Pearson Edexcel

Pearson Edexcel
Level 3 Advanced GCE in Statistics (9ST0)

May/June 2022 Exemplar
9ST0-01 A level Statistics
Paper 01: Data and Probability
Senior Examiner's feedback on student responses

Contents

About this booklet 2

How to use this booklet 3

General Examiner Feedback 4

Question 1 5

Question 2 17

Question 3 28

Question 4 34

Question 5 47

Question 6 60

Question 7 67

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

© Pearson Education Ltd 2023

Skip to Main Contents

About this booklet

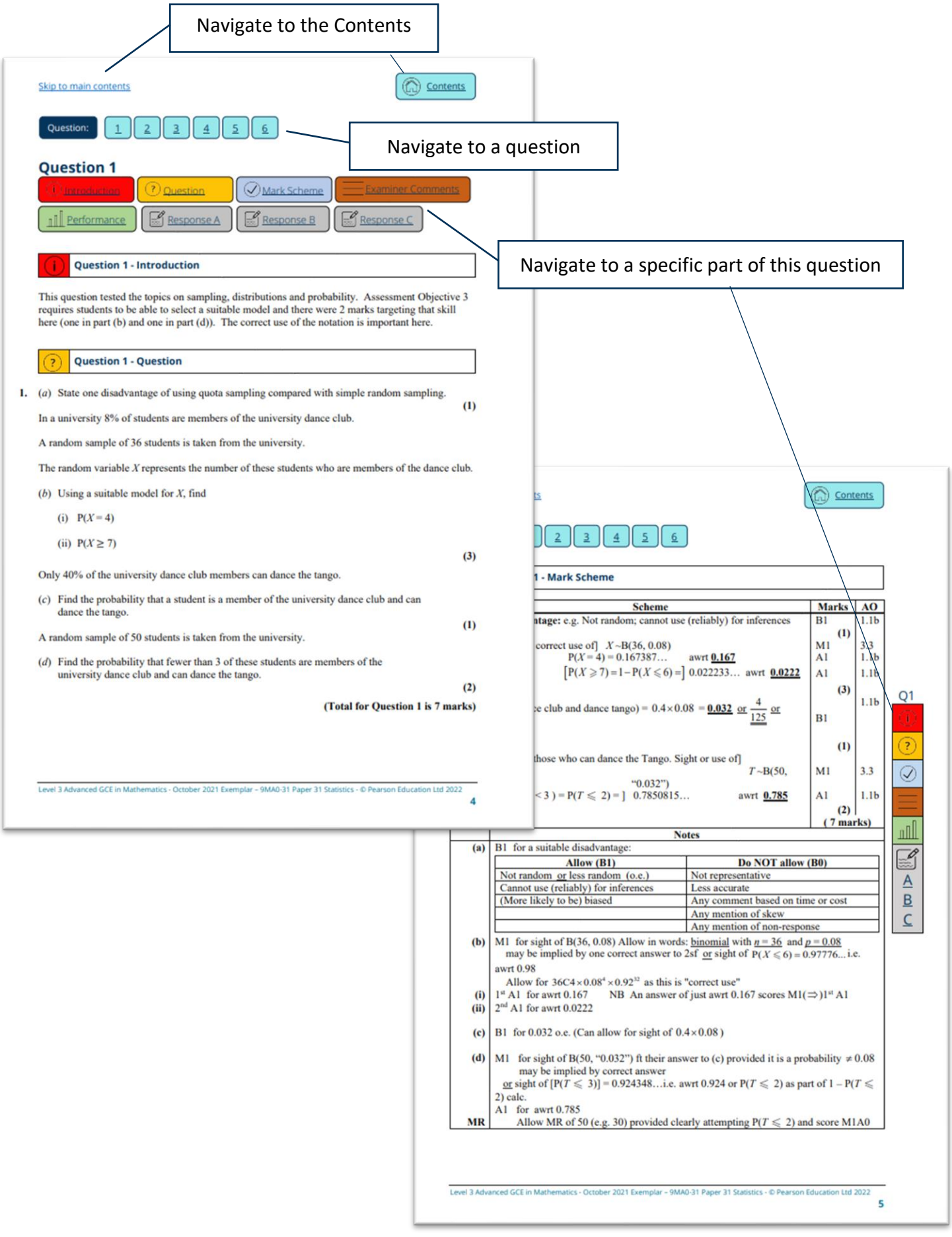
This document has been produced to support mathematics teachers delivering the GCE Statistics specification.

This document looks at questions from the 9ST0-01 A level Statistics Paper 1: Data and Probability June 2022 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.

* 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

This paper was accessible to most candidates, though most questions tended to also include some more challenging aspects that allowed more able candidates to demonstrate their knowledge. A common theme across the paper is use of context; in general it is not sufficient to simply state the assumptions required for a test to be valid, but rather how those apply to the situation in the question.

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

- give all explanations in context.
- look out for words shown in bold type.
- keep their working to more than 3 significant figures accuracy, throughout only rounding their final answer.
- use bullet points, each written in clear, specific, and concise sentences for explanation questions.

Question 1

Question | Mark Scheme | Examiner Comments

Performance | Response A | Response B | Response C

Question 1 - Question

1 The journey times of commuters for a selection of those travelling into two London train stations are given in **Figure 1** and **Figure 2**.

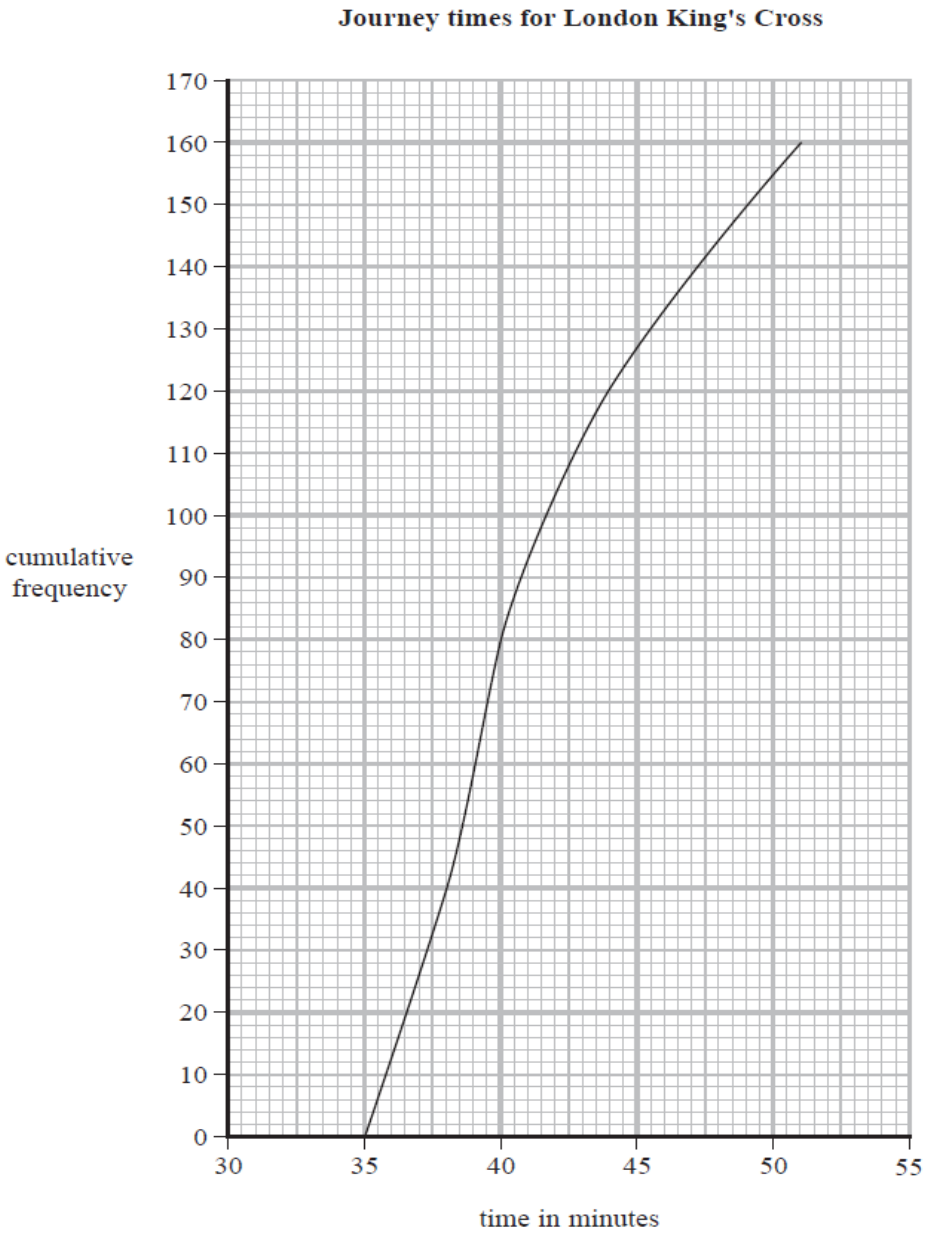


Figure 1

Journey time (min) for London Waterloo

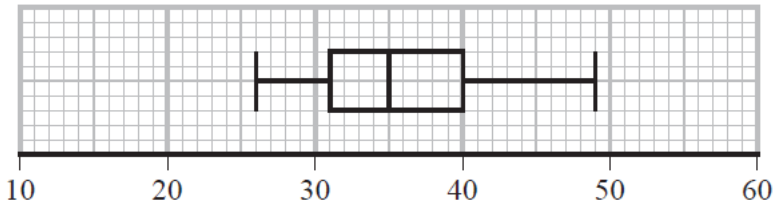


Figure 2

These times are the time taken from **leaving their house** to arriving at their destination station.

(a) Compare the average and spread of journey times for commuters travelling to these two stations.

(6)

These journeys are made up of two parts.

The first part of the total journey is the journey travelling from the commuter's home to their local station, where they take the train to London King's Cross or London Waterloo. This part of the journey may include waiting for this train to arrive.

The second part of the total journey is the journey by train to London King's Cross or London Waterloo.

All passengers in **Figure 1** took the 7:15 train from Stevenage to London King's Cross.

All passengers in **Figure 2** took the 7:45 train from Wimbledon to London Waterloo.

All the data is from the same day.

Lyra looks at the information in **Figure 1** and **Figure 2** and concludes that the train journey time from Wimbledon to London Waterloo is shorter than that from Stevenage to London King's Cross.

(b) State an assumption that Lyra has made in order for this conclusion to be valid.

(1)

(c) Identify **two** sources of variability for these commuters from Wimbledon to London Waterloo?

(2)

(Total for Question 1 is 9 marks)

Question 1 - Mark Scheme

Question	Scheme	Marks	AO	Notes
1(a)	Median from box plot=35	A1	1.1	
	IQR from box plot=9	A1	1.1	Or range=23
	Median from graph=40	A1	1.1	Or comment such as "obviously greater than 35 as it starts at 35"
	IQR from graph=6 or 7	A1	1.1	Or range=16
	So Kings Cross has a higher average journey time	E1	1.1	Must have context
	The spread is larger for Waterloo	E1	1.1	Must have context
1(b)	The average journey times for Wimbledon and Stevenage commuters from their home to the train are similar.	E1	2.1a	e.g. the distributions are similar.
1(c)	**Possible sources (not exhaustive)** Waiting time at the station. Distance lived from the station. Traffic during journey to the station. Walking speed.	E1, E1	2.1a, 2.1a	Must be related to the first part of the journey.
	Total	**9**		

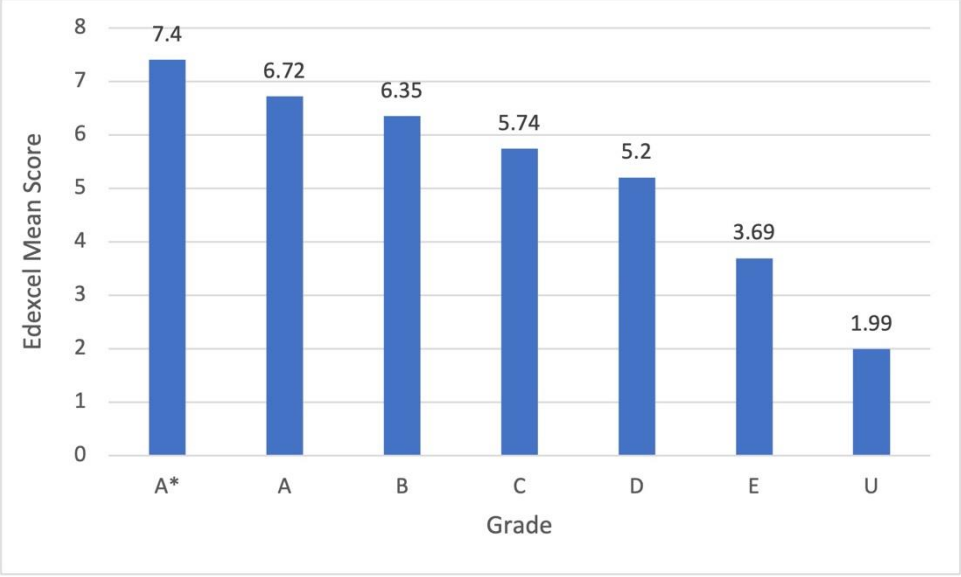
Q1
A
B
C

Question 1 - Examiner Comments

In part (a) candidates were generally able to obtain the median and interquartile range or range from the cumulative frequency diagram and box plot, though some candidates incorrectly used the maximum frequency shown on the y-axis as the total number. Some candidates incorrectly referred to these as the mean and standard deviation, neither of which can be obtained from the diagram. When asked to compare spread a large number of candidates were providing comments comparing maximum and minimum values, rather than either measure of spread. Candidates should be reminded that comparisons should be done in context, referring to the average journey times from Stevenage and Wimbledon, rather than simply the averages from Figure 1 and Figure 2.
Part (b) was difficult, and in part (c) a large number of students had not read all the information and gave comments about different train journeys. As all the commuters were on the same train from Wimbledon to Waterloo, there is no variability in that part of the journey.

Question 1 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade:							
			ALL	A*	A	B	C	D	E	U
5.64	9	63	5.64	7.40	6.72	6.35	5.74	5.20	3.69	1.99



Question 1 - Response A

1 The journey times of commuters for a selection of those travelling into two London train stations are given in **Figure 1** and **Figure 2**.

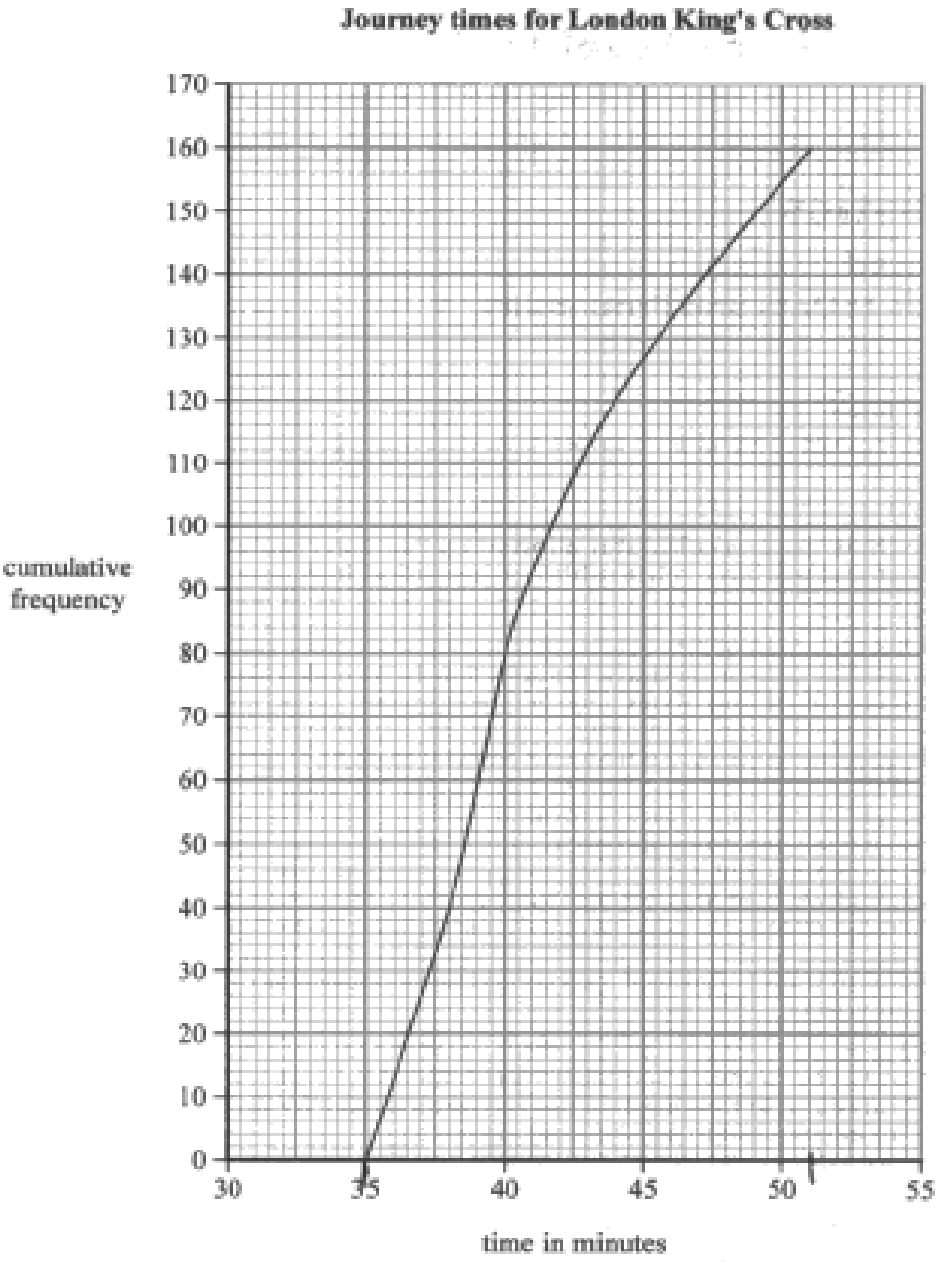


Figure 1

Journey time (min) for London Waterloo

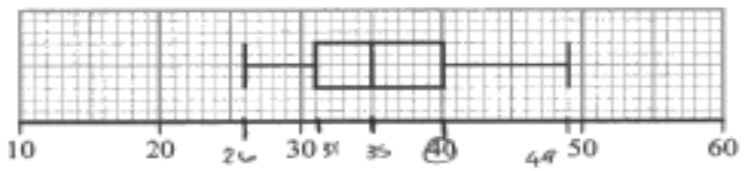


Fig 2
26 - 49 = 23
Fig 1
35 - 51 = 16

Figure 2

These times are the time taken from **leaving their house** to arriving at their destination station.

(a) Compare the average and spread of journey times for commuters travelling to these two stations.

(6)

- Figure 1 has a lower range than figure 2
- Figure 1 has a more even spread

(b) State an assumption that Lyra has made in order for this conclusion to be valid.

(1)

people have to be where they need to at similar times

(c) Identify **two** sources of variability for these commuters from Wimbledon to London Waterloo?

(2)

2 /9

Q1

Part (a)

A0A1A0A1: The ranges can be seen beside the diagram. Though 35 is labelled on the diagram it is not clear that this candidate knows that this is the median. There is no indication of the median for the Kings Cross journey times.

E0E0: While figure 1 does have a lower spread this comment is not contextualised.

Part (b)

E0: The time people need to be at certain places is not relevant to this question.

Part (c)

E0E0: The candidate did not attempt this part.

Question 1 - Response B

1 The journey times of commuters for a selection of those travelling into two London train stations are given in **Figure 1** and **Figure 2**.

Journey times for London King's Cross

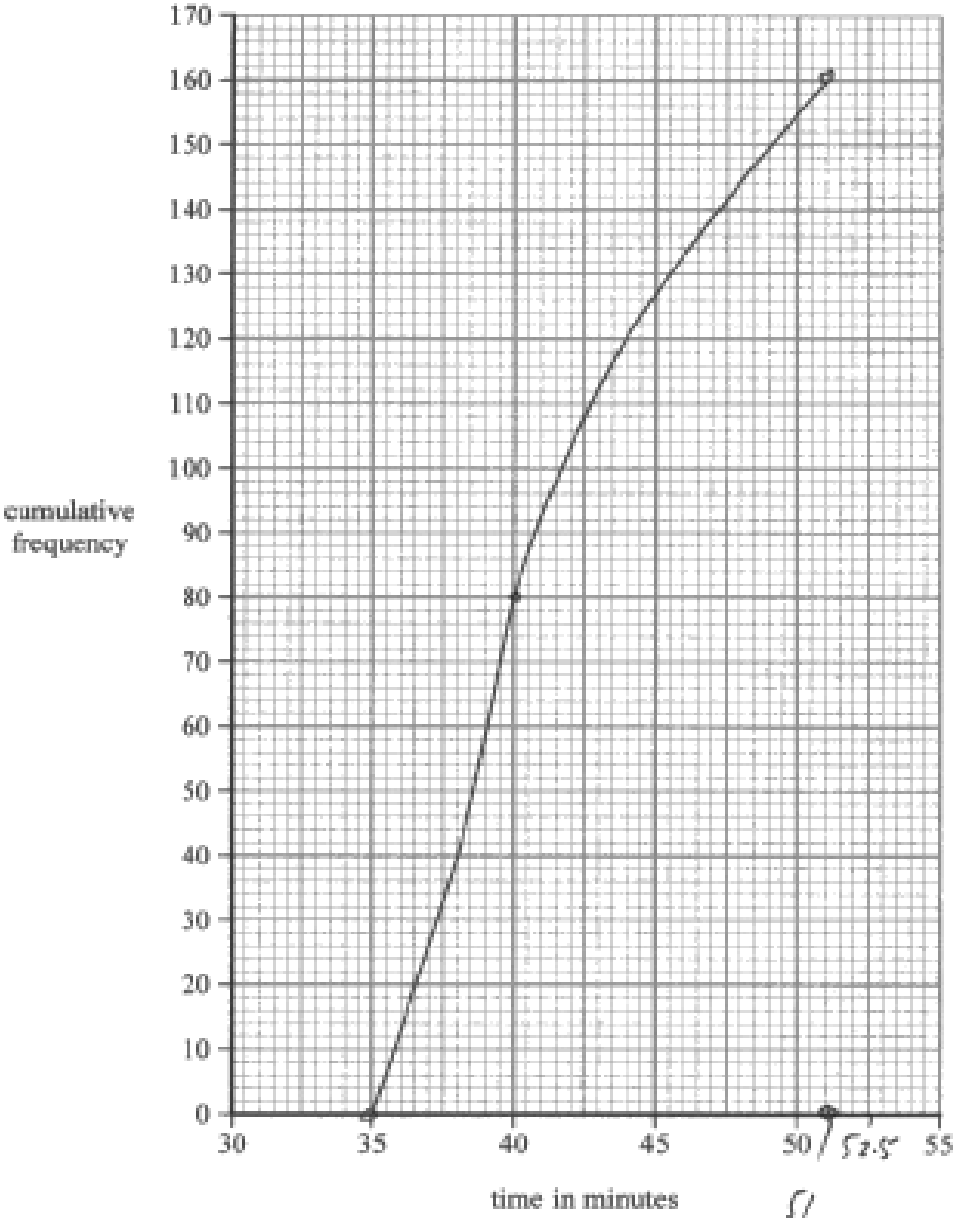


Figure 1

Journey time (min) for London Waterloo

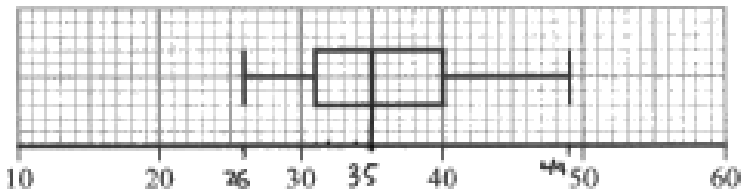


Figure 2

These times are the time taken from **leaving their house** to arriving at their destination station.

(a) Compare the average and spread of journey times for commuters travelling to these two stations.

(6)

- There is a smaller spread in the King's Cross journey times than there is for London Waterloo. (16 < 23), on average.
- On average, London's Kings Cross journey times are slower than for London Waterloo's journey times.
- On average, the King's Cross' fastest journey time (35m) is slower than the fastest Waterloo journey time (26m).
- London King's Cross average journey time is 43 compared to Waterloo's 35 minutes.
- London Waterloo has a larger interquartile range than King's Cross' data on journey times.
- The slowest journey at Waterloo recorded is at 49 mins, whereas it is 51 minutes for King's Cross (average).
- Both sets of journey times have more data ~~at~~ skewed towards longer journey times, on average.

Q1

A
B
C

(b) State an assumption that Lyra has made in order for this conclusion to be valid.

(1)

- Randomly sampled day of testing (all tested in 1 day).
- Independent of each other, the journey times i.e. no disruption for example a crash.
- All travel time is consistent inclusive of waiting for train to arrive.

(c) Identify **two** sources of variability for these commuters from Wimbledon to London Waterloo?

(2)

- May be less frequent trains timetabled from Wimbledon, to Waterloo, than from Wimbledon.

- Part of the journey may include waiting, which may give an increased time, or some may not wait, creating lower time

- Commuters at 7:15 AM to King's

(Total for Question 1 is 9 marks)

Costs may be incredibly different to arrive at 7:45 AM towards Waterloo; i.e. may be more tourists going towards Kings Cross who likely wait longer for example.

- May not be independently commuting of each other i.e. families or friends travel together.

7 /9

Q1

Part (a)

A1A1A1A1: The range and median for all four values clearly appear in the working.

E1E1: The average and spread are compared in context. It would been preferable to have stated the spread is larger, but specifically mentioning the interquartile range was condoned. Note that this candidate has written more than was needed; the question asked only asked for a comparison of the average and spread, yet the candidates has given far more comments.

Part (b)

E0: The candidate is listed common conditions from statistics which were not relevant to the scenario.

Part (c)

E1E0: More than two possibilities are mentioned here. It is not recommended to list more than requested by the question as incorrect answers will be penalised even if there are two correct ones. In this situation the mark was awarded for waiting, as different commuters may arrive at a station at different times.

Question 1 - Response C

1 The journey times of commuters for a selection of those travelling into two London train stations are given in **Figure 1** and **Figure 2**.

Journey times for London King's Cross

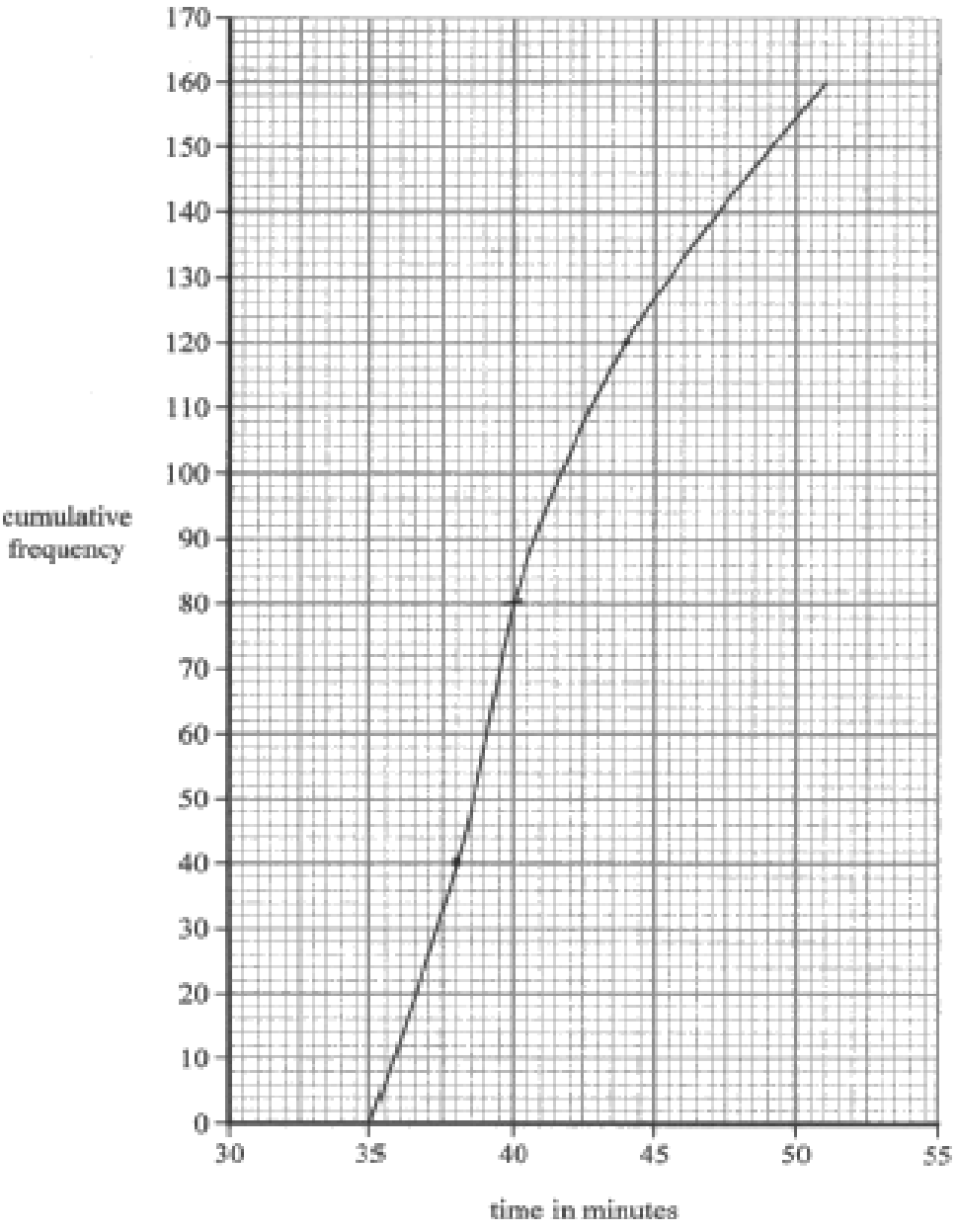
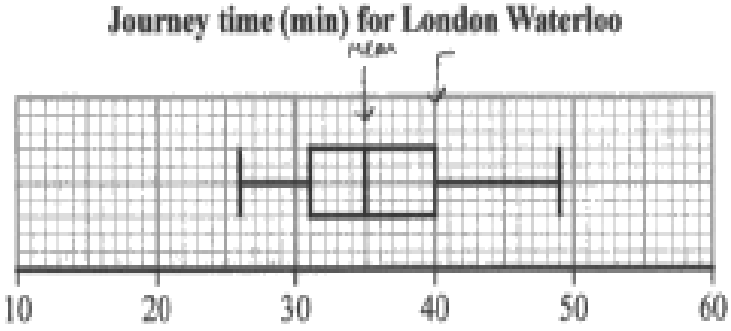


Figure 1

44 − 38 = 6



40 - 31 = 9.

Figure 2

These times are the time taken from **leaving their house** to arriving at their destination station.

IQR

(a) Compare the average and spread of journey times for commuters travelling to these two stations.

(6)

The Range for London waterloo's line ~~is~~ is 23 whereas London King's cross' range is 1~~6~~, meaning that the journey time is more spread out and more inconsistent at London Waterloo.

London King's cross' Interquartile range is 6 whereas London waterloo's range is 9, therefore meaning that London king's cross' journey times are more consistent as the data is less spread out.

London waterloo's journey time's mean is 35 whereas London King's cross' mean is 40, meaning that on average London king's cross' journey time takes longer with an average difference of 5 minutes.

Q1

A
B
C

(b) State an assumption that Lyra has made in order for this conclusion to be valid.

(1)

The time it takes from the journey of the home to the station is the same time and that both trains arrive at the correct, scheduled time.

(c) Identify **two** sources of variability for these commuters from Wimbledon to London Waterloo?

(2)

The journey from their home to the station.

Number of people boarding the train

8 /9

Q1

Part (a)
A1A1A1A1: They give the median, range and interquartile range.
E1E1: These are compared contextually.

Part (b)
E1: They have identified the correct condition.

Part (c)
E1E0: The journey time for each commuter is certainly different, but the number of people boarding the train is the same for all of them and this will have the same affect for all commuters.

Question 2

Question | Mark Scheme | Examiner Comments

Performance | Response A | Response B | Response C

Question 2 - Question

2 Alana is researching show dogs.

Show dogs have a desired range of heights.

The maximum and minimum desired male and female heights, in inches, for a random sample of breeds of show dog are shown in **Figure 3**.

Alana plots the minimum desired male heights and minimum desired female heights in **Figure 4**.

		Male		Female	
		Minimum desired height	Maximum desired height	Minimum desired height	Maximum desired height
Breed	Affenpinscher	9	11.5	9	11.5
	Afghan Hound	26	28	24	26
	Akita	26	28	24	26
	American Eskimo Dog	9	19	9	19
	Australian Shepherd	20	23	18	21
	Bearded Collie	21	22	20	21
	Beauceron	25.5	27.5	24	26.5
	Belgian Malinois	24	26	22	24
	Bernese Mountain Dog	25	27.5	23	26
	Black Russian Terrier	27	30	24	26.5
	Cattle Dog	18	20	17	19
	Chesapeake Bay Retriever	21	24	23	26
	Chinese Shar-Pei	18	20	18	20
	Doberman Pinscher	26	28	24	26
	Flat-Coated Retriever	23	24.5	22	23.5
	Foxhound	22	25	21	24
	Greater Swiss Mountain Dog	25.5	28.5	23.5	27
	Irish Red and White Setter	24.5	26	22.5	24
	Irish Wolfhound	32	No max	30	No max
	Italian Greyhound	13	15	13	15
	Keeshound	17	19	16	18
	Neapolitan Mastiff	26	31	24	29
	Pointer	25	28	23	26
	Shetland Sheepdog	13	16	13	16
	Staffordshire Terrier	18	19	17	18
	Water Spaniel	15	18	15	18
	Welsh Springer Spaniel	18	19	17	18
	Wirehaired Pointing Griffon	22	24	20	22

[Source: American Kennel Club Breed Statistics]

Figure 3

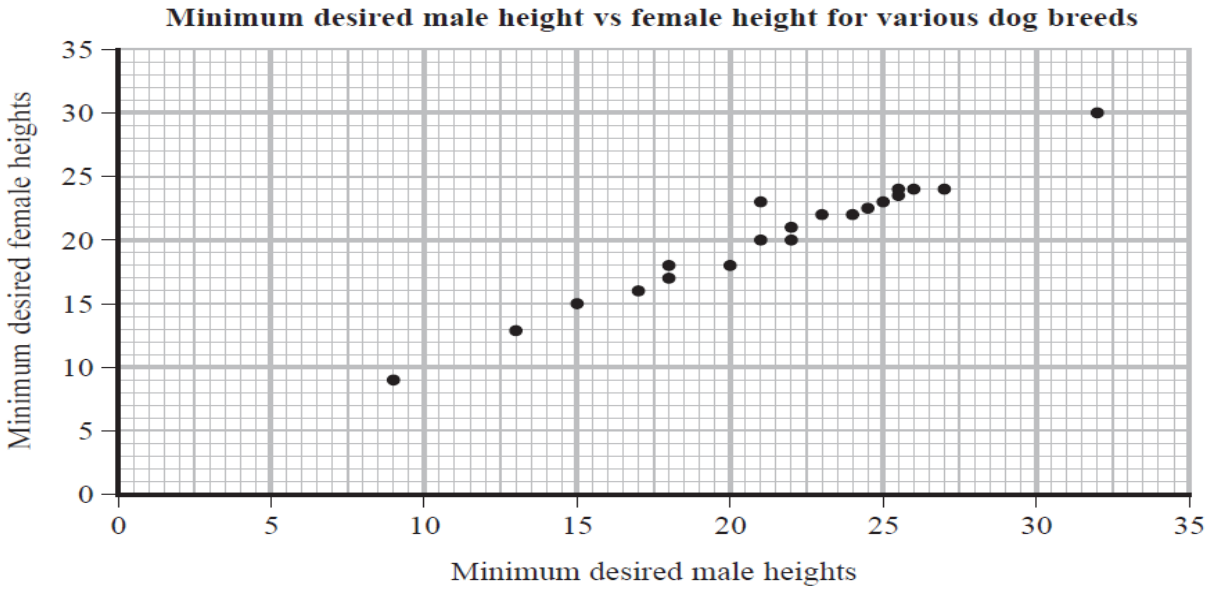


Figure 4

(a) Explain why Alana might have chosen to plot her data before calculating a correlation coefficient.

(2)

After examining **Figure 4** Alana realises that one of the dog breeds has had the male and female columns recorded the wrong way round.

(b) State, with a reason, which breed you believe has been recorded incorrectly.

(2)

The PMCC between the minimum desired height for males and females was calculated as r = 0.988

Alana says that the PMCC will be higher if the heights were measured in centimetres.

(c) Explain why Alana is incorrect.

(1)

Alana edits the data, swapping the values she realised had been recorded incorrectly.

(d) How would you expect the PMCC of her corrected data to compare to the PMCC of the original data, r = 0.988?

(1)

Alana states that it would be impossible for there to be a negative correlation between the **minimum desired female height** and **maximum desired female height** for all breeds of show dog.

(e) Discuss Alana's statement.

You should include reasoned comments on whether or not you believe she is correct.

(4)

(Total for Question 2 is 10 marks)

Question 2 - Mark Scheme

Question	Scheme	Marks	AO	Notes
2(a)	To check if the relationship is **linear** To check if correlation is positive or negative To see if the PMCC is an appropriate measure to use To get a rough idea of the PMCC before calculation as a check To check for anomalies To see if there is only one trend Any 2 reasons	E1,E1	1.1, 1.1	Condone type of correlation
2(b)	Chesapeake Bay Retriever As it doesn't fit trend of the other data on the **graph** or As it is the only breed with male lower than female in the **table**	E1 E1	1.1 1.1	
2(c)	PMCC is independent of units	B1	2.1b	oe e.g. the relationship
2(d)	It would be a value bigger than 0.988 but less than 1.	B1	2.1b	

2(e)	**Statements in support** Certain breeds of dogs may have a large minimum height and a small maximum height **Statements against** Breeds of show dog with a larger minimum desired height would be larger dogs in general. You would expect these to have a larger desired max height also.			

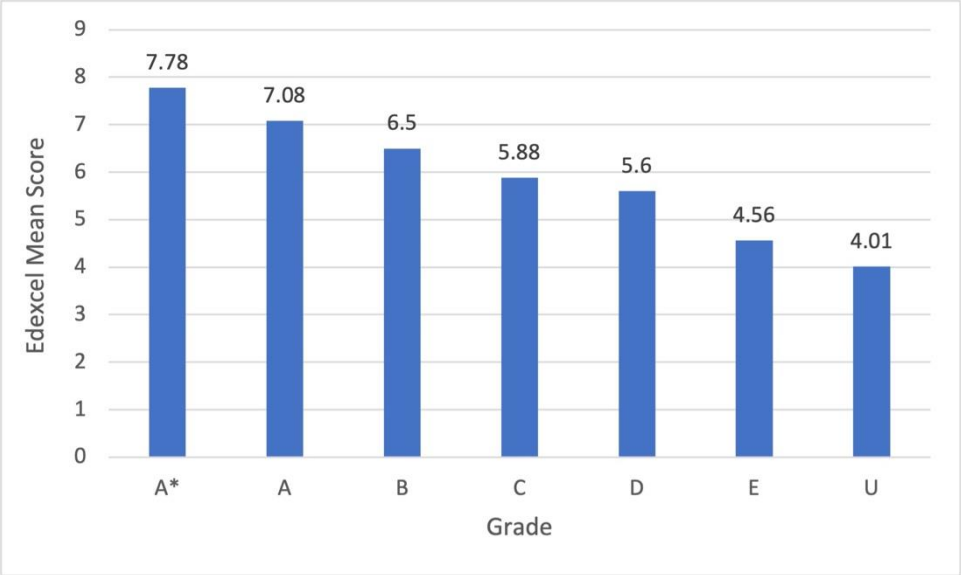
	This would require the maximum heights to decrease as minimum heights increase which does not happen in our sample Sample has a positive correlation Minimum desired height can't be higher than maximum desired height which could happen with a negative correlation	E1,E1, E1	2.1a	Each of up to 3 statements (may include a mix of support/against)
	Alana is **incorrect**	E1	3.1b	Awarded for overall disagreement with Alana with attempt at a reason SC supporting statement and Alana is correct scores E1E0E0E1
	Total	**10**		

Question 2 - Examiner Comments

In part (a), many candidates only gave only one reason. The number of marks on a question is a good indicator of the depth that should be given when answering.
Candidates generally answered parts (b), (c) and (d) well, though many struggled with (e). It is important to note the difference between what is **impossible** and what is simply **unlikely** or unexpected; although a negative correlation in the population is extremely unlikely, especially when compared to the given data, it is not impossible and hence Alana is incorrect. It was encouraging to read responses from the most able candidates who indeed realised this.

Question 2 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade:							
			ALL	A*	A	B	C	D	E	U
5.98	10	60	5.98	7.78	7.08	6.50	5.88	5.60	4.56	4.01



Question 2 - Response A

(a) Explain why Alana might have chosen to plot her data before calculating a correlation coefficient.

(2)

So she is able to compare her figure to the graph and check on see if the show a similar result. ~~her~~ Making it slightly more reliable

After examining **Figure 4** Alana realises that one of the dog breeds has had the male and female columns recorded the wrong way round.

(b) State, with a reason, which breed you believe has been recorded incorrectly.

(2)

Irish Wolfhound as there is no max desired height suggesting a biggest Irish wolfhound would be acceptable.

The PMCC between the minimum desired height for males and females was calculated as $r = 0.988$

Alana says that the PMCC will be higher if the heights were measured in centimetres.

(c) Explain why Alana is incorrect.

(1)

All values will still stay proportionate to each other

Alana edits the data, swapping the values she realised had been recorded incorrectly.

(d) How would you expect the PMCC of her corrected data to compare to the PMCC of the original data, $r = 0.988$?

(1)

It would ~~decrease slightly~~, it would have no effect, be the same

Q2

A B C

Alana states that it would be impossible for there to be a negative correlation between the **minimum desired female height** and **maximum desired female height** for all breeds of show dog.

(e) Discuss Alana's statement.

You should include reasoned comments on whether or not you believe she is correct.

(4)

Alana could be correct as it is impossible to have a minimum height be above a maximum height however, there could be inconsistencies in the size of class for desired height.

3 /10

Part (a)

E0E0: This candidate has suggested that Alana might want to plot her data so she can compare the plotted data to her original data. This would not aid her in analysing the data she has.

Part (b)

E0E0: The incorrect breed has been identified. The graph was for minimum heights, not maximum, so the lack of maximum heights is not an issue.

Part (c)

B1: This statement explains why the PMCC is independent of units, so is correct.

Part (d)

B0: The PMCC would increase, so the candidate is incorrect.

Part (e)

E1E0E0E1: This candidate received the special case marks. They are correct that it is impossible for a minimum height to be above a maximum height which is correct.

Question 2 - Response B

(a) Explain why Alana might have chosen to plot her data before calculating a correlation coefficient.

(2)

To see if the data was linear or not. Can only carry out a product moment correlation coefficient if the data is linear.

After examining **Figure 4** Alana realises that one of the dog breeds has had the male and female columns recorded the wrong way round.

(b) State, with a reason, which breed you believe has been recorded incorrectly.

(2)

The Chesapeake Bay Retriever as this is the only breed of dog where the female exceeds the male minimum and maximum desired height. It seems like an outlier to the rest of the female breeds.

The PMCC between the minimum desired height for males and females was calculated as $r = 0.988$.

Alana says that the PMCC will be higher if the heights were measured in centimetres.

(c) Explain why Alana is incorrect.

(1)

The correlation would be different using a different measurement as the trend would be the same.

Alana edits the data, swapping the values she realised had been recorded incorrectly.

(d) How would you expect the PMCC of her corrected data to compare to the PMCC of the original data, $r = 0.988$?

(1)

The corrected PMCC would be greater because the outlier would be corrected to follow the very strong positive correlation.

Alana states that it would be impossible for there to be a negative correlation between the **minimum desired female height** and **maximum desired female height** for all breeds of show dog.

(e) Discuss Alana's statement.

You should include reasoned comments on whether or not you believe she is correct.

(4)

~~because~~ ~~As the minimum desired female height increases~~

If it was a negative correlation then this could be for two reasons:

- As the minimum desired height increases and the maximum desired height ~~increases~~ decreases, it could be show dogs for a small breed competition.
- As the minimum desired height decreases and the maximum desired height increases, it could be show dogs for a large breed competition.

I don't believe Alana's statement that it would be impossible for there to be a negative correlation is correct.

5 /10

Q2

A B C

Part (a)

E1E0: The mark for checking if linear has been awarded. While the candidate is correct that we should only use the PMCC for linear data that is the same point they have just made, so no further mark is awarded.

Part (b)

E1E1: Correctly identified with the correct reason.

Part (c)

B0: The correlation would be the same, so the candidate is incorrect.

Part (d)

B1: The correlation would be greater. The candidate has not mentioned it should still be less than or equal to 1 but we condoned this.

Part (e)

E0E0E0: The candidate has attempted to explain what a negative correlation is but has not explained why this is unlikely in the context.

E1: Awarded for disagreeing with Alana. The candidate has attempted to reason, and though no marks were awarded for the first 3 marks, they have made an attempt rather than guessing.

Question 2 - Response C

(a) Explain why Alana might have chosen to plot her data before calculating a correlation coefficient.

(2)

From a graph you can visibly see wether the correlation is positive or negative So you can easily spot a miss calculation and to see wether the shape of the data makes it valid to is linear otherwise a correlation coefficient maybe deemed invalid.

After examining **Figure 4** Alana realises that one of the dog breeds has had the male and female columns recorded the wrong way round.

(b) State, with a reason, which breed you believe has been recorded incorrectly.

(2)

The Flat-Coated Retriever as the minimum male height is 23 but on the graph it reads the Female height as 23. Also The female height is 22 but on the graph reads the male height as 22.
Also the dot sticks out above the data.

The PMCC between the minimum desired height for males and females was calculated as r = 0.988

Alana says that the PMCC will be higher if the heights were measured in centimetres.

(c) Explain why Alana is incorrect.

(1)

The PMCC is a measure of wether the data has a positive or negitive correlation and does not matter wether the actual data all increase or decrease in numerical size by the same amount.

Alana edits the data, swapping the values she realised had been recorded incorrectly.

(d) How would you expect the PMCC of her corrected data to compare to the PMCC of the original data, r = 0.988?

(1)

The PMCC would increase as the data the was wrong would move closer to the other values.

Alana states that it would be impossible for there to be a negative correlation between the **minimum desired female height** and **maximum desired female height** for all breeds of show dog.

(e) Discuss Alana's statement.

You should include reasoned comments on whether or not you believe she is correct.

(4)

I beleive she is ~~wrong~~ as ~~however~~ if another dog breeds minimum desired female height was to be small and the minimum desired male height was to be large the this would heavily skew the data potentially making the PMCC negative.
But if all other dog breeds are similarly judged then the data ~~can be~~ is very likly to ~~be~~ have a ~~po~~ positve correlation. I wouldn't say impossible tho as if just two or ~~three~~ three are judged differently then the data can be skew and change it to a negative correlation or close to 0 where it may be seen as no correlation.

8 /10

Q2

Part (a)
E1E1: Discusses seeing if positive or negative, and linear or non-linear.

Part (b)
E0E1: Incorrectly identified the breed but identified the transposition of values.

Part (c)
B1: While they seem to be discussing an addition or subtraction, they broadly have the right idea.

Part (d)
B1: Correctly identifies increase in PMCC.

Part (e)
E1E1E0: Awarded a mark for discussing it is not impossible and only unlikely due to the correlation of our sample and explains how a negative PMCC could occur.
E1: Awarded for disagreeing with Alana.

Question 3

Question | Mark Scheme | Examiner Comments | Performance | Response A | Response B | Response C

Question 3 - Question

3 In 2011 Boston College psychologists, Angelina Hawley-Dolan and Ellen Winner, investigated whether paintings of professional abstract artists could be differentiated from those of a five-year-old child.

Leo, a psychology student, decides to do his own version of this experiment.

Suggest how Leo might set up his experiment.

You should include comments about

- how he will source the data
- the data he should collect
- how he could minimise bias.

(7)

(Total for Question 3 is 7 marks)

Question 3 - Mark Scheme

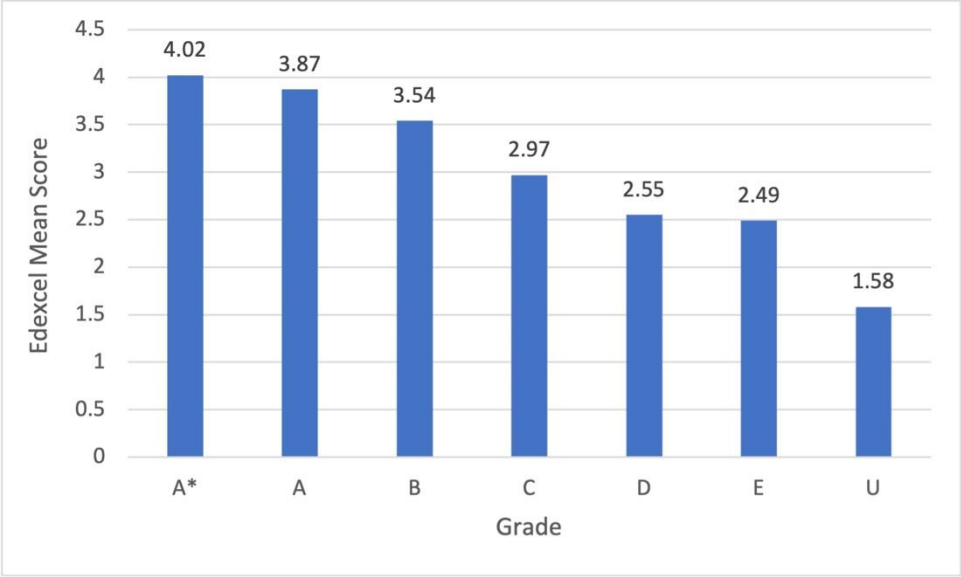
Question	Scheme	Marks	AO	Notes
3	Leo should choose a **large sample** of participants His sample of participants should be **randomly selected** He should show them one painting at a time and ask if they think it was done by a professional or by a child Mention of blocking generally Mention of a sensible specific blocking factor e.g. background in art Leo should collect paintings from several different artists/several different children Leo should collect an equal number of paintings from artists and children He could try to pair similar paintings from a child/professional artist Blind/double blind trial mentioned in context He should remove identifying information such as signatures from the artists' work He should keep participants separate so they do not discuss Use same paintings for all participants	E1, E1, E1, E1, E1, E1, E1	3.1a	Condone $n \geq 30$ or implied by context Or He should show them two paintings at a time and ask which they think is by a professional e.g. randomised block design Consideration of the bias caused by the painters One mark for each comment up to 7
	Total	**7**		

Question 3 - Examiner Comments

Candidates are advised to use bullet points when answering a question like this to be certain they have made enough points. There are some straightforward ideas that candidates should be familiar with when it comes to planning an experiment like using a large sample, which were not always mentioned. A large number of candidates went into great of detail regarding how to generate a random sample, which was not the question that was asked. Many candidates did not address all three points, with some not mentioning what his sample should be doing or suggested that Leo should ask people to rate how good they thought the paintings were rather than asking them which they thought was painted by a professional. There were, however, some excellent responses seen.

Question 3 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade:							
			ALL	A*	A	B	C	D	E	U
3.07	7	44	3.07	4.02	3.87	3.54	2.97	2.55	2.49	1.58



Question 3 - Response A

3 In 2011 Boston College psychologists, Angelina Hawley-Dolan and Ellen Winner, investigated whether paintings of professional abstract artists could be differentiated from those of a five-year-old child.

Leo, a psychology student, decides to do his own version of this experiment.

Suggest how Leo might set up his experiment.

You should include comments about

- how he will source the data
- the data he should collect
- how he could minimise bias.

(7)

He could input all his abstract art into a random generator and have 5 selected.

He could then repeat this with the 5 year olds paintings and select 5.

~~A~~ Ignoring ~~one~~ any repeats.

This data could be collected from the internet for the abstract and local 5 year olds paintings.

He could then select 15 people at random who could then ~~go~~ try to point out which ones are famous.

A Sign test could then be used to see how many people thought 5 year olds where better / the same ~~as~~ as famous abstract artists.

3 /7

E1: Selecting 5 professionals and 5 children is equal numbers of each.
E1: Selecting his sample of 15 at random.
E1: Asking them to decide which they think is famous.
E0E0E0E0: No further valid points.

Question 3 - Response B

3 In 2011 Boston College psychologists, Angelina Hawley-Dolan and Ellen Winner, investigated whether paintings of professional abstract artists could be differentiated from those of a five-year-old child.

Leo, a psychology student, decides to do his own version of this experiment.

Suggest how Leo might set up his experiment.

You should include comments about

- how he will source the data
- the data he should collect
- how he could minimise bias.

Sample of 50
Number each 1 to 50
Random number generator
Pick 5 from each group
Ask group to say which is which

(7)

Leo should first get a sample of 50 professional abstract artists ~~and 50~~ paintings and 50 50 paintings from five year old children. Leo should then number each of professional abstract artist painting from 1 to 50 and number the five year old children's painting from 1 to 50. Then get a random number generator and pick 5 from each group. ~~Next Leo should get someone else to~~ Leo would be able to minimise bias if he lets someone else collect the results from another sample of 50 people. The sample of 50 people should pick which pictures are from a professional artist and a 5 year old child. Leo should do this as it creates a double blind test as the person collecting the data and the sample should not know which are is which. Then he should put the results into a table and see the correlation between which one people thought was done by a professional or a child.

Q3

A B C

4 /7

E1: Equal numbers of paintings by children and professional artists.
E1: Sample of 50 people implies large sample.
E1: Asking sample to decide which is professional and which is not.
E1: Mentions double blind trial in context.
E0E0E0: No further valid points.

Question 3 - Response C

3 In 2011 Boston College psychologists, Angelina Hawley-Dolan and Ellen Winner, investigated whether paintings of professional abstract artists could be differentiated from those of a five-year-old child.

Leo, a psychology student, decides to do his own version of this experiment.

Suggest how Leo might set up his experiment.

You should include comments about

- how he will source the data
- the data he should collect
- how he could minimise bias

(7)

Professional Painting

Child Paint	1	2	3
1	A C	BF	LQ
2	D E	I J	KO
3	H JH	MN	PR

Source data :

- Collect a random sample of people.
- Gather a sampling frame, e.g. all boston residents on a local council list.
- Use simple random sampling and a random number generater.
- Sufficientiy larger sample size.

Type of Data

- Paired data - ask samplee which one they believe to be professional and which is the child's art.
- Quantitative data - so use sign test.

Minimise Bias

- Double blind trial, neither samplee or investigator knows which painting is which.
- Blocking factor - use same type of paint and same size canvas. To minimise effect size.
- Use of randomisation as to which two painting a samplee see. Randomised Block Design.
  - see table above.
    A, B, C are samplees.
- use three different child and three different professional paintings.

7 /7

Q3

A B C

E1: Random Sample specifically mentioned.
E1: Sampling frame implies large sample, and large sample also explicitly mentioned.
E1: Ask which painting is by a child and which by a professional.
E1: Double blind trial mentioned in context.
E1: Blocking mentioned.
E1: Specific blocking factor mentioned.
E1: Same number of paintings by child and professional mentioned (3 of each).

Question 4

Question | Mark Scheme | Examiner Comments | Performance | Response A | Response B | Response C

Question 4 - Question

4 Shai has created a video game. She makes her game free to play, with players able to buy gems to spend in the game for bonuses.

She creates a model to predict the income that she can expect from her game.

Shai's data shows that the majority of players do not spend any money on gems.

A small number of players, known as 'whales', make up the majority of the revenue for games by buying the gems.

[Source: https://deltadna.com/blog/how-whales-spend/]

(a) Give a reason why the normal distribution is unlikely to be a suitable model for the amount a randomly selected player spends in a given month.

(1)

Shai instead chooses to model the spend, in pounds, of a randomly selected whale in a given month. She uses the distribution $W \sim N(50, 100)$

She assumes that the amount a whale spends each month is independent from their spend in previous months.

(b) Find $P(W < 40)$

(1)

(c) Using Shai's model, calculate the probability that a randomly selected whale has a mean spend of less than £40 a month over a year.

(2)

(d) Explain why the probability calculated in part (*c*) **must** be smaller than the probability calculated in part (b).

(1)

(e) Explain why $P(W \leq 40)$ and $P(W < 40)$ are equal.

(1)

(f) Calculate the probability that a randomly selected whale spends **more** than £40 in each of four consecutive months.

(2)

(g) Calculate the probability that a randomly selected whale spends more than £40 in at least 8 of the months in a one-year period.

(3)

(h) Calculate the probability that a randomly selected whale spends less than £40 in a month that they spent more than £30

(3)

In Shai's game, a bundle of gems costs £5. The game also offers unlimited gems for any player spending £60 a month.

(*i*) Discuss the appropriateness or otherwise of the normal distribution to model the spend of a whale in light of this information.

(3)

(Total for Question 4 is 17 marks)

Question 4 - Mark Scheme

Question	Scheme	Marks	AO	Notes
4(a)	Players with a spend of 0 give the data a heavy skew Revenue cannot be negative and normal is unbounded	E1	2.1a	Or reverse with whales to right
4(b)	0.1587	B1	1.2	awrt 0.159
4(c)	$\overline{X} \sim N\left(50, \frac{100}{12}\right)$ $P(\overline{X} < 40) = 0.0003$ **Alternative** $X_1 + \cdots + X_{12} \sim N(600, 1200)$ $P(X_1 + \cdots + X_{12} < 480) = 0.0003$	M1 A1 (M1) (A1)	1.2 1.2	Divides variance by 12 or standard deviation by $\sqrt{12}$ awrt
4(d)	Because the standard deviation of $\overline{X}$ is smaller than that of X	E1	2.1b	Or 40 is more standard deviations from the mean
4(e)	Because it's a normal model Because $P(W = 40) = 0$ Because W is continuous Any of these	E1	2.1b	
4(f)	$(1 - "0.1587")^4$ 0.50096	M1 A1	1.2 1.2	$(1 - 'their\ (b)')^4$ awrt 0.5
4(g)	$X \sim B(12, 0.8413)$	M1	1.2	Binomial with n=12 seen or used
		M1	1.2	$p = 1 - 'their\ (b)'$

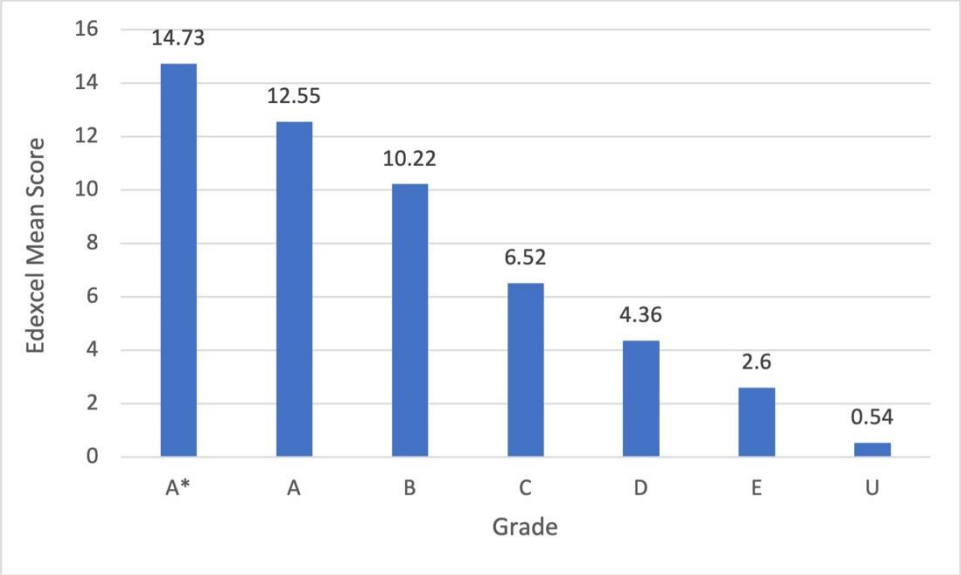
	$P(X \geq 8) = (1 - P(X \leq 7))$ $= 0.970$	A1	1.2	awfw 0.969-0.972
4(h)	$P(30 < X < 40) = 0.1359$	B1	1.2	
	$\frac{P(30 < X < 40)}{P(X > 30)} = \frac{0.1359}{0.9772}$	M1	1.2	Divides by 0.9772
	0.139	A1	1.2	
4(i)	**Not appropriate**			
	The actual distribution would be discrete not continuous			
	The normal distribution would not have an upper limit of £60			Or lower limit of £0
	Whales may be inclined to spend £60 meaning the distribution is skew			
	Appropriate			
	Could model using e.g. $P(X = 5) = P(2.5 < W < 7.5)$			Continuity correction suggested
	Conclusion			
	Conclusion that normal may be appropriate/inappropriate			Only valid if agrees with at least one of their reasons
	Three comments	E1, E1, E1	2.1a	
	Total	17		

Question 4 - Examiner Comments

This question had several parts that most candidates did well on – (a), (b), (e) and (f) were well answered. Parts (c), (g) and (h) demonstrated some of the harder aspects of the normal distribution – sample mean, a hidden binomial and conditional probability. Candidates would do well to be prepared for these, as standard probability calculations are straightforward due to the calculators available and therefore do not gain a large number of marks. In part (i) many candidates identified either the discrete nature of the data, or the upper bound as a reason the normal distribution was not appropriate but did not provide sufficient comments for 3 marks.

Question 4 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade:							
			ALL	A*	A	B	C	D	E	U
7.46	17	44	7.46	14.73	12.55	10.22	6.52	4.36	2.60	0.54



Question 4 - Response A

4 Shai has created a video game. She makes her game free to play, with players able to buy gems to spend in the game for bonuses.

She creates a model to predict the income that she can expect from her game.

Shai's data shows that the majority of players do not spend any money on gems.

A small number of players, known as 'whales', make up the majority of the revenue for games by buying the gems.

[Source: https://deltadna.com/blog/how-whales-spend/]

(a) Give a reason why the normal distribution is unlikely to be a suitable model for the amount a randomly selected player spends in a given month.

(1)

The data is not symmetrical because more people don't ~~buy gems~~ buy gems compared to the people who do buy gems.

Shai instead chooses to model the spend, in pounds, of a randomly selected whale in a given month. She uses the distribution $W \sim N(50, 100)$

no

She assumes that the amount a whale spends each month is independent from their spend in previous months.

(b) Find $P(W < 40)$

(1)

P = 0.4602

= 0.46

(c) Using Shai's model, calculate the probability that a randomly selected whale has a mean spend of less than £40 a month over a year.

(2)

P (W < 40) = 0.46

$(0.46)^{12}$ = 0.00008898

(d) Explain why the probability calculated in part (c) **must** be smaller than the probability calculated in part (b).

(1)

The average money spent is £50 as shown by M.

Q4

A B C

(e) Explain why $P(W \leqslant 40)$ and $P(W < 40)$ are equal.

(1)

There might not be an option to spend £41.

(f) Calculate the probability that a randomly selected whale spends **more** than £40 in each of four consecutive months.

(2)

$P(W > 40) = 0.53982783$

$(0.53982783)^4 = 0.0849$

(g) Calculate the probability that a randomly selected whale spends more than £40 in at least 8 of the months in a one-year period.

(3)

$P(W > 40) = 0.53982783$

$(0.53982783)^{8/12} = 0.663$

(h) Calculate the probability that a randomly selected whale spends less than £40 in a month that they spent more than £30

(3)

$P(30 < W < 40)$

$= 0.0394$

In Shai's game, a bundle of gems costs £5. The game also offers unlimited gems for any player spending £60 a month.

(i) Discuss the appropriateness or otherwise of the normal distribution to model the spend of a whale in light of this information.

(3)

Data may not be symmetrical as more people might want ~~a~~ unlimited gems compared to a bundle of gems. This means that a normal distribution is not appropriate.

4 /17

Part (a)
E1: Not symmetrical equivalent to skew.

Part (b)
B0: Incorrect answer.

Part (c)
M0 A0: Incorrect answer.

Part (d)
E0: The response does not answer the question asked.

Part (e)
E0: £41 is not relevant to the question, however, is a misunderstanding of the questions.

Part (f)
M1 A0: They calculated $(1-\text{their}\,b)^4$ but do not get an accuracy mark as the answer is incorrect.

Part (g)
M0 M0 A0: No binomial distribution used, so even though we see 0.539 being used it is not being used as *p*.

Part (h)
B0 M0 A0: Incorrect value.

Part (i)
E1: Not symmetrical explained, this is equivalent to skew.
E1: Concluded normal may not be appropriate.
E0: No further points made.

Question 4 - Response B

4 Shai has created a video game. She makes her game free to play, with players able to buy gems to spend in the game for bonuses.

She creates a model to predict the income that she can expect from her game.

Shai's data shows that the majority of players do not spend any money on gems.

A small number of players, known as 'whales', make up the majority of the revenue for games by buying the gems.

[Source: https://deltadna.com/blog/how-whales-spend/]

(a) Give a reason why the normal distribution is unlikely to be a suitable model for the amount a randomly selected player spends in a given month.

(1)

the ~~median~~ mean spending is unlikely to be the same as the ~~mean~~ mode as majority of players do not spend anything making the mode equal 0 however the mean is impacted by the "whales"

Shai instead chooses to model the spend, in pounds, of a randomly selected whale in a given month. She uses the distribution $W \sim N(50, 100)$

She assumes that the amount a whale spends each month is independent from their spend in previous months.

(b) Find $P(W < 40)$

(1)

$P(W<40) = 0.1587$

(c) Using Shai's model, calculate the probability that a randomly selected whale has a mean spend of less than £40 a month over a year.

(2)

$P(12W<480) =$

$12(50, 600, 1200) = 2.660 \times 10^{-4}$

(d) Explain why the probability calculated in part (c) **must** be smaller than the probability calculated in part (b).

(1)

~~As the cumulative buying patterns~~ as there is a ~~smaller variance~~ longer time frame the whales are more likely to ~~spend more~~ have a 15% probability in year

Q4

A B C

(e) Explain why $P(W \leqslant 40)$ and $P(W < 40)$ are equal.

(1)

as the normal distribution is continuous so only an change in the range will change the probability.

(f) Calculate the probability that a randomly selected whale spends **more** than £40 in each of four consecutive months.

(2)

$4W \sim N(200, 20^2)$
$P(X > 160) = 0.9772$

$Z \sim B(4, 0.8413)$
$P(Z = 4)$
$= 0.5000$

(g) Calculate the probability that a randomly selected whale spends more than £40 in at least 8 of the months in a one-year period.

(3)

$Z \sim B(12, 0.8413)$
$P(Z \geq 8) = 1 - P(Z \leq 7)$
$= 1 - 0.7490$
$= 0.2510$

(h) Calculate the probability that a randomly selected whale spends less than £40 in a month that they spent more than £30

(3)

$P(30 < W < 40) = 0.1359$

Q4

A
B
C

(i) Discuss the appropriateness or otherwise of the normal distribution to model the spend of a whale in light of this information.

(3)

The normal distribution is inappropriate for modelling the income from sharks as standard deviation from the mean is £60 which means roughly 16% of players will be only paying £60 and nothing above that meaning there is a sharp cut off which is not in a normal distribution which usually has a bell shaped curve.

12 /17

Part (a)
E1: Valid argument, equivalent to skew.

Part (b)
B1: Correct answer.

Part (c)
M1 A1: Alternative method used correctly.

Part (d)
E0: Incorrect explanation.

Part (e)
E1: Correct mention of continuous.

Part (f)
M1 A1: Correctly calculated using binomial distribution.

Part (g)
M1 M1: Use of B(12, 0,8413).
A0: Incorrect probability given.

Part (h)
B1: Calculates P(30 < X < 40).
M0 A0: Does not calculate conditional probability.

Part (i)
E1: Explains about the upper limit.
E1: Concludes normal distribution is not appropriate.
E0: No further points made.

Question 4 - Response C

4 Shai has created a video game. She makes her game free to play, with players able to buy gems to spend in the game for bonuses.

She creates a model to predict the income that she can expect from her game.

Shai's data shows that the majority of players do not spend any money on gems.

A small number of players, known as 'whales', make up the majority of the revenue for games by buying the gems.

[Source: https://deltadna.com/blog/how-whales-spend/]

(a) Give a reason why the normal distribution is unlikely to be a suitable model for the amount a randomly selected player spends in a given month. (1)

there is a minimum limit on how much someone can spend of 0

Shai instead chooses to model the spend, in pounds, of a randomly selected whale in a given month. She uses the distribution $W \sim N(50, 100)$

She assumes that the amount a whale spends each month is independent from their spend in previous months.

(b) Find $P(W < 40)$ (1)

P = 0.158655

(c) Using Shai's model, calculate the probability that a randomly selected whale has a mean spend of less than £40 a month over a year. (2)

$(50, \sqrt{\frac{100}{12}})$

P(x̄ < 40) =

0.0002660028

(d) Explain why the probability calculated in part (c) **must** be smaller than the probability calculated in part (b). (1)

the standard error is smaller than the standard deviation and b) is in the lower tail [illegible] of the distribution

Q4

A B C

(e) Explain why $P(W \leqslant 40)$ and $P(W < 40)$ are equal. (1)

the normal distribution is continuous, $P(W = 40) = 0$ so the same $P(W \leq 40)$ and $P(W < 40)$ value

(f) Calculate the probability that a randomly selected whale spends **more** than £40 in each of four consecutive months. (2)

4 months

$P(X > 40)^4 = (0.841345)^4 =$

0.50107

(g) Calculate the probability that a randomly selected whale spends more than £40 in at least 8 of the months in a one-year period. (3)

$p = 0.8413$ $X \sim B(12, 0.8413)$

$P(X \geq 8) =$ $1 - P(X \leq 7) =$

0.970028

(h) Calculate the probability that a randomly selected whale spends less than £40 in a month that they spent more than £30 (3)

$P(A|B) = \frac{P(A \cap B)}{P(B)}$

$P(30 < X < 40)$ $X \sim N(50, 10)$

$= 0.1359$ of total whales.

$P(X > 30) = 0.97725$

$P(<40 | >30)$

$\frac{0.1359}{0.97725} =$

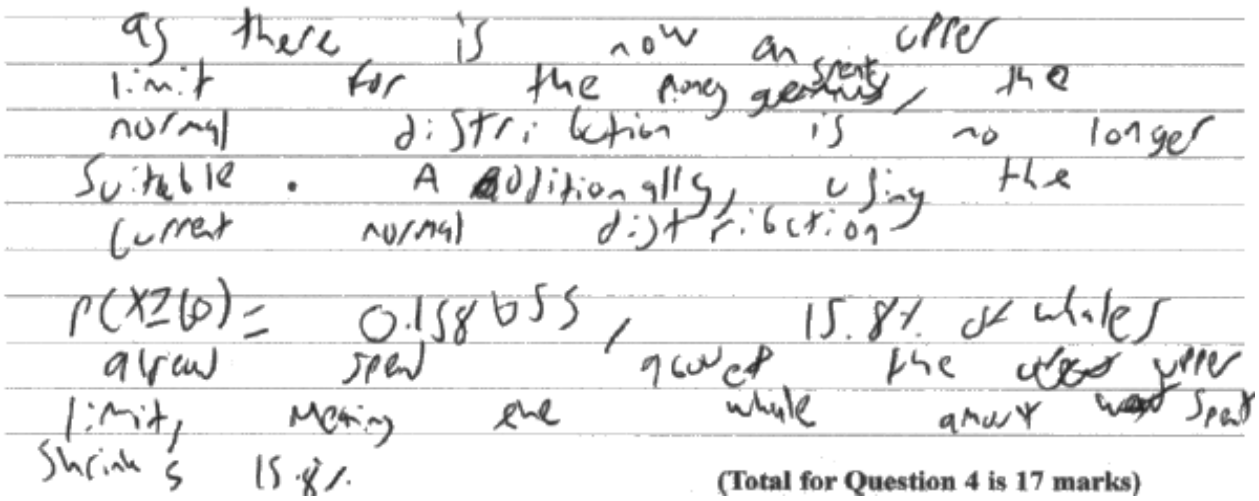
0.13906

Q4

A B C

(i) Discuss the appropriateness or otherwise of the normal distribution to model the spend of a whale in light of this information.

(3)



(Total for Question 4 is 17 marks)

16 /17

Q4

Part (a)
E1: Minimum limit equivalent to normal unbounded below.

Part (b)
B1: Correct answer.

Part (c)
M1A1: Though notation in the normal distribution is incorrect, the answer shows they did the correct calculation so we award both marks.

Part (d)
E1: Correct statement about size of standard deviation.

Part (e)
E1: Correct statement about continuity.

Part (f)
M1 A1: Correct calculation.

Part (g)
M1 M1 A1: Correct calculation.

Part (h)
B1 M1 A1: Correct calculation.

Part (i)
E1: Explains about the upper limit.
E1: Concludes normal distribution is not appropriate.
E0: No further points made.

Question 5

Question 5 - Question

5 Tom owns a busy cinema with two screens. Screen 1 has 120 seats, and Screen 2 has 300 seats. He shows a different film on each screen. Children and pensioners are eligible for a concession ticket that is cheaper than a standard ticket.

From past experience Tom estimates that the proportion of concession tickets sold for the most popular 7 pm showing for the film on Screen 1 is 35% and the proportion for the film on Screen 2 is 15%.

Both screens show films at 7 pm one night.

The random variable X denotes the number of concession tickets purchased for the 7 pm showing of the film on **Screen 1**.

The random variable Y denotes the number of concession tickets purchased for the 7 pm showing of the film on **Screen 2**.

The random variable T denotes the **total** number of concessions tickets purchased for **both** showings of films at 7 pm that night.

Tom assumes that all the tickets will be purchased, and that the probability of a concession ticket being purchased is independent from one customer to the next.

(a) By modelling each of X and Y as an independent binomial distribution,

(i) show that the distribution of T can be approximated as N(87, 65.55)

(4)

(ii) estimate the probability that at least 100 concessions tickets are purchased in total.

(2)

(b) Assuming that it is valid to model X and Y as binomial distributions, give **two** reasons to justify the use of the **approximation** in part (a).

(2)

(c) Give **two** reasons in context why the assumptions that Tom made may **not** be justified.

(2)

Tom charges £4 for a concession ticket, and £6 for a regular ticket. It costs him £200, in total, to run the cinema for the 7pm showing.

(d) Assuming that all tickets are sold for both screens,

(i) write an expression for the number of regular tickets sold in terms of T

(1)

(ii) find the probability Tom earns at least £2 125 of profit.

Note that profit is total income from sales minus costs.

(5)

(Total for Question 5 is 16 marks)

Question 5 - Mark Scheme

Question	Scheme	Marks	AO	Notes
5(a)(i)	X~B(120,0.35) or Y~B(300,0.15)	M1	2.1a	PI
	$X' \sim N(42, 27.3)$ *or* $Y' \sim N(45, 38.25)$	M1	2.1a	PI
	42+45=87	E1	1.2	Addition of means (no variances seen scores M1M0E1E0)
	27.3+38.25=65.55	E1	1.2	Addition of variances not standard deviations
5(a)(ii)	$P(T \geq 100) \approx P(T' > 99.5)$	M1	1.2	Continuity correction used
	0.0613	A1	1.2	sc 0.0542~0.0543 No continuity correction scores M0A1
5(b)	The binomial distributions X & Y can be approximated as a normal distribution as n is sufficiently large	E1	3.1a	Equivalently correct comment regarding np or the mean
	Two independent normal distributions can be added to give a normal distribution, and we assumed independence above	E1	3.1a	
5(c)	All tickets will sell – may not be justifiable as it's unlikely every single ticket will sell every night as some nights may be busier than others	E1	3.1a	Reason why assumption all tickets sell may be incorrect
	Independence of tickets being concessions – may not be justifiable as people may come in groups of friends of similar ages	E1	3.1a	Reason why assumption of independence may be incorrect e.g. not independent because people seeing one film can't see the other
5(d)(i)	$420 - T$	B1	1.2	
5(d)(ii)	$P = 6(420 - T) + 4T[-200]$	M1	1.2	
	$P = 2320 - 2T$	M1	1.2	Or 2520
	$P \sim N(2146, \ldots$	A1	1.2	Mean (or 2346)
	$\ldots, 262.2)$	A1	1.2	Variance
	$P(P > 2125) = 0.9027$	A1	1.2	Or >2325
	Total	**16**		

Q5

A
B
C

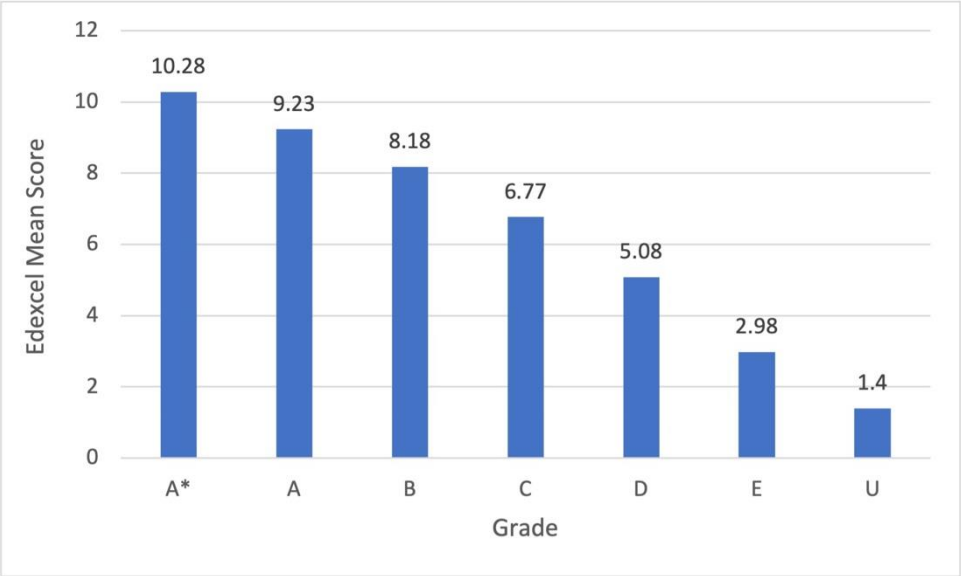
Question 5 - Examiner Comments

Part (a) was well answered, though a large number of candidates forgot their continuity correction so didn't gain full marks. Most candidates were able to justify the use of approximating each binomial to a normal distribution but struggled to realise that the combination of the two normal distributions was part of Tom's approximation. In part(c), most candidates answered well, but should again be reminded to give answers in context; they were asked to justify why his assumptions may not be true so simply saying "may not be independent" is not enough as they have been told independence may not be likely and are being asked **why** that is the case in this scenario.

Parts (d) and (e) involved some basic algebra. It is a good reminder of the additional skills in appendices 3 & 4 of the specification.

Question 5 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade:							
			ALL	A*	A	B	C	D	E	U
6.63	16	41	6.63	10.28	9.23	8.18	6.77	5.08	2.98	1.40



Question 5 - Response A

5 Tom owns a busy cinema with two screens. Screen 1 has 120 seats, and Screen 2 has 300 seats. He shows a different film on each screen. Children and pensioners are eligible for a concession ticket that is cheaper than a standard ticket.

From past experience Tom estimates that the proportion of concession tickets sold for the most popular 7 pm showing for the film on Screen 1 is 35% and the proportion for the film on Screen 2 is 15%.

Both screens show films at 7 pm one night.

The random variable X denotes the number of concession tickets purchased for the 7 pm showing of the film on **Screen 1**.

The random variable Y denotes the number of concession tickets purchased for the 7 pm showing of the film on **Screen 2**.

The random variable T denotes the **total** number of concessions tickets purchased for **both** showings of films at 7 pm that night.

Tom assumes that all the tickets will be purchased, and that the probability of a concession ticket being purchased is independent from one customer to the next.

(a) By modelling each of X and Y as an independent binomial distribution,

(i) show that the distribution of T can be approximated as N(87, 65.55)

(4)

$X \sim B(0.35, 0.15)$

$E(T) = 87$ $\quad E(X) = 0.35 \times 120 = 42$

$Var(T) = 65.55^2$ $\quad E(Y) = 0.15 \times 300 = 45$

$E(T) = E(X) + E(Y) = 42 + 45 = 87$

$Var(T) \Rightarrow Var(X) = 0.35 \times 0.65 \times 120 = 27.3$

$Var(Y) = 0.15 \times 0.85 \times 300 = 38.25$

$Var(T) = Var(X) + Var(Y) = 27.3 + 38.25 = 65.55$

(ii) estimate the probability that at least 100 concessions tickets are purchased in total.

(2)

$T \sim N(87, 65.55)$

$P(T \geqslant 99.5) = 0.0841$ 0.0613

(b) Assuming that it is valid to model X and Y as binomial distributions, give **two** reasons to justify the use of the **approximation** in part (a).

(2)

The number of seats x probability > 10
120 x 0.35 = 42
300 x 0.15 = 45
The Central Limit Theorem states that in a sample bigger than 30, the means of variables in any distribution will be normally distributed.

(c) Give **two** reasons in context why the assumptions that Tom made may **not** be justified.

(2)

The type of film showing will attract an audience appropriate to its themes. Tom cannot assume the proportion of concession tickets sold will be the same for each film.
The day of the week will also have an effect on the number of people at different ages who attend.

Tom charges £4 for a concession ticket, and £6 for a regular ticket. It costs him £200, in total, to run the cinema for the 7pm showing.

(d) Assuming that all tickets are sold for both screens,

(i) write an expression for the number of regular tickets sold in terms of T

(1)

number of regular tickets sold = 420 - T = R

R =

(ii) find the probability Tom earns at least £2 125 of profit.

Note that profit is total income from sales minus costs.

(5)

$T \sim N(87, 65.55)$

Profit = total income from sales + costs =

$2125 + 200 = 2325$

$E(T) = 87$

$E(R) = 420 - 87 = 333$

8 /16

Q5

Part (a)(i)
M1 M1: Correct calculation of the variances implies the binomial distribution has been used.
E1 E1: Expectation and variances have been added correctly.

Part (a)(ii)
M1 A1: Correct calculation including continuity correction.

Part (b)
E1: Demonstrated that a normal approximation is appropriate by calculation.
E0: Further remarks do not explain the summation being valid.

Part (c)
E0 E0: This candidate has not addressed the assumptions that Tom makes – these were specifically labelled as assumptions in the initial stem.

Part (d)(i)
B1: 420-*T* seen.

Part (d)(ii)
M0 M0 A0 A0 A0: No relevant working seen.

Question 5 - Response B

5 Tom owns a busy cinema with two screens. Screen 1 has 120 seats, and Screen 2 has 300 seats. He shows a different film on each screen. Children and pensioners are eligible for a concession ticket that is cheaper than a standard ticket.

From past experience Tom estimates that the proportion of concession tickets sold for the most popular 7 pm showing for the film on Screen 1 is 35% and the proportion for the film on Screen 2 is 15%.

Both screens show films at 7 pm one night.

The random variable X denotes the number of concession tickets purchased for the 7 pm showing of the film on **Screen 1**.

The random variable Y denotes the number of concession tickets purchased for the 7 pm showing of the film on **Screen 2**.

The random variable T denotes the **total** number of concessions tickets purchased for **both** showings of films at 7 pm that night.

Tom assumes that all the tickets will be purchased, and that the probability of a concession ticket being purchased is independent from one customer to the next.

(a) By modelling each of X and Y as an independent binomial distribution,

(i) show that the distribution of T can be approximated as N(87, 65.55)

(4)

$\frac{35}{100} \times 120 = 42 \qquad \frac{15}{100} \times 300 = 45$

$42 + 45 = 87$

(ii) estimate the probability that at least 100 concessions tickets are purchased in total.

(2)

$N(87, 65.55)$

$P(T \geq 100) = 0.0512$ (to 3sf)

(b) Assuming that it is valid to model X and Y as binomial distributions, give **two** reasons to justify the use of the **approximation** in part (a).

(2)

Q5

A B C

(c) Give **two** reasons in context why the assumptions that Tom made may **not** be justified.

(2)

One reason why the assumptions that tom made may not be justified is that the probability of each concession ticket being independent from the next is true as there may be a family with more than one child. Another reason is that all tickets may not be sold as people may not want to watch one of the films being screened.

Tom charges £4 for a concession ticket, and £6 for a regular ticket. It costs him £200, in total, to run the cinema for the 7pm showing.

(d) Assuming that all tickets are sold for both screens,

(i) write an expression for the number of regular tickets sold in terms of T

(1)

$470 - 87 = 333$ $333 \times 6 = £1998$

(ii) find the probability Tom earns at least £2 125 of profit.

Note that profit is total income from sales minus costs.

(5)

$333 \times 6 = 1998 + = £2346$

$87 \times 4 = 348$

$£2346 - £200 = £2146$

$X \sim N(2146, 65.55)$

$P(X > 2125) = 0.995$ (to ~~3~~)

Q5

A B C

8 /16

Part (a)(i)
M1 M0 E1 E0: Addition of means (no variance seen) special case.

Part (a)(ii)
M0 A1: Special case, no continuity correction used.

Part (b)
E0 E0: No answers given.

Part (c)
E1: Contextual reason all tickets may not sell.
E1: Contextual reason we may not have independence.

Part (d)(i)
B0: Not seen.

Part (d)(ii)
M1: Method correct using their answer to (i).
M1 M1: 1st mark implied by the 2146.
A0 A0: Variance not seen and answer not correct.

Question 5 - Response C

5 Tom owns a busy cinema with two screens. Screen 1 has 120 seats, and Screen 2 has 300 seats. He shows a different film on each screen. Children and pensioners are eligible for a concession ticket that is cheaper than a standard ticket.

From past experience Tom estimates that the proportion of concession tickets sold for the most popular 7 pm showing for the film on Screen 1 is 35% and the proportion for the film on Screen 2 is 15%.

Both screens show films at 7 pm one night.

The random variable X denotes the number of concession tickets purchased for the 7 pm showing of the film on **Screen 1**. $X \sim B(120, 0.35)$

The random variable Y denotes the number of concession tickets purchased for the 7 pm showing of the film on **Screen 2**. $Y \sim B(300, 0.15)$

The random variable T denotes the **total** number of concessions tickets purchased for **both** showings of films at 7 pm that night.

Tom assumes that all the tickets will be purchased, and that the probability of a concession ticket being purchased is independent from one customer to the next.

(a) By modelling each of X and Y as an independent binomial distribution,

(i) show that the distribution of T can be approximated as N(87, 65.55)

(4)

$T = X + Y$

For X: $\bar{x} = 120 \times 0.35 = 42$ $\quad X \sim N(42, 27.3)$

$S_x^2 = 120(0.35)(1 - 0.35) = 27.3$

For Y: $\bar{y} = 300 \times 0.15 = 45$ $\quad Y \sim N(45, 38.25)$

$S_y^2 = 300(0.15)(1 - 0.15) = 38.25$

$E(X+Y) = 42 + 45 = 87$

$\text{Var}(X+Y) = 27.3 + 38.25 = 65.55$

$\therefore T \sim N(87, 65.55)$

(ii) estimate the probability that at least 100 concessions tickets are purchased in total.

(2)

$P(T > 100) = 0.0542$ (to 3SF)

upper: 1×10^{10}

lower: 100

$\sigma = \sqrt{65.55}$

$\mu = 87$

Q5

A B C

C - Constant probability
I - Independent
F - Fixed
2 - 2 outcomes

Question 5 continued

(b) Assuming that it is valid to model X and Y as binomial distributions, give **two** reasons to justify the use of the **approximation** in part (a).

(2)

- $n \geq 30$ in both conditions ($120 > 30$ and $300 > 30$)
- ~~The probability is constant~~
- The samples are random

(c) Give **two** reasons in context why the assumptions that Tom made may **not** be justified.

(2)

~~The probability one~~ ~~The prob~~

The number of concession tickets bought may not be independent of eachother, likely that more than one child may go to the cinema to watch the film together.

There may not be a constant probability, this could change depending on what film is screened (eg a child's film likely to attract more children ∴ more concessions.

It is unlikely that all tickets for both screens ~~so~~ to be sold

Question 5 continued

$T \sim N(87, 65.55)$

$\hookrightarrow T \sim N(348, 262.2)$ in pound

Tom charges £4 for a concession ticket, and £6 for a regular ticket. It costs him £200, in total, to run the cinema for the 7pm showing.

S_1 regular $\sim B(120, 0.65)$

$S_2 \sim B(300, 0.85)$

(d) Assuming that all tickets are sold for both screens,

→ 420

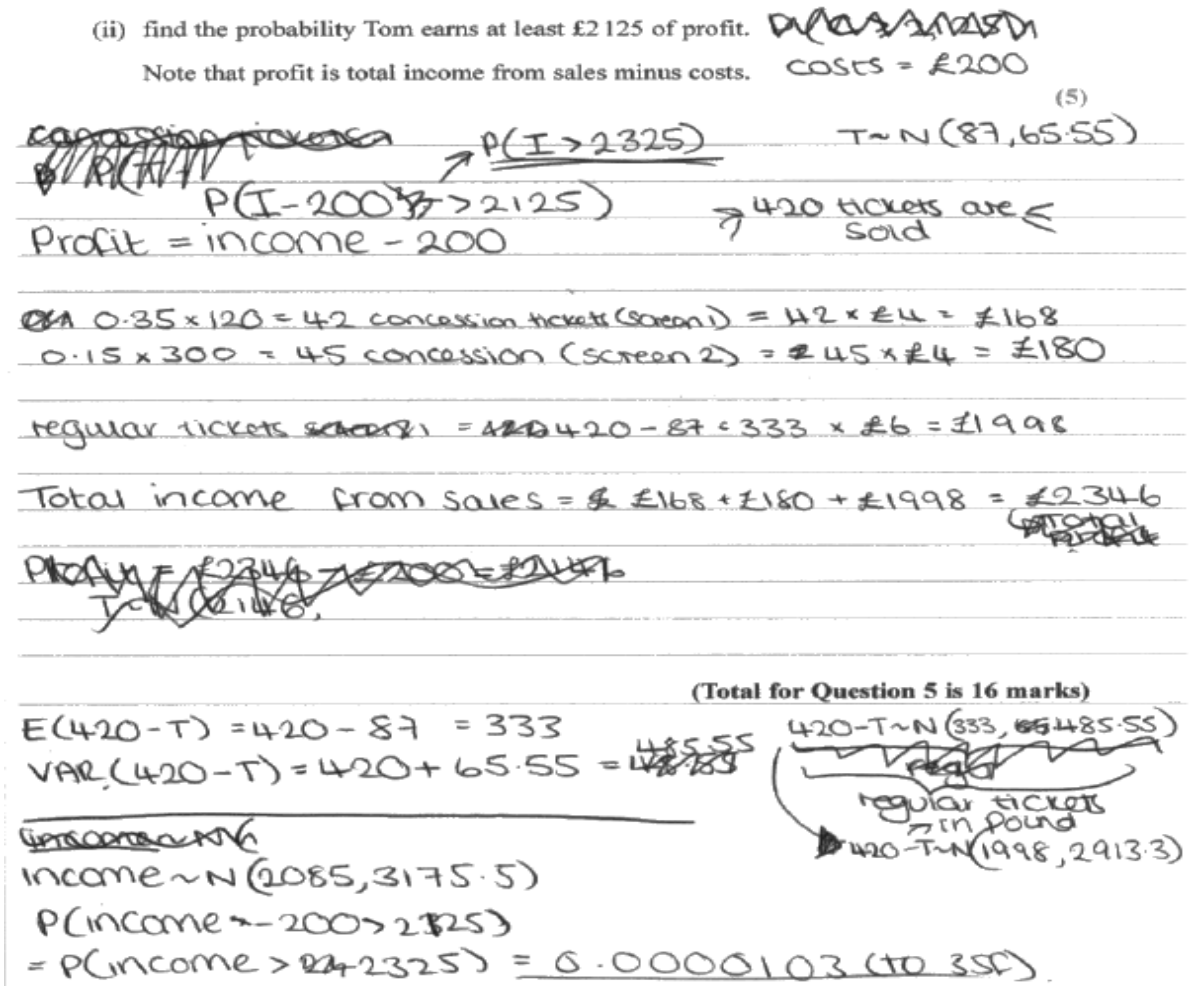
(i) write an expression for the number of regular tickets sold in terms of T

(1)

regular tickets = ~~[illegible]~~ $420 - T$

Q5

A B C



Q5

12 /16

Part (a)(i)
M1 M1 E1 E1: Clearly shows each distribution approximated and added together.

Part (a)(ii)
M0 A1: Special case, no continuity correction used.

Part (b)
E1: Condone $n \geq 30$ for the mark for large sample.
E0: No mark awarded for random sample.

Part (c)
E1: Condone for unlikely that all tickets will be sold.
E1: Contextual reason to doubt independence.

Part (d)(i)
B1: Correct expression.

Part (d)(ii)
M1 M1 A1: The 2346 seen implies these first 3 marks.
A0 A0: Variance incorrect and probability incorrect.

Question 6

Question | Mark Scheme | Examiner Comments | Performance | Response A | Response B | Response C

Question 6 - Question

6 Caroline runs a hairdressing salon in London. She reads the following article.

How much hair loss is normal in a day

According to the American Academy of Dermatologists, an adult will normally lose anywhere From 50 to 100 strands of hair per day. For people with longer hair strands, losing them May be more noticeable. Since there are 100,000 hair follicles – or more – on each person's scalp, the loss of 100 or so hair strands a day doesn't make a big difference in appearance.

Average hair loss in men and women

Women tend to lose more hair strands per day than men. There's no way to measure the difference objectively, because daily heat styling and frequent hair colouring plays a big part in how much of your hair sheds. Approximately 40 percent of women lose extra hair every day because of the way they style it.

What causes hair to fall out

Some daily hair loss is normal. Increased hair loss can be a result of stress or a health condition.

[Source: https://www.healthline.com/health/how-much-hair-loss-is-normal]

Caroline states that the random variable $X \sim B(100\,000, 0.001)$ could be used to model the number of hairs lost by a randomly selected customer during one day.

(a) Explain how Caroline chose her parameters for the distribution of X

(2)

(b) Comment on the assumptions underlying Caroline's model indicating whether or not these are reasonable.

Your answer should include **three** different statements.

(3)

(Total for Question 6 is 5 marks)

Question 6 - Mark Scheme

Question	Scheme	Marks	AO	Notes
6(a)	**Explanation of how Caroline chose her parameters** $n = 100\,000$ represents number of hairs on a head $p = 0.001 = \frac{100}{100000}$ 100 hairs out of 100000 lost per day	 E1 E1	 1.3 1.3	
6(b)	**Reasonable** Assuming that values for Britain are similar to America $n = 100\,000$ is reasonable from the article Caroline may have chosen 100 hairs (at the upper end) as she has mostly female customers who lose more hair **Unreasonable** Assuming that values for Britain are different to America Assumes that everyone is treating their hair similarly e.g. similar products Probability is different for men and women as women tend to lose more hair $n = 100\,000$ unlikely to be valid as likely to be substantial differences between individuals Caroline chose 100 hairs when that was an upper end of a range Family members may book appointments on the same day so unlikely to be independent events Probability doesn't take into account effect of stress or health conditions so not valid if customer suffers from one of these			Or mention of how people differ

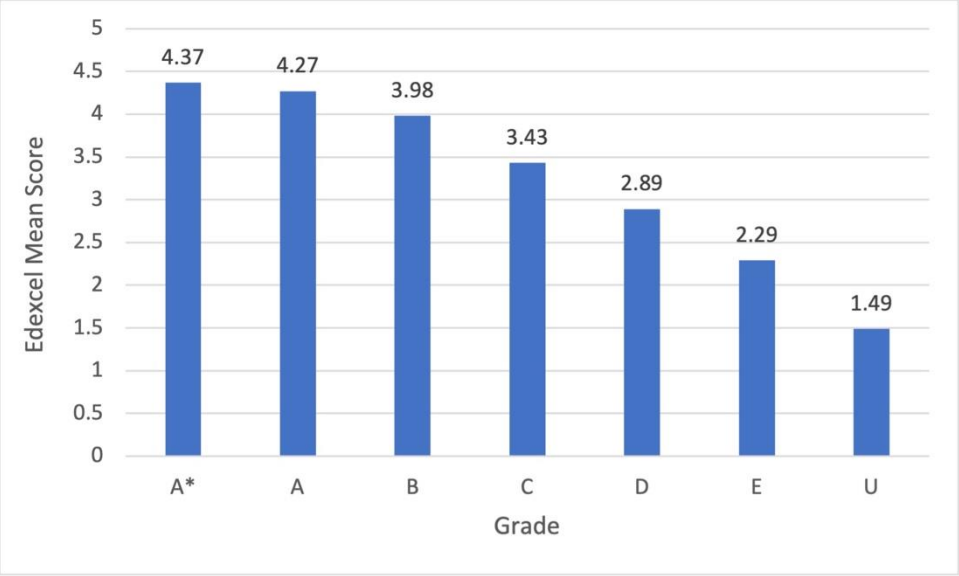
	The falling out of hairs may not be independent of one another (e.g. one hair falling out may weaken the surrounding hairs)			
		E1, E1, E1	3.1b, 3.1b, 3.1b	Any 3 of these comments. Cannot score 3 without a comment about whether this is reasonable or not
	Total	**5**		

Question 6 - Examiner Comments

This question was answered well, although some candidates did not provide contextual reasons as to why assumptions were not reasonable, simply saying that loss of one hair compared to another may not be independent. It is important to challenge the assumptions in context, rather than simply just restate the assumptions required for a binomial distribution, when candidates are asked to comment on the assumptions.

Question 6 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade:							
			ALL	A*	A	B	C	D	E	U
3.41	5	68	3.41	4.37	4.27	3.98	3.43	2.89	2.29	1.49



Question 6 - Response A

(a) Explain how Caroline chose her parameters for the distribution of X

(2)

100,000 is from hair follicles

0.001 is the 100 that is lost

(b) Comment on the assumptions underlying Caroline's model indicating whether or not these are reasonable.

Your answer should include **three** different statements.

(3)

- There is no way of knowing how much ~~hair~~ out Hair you lose Comed from Man to female

Question 6 continued

- That everyones hair falls on the same away or how much is lost

- That men dont use as much hair product as women thats why women lose more hair

Q6

A B C

2 /5

Part (a)

E1: Explains $n = 100{,}000$

E0: Does not explain calculation behind the 0.001

Part (b)

E0: "There is no way to know how much hair you lose from male to female" does not address any assumption.

E0: "That everyone's hair…" does not address an assumption.

E1: "Men don't use as much hair product…" is equivalent to doubting the assumption that everyone treats their hair similarly.

Question 6 - Response B

(a) Explain how Caroline chose her parameters for the distribution of X

(2)

100 000 -> number of hair folicles

then -> $\frac{100}{100\,000}$ -> 0.001 -> the loss of 100 on day

(b) Comment on the assumptions underlying Caroline's model indicating whether or not these are reasonable.

Your answer should include **three** different statements.

(3)

1 - fixed number of trials

- ~~probability of success is constant~~

2 - events are independent to each other

~~only two outcomes can occur~~

3 - we can have only two outcomes

Q6

2 /5

Part (a)

E1: Explains 100,000 hairs.

E1: Explains $\frac{100}{100,000}$ as probability.

Part (b)

E0 E0 E0: Though they mention the conditions of a binomial distribution, there is no attempt to tie them to the context or discuss if they are reasonable or not.

Question 6 - Response C

Caroline states that the random variable $X \sim B(100\,000, 0.001)$ could be used to model the number of hairs lost by a randomly selected customer during one day.

(a) Explain how Caroline chose her parameters for the distribution of X

(2)

- Caroline as used the data of 100,000 hair follicles for her n value
- She has used $\frac{100}{100,000} = 0.001$ as her probability as between 50 to 100 strands fall out a day.

(b) Comment on the assumptions underlying Caroline's model indicating whether or not these are reasonable.

Your answer should include **three** different statements.

(3)

- There may not be a fixed number of trials as it states some people may have more than 100,000 hair follicles.
- A hair strand falling out may not be independent of another as a group may fall out together
- there may not be the same probabilities of a strand of hair falling out as a person may style or heat one part of their hair more than the rest which may increased the probability of falling out.

Q6

A B C

5 /5

Part (a)

E1: Explains 100,000 hairs.

E1: Explains $\frac{100}{100,000}$ as probability.

Part (b)

E1: Contextual reason that n = 100,000 may not be reasonable.

E1: Contextual reason that independence may not be reasonable.

E1: Contextual reason that probability may not be constant.

Question 7

Question | Mark Scheme | Examiner Comments | Performance | Response A | Response B | Response C

Question 7 - Question

7 Animal Crossing is a video game in which players visit desert islands. A single animal lives on each desert island. A player visiting an island will always meet the animal that lives there. Each animal is either male or female.

Jonny plays a simplified version of Animal Crossing in which there are only four animal species. Data on the number of animals of each species is shown in Figure 5.

		Species				
		Alligator	**Anteater**	**Bear**	**Bird**	**Total**
Gender	**Male**	6	3	10	15	34
	Female	3	7	6	6	22
	Total	9	10	16	21	56

[Source: https://animalcrossing.fandom.com/wiki/Animal_Crossing_Wiki]

Figure 5

Jonny visits a desert island at random.

(a). Write down the probability that the animal he meets is a female anteater.

(1)

(b). Write down the probability that the animal he meets is a male, given that it is an alligator.

(1)

Jonny assumes that once he has visited a particular desert island, he will never visit that island again.

(c) Under Jonny's assumption, find the probability that when Jonny visits four islands, he meets two birds and two bears, given that all the animals he meets are either birds or bears.

(5)

In the full version of Animal Crossing there are **more** than four species of animal.

Jonny's assumption in part (c) is untrue, and it is possible to meet the same animal **multiple times**.

The data for alligators, anteaters, bears and birds in **Figure 5** are correct for both

versions of the game.

(d) Explain **why**, in the full version of the game,

(i) the probability in part (a) would be smaller than the calculated value

(1)

(ii) the probability in part (b) would be unchanged

(1)

(iii) the probability in part (c) would be smaller than the calculated value.

(1)

In Animal Crossing players are able to buy and sell turnips each week.

The prices at which players can sell turnips during a week follow one of three patterns:

Spike Fluctuating Decreasing

The probability of the prices following a certain pattern one week depends only on the previous week's pattern, as shown in **Figure 6**.

		This week's pattern		
		Spike	**Fluctuating**	**Decreasing**
Previous week's pattern	**Spike**	0.3	0.5	0.2
	Fluctuating	0.65	0.2	0.15
	Decreasing	0.7	0.25	0.05

[Source: https://animalcrossing.fandom.com/wiki/Animal_Crossing_Wiki]

Figure 6

During the first week of May, the prices of Jonny's turnips are following the fluctuating pattern.

(e) Find the probability that, during the **third** week of May, the prices of Jonny's turnips are following the fluctuating pattern again.

(2)

During the first week of May, the prices of Jonny's turnips were following the fluctuating pattern.
During the third week of May, the prices of Jonny's turnips were following the fluctuating pattern again.

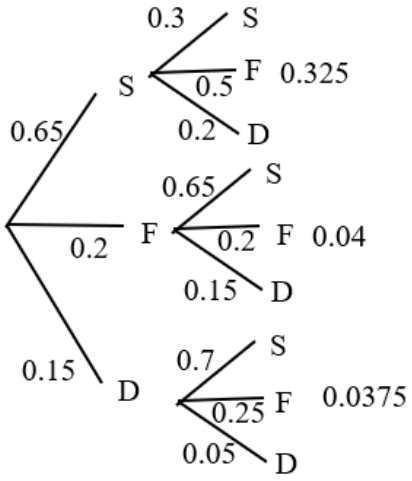
(f) Find the probability that during the **second** week of May, the prices of Jonny's turnips were following the spike pattern.

(4)

(Total for Question 7 is 16 marks)

Question 7 - Mark Scheme

Question	Scheme	Marks	AO	Notes
7(a)	$\frac{7}{56}\left(=\frac{1}{8}\right)$	B1	1.2	oe
7(b)	$\frac{6}{9}\left(=\frac{2}{3}\right)$	B1	1.2	oe
7(c)	$21 \times 20 \times 16 \times 15$	M1	1.2	Numerator
	37 seen or used in denominator	B1	1.2	
	$37 \times 36 \times 35 \times 34$	M1	1.2	Denominator decreases
	$\times 6$	M1	1.2	6 orders (or $4C2$)
	$\frac{21}{37} \times \frac{20}{36} \times \frac{16}{35} \times \frac{15}{34} \times 6 = \frac{240}{629}$	A1	1.2	Or 0.382
7(d)(i)	Probability in (a) would be smaller as more species means more animals overall	E1	2.1b	
7(d)(ii)	(b) would be unchanged as it is conditional on animal being an alligator and the number of male and female alligators is the same in each game	E1	2.1b	
7(d)(iii)	The probability in (c) would be smaller because without replacement the probability gets smaller for each successive animal of each type	E1	2.1b	Or by explicit calculation

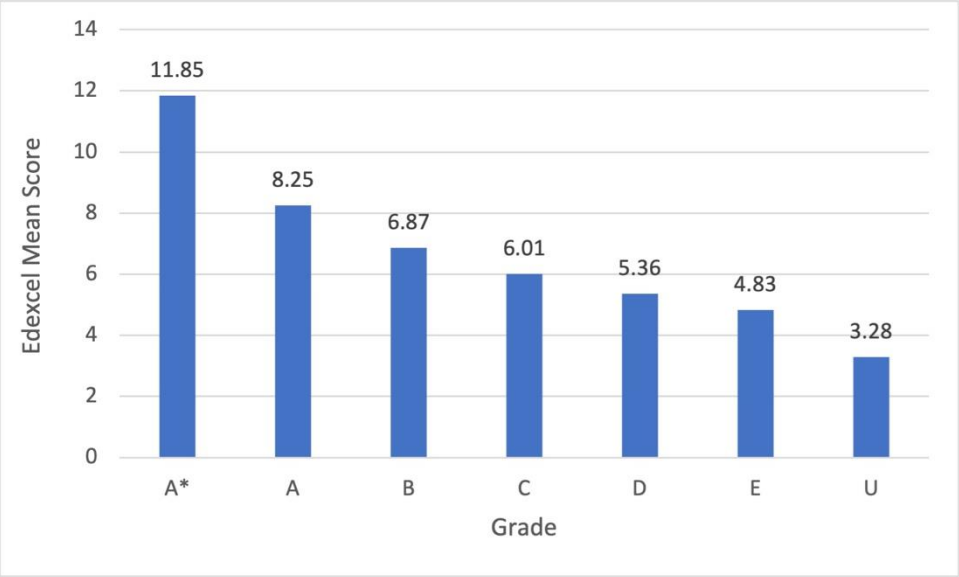
7(e)	$0.65 \times 0.5 + 0.2 \times 0.2 + 0.15 \times 0.25$ 0.4025 **Tree Diagram Method**  0.4025	M1 A1 (M1) (A1)	1.2 1.2	Tree with at least the 0.65, 0.2, 0.15 as the first layer or 0.5, 0.2 and 0.25 on the fluctuating branch of 2nd layer

Question 7 - Examiner Comments

Parts (a) and (b) were well answered, but part (c) had lots of possibilities for mistakes. Some candidates forgot to take into account that there was no replacement in this scenario, and some either calculated the wrong number of orderings, or forgot about them entirely. In part (d), (i) and (ii) were well answered, but many candidates gave the same answer for (iii) as they had for (i), missing the fact that the sampling with replacement was the important factor here. It is a good idea for candidates to make sure all the information provided has been considered. Parts (e) and (f) were better answered by candidates who chose to use a tree diagram than those who attempted them purely algebraically. It was pleasing to see some full mark responses on this question.

Question 7 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade:							
			ALL	A*	A	B	C	D	E	U
6.42	16	40	6.42	11.85	8.25	6.87	6.01	5.36	4.83	3.28



Question 7 - Response A

7 Animal Crossing is a video game in which players visit desert islands. A single animal lives on each desert island. A player visiting an island will always meet the animal that lives there. Each animal is either male or female.

Jonny plays a simplified version of Animal Crossing in which there are only four animal species. Data on the number of animals of each species is shown in Figure 5.

		Species				
		Alligator	Anteater	Bear	Bird	Total
Gender	Male	6	3	10	15	34
	Female	3	7	6	6	22
	Total	9	10	16	21	56

[Source: https://animalcrossing.fandom.com/wiki/Animal_Crossing_Wiki]

Figure 5

Jonny visits a desert island at random.

(a) Write down the probability that the animal he meets is a female anteater.

(1)

$\frac{7}{56} = 0.125$

(b) Write down the probability that the animal he meets is a male, given that it is an alligator.

(1)

$\frac{6}{9} = 0.667$

Jonny assumes that once he has visited a particular desert island, he will never visit that island again.

(c) Under Jonny's assumption, find the probability that when Jonny visits four islands, he meets two birds and two bears, given that all the animals he meets are either birds or bears.

(5)

$\frac{16}{37} \times \frac{16}{37} \times \frac{15}{37} \times \frac{15}{37} = 0.0307$

BBbb
BbBb
bbBB
bBbB

$0.0307 \times 4 = 0.1229$

Q7

A
B
C

In the full version of Animal Crossing there are **more** than four species of animal.

Jonny's assumption in part (c) is untrue, and it is possible to meet the same animal **multiple times**.

The data for alligators, anteaters, bears and birds in **Figure 5** are correct for both versions of the game.

(d) Explain **why**, in the full version of the game,

(i) the probability in part (a) would be smaller than the calculated value

(1)

because there are more than four species of animals so the total number of animals is going to differ.

(ii) the probability in part (b) would be unchanged

(1)

because the total number of alligators hasn't changed so the probability would be the same.

(iii) the probability in part (c) would be smaller than the calculated value.

(1)

because there are more animals and you can meet the same animal twice

In Animal Crossing players are able to buy and sell turnips each week.

The prices at which players can sell turnips during a week follow one of three patterns:

Spike Fluctuating Decreasing

The probability of the prices following a certain pattern one week depends only on the previous week's pattern, as shown in **Figure 6**.

		This week's pattern		
		Spike	**Fluctuating**	**Decreasing**
Previous week's pattern	**Spike**	0.3	0.5	0.2
	Fluctuating	0.65	0.2	0.15
	Decreasing	0.7	0.25	0.05

1.65 0.95 0.4

[Source: https://animalcrossing.fandom.com/wiki/Animal_Crossing_Wiki]

Figure 6

During the first week of May, the prices of Jonny's turnips are following the fluctuating pattern.

(e) Find the probability that, during the **third** week of May, the prices of Jonny's turnips are following the fluctuating pattern again.

(2)

$\frac{0.95}{1.65 + 0.95 + 0.4} = 0.3167$

0.2

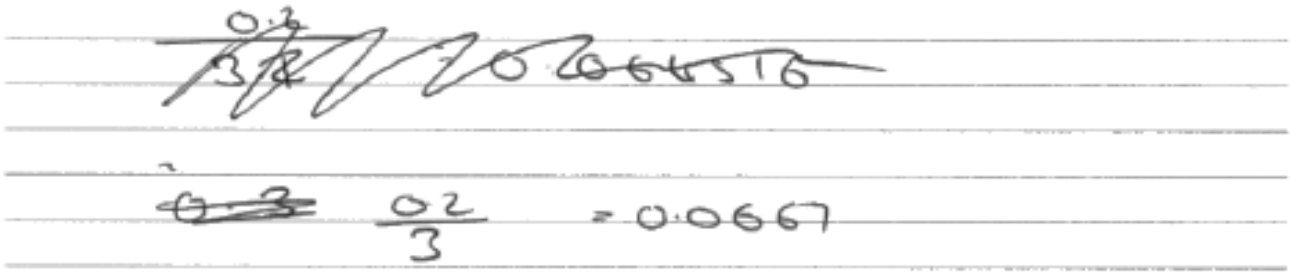
Q7

A B C

During the first week of May, the prices of Jonny's turnips were following the fluctuating pattern.
During the third week of May, the prices of Jonny's turnips were following the fluctuating pattern again.

(f) Find the probability that during the **second** week of May, the prices of Jonny's turnips were following the spike pattern.

(4)



6 /16

Part (a)
B1: Correct answer.

Part (b)
B1: Correct answer.

Part (c)
M0: Numerator incorrect.
B1: 37 used.
M0: denominator incorrect.
M0: Incorrect number of combinations.
A0: Incorrect answer.

Part (d)(i)
E1: Correct answer.

Part (d)(ii)
E1: Correct answer.

Part (d)(iii)
E1: Correct answer.

Part (e)
M0 A0: No relevant working.

Part (f)
M0 B0 M0 A0: No relevant working.

Question 7 - Response B

7 Animal Crossing is a video game in which players visit desert islands. A single animal lives on each desert island. A player visiting an island will always meet the animal that lives there. Each animal is either male or female.

Jonny plays a simplified version of Animal Crossing in which there are only four animal species. Data on the number of animals of each species is shown in Figure 5.

		Species				
		Alligator	Anteater	Bear	Bird	Total
Gender	Male	6	3	10	15	34
	Female	3	7	6	6	22
	Total	9	10	16	21	56

[Source: https://animalcrossing.fandom.com/wiki/Animal_Crossing_Wiki]

Figure 5

Jonny visits a desert island at random.

(a) Write down the probability that the animal he meets is a female anteater.

(1)

$\frac{7}{56} = 0.125$

(b) Write down the probability that the animal he meets is a male, given that it is an alligator.

(1)

m | A

$\frac{6}{9} = 0.667$

Jonny assumes that once he has visited a particular desert island, he will never visit that island again. no replacement

(c) Under Jonny's assumption, find the probability that when Jonny visits four islands, he meets two birds and two bears, given that all the animals he meets are either birds or bears.

(5)

$\frac{21}{37} \times \frac{20}{36} \times \frac{16}{35} \times \frac{15}{34} = 0.0636$

$\frac{16}{37} \times \frac{15}{36} \times \frac{21}{35} \times \frac{20}{34} = 0.0636$

$= 0.1272$

OR

Q7

A B C

In the full version of Animal Crossing there are **more** than four species of animal.

Jonny's assumption in part (c) is untrue, and it is possible to meet the same animal **multiple times**.

The data for alligators, anteaters, bears and birds in **Figure 5** are correct for both versions of the game.

(d) Explain **why**, in the full version of the game,

(i) the probability in part (a) would be smaller than the calculated value

(1)

More animals so total would be bigger as more species of animal

(ii) the probability in part (b) would be unchanged

(1)

there are no more alligators added but more species so the number of male alligators is the same as well as number of alligators is same.

(iii) the probability in part (c) would be smaller than the calculated value.

(1)

He would have to visit more than 4 islands as there would be more species. so prob would be lower.

In Animal Crossing players are able to buy and sell turnips each week.

The prices at which players can sell turnips during a week follow one of three patterns:

Spike Fluctuating Decreasing

The probability of the prices following a certain pattern one week depends only on the previous week's pattern, as shown in **Figure 6**.

		This week's pattern		
		Spike	**Fluctuating**	**Decreasing**
Previous week's pattern	**Spike**	0.3	0.5	0.2
	Fluctuating	0.65	0.2	0.15
	Decreasing	0.7	0.25	0.05

[Source: https://animalcrossing.fandom.com/wiki/Animal_Crossing_Wiki]

Figure 6

During the first week of May, the prices of Jonny's turnips are following the fluctuating pattern.

(e) Find the probability that, during the **third** week of May, the prices of Jonny's turnips are following the fluctuating pattern again.

(2)

this week 0.2 0.5 0.2

Week 1 = fluctuating

0.5 x 0.2 x 0.2 = 0.04

Q7

A B C

During the first week of May, the prices of Jonny's turnips were following the fluctuating pattern.
During the third week of May, the prices of Jonny's turnips were following the fluctuating pattern again.

(f) Find the probability that during the **second** week of May, the prices of Jonny's turnips were following the spike pattern.

(4)

~~0.3 × 0.5 = 0.15~~

W1 W2 W3

Fluc Fluc

Fluc → Spik → Fluc

0.65 × 0.5 = 0.325

8 /16

Q7

Part (a)
B1: Correct answer.

Part (b)
B1: Correct answer.

Part (c)
M1 B1 M1: Correct numerator and denominator.
M0: Incorrect number of combinations.
A0: Incorrect answer.

Part (d)(i)
E1: Correct answer.

Part (d)(ii)
E1: Correct answer.

Part (d)(iii)
E0: Incorrect answer.

Part (e)
M0 A0: No relevant working.

Part (f)
M0 B1 M0 A0: Use of 0.65×0.5 awarded, no other relevant working.

Question 7 - Response C

7 Animal Crossing is a video game in which players visit desert islands. A single animal lives on each desert island. A player visiting an island will always meet the animal that lives there. Each animal is either male or female. NOT TRUE

Jonny plays a simplified version of Animal Crossing in which there are only four animal species. Data on the number of animals of each species is shown in Figure 5.

		Species				
		Alligator	Anteater	Bear	Bird	Total
Gender	Male	6	3	10	15	34
	Female	3	7	6	6	22
	Total	9	10	16	21	56

[Source: https://animalcrossing.fandom.com/wiki/Animal_Crossing_Wiki]

Figure 5

Jonny visits a desert island at random.

(a) Write down the probability that the animal he meets is a female anteater.

(1)

$\frac{7}{56} = 0.125$

(b) Write down the probability that the animal he meets is a male, given that it is an alligator.

(1)

$\frac{6}{9} = 0.666...$

Jonny assumes that once he has visited a particular desert island, he will never visit that island again.

(c) Under Jonny's assumption, find the probability that when Jonny visits four islands, he meets two birds and two bears, given that all the animals he meets are either birds or bears.

(5)

$\frac{16}{37} \times \frac{15}{36} \times \frac{21}{35} \times \frac{20}{34}$

$\frac{16}{37} \times \frac{21}{36} \times \frac{15}{35} \times \frac{20}{34}$

$\frac{16}{37} \times \frac{21}{36} \times \frac{20}{35} \times \frac{15}{34}$

$\frac{60}{624}$

0.3816

Q7

A B C

In the full version of Animal Crossing there are **more** than four species of animal.

Jonny's assumption in part (c) is untrue, and it is possible to meet the same animal **multiple times**.

The data for alligators, anteaters, bears and birds in **Figure 5** are correct for both versions of the game.

(d) Explain **why**, in the full version of the game,

(i) the probability in part (a) would be smaller than the calculated value

(1)

more possible animals to find so prob of picking alligators decreases

(ii) the probability in part (b) would be unchanged

(1)

still same number of alligators made available

(iii) the probability in part (c) would be smaller than the calculated value.

(1)

more possible animals to find and could read the same animal multiple times

In Animal Crossing players are able to buy and sell turnips each week.

The prices at which players can sell turnips during a week follow one of three patterns:

Spike Fluctuating Decreasing

The probability of the prices following a certain pattern one week depends only on the previous week's pattern, as shown in **Figure 6**.

		This week's pattern		
		Spike	**Fluctuating**	**Decreasing**
Previous week's pattern	**Spike**	0.3	0.5	0.2
	Fluctuating	0.65	0.2	0.15
	Decreasing	0.7	0.25	0.05

[Source: https://animalcrossing.fandom.com/wiki/Animal_Crossing_Wiki]

Figure 6

During the first week of May, the prices of Jonny's turnips are following the fluctuating pattern.

(e) Find the probability that, during the **third** week of May, the prices of Jonny's turnips are following the fluctuating pattern again.

(2)

$0.5 \times 0.65 = 0.325$
$0.2 \times 0.2 = 0.04$
$0.15 \times 0.25 = 0.0375$
0.4025

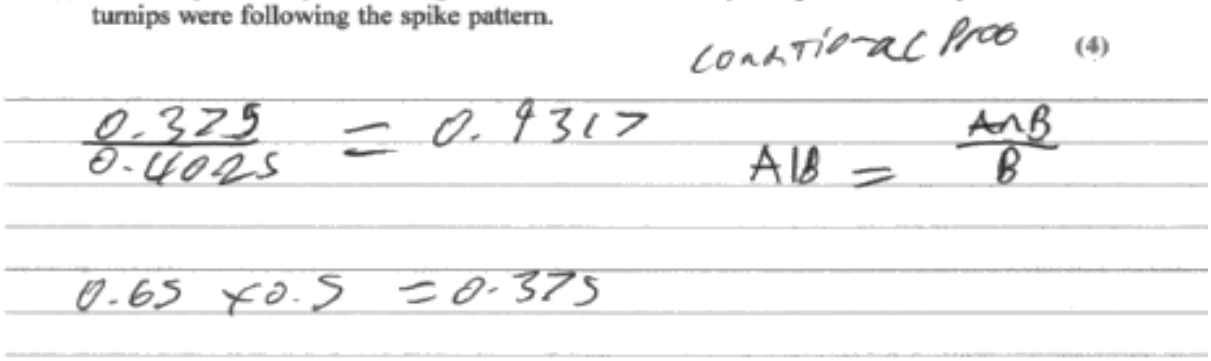
prob that in 3rd week of May that price is fluctuating $= 0.4025$

Q7

A
B
C

During the first week of May, the prices of Jonny's turnips were following the fluctuating pattern.
During the third week of May, the prices of Jonny's turnips were following the fluctuating pattern again.

(f) Find the probability that during the **second** week of May, the prices of Jonny's turnips were following the spike pattern. (4)



15 /16

Q7

Part (a)
B1: Correct answer.

Part (b)
B1: Correct answer.

Part (c)
M1 B1 M1 M1 A1: Correct answer.

Part (d)(i)
E1: Correct answer.

Part (d)(ii)
E1: Correct answer.

Part (d)(iii)
E1: Correct answer – mentions multiple times at the end.

Part (e)
M1 A1: Correct answer.

Part (f)
M1 B1 M1 A0: Correct calculation performed but incorrect answer given.