of: <u>prizes</u> in <u>packets</u> **or** <u>packets</u> in <u>boxes</u> **May use idea of constant probability**. Must see key words underlined in scheme. Idea of probability with "independence" or "not affected by other packets" is B0 B0 for: **Idea of only 2 cases**. E.g. <u>Packet</u> contains a <u>prize</u> or not <u>or</u> **Idea of a fixed number of trials**. E.g. Need a <u>fixed</u> number of <u>packets</u> in each <u>box</u>
(b)(i)	B1 for awrt 0.173
(ii)	B1 for awrt 0.0616
(c)	M1 for sight of B(5, "0.0616") <u>or</u> ${}^5C_2\,(\text{"0.0616"})^2\,(1-\text{"0.0616"})^3$ ft their answer to (b)(ii). A1 for an answer in the range [0.0313 to 0.0314] Use of 0.0616 gives 0.031356..ans only 2/2
(d)	B1 for both hypotheses correct in terms of p or π M1 for selecting an appropriate model, may be implied by 1st A1 or P($X = 9$) = 0.0199(2…) 1st A1 for 0.038 or better <u>or</u> allow 0.04 with sight of P($X \leq 9$)
ALT	**Critical Region.** Allow CR of $X \leq 9$ (or $X < 10$) provided a supporting probability is seen e.g. A1 for correct CR plus P($X \leq 10$) = 0.0718… (accept 2sf or 1sf if prob statement seen) 2nd A1 (dep on 1st A1 but indep of hyp's) for a suitable conclusion in context that suggests <u>support</u> for (Kamil's) <u>claim</u> **or** states that there is evidence that <u>proportion /probability/chance</u> of packets containing a <u>prize</u> is less than $\frac{1}{7}$ Do not award 2nd A1 for contradictory statements e.g. "not significant" so "supports claim"
Normal	Sight of $N\left(\frac{110}{7}, \frac{660}{49}\right.$ <u>or</u> awrt 13.5$\left.\right)$ <u>or</u> probability of 0.045(20..) <u>or</u> 0.033(66..) scores M1

Question 2 - Examiner Comments

This question was also answered well.

In part (a) students were often aware of the conditions needed for a binomial distribution but did not always present their answers in context with the precision required. Some focussed on the need for a fixed number of trials or that packets either did or did not contain a prize but these features were clearly identified in the question. Those scoring the mark either identified the need for prizes to be placed in packets at random or that the probability of a packet containing a prize needed to be constant. There was some confusion between "events", such as a prize being placed in a packet, being independent and the "probability" of such an event being constant. We sometimes saw a statement such as "the probability of a packet containing a prize is independent" which didn't score this first mark.

Part (b) was generally answered well though some gave answers to 2dp or 2sf rather than the 3sf as instructed on the front of the paper. Most answered (i) correctly but in (ii) common incorrect answers were from those who found $P(T = 3)$ or $P(T \leqslant 3)$. In part (c) many were able to identify the correct distribution and those who stated this were guaranteed the method mark even if their answer to (b)(ii) was incorrect. Whilst many did obtain the correct answer here those with an incorrect answer to (b)(ii) who simply wrote down a probability could not be given either mark. Some students in part (c) didn't identify the binomial model but wrote an expression for the probability of the form ${}^5C_2\, p^2(1-p)^3$ but a few forgot the ${}^5C_2$ term.

The hypothesis test in part (d) was answered well with correct notation used for the hypotheses and the appropriate one-tailed test identified. Most selected the correct binomial model but it was disappointing to see a number finding $P(X = 9)$ rather than $P(X \leq 9)$. A correct conclusion in context was often given: some using the simple response that there is evidence to support Kamil's claim and others stating that there was evidence that the proportion of packets containing a prize is less than $\frac{1}{7}$. It was encouraging to see that most of those using the second approach were referring to "proportion" rather than "number of" this time.

Q2

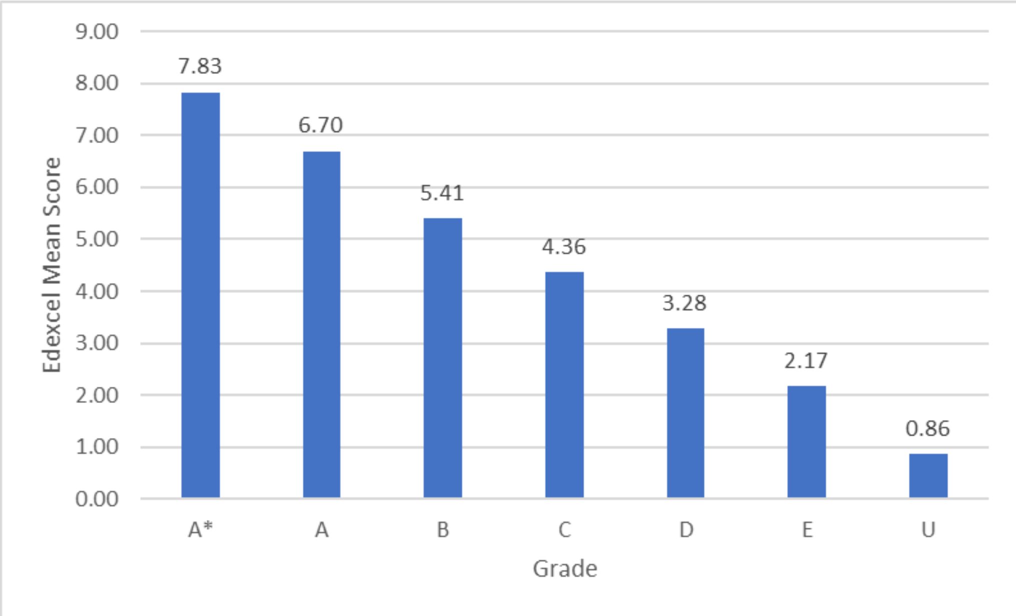
A

B

C

Question 2 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade: A*	A	B	C	D	E	U
5.27	9	59%	7.83	6.70	5.41	4.36	3.28	2.17	0.86



Q2
?
A
B
C

Question 2 - Response A

P of packet containing prize : $\frac{1}{7}$

a) The ~~sex~~ process of choosing a packet to fill with a prize is random.

b) i) $X \sim B(40, \frac{1}{7})$

~~$6/7$~~

$P(x=6)$

$P(x=6) = \cancel{0.1723}$ $1-0.1723$

$\cancel{= 0.827}$ $\underline{0.83} = P$

ii) $P(x<3)$

$= P(x \leq 2)$

$= \underline{0.0616}$

Q2

?

A
B
C

c) ~~P(~~

0.616 prob of box having < 3 prizes

$I \sim B(5, 0.616)$

$P(I=2)$

$= \underline{0.215}$

d) $H_0 : p = \frac{1}{7}$ $H_1 : p < \frac{1}{7}$

$n = 110$ 9 contain prize

5% sig level

Crit value:

~~$\underline{< 0.02}$~~

~~0.02~~

~~$P(x=9) = 0.0199$~~

~~$= \underline{0.98}$~~

$P(x=9) : = 0.02$

~~$\underline{< 0.02}$~~

Critical value
At 5%
$= \underline{0.051}$

$0.02 < 0.051$, there is enough evidence to reject the null hypothesis.

Q2
?
A
B
C

4/9 marks

Part (a)
They have a correct statement about the need for random packing and they have put this in context using the words "packet" and "prize" (B1)

Part (b)

In (i) they have found 1 – P(X = 6) (B0) but their answer to (ii) is correct (B1)

Part (c)

They have miscopied their 0.0616 from (b)(ii) as 0.616 and used this figure to define a binomial distribution with n = 5. This would be regarded as a slip that does not affect the method so (M1) but of course their answer is incorrect (A0).

Part (d)

The hypotheses are given in terms of p and are correct (B1) but the B(110, $\frac{1}{7}$) model is not clearly stated, just stating n = 110 and the value of p is not sufficient. However, they have attempted to find P(X = 9) but give the answer as 0.02, this is not P(X = 9) = 0.0199… for which the MS would allow the method mark (M0). They are not attempting P($X \leq 9$) and do not have the correct probability (1st A0 and therefore 2nd A0) because the 2nd A mark depends upon the 1st A mark. Once M0 has been scored the following A marks cannot be scored.

Examiner comments

In an exam the conditions for a binomial distribution are expected to be given in the context of the question. In part (a) students were allowed to use the independence of the events (a prize being placed in a packet, and this being carried out at random) or the constant nature of the probability (the probability of a packet containing a prize equalling $\frac{1}{7}$). The success or failure idea (a packet either contains a prize or not) was not allowed as this feature was clearly given in the context of the question. Nor was the idea of a fixed number of trials allowed since the question clearly stated that there were 40 packets in each box. Mixing "independence" with "probability" was not condoned for the first mark.

Q2

A
B
C

Question 2 - Response B

a) The events need to be independent.

b) $T \sim B(40, \frac{1}{7})$

i) $P(T=6) = 0.1727$ ~~(4dp)~~ (4s.f)

ii) $P(T<3) = P(T \leq 2) = 0.06159$ (4s.f)

c) 5 boxes = 5×40
$= 200$ sweets

$\frac{1}{7} \times 40 = \frac{40}{7}$ average no. of prizes per box

d) $H_0 : p = \frac{1}{7}$ $\quad H_1 : p < \frac{1}{7}$

Assume H_0 to be true $\quad$ 5% sig level one tailed $= 0.05$

$X \sim B(110, \frac{1}{7})$

$110 \times \frac{1}{7} = 15.7$

~~∞~~ $P(X \leq 9) = 0.03829 < 0.05$

$P(X \leq 10) = 0.07183$

~~[illegible]~~

$P(X \leq 11) = 0.1226$

Since $P(X \leq 9) = 0.03829 < 0.05$

Reject H_0.

~~[illegible]~~ Therefore the proportion of packets containing a prize is less than $\frac{1}{7}$, and there is sufficient evidence to reject H_0.

6/9 marks

Part (a)

They have realised that the "events" need to be independent but there is no attempt to use context (B0). A comment such as "(The event) a packet contains a prize must be independent of other packets." would have been acceptable.

Q2 ? A B C

Part (b)

This is fully correct, both probabilities round to the answers in the mark scheme (i.e. 3sf as required on the front of the examination paper). (B1B1)

Part (c)

They do not select a suitable binomial model. (M0 and therefore A0)

Part (d)

The hypotheses are given in terms of *p* and are correct (B1) and the B(110, $\frac{1}{7}$) model is clearly stated (M1). The correct probability of 0.038… is seen (1st A1) and they have a correct conclusion in context mentioning "proportion" and "prize". Their other statements such as "reject H_0" are not necessary but they are correct and so there are no contradictory statements (2nd A1).

Question 2 - Response C

(a) the events are independant of previous outcomes

(b)(i) $P(T=6) = 0.1727 \approx 0.173$

(ii) $P(T<3) = P(T \leq 2) = 0.0616$

Q2

?

A
B
C

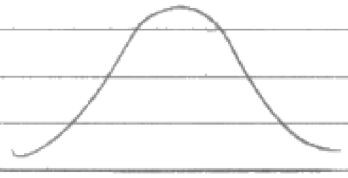
(c) $X \sim N(5, 0.0616)$

$P(x=2) = \cancel{0.2684}\ \cancel{0.284}\ 0.0314$

(d) $H_0 : p = 1/7$ $H_1 : p < 1/7$

$n = 110$ 5%

assuming H_0 is true



$\cancel{P(X \leq 9)}$ $P(x \leq 9) = 3.8\%$

critical region is $x \geq 10$

9 is beyond the critical region and therefore we have sufficient evidence to reject H_0

as $3.8\% < 5\%$

7/9 marks

Part (a)

They have realised that the “events” need to be independent but there is no attempt to use context (B0).

Part (b)

This is fully correct (B1B1)

Part (c)

They stated N(5, 0.0616) as their model but give the correct probability so the marks are awarded and the incorrect notation is condoned (M1A1)

Part (d)

The hypotheses are correct (B1) and, although the correct model is not stated, they have $P(X \leq 9) = 3.8\%$ (which is equivalent to 0.038) (M1 and 1st A1). They go on to state that the critical region is ≥ 10 but still reject H_0. Their conclusion is not given in context (2nd A0). The accuracy mark would not be awarded even if they had a correct statement in context because of the contradictory statement about the critical region.

Q2

?

A

B

C

Question 3

Question 3 - Question

3. Ben is studying the Daily Total Rainfall, x mm, in Leeming for 1987

He used all the data from the large data set and summarised the information in the following table.

x	0	0.1–0.5	0.6–1.0	1.1–1.9	2.0–4.0	4.1–6.9	7.0–12.0	12.1–20.9	21.0–32.0	tr
Frequency	55	18	18	21	17	9	9	6	2	29

(*a*) Explain how the data will need to be cleaned before Ben can start to calculate statistics such as the mean and standard deviation.

(2)

Using all 184 of these values, Ben estimates $\sum x = 390$ and $\sum x^2 = 4336$

(*b*) Calculate estimates for

(i) the mean Daily Total Rainfall,

(ii) the standard deviation of the Daily Total Rainfall.

(3)

Ben suggests using the statistic calculated in part (b)(i) to estimate the annual mean Daily Total Rainfall in Leeming for 1987

(*c*) Using your knowledge of the large data set,

(i) give a reason why these data would not be suitable,

(ii) state, giving a reason, how you would expect the estimate in part (b)(i) to differ from the actual annual mean Daily Total Rainfall in Leeming for 1987

(2)

(Total for Question 3 is 7 marks)

Question 3 - Mark Scheme

Qu 3	Scheme	Marks	AO
(a)	Need to replace tr with a numerical value	M1	1.2
	Value of tr is between 0 and 0.05 suggest using e.g 0.025 , 0 <u>or</u> value „ 0.05	A1	1.1b
		(2)	
(b)(i)	$\left[\bar{x} = \frac{389.3 \sim 390.8}{184}\right] = 2.119\ldots$ awrt **<u>2.12</u>** allow $\frac{195}{92}$ or $2\frac{11}{92}$	B1	1.1b
(ii)	$[\sigma =]\sqrt{\frac{(\text{awrt})4336}{184} - "\bar{x}^2"}$ <u>or</u> allow $\left[\sigma^2 =\right]\frac{(\text{awrt})4336}{184} - "\bar{x}^2"$ <u>or</u> awrt 19.1	M1	1.1b
	$= 4.367\ldots$ awrt **<u>4.37</u>**	A1	1.1b
		(3)	
(c)(i)	Only covers May~Oct (so not a suitable sample)	B1	1.1b
(ii)	e.g. Winter months are <u>missing</u> when we'd expect <u>more rain</u> so expect estimate in (b)(i) to be an <u>underestimate</u> (oe)	B1	2.4
		(2)	
		(7 marks)	

Notes

(a) M1 for recognising that tr must be replaced (oe) with a numerical value
The following examples would score M0: The tr values are worth 0 so ignore (not replacing) <u>or</u> must remove outliers <u>or</u> fill gaps in table <u>or</u> make widths the same <u>or</u> need to find mid-points
A1 for using a suitable value: e.g. 0.025 (or allow 0) i.e. any value in [0, 0.05]
(these give $\sum x = 390$ (3sf), use of 0.05 gives 390.8, use of 0 gives 389.3 allow in (b)(i))

(b)(i) B1 for awrt 2.12 <u>or</u> allow simplified fraction or mixed number. B0 for $\frac{390}{184}$

(ii) M1 for a correct expression for standard deviation **or** variance. Allow $\sum x^2 =$ awrt 4336
Ignore their label σ or σ^2 Can ft their mean

A1 for awrt 4.37 [Use of s gives 4.3791… so for correct use seen allow awrt 4.38]

SC **Using $n = 155$** Allow M1 for expression $[\sigma =]\sqrt{\frac{(\text{awrt})4336}{155} - "\bar{x}^2"} = \sqrt{21.64\ldots}$ or 4.65…

Part (c) can effectively be marked together.

(c)(i) B1 for a comment mentioning that data is just from May~Oct (so not representative of the whole year).
Just saying "only 184 days so not representative" is B0, **must mention May ~ Oct**

(ii) B1 for comment that <u>missing</u>/<u>winter</u> months expected to have more rain (oe) **and** "underestimate"(oe)

We are looking for all 3 of these ideas here:

1. A statement or implication that missing data is from winter or different months.
2. A suggestion about the rainfall in these months (probably more rain).
3. A statement about the impact on the estimate in (b)(i) <u>equivalent</u> to saying it would be an underestimate or the (actual) mean will be higher.

SC If you see "Leeming or N or NE has <u>less</u> rain in winter months" – please send to review

Q3

?

A

B

C

Question 3 - Examiner Comments

Once again the large data set aspects of this question caused problems for some students but most scored some marks here.

In part (a) a few responses simply quoted standard textbook answers about cleaning data by removing outliers and others were distracted by the class intervals for the data. Many though did realise that the "tr" values needed attention but a number didn't seem to realise that trace entries had a value between 0 and 0.05 and therefore removing them would have a big effect on the calculations Ben wanted to carry out. Those that did appreciate the value of the "tr" usually suggested a suitable numerical value (usually 0) to replace them with and scored both marks.

Part (b) was, unsurprisingly, answered very well with many scoring full marks. A few lost marks for truncating rather than rounding their answers for example 2.11 in (i) whilst others were convinced that the "tr" values should have been removed and had a denominator of 155 despite the question telling them that all 184 values were used. It is still surprising to see a small number of students using a denominator of 9 or 10 (the number of classes in the table) and a slightly larger proportion who couldn't use the standard deviation formula, given in the formula booklet, correctly.

It was encouraging to see a good number of students demonstrating some familiarity with the large data set and scoring some marks in part (c). In (i) we were simply looking for an awareness that because the large data set only had data for May to October this would not be representative of the whole population. Some demonstrated a lack of understanding of the principle of using a sample to make inferences by stating that 184 days was not a sufficiently large sample to represent the weather for the whole year. In this case the problem was not the sample size but the fact that the sample was not representative. In (ii) even some of those who had failed to mention May to October in (i) were able to score here by stating that because the missing months were in winter and they expected more rainfall then, the actual mean would be higher. The great storm of October 1987 featured in many incorrect answers to part (c).

Q3

A

B

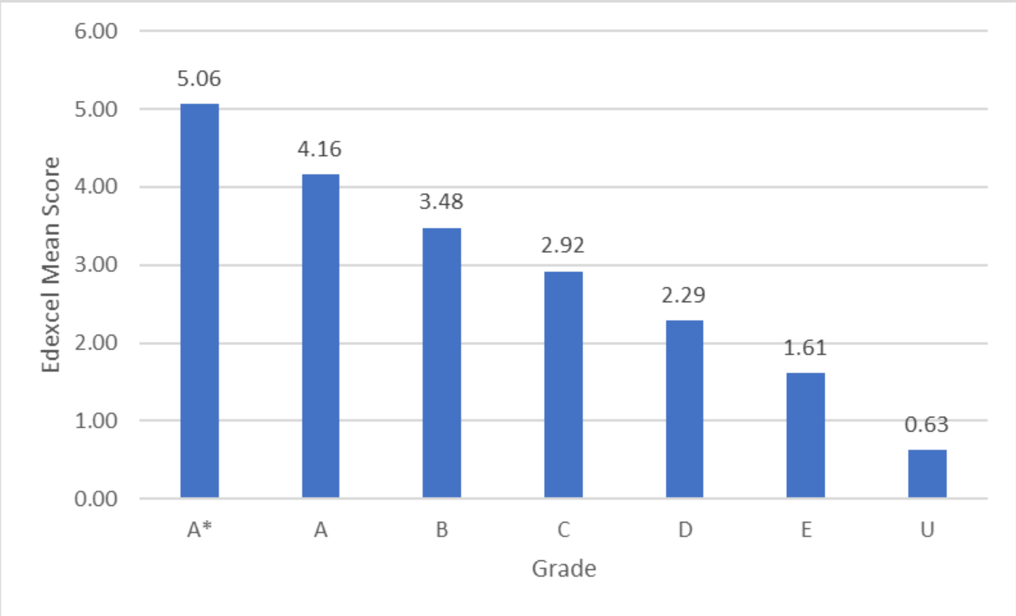
C

Skip to Main Contents

Question: 1 2 3 4 5 6

Question 3 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade: A*	A	B	C	D	E	U
3.43	7	49%	5.06	4.16	3.48	2.92	2.29	1.61	0.63



Q3
?
A
B
C

Question 3 - Response A

(a) Ben will need to get rid of any days where the rain was not recorded for some reason, this will reduce the number Ben divides by to find the mean

(b) (i) $\bar{x} = \frac{\sum x}{n}$ $\quad \frac{390}{184} = 2.1$ mm

(ii) $\sqrt{\frac{\sum x^2}{n} - \bar{x}^2} = \sqrt{\frac{4336}{184} - (2.119565)^2}$

$= 4.37$

Standard deviation $= 4.4$ 2dp.

(c)(i) this is only 49.59% of the year so not sensible to apply to whole year as some months will have much more rainfall

(c)(ii) in 1987 there was a huge storm in the UK which includes Leeming so the actual amount of rainfall would be more than the estimated

2/7 marks

Part (a)
Removing the tr values was not an appropriate cleaning process. Rather they should be replaced by a suitable value (M0 and therefore A0)

Part (b)
Despite the front cover telling students to give answers from a calculator to 3sf there was only a 2sf answer here (B0). In part (ii) a correct expression was given scoring the method mark (M1) and a correct 3sf answer was given which would score the accuracy mark (A1). The rounded answer of 4.4 is ignored since the mark can be awarded for 4.37 in the line above.

Part (c)
(i) The student has realised that only part of the year is used to derive the estimates but hasn’t realised that it is not the size of the sample that is an issue here but the fact that it is not a random (representative) sample of the whole year. The mention of the fact that just the months May to October were being considered was required. (B0)
(ii) The storm wasn’t the key issue here but rather the rainfall in the missing months and this was not mentioned. (B0)

Q3
?
A
B
C

Question 3 - Response B

a) instead of having a range of values for x, midpoints need to be found to represent x and value for x will need to be in the final column instead of 'tr'.

(2)

bi) $\bar{x} = \frac{390}{184} = \frac{195}{92} = 2.12$

bii) $\sqrt{\frac{2336}{184} - \frac{195}{92}} = 4.63$

ci) ~~some days rainfall was not~~ rainfall was only observed/recorded from may - October so this isn't representative of the whole year so can't account for annual rainfall

ii) increase as outside of May-October, it is winter/ ~~spring~~ in Leeming so there's more rainfall

4/7 marks

Part (a)

They have identified the fact that tr values need replacing with a numerical value (M1) but have not suggested what the value should be (A0)

Part (b)

The mean is correct (B1) but their formula for the standard deviation is not correct – the mean is not squared (M0 and therefore A0).

Part (c)

In **(i)** they have mentioned that the data is only for May~Oct and therefore not representative (B1). In **(ii)** they have identified that the missing months are winter and there is likely to be more rain then. They have said that the estimate should "increase" and this is equivalent to saying that the estimate would be an "underestimate"(B1). Equivalent wording is accepted for this mark.

Q3 ? A B C

Question 3 - Response C

a) Ben will need to clean the data by ~~ø~~ finding the data between May to October of 1987

a) He will need to change ~~the~~ Tr (trace) to less than 0.05 mm.

n = 184

b) (i) $\bar{x} = \frac{390}{184}$ ~~ø~~ $\bar{x} = \frac{195}{92}$ mm.

(ii) $\sigma_x = \sqrt{\frac{4336}{184} - \left(\frac{195}{92}\right)^2} = 4.37$ (3 sf).

c) ~~(i)~~ (i) Because the data recorded is only between May and October 1987 ∴ unsuitable.

~~(ii) I would expect the mean estimate in part (b)(i) to be a lower expectation of the actual mean Daily Total Rainfall in Leeming 1987 because in 1987 there was the great storm.~~

(ii) the estimation would differ with an increase as the recording is only between May-October of 1987 therefore not taking into account the annual rainfall ∴ differing from the actual mean Daily Total Rainfall.

6 / 7 marks

Part (a)

The first comment doesn't score (but it is not contradictory so it can be ignored) but their second comment states that the "tr" items need replacing with a value (M1) and suggests a suitable value (less than 0.05 mm is acceptable) (A1)

Part (b)

A simplified fraction for (i) was accepted (B1) and (ii) is fully correct (M1A1)

Part (c)

(i) They have said that the data is unsuitable and mentioned the months May to October (B1)

(ii) They have not suggested what the rainfall in the missing months was likely to be and then deduced what impact this would have on the estimate in (b)(i) (B0)

Q3

A

B

C

Question 4

Question 4 - Question

4. A study was made of adult men from region *A* of a country.
It was found that their heights were normally distributed with a mean of 175.4 cm and standard deviation 6.8 cm.

(*a*) Find the proportion of these men that are taller than 180 cm.

(1)

A student claimed that the mean height of adult men from region *B* of this country was different from the mean height of adult men from region *A*.

A random sample of 52 adult men from region *B* had a mean height of 177.2 cm

The student assumed that the standard deviation of heights of adult men was 6.8 cm both for region *A* and region *B*.

(*b*) Use a suitable test to assess the student's claim.
You should

- state your hypotheses clearly
- use a 5% level of significance

(4)

(*c*) Find the *p*-value for the test in part (*b*)

(1)

(Total for Question 4 is 6 marks)

Question 4 - Mark Scheme

Qu 4	Scheme	Marks	AO
(a)	[Let N = height from region A; P(N > 180) =] 0.24937… awrt **0.249**	B1 (1)	1.1b
(b)	$H_0 : \mu = 175.4 \quad H_1 : \mu \neq 175.4$	B1	2.5
	[S = height from region B] $\bar{S} \sim N\left(175.4, \frac{6.8^2}{52}\right)$ Allow σ^2 = awrt 0.889	M1	3.3
	[P($\bar{S}$ > 177.2)] = 0.02814…	A1	3.4
	[0.028… > 0.025, Not sig, do not reject H_0] Insufficient evidence to support student's claim	A1 (4)	2.2b
(c)	[p-value = $2 \times 0.02814...$ =] 0.05628… in range **0.056~0.06** or **5.6(%)~6(%)**	B1ft (1)	1.2
		(6 marks)	

	Notes
(a)	B1 for awrt 0.249
(b)	B1 for both hypotheses correct in terms of μ **(See below for one-tail test)** M1 for selecting the correct model, may be implied by standardisation using correct values or may be implied by a correct value in 1st A1 e.g.(Prob =) 0.028 or awrt 0.972, (Z =) 1.9(08..) (CV=) 177.25 Condone use of S (or any other letter) instead of $\bar{S}$ Condone use of $\bar{S} \sim N\left(177.2, \frac{6.8^2}{52}\right)$ but **this will lose 2nd A mark** 1st A1 for probability of awrt 0.028 (allow 0.03 if P($\bar{S}$ > 177.2) is seen) Condone 1 – 0.02814 … = 0.9718…(awrt 0..972) **only if clearly compared with 0.975**
ALT	Allow Z = 1.9(088…) and comparison with 1.96 (or better: calc gives 1.95996…) or CR of $\left[\bar{S}\right]$… 177.248…(awrt 177.25) Allow $\left[\bar{S}\right]$ > 177.248…(awrt 177.25) Implied by diagram or correct interpretation of inequality with their CV (Ignore any attempt at a lower CR for $\bar{S}$) 2nd A1 (dep on 1st A1 and use of correct model. Use of N(177.2, …) scores A0) for a conclusion using context: e.g. does not support student's claim or e.g. insufficient evidence of a difference in heights Do not allow 2nd A mark for contradictory statements e.g. "significant" so "no support for claim"
(c)	B1ft for answer in range 0.056~0.06 or 5.6%~6% (Ranges are inclusive, condone missing %) (can ft their probability, provided < 0.5, from part (b) but not 0.025 leading to 5%)
NB	**One-tail test [Max of 3/5 for (b) and (c)]** In (b) B0 (hypotheses) M1(model as above) 1st A1[for probability or Z compared with 1.6449 or CR $\left[\bar{S}\right]$… or > 176.95… (awrt 177)] 2nd A1 for conclusion in context that supports claim **or** "heights of men from B is different from/greater than from A" In (c) B0

Question 4 - Examiner Comments

Whilst it was rare to see an incorrect answer to part (a) a minority of students were unable to use their calculator to evaluate this simple probability using a normal distribution.

In part (b) most students stated their hypotheses correctly in terms of μ with very few using a one-tailed test. To score the method mark in this type of question the students simply need to select and write down the correct model (distribution). A number seemed to be listing numbers they needed for their calculator, but we expect to see the appropriate mathematical notation used. Fortunately, many obtained a correct probability and would therefore receive the method mark by implication. Some chose to standardise and find a z value (1.9088… which should be compared with 1.96) or a critical region $\left(\bar{S} \geq 177.248...\right)$ but given that a p-value was required in part (c) this was not the most efficient route. A common error in stating the model is to use an incorrect value for the mean. On this occasion we condoned $\bar{S} \sim \mathrm{N}\left(177.2, \frac{6.8^2}{52}\right)$ for the M1 and first A1 but penalised the final accuracy mark. In a hypothesis test the probability (or critical region) is always calculated assuming H_0 is true and so the mean should be 175.4 and the probability calculated is $\mathrm{P}\left(\bar{S} > 177.2\right)$. Some students stopped after stating that the result was not significant or not rejecting H_0 but a good number gave a correct conclusion in context.

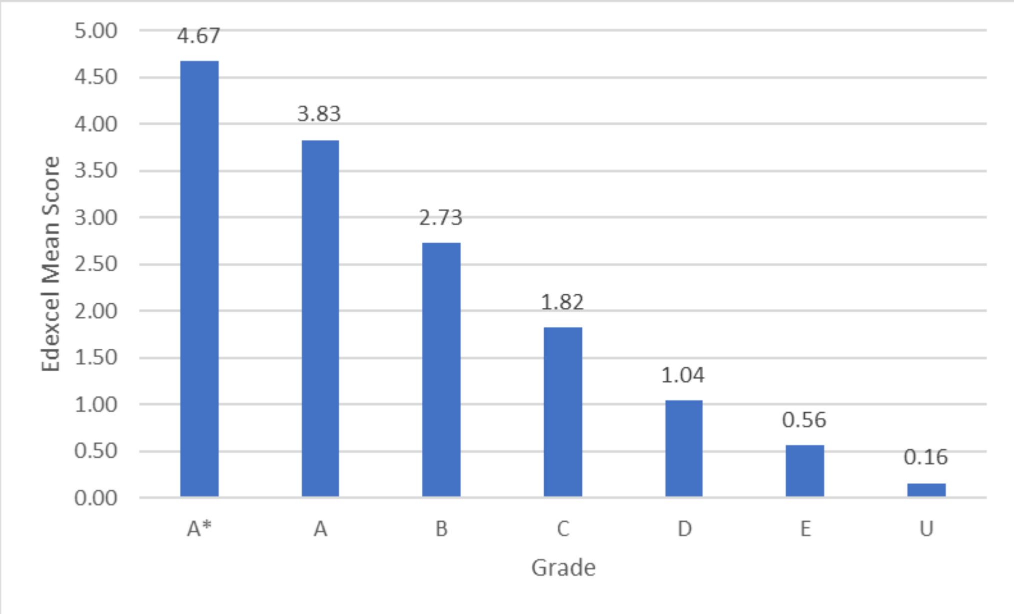
Part (c) was not answered well with only a few scoring this mark. The common error was to use 0.028 or 0.025 but others used values like 175.4

Q4

A
B
C

Question 4 - Performance

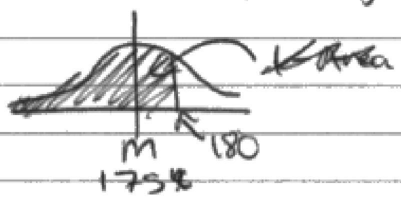
			Edexcel averages: mean scored by candidates achieving grade:						
Mean score	Max score	Mean %	A*	A	B	C	D	E	U
2.71	6	45%	4.67	3.83	2.73	1.82	1.04	0.56	0.16



Question 4 - Response A

a) $\mu = 175.4$ $\quad \sigma = 6.8$

$Y \sim N(179.4, 6.8^2)$



$= 0.2494$

b) sample: 52 $\quad \mu = 177.2$ $\quad$ ~~σ = 6.8~~ $\quad \sigma = \frac{6.8}{\sqrt{52}}$

$H_0: \mu = 175.4$

$H_1: \mu > 175.4$

~~178.0252m~~ $= 176.95$

Accept H_0 as no evidence to suggest students claim

c) $\mu = 175.4$

2/6 marks

Part (a)

This is correct (B1)

Part (b)

The hypotheses are not clear but sight of the value 176.95 (the critical value for a one-tail test - see the bottom of the mark scheme) suggests that the student intends to be specifying a one-tail test which is incorrect (B0). There is no sight of the correct model stated but the M mark is implied by the critical value of 176.95 (M1). They do not state the correct critical region (CR), we'd need to see something like $\bar{S} \geq 176.95$, (1st A0 and therefore 2nd A0 as this mark depends on 1st A mark being scored.)

Part (c)

If they are using a one-tail test they cannot score this mark however it is clear that this student does not understand the concept of *p*-value. (B0)

Q4
A
B
C

Question 4 - Response B

(1)

4a) $P(X > 180)$

lower $= 180$, upper $= 1 \times 10^6$, $\mu = 175.4$, $\sigma = 6.8$

$= 0.2493709466$

b) $H_0: \mu = 177.2$ $H_1: \mu \neq 177.2$

$\bar{X} \sim N(177.2, \frac{6.8^2}{5^2})$

$P(\bar{X} \leq 175.4) = 0.02814$

as change, significance level is halved to 2.5% / 0.025

$0.028 > 0.025$ ∴ not significant,

do not reject Ho, insufficient evidence at 2.5% level that the mean height of men from region B was different from region A

c) 0.028?

3/6 marks

Part (a)

This is correct (B1)

Part (b)

They have not used the correct value of 175.4 (B0) but are still using a two-tail test. The model stated is incorrect (mean should be 175.4 not 177.2). This error was condoned for the M mark and 1st A mark but students making this error will be penalised by losing the 2nd A mark. The correct probability of 0.028 is seen (M1 1st A1) but the error in mean loses the final mark (2nd A0)

Part (c)

This is incorrect, the student knows it is something to do with probability but didn't adjust for the two-tailed test. (B0)

Examiner comments

In part (b) there were many places where students were getting confused. Firstly, some mis-interpreted the question as a one-tailed test (see the previous example) despite the clear use of the phrase "different from" in the question. Secondly some students (as here) failed to appreciate that the hypotheses should be stated in terms of the population parameter not the sample mean (175.4 – the hypothesis is that the mean height from region *B* is the same as the mean height from region *A* which was given as 175.4). This often led students to state or use an incorrect model with mean of 177.2 rather than 175.4 but that was condoned for the method mark and first accuracy mark. Finally, some students didn't know how to interpret their probability (as here).

Q4
?
A
B
C

Skip to Main Contents

Question: 1 2 3 4 5 6

Question 4 - Response C

4 (a) X = men from region A

$X \sim N(175.4, 6.8^2)$

$P(X > 180) = 0.1617092675$

(b) X = mean height of men from region B

~~$X \sim N(\mu, \sigma$~~ $X \sim N(\mu, 6.8^2)$

$H_0: \mu = 175.4$ null hypothesis asserts mean height for men in region B is 175.4cm

$H_1: \mu \neq 175.4$ alternate hypothesis asserts mean height of men in region B is not 175.4 cm

Assume Ho is true, then $X \sim N(175.4, 6.8^2)$

$\bar{X}_{52} \sim N\left(175.4, \frac{6.8^2}{52}\right)$ ~~Nnot~~

$\bar{X}_{52} \sim N\left(175.4, \frac{289}{325}\right)$ $\sigma = \sqrt{\frac{289}{325}}$

$P(\bar{X} > 177.2) = 0.02814257647$

this is a two tailed test at 5% level signifcance so p value $= \frac{0.05}{2} =$ ~~0.0~~ 0.025

$0.02814257647 > 0.025$ so there is signficant evidence at 5% signifcance to not reject the null hypothesis that the mean height for men in region B is 175.4cm

(c) $0.02814257647 \times 2 = 0.05628515294$

4/6 marks

Part (a)

The answer is incorrect (B0)

Part (b)

The hypotheses are correct (B1) note that we would condone the very short "tail" on the *m*, the correct model is stated (M1) and the correct probability is seen (1st A1). Their conclusion is that they should not reject H0 and therefore that the mean height of men from region *B* is 175.4. However, they have said that "there is significant evidence" and this is a technical phrase that students would be expected to use correctly. There are contradictory statements (2nd A0).

Part (c)

This is a rare case where the correct answer is given (B1)

Q4

A
B
C

Question 5

Question 5 - Question

5. Tisam is playing a game.
She uses a ball, a cup and a spinner.

The random variable X represents the number the spinner lands on when it is spun.
The probability distribution of X is given in the following table

x	20	50	80	100
$\mathrm{P}(X = x)$	a	b	c	d

where a, b, c and d are probabilities.

To play the game

- the spinner is spun to obtain a value of x
- Tisam then stands x cm from the cup and tries to throw the ball into the cup

The event S represents the event that Tisam successfully throws the ball into the cup.

To model this game Tisam assumes that

- $\mathrm{P}(S \mid \{X = x\}) = \dfrac{k}{x}$ where k is a constant
- $\mathrm{P}(S \cap \{X = x\})$ should be the same whatever value of x is obtained from the spinner

Using Tisam's model,

(*a*) show that $c = \dfrac{8}{5}b$

(2)

(*b*) find the probability distribution of X

(5)

Nav tries, a large number of times, to throw the ball into the cup from a distance of 100 cm.
He successfully gets the ball in the cup 30% of the time.

(*c*) State, giving a reason, why Tisam's model of this game is not suitable to describe Nav playing the game for all values of X

(1)

(Total for Question 5 is 8 marks)

Question 5 - Mark Scheme

Qu 5	Scheme	Marks	AO
(a)	$\text{P}(S \cap \{X = 50\}) = \text{P}(S \cap \{X = 80\})[= \text{a constant, } V] \Rightarrow b \times \frac{k}{50} = c \times \frac{k}{80}$ May see: $\frac{k}{50} = \frac{V}{b}$ and $\frac{k}{80} = \frac{V}{c}$ (condone any letter for V even S)	M1	3.1a
	So $c = \frac{8}{5}b$ *	A1cso*	1.1b
		(2)	
(b)	$d = 2b$ **or** $a = \frac{2}{5}b$ **or** $c = 4a$ **or** $d = 5a$ **or** $d = \frac{5}{4}c$	M1 A1	2.1 3.3
	$\frac{2}{5}b + b + \frac{8}{5}b + 2b = 1$	M1	2.1
	$\Rightarrow 5b = 1$ so $b = \frac{1}{5}$ (o.e.)	A1	1.1b
	$a = \frac{2}{25}$ $b = \frac{1}{5}$ $c = \frac{8}{25}$ $d = \frac{2}{5}$	A1	3.2a
		(5)	
(c)	[Experiment suggests for Nav] $\text{P}(S \mid \{X = 100\}) = 0.3 \Rightarrow k = 30$ **or** $0.3 = \frac{V}{0.4} \Rightarrow V = 0.12$ So model won't work since $\text{P}(S \mid X = 20) = \frac{30}{20}$ or $\frac{0.12}{0.08}$ and so would be greater than 1	B1	2.4
		(1)	
		(8 marks)	

	Notes
(a)	M1 for use of $\text{P}(S \mid X = x) \times \text{P}(X = x)$ for $x = 50$ and $x = 80$ (Must see k or their V) Any expression or equation MUST be based on the probability statements in qu.
*	A1cso for rearranging to required result, no incorrect work seen, condone poor notation
NB	Use of values e.g. $b = \frac{50}{20 + 50 + 80 + 100}$ to prove (a) is M0A0 but scores 2nd M1A1 in (b)
	Marks for (b) may be awarded for work seen in (a)
(b)	1st M1 for at least one other relationship (either probability the subject) from the list. 1st A1 for a second different relationship (either probability the subject) from the list. or Allow for: $\frac{ak}{20} = \frac{bk}{50} = \frac{ck}{80} = \frac{dk}{100}$ for 1st M1 1st A1 2nd M1 for using or stating sum of prob's = 1 May be implied by one correct probability. 2nd A1 for one correct probability e.g. $b = \frac{1}{5}$ or exact equivalent such as 0.2 3rd A1 for all correct probabilities. Allow exact equivalents e.g. $c = 0.32$ **Sight of correct distribution or list of probs with no obvious incorrect working is 5/5**
(c)	B1 for deducing $k = 30$ **and** giving a suitable example to show model breaks down

Q5

?

A

B

C

Question 5 - Examiner Comments

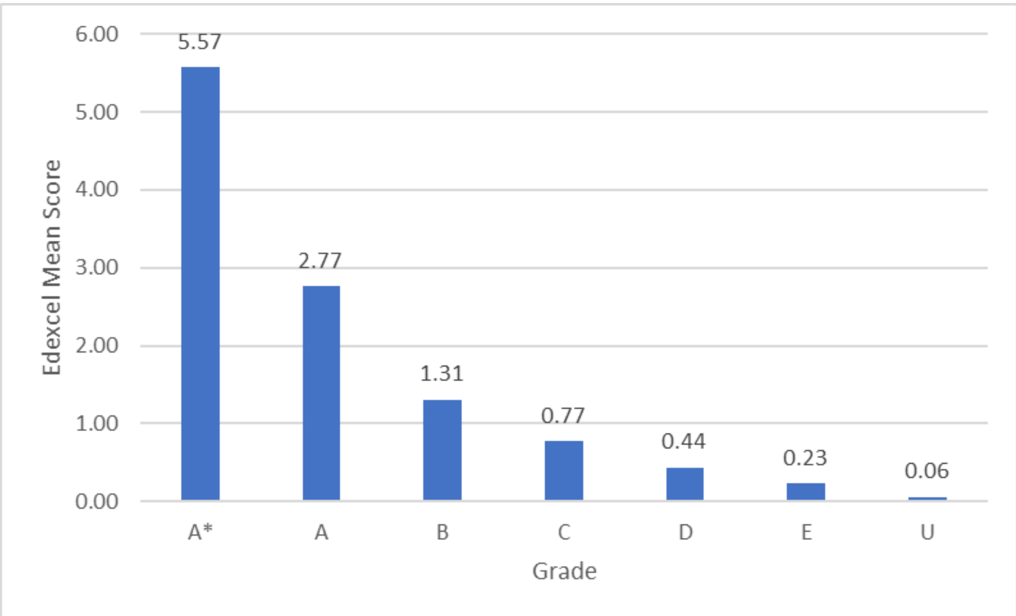
This question proved to be the most challenging on the paper for students.

In part (a) few could see how to use the two bullet points to reach the required result. A good number realised that $\mathrm{P}(S\,|\,\{X=x\}) = \frac{k}{x}$ but couldn't always use the conditional probability formula to reach $\mathrm{P}(S \cap X = 50) = b \times \frac{k}{50}$. Those that did get this far were often able to give a convincing proof of the given result. A common misconception was believing that $a = \frac{k}{20}, b = \frac{k}{50}$ etc and then using the sum of the probabilities equalling 1 but this scored no marks. It was good to see a number of students using the result from part (a) to form similar relationships between the probabilities, such as $d = 2b$ and going on to solve their equations to find the probability distribution.

The final part was particularly challenging with many misinterpreting the 30% probability given in the stem to part (c). Many confused this 0.3 [= $\mathrm{P}(S\,|\,X = 100)$] with the 0.4 [= $\mathrm{P}(X = 100)$] from the probability distribution and simply compared these two values. Those that did realise that Nav's 30% meant that $k = 30$ often went on to show that when $x = 20$ this would give a probability [$\mathrm{P}(S\,|\,X = 20)$] > 1 and so the model was not suitable.

Question 5 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade: A*	A	B	C	D	E	U
2.03	8	25%	5.57	2.77	1.31	0.77	0.44	0.23	0.06



Q5
?
A
B
C

Question 5 - Response A

a)

x	20	50	80	100	250
$p(X=x)$	a	b	c	d	(

$p(X=80) = c \qquad \frac{80}{250} = c = 0.32$

$p(X=50) = b \qquad \frac{50}{250} = 0.2 = b$

$$\boxed{\begin{gathered} \text{as } c = \tfrac{8}{5}b \\ 0.32 = \tfrac{8}{5} \times 0.2 \end{gathered}}$$

2/8 marks

Part (a)

This candidate has used the summation property noted on the mark scheme. The conditions for this distribution given in the question can be shown to yield this result (see the technical note as an appendix to the mark scheme) but they have not shown this and so no marks can be awarded in (a), there is no k or V seen. (M0A0)

Part (b)

b = 0.2 is seen in part (a) and marks for (b) seen there can be awarded. One correct probability (in fact they have two) is sufficient to imply 2nd M mark and then also they score 2nd A mark. The first M mark is not scored as another of the given relationships between a, b, c and d is not seen and a fully correct distribution is not seen. (M0A0M1A1A0).

Part (c)

There is no attempt (B0)

Examiner comments

This question was found to be challenging for many students. Marks for part (b) were allowed to be scored if they were seen in the working for part (a) and there were some unusual ways of achieving the correct answers for part (b) (as in this case) which were allowed (because they applied under the conditions given for this question) but did not gain full credit unless their justification was shown.

Q5 ? A B C

Question 5 - Response B

5a) ~~a~~ $b=50$ $c=80$

$a+b+c+d=1$

~~$a=20$~~ $b=\frac{50}{250}=\frac{1}{5}$

$c=\frac{80}{250}=\frac{8}{25}$

$b=\frac{1}{5}$

$c=\frac{8}{25}$

$\therefore \frac{\frac{8}{25}}{\frac{1}{5}} = \frac{8}{5}$

$\boxed{\therefore c = \frac{8}{5}B}$

b) $\frac{2}{5}b + b + \frac{8}{5}b + 2b = 1$

$\therefore 5b = 1$

~~$\therefore b = 0.2$~~

~~$\therefore A =$~~

$\frac{p(S \cap (X=x))}{P(X=x)} = \frac{k}{x}$

$\frac{p(S \cap (x=x))}{1} = \frac{k}{x}$

$P(S \cap (x=x)) = \frac{k}{x}$

c) In Tijams model we are assuming the probability of getting it in after each try is constant but with each go Nav is getting better $\therefore$ Not suitable.

4/8 marks

Q5
?
A
B
C

Part (a)

This candidate has also used the summation property noted on the mark scheme. The conditions for this distribution given in the question can be shown to yield this result but this student has not shown this and so no marks can be awarded in (a) (M0A0)

Part (b)

At the start of their part (b) they have used $a + b + c + d = 1$ with two other correct relationships ([a =] $\frac{2b}{5}$ and [d =] $2b$ the a =… and d = … can be implied by the position/context) (1st M1 1st A1). They are using the sum of probabilities = 1 (2nd M1 this could have been given on the first page when $a + b + c + d = 1$ was seen) and have a correct value for b (and also c, both seen in the working to their part (a)) (2nd A1) but they do not find the other values (3rd A0).

Part (c)

Their comment does not make use of Nav's results and there is no attempt to find k or P(S and X = 100) (B0).

Question 5 - Response C

a) ~~80k~~ $\frac{S}{b} = \frac{k}{80}$ $\quad \frac{S}{c} = \frac{k}{80}$

$\frac{S}{k} = \frac{b}{50}$ $\quad \frac{S}{k} = \frac{c}{80}$

$\frac{b}{50} = \frac{c}{80}$ $\quad c = \frac{80b}{50} = \frac{8}{5}b$

$c = \frac{8}{5}b$

b) $c = \frac{8}{5}b$ $\quad 5c = 8b$

$\frac{S}{k} = \frac{a}{20} = \frac{d}{100} = \frac{b}{50}$

$d = 2b$

$a = 0.4b$

$0.4b + b + 1.6b + 2b$

$= 1$

$5b = 1$ $\quad b = 0.2$

$a = 0.08$

$b = 0.2$

$c = 0.32$

$d = 0.40$

Probability distribution of x:

x	20	50	80	100
$P(X=x)$	0.08	0.2	0.32	0.4

c) As $P(S \cap \{X=x\})$ is the same regardless of x.

The probability Nav successfully gets the ball into the cup at 100cm is $0.3 \times 0.4 = 0.12$.

This means that at a distance of 20cm, to keep the model complete, Nav must successfully score 150% of the time which is not possible. therefore the model isn't suitable to describe Nav playing the game.

$\frac{0.12}{0.08} = 1.5$
$= 150\%$

8 / 8 marks

Part (a)

This solution is basically the one given in the scheme with *S* used for our *V* (this poor use of notation is condoned as *S* is already defined as an event in the question). They have used the formula with $x = 50$ and $x = 80$ (M1) and formed a suitable equation to simplify to the given result (A1).

Part (b)

Two of the other relationships are given (M1A1) and then the sum of probabilities property is stated (2^{nd} M1) and the correct distribution follows (2^{nd} A1 3^{rd} A1). Note that the distribution completed in a table is not needed for the mark to be awarded, just a list of the values for *a*, *b*, *c* and *d*. However, this student does give a correct probability distribution in a table on their next page.

Part (c)

They have used the 30% correctly as a conditional probability and used this to find the equivalent value of $V = 0.12$. They have then used this with $x = 20$ to show that the conditional probability would be greater than 1 (B1). A rare case of a correct solution to this part.

Q5

?

A

B

C

Question 6

Question | Mark Scheme | Examiner Comments

Performance | Response A | Response B | Response C

Question 6 - Question

6. A medical researcher is studying the number of hours, T, a patient stays in hospital following a particular operation.

The histogram on the next page summarises the results for a random sample of 90 patients.

(*a*) Use the histogram to estimate $P(10 < T < 30)$

(2)

For these 90 patients the time spent in hospital following the operation had

- a mean of 14.9 hours
- a standard deviation of 9.3 hours

Tomas suggests that T can be modelled by $N(14.9, 9.3^2)$

(*b*) With reference to the histogram, state, giving a reason, whether or not Tomas' model could be suitable.

(1)

Xiang suggests that the frequency polygon based on this histogram could be modelled by a curve with equation

$$y = kxe^{-x} \qquad 0 \leqslant x \leqslant 4$$

where

- x is measured in **tens of hours**
- k is a constant

(*c*) Use algebraic integration to show that

$$\int_0^n xe^{-x}\,dx = 1 - (n + 1)e^{-n}$$

(4)

(*d*) Show that, for Xiang's model, $k = 99$ to the nearest integer.

(3)

(*e*) Estimate $P(10 < T < 30)$ using

(i) Tomas' model of $T \sim N(14.9, 9.3^2)$

(1)

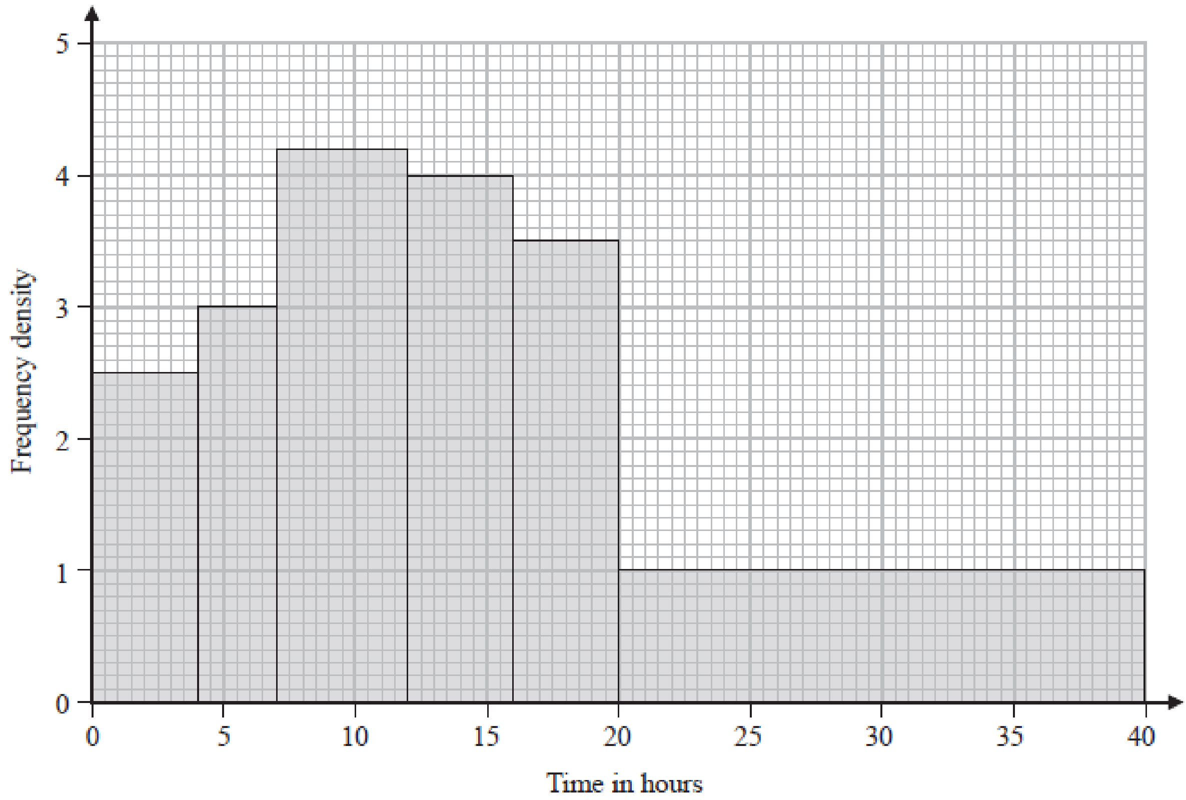
(ii) Xiang's curve with equation $y = 99xe^{-x}$ and the answer to part (*c*)

(2)

The researcher decides to use Xiang's curve to model P($a < T < b$)

(f) State one limitation of Xiang's model.

(1)



(Total for Question 6 is 14 marks)

Question 6 - Mark Scheme

Qu 6	Scheme	Marks	AO
(a)	$2\times 4.2\,,\,4\times 4\,,\,4\times 3.5\,,10\times 1\quad (= 8.4 + 16 + 14 + 10 = 48.4)$	M1	1.1b
	[So $P(10 < T < 30) =$] $\left[\dfrac{48.4}{90}\right] = \dfrac{121}{225} = 0.53777...$ $\quad$ **0.53~0.54** (2sf OK)	A1	1.1b
		(2)	
(b)	(Not suitable as) data is not symmetric <u>or</u> is skew (normal is symmetric) ("Even" distribution or a diagram <u>on its own</u> is not enough so B0)	B1 **(1)**	2.4
(c)	$\int xe^{-x}\ (dx) = \int x\,d(-e^{-x})$	M1	2.1
	$= \left[-xe^{-x}\right] - \int\left(-e^{-x}\right)\ (dx)\quad (+c)$	A1	1.1b
	$\int_0^n xe^{-x}\ (dx) = \left[-xe^{-x} - e^{-x}\right]_0^n = \left(-ne^{-n} - e^{-n}\right) - \left[-(0) - 1\right]$	dM1	1.1b
	$= \underline{1 - (n+1)e^{-n}}\quad (*)$	A1cso*	1.1b
		(4)	
(d)	Require area = 90 i.e. $k\int_{(0)}^{(n)} xe^{-x}\ dx = 90\quad$ (ignore limits)	M1	3.1a
	Using the result in part (c) with $n = 4$ gives $k\left[1 - 5e^{-4}\right] = 90$	M1	2.1
	$(k =)$ <u>99</u>(.0729…) (*)	A1cso*	1.1b
		(3)	
(e)(i)	[$P(10 < T < 30) =$] 0.64863… $\quad$ awrt **<u>0.649</u>**	B1 **(1)**	1.1b
(ii)	[No. of patients =] $\quad (99)\left[\left(1-4e^{-3}\right) - \left(1-2e^{-1}\right)\right]\quad (= 53.1..)$	M1	3.4
	Prob $= \dfrac{0.5366...\times 99}{90} = 0.59027...$[or 0.5907...] $\quad$ = awrt **<u>0.590</u>** or **<u>0.591</u>**	A1 **(2)**	3.2a
(f)	eg Patients might stay longer than 40 hours (Can ignore other comments unless clearly contradictory.)	B1 **(1)**	3.5b
		(14 marks)	

	Notes
(a)	M1 for an attempt to find the number between 10 and 30 (2 correct products or 48 or 48.4 seen) A1 for 2sf answer in [0.53 ~ 0.54] NB use of 48 gives 0.5333… [Correct ans implies 2/2]
(b)	B1 for a comment suggesting not suitable based on (lack of) symmetry or "not bell shaped"
(c)	1st M1 for attempting integration by parts in right direction. Must have $u = x$ and $v = \pm\ e^{-x}$ 1st A1 for a correct first step, correct first integration and expression for second integral 2nd dM1 (dep on 1st M1) for all integration attempted and some use of at least one limit
*	2nd A1 for cso with no incorrect working seen. Minimum is correct int and use of limits seen.
(d)	1st M1 for realising need area under the curve (implied by the integral) = 90 2nd M1 for use of (c) with $n = 4$ and set = 90 May be implied by sight of 99.07… or better
*	A1cso for $k = 99$ or awrt 99.1
NB	Allow use of $k = 99$ and show area = awrt 89.9 with a conclusion to score 3/3
(e)(i)	B1 for awrt 0.649
(ii)	M1 for use of (c) with $n = 1$ and $n = 3$ Don't need the 99. Implied by sight of awrt 0.54 A1 for awrt 0.590 or awrt 0.591 Allow 0.59 from correct working seen.
(f)	B1 eg for comment, in context, about the upper limit for time (t or x)(time/hour may be implied)

Q6 ? A B C

Question 6 - Examiner Comments

This question also proved to be quite challenging.

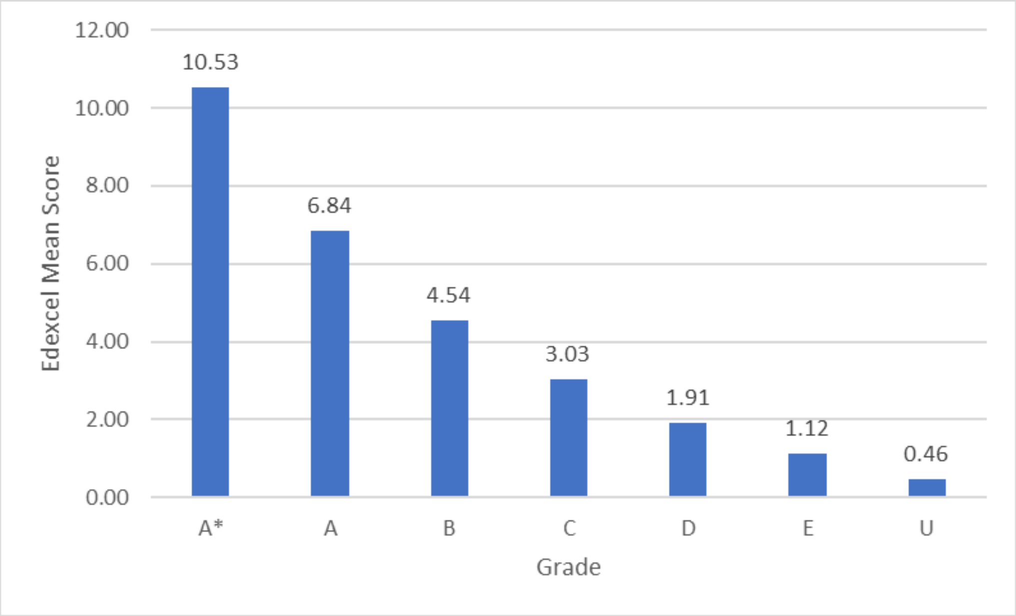
In part (a) most students could use the histogram to calculate the number of patients spending between 10 and 30 hours in hospital but many failed to convert this to a probability by dividing by 90. Some students tried using small squares, but errors were often made and a divisor of 1600 not 1800 was often seen. Part (b) though was not answered so well. Some did explain that the histogram was not symmetric or was skewed and so a normal distribution wouldn't be suitable. Some focussed on the continuous nature of the variable and thought it was therefore suitable whilst others thought that T was discrete and therefore a normal wasn't suitable. The question wanted reference to the histogram and there was no expectation or requirement to find probabilities from the normal distribution nor compare the mean of the data with 14.9. If further calculations were expected these would be asked for and reflected in the mark allocation.

Many students were able to use integration by parts correctly to establish the result in part (c). Some started the wrong way around (using $u = e^{-x}$) and soon realised they were not heading for the given answer and there was some muddling over the signs when applying the limits. Part (d) was not answered so well with only a minority of students realising that the area under the curve should equal 90. Those who did realise this were often able to apply the answer to part (c) though some used $n = 40$ rather than $n = 4$.

The first part of (e) was often answered well and some realised that the answer to part (c) could be used to answer (e)(ii) but some tried using $n = 10$ and 30 and others stopped after find an estimate for the number of patients as 53.1 and failed to complete the question to find the required probability. In part (f) only a small number of candidates realised that the fact that some patients might stay longer than 40 hours was a limitation of the model. Others did not give any context in their answers or discussed the shape of the curve without considering the stated upper limit.

Question 6 - Performance

			Edexcel averages: mean scored by candidates achieving grade:						
Mean score	**Max score**	**Mean %**	**A***	**A**	**B**	**C**	**D**	**E**	**U**
5.09	14	36%	10.53	6.84	4.54	3.03	1.91	1.12	0.46



Question 6 - Response A

Total of 90 patients

$40 + 20 + 0.5 \times 20 + 0.5 \times \cancel{10} 6 + 0.5 \times 5$
$+ 0.5 \times 6.5 + 1$

a) $P(10 < T < 30)$

Freq 4.2 1

$1 \times 20 + 10 + 10 + 5 + 2.58 + 1 \quad \dfrac{48.55}{90} = 0.539$

b) as calculated above probability of ~~the~~ time in hospital between 10 and 30 is 0.539 however according to Tomas' model it is 0.399 (3sf) $\therefore$ the model may not be suitable

c) $y = kxe^{-x} \quad 0 \le x \le 4$

$x = h \ x{=}0$

$\hookrightarrow$ continued on next page.

c) $\displaystyle\int_0^n xe^{-x}\,dx = 1-(n+1)e^{-n}$

$u = x \qquad v = -e^{-x}$

$\dfrac{du}{dx} = 1 \qquad \dfrac{dv}{dx} = e^{-x}$

$-xe^{-x} - \int -e^{-x}(1)\,dx$

$\left[-xe^{-x} - e^{-x}\right]_0^n$

$\left[-(n)e^{-(n)} - e^{-(n)}\right] - \left[-(0)e^{-(0)} - e^{-(0)}\right]$

$-ne^{-n} - e^{-n} + 1$

$1 - ne^{-n} - e^{-n}$

$1 - (n+1)e^{-n}$

d) $y = kxe^{-x} \quad 0 \leq x \leq 4$ Show that $k = 99$

$y = k(4)e^{(4)}$

$y = k(0.073)$

~~[illegible]~~ k

$y = k\left[1-(x+1)e^{-x}\right]$

$x = 4$

$y = k\left[1-(4+1)e^{-4}\right]$

$1 - \cancel{0.073}\ 0.05$

$y = k\ 0.95$

e) i ≈ 0.349 (3sf)

ii $\approx$ $n = 10$ $\quad 1-(10+1)e^{-10} = 0.9995$

$1-(30+1)e^{-30} \approx 1$

$1 - 0.9995 = 1.41 \times 10^{-3}$

f) ~~the~~ Xiang's model uses a curve which may overestimate the ~~[illegible]~~ histogram's probability in some locations as well as underestimate the ~~proportion~~ probability in other locations

6/14 marks

Part (a)

They have a final answer in the range 0.53~0.54 so we award both marks (M1A1). This is an uncommon approach, but the method is perfectly acceptable: they have used rows rather than columns. The 20 + 10 + 10 + 5 is all fine but their “2.55” should be 6×0.5 = 3 and the final “1” should be 2×0.2 = 0.4. The mark scheme allows for small errors in their calculations but as long as the final answer is in the range to 2sf we can award the marks.

Part (b)

Their comment relates to probabilities and not the (shape of the) histogram. (B0) A correct answer to (e)(i) was occasionally embedded here but the mark was not awarded unless it was in part (e).

Part (c)

This is fully correct (M1A1dM1A1)

Part (d)

There is no attempt to set the area = 90 (1st M0) and although they use the result from part (c) they do not use this to form an equation in *k* (2nd M0 and therefore A0)

Part (e)

Part (i) is incorrect (B0) and they are using n = 30 and 10 in (ii) (M0 and therefore A0)

Part (f)

There is no mention of the upper limit for time (B0)

Examiner comments

The specification requires the bringing together of some topics from other areas into a question from time to time. Some students didn't seem prepared to use integration by parts in this paper, but part (c) was for many other students (such as this one) a good source of marks here.

Skip to Main Contents

Question: 1 2 3 4 5 6

Question 6 - Response B

a) ~~2.5 × 4.2 = 10.5~~

$2 \times 4.2 = 8.4$
$4 \times 4 = 16$
$4 \times 3.5 = 14$
$10 \times 1 = 10$

$8.4 + 16 + 14 + 10 = 48.4$

$\frac{48.4}{90} = 0.538$

b) No, as the graph does not follow a bell-shaped curve.

c) $\int_0^n xe^{-x} \, dx$

$u = x \qquad dv/dx = e^{-x}$
$du/dx = 1 \qquad v = -e^{-x}$

$I = -xe^{-x} - \int -e^{-x}$

$= -xe^{-x} + \int e^{-x}$

$= [-xe^{-x} - e^{-x}]_0^n$

$= (-ne^{-n} - e^{-n}) - (-1)$

$= -ne^{-n} - e^{-n} + 1$

$= e^{-n}(-n-1) + 1$

$= -e^{-n}(n+1) + 1$

$= 1 - e^{-n}(n+1)$

$= 1 - (n+1)e^{-n}$ as required.

Q6

A
B
C

d) ~~$1-(4+1)e^{-4} = 0.90842$~~

~~$1-(0+1)e^{\ } = 0.623212$~~

~~0.90842 -~~

~~$1-(40+1)e^{-x} = 0.74905\ldots$~~

~~1- (~~

e) i) $T \sim N(14.9, 9.3^2)$

$P(10<T<30) = 0.649$

ii) $y = 99xe^{-x}$ $\quad 0 \le x \le 4$

$(1-(3+1)e^{-3}) - (1-(1+1)e^{-1}) = 0.80085 - 0.264241\ldots = 0.537$

f) ~~$x = 1,\ y = 36$. This suggests 36 ... kind of exactly 10 hours which is unrealistic~~

It cannot be used to predict times of over 40 hours.

10/14 marks

Part (a)

This is fully correct (M1A1). The M1 would be awarded for sight of the correct products before they are added.

Part (b)

They have a comment relating to shape and have said the histogram does not follow a “bell shaped curve” which was acceptable on this occasion (B1)

Part (c)

This is fully correct (M1A1dM1A1)

Part (d)

They have an attempt at part (d) which is crossed out and not replaced so it is marked. They do not set the area = 90 and no progress is made (M0M0A0)

Part (e)

Part (i) is correct (B1) and for part (ii) they have used the result in (c) with n = 3 and n = 1 correctly as can be seen from the 0.537 (M1) but they do not go on to form a correct probability (A0)

Part (f)

They have a comment stating that the model cannot be used for times greater than 40 hours (B1)

Question 6 - Response C

Time in hours

Area

Freq.den. × width

a) $(2 \times 4.2) + (4 \times 4) + (4 \times 3.5) + (10 \times 1) = 48.4$

$\frac{48.4}{90} = \underline{0.5378}$

b) No, as the histogram does ~~...~~ not appear to follow a normal distribution.

c) $\int_0^n x \times e^{-x} \, dx = 1 - (n+1)e^{-n}$

$\int u \times \frac{dv}{dx} \, dx = uv - \int v \frac{du}{dx}$. $\quad u = x \quad \frac{dv}{dx} = e^{-x}$

$\frac{du}{dx} = 1 \quad v = -e^{-x}$

$= -xe^{-x} - \int -e^{-x} \, dx = -xe^{-x} - e^{-x}$

$= -(x+1)e^{-x}$

Q6

A

B

C

$$\left[-(x+1)e^{-x}\right]_0^n$$

$$= \left(-(n+1)e^{-n}\right) - \left(-(1)e^{0}\right)$$

$$= 1-(n+1)e^{-n}.$$

d) ~~[illegible]~~

$$90 = k\int_0^{4} xe^{-x}\,dx$$

$$90 = k\left(1-(5)e^{-4}\right)$$

$$90 = k \times 0.9084\ldots$$

$$k = 99.0729\ldots$$

$$k = 99.$$

e) i) $T \sim N(14.9, 9.3^2)$

$P(10 < T < 30) = \underline{0.6486}$

ii) $\cancel{99}\int_1^3 xe^{-x}\,dx$ ~~$= 99\int_0^3 xe^{-x}\,dx$~~

$= \left(\int_0^3 xe^{-x}\,dx - \int_0^1 xe^{-x}\,dx\right)$

$= \left((1-4e^{-3})-(1-2e^{-1})\right)$

$= 0.5366 \times \frac{99}{90} = \underline{0.5903}$

f) it is only suitable upto a 40 hour stay.

13 / 14 marks

Q6

?

A

B

C

Part (a)

This is fully correct (M1A1)

Part (b)

They have stated that Tomas' model is not suitable but not mentioned the idea of symmetry (B0)

Part (c)

They have made a correct start to integration by parts with a correct value for u and v seen (1^{st} M1) and the first step of integration and the second integral is correct (1^{st} A1) and in fact the full integration is correct too. On the next page all the integration is attempted and at least one correct limit used (2^{nd} dM1 this is dependent on 1^{st} M1) and they complete the solution without any incorrect working seen (2^{nd} A1 cso)

Part (d)

They have a correct statement setting the area = 90 (1^{st} M1) then apply the result from part (c) correctly (2^{nd} M1) and correctly show that k = awrt 99.1 (A1). This final mark can be given if $k = 99$ or awrt 99.1

Part (e)

Part (i) is correct (B1) and for part (ii) they have used the result in (c) with $n = 3$ and $n = 1$ correctly (M1) and then go on to find the correct probability (A1).

Part (f)

They have stated that the model is only valid for stays of up to 40 hours (B1)