

## **Pearson Edexcel**

### **Level 3 Advanced GCE in Statistics (9ST0)**

**Summer 2023 Exemplar**

**9ST0-01 A level Statistics**

**Paper 01: Data and Probability**

**Senior Examiner's feedback on student responses**

### **Contents**

<b>About this booklet .....</b>	<b>2</b>
<b>How to use this booklet .....</b>	<b>3</b>
<b>General Examiner Feedback .....</b>	<b>4</b>
<b>Question 1 .....</b>	<b>5</b>
<b>Question 2 .....</b>	<b>16</b>
<b>Question 3 .....</b>	<b>23</b>
<b>Question 4 .....</b>	<b>38</b>
<b>Question 5 .....</b>	<b>51</b>
<b>Question 6 .....</b>	<b>63</b>
<b>Question 7 .....</b>	<b>74</b>

**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](http://www.pearson.com/uk)



## About this booklet

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

This document looks at questions from the 9ST0-01 A level Statistics Paper 1: Data and Probability Summer 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.



# How to use this booklet

Navigate to the Contents

[Skip to main contents](#)

[Contents](#)

Question:

1 2 3 4 5 6

Navigate to a question

## Question 1

[Introduction](#) [Question](#) [Mark Scheme](#) [Examiner Comments](#)  
[Performance](#) [Response A](#) [Response B](#) [Response C](#)

### Question 1 - Introduction

This question tested the topics on sampling, distributions and probability. Assessment Objective 3 requires students to be able to select a suitable model and there were 2 marks targeting that skill here (one in part (b) and one in part (d)). The correct use of the notation is important here.

### Question 1 - Question

1. (a) State one disadvantage of using quota sampling compared with simple random sampling. (1)
- In a university 8% of students are members of the university dance club.  
 A random sample of 36 students is taken from the university.  
 The random variable  $X$  represents the number of these students who are members of the dance club.
- (b) Using a suitable model for  $X$ , find
- (i)  $P(X = 4)$   
 (ii)  $P(X \geq 7)$  (3)
- Only 40% of the university dance club members can dance the tango.
- (c) Find the probability that a student is a member of the university dance club and can dance the tango. (1)
- A random sample of 50 students is taken from the university.
- (d) Find the probability that fewer than 3 of these students are members of the university dance club and can dance the tango. (2)
- (Total for Question 1 is 7 marks)

Level 3 Advanced GCE in Mathematics - October 2021 Exemplar - 9MA0-31 Paper 31 Statistics - © Pearson Education Ltd 2022

4

Navigate to a specific part of this question

[Contents](#)

2 3 4 5 6

### 1 - Mark Scheme

Scheme	Marks	AO
Disadvantage: e.g. Not random; cannot use (reliably) for inferences	B1	1.1b
correct use of] $X \sim B(36, 0.08)$	M1	(1) 3.3
$P(X = 4) = 0.167387 \dots$ awrt <b>0.167</b>	A1	1.1b
$[P(X \geq 7) = 1 - P(X \leq 6) = ] 0.022233 \dots$ awrt <b>0.0222</b>	A1	1.1b
	(3)	1.1b
club and dance tango) = $0.4 \times 0.08 = \underline{0.032}$ or $\frac{4}{125}$ or	B1	1.1b
	(1)	3.3
those who can dance the Tango. Sight or use of]	M1	1.1b
"0.032") $T \sim B(50, \dots)$	A1	(2) 1.1b
$< 3) = P(T \leq 2) = ] 0.7850815 \dots$ awrt <b>0.785</b>		(7 marks)

### Notes

- (a) B1 for a suitable disadvantage:
- | Allow (B1)                           | Do NOT allow (B0)                 |
|--------------------------------------|-----------------------------------|
| Not random or less random (o.e.)     | Not representative                |
| Cannot use (reliably) for inferences | Less accurate                     |
| (More likely to be) biased           | Any comment based on time or cost |
|                                      | Any mention of skew               |
|                                      | Any mention of non-response       |
- (b) M1 for sight of  $B(36, 0.08)$  Allow in words: binomial with  $n = 36$  and  $p = 0.08$  may be implied by one correct answer to 2sf or sight of  $P(X \leq 6) = 0.97776 \dots$  i.e. awrt 0.98  
 Allow for  $36C4 \times 0.08^4 \times 0.92^{32}$  as this is "correct use"
- (i) 1<sup>st</sup> A1 for awrt 0.167 NB An answer of just awrt 0.167 scores M1(=) 1<sup>st</sup> A1  
 (ii) 2<sup>nd</sup> A1 for awrt 0.0222
- (c) B1 for 0.032 o.e. (Can allow for sight of  $0.4 \times 0.08$ )
- (d) M1 for sight of  $B(50, "0.032")$  ft their answer to (c) provided it is a probability  $\neq 0.08$  may be implied by correct answer  
 or sight of  $[P(T \leq 3)] = 0.924348 \dots$  i.e. awrt 0.924 or  $P(T \leq 2)$  as part of  $1 - P(T \leq 2)$  calc.  
 A1 for awrt 0.785  
 MR Allow MR of 50 (e.g. 30) provided clearly attempting  $P(T \leq 2)$  and score M1A0

Level 3 Advanced GCE in Mathematics - October 2021 Exemplar - 9MA0-31 Paper 31 Statistics - © Pearson Education Ltd 2022

5



## General Examiner Feedback

This paper was accessible to all candidates.

Based on their performance on this paper, candidates should be advised to:

- give explanation answers within the context of the question
- show full working on questions requiring calculations
- use bullet points when answering questions requesting multiple reasons
- ensure they are answering the exact question asked
- attempt the whole question, even if they have not managed earlier parts

Question:

1

2

3

4

5

6

7

## Question 1

 Question

 Mark Scheme

 Examiner Comments

 Performance

 Response A

 Response B

 Response C



### Question 1 - Question

- 1 Yin is the manager of a busy fast-food restaurant. He wants to get customer feedback on the quality of the service at the restaurant.

Three different sampling methods are considered by Yin.

#### Method 1

A staff member stands at the main exit to survey customers as they leave the restaurant.

- (a) For **Method 1**, give

- (i) **one** advantage, (1)
- (ii) **one** disadvantage. (1)

#### Method 2

An internet link to the survey is printed on the back of customer receipts.

- (b) For **Method 2**, give

- (i) **one** advantage, (1)
- (ii) **one** disadvantage. (1)

#### Method 3

Customers are asked for their email address when they pay their bill, and a survey is emailed to them afterwards.

- (c) For **Method 3**, give

- (i) **one** advantage, (1)
- (ii) **one** disadvantage. (1)

Yin would like the survey to be ongoing. He wants to minimise expense to himself and inconvenience to customers.

- (d) Select which one of the three methods, 1, 2 or 3, you would recommend for Yin to use.

Briefly explain your selection. (1)

Yin would like the response rate to this survey to be as high as possible.

- (e) Make **two** suggestions for Yin to improve the response rate of his survey. (2)

(Total for Question 1 is 9 marks)



Question:

1

2

3

4

5

6

7

**Question 1 - Mark Scheme**

Question	Scheme	Marks	AO	Notes
1(a)(i)	<b>Advantages (not exhaustive)</b>  The response rate is likely to be good.  Customers will be able to feed back as soon as the meal is finished.	E1	3.1a	Quick feedback  Any sensible advantage
1(a)(ii)	<b>Disadvantages (not exhaustive)</b>  It will be expensive to pay a staff member to be there permanently.  The restaurant may not have the layout to station someone at the exit.  Only asks customers using the main exit  Customers may not like being accosted on the way out.  Customers may overhear and repeat each other's answers	E1	3.1a	Any sensible disadvantage
1(b)(i)	<b>Advantages (not exhaustive)</b>  They are contacting everyone.  There will be minimal ongoing costs after the initial start-up.  Provides online data in an easy to use form	E1	3.1a	or relatively cheap  Any sensible advantage

Q1



A

B

C



Question:

1

2

3

4

5

6

7

Question	Scheme	Marks	AO	Notes
1(b)(ii)	<b>Disadvantages (not exhaustive)</b>  The response rate is likely to be low  Lots of customers do not take receipts away.  Lots of customers do not look at the back of receipts.  Customers may not have access to the internet  Only the person who pays will have the receipt and therefore the survey	E1	3.1a	oe          Any sensible disadvantage
1(c)(i)	<b>Advantages (not exhaustive)</b>  The company may be able to contact the customers in the future for feedback or marketing purposes.  Quick to get the survey out.  Very cheap.  Easy to set up.  Provides online data in an easy to use form	E1	3.1a	Any sensible advantage
1(c)(ii)	<b>Disadvantages (not exhaustive)</b>  Lots of customers may not want to give their email addresses.  Asking for email addresses may make customers unhappy.  There may be data protection issues.  It will take longer to take orders in the restaurant.  Only the person who pays will have the receipt and therefore the survey	E1	3.1a	or GDPR oe          Any sensible disadvantage

Q1



A

B

C



Question:

1

2

3

4

5

6

7

Question	Scheme	Marks	AO	Notes
1(d)	<p>Method 2, as it is both cheap and inconvenient to customers as you aren't asked for your email address</p> <p>Method 3, is cheap, and not inconvenient to customers as you can't lose your receipt</p>	E1	2.1b	Method 2 or 3 must be chosen as these minimise expense, and a comment must also be made about inconvenience
1(e)	<p><b>Possible suggestions (not exhaustive)</b></p> <p>Offer some free food if people take the survey.</p> <p>Enter survey-takers into a prize draw.</p> <p>Make sure the survey is clear and concise.</p> <p>Tell customers that the survey will only take 2 minutes.</p> <p>Send survey multiple times</p>	E1, E1	3.1a, 3.1a	<p>E1 for each sensible suggestion (max E2)</p> <p><b>Not</b> dep on (d)</p>
Total		9		

Q1



A

B

C

Question:

1

2

3

4

5

6

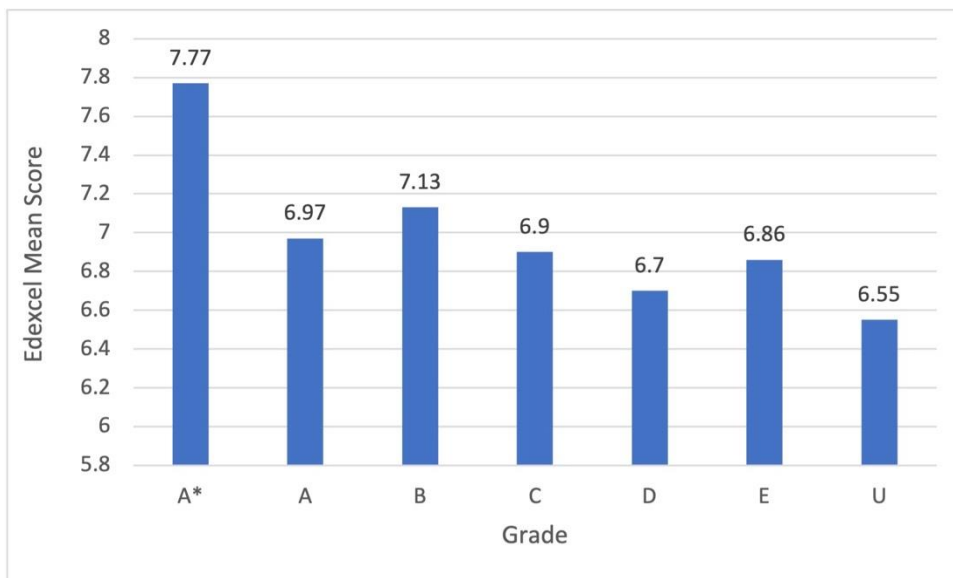
7

## Question 1 - Examiner Comments

This question served as a straightforward introduction to the paper, with at least one of the three methods accessible to all candidates. Part (d) required candidates to refer to both minimising expense and minimising inconvenience to customers. Candidates generally lost marks for either only referring to one of these or failing to realise there was a cost associated to a staff member waiting by a door. For questions such as (e) it is important to make sure that candidates are answering the exact question asked. For example, in this case a large sample of people to complete the survey would result in a large sample and more responses, but this does not impact response rate.

## Question 1 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade:							
			ALL	A*	A	B	C	D	E	U
6.90	9	77	6.90	7.77	6.97	7.13	6.90	6.70	6.86	6.55



**\*Note:** As the data presented is real candidate performance data the averages will not necessarily follow a linear pattern.

Q1

?

✓

≡

Bar chart icon

✍

A

B

C

 **Question 1 - Response A**

(a) For **Method 1**, give  
(i) **one** advantage, (1)

*Convenient*

(ii) **one** disadvantage. (1)

*Not random ; potentially biased.*

(b) For **Method 2**, give  
(i) **one** advantage, (1)

*Convenient*

(ii) **one** disadvantage. (1)

*Opt in, chance people won't respond*

(c) For **Method 3**, give  
(i) **one** advantage, (1)

*Convenient*

(ii) **one** disadvantage. (1)


*May not respond*


Q1

?

✓

≡





A

B

C

Question:

1

2

3

4

5

6

7

(d) Select which one of the three methods, 1, 2 or 3, you would recommend for Yin to use.

Briefly explain your selection.

(1)

Method 3 would be better as it ensures customers are aware of the survey, meaning it may be more likely for them to respond.

(e) Make two suggestions for Yin to improve the response rate of his survey.

(2)

Add an incentive to completing the survey such as an offer or prize draw.  
Send the survey multiple times, e.g. 2, to ensure they have received it in case they forget to respond.

4 / 9

Q1



A

B

C

**Part (a)**

**E0:** This answer is too vague to qualify as an advantage. There is no contextual explanation of why this would be convenient.

**E1:** This is a minimal acceptable explanation for a disadvantage.

**Part (b)**

**E0:** As above, this answer is too vague to qualify as an advantage. There is no contextual explanation of why this would be convenient.

**E1:** Another minimal acceptable explanation, specific to this method.

**Part (c)**

**E0:** As above, this answer is too vague to qualify as an advantage. There is no contextual explanation of why this would be convenient.

**E0:** No contextual explanation as to why they won't respond, e.g. not checking email or spam filter etc.

**Part (d)**

**E0:** While it is true that method 3 makes the customers aware of the survey this does not answer the question. Yin wishes to minimise expense and inconvenience, and neither is referred to.

**Part (e)**

**E1:** Reasonable suggestion of incentive.

**E1:** Reasonable suggestion of multiple emails as reminder.

Question:

1

2

3

4

5

6

7



## Question 1 - Response B

(a) For **Method 1**, give

(i) **one** advantage,

(1)

Quick to carry out

(ii) **one** disadvantage.

(1)

may be bias as people usually go as a family to restaurants

(b) For **Method 2**, give

(i) **one** advantage,

(1)

allows for a larger sample to be obtained

(ii) **one** disadvantage.

(1)

sample obtained may not be representative as not everyone who visits the restaurants will have a mobile device to access one with/may not see it.

(c) For **Method 3**, give

(i) **one** advantage,

(1)

cheap to carry out

(ii) **one** disadvantage.

(1)

may not receive or they may ignore the email sent to them.

Q1

?

✓

≡

▮

▮

A

B

C

Question:

1

2

3

4

5

6

7

(d) Select which one of the three methods, 1, 2 or 3, you would recommend for Yin to use.

Briefly explain your selection.

Method 1 as its cheap ~~and easy~~ <sup>and easy</sup> (1)  
to do and <sup>Yin</sup> can obtain his  
sample much quicker this way.

(e) Make two suggestions for Yin to improve the response rate of his survey.

Simple, understandable questions to be (2)  
used in his survey.  
Take no longer than 5 mins to carry  
out as people have busy days  
and will get bored quickly.

7 / 9

#### Part (a)

**E1:** This mark is awarded as the feedback is instant which makes the whole process of surveying quicker.

**E1:** This is a good contextual disadvantage.

#### Part (b)

**E0:** A large sample may be obtained with any method, so this is not an advantage of this method.

**E1:** Good contextual explanation for disadvantage.

#### Part (c)

**E1:** This is an advantage over method 1 so is acceptable.

**E1:** Good contextual explanation for disadvantage.

#### Part (d)

**E0:** Method 1 would not be cheap, so the student is incorrect about cost, and they do not address inconvenience.

#### Part (e)

**E1:** Reasonable suggestion of simple questions

**E1:** Reasonable suggestion of survey being short. This is very similar to the previous suggestion, but just different enough.

Q1



A

B

C



Question:

1

2

3

4

5

6

7



## Question 1 - Response C

(a) For **Method 1**, give(i) **one** advantage,

(1)

Easy to administer.

(ii) **one** disadvantage.

(1)

Answers may be rushed as they want to leave.

(b) For **Method 2**, give(i) **one** advantage,

(1)

Gives his customers freedom to answer the survey whenever they want.

(ii) **one** disadvantage.

(1)

Not all customers will want to go on the survey as they might be unaware of what the link means.  
throw away the receipt receipts(c) For **Method 3**, give(i) **one** advantage,

(1)

~~Can target their survey~~ Easier to send survey to a lot of people at once instead of individually so it is time-efficient.(ii) **one** disadvantage.

(1)

Not everyone will give out their email address as it might be a scam.

Q1



A

B

C



Question:

1

2

3

4

5

6

7

(d) Select which one of the three methods, 1, 2 or 3, you would recommend for Yin to use.

Briefly explain your selection.

(1)

Method 3 - This is because email's are free to send to his customers and there is no rush for the customers to complete the survey straight away.

(e) Make two suggestions for Yin to improve the response rate of his survey.

(2)

- He could add an incentive for those customers who completed the survey such as discounts for their next meal.

- He could make physical copies of the surveys and leave them on the customer's tables to fill out whilst waiting for their food.

(Total for Question 1 is 9 marks)

9 /9

### Part (a)

E1: This answer is sufficient when only one method has been provided.

E1: Good explanation for disadvantage in context.

### Part (b)

E1: An advantage which applies specifically to this method.

E1: A disadvantage which applies specifically to this method.

### Part (c)

E1: An advantage which applies specifically to this method.

E1: A disadvantage which applies specifically to this method.

### Part (d)

E1: The answer refers directly to cost and indirectly to inconvenience.

### Part (e)

E1: Reasonable suggestion.

E1: Reasonable suggestion.

Q1



A

B

C

Question:

1

2

3

4

5

6

7

## Question 2

? Question

✓ Mark Scheme

≡ Examiner Comments

Performance

Response A

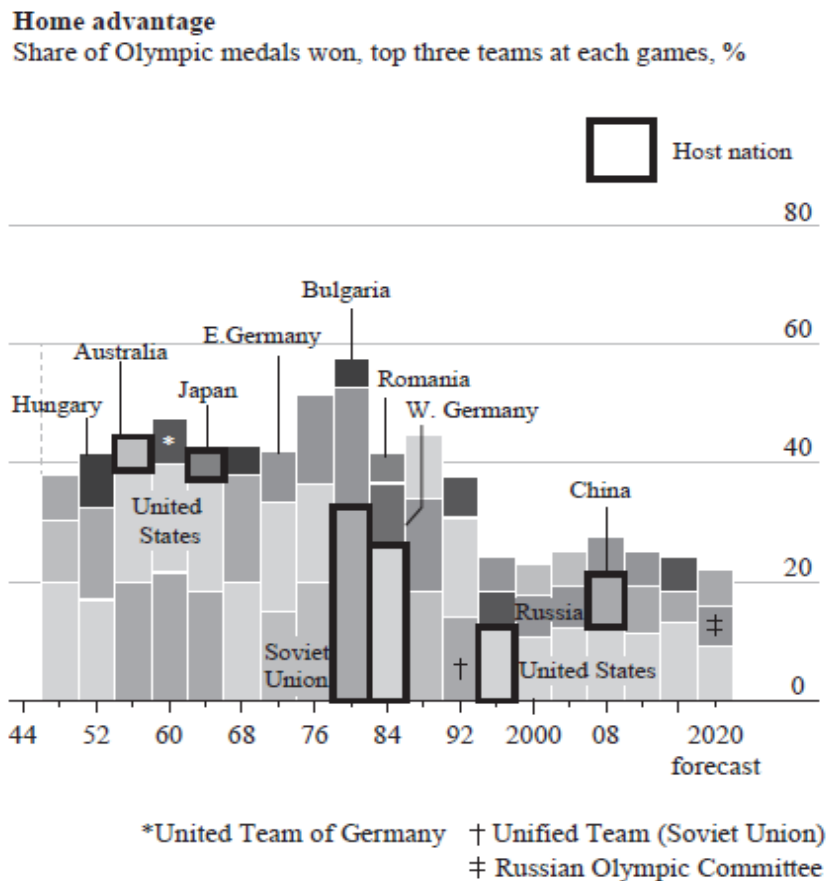
Response B

Response C

### ? Question 2 - Question

- 2 A graph was shown in an article about Summer Olympic teams in *The Economist* in July 2021

A section of the graph is shown in **Figure 1**.



**Figure 1**

- (a) Estimate the highest proportion of medals won by a single nation in one Summer Olympic Games.

(1)

- (b) State how many Summer Olympic Games saw the host nation score **more** medals than any other team.

(1)

Question:

1

2

3

4

5

6

7



## Question 2 – Question (Cont.)

In the original article, the graph in **Figure 1** was printed in colour.

(c) Explain how colour could be used to make the graph clearer and easier to understand.

(1)

(d) State **four** criticisms, other than the lack of colour, of the graph.

(4)

(Total for Question 2 is 7 marks)



## Question 2 - Mark Scheme

Question	Scheme	Marks	AO	Notes
2(a)	33%	B1	1.1	awfw 30~35%
2(b)	3	B1	1.1	
2(c)	The <b>nations</b> could be colour-coded. A key could be added	E1	1.1	Any sensible suggestion
2(d)	<b>Possible criticisms (not exhaustive)</b>  There is no key/legend (rather than having nation labels on the graph).  No consistent labelling on time (x) axis (e.g., every other Olympics).  No label on (x) axis.  No label on (y) axis  There is no line for the proportion (y) axis.  No (y) axis on the left side.  Use of †‡* symbols with no explanation  No source			

Q2



A

B

C

Question:

1

2

3

4

5

6

7

Question	Scheme	Marks	AO	Notes
	y-axis goes	E1, E1, E1, E1	3.1a, 3.1a, 3.1a, 3.1a	E1 for each sensible criticism (max E4)
		<b>Total</b>	<b>7</b>	

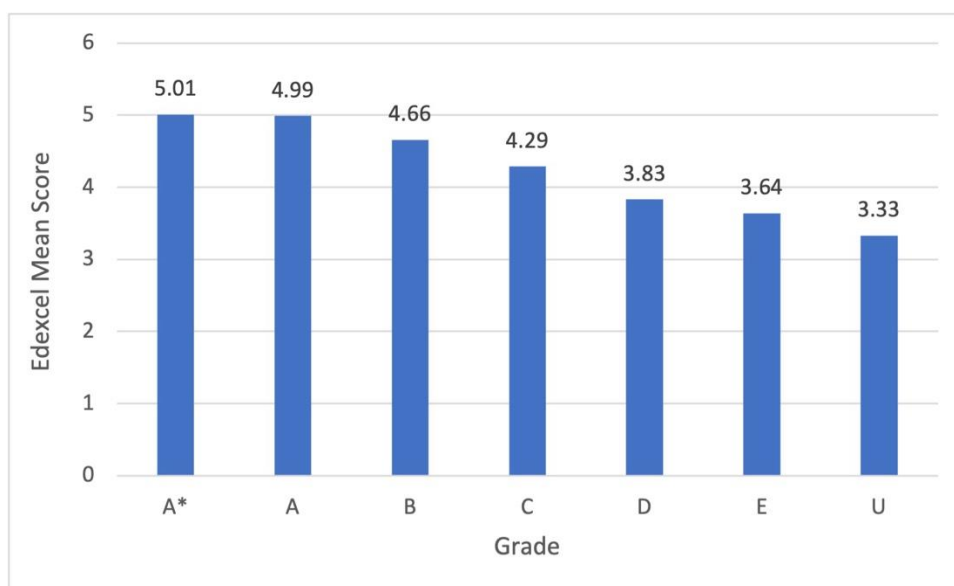
## Question 2 - Examiner Comments

This question was well attempted by candidates, and in general they were able to perform well. In (c) many candidates did not explain how colour would make the graph easier to understand, such as by giving each country a different colour, and merely restated that it would make the graph easier to understand.

(d) was well accessed with candidates able to give a range of criticisms. Candidates are reminded that in questions where they are required to give a number of criticisms, such as (d), they are advised to use bullet points, and to be careful not to repeat the same point with different wording.

## Question 2 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade:							
			ALL	A*	A	B	C	D	E	U
4.19	7	60	4.19	5.01	4.99	4.66	4.29	3.83	3.64	3.33



Q2



A  
B  
C



Question:

1

2

3

4

5

6

7



## Question 2 - Response A

- (a) Estimate the highest proportion of medals won by a single nation in one Summer Olympic Games.

(1)

36.1

- (b) State how many Summer Olympic Games saw the host nation score **more** medals than any other team.

(1)

2

- (c) Explain how colour could be used to make the graph clearer and easier to understand.

(1)

it would make each bar stand out more and easier to tell the difference

- (d) State **four** criticisms, other than the lack of colour, of the graph.

(4)

The x-axis is not consistent as it ~~jumps~~ jumps from 2008 - 2020.

There is no title

not all bars are labeled and the labels are confusing

Can't get accurate percent from the y axis

Q2



A

B

C

3 / 7



Question:

1

2

3

4

5

6

7

**Part (a)**

**B0:** Not within range of 30-35

**Part (b)**

**B0:** Incorrect value.

**Part (c)**

**E0:** No explanation of how colour makes it easier to tell the difference.

**Part (d)**

**E1:** Jump from 2008-2020 is a different size to others.

**E1:** Lack of labels on bars (this is the third comment the candidate makes).

**E1:** Difficult to read the y-axis (this is the fourth comment the candidate makes).

**E0:** There is a title, but the candidate states there is not.

Q2



A

B

C

Question:

1

2

3

4

5

6

7



## Question 2 - Response B

(a) Estimate the highest proportion of medals won by a single nation in one Summer Olympic Games.

(1)

(b) State how many Summer Olympic Games saw the host nation score **more** medals than any other team.

(1)

(c) Explain how colour could be used to make the graph clearer and easier to understand.

(1)

Each nation could have their own specific ~~own~~ colour so it's easier to tell the difference

(d) State **four** criticisms, other than the lack of colour, of the graph.

(4)

The x and y axes need to be labelled

Each nation should have a key to tell what colour matches it

4 / 7

**Part (a)**

**B0:** No answer given.

**Part (b)**

**B0:** No answer given.

**Part (c)**

**E1:** Explains how colour is used by each nation having their own colour.

**Part (d)**

**E1:** x-axis not labelled.

**E1:** y-axis not labelled.

**E1:** Lack of key.

**E0:** No fourth point made.

Q2

?

✓

≡

Bar chart icon

A

B

C

Question:

1

2

3

4

5

6

7



## Question 2 - Response C

- (a) Estimate the highest proportion of medals won by a single nation in one Summer Olympic Games.

(1)

~~20~~ <sup>30</sup> medals

- (b) State how many Summer Olympic Games saw the host nation score **more** medals than any other team.

(1)

2

- (c) Explain how colour could be used to make the graph clearer and easier to understand.

(1)

It would be ~~or~~ easier to see the different medals and therefore would be clearer on total number of medals won.

- (d) State **four** criticisms, other than the lack of colour, of the graph.

(4)

- The x and y axis are not labelled.
- No Title for the graph
- The time dates don't go up the same at the end
- No key.

5 / 7

### Part (a)

**B1:** Falls between 30-35 inclusive.

### Part (b)

**B0:** Incorrect value.

### Part (c)

**E0:** Simply repeats the question; doesn't explain how colour helps to distinguish.

### Part (d)

**E1:** x-axis not labelled.

**E1:** y-axis not labelled.

**E1:** Gaps in dates are irregular.

**E1:** No key.

Note there is no penalty here for the incorrect "No title" as the candidate has not been hedging their bets by e.g., giving more answers than required, they simply didn't realise their first response was worth two marks.

Q2

?

✓

≡

Bar chart icon



A

B

C

Question:

1

2

3

4

5

6

7

## Question 3

 Question

 Mark Scheme

 Examiner Comments

 Performance

 Response A

 Response B

 Response C

### Question 3 - Question

- 3 Many websites allow various filters, such as product features, to be applied to searches by online shoppers.

In November 2021, the Currys website had 400 models of television for sale.

Three different filters for televisions were considered. They are described below.

**8K resolution (R)** A very high-resolution screen with many pixels.  
This technology is fairly new and relatively expensive.

**Screen size  $\geq 75''$  (S)** The screen size is at least 75 inches.  
Generally a larger screen is more expensive.

**Voice control (V)** The television can be controlled using voice commands.  
This technology is fairly new.

The three filters were applied singly and in combination with other filters. The number of models of television found on the website after these filters were applied can be seen in **Figure 2**.

Filter applied to			Models of television
8K resolution (R)	Screen size $\geq 75''$ (S)	Voice control (V)	
✓			16
	✓		74
		✓	304
✓	✓		9
✓		✓	16
	✓	✓	69
✓	✓	✓	9

(Date source: [www.currys.co.uk](http://www.currys.co.uk), accessed 25/11/2021)

**Figure 2**

Question:

1

2

3

4

5

6

7

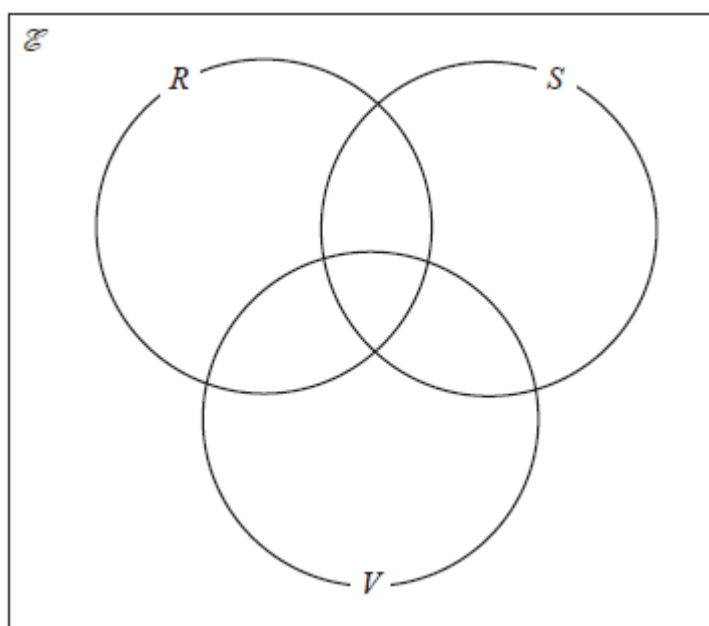


### Question 3 – Question (Cont.)

- (a) Use the data in **Figure 2** to complete the Venn diagram frequencies below.

You are reminded that there are 400 models in total.

(6)



One model of television is selected at random from those listed on the website.

- (b) Find the probability that the television
- has 8K resolution
  - has a screen size **smaller** than 75 inches and voice control
  - will not be displayed if any of the filters are applied
  - has 8K resolution, given that the screen size is at least 75 inches.

(1)

(2)

(1)

(2)

Three events are defined as follows:

- $R$       The television has 8K resolution  
 $S$       The television has a screen size of at least 75 inches  
 $V$       The television has voice control

- (c) Show that  $R$  and  $S$  are **not** statistically independent.

You should provide numerical justification.

(2)

- (d) Suggest a reason, in context, why  $R$  and  $S$  are **not** likely to be independent.

(1)

(Total for Question 3 is 15 marks)

Q3



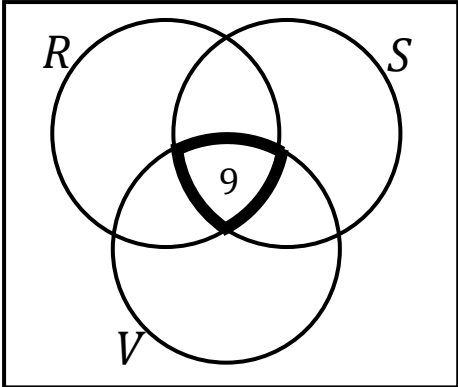
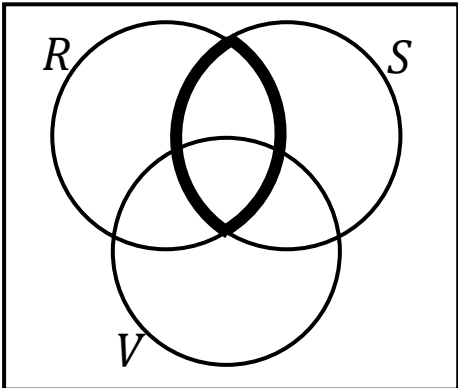
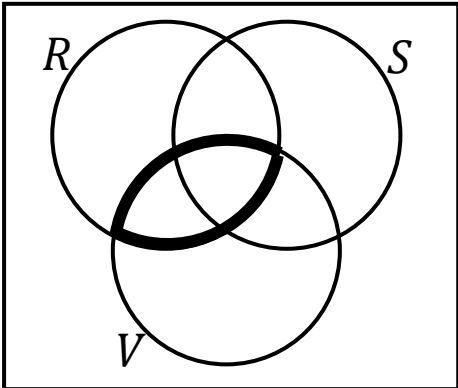
A

B

C



Question 3 - Mark Scheme


Question	Scheme	Marks	AO	Notes
3(a)		B1	1.1	cao
		M1	1.1	Cells add to 9
				or Cells add to 16


Q3

?

✓

☰

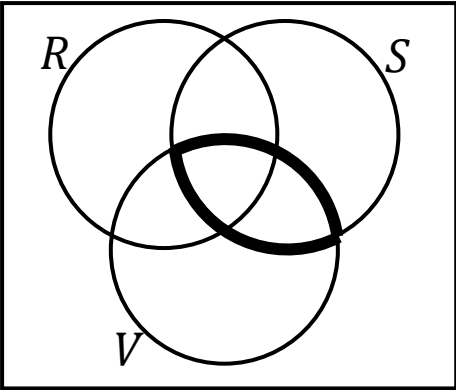
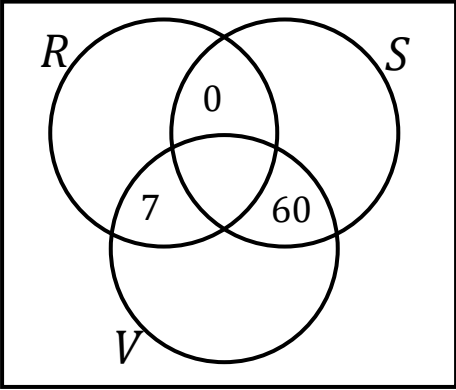
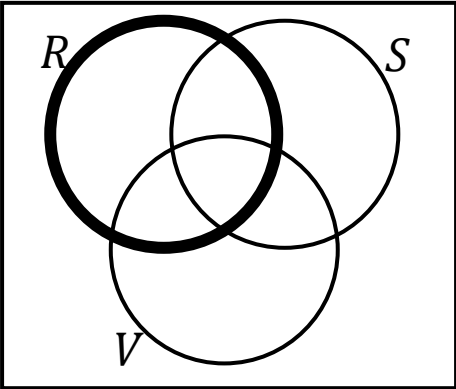
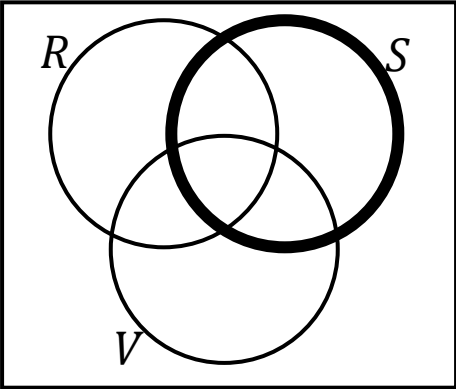




A

B

C


Question	Scheme	Marks	AO	Notes
				<b>or</b> Cells add to 69
		A1	1.1	All three correct
		M1	1.1	Cells add to 16
				<b>or</b> Cells add to 74


Q3

?

✓

≡

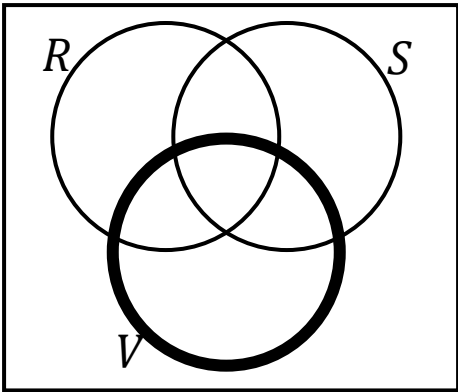
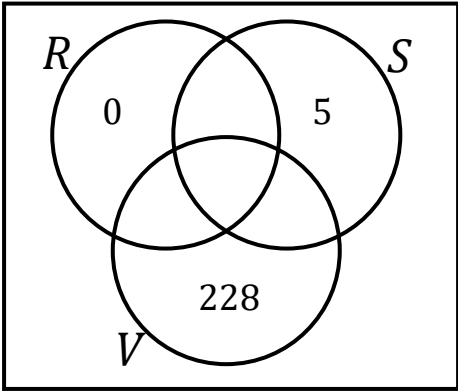
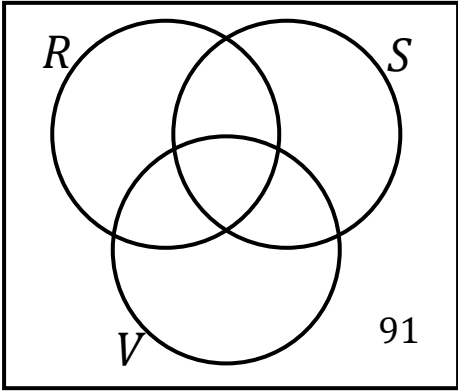
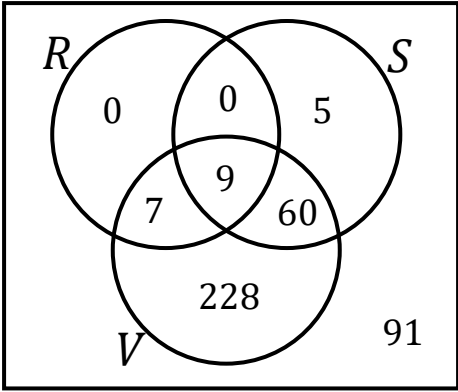




A

B

C

Question	Scheme	Marks	AO	Notes
				<p>or</p> <p>Cells add to 304</p>
		A1	1.1	All three correct
		B1	1.1	cao
				<p>Fully correct diagram</p> <p>[scores full marks]</p>

Q3

?

✓

≡

Bar chart icon

Document icon

A

B

C



Question:

1

2

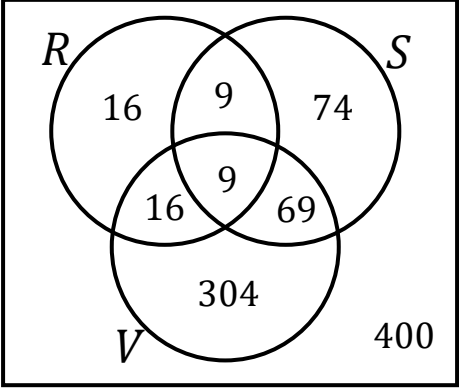
3

4

5

6

7

Question	Scheme	Marks	AO	Notes
	<b>SC</b> 			Scores max B1M1A0M1A0B0 Remove 1 mark for each small slip
3(b)(i)	$P(R) = \frac{16}{400} = \frac{1}{25} = 0.04$	B1	1.2	Any of these, oe <b>Condone SC</b> $\frac{50}{400} = \frac{1}{8} = 0.125$ <b>or</b> $\frac{50}{897} = 0.0557$
3(b)(ii)	$P(S' \cap V)$ <hr/> $= \frac{7 + 228}{400}$	M1	1.2	PI $P(S' \cap V)$ <b>or</b> correct numerator seen <b>Condone SC</b> $16 + 304$
	$= \frac{235}{400} = \frac{47}{80} = 0.588 \text{ (3sf)}$	A1	1.2	awfw 0.587~0.588 <b>Condone SC</b> $\frac{320}{400} = \frac{4}{5} = 0.8$ <b>or</b> $\frac{320}{897} = 0.357$
3(b)(iii)	$P(R' \cap S' \cap V')$ $= \frac{91}{400} = 0.228 \text{ (3sf)}$	B1	1.2	awfw 0.227~0.228 <b>Condone SC</b> $\frac{400}{897} = 0.446$
3(b)(iv)	$P(R S)$			
	$= \frac{9}{74} = 0.122 \text{ (3sf)}$	M1	1.2	PI Numerator or denominator correct or correct use of $P(R S) = \frac{P(R \cap S)}{P(S)}$

Q3


 A  
 B  
 C



Question:

1

2

3

4

5

6

7

		A1	1.2	Fully correct <b>Condone SC</b> $\frac{18}{161} = 0.112$
3(c)	$P(R) = 0.04$ $P(R S) = 0.122$	M1	2.1b	PI oe $P(R)$ and $P(R S)$ considered
	$P(R) \neq P(R S)$ so $R$ and $S$ are not statistically independent.	A1	2.1b	Correct comparison and conclusion.
	<b>Alternative</b>			
	$P(S) = \frac{74}{400} = \frac{37}{200} = 0.185$	(M1)		PI Attempt at finding $P(R \cap S)$ and $P(R) \times P(S)$
	$P(R \cap S) = \frac{9}{400} = 0.0225$			
	$P(R) \times P(S) = \frac{37}{5000} = 0.0074$			
	$P(R) \times P(S) \neq P(R \cap S)$ so $R$ and $S$ are not statistically independent.	(A1ft)		Correct comparison and conclusion. Dep on correct calculations in M1 ft their (a) and (b)(i)
3(d)	<b>Possible reasons (not exhaustive)</b>			
	Larger screens have more space for pixels.			
	More pixels need a larger screen to fit.			
	More expensive televisions are more likely to have bigger screens and higher resolution.			
		E1	2.1a	Any sensible reason linking higher resolution and larger screen
Total			15	

Q3



A

B

C

Question:

1

2

3

4

5

6

7

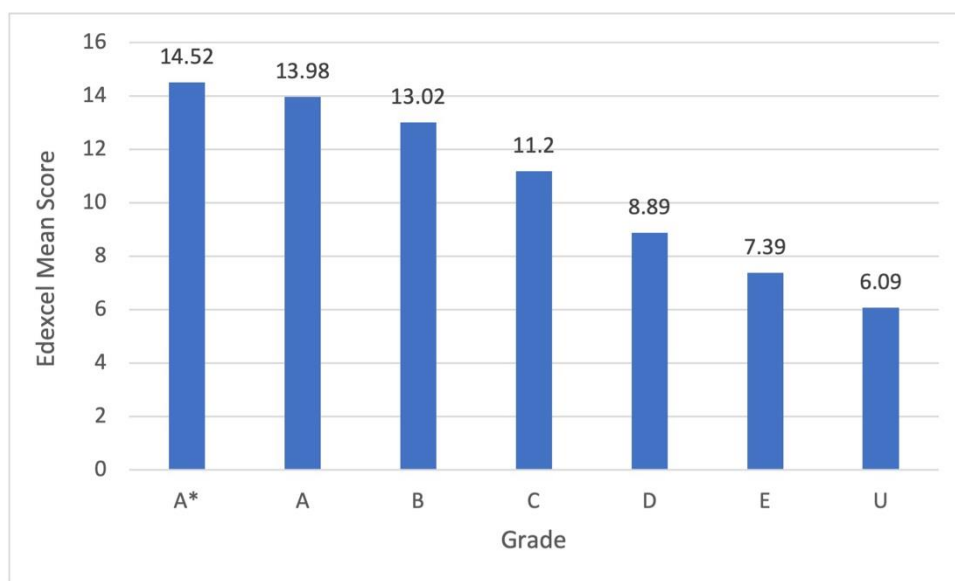
### Question 3 - Examiner Comments

This was a relatively standard Venn diagram question. Candidates are reminded to show full working. For example, if a candidate has filled out the diagram incorrectly in (a) they still received method marks in (c) for showing full working including the use of formula, but not for simply writing down an answer with the values from their diagram.

Parts (d) and (e) were answered well, with candidates who lost marks in (d) generally doing so for not writing a conclusion referring to the independence of the events.

### Question 3 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade:							
			ALL	A*	A	B	C	D	E	U
10.49	15	70	10.49	14.52	13.98	13.02	11.20	8.89	7.39	6.09



Q3



A

B

C

Question:

1

2

3

4

5

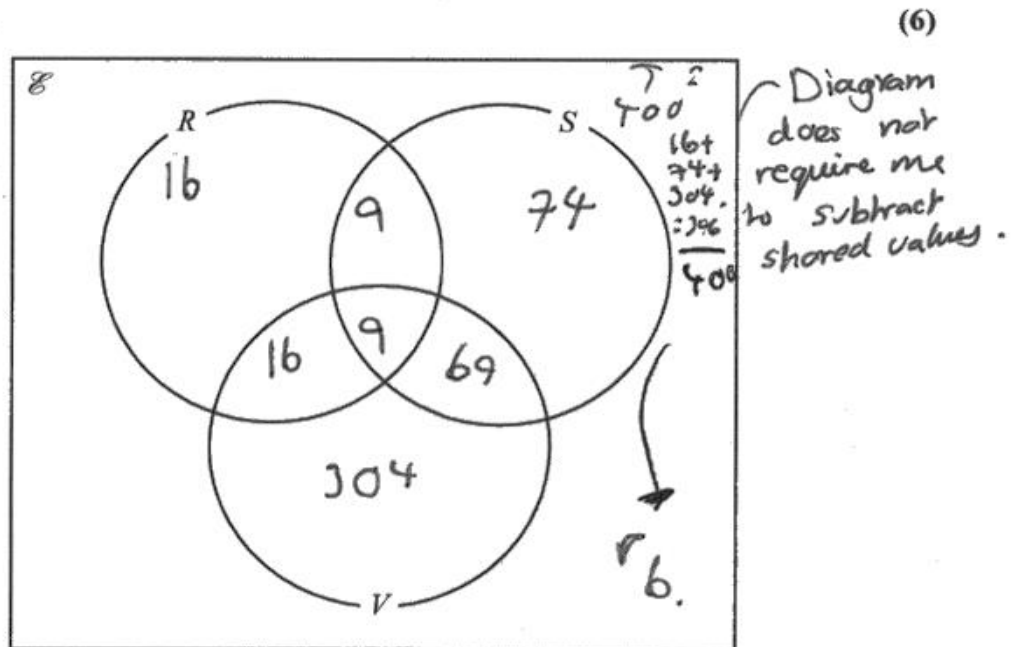
6

7

### Question 3 - Response A

(a) Use the data in **Figure 2** to complete the Venn diagram frequencies below.

You are reminded that there are 400 models in total.



(b) Find the probability that the television

(i) has 8K resolution

(1)

$$\frac{16}{400} = 0.04.$$

(ii) has a screen size **smaller** than 75 inches and voice control

(2)

$$\frac{326}{400} \times \frac{304}{400} = 0.6194$$

(iii) will not be displayed if any of the filters are applied

(1)

$$\frac{6}{400} = 0.015.$$

Q3

?

✓

≡

Bar chart icon

✎

A

B

C

Question:

1

2

3

4

5

6

7

(iv) has 8K resolution, given that the screen size is at least 75 inches.

(2)

$$\frac{16}{400} \times \frac{74}{400} = 0.04$$

$$\frac{74}{400}$$

$$\frac{9}{14400} = 0.000625$$

$$\frac{74}{14400} = 0.005139$$

$$0.005139 - 0.000625 = 0.004514$$

Three events are defined as follows:

R The television has 8K resolution

(c) Show that R and S are **not** statistically independent.

You should provide numerical justification.

(2)

$$r = \frac{16}{400}$$

$$\frac{9}{90} = P(R \cap S)$$

$$S = \frac{74}{400}$$

$$P(R \cap S) = \frac{9}{400}$$

Over 30% of TVs with 8K are 5 TVs. This infers a correlation between the two features. This is easily explainable as both are stated by the source to be "expensive features" and so a person with enough money to buy a tv with one feature would surely go for both of them in a product.

(d) Suggest a reason, in context, why R and S are **not** likely to be independent.

(1)

(Total for Question 3 is 15 marks)

6 /15

Part (a)

B1: Awarded as special case.

M1: Awarded as special case.

A0: Not awarded in special case.

M0: Not awarded for special case as slip on value of 400.

A0: Not awarded in special case.

B0: Not awarded in special case.



Question:

1

2

3

4

5

6

7

**Part (b)**

**B1:** Correct value.

**M0:** Incorrect value and not correct for values on their diagram.

**A0:** Incorrect value.

**B0:** Incorrect value.

**M1:** Correct answer implies correct method, and  $\frac{74}{400}$  seen.

**A1:** Correct answer.

**Part (c)**

**M0:** While  $P(R)$ ,  $P(S)$  and  $P(R \cap S)$  are considered,  $P(R)$  and  $P(S)$  are not multiplied.

**A0:** No comparison and conclusion.

**Part (d)**

**E1:** Contextual reason of expensive features given.

Q3



Question:

1

2

3

4

5

6

7

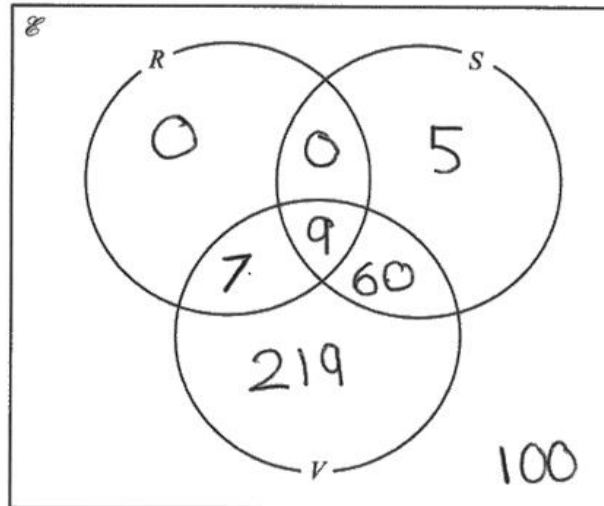


### Question 3 - Response B

(a) Use the data in **Figure 2** to complete the Venn diagram frequencies below.

You are reminded that there are 400 models in total.

$R = 8K$   
 $S = 75\text{ inches}$   
 $V = \text{voice}$



$T = 400$  (6)

$R = 16$   
 $S = 74$   
 $V = 304$

$219 + 9 + 7 + 60 + 5 = 300$

(b) Find the probability that the television

(i) has 8K resolution

$$P(R) = \frac{16}{400} = 0.04 \quad (1)$$

(ii) has a screen size **smaller** than 75 inches and voice control

$$P(V \cap S') = \frac{226}{400} = 0.565 \quad (2)$$

(iii) will not be displayed if any of the filters are applied

$$P(V' \cap S' \cap R') = \frac{100}{400} = 0.25 \quad (1)$$

(iv) has 8K resolution, given that the screen size is at least 75 inches.

$$P(R|S) = \frac{P(R \cap S)}{P(S)} = \frac{0}{74} = 0 \quad (2)$$

Q3

?

✓

≡

Bar chart icon

Pencil icon

A

B

C

Question:

1

2

3

4

5

6

7

Three events are defined as follows:

$R$  The television has 8K resolution

$S$  The television has a screen size of at least 75 inches

$V$  The television has voice control

(c) Show that  $R$  and  $S$  are **not** statistically independent.

You should provide numerical justification.

$$P(R \cap S) = \frac{0}{400} \quad P(R) = \frac{16}{400} \times P(S) \frac{74}{400} \quad (2)$$

$$0 \neq 0.0074$$

(d) Suggest a reason, in context, why  $R$  and  $S$  are **not** likely to be independent.

Because they're both expensive

9 /15

**Part (a)**

**B1:** Central cell is 9

**M1A1:** All three values of 0, 7 & 60 correct.

**M1:** Totals for  $R$  and  $S$  are correct (but not  $V$ )

**A0:**  $V$  incorrect.

**B0:** 91 incorrect.

**Part (b)**

**B1:** Correct value.

**M1:** Correct value from their diagram (219+7)

**A0:** Incorrect answer, and  $P(S)$  is not stated as  $\frac{74}{400}$

**B0:** Incorrect value.

**M1:** Correct answer implies correct method, and  $\frac{74}{400}$  seen.

**A1:** Correct answer.

**Part (c)**

**M1:** While  $(R \cap S)$  is calculated incorrectly,  $P(R)$  and  $P(S)$  are multiplied and compared to this.

**A0:** No conclusion.

**Part (d)**

**E1:** Contextual reason of expensive features given.

Q3

?

✓

≡

Bar chart icon

✎

A

B

C

Question:

1

2

3

4

5

6

7

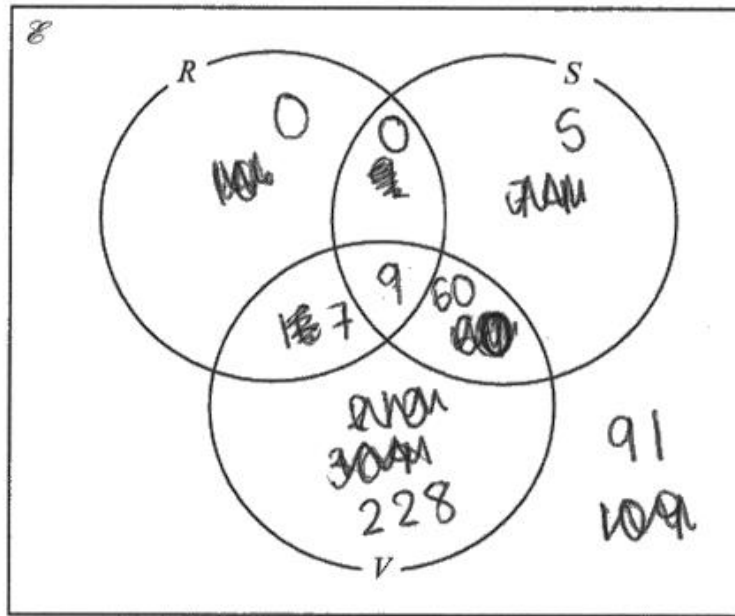


### Question 3 - Response C

(a) Use the data in **Figure 2** to complete the Venn diagram frequencies below.

You are reminded that there are 400 models in total.

(6)



(b) Find the probability that the television

(i) has 8K resolution

(1)

$$\frac{16}{400} = 0.04$$

(ii) has a screen size **smaller** than 75 inches and voice control

(2)

$$P(S' \cap V) = \frac{7}{400} = 0.0175$$

(iii) will not be displayed if any of the filters are applied

(1)

$$\frac{91}{400} = 0.2275$$

Q3

?

✓

≡

Bar chart icon

Pencil icon

A

B

C

Question:

1

2

3

4

5

6

7

(iv) has 8K resolution, given that the screen size is at least 75 inches.

(2)

$$P(R|S) = \frac{P(R \cap S)}{P(S)} = \frac{0}{74/400}$$

$$\frac{0}{0.185} = 0 \quad 0.122 = \frac{9/400}{74/400}$$

(c) Show that  $R$  and  $S$  are **not** statistically independent.

You should provide numerical justification.

$$\text{If independent } P(R \cap S) = P(R) \times P(S) \quad (2)$$

$$P(R \cap S) = 0 \quad P(R) = \frac{16}{400} \quad P(S) = 0.185$$

$$0.04 \times 0.185 = 0.0074 \neq 0$$

$\therefore$  Not independent

(d) Suggest a reason, in context, why  $R$  and  $S$  are **not** likely to be independent.

(1)

As televisions with 8K resolution are likely to have smaller screens

11 /15

### Part (a)

**B1:** Central cell is 9

**M1A1:** All three values of 0, 7 & 60 correct.

**M1A1:** All three totals correct.

**B1:** 91 correct.

### Part (b)

**B1:** Correct value.

**M0:** Incorrect value.

**A0:** Incorrect answer.

**B1:** Correct value.

**M1:** Correct denominator of  $\frac{74}{400}$  seen.

**A1:** Correct answer seen – this qualifies as attempting again and trying.

### Part (c)

**M1:** While  $(R \cap S)$  is calculated incorrectly,  $P(R)$  and  $P(S)$  are multiplied and compared to this.

**A0:** Concludes but values incorrect.

### Part (d)

**E0:** This is the reverse of the answer and not backed by the data.

Q3

?

✓

≡

Bar chart icon

✎

A

B

C

Question:

1

2

3

4

5

6

7

## Question 4

? Question

✓ Mark Scheme

≡ Examiner Comments

Performance

Response A

Response B

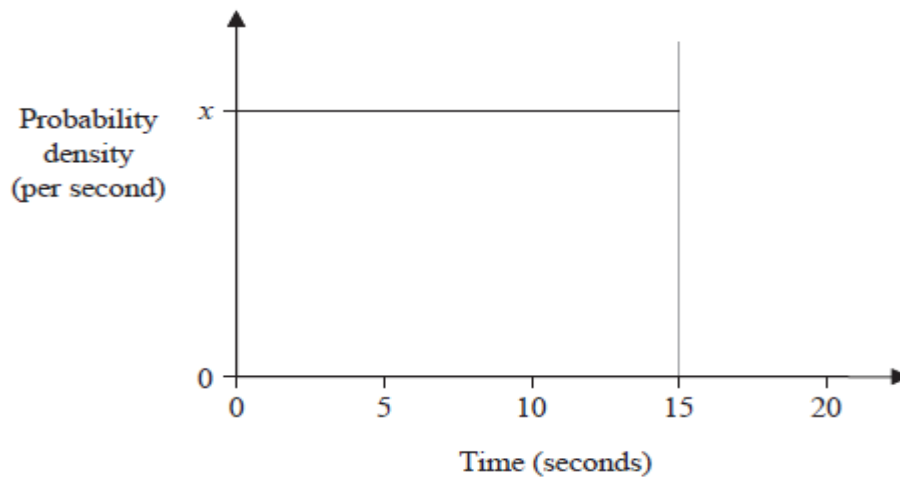
Response C

### ? Question 4 - Question

- 4 A video game is being created by a game developer. In the game, zombies are spawned into the game at random (a zombie will appear in the game at random).

Zombies are spawned one at a time.

In Version 1 of the game, the time between spawns (in seconds) follows the probability distribution in **Figure 3**.



- (a) Write down the value of  $x$  (1)

Ahmed is playing Version 1 of the game, and a zombie has just spawned into the game.

- (b) Calculate
- (i) the probability that no zombies will spawn in the next 10 seconds, (1)
  - (ii) the probability that **each** of the next three zombies will spawn more than 10 seconds after the last. (2)

In Version 2 of the game, the time between spawns (in seconds) follows the exponential distribution with **mean** 5 seconds.

Zara is playing Version 2 of the game, and a zombie has just spawned into the game.

- (c) Calculate the probability that
- (i) at least one zombie will spawn in the next 5 seconds, (3)
  - (ii) no zombies will spawn in the next 10 seconds, (2)
  - (iii) no more than 10 zombies will spawn in the next minute. (4)

Question:

1

2

3

4

5

6

7



### Question 4 – Question (Cont.)

The game developer decides that the probability found in (c)(iii) is too low.

She decides the probability that no more than 10 zombies will spawn within a minute of another zombie should be as close to 0.95 as possible, but no less than 0.95

(d) Using this information, find, to the nearest integer, the average number of zombies that will spawn in the next minute.

You may use trial and improvement.

(1)

(e) Using your answer to (d) state the new mean value for the exponential distribution in (c)

(1)

(Total for Question 4 is 15 marks)



### Question 4 - Mark Scheme

Question	Scheme	Marks	AO	Notes
4(a)	$x = \frac{1}{15}$	B1	1.1	or awrt 0.067
4(b)(i)	$P(X > 10)$			
	$= P(10 < X < 15)$			
	$= 5 \times \frac{1}{15} = \frac{1}{3}$	B1	1.1	oe or awrt 0.33
4(b)(ii)	$\left(\frac{1}{3}\right)^3$	M1	1.2	PI Their (i) cubed
	$= \frac{1}{27}$	A1	1.2	oe or awrt 0.037
4(c)(i)	[X = Number of Zombies spawning in the next 5 seconds]			
	[X ~ Po( $\lambda$ )]			
	$\lambda = 1$	B1	1.2	
	$P(X \geq 1) = 1 - P(X = 0)$	M1	1.2	PI
	$= 0.632$ (3sf)	A1	1.2	awrt
	Alternative			

Q4



A

B

C

Question:

1

2

3

4

5

6

7

Question	Scheme	Marks	AO	Notes
	[X = Time before next zombie spawn (seconds)]			
	[X ~ Exp( $\lambda$ )]			
	$\lambda = \frac{1}{5} = 0.2$	(B1)		oe
	$P(X \leq 5) = 1 - e^{-0.2 \times 5}$	(M1)		PI Correct use of exponential formula <b>or</b> 0.368 seen <b>Condone</b> $\lambda = 5$ <b>Condone</b> one small slip
	= 0.632 (3sf)	(A1)		
4(c)(ii)	$X \sim Po(2)$ $P(X = 0)$	M1	1.2	PI
	= 0.135 (3sf)	A1	1.2	awrt
	<b>Alternative</b>			
	$P(X > 10) = 1 - P(X \leq 10)$			
	$= 1 - (1 - e^{-0.2 \times 10})$ $= e^{-0.2 \times 10}$	(M1)		PI
	= 0.135 (3sf)	(A1)		
4(c)(iii)	[Y = Number of zombies spawning in the next minute (60 seconds)]			
	$Y \sim Po(\lambda)$	B1	2.1b	PI Poisson distribution stated or clearly used
	Mean: $\lambda = 60 \times 0.2 = 12$	B1	1.2	PI
	$P(Y \leq 10)$	M1	1.2	PI Attempt to find this probability
	= 0.347 (3sf)	A1	1.2	awrt
4(d)	From tables or calculator, use $\lambda = 6$ [per minute]	B1	1.2	

Q4



A

B

C

Question:

1

2

3

4

5

6

7

Question	Scheme	Marks	AO	Notes
4(e)	Exponential mean = $\frac{1}{6}$ minute	B1	1.2	Or equivalent in seconds
		<b>Total</b>	<b>15</b>	

## Question 4 - Examiner Comments

Parts (a) and (b) were well answered in general, although some candidates mistakenly gave the answer a zombie would spawn in the next 10 seconds instead for (b)(i).

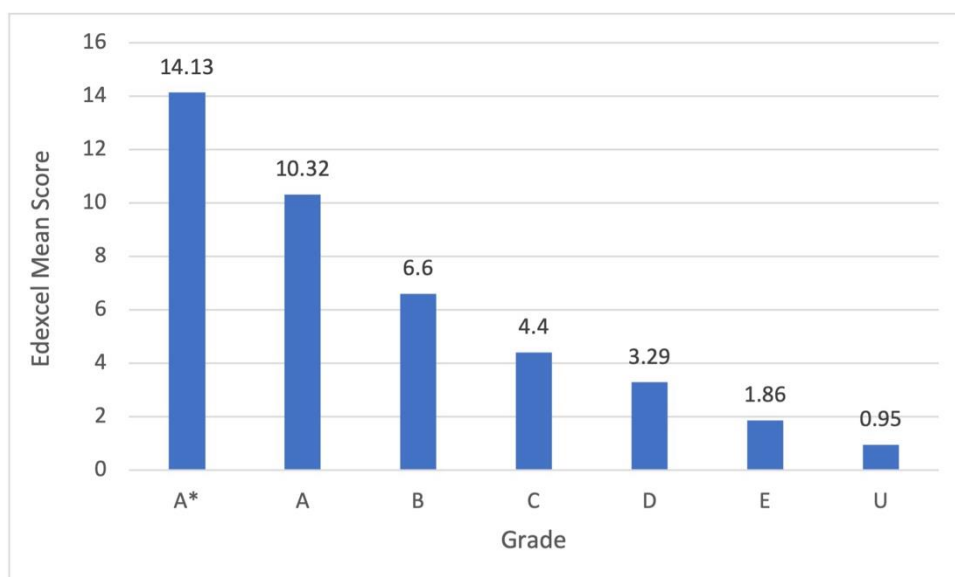
Candidates had a choice of the Poisson or exponential distribution for (c)(i) and (c)(ii) with most choosing to use the exponential distribution. Relatively few candidates answered (c)(ii) correctly.

Part (c)(iii) was more accessible to some than the preceding parts, but some candidates who had not been able to access (c)(ii) chose not to continue with the question. These candidates missed out on some relatively straightforward method marks in (c)(iii)

Parts (d) and (e) were good at discriminating between candidates due to the difficulty of the question. Candidates were expected to use trial and improvement, as mentioned in the notes, rather than to use logarithms although this was a valid method.

## Question 4 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade:							
			ALL	A*	A	B	C	D	E	U
4.69	15	31	4.69	14.13	10.32	6.60	4.40	3.29	1.86	0.95



Q4



A

B

C



Question:

1

2

3

4

5

6

7



## Question 4 - Response A

a) Write down the value of  $x$ 

(1)

$$\frac{1}{15} = 0.0667$$

Ahmed is playing Version 1 of the game, and a zombie has just spawned into the game.

b) Calculate

(i) the probability that no zombies will spawn in the next 10 seconds,

(1)

$$10 \times \frac{1}{15} = \frac{2}{3} \quad 1 - \frac{2}{3} = \frac{1}{3} = 0.333$$

(ii) the probability that **each** of the next three zombies will spawn more than 10 seconds after the last.

(2)

$$X \sim B(3, \frac{1}{3})$$

$$P(X=3) = 0.0370$$

Q4



A

B

C

Question:

1

2

3

4

5

6

7

(c) Calculate the probability that

(i) at least one zombie will spawn in the next 5 seconds,

(3)

$$\begin{aligned} \lambda &= 5 & 1 - e^{-5} & P(X \geq 1) = 1 - e^{-5 \times 1} \\ P(X \geq 1) &= 1 - (1 - e^{-1/5 \times 5}) & & 0.9933 \\ &= e^{-1/5 \times 5} & & 0.3679 \quad 0.8187 \\ & & & 0.3679 \end{aligned}$$

(ii) no zombies will spawn in the next 10 seconds,

(2)

$$\begin{aligned} P(X \leq 0) &= 1 - e^{-1/5 \times 10} \\ &= 0.8187 \end{aligned}$$

(iii) no more than 10 zombies will spawn in the next minute.

(4)

$$\begin{aligned} P(X \leq 10) & & 12\lambda &= 2.4 \\ 1 - e^{-2.4 \times 10} &= 0.9093 \end{aligned}$$

(d) Using this information, find, to the nearest integer, the average number of zombies that will spawn in the next minute.

You may use trial and improvement.

(1)

$$\begin{aligned} \lambda &= 0.25 & 12\lambda &= 3 \\ \lambda &= 0.25 & \text{mean} &= 1 \\ & & 0.25 & \end{aligned}$$

(e) Using your answer to (d) state the new mean value for the exponential distribution in (c)

(1)

$$\frac{1}{0.25} = 4$$

Q4

?

✓

≡

Bar chart icon

Icon with a notepad and pencil

A

B

C



Question:

1

2

3

4

5

6

7

8 /15

**Part (a)****B1:** Correct value.**Part (b)****B1:** Correct value.**M1A1:** Correct answer using binomial method.**Part (c)****B1M1A0:** Uses exponential distribution with 0.368 seen but incorrect answer.**M1A1:** Correct value and working for exponential method.**M0B0:** Poisson distribution not seen.**M0A0:** Incorrect calculation.**Part (d)****B0:** Incorrect value.**Part (e)****B0:** Incorrect value.

Q4



A

B

C



Question:

1

2

3

4

5

6

7



## Question 4 - Response B

(a) Write down the value of  $x$ 

(1)

1

Ahmed is playing Version 1 of the game, and a zombie has just spawned into the game.

(b) Calculate

(i) the probability that no zombies will spawn in the next 10 seconds,

(1)

$$\frac{1}{3} \times 1 = 0.667$$

(ii) the probability that **each** of the next three zombies will spawn more than 10 seconds after the last.

(2)

Q4



A

B

C

Question:

1

2

3

4

5

6

7

(c) Calculate the probability that

(i) at least one zombie will spawn in the next 5 seconds,

(3)

$P_0: CD$

$$\begin{aligned} x &= 0 \\ \lambda &= 1 \end{aligned} \Bigg] 0.368$$

~~$P_0: CD$~~

$$1 - 0.368 = 0.632$$

$$X \sim P_0(5)$$

(ii) no zombies will spawn in the next 10 seconds,

(2)

$P_0: CD$

$$\begin{aligned} x &= 0 \\ \lambda &= 2 \end{aligned} \Bigg] 0.135$$

(iii) no more than 10 zombies will spawn in the next minute.

(4)

$P_0: CD$

$$\begin{aligned} x &= 10 \\ \lambda &= 12 \end{aligned} \Bigg] 0.347$$

$$1 - 0.347 = 0.653$$

(d) Using this information, find, to the nearest integer, the average number of zombies that will spawn in the next minute.

You may use trial and improvement.

(1)

12

(e) Using your answer to (d) state the new mean value for the exponential distribution in (c)

(1)

11.5

Q4

?

✓

≡

Bar chart icon

✎

A

B

C

8 / 15



Question:

1

2

3

4

5

6

7

**Part (a)**

**B0:** Incorrect value.

**Part (b)**

**B0:** Incorrect value.

**M0A0:** Left blank.

**Part (c)**

**B1M1A1:** Correct value and working for Poisson method.

**M1A1:** Correct value and working for Poisson method.

**M1B1:** Use of Poisson with 12

**M1:** Correctly calculates with 10

**A0:** Incorrect final answer.

**Part (d)**

**B0:** Incorrect value.

**Part (e)**

**B0:** Incorrect value.

Q4



A

B

C

Question:

1

2

3

4

5

6

7



### Question 4 - Response C

(a) Write down the value of  $x$

(1)

$$\frac{1}{15}$$

Ahmed is playing Version 1 of the game, and a zombie has just spawned into the game.

(b) Calculate

(i) the probability that no zombies will spawn in the next 10 seconds,

(1)

$$P(>10) = \frac{5}{15}$$

(ii) the probability that **each** of the next three zombies will spawn more than 10 seconds after the last.

(2)

$$\frac{5}{15} \times \frac{5}{15} \times \frac{5}{15} = \frac{1}{27}$$

Q4



A

B

C



Question:

1

2

3

4

5

6

7

(c) Calculate the probability that

(i) at least one zombie will spawn in the next 5 seconds,

$$P(X \leq 5) = 1 - e^{-0.2 \times 5}$$

$$\text{mean} = 5 = \frac{1}{0.2}$$

(3)

$$P(X \leq 5) = 0.6321$$

$$P(X \leq 5) = 1 - e^{-0.2 \times 5} = 0.6321$$

(ii) no zombies will spawn in the next 10 seconds,

$$P(X \leq 10) = 1 - e^{-0.2 \times 10}$$

$$P(X \geq 10) = e^{-0.2 \times 10} = 0.1353$$

(iii) no more than 10 zombies will spawn in the next minute.

$$X \sim P(12)$$

$$P(X \leq 10) = 0.3472$$

(4)

(d) Using this information, find, to the nearest integer, the average number of zombies that will spawn in the next minute.

You may use trial and improvement.

(1)

$$Y \sim P(6) \quad P(Y \leq 10) = 0.2574$$

$$Y \sim P(7) \quad P(Y \leq 10) = 0.2015$$

$\therefore$  average number of zombies should be 6

(e) Using your answer to (d) state the new mean value for the exponential distribution in (c)

(1)

$$\frac{1}{\frac{1}{60}} = 60$$

$$\frac{1}{10} \quad \frac{1}{0.1} = 10$$

mean = 10 seconds

Q4



A

B

C



Question:

1

2

3

4

5

6

7

15 /15

**Part (a)****B1:** Correct value.**Part (b)****B1:** Correct value.**M1A1:** Correct value and method.**Part (c)****B1M1A1:** Correct value and working for exponential method.**M1A1:** Correct value and working for exponential method.**M1B1:** Use of Poisson with 12**M1A1:** Correctly calculates.**Part (d)****B1:** Correct value.**Part (e)****B1:** Correct value in seconds.

Q4



A

B

C

Question:

1

2

3

4

5

6

7

## Question 5

[Question](#)

[Mark Scheme](#)

[Examiner Comments](#)

[Performance](#)

[Response A](#)

[Response B](#)

[Response C](#)

### Question 5 - Question

- 5 In 1950, W.S. Robinson conducted a statistical analysis of some data from the 1930 census in the USA. Part of his analysis investigated correlation between rates of literacy and place of birth.

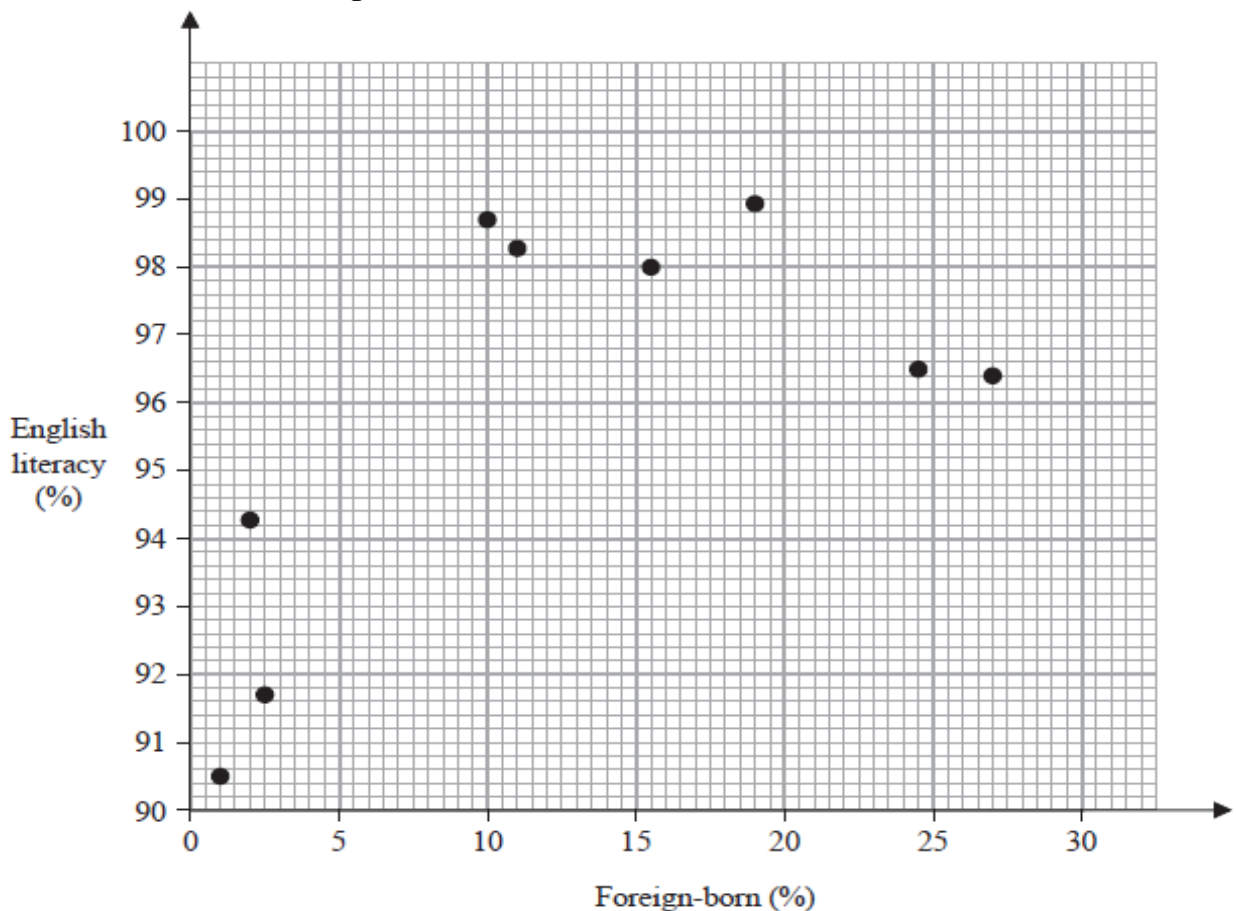
In 1930, USA census data was split into 9 geographical regions.

For each region, the following two percentages were calculated.

**Foreign-born** Percentage of the adult population born outside the USA.

**English literacy** Percentage of the adult population able to read English to a standard sufficient to function in society.

In Robinson's paper, the full data was not provided, though a scatter diagram was given similar in form to the one in **Figure 4**.



[source: <https://doi.org/10.2307/Data/2087176>]

Figure 4

 **Question 5 – Question (Cont.)**

- (a) Complete the table below using **Figure 4**. (2)

Foreign-born (%)	1	2	2.5	10	11	15.5	18	24.5	27
English literacy (%)									

Robinson calculated a correlation coefficient for this data.

- (b) Calculate the value for the Pearson’s product-moment correlation coefficient between Foreign-born (%) and English literacy (%) for adults in the 9 regions. (1)
- (c) State **one** reason why it may be inappropriate to calculate the Pearson’s product-moment correlation coefficient between Foreign-born (%) and English literacy (%) for adults in the 9 regions. (1)

In the same paper, Robinson considers correlation between Foreign-born (%) and English literacy (%) for **all of the adults** in the census, considered **at individual level** (so each data point represents one person, not one region).


He produces the following statistic.


$$r = -0.118$$


- (d) Explain why this statistic may be unexpected when compared with the result obtained in (b) (2)
- (e) Using your understanding of the context, give a short explanation to explain how the unexpected situation regarding the results in (b) and (d) may have occurred. (4)


**(Total for Question 5 is 10 marks)**


**Q5**











**A**

**B**

**C**



Question:

1

2

3

4

5

6

7

**Question 5 - Mark Scheme**

Question	Scheme				Marks	AO	Notes			
5(a)	Foreign Born	1	2	2.5	10	11	15.5	18	24.5	27
	English Literacy	90.5	94.3	91.7	98.7	98.3	98	98.9	96.5	96.4
					M1	1.1	PI At least one point correctly located (in both variables) Condone $\pm 0.1$			
					M1	1.1	PI All points correctly located Condone $\pm 0.1$			
	Printing Error									
	<ul style="list-style-type: none"><li>Foreign born 18 replaced with 19 (leads to <math>r=0.624</math>)</li><li>English literacy recorded as 0 for foreign born=18 (leads to <math>r=-0.165</math>)</li><li>Foreign born 18 has English literacy left blank (leads to <math>r=0.59</math>)</li></ul>									
5(b)	$r = 0.616$ (3sf)				A1	1.2	awfw 0.59~0.64			
5(c)	Possible reasons									
	The data is not linear.									
	There appear to be two different groups.									
					E1	3.1a				

Q5



A

B

C



Question:

1

2

3

4

5

6

7

Question	Scheme	Marks	AO	Notes
5(d)	The same data, when considered at different levels...	E1	3.1b	Mention that the same data is used <b>or</b> Mention of different levels
	...is producing both positive and negative values of $r$ .	E1	3.1b	Mention of positive and negative $r$ <b>or</b> Mention of <b>extreme</b> difference in values (oe)
5(e)	Foreign-born people moving to the USA are likely to move to regions with lots of urban areas (where there are more jobs).			Linking foreign-born with a third variable <b>Examples</b> Urbanness Availability of jobs Money/prosperity Better schools Nicer area Others of their nationality in area
	People in urban areas are generally better educated than those in rural areas.			Linking education (or literacy skills) with same variable <b>or</b> Linking education (or literacy skills) with one of above variables if first E1 not awarded
	So, we would expect regions with lots of foreign-born population to have a high literacy.			Association between foreign-born and English literacy due to a common third variable
	However, at an individual level, people who are foreign-born are more likely to have another language as their first language, so their English literacy is likely to be poorer.			Implication that English may not be first language
	An individual can't have a foreign born %, they are either foreign born or not			
	The grouped data is likely to have fewer outliers			Or anomalies
	Looking at individuals would have a larger sample size with more data points			

Q5



A

B

C

Question:

1

2

3

4

5

6

7

Question	Scheme	Marks	AO	Notes
	9 regions might not cover the whole of the USA while individual data does			
	Literacy rate of American born population in an area may skew figures			
	May not have data for all adults in a region			
		E1, E1, E1, E1	2.1a, 2.1a, 2.1a, 2.1a	One mark per comment Not exhaustive
<b>Total</b>			<b>10</b>	

## Question 5 - Examiner Comments

Parts (a) and (b) were answered well in spite of the fact that there was a misprint, though appropriate steps were taken to ensure that candidates were not disadvantaged due to this.

Part (c) was a straightforward question for most candidates, provided they did not get confused between the product moment correlation coefficient and Spearman's correlation coefficient.

In (d) most candidates were able to identify the fact there was a difference in the correlation coefficients, but few fully explained that this was particularly surprising as it was exactly the same data.

A full answer to (e) was an indicator of exceptional understanding and rare to come across, however some candidates did give sophisticated answers, linking together multiple ideas such as the ruralness of an area and the quality of education.

Q5



A

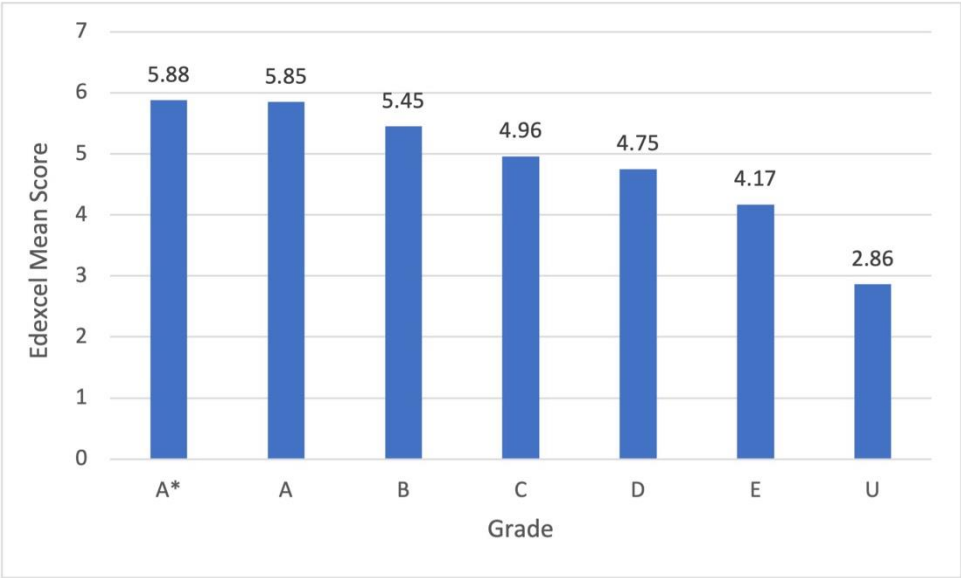
B

C



Question 5 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade:							
			ALL	A*	A	B	C	D	E	U
4.86	10	49	4.86	5.88	5.85	5.45	4.96	4.75	4.17	2.86





Q5

?

✓

≡





A

B

C

Question:

1

2

3

4

5

6

7



## Question 5 - Response A

(a) Complete the table below using **Figure 4**.

(2)

Foreign-born (%)	1	2	2.5	10	11	15.5	18	24.5	27
English literacy (%)	90.6	94.7	91.8	98.8	98.7	98.4	98.9	96.8	96.7

Robinson calculated a correlation coefficient for this data.

(b) Calculate the value for the Pearson's product-moment correlation coefficient between Foreign-born (%) and English literacy (%) for adults in the 9 regions.

(1)

0.627

(c) State **one** reason why it may be inappropriate to calculate the Pearson's product-moment correlation coefficient between Foreign-born (%) and English literacy (%) for adults in the 9 regions.

(1)

foreign born only goes up to 30% and english literacy goes up to 100%, isn't a clear graph

(d) Explain why this statistic may be unexpected when compared with the result obtained in (b)

(2)

it is a negative -0.118

Q5



A

B

C

Question:

1

2

3

4

5

6

7

(e) Using your understanding of the context, give a short explanation to explain how the unexpected situation regarding the results in (b) and (d) may have occurred.

(4)

- results are independent  
 - data points represent one person  
 - the graph is very confusing and may have equalled different values, results are independent.

3 /10

**Part (a)**

**M1:** 90.6 correct within 0.1 of 90.5

**M0:** 96.8 not within 0.1 of 96.5

**Part (b)**

**A1:** r falls within 0.59-0.64

**Part (c)**

**E0:** Incorrect reason.

**Part (d)**

**E0:** No identification that same data is being used in both calculations.

**E1:** Minimum answer expressing difference in r values.

**Part (e)**

**E0E0E0E0:** No relevant points made.

Q5



A

B

C

Question:

1

2

3

4

5

6

7



## Question 5 - Response B

(a) Complete the table below using **Figure 4**.

(2)

Foreign-born (%)	1	2	2.5	10	11	15.5	18	24.5	27
English literacy (%)	90.5	<del>91.75</del> 94.3	91.75	98.7	98.3	98	98.9	96.5	96.4

Robinson calculated a correlation coefficient for this data.

(b) Calculate the value for the Pearson's product-moment correlation coefficient between Foreign-born (%) and English literacy (%) for adults in the 9 regions.

(1)

$$r = 0.616$$

(c) State **one** reason why it may be inappropriate to calculate the Pearson's product-moment correlation coefficient between Foreign-born (%) and English literacy (%) for adults in the 9 regions.

(1)

Some foreign countries may speak more English or may have English as a more common language than other foreign countries meaning some people foreign born may have already spoken fluent English before arriving to ~~the~~ USA.

(d) Explain why this statistic may be unexpected when compared with the result obtained in (b)

(2)

This is unexpected as the result obtained in (b) is moderate positive showing a positive correlation however this shows a weak negative correlation implying that the lower % foreign born the higher their English literacy is.

Q5



A

B

C



Question:

1

2

3

4

5

6

7

- (e) Using your understanding of the context, give a short explanation to explain how the unexpected situation regarding the results in (b) and (d) may have occurred.

(4)

Robinson uses each data point as a person not a whole region  
 this means his data is smaller and weaker ~~more~~  
 which implies his results aren't reliable enough to represent  
 the regions as a whole.

5 /10

**Part (a)****M1M1:** All correct.**Part (b)****A1:** r correct.**Part (c)****E0:** Incorrect reason.**Part (d)****E0:** No identification that same data is being used in both calculations.**E1:** Identifies difference in r values.**Part (e)****E1:** “smaller and weaker” considered equivalent to “larger sample size with more data points”**E0E0E0:** No further relevant points made.

Q5



A

B

C

Question:

1

2

3

4

5

6

7



## Question 5 - Response C

(a) Complete the table below using **Figure 4**.

(2)

Foreign-born (%)	1	2	2.5	10	11	15.5	18	24.5	27
English literacy (%)	90.5	94.3	91.7	98.7	98.3	98.0	98.9	96.5	96.4

Robinson calculated a correlation coefficient for this data.

(b) Calculate the value for the Pearson's product-moment correlation coefficient between Foreign-born (%) and English literacy (%) for adults in the 9 regions.

(1)

$r = 0.616$  (3sf) - moderate positive correlation

(c) State **one** reason why it may be inappropriate to calculate the Pearson's product-moment correlation coefficient between Foreign-born (%) and English literacy (%) for adults in the 9 regions.

(1)

the data is not linear

(d) Explain why this statistic may be unexpected when compared with the result obtained in (b)

(2)

This  $r$  value is negative, suggesting a whereas the value in (b) is positive, and suggests a much stronger correlation, which is unexpected, especially since the same data is used for two values come from the same data.

Q5



A

B

C

Question:

1

2

3

4

5

6

7

(e) Using your understanding of the context, give a short explanation to explain how the unexpected situation regarding the results in (b) and (d) may have occurred.

(4)

~~In part (b)~~ Part (b) uses data from Figure 4, which shows areas with the lowest Foreign-born (%) to have the lowest English Literacy (%). This may be because these regions are more rural ~~then~~ and sparsely populated than the others, so although few immigrants live there, the literacy level is lower due to worse lower levels of education. In the other regions, the correlation ~~However~~ suggests that with more Foreign-born people (%), the Literacy rate decreases. These other regions may be more well represented in the more detailed data for part (d) so than the in part (b), so will affect the  $r$  value more, resulting in the different  $r$  value seen in part (d).

8 /10

**Part (a)**

**M1M1:** All values correct.

**Part (b)**

**A1:** Correct  $r$ .

**Part (c)**

**E1:** Correct reason.

**Part (d)**

**E1:** Identifies same data is being used in both calculations.

**E1:** Identifies difference in  $r$  values.

**Part (e)**

**E1:** Mention of rural/sparsely populated.

**E1:** Links this to education level.

**E0E0:** No further relevant points.

Q5



A

B

C

Question:

1

2

3

4

5

6

7

## Question 6

 Question

 Mark Scheme

 Examiner Comments

 Performance

 Response A

 Response B

 Response C

### Question 6 - Question

- 6 Asbestos is a type of mineral used in construction that has been linked to health risks. Riddhi works for an asbestos removal company. She is researching the removal of two types of asbestos,
- White asbestos (or chrysotile)
  - Blue asbestos (or crocidolite)

When there is an asbestos find during an inspection, its location and type are recorded.

An asbestos find is called an **instance** of asbestos.

**Figure 5** shows instances of the types of asbestos as a proportion of **total** instances of asbestos recorded by the company in 2022.

Asbestos type	Proportion of total instances
White	0.972
Blue	0.018
Other	0.010

**Figure 5**

Not all instances of asbestos are removed, as sometimes it is safer to cover the asbestos up and leave it in place.

**Figure 6** shows the proportion of instances in 2022 that were removed, by type of asbestos.

Asbestos type	Proportion of instances that are removed
White	0.015
Blue	0.971
Other	0.153

**Figure 6**

Question:

1

2

3

4

5

6

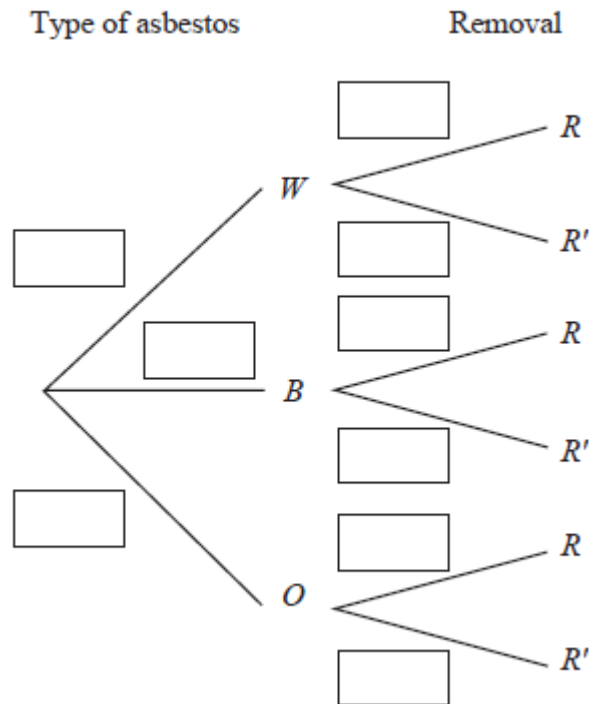
7



## Question 6 – Question (Cont.)

(a) Enter the correct probability values into each box in the tree diagram below.

(2)



(b) Calculate the probability that a randomly selected instance was removed.

(2)

(c) Find the probability that a randomly selected instance is blue asbestos, given that it was removed by the company.

(4)

(d) Suggest a possible reason, in context, why the probability found in (c) is greater than the probability that a randomly selected instance is blue asbestos.

(1)

The removal of different types of asbestos requires different types of safety equipment.

Riddhi states that the company should invest mostly in safety equipment for removing white asbestos, as more than 97% of asbestos instances are of this type.

(e) Explain whether you agree with Riddhi's statement.

You should include **numerical justification** from (c)

(3)

(Total for Question 6 is 12 marks)

Q6



A

B

C

Question:

1

2

3

4

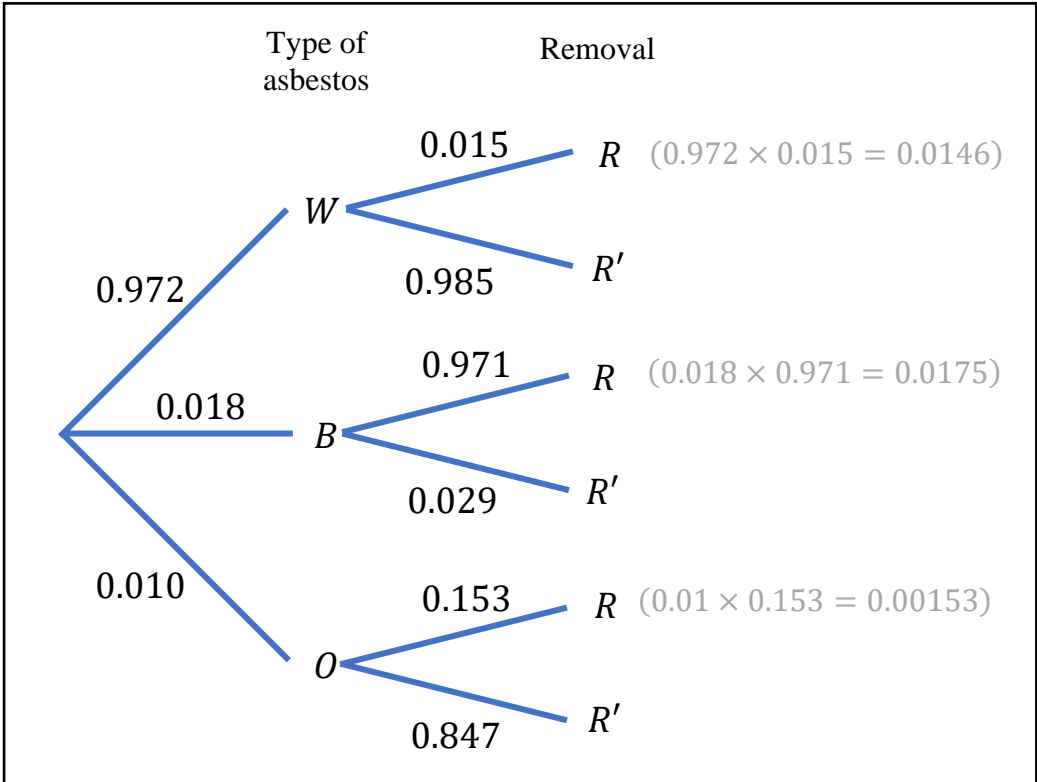
5

6

7



## Question 6 - Mark Scheme

Question	Scheme	Marks	AO	Notes
6(a)				
		A1	1.2	Any correct pair of conditional values given (0.015/0.985 after 0.972, 0.971/0.029 after 0.018 or 0.153/0.847 after 0.010)
		A1	1.2	Completely correct
6(b)	$0.972 \times 0.015 + 0.018 \times 0.971 + 0.01 \times 0.153$ $= 0.033588$	M1		Or $0.0146 + 0.0175 + 0.00153$ May be seen on diagram Allow one small slip
		A1		awrt 0.0336

Q6



A  
B  
C



Question:

1

2

3

4

5

6

7

Question	Scheme	Marks	AO	Notes
6(c)	$P(B R) = \frac{P(B \cap R)}{P(R)}$ $= \frac{0.018 \times 0.971}{0.972 \times 0.015 + 0.018 \times 0.971 + 0.01 \times 0.153}$ $= 0.520 \text{ (3sf)}$	B1	1.2	
		B1	1.2	Numerator correct <b>or</b> awrt 0.0175 seen anywhere (including on tree diagram)
		B1	1.2	Denominator correct or is 'their (b)'
		A1	1.2	awfw 0.52~0.55
6(d)	<b>Possible reasons (not exhaustive)</b>  Blue asbestos may be easier to remove (so it is removed more often).  Blue asbestos may be more dangerous (so it is removed more often).	E1	2.1a	Any sensible reason
6(e)	No, I disagree with Riddhi's statement...	E1dep	3.1b	dep on good attempt at explanation ft their (a)
	...as about 52% of the company's removals are of blue asbestos...	E1ft	3.1b	Reference to 0.520 ft their (a)
	...so lots of safety equipment for blue asbestos will be required.	E1	3.1b	Relating back to safety equipment correctly.
<b>Total</b>		<b>12</b>		

Q6



A

B

C

Question:

1

2

3

4

5

6

7

## Question 6 - Examiner Comments

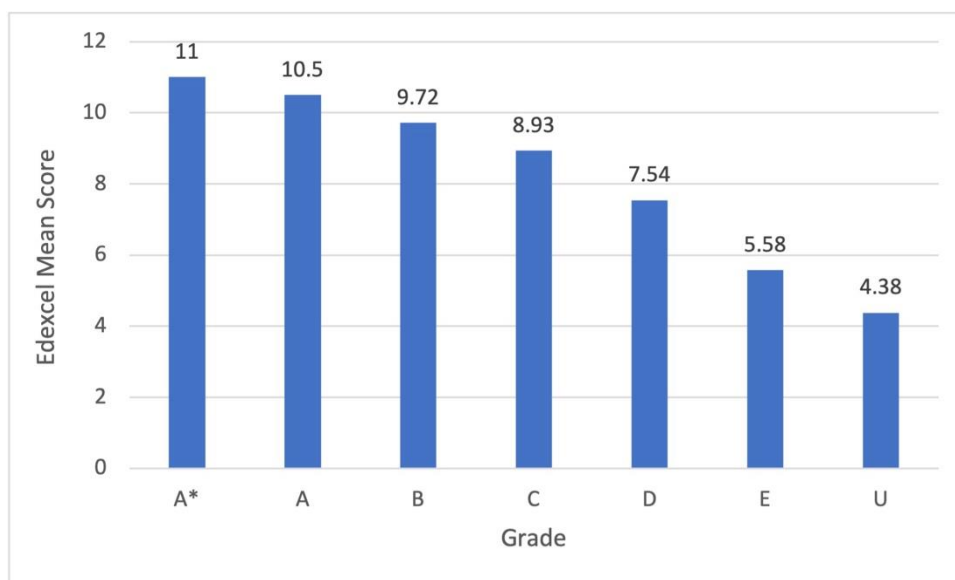
Most candidates filled in the tree diagram correctly, and were therefore able to access (b). The tree diagram was used by many candidates to complete (c) rather than using Bayes theorem, and as in question 3 candidates are advised to show full working in calculation based questions to earn method marks if they made a mistake in their diagram.

In (d) many candidates pointed out that the probability blue asbestos was removed was extremely high, but this is not a contextual reason – rather, this is a consequence of the reason it is removed more frequently.

Part (e) was generally answered well, but candidates need to make sure to link back to the statement you are commenting on in the question. Some candidates did not include numerical justification as they were instructed to and were therefore penalised. Other candidates did not refer back to safety equipment and also lost a mark for this.

## Question 6 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade:							
			ALL	A*	A	B	C	D	E	U
8.19	12	68	8.19	11.00	10.50	9.72	8.93	7.54	5.58	4.38



Q6



A

B

C

Question:

1

2

3

4

5

6

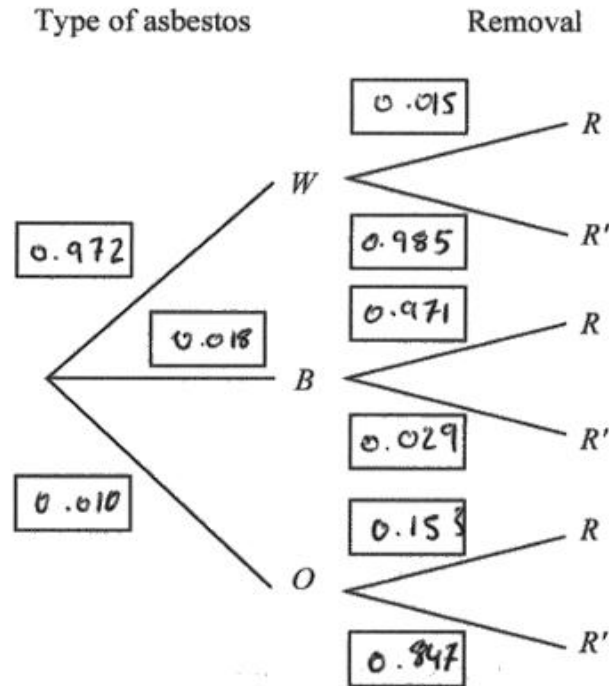
7



## Question 6 - Response A

(a) Enter the correct probability values into each box in the tree diagram below.

(2)



(b) Calculate the probability that a randomly selected instance was removed.

(2)

$$(0.972 \times 0.015) + (0.018 \times 0.971) + (0.010 \times 0.153) = 0.033588$$

(c) Find the probability that a randomly selected instance is blue asbestos, given that it was removed by the company.

(4)

$$0.018 \times (0.971) = 0.17478$$

$$0.033588 - 0.17478 = 0.01611$$

$$0.17478 - 0.033588 = 0.141192$$

(d) Suggest a possible reason, in context, why the probability found in (c) is greater than the probability that a randomly selected instance is blue asbestos.

(1)

~~It shows mathematical error~~ 97%  
of fine blue asbestos removed

Q6



A

B

C

Question:

1

2

3

4

5

6

7

(e) Explain whether you agree with Riddhi's statement.

You should include **numerical justification** from (c)

(3)

No, I do not agree, however there is 97% of white asbestos, only 1.5% of it is removed whereas both Blue and ~~white~~ Other Asbestos have 97.1% and 15.3% removed respectively, which is a huge increase compared to the 1.5% of Blue asbestos removed, however there is also a lower % of instances for both Blue and other asbestos, the company should invest a high % of safety equipment on white asbestos, but they should also invest in Blue and other asbestos as they are more likely to be removed.

(Total for Question 6 is 12 marks)

7 / 12

Q6

?

✓

≡

Bar chart icon

✎

A

B

C

**Part (a)**

**A1A1:** Fully correct tree diagram.

**Part (b)**

**M1A1:** Correct working and answer.

**Part (c)**

**B0:** No use of formula.

**B1:** 0.175 seen.

**B0:** Denominator not seen.

**A0:** Final answer incorrect.

**Part (d)**

**E0:** Not a contextual reason as to why blue is highly removed.

**Part (e)**

**E1:** Good attempt at explanation.

**E0:** Does not refer to their value from (c).

**E1:** Refers back to safety equipment.

Question:

1

2

3

4

5

6

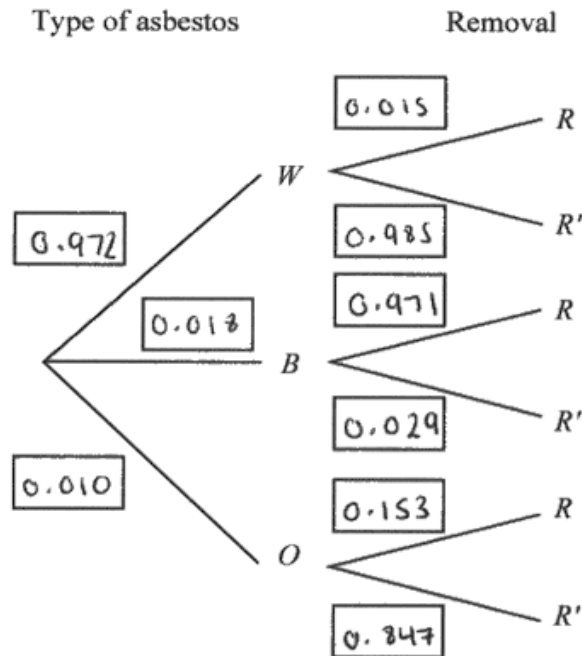
7



## Question 6 - Response B

(a) Enter the correct probability values into each box in the tree diagram below.

(2)



(b) Calculate the probability that a randomly selected instance was removed.

(2)

$$P(R) = (0.972 \times 0.015) + (0.018 \times 0.971) + (0.010 \times 0.153) = 0.336$$

(c) Find the probability that a randomly selected instance is blue asbestos, given that it was removed by the company.

(4)

$$P(B|R) = \frac{P(B \cap R)}{P(R)} = \frac{0.018 \times 0.971}{(0.972 \times 0.015) + (0.018 \times 0.971) + (0.010 \times 0.153)} = \frac{0.017478}{0.336} = 0.0520$$

Q6

?

✓

≡

Bar chart icon

Pencil icon

A

B

C

Question:

1

2

3

4

5

6

7

(d) Suggest a possible reason, in context, why the probability found in (c) is greater than the probability that a randomly selected instance is blue asbestos.

(1)

This is because the blue asbestos type has the highest proportion of instances removed.

(e) Explain whether you agree with Riddhi's statement.

You should include **numerical justification** from (c)

(3)

I disagree with Riddhi's statement because

$0.0520 \times 0.97 = 0.05044$  which is less

and tells us that Riddhi's statement is incorrect.

7 /12

**Part (a)**

**A1A1:** Fully correct tree diagram.

**Part (b)**

**M1A0:** Correct working but answers 0.336 not 0.0336

**Part (c)**

**B1:** Use of formula.

**B1:** Numerator correct.

**B1:** Denominator is their (b)

**A0:** Final answer incorrect.

**Part (d)**

**E0:** Not a contextual reason as to why blue is highly removed.

**Part (e)**

**E0:** Does not qualify as good attempt at explanation as not detailed enough.

**E1:** Uses their value from (c)

**E0:** No reference to safety equipment.

Q6



A

B

C

Question:

1

2

3

4

5

6

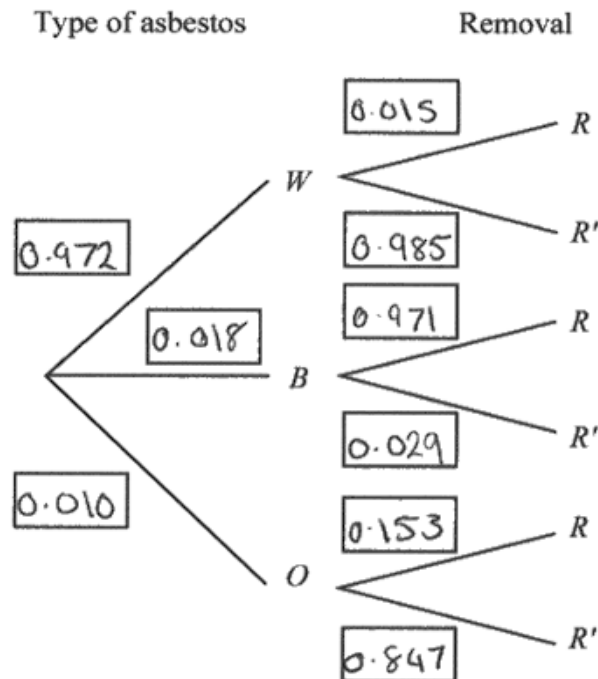
7



## Question 6 - Response C

(a) Enter the correct probability values into each box in the tree diagram below.

(2)



(b) Calculate the probability that a randomly selected instance was removed.

(2)

$$\begin{aligned}
 P(R) &= (0.972 \times 0.015) + (0.018 \times 0.971) + (0.01 \times 0.153) \\
 &= 0.01458 + 0.001748 + 0.00153 \\
 &= 0.03359 \\
 &= 0.0336 \text{ (3.s.f.)}
 \end{aligned}$$

(c) Find the probability that a randomly selected instance is blue asbestos, given that it was removed by the company.

(4)

$$\begin{aligned}
 \frac{P(B \cap R)}{P(R)} &= \frac{0.018 \times 0.971}{0.03359} \\
 &= \frac{0.017478}{0.03359} \\
 &= 0.5210
 \end{aligned}$$

Q6



A

B

C

Question:

1

2

3

4

5

6

7

(d) Suggest a possible reason, in context, why the probability found in (c) is greater than the probability that a randomly selected instance is blue asbestos.

(1)

If ~~B~~ The asbestos is blue it is much more likely to be removed than the other two asbestos types.

(e) Explain whether you agree with Riddhi's statement.

You should include **numerical justification** from (c)

(3)

I would disagree with Riddhi's statement as even though white Asbestos<sup>a</sup> is more common (97% to account 2%), the company is ~~more~~ more likely to need to remove blue asbestos. ~~With~~ If the Asbestos is being removed its around a 32% chance its blue asbestos and 43% it is white asbestos.

10 /12

Part (a)

A1A1: Fully correct tree diagram.

Part (b)

M1A1: Correct working and answer.

Part (c)

B1B1A1: Fully correct working and answer.

Part (d)

E0: Not a contextual reason as to why blue is highly removed.

Part (e)

E1: Disagreement and good explanation.

E1: Uses their value from (c)

E0: No reference to safety equipment.

Q6



A

B

C

Question:

1

2

3

4

5

6

7

## Question 7

 Question

 Mark Scheme

 Examiner Comments

 Performance

 Response A

 Response B

 Response C



### Question 7 - Question

- 7 Divya carries out a large-sample study to investigate the 2D:4D digit ratio for adult females

The 2D:4D digit ratio is calculated by dividing the length of the index finger (second digit) by the length of the ring finger (fourth digit) on their dominant hand.

Previous studies suggest that the 2D:4D digit ratio for adult females is normally distributed.

**Figure 7** shows the values of the **sextiles** in Divya's large-sample study.

[A **sextile** is similar to a quartile, but the data has been split into six equal proportions instead of 4.]

Sextile	Approximate percentile	Digit ratio
$S_1$	17th	0.942
$S_2$	33rd	0.958
$S_3$	50th	0.971
$S_4$	67th	0.984
$S_5$	83rd	1.000

**Figure 7**

- (a) State the approximate value of the mean of the digit ratio in Divya's study.

Explain your answer.

- (b) Explain why the approximate value of the standard deviation of the digit ratio in Divya's study is 0.029

(3)

Question:

1

2

3

4

5

6

7



## Question 7 – Question (Cont.)

Divya also carries out a second large-sample study to investigate the 2D:4D digit ratio for adult males.

Previous studies suggest that the 2D:4D digit ratio for adult males is also normally distributed.

Her second large-sample study yields the following statistics:

$$\bar{x} = 0.952$$

$$s = 0.034$$

(c) Find the probability that a randomly chosen male adult has a 2D:4D digit ratio

(i) greater than 1

(1)

(ii) between 0.9 and 1

(1)

(d) Find the probability that a randomly chosen adult male has a 2D:4D digit ratio lower than that of a randomly chosen adult female.

You should use your answers to (a) and (b)

(5)

(Total for Question 7 is 12 marks)



## Question 7 - Mark Scheme

Question	Scheme	Marks	AO	Notes
7(a)	$\mu \approx 0.971$  The normal distribution is symmetrical, so the mean equals the median.	B1  E1	2.1b  2.1b	
7(b)	Approximately $\frac{2}{3}$ of data lies within $\mu \pm \sigma \dots$  ...so $\mu + \sigma = S_5$  $\sigma \approx 1.000 - 0.971$ $= 0.029$  SC  $1.000 - 0.971 = 0.029$ or $0.971 - 0.942 = 0.029$ seen with no explanation scores E0M1A0	E1  M1  A1*dep	2.1b  2.1b  2.1b	oe  PI oe for example $\mu - \sigma = S_1$  or $0.971 - 0.942$ dep on E1

Q7



A  
B  
C

Question:

1

2

3

4

5

6

7

Question	Scheme	Marks	AO	Notes
7(c)(i)	$[M = 2D:4D \text{ ratio of a randomly selected male adult}]$ $[M \sim N(0.952, 0.034^2)]$ $P(M > 1) = 0.0790 \text{ (3sf)}$	B1	1.2	awrt 0.079
7(c)(ii)	$P(0.9 < M < 1) = 0.858 \text{ (3sf)}$	B1	1.2	awrt 0.86
7(d)	$[F = 2D:4D \text{ ratio of a randomly selected female adult}]$ $[F \sim N(0.971, 0.029^2)]$ $[D = \text{Difference between 2D:4D ratio of a randomly selected male and female adult}]$			PI <b>or</b> $D = M - F$ PI Normal distribution used for difference PI $\mu = \pm 0.019$ cao PI $\sigma = 0.0447 \text{ (awrt)}$ <b>or</b> $\sigma^2 = 0.00200 \text{ (awrt)}$ awfw 0.664~0.665 awfw 0.664~0.665
	$D = F - M$	M1	2.1b	
	$D \sim N(0.019, 0.0447^2)$	M1	2.1b	
		B1	1.2	
		B1	1.2	
	$P(D > 0) = 0.665 \text{ (3sf)}$	A1	1.2	
	$P(D > 0) = 0.665 \text{ (3sf)}$	A1	1.2	
	<b>Total</b>	<b>12</b>		

Q7



A

B

C

Question:

1

2

3

4

5

6

7

## Question 7 - Examiner Comments

Most candidates were able to correctly find the mean was 0.971, though some incorrectly justified this by averaging the sextiles and thus only scored one mark.

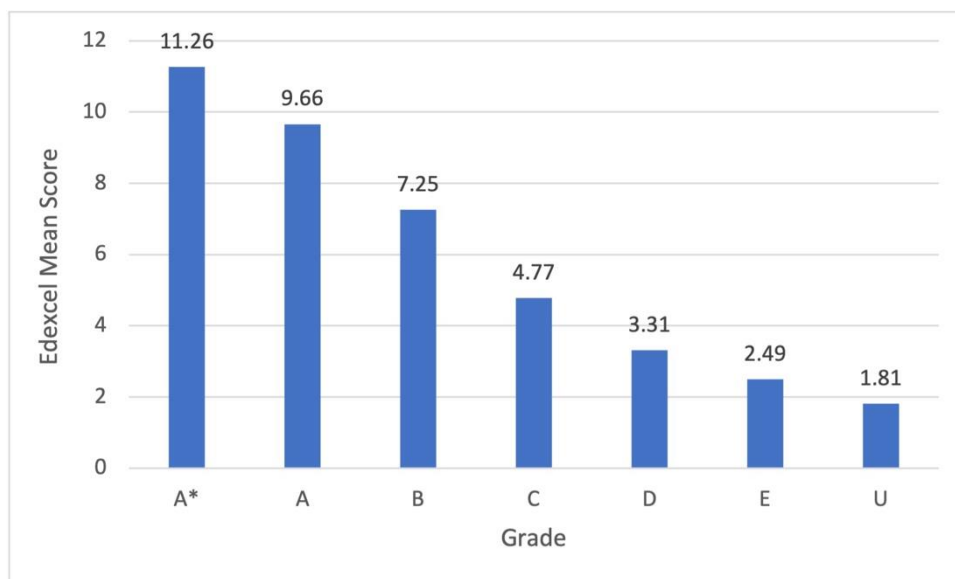
Explanations in (b) varied in quality, with some candidates manipulating numbers to attempt to find the stated value. Most correct answers considered  $\mu \pm \sigma$  but there were other correct alternatives possible, by considering the z-score of a given sextile.

Part (c) was a simple calculation that some candidates missed as they had given up after (b)

Part (d) required combining two normal distributions. If candidates realised this they were generally successful as this was a straightforward example of combining two normal distributions.

## Question 7 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade:							
			ALL	A*	A	B	C	D	E	U
4.99	12	42	4.99	11.26	9.66	7.25	4.77	3.31	2.49	1.81



Q7



A

B

C

Question:

1

2

3

4

5

6

7



## Question 7 - Response A

(a) State the approximate value of the mean of the digit ratio in Divya's study.

Explain your answer.

$$F \sim N(0.930, 0.029) \quad (2)$$

$$99 \div 2 = 49.5$$

$$\frac{16.5}{17} \times 0.958 = 0.9298 \rightarrow 0.930$$

(b) Explain why the approximate value of the standard deviation of the digit ratio in Divya's study is 0.029

(3)

$$0.013 + 0.016 = 0.029$$

Sum of difference between ~~the~~ value ~~appe~~ above and below each value.

(c) Find the probability that a randomly chosen male adult has a 2D:4D digit ratio

(i) greater than 1

(1)

$$P(X \geq 1) = 0.0790 \quad (3sf)$$

(ii) between 0.9 and 1

(1)

$$P(0.9 \leq X \leq 1) = 0.858 \quad (3sf)$$

(d) Find the probability that a randomly chosen adult male has a 2D:4D digit ratio lower than that of a randomly chosen adult female.

You should use your answers to (a) and (b)

(5)

$$P(M < F)$$

$$M \sim N(0.952, 0.034^2)$$

$$F \sim N(0.930, 0.029^2)$$

$$P(M - F < 0) = 0.311 \quad (3sf)$$

$$M - F \sim \left( \overset{0.022}{\cancel{0.882}}, 0.001997 \right) \sigma = \sqrt{0.001997}$$

Q7



A

B

C



Question:

1

2

3

4

5

6

7

4 /12

**Part (a)****B0:** Incorrect value given.**E0:** No correct explanation seen.**Part (b)****E0M0A0:** No relevant points made, just attempts at calculation to find the answer given.**Part (c)****B1:** Correct value.**B1:** Correct value.**Part (d)****M1:** Rearranges  $M < F$  to  $M - F < 0$ **M0:** While brackets are seen, the normal distribution is not specified for the difference.**B0:** Incorrect mean.**B1:** Correct variance.**A0:** No final answer.

Q7



A

B

C

Question:

1

2

3

4

5

6

7



## Question 7 - Response B

(a) State the approximate value of the mean of the digit ratio in Divya's study.

Explain your answer.

(2)

0.971, the 50th percentile is the middle and equal proportion mean is symmetrical and therefore the median is also the mean

(b) Explain why the approximate value of the standard deviation of the digit ratio in Divya's study is 0.029

(3)

~~1 - 0.971 = 0.029~~  $1 - 0.971 = 0.029$

Symmetrical, equal proportion mean, the data each data point is an equal and same distance away from 0.971

(c) Find the probability that a randomly chosen male adult has a 2D:4D digit ratio

(i) greater than 1

(1)

$$P(X > 1) = 0.079$$

(ii) between 0.9 and 1

(1)

$$P(0.9 \leq X \leq 1) = 0.858$$

(d) Find the probability that a randomly chosen adult male has a 2D:4D digit ratio lower than that of a randomly chosen adult female.

You should use your answers to (a) and (b)

Male = X, Female = Y

(5)

$$X < Y \Rightarrow X - Y < 0$$

$$X(0.952, 0.034)$$

$$Y(0.971, 0.029)$$

$$X - Y \sim N(0.952 - 0.971, 0.034 + 0.029)$$

$$= X - Y \sim N(-0.019, 0.063)$$

$$P(X < 0) = 0.6185 \dots P = 0.619$$

Q7



A

B

C

**Part (a)**  
**B1:** Correct value given.  
**E1:** Refers to symmetry of normal distribution.

**Part (b)**  
**E0:** No comment on  $\frac{2}{3}$  of data seen.  
**M0:** No reference to  $S_5$   
**A0:** This is not awarded despite seeing  $1 - 0.971 = 0.029$  as it is dependent on the E1

**Part (c)**  
**B1:** Correct value.  
**B1:** Correct value.


**Part (d)**  
**M1:** Rearranges  $X < Y$  to  $X - Y < 0$  with  $X$  and  $Y$  defined.  
**M1:**  $X - Y \sim N$  seen.  
**B1:** Correct mean.  
**B0:** Incorrect standard deviation.  
**A0:** Incorrect answer.


Q7

?

✓

≡





A

B

C

Question:

1

2

3

4

5

6

7



## Question 7 - Response C

- (a) State the approximate value of the mean of the digit ratio in Divya's study.

Explain your answer.

(2)

The mean is 0.971 meaning that the average woman's digit ratio lies within Sextile 3.

- (b) Explain why the approximate value of the standard deviation of the digit ratio in Divya's study is 0.029

(3)

$$1 - 0.971 = 0.029$$

$$0.971 - 0.942 = 0.029$$

The difference between the quartile mean in the lower and middle & upper and middle quartiles is 0.029

- (c) Find the probability that a randomly chosen male adult has a 2D:4D digit ratio

- (i) greater than 1

(1)

$$P(X > 1) = 0.0790$$

- (ii) between 0.9 and 1

(1)

$$P(0.9 \leq X \leq 1) = 0.8579 \rightarrow 0.858 \text{ (3sf)}$$

- (d) Find the probability that a randomly chosen adult male has a 2D:4D digit ratio lower than that of a randomly chosen adult female.

You should use your answers to (a) and (b)

(5)

$$P(M - F < 0) \quad P(F - M > 0)$$

$$M - F \sim N(0.952 - 0.971, 0.029^2 + 0.034^2)$$

$$M - F \sim N$$

$$M - F \sim N(0.971 - 0.952, 0.029^2 + 0.034^2)$$

$$F - M \sim N(0.019, 0.0447^2)$$

$$P(F - M > 0) = 0.6646 \rightarrow 0.665 \text{ (3sf)}$$

Q7



A

B

C



Question:

1

2

3

4

5

6

7

9 /12

**Part (a)**

**B1:** Correct value given.

**E0:** No correct explanation seen.

**Part (b)**

**E0M1A0:** Special case awarded.

**Part (c)**

**B1:** Correct value.

**B1:** Correct value.

**Part (d)**

**M1M1B1B1A1:** Fully correct working and answer.

Q7



A

B

C

