Pearson Edexcel

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

May/June 2022 Exemplar
9ST0-02 A level Statistics
Paper 03: Statistics in Practice
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 20

Question 3 30

Question 4 47

Question 5 60

Question 6 73

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

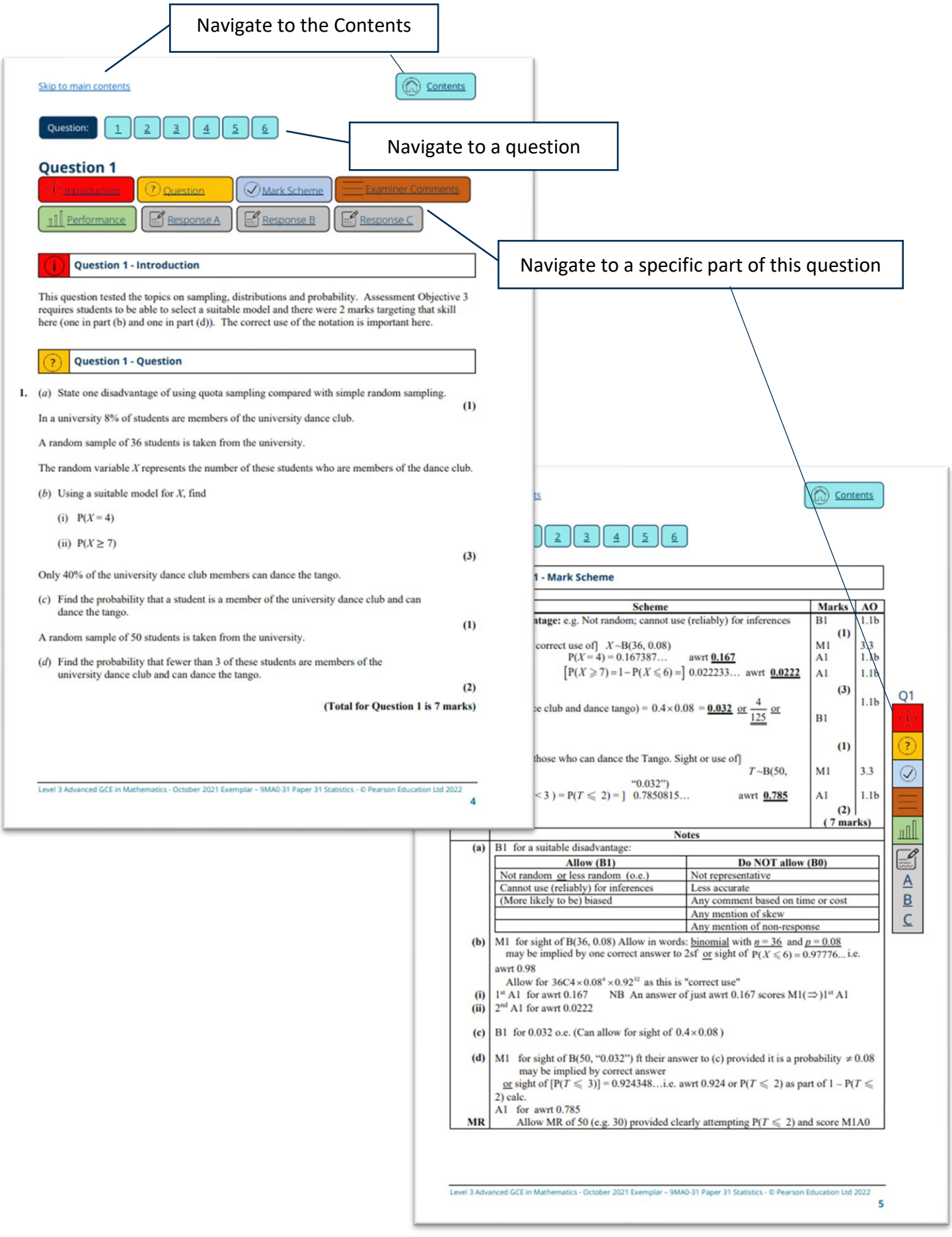
This document has been produced to support mathematics teachers delivering the GCE Statistics specification.

This document looks at questions from the 9ST0-03 A level Statistics Paper 3: Statistics in Practice 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 the most accessible of the three papers. As with both paper 1 and paper 2, a common area for improvement is ensuring that answers are given in context, in context, rather than simply being stated in general terms.

Summary

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. Following an illness, patients might need therapy in order to help improve their balance.
The therapy offered is called standard balance training.

Giovanni, a physiotherapist, thinks that his patients may be helped further by using a video game designed to improve balance.
He designs an experiment to test his theory.

He randomly assigns patients to one of two groups.
One group receives a course of therapy including standard balance training and the video game. The other group receives only standard balance training.

(a) State which of the two groups described above is the **control** group.

(1)

(b) Explain, in context, why Giovanni **randomly** assigns each patient to one of the two groups.

(1)

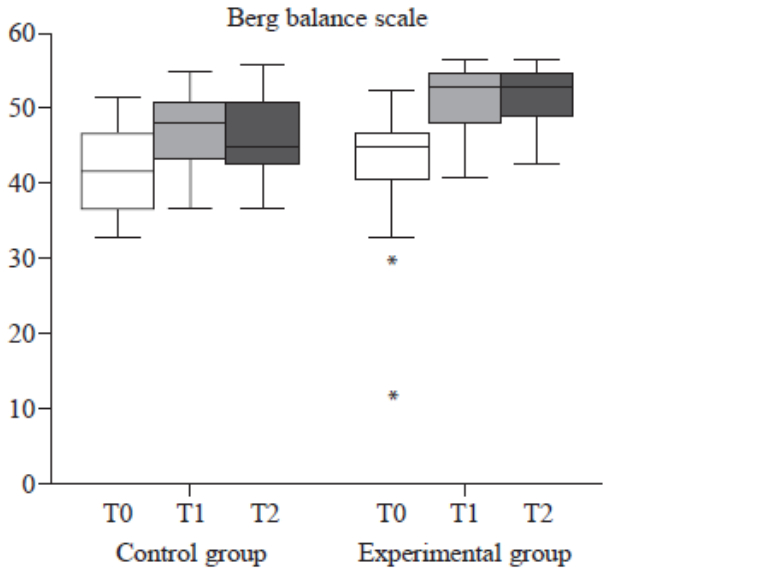
(c) Explain why Giovanni's experiment design is not double-blind.

(1)

At the start of the therapy (T0), Giovanni measures the patients' ability to balance using the "Berg balance scale".

He repeats this measure at the end of the course of therapy (T1), and at a follow-up appointment one month later (T2).

Giovanni displays his results using box and whisker plots, shown in **Figure 1**.



[Source: http://dx.doi.org/10.1155/2014/580861]

Figure 1

A **higher score** on the Berg balance scale indicates **better** balance.

(d) Make **four** distinct comments, in context, about Giovanni's experiment and on his results using the information shown in **Figure 1 only**.

(4)

(e) Highlight **one** advantage and **one** disadvantage of Giovanni's choice of box and whisker plots to display his results.

(2)

(Total for Question 1 is 9 marks)

Question 1 - Mark Scheme

Question	Scheme	Marks	AO	Notes
1(a)	The group that only receives the **standard** balance training.	B1	1.1	Accept "the second group"
1(b)	So that any difference (in outcome) found between the two groups can be more confidently attributed to the video game.			oe Accept "To reduce the likelihood that some factor other than the video game results in a difference (in outcome) between the two groups" or "Reduces experimental error"

	To avoid the **bias** that could arise if patients were not assigned randomly.			oe Reference specifically to removing **bias**.
		E1	3.1a	Either Do not accept "to make the experiment more fair" or similar.
1(c)	Because Giovanni knows which patients are assigned to each group.			oe
	Because the patients know if they are playing the video game or not			
		B1	1.1	Either
1(d)	**List not exhaustive**			
	Both the control and experimental groups **show improvement** in their (average Berg balance) scores at the **end** (T1) of the experiment.			Accept "Both groups have better (average) balance at the end of the experiment"
	Both the control and experimental groups **show improvement** on their **initial** (average Berg balance) scores **one month after** (T2) the experiment.			Must specify that this is an improvement on **initial** (average) scores or balance.
	The **control group's improvement** in average (Berg balance) score at the end of the experiment **does not remain** one month after (T2) the experiment			Accept "The control group's better average balance is not maintained one month after the experiment"
	The **experimental group's improvement** in average (Berg balance) score at the end of the experiment **was maintained** one month after (T2) the experiment			
	The control and experimental groups have **different distributions** of (Berg balance) scores at the **start** (T0) of the experiment.			This is **not** distinct from comments on the differences in spread or average of the scores at the start of the experiment. (41 vs 45)
	The **control group** has a slightly **lower average** (Berg balance) score at the **start** (T0) of the experiment than the experimental group.			Accept "The control and experimental groups have similar average scores at the start of the experiment."

1(d) cont.	The **experimental group** has a higher median balance score than the control group at **all stages**			
	The experimental group has **two outliers**			
	The **experimental group** has a **bigger range** of (Berg balance) scores at the **start** (T0) of the experiment than the control group.			
	The **experimental group** has a **smaller** range of (Berg balance) scores at the **end** (T1) of the experiment than the control group.			
	The **experimental group** has a **smaller IQR** of (Berg balance) scores at the **start** (T0) of the experiment than the control group.			
	Control group **one month after the experiment** (T2) has the **same median** as the experimental group had at the **start** of the experiment (T0)			
	The **spread decreases** in the **experimental group** from the end of the experiment (T1) to one month after the experiment (T2).			
	Experimental group are **positively skewed**			
		E1, E1, E1, E1	1.1, 1.1, 1.1, 1.1	One mark for each **distinct** correct comment up to a maximum of 4 marks Maximum 3 if no context attempted **Accept** median or average but **not** mean throughout – penalise only **once**

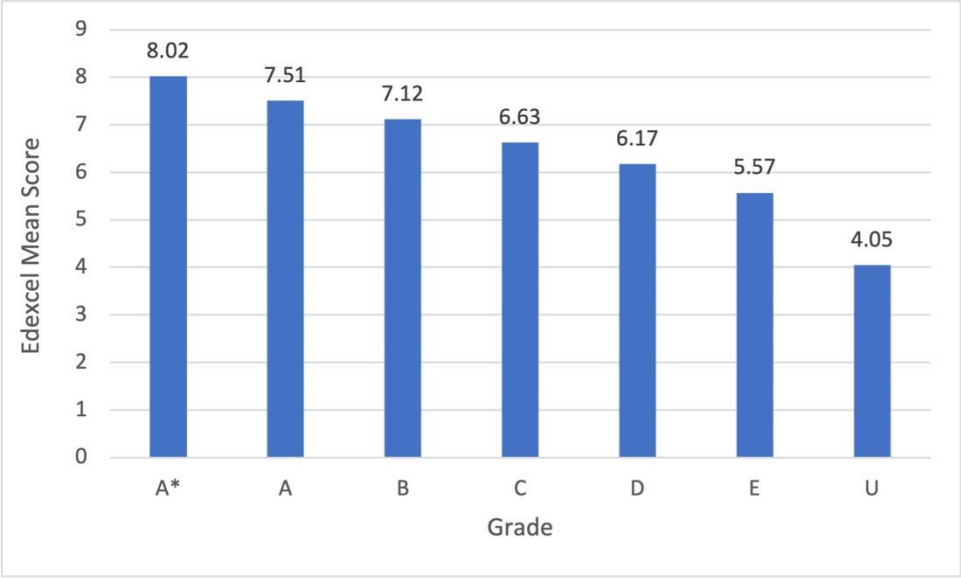
Question	Scheme	Marks	AO	Notes
1(e)	**Advantages**			
	The box and whisker plots clearly show the differences between the distributions (of the Berg balance scores) before and after the therapies.			oe
	The box and whisker plots clearly show the differences (of the Berg balance scores) between the distributions of the control and experimental groups.			oe control v experimental or e.g. easy to compare averages
	It can show the scores at 3 different times on the same diagram			
	Uses statistics that are not distorted by outliers and shows outlying observations clearly.			
	Easy to compare groups			
		E1	3.1a	Any advantage
	Disadvantages			
	The meaning of box and whisker plots may need to be explained to a non-specialist audience.			oe Specialist knowledge is required to understand box and whisker plots.
	The box and whisker plot only provides 5 statistics.			oe Limits the information provided
	Doesn't tell us how large the sample was			
	Exact values are difficult to read			
		E1	3.1a	Any disadvantage
	Total	**9**		

Question 1 - Examiner Comments

Candidates answered the beginning of this question well. However, when it came to commenting on the diagram, some candidates gave comments about the mean and standard deviation, as opposed to the median and the interquartile range or range. Some comments were also too vague, referring, for example, simply to the median of the control group rather than specifically relating that information to the median of the control group one month after the experiment. As a general rule, candidates should focus comments on differences and changes seen in spread and average, rather than making mention of specific quartiles or medians.

Question 1 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade:							
			ALL	A*	A	B	C	D	E	U
6.66	9	74	6.66	8.02	7.51	7.12	6.63	6.17	5.57	4.05



Question 1 - Response A

Answer ALL questions. Write your answers in the spaces provided.

1 Following an illness, patients might need therapy in order to help improve their balance. The therapy offered is called standard balance training.

Giovanni, a physiotherapist, thinks that his patients may be helped further by using a video game designed to improve balance.
He designs an experiment to test his theory.

He randomly assigns patients to one of two groups.
One group receives a course of therapy including standard balance training and the video game. The other group receives only standard balance training.

(a) State which of the two groups described above is the **control** group.

(1)

The group only recieving standard bblance training.

(b) Explain, in context, why Giovanni **randomly** assigns each patient to one of the two groups.

(1)

The patients are assigned randomly to elimenate bias

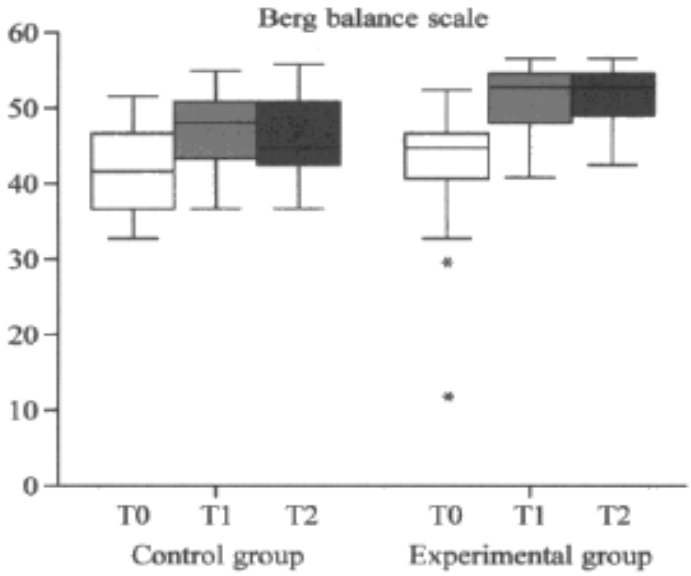
(c) Explain why Giovanni's experiment design is not double-blind.

(1)

~~bea~~ Each group is aware of the groups they are in as the actions they're preforming are different

At the start of the therapy (T0), Giovanni measures the patients' ability to balance using the "Berg balance scale".
He repeats this measure at the end of the course of therapy (T1), and at a follow-up appointment one month later (T2).

Giovanni displays his results using box and whisker plots, shown in **Figure 1**.



[Source: http://dx.doi.org/10.1155/2014/580861]

Figure 1

A **higher score** on the Berg balance scale indicates **better** balance.

(d) Make **four** distinct comments, in context, about Giovanni's experiment and on his results using the information shown in **Figure 1 only**.

(4)

~~The experimental group shows a higher mean/average improvement and this~~ .

The ~~average~~ mean measurement in the experimental group stays the same ~~at~~ the end of therapy (T1) and at the follow up (T2) while ~~that~~ the average measurement decreases between these in the control group

The measurements at the start of therapy (T0) for the experimental group have on average a smaller range though its spread is larger than the control group

Question 1 continued

The experimental group has a higher average improvement in the measure of balance than the control group

Q1

A
B
C

(e) Highlight **one** advantage and **one** disadvantage of Giovanni's choice of box and whisker plots to display his results.

(2)

The averages and spread of each group ~~data~~ of data is displayed in a way that's easy to compare to each other.

The exact measurements aren't included so accurate comparisons can't be made

6 /9

Q1

Part (a)
B1: **Standard** balance training has been recognised.

Part (b)
E1: Reduction of bias has been recognised.

Part (c)
B1: Although the response is a little vague in the detail it recognises that the patients are aware of how they have been assigned by comparing the different actions.

Part (d)
E1: Comment that control group average decreases between T1 and T2.
E0: Comment that experimental group average stays the same not given credit as candidate has used the term 'mean' not 'median/average'.
E0: Comment about smaller range for experimental group not given credit as this is not true due to the outliers.
E0: Comment about higher average improvement not given credit, it is too vague as there are points where the median score for the control group is higher than experimental group at different stages of the experiment.

Part (e)
E1: First comment would gain no credit but 'easy to compare' achieves the mark.
E1: Comment that exact measurements aren't included implies the difficult to read comment.

Question 1 - Response B

1 Following an illness, patients might need therapy in order to help improve their balance.
The therapy offered is called standard balance training.

Giovanni, a physiotherapist, thinks that his patients may be helped further by using a video game designed to improve balance.
He designs an experiment to test his theory.

He randomly assigns patients to one of two groups.
One group receives a course of therapy including standard balance training and the video game. The other group receives only standard balance training.

(a) State which of the two groups described above is the **control** group.

(1)

The other grap that has only Standard Balance training

(b) Explain, in context, why Giovanni **randomly** assigns each patient to one of the two groups.

(1)

Giovanni assigns patients randomly to either group with video game and balance training or the grap with just standard balance training to remove bias & reduce as much bias as possible.

(c) Explain why Giovanni's experiment design is not double-blind.

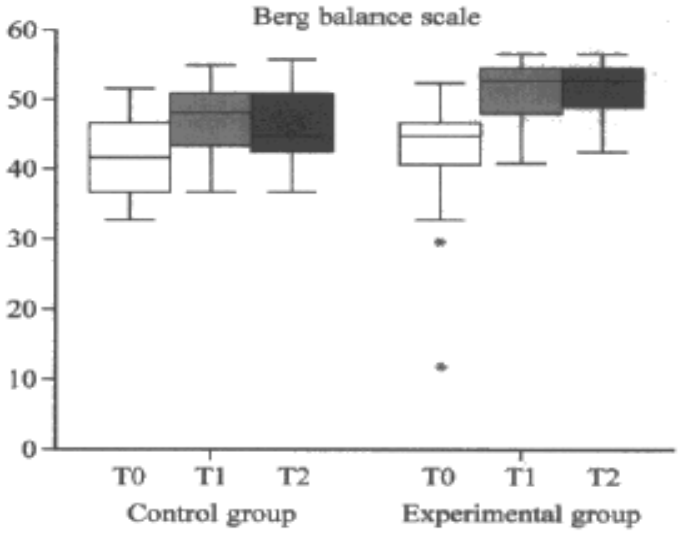
(1)

His experiment is not double blind as he knows who goes to which grap, known as blind trial.

Question 1 continued

At the start of the therapy (T0), Giovanni measures the patients' ability to balance using the "Berg balance scale".
He repeats this measure at the end of the course of therapy (T1), and at a follow-up appointment one month later (T2).

Giovanni displays his results using box and whisker plots, shown in **Figure 1**.



[Source: http://dx.doi.org/10.1155/2014/580861]

Figure 1

A **higher score** on the Berg balance scale indicates **better** balance.

(d) Make **four** distinct comments, in context, about Giovanni's experiment and on his results using the information shown in **Figure 1 only**.

(4)

The experimental group's start of therapy has two outliers at patient's ability 30 and 13.

The follow-up appointment one month later for the control group and start of therapy for experimental group have the same median at patient's ability 45.

The start of therapy for the control group and experimental group have the same IQR, starting from 32 to 52.

Question 1 continued

The end of the course therapy and follow-up appointment for both control group and experimental group have the highest score indicating as better balance.

Q1

A
B
C

(e) Highlight **one** advantage and **one** disadvantage of Giovanni's choice of box and whisker plots to display his results.

(2)

Advantage: Based on Giovanni's choice of box and whisker plots, we can compare and contrast between the control group and experimental group, their differences in patient's ability from start of therapy to the follow-up appointment.

Disadvantage: The y-axis is the patients' ability score, although it has equal spaces between each number, there should be smaller intervals i.e a smaller scale. For example, instead of every 10 do every 5.

7 /9

Q1

Part (a)

B1: Standard balance training has been recognised.

Part (b)

E1: Reduction of bias has been recognised.

Part (c)

B1: Although the response is a little vague in the detail it recognises that Giovanni is aware of how the patients have been assigned.

Part (d)

E1: Comment that experimental group's start of therapy has two outliers.

E1: Comment comparing the median values for T2 of control group and T0 of experimental group.

E0: Comment stating interquartile range is the same at the start of the experiment, not true so no credit given.

E0: Comment about the highest score for control and experimental group, no credit given as comments on highest and lowest values were not given credit.

Part (e)

E1: The use of box plots for comparison has been recognised.

E1: Although the response is a little vague the candidate is implying through their answer that there may be difficulty in reading the values from the *y*-axis.

Question 1 - Response C

1 Following an illness, patients might need therapy in order to help improve their balance. The therapy offered is called standard balance training.

Giovanni, a physiotherapist, thinks that his patients may be helped further by using a video game designed to improve balance.
He designs an experiment to test his theory.

He randomly assigns patients to one of two groups.
One group receives a course of therapy including standard balance training and the video game. The other group receives only standard balance training.

(a) State which of the two groups described above is the **control** group.

(1)

The control group would be those who only receive standard balance training.

(b) Explain, in context, why Giovanni **randomly** assigns each patient to one of the two groups.

(1)

This minimises experimental error, as it decreases the variability of the results. It is done in order to achieve the least bias sample.

(c) Explain why Giovanni's experiment design is not double-blind.

(1)

It is not a double-blind theory as both the physiotherapist and the patient knows what therapy they are recieving, and both will have an effect.

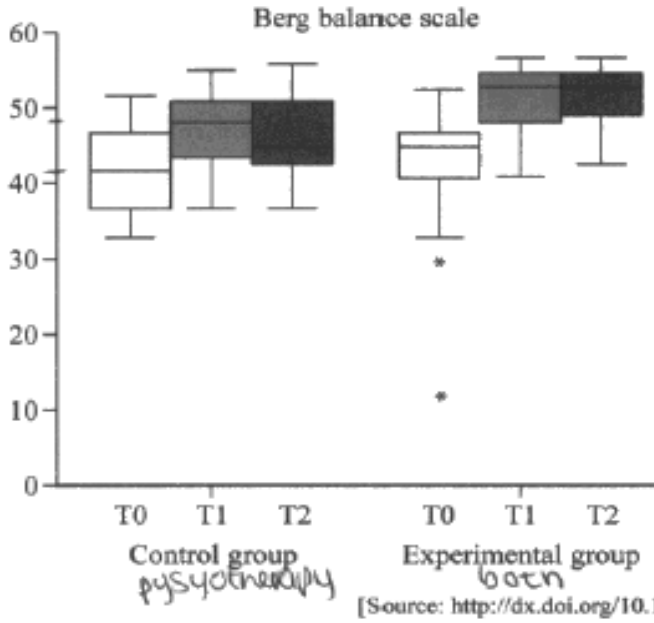
Q1

A
B
C

Question 1 continued

At the start of the therapy (T0), Giovanni measures the patients' ability to balance using the "Berg balance scale".
He repeats this measure at the end of the course of therapy (T1), and at a follow-up appointment one month later (T2).

Giovanni displays his results using box and whisker plots, shown in **Figure 1**.



[Source: http://dx.doi.org/10.1155/2014/580861]

Figure 1

A **higher score** on the Berg balance scale indicates **better** balance.

(d) Make **four** distinct comments, in context, about Giovanni's experiment and on his results using the information shown in **Figure 1 only**.

(4)

Figure 1 indicates that the control group only doing standard balance training do improve by the end of the course. Changing from approximately 41 to a mean score of 48. However, after 1 month (T2) although the range of results is bigger the mean/average balance is lower than the end of the course. Suggesting it may not have long term effects. In comparision to the experimental group whos score increased from around 45 to 53 for both T1 & T2. This suggests using both the standard traing and video games ~~has~~ is more effective and has long term effects/results.

Q1

A
B
C

(e) Highlight **one** advantage and **one** disadvantage of Giovanni's choice of box and whisker plots to display his results.

(2)

One advantage of using box & whisker plots is that it does not just show 1 measure of central tendancy. It shows the mean, the IQR and the range. This can provide lots of data to analyse and doesn't simplify it to one average.

One disadvantage of displaying the results in this way is that it may be misinterpreted, for example for control group T2 the range is a lot larger than for T2 in the experimental group. Someone who doesn't know a lot about graphs may assume this shows better results, whereas it's actually showing the results are more varied. **(Total for Question 1 is 9 marks)**

7 /9

Q1

Part (a)

B1: Standard balance training has been recognised.

Part (b)

E1: Minimises experimental error has been recognised as well as reduction in bias (only one of these comments was required to gain credit).

Part (c)

B1: Candidate has recognised that both Giovanni and the patients will be aware of who is receiving which treatment (again, only one of these comments was required to gain credit).

Part (d)

E1: Comment that both groups improve over the course, this comment was stated separately in the candidate's answer, but then brough together in final statement.

E1: Comment on median decreasing for control group between T1 and T2.

E1: Comment on median staying the same for experimental group between T1 and T2.

E0: Comment on the range being larger for control group at T2 not given credit as not obvious change from the diagram and the candidate has not been clear whether they are comparing T2 with T0 or T1.

Examiner comments: Although the candidate used the word mean in their response to (d) this was not penalised as they had used the word average as well.

Part (e)

E0: Candidate has described what is contained within a box plot (ignoring the mean comment), the use of the word 'analyse' does not give enough evidence that this information can be used for comparison, no credit given.

E1: Recognises specific statistical knowledge required.

Question 2

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

Question 2 - Question

2 Shelly works as a project manager for a clothing design company. She is in charge of a project to produce a new jacket design. A pattern is then made of her design.

Shelly sends the pattern to the sample production department to be made into a sample jacket. From past experience, it is known that it should take between 5 and 13 working hours for a sample jacket to be produced from a pattern.

The uniform distribution is considered a suitable model for the time taken to produce such a sample jacket from a pattern.

A diagram of Shelly's modelling distribution is shown in **Figure 2**.

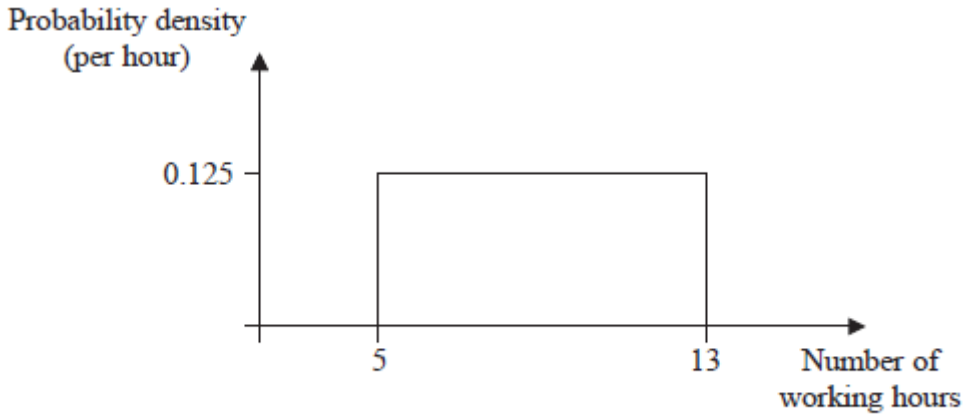


Figure 2

(a) Explain why the height of the modelling distribution in **Figure 2** is 0.125

(1)

Use Shelly's model and **Figure 2** to answer parts (b) to (e).

(b) State the expected time that it will take to produce her sample jacket from the pattern.

(1)

(c) State the probability that it will take less than 4 hours to produce her sample jacket from the pattern.

(1)

One week, Shelly sends the pattern to the sample production department on Tuesday, 7 hours before it closes for the day.

(d) Calculate the probability that her sample jacket will be produced from the pattern before the sample production department closes on Tuesday.

(1)

That same week, Shelly's boss would like the sample jacket to be produced from the pattern by 11 am on Wednesday.
The sample production department opens at 8 am.

(e) Given that the sample jacket was **not** produced from the pattern before the sample production department closed on Tuesday, calculate the probability that it is ready by 11 am on Wednesday.

(2)

(Total for Question 2 is 6 marks)

Question 2 - Mark Scheme

Question	Scheme	Marks	AO	Notes
2(a)	$\frac{1}{(13-5)} = 0.125$	E1	2.1a	oe May be stated in words: "Total probability/area is one. The base of the rectangle is 8. One divided by 8 is 0.125" Accept "1/8 = 0.125"
2(b)	9 (working) hours	B1	1.2	
2(c)	0	B1	2.1a	oe "zero"
2(d)	$((7-5) \times 0.125 =) \ 0.25$	B1	1.2	
2(e)	$\frac{3 \times 0.125}{1 - "0.25"}$	M1ft	1.2	PI 0.375/0.75 ft on candidate's 0.25 Other appropriate methods allowed.
	0.5	A1	1.2	oe
	Total	**6**		

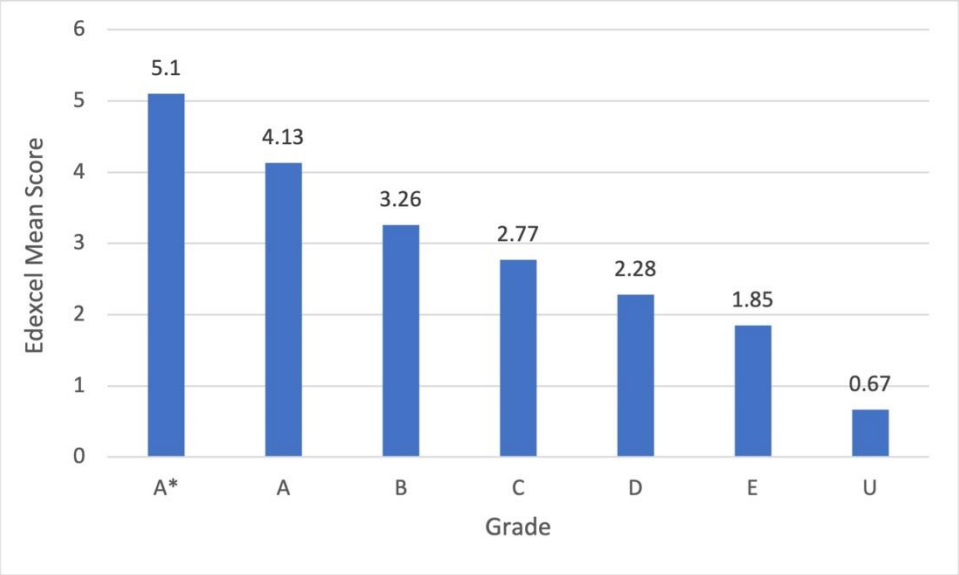
Q2

Question 2 - Examiner Comments

Candidates did very well on this question in general, though conditional probability caused issues for some candidates in the last part of this question.

Question 2 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade:							
			ALL	A*	A	B	C	D	E	U
2.92	6	49	2.92	5.10	4.13	3.26	2.77	2.28	1.85	0.67



Question 2 - Response A

2 Shelly works as a project manager for a clothing design company. She is in charge of a project to produce a new jacket design. A pattern is then made of her design.

Shelly sends the pattern to the sample production department to be made into a sample jacket. From past experience, it is known that it should take between 5 and 13 working hours for a sample jacket to be produced from a pattern.

The uniform distribution is considered a suitable model for the time taken to produce such a sample jacket from a pattern.

A diagram of Shelly's modelling distribution is shown in **Figure 2**.

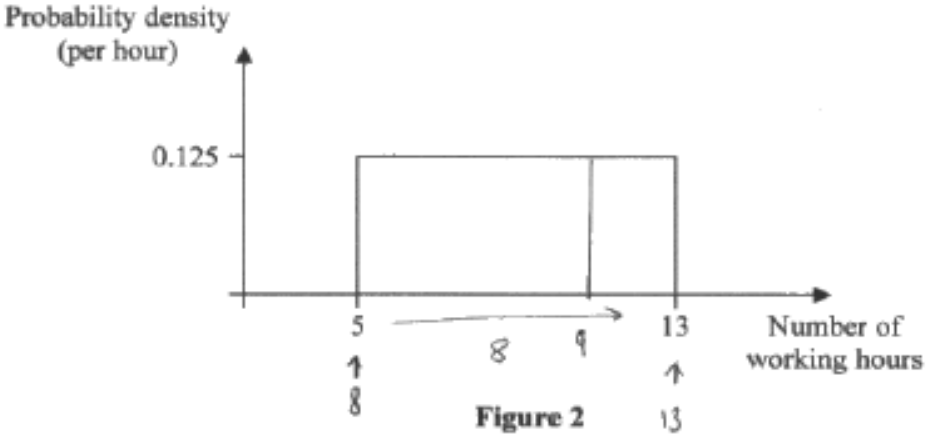


Figure 2

(a) Explain why the height of the modelling distribution in **Figure 2** is 0.125

(1)

$w \times h = 1 \Rightarrow h = \frac{1}{w} \rightarrow \frac{1}{8} = 0.125$

$13 - 5 = 8$

Use Shelly's model and **Figure 2** to answer parts (b) to (e).

(b) State the expected time that it will take to produce her sample jacket from the pattern.

(1)

8 hours

5 → 13 = 8 hours

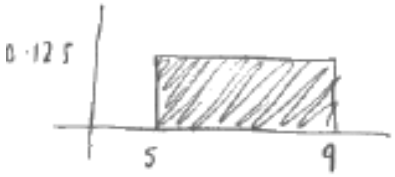
(c) State the probability that it will take less than 4 hours to produce her sample jacket from the pattern.

(1)

$P(X < 4)$ $0.125 = 0.013888$

$13 - 4 = 9$ $4 \times 0.125 = 0.5$

$= 0.0139$

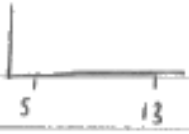


Question 2 continued $5 \rightarrow 6$ 1

One week, Shelly sends the pattern to the sample production department on Tuesday, 7 hours before it closes for the day.

(d) Calculate the probability that her sample jacket will be produced from the pattern before the sample production department closes on Tuesday.

(1)



7 hours before

That same week, Shelly's boss would like the sample jacket to be produced from the pattern by 11 am on Wednesday.
The sample production department opens at 8 am. 3 hours. > 11

(e) Given that the sample jacket was **not** produced from the pattern before the sample production department closed on Tuesday, calculate the probability that it is ready by 11 am on Wednesday.

(2)

take 3 hours $= 3 \times 0.125 = 0.375$. after 11.

given that not produced =

$P(R \mid NP) = \frac{0.375}{0.625} = 0.6$

$3 \times 0.125 = 0.375$

$1 - 0.375 = 0.625$.

Q2

A B C

2 /6

Part (a)

E1: Candidate has recognised total area of 1 and calculated the width to get 8 (this would have been acceptable through the diagram as well).

Part (b)

B0: A few candidates did not recognise the word expected as the mean (or median due to distribution being symmetric). The width being 8 was a common misconception on this question.

Part (c)

B0: Again 0.5 was a common misconception on this part. The important thing to note on this question is that the distribution for how long it takes to make a jacket is specifically between 5 and 13 therefore the answer for $\mathrm{P}(X < 4)$ is impossible.

Part (d)

B0: No attempt made by candidate.

Part (e)

M1: Credit given as candidate has calculated $3 \times 0.125 = 0.375$.

A0: Did not reach correct solution

Question 2 - Response B

2 Shelly works as a project manager for a clothing design company. She is in charge of a project to produce a new jacket design. A pattern is then made of her design.

Shelly sends the pattern to the sample production department to be made into a sample jacket. From past experience, it is known that it should take between 5 and 13 working hours for a sample jacket to be produced from a pattern.

The uniform distribution is considered a suitable model for the time taken to produce such a sample jacket from a pattern.

A diagram of Shelly's modelling distribution is shown in **Figure 2**.

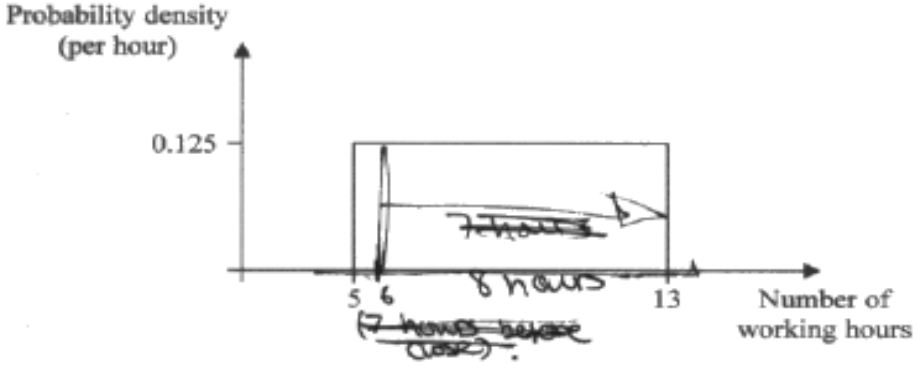


Figure 2

(a) Explain why the height of the modelling distribution in **Figure 2** is 0.125

(1)

13 − 5 = 8 1 ÷ 8 = 0·125 (Area ÷ base = height)

Area = 1 Area = base × height.

8 × 0·125 = 1

Use Shelly's model and **Figure 2** to answer parts (b) to (e).

(b) State the expected time that it will take to produce her sample jacket from the pattern.

(1)

(c) State the probability that it will take less than 4 hours to produce her sample jacket from the pattern.

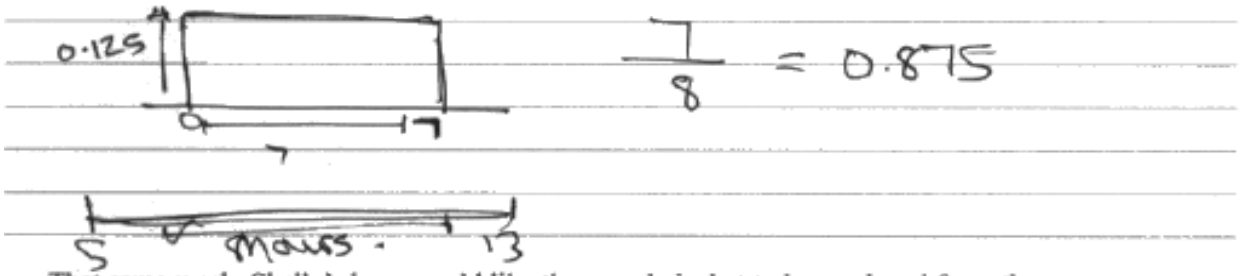
(1)

P(x < 4) = 0 Takes her 5 hours minimum.

One week, Shelly sends the pattern to the sample production department on Tuesday, 7 hours before it closes for the day.

(d) Calculate the probability that her sample jacket will be produced from the pattern before the sample production department closes on Tuesday.

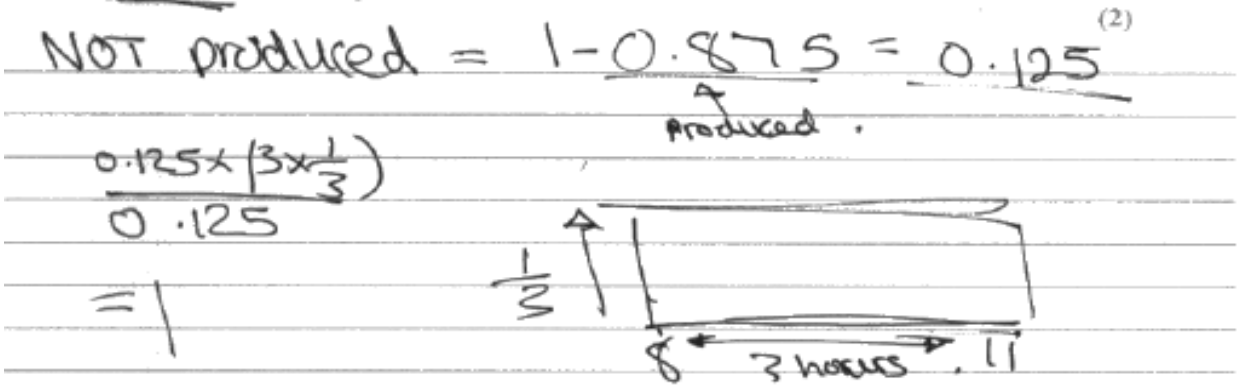
(1)



That same week, Shelly's boss would like the sample jacket to be produced from the pattern by 11 am on Wednesday.
The sample production department opens at 8 am.

(e) Given that the sample jacket was **not** produced from the pattern before the sample production department closed on Tuesday, calculate the probability that it is ready by 11 am on Wednesday.

(2)



3 /6

Q2

Part (a)

E1: Candidate has recognised total area of 1 and calculated the width to get 8. The candidate then performed a check at the end $0.125 \times 8 = 1$, which is good practice but not necessary for the mark to be awarded.

Part (b)

B0: No attempt made by candidate.

Part (c)

B1: '0' gained the credit but again a good explanation included.

Part (d)

B0: Candidate has not used the initial information given about the distribution for how long it takes to make a jacket is specifically between 5 and 13. The diagram they have drawn is not the correct interpretation of this question.

Part (e)

M1: Credit given as candidate has calculated $1 - 0.875$.

A0: Did not reach the correct solution.

Question 2 - Response C

2 Shelly works as a project manager for a clothing design company. She is in charge of a project to produce a new jacket design. A pattern is then made of her design.

Shelly sends the pattern to the sample production department to be made into a sample jacket. From past experience, it is known that it should take between 5 and 13 working hours for a sample jacket to be produced from a pattern.

The uniform distribution is considered a suitable model for the time taken to produce such a sample jacket from a pattern.

A diagram of Shelly's modelling distribution is shown in **Figure 2.**

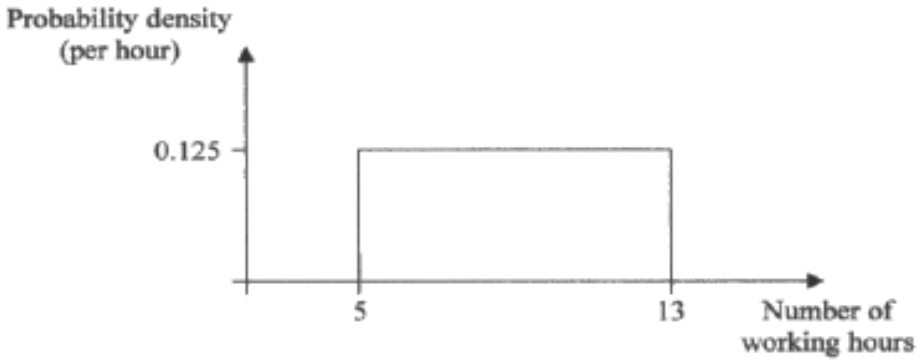


Figure 2

(a) Explain why the height of the modelling distribution in **Figure 2** is 0.125

(1)

$\frac{1}{b-a} \rightarrow \frac{1}{13-5} = \frac{1}{8} = 0.125$

Use Shelly's model and **Figure 2** to answer parts (b) to (e).

(b) State the expected time that it will take to produce her sample jacket from the pattern.

(1)

7.5 minutes.

(c) State the probability that it will take less than 4 hours to produce her sample jacket from the pattern.

(1)

0

One week, Shelly sends the pattern to the sample production department on Tuesday, 7 hours before it closes for the day.

(d) Calculate the probability that her sample jacket will be produced from the pattern before the sample production department closes on Tuesday.

(1)

$1 \times 0.125 = 0.125$

That same week, Shelly's boss would like the sample jacket to be produced from the pattern by 11 am on Wednesday.
The sample production department opens at 8 am.

(e) Given that the sample jacket was **not** produced from the pattern before the sample production department closed on Tuesday, calculate the probability that it is ready by 11 am on Wednesday.

(2)

$\frac{0.375}{0.875} = 0.429$

3 /6

Q2

Part (a)
E1: Candidate has recognised total area of 1 and calculated the width to get 8.

Part (b)
B0: Incorrect answer.

Part (c)
B1: Correct solution.

Part (d)
B0: Incorrectly calculated $7-5$ as 1.

Part (e)
M1: Candidate could have achieved this mark either for $3\times 0.125 = 0.375$ or for $1-0.125$
A0: Did not reach correct solution.

Question 3

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

Question 3 - Question

3 Klazine is a researcher interested in nutrition. She believes that involving a child in preparing their own meal affects what they choose to eat at that meal.

She conducted a study involving 47 children aged 6 to 10 years.

25 children ate a meal that they had prepared with their parent.

The other 22 children ate a meal that their parent had prepared alone.

All of the meals contained the same four ingredients: pasta, chicken, cauliflower, and salad.

Klazine measured the total weight, in grams (g), of each ingredient eaten by the children in each group. Her results for **salad** are shown in **Figure 3**.

	Salad eaten (g)		
	n	**Mean**	**SD**
Child prepared with parent	25	96.4	61.5
Parent prepared alone	22	54.8	35.4

[Data source: https://doi.org/10.1016/j.appet.2014.03.030]

Figure 3

(a) Calculate a 95% confidence interval for the weight of salad eaten by the children whose parent prepared the meal alone.

(3)

A 95% confidence interval for the weight, in grams, of salad eaten by the children who prepared the meal with their parent is (71.0, 121.8)

(b) Give a full explanation of the evidence provided about Klazine's belief by the confidence interval in (a), together with the interval given above.

(3)

Klazine measured the weight of cauliflower, in grams (*g*), eaten by the children in each group. Her results are shown in **Figure 4**.

	Cauliflower eaten (g)		
	n	**Mean**	**SD**
Child prepared with parent	25	110.5	50.1
Parent prepared alone	22	89.7	51.3

[Data source: https://doi.org/10.1016/j.appet.2014.03.030]

Figure 4

(c) Making any necessary assumptions, use a ***t*-test** to investigate whether the children who prepared their meal with a parent ate over 10 grams more cauliflower, on average, than the children whose parent prepared the meal alone.

(8)

There is an assumption about variance that is necessary for the *t*-test in (c) to be valid.

(d) State the assumption about variance, in context, and comment on its appropriateness, given the data in **Figure 4**.

(2)

Klazine's colleague, Aurora, is concerned that the sample sizes used in their study are too small to be able to detect a difference between the two groups of children.

(e) Discuss Aurora's concern about **sample sizes**.

Include **one** reference to the statistical techniques applied in (a) and in (c) and their conclusions and **one** reference to the effect of increasing sample sizes.

(2)

(f) If Klazine uses a much larger sample of children in her study, what effect would you expect this to have on the confidence intervals in (a) and in (b)?

(1)

(Total for Question 3 is 19 marks)

Question 3 - Mark Scheme

Question	Scheme	Marks	AO	Notes
3(a)	(n = 22) t = 2.080	B1	1.3	t-value Condone z=1.96
	$54.8 \pm 2.080\left(\frac{35.4}{\sqrt{22}}\right)$	M1ft	1.3	Formula correct ft on B1
	(39.1, 70.5)	A1	1.3	awfw 39~39.1, 70.4~70.5
3(b)	The confidence intervals do not overlap.	M1	2.1b	Comparison of candidates CI from (a) with given CI
	So there is significant evidence to support Klazine's belief.	E1dep	2.1b	oe dep M1 Evidence supports Klazine's belief
	Involving a child in preparing their own meal affects what they choose to eat at that meal (this can be seen for salad)	E1dep	2.1a	dep on M1 Response can be more specific, e.g. "children who help prepare their own meal eat more salad" or "children who do not prepare their own meal eat less salad"
				Award E marks independently from one another. May be seen in one sentence. Disregard references to health.

Q3

A
B
C

Question	Scheme	Marks	AO	Notes
3(c)	C signifies child prepared with parent D signifies parent prepared alone			Can be the other way around with a negative ts and cv
	H_0: $\mu_C - \mu_D = 10$ H_1: $\mu_C - \mu_D > 10$	B1	1.3	oe both, subscripts clearly defined
	$s_p^2 = \frac{(25-1)50.1^2 + (22-1)51.3^2}{25+22-2}$	M1	1.3	PI
	= 2567	A1	1.3	PI awrt 2560~2570 or s_p = awfw 50.5~50.7
	$ts = \frac{(110.5-89.7)-10}{\sqrt{2567\left(\frac{1}{25}+\frac{1}{22}\right)}}$	M1	1.3	PI Numerator or denominator correct (ignoring -10)
		M1dep	1.3	PI Fully correct numerator with candidates s_p^2
	Critical value method: $10 + 1.679 \times \sqrt{219.344} = 34.866$			1.679 scores M1 Formula scores M1dep
	= 0.729	A1	1.3	ts awfw 0.728~0.730 or 34.9
	cv = 1.679 or p-value = 0.235	B1	1.3	Any of 1.679, 0.235, 20.8 selected for comparison
	(0.729<1.679, No significant evidence to reject H_0) There is insufficient evidence that the children who prepared their meal with a parent ate over 10 grams more cauliflower, on average, than the children who did not.	E1dep	2.1a	oe 0.235>0.05 or 20.8<34.9 in context. Must contain element of doubt. dep on whole test correct

Question	Scheme	Marks	AO	Notes
3(d)	The (population) **variance** of the **weight** of cauliflower **eaten** by **children who prepare their meal with a parent** should be very **similar** to the (population) **variance** of the **weight** of cauliflower **eaten** by **children whose parent prepared their meal alone.**	B1	1.3	oe Variances equal/similar/close in context. Accept "Variance of cauliflower should be the same for children who prepare and those that don't."
	(The sample variances/sds are very close so) there is no evidence that the assumption is not valid	E1dep	2.1b	oe Clear indication that the assumption is valid. Dep "variances similar" assumption stated. Not dep gaining B1 mark
3(e)	The t-distribution techniques used in (a) and (c) can be (validly) applied to small samples.			oe Accept "t-test can be used with small samples"
	Populations might not be normally distributed (and n is not large enough for CLT to apply)			Do not accept CLT doesn't apply by itself
		E1	3.1a	Either
	Using larger sample sizes will increase the power of the test (to detect a given difference in means).			oe Accept "large sample means more likely to find a small difference" or "large sample means less likely to make a Type II error" or "more reliable"
	Would use z instead of t			
		E1	3.1a	Either
3(f)	The width of the confidence intervals should decrease.	B1	2.1b	oe Confidence intervals narrower. Condone smaller
	Total	**19**		

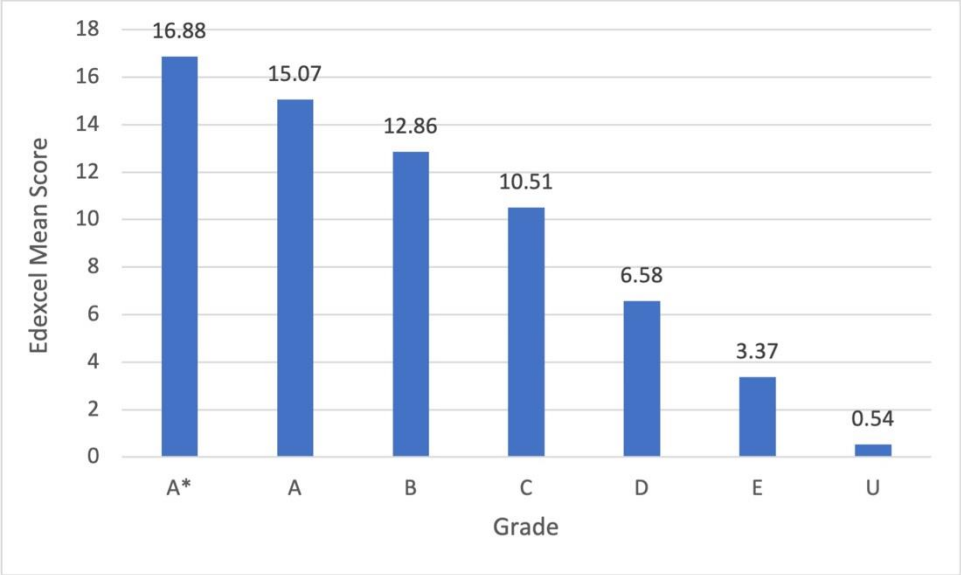
Q3

Question 3 - Examiner Comments

Some candidates failed to realise that a t value was required for this question, rather than a z value. In part (b), most candidates compared their confidence intervals, and some candidates gave answers commenting on what this showed, but then did not comment on Klazine's belief as they were asked to do. There were lots of possible mistakes in (c) that could be made if candidates weren't taking care, but this is a standard hypothesis test and done well by many. Parts (d) and (e) required comments, and it should be noted that whilst it is possible to use t when $n \geq 30$, a z-test is more powerful in this situation. Some excellent comments were seen from the most able candidates in these parts.

Question 3 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade:							
			ALL	A*	A	B	C	D	E	U
10.13	19	53	10.13	16.88	15.07	12.86	10.51	6.58	3.37	0.54



Question 3 - Response A

3 Klazine is a researcher interested in nutrition. She believes that involving a child in preparing their own meal affects what they choose to eat at that meal.

She conducted a study involving 47 children aged 6 to 10 years.

25 children ate a meal that they had prepared with their parent.

The other 22 children ate a meal that their parent had prepared alone.

All of the meals contained the same four ingredients: pasta, chicken, cauliflower, and salad.

Klazine measured the total weight, in grams (g), of each ingredient eaten by the children in each group. Her results for **salad** are shown in **Figure 3**.

	Salad eaten (g)		
	n	Mean	SD
Child prepared with parent	25	96.4	61.5
Parent prepared alone	22	54.8	35.4

[Data source: https://doi.org/10.1016/j.appet.2014.03.030]

Figure 3

(a) Calculate a 95% confidence interval for the weight of salad eaten by the children whose parent prepared the meal alone. (3)

$\bar{x} \pm z\frac{\sigma}{\sqrt{n}}$ $z = 1.96$ $\bar{x} = 54.8$ $SD = 35.4$ $n = 22$

$54.8 \pm 1.96 \times \frac{35.4}{\sqrt{22}}$

(40.01, 69.59) to 2dp.

A 95% confidence interval for the weight, in grams, of salad eaten by the children who prepared the meal with their parent is (71.0, 121.8)

(b) Give a full explanation of the evidence provided about Klazine's belief by the confidence interval in (a), together with the interval given above. (3)

Prepared = (71.0, 121.8)

alone = (40.01, 69.59) = (40.0, 69.6)

when prepared with parents confidence interval is much higher there is no overlap inbetween the two intervals suggesting a massive difference in results significant evidence to suggest Klazines belief has merit. and

Q3

A B C

Klazine measured the weight of cauliflower, in grams (g), eaten by the children in each group. Her results are shown in **Figure 4**.

		Cauliflower eaten (g)		
		n	Mean	SD
x	Child prepared with parent	25	110.5	50.1
y	Parent prepared alone	22	89.7	51.3

[Data source: https://doi.org/10.1016/j.appet.2014.03.030]

Figure 4

(c) Making any necessary assumptions, use a ***t*-test** to investigate whether the children who prepared their meal with a parent ate over 10 grams more cauliflower, on average, than the children whose parent prepared the meal alone.

(8)

$$\frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}}$$

$\bar{x} = 110.5$ $H_0: \mu_x - \mu_y \leq 10$

$\bar{y} = 89.7$ $H_1: \mu_x - \mu_y > 10$

$SD\ \sigma_x = 50.1$

$SD\ \sigma_y = 51.3$

$n_x = 25$

$n_y = 22$

$$= \frac{(110.5-89.7)-(11)}{\sqrt{\frac{50.1^2}{25}+\frac{51.3^2}{22}}}$$

$$= \frac{(20.8)-(11)}{14.83316125} = \frac{9.8}{14.83316125}$$

$= 0.660681822$

CV = 1.679

1.679 > 0.66

TS = 0.66

Accept Ho insignificant evidence to suggest that on average children who prepare their own meals are likely to eat more than 10 grams more.

Q3

A B C

There is an assumption about variance that is necessary for the *t*-test in (c) to be valid.

(d) State the assumption about variance, in context, and comment on its appropriateness, given the data in **Figure 4**.

(2)

Variance or standard deviation is known for sample and isn't estimated.

Klazine's colleague, Aurora, is concerned that the sample sizes used in their study are too small to be able to detect a difference between the two groups of children.

(e) Discuss Aurora's concern about **sample sizes**.

Include **one** reference to the statistical techniques applied in (a) and in (c) and their conclusions and **one** reference to the effect of increasing sample sizes.

(2)

Sample sizes should be larger for a t-test. t-test are usually used for samples that are large this sample is classed as large but it isn't a massive sample and the bigger the sample the more reliable. A larger sample will change results and could be enough data to be more representative of the sample.

(f) If Klazine uses a much larger sample of children in her study, what effect would you expect this to have on the confidence intervals in (a) and in (b)?

the confidence intervals will change but will also become more suited for the new sample and will change more and more until close to the population.

Q3

A B C

8 /19

Part (a)

B1: Use of $z = 1.96$

M1: Use of the correct confidence interval formula, although not substituted into.

A0: Based on the content of the specification the *t*-distribution should be used for samples less than 30.

Part (b)

M1ft: Although the incorrect confidence interval was found in (a) the candidate has correctly compared the overlap.

E1: There is a clear statement about Klazine's belief, it was common for candidates not to include this statement which is what the question is actually asking for.

E0: The confidence intervals not overlapping has not been interpreted within the context of this question. From the candidate's response we cannot tell what is being investigated just that results are significant.

Part (c)

B1: Hypotheses given, although unusual to see H_0 written as $\leq$ it is statistically correct.

M0: As candidate has used a *z*-distribution no pooled variance was calculated.

A0: As candidate has used a *z*-distribution no pooled variance was calculated.

M1: Attempt has been made to calculate the test statistic.

M0: Could not gain credit as had subtracted 11 rather than the 10 stated in hypotheses.

A0: Did not obtain correct test statistic for their method.

B1: Even though candidate has now found a *t*-critical value (disagreeing with previous work) credit was still given.

E0: Conclusion has context based on the information from the question and candidate has recognised this as a test on the average, but this final mark was dependent on whole test being correct.

Part (d)

B0: Candidate has not stated the assumption that variances must be similar/equal for an unpaired t test to be valid.

E0: No attempt made by candidate.

Part (e)

E0: Candidate has made the opposite statement to that required.

E1: Correct comment that a bigger sample will create a more reliable conclusion

Part (f)

B0: The candidate recognised that there would be a change but haven't been specific in what change would occur.

Question 3 - Response B

3 Klazine is a researcher interested in nutrition. She believes that involving a child in preparing their own meal affects what they choose to eat at that meal.

She conducted a study involving 47 children aged 6 to 10 years.

25 children ate a meal that they had prepared with their parent.

The other 22 children ate a meal that their parent had prepared alone.

All of the meals contained the same four ingredients: pasta, chicken, cauliflower, and salad.

Klazine measured the total weight, in grams (g), of each ingredient eaten by the children in each group. Her results for **salad** are shown in **Figure 3**.

	Salad eaten (g)		
	n	Mean	SD
Child prepared with parent	25	96.4	61.5
Parent prepared alone	22	54.8	35.4

[Data source: https://doi.org/10.1016/j.appet.2014.03.030]

Figure 3

(a) Calculate a 95% confidence interval for the weight of salad eaten by the children whose parent prepared the meal alone.

(3)

left = 36.582
~~right = 73.018~~

$\bar{x} \pm t_{n-1} \frac{s}{\sqrt{n}}$

$t_{21} = 2.080$

~~$\mu = (36.582, 73.018)$~~

$\mu = (39.105, 70.495)$

mean s.d
w 96.4 61.5
54.8 35.4

Question 3 continued

A 95% confidence interval for the weight, in grams, of salad eaten by the children who prepared the meal with their parent is (71.0, 121.8)

(39.1, 70.5)

(b) Give a full explanation of the evidence provided about Klazine's belief by the confidence interval in (a), together with the interval given above.

(3)

- both means lie between their confidence intervals. Therefore, it can be suggested that children who prepare with their parents will eat more of the salad on average. Klazine's belief is correct.

Q3

A B C

Klazine measured the weight of cauliflower, in grams (g), eaten by the children in each group. Her results are shown in **Figure 4**.

		Cauliflower eaten (g)		
		n	Mean	SD
A	Child prepared with parent	25	110.5	50.1
B	Parent prepared alone	22	89.7	51.3

19.7

[Data source: https://doi.org/10.1016/j.appet.2014.03.030]

Figure 4

(c) Making any necessary assumptions, use a ***t*-test** to investigate whether the children who prepared their meal with a parent ate over 10 grams more cauliflower, on average, than the children whose parent prepared the meal alone.

(8)

$H_0: \mu_A = \mu_B + 10$

$H_1: \mu_A < \mu_B + 10$

where A = with Parent

B = Alone

t = 0.7292

p = 0.765

df = 45

sp = 50.664

0.765 > 0.05

Do not reject H_0, insufficient evidence to suggest children who prepared the meal with parent ate over 10 grams more cauliflower than children whose parent prepared alone.

Q3

There is an assumption about variance that is necessary for the *t*-test in (c) to be valid.

(d) State the assumption about variance, in context, and comment on its appropriateness, given the data in **Figure 4**.

(2)

Variance must be constant.
$\sigma_a^2 = \sigma_b^2 = \sigma^2$

Not as appropriate as $n > 30$ & so CLT applies. does not apply.

Klazine's colleague, Aurora, is concerned that the sample sizes used in their study are too small to be able to detect a difference between the two groups of children.

(e) Discuss Aurora's concern about **sample sizes**.

Include **one** reference to the statistical techniques applied in (a) and in (c) and their conclusions and **one** reference to the effect of increasing sample sizes.

(2)

- both groups should have an equal sample size so results are fair. The extra 3 participants in group b could skew the data.

- Increasing sample size more than 30 will improve the validity of results. Generates more accurate findings that are more representative; and can be generalised.

(f) If Klazine uses a much larger sample of children in her study, what effect would you expect this to have on the confidence intervals in (a) and in (b)?

(1)

As sample size increases, ~~confidence intervals~~ the range between confidence intervals will decrease and they become more specific.

Q3

A
B
C

11 /19

Part (a)

B1: Use of $t = 2.080$

M1: Use of the correct confidence interval formula, although not substituted into.

A1: Confidence interval is correct to 3 significant figures.

Part (b)

M0: The candidate has not compared the confidence intervals in terms of overlap, contained entirely within etc. Therefore, no marks can be gained in this part.

E0: There is a clear statement about Klazine's belief. However, this is a dependent mark so without a comparison of the confidence intervals no credit can be given.

E0: Their conclusion has been interpreted into the context of the question. However, this is a dependent mark so without a comparison of the confidence intervals no credit can be given.

Part (c)

B0: Candidate has correct H_0 but H_1 has an incorrect inequality.

M1: Pooled variance not calculated but given due to correct s_p stated in working.

A1: Correct pooled standard deviation stated.

M1: Implied by $t = 0.7292$

M1: Implied by $t = 0.7292$

A1: Obtained correct value for $t = 0.7292$

B1: 0.765 selected for comparison.

E0: Although candidate has obtained the 'incorrect' p-value (it does agree with their hypotheses) this mark did not require a comparison. Conclusion has context based on the information from the question, but this final mark was dependent on whole test being correct.

Part (d)

B1: Although a little vague the candidate has implied the assumption that variances must be similar/equal for an unpaired t test to be valid.

E0: Candidate has not used their correct previous statement to inform testing the appropriateness.

Part (e)

E0: Candidate has not commented about the techniques that have been used in (a) or (c). Their comment is to do with equal sample sizes.

E0: The candidate has again made the common misunderstanding between the terms accurate and reliable; it is not that using a z-distribution over a t-distribution creates a more accurate conclusion but that it more powerful/reliable.

Part (f)

B1: Correct answer.

Question 3 - Response C

3 Klazine is a researcher interested in nutrition. She believes that involving a child in preparing their own meal affects what they choose to eat at that meal.

She conducted a study involving 47 children aged 6 to 10 years.

25 children ate a meal that they had prepared with their parent.

The other 22 children ate a meal that their parent had prepared alone.

All of the meals contained the same four ingredients: pasta, chicken, cauliflower, and salad.

Klazine measured the total weight, in grams (g), of each ingredient eaten by the children in each group. Her results for **salad** are shown in **Figure 3**.

	Salad eaten (g)		
	n	Mean	SD
Child prepared with parent	25	96.4	61.5
Parent prepared alone	22	54.8	35.4

sample

[Data source: https://doi.org/10.1016/j.appet.2014.03.030]

Figure 3

(a) Calculate a 95% confidence interval for the weight of salad eaten by the children whose parent prepared the meal alone.

(3)

22 < 30 ∴ small ⇒ t · $\nu = 22 - 1 = 21$

$t = 2.080$

$54.8 \pm 2.080 \times \frac{35.4}{\sqrt{22}}$

$= (39.1, 70.5)$ g

A 95% confidence interval for the weight, in grams, of salad eaten by the children who prepared the meal with their parent is (71.0, 121.8)

(b) Give a full explanation of the evidence provided about Klazine's belief by the confidence interval in (a), together with the interval given above.

(3)

∵ the confidence interval were kids helped prepare the meal was higher than the other ∴ significant there is evidence that when kids prepare their own food it affects what they choose to eat

however there was more variation and longer so it may not be completely true

∴ klazines belief is true.

Q3

A B C

Klazine's colleague, Aurora, is concerned that the sample sizes used in their study are too small to be able to detect a difference between the two groups of children.

(e) Discuss Aurora's concern about **sample sizes**.

Include **one** reference to the statistical techniques applied in (a) and in (c) and their conclusions and **one** reference to the effect of increasing sample sizes.

(2)

• the conclusions for (A) and (C) differed but the amount extra differ

• increasing the sample sizes will make the data more accurate and representative but also means central limit theorem applies. so the sample means will be normally distributed so we wont have to assume if they are normally distributed

• the conclusions to (a) and (c) differed so maybe using a larger sample would data more representative could change that so a definitive answer could occur.

(f) If Klazine uses a much larger sample of children in her study, what effect would you expect this to have on the confidence intervals in (a) and in (b)?

(1)

the confidence intervals would be smaller as $\frac{\sigma}{\sqrt{n}}$ the sample dividing the standard deviation is large however more people usually causes more variation.

16 /19

Q3

Part (a)

B1: Use of $t = 2.080$

M1: Use of the correct confidence interval formula

A1: Confidence interval is correct to 3 significant figures (units are not necessary but good practice).

Part (b)

M1: The response is a little vague but the word 'higher' was given credit. Clearly talking about overlaps or being included is preferred.

E1: There is a clear statement about Klazine's belief, it was common for candidates not to include this statement which is what the question is asking for.

E1: The confidence intervals not overlapping has been interpreted correctly within the context of this question.

Part (c)

B1: Hypotheses given but would be better to label the table, so it is clearer what W and A represent.

M1: Pooled variance formula calculated

A1: Correct pooled variance found

Candidate uses an alternative valid method than that shown in the mark scheme.

M1: *t*-value of 1.679 seen

M1: Using critical value formula correctly

A1: Obtained correct critical value of 34.9

B1: Calculating the difference of sample means and using it as comparison.

E1: There was no mark here for comparison of test statistic and critical value available in this question. Conclusion has context based on the information from the question and candidate has recognised that this is a test on the average.

Part (d)

B1: Candidate has stated the assumption that variances must be similar/equal for an unpaired *t* test to be valid.

E0: By calculating the variances the candidate has interpreted the difference between them as large (121.68). However, compared to the values of the variances this would not be considered to be a large difference, for example if the values had been 25.10 and 26.31 it is likely the candidate would have reached the opposite conclusion.

Part (e)

E0: Candidate has not commented about the techniques that have been used in (a) or (c)

E0: Comment about central limit theorem is correct. However, it is not relevant in terms of this question specifically. The candidate may have gained credit if they had not used to word 'accurate' it is not that using a *z*-distribution over the *t*-distribution create a more accurate conclusion but that it more powerful/reliable.

Part (f)

B1: Even though candidate has misquoted the standard error they have stated that it would be 'smaller'. Therefore, further comments were ignored.

Question 4

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

Question 4 - Question

4 Petra is an engineer who works on an offshore wind farm.

Petra organises the repairs of wind turbines that fail. The wind farm contract states that wind turbines must be repaired on the same day that they fail.

Failures appear to happen at random, and at a fairly consistent rate over the year.

In order to organise repairs, Petra uses a Poisson model with $\lambda = 2.8$ for the number of failures per year for each wind turbine.

[Data source: https://doi.org/10.1177%2F0957650915597560]

Use Petra's model to answer (*a*) to (*f*).

(a) Calculate the standard deviation of the number of failures per year for a wind turbine at the wind farm.

(1)

(b) Find the probability that a wind turbine at the wind farm fails exactly three times in a given year.

(1)

(c) Find the probability that a wind turbine at the wind farm fails more than seven times over a three-year period.

(2)

(d) State fully the distribution of the time, in years, between successive failures of a wind turbine at the wind farm.

(2)

(e) Calculate the expected time, in years, between successive failures of a wind turbine at the wind farm.

(1)

(f) Given that a particular wind turbine at the wind farm has not failed for two consecutive years, calculate the probability that it fails within the next six months.

(3)

The wind farm has six wind turbines. During the past five years, there have been 84 wind turbine failures at the wind farm.

(g) Show how Petra used this information to calculate the value of λ

(1)

(h) Make **three** criticisms, in context, of Petra's use of a Poisson model with λ = 2.8 for the number of failures per year for each wind turbine.

You may consider both the use of a Poisson model and the value of λ

(3)

(Total for Question 4 is 14 marks)

Question 4 - Mark Scheme

Question	Scheme	Marks	AO	Notes
4(a)	$(\sqrt{2.8} =)$ 1.67	B1	1.2	awrt 1.67
4(b)	0.222	B1	1.2	awfw 0.222~0.223
4(c)	$\lambda = 8.4$	M1	1.2	PI rescaling
	$(1 - 0.399 =)$ 0.601	A1	1.2	awfw 0.600~0.602
4(d)	Exponential (distribution)	B1	2.1b	
	with parameter 2.8	B1dep	2.1b	$\lambda = 2.8$ **or** **mean** = 1/2.8 = 0.357 *must* state that this is the mean Dep on previous B1
				Accept Exp(2.8) for both marks
4(e)	$\left(\frac{1}{2.8} =\right) = \frac{5}{14} = 0.357$ (years)	B1	1.2	oe
4(f)	Use of memoryless property	M1	2.1a	PI by correct working May be stated or demonstrated by candidate clearly disregarding the wind turbine history e.g. "P(X < 0.5)" with no conditional probability used.
	$1 - e^{-2.8\times0.5}$	M1ft	1.2	PI. $1 - e^{-1.4}$

				Accept ft of candidate's λ
	0.753	A1	1.2	awrt 0.753
	Alternative			
	Use of Poisson distribution with $\lambda = 1.4$	(M1)		PI
	$1 - P(X = 0)$	(M1)		PI. $1 - 0.247$
	0.753	(A1)		awrt 0.753
4(g)	$\frac{84}{6 \times 5} = 2.8$	B1	1.2	oe working Full calculation must be demonstrated. May state in words.
4(h)	A wind turbine may not fail at a constant average rate as older turbines may be more likely to fail.			Challenge to the assumption that failure rate for a wind turbine is constant.
	The failure of a wind turbine may not be a random event because it may be caused by weather.			Challenge to the assumption that wind turbines fail at random.
	The failure of a particular wind turbine may not be independent of the failure of another wind turbine as one may fall onto another.			Challenge to the assumption that wind turbines fail independently.
	Petra's calculation of λ made the assumption that all of the wind turbines fail at the same average rate per year. This might not be true.			Challenge to the assumption that **all wind turbines fail** at the **same rate**.
	Petra based her value of 2.8 on a limited amount of data			Challenging Petra's value of 2.8 e.g. When a wind turbine has failed it can't fail again until repaired
	More than one turbine could fail at exactly the same time			
	There would be an upper limit to the number of failures in one year			
		E1, E1, E1	3.1b, 3.1b, 3.1b	Any three **distinct** answers from the above, in context of wind turbine failure. Max E1E0E0 if no context
	Total	**14**		

Q4

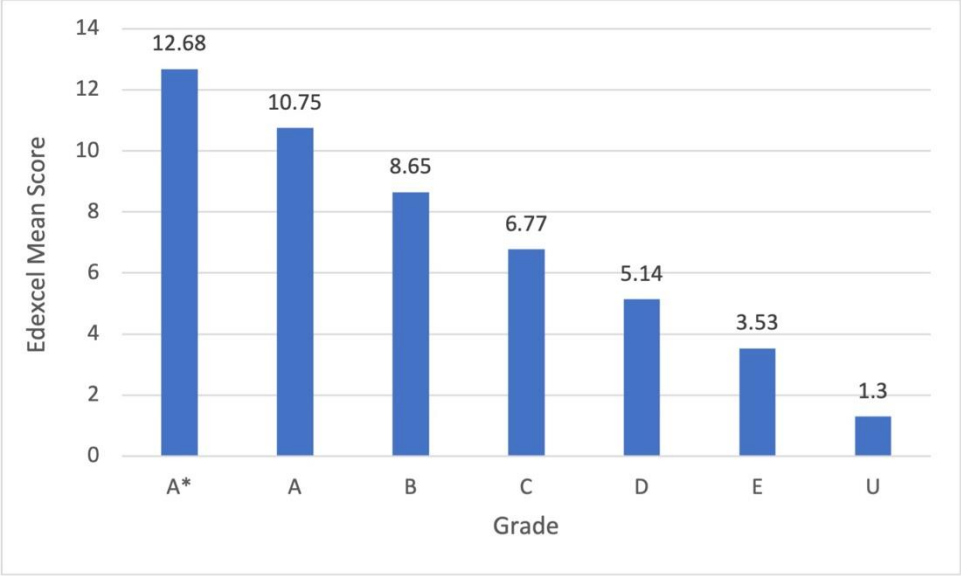
A B C

Question 4 - Examiner Comments

Candidates approached this question well, with many able to do well on the calculations. Some candidates were incorrect in stating the parameter of the exponential distribution, using $\frac{1}{2.8}$, the mean, instead of 2.8. In part (h) candidates were able to identify the assumptions required for a Poisson distribution but did not always tie these to contextual reasons.

Question 4 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade:							
			ALL	A*	A	B	C	D	E	U
7.17	14	51	7.17	12.68	10.75	8.65	6.77	5.14	3.53	1.30



Question 4 - Response A

4 Petra is an engineer who works on an offshore wind farm.

Petra organises the repairs of wind turbines that fail. The wind farm contract states that wind turbines must be repaired on the same day that they fail.

Failures appear to happen at random, and at a fairly consistent rate over the year.

In order to organise repairs, Petra uses a Poisson model with $\lambda = 2.8$ for the number of failures per year for each wind turbine.

[Data source: https://doi.org/10.1177%2F0957650915597560]

Use Petra's model to answer (a) to (f).

(a) Calculate the standard deviation of the number of failures per year for a wind turbine at the wind farm.

(1)

2.8

(b) Find the probability that a wind turbine at the wind farm fails exactly three times in a given year.

(1)

P(x = 3) = P = 0.222

(c) Find the probability that a wind turbine at the wind farm fails more than seven times over a three-year period.

(2)

2.8 × 3 = 8.4

Po(8.4)

P(x > 7)

P = 0.601348123

P = 0.601

(d) State fully the distribution of the time, in years, between successive failures of a wind turbine at the wind farm.

(2)

1 ÷ 2.8 = 0.371428571

Po(0.371)

(e) Calculate the expected time, in years, between successive failures of a wind turbine at the wind farm.

(1)

1 ÷ 2.8 = 0.371428571

0.357 years

Q4

A B C

(f) Given that a particular wind turbine at the wind farm has not failed for two consecutive years, calculate the probability that it fails within the next six months.

(3)

$2.8 \div 2 = 1.4$

$Po(1.4)$

$P(x=1)$

$P = 0.34523575$

$P = 0.345$

$Po(5.6) = 3.6979 \times 10^{-3}$

$P(x=0)$

$3.6979 \times 10^{-3} \times 0.345 =$

0.00128

The wind farm has six wind turbines. During the past five years, there have been 84 wind turbine failures at the wind farm.

(g) Show how Petra used this information to calculate the value of λ

(1)

$6 \times 5 = 30$

$84 \div 30 = 2.8$

(h) Make **three** criticisms, in context, of Petra's use of a Poisson model with $\lambda = 2.8$ for the number of failures per year for each wind turbine.

You may consider both the use of a Poisson model and the value of λ

(3)

he is using a small sample so the value for λ may not be very accurate

2.8 wind turbines can't fail every year (not a whole integer)

the poisson model isn't very accurate

Q4

A B C

6 /14

Part (a)
B0: Mean/variance stated not standard deviation for Poisson distribution.

Part (b)
B1: Correct probability calculated.

Part (c)
M1: Candidate has calculated $\lambda = 8.4$
A1: Correct probability calculated.

Part (d)
B0: Incorrectly identified the distribution, this was a common mistake where candidates did not recognise that the question is now working with time, a continuous variable.
B0: Used the mean for an exponential (although incorrectly calculated) not the correct parameter.

Part (e)
B1: Correct value for the mean of the exponential found units were not required.

Part (f)
Candidate has used Alternative method
M1: Candidate has used the Poisson distribution with $\lambda = 1.4$
M0: Although $\mathrm{P}(X = 0)$ has been calculated it has not been subtracted from 1
A0: No obvious attempt to find a final answer.

Part (g)
B0: A common approach to show where the value of 2.8 was derived from.

Part (h)
E1: Criticism of the use of 2.8, although a little vague, was given credit.
E0: The second comment about 2.8 not being an integer was a common misconception shown by candidates.
E0: No third criticism given.

Question 4 - Response B

4 Petra is an engineer who works on an offshore wind farm.

Petra organises the repairs of wind turbines that fail. The wind farm contract states that wind turbines must be repaired on the same day that they fail.

Failures appear to happen at random, and at a fairly consistent rate over the year.

In order to organise repairs, Petra uses a Poisson model with $\lambda = 2.8$ for the number of failures per year for each wind turbine.

[Data source: https://doi.org/10.1177%2F0957650915597560]

Use Petra's model to answer (a) to (f).

(a) Calculate the standard deviation of the number of failures per year for a wind turbine at the wind farm.

(1)

Var $= \frac{1}{\lambda^2} = \frac{1}{2.8^2} = 0.128$

standard deviation $= 0.357$

(b) Find the probability that a wind turbine at the wind farm fails exactly three times in a given year.

(1)

$P(X=3)$ PD: x: 3, λ: 2.8

$= 0.22$

(c) Find the probability that a wind turbine at the wind farm fails more than seven times over a three-year period.

(2)

$\lambda = 2.8 \times 3 = 8.4$ $X \sim Po(8.4)$

$P(X>7) = 1 - P(X<7)$

Poisson cd: x: 7, λ: 8.4

$= 1 - 0.399$

$= 0.601$

(d) State fully the distribution of the time, in years, between successive failures of a wind turbine at the wind farm.

(2)

$X \sim Po(2.8)$

(e) Calculate the expected time, in years, between successive failures of a wind turbine at the wind farm.

(1)

$E(x) = \lambda = \frac{1}{2.8} = \frac{5}{14} = 0.357$

Q4

A B C

(f) Given that a particular wind turbine at the wind farm has not failed for two consecutive years, calculate the probability that it fails within the next six months.

(3)

~~not failed for two years~~)

P(fails within next six months)

P(fails next 6 month | not failed) = P(fails ∩ not failed) / P(not failed)

The wind farm has six wind turbines. During the past five years, there have been 84 wind turbine failures at the wind farm.

(g) Show how Petra used this information to calculate the value of λ

(1)

84 ÷ 2.8 = 30

6 × 5 = 30

(h) Make **three** criticisms, in context, of Petra's use of a Poisson model with $\lambda = 2.8$ for the number of failures per year for each wind turbine.

You may consider both the use of a Poisson model and the value of λ

(3)

- We are unsure of the area ~~of~~ of the wind farm.

- This figure of 2.8, may be as estimate, therefore may not be fully representative.

- Although, they are fairly at a constant rate over the year, this may have some error.

6 /14

Q4

A
B
C

Part (a)
B0: Candidate has calculated the standard deviation for the exponential distribution instead.

Part (b)
B1: Correct probability calculated, 0.22 was accepted.

Part (c)
M1: Candidate has clearly restated the Poisson distribution for $\lambda = 8.4$
A1: Correct probability calculated.

Part (d)
B0: Incorrectly identified the distribution, this was a common mistake where candidates did not recognise that the question is now working with time, a continuous variable.
B1: 2.8 clearly stated as the parameter.

Part (e)
B1: Correct value for the mean of the exponential found.

Part (f)
M0: Candidate has not recognised that the question requires use of the memoryless property or of Poisson.
M0: No attempt made by candidate.
A0: No attempt made by candidate.

Part (g)
B0: Candidate has not shown how the value of 2.8 was obtained, they have shown two calculations that both make 30.

Part (h)
E1: Criticism of the use of 2.8, although a little vague, was given credit.
E0: The first comment about area of wind farm is not directly linking to any Poisson conditions.
E0: Final comment has not got enough detail to be given credit, an example in context would have helped here.

Question 4 - Response C

4 Petra is an engineer who works on an offshore wind farm.

Petra organises the repairs of wind turbines that fail. The wind farm contract states that wind turbines must be repaired on the same day that they fail.

Failures appear to happen at random, and at a fairly consistent rate over the year.

In order to organise repairs, Petra uses a Poisson model with $\lambda = 2.8$ for the number of failures per year for each wind turbine.

[Data source: https://doi.org/10.1177%2F0957650915597560]

Use Petra's model to answer (a) to (f).

(a) Calculate the standard deviation of the number of failures per year for a wind turbine at the wind farm.

(1)

$\sqrt{2.8} = 0.1276 = 0.3574$

variance = 2.8 standard deviation = $\sqrt{2.8} = 1.67$

(b) Find the probability that a wind turbine at the wind farm fails exactly three times in a given year.

(1)

$X \sim Po(2.8)$

$P(x = 3) = 0.2225$

(c) Find the probability that a wind turbine at the wind farm fails more than seven times over a three-year period.

(2)

$X_1 \sim Po(2.8) \rightarrow$ one year

$X_3 \sim Po(8.4) \rightarrow$ three years

$P(x \geq 7) = 1 - P(x \leq 6)$

$= 1 - 0.26699$

$= 0.733$

(6) (7 8)

Question 4 continued

→ continuous. ⇒ Exp.

(d) State fully the distribution of the time, in years, between successive failures of a wind turbine at the wind farm.

(2)

$X \sim Exp\left(\frac{1}{2.8}\right)$ → per year.

~~$X \sim Exp($~~

(e) Calculate the expected time, in years, between successive failures of a wind turbine at the wind farm.

(1)

$\frac{1}{2.8}$

Q4

A B C

(f) Given that a particular wind turbine at the wind farm has not failed for two consecutive years, calculate the probability that it fails within the next six months.

(3)

Since exponential has no memory....

$X \sim Exp\left(\frac{1}{2.8}\right)$ → per year.

$P(a \leq x \leq b) = e^{-\lambda a} - e^{-\lambda b}$

$P\left(x < \frac{6}{12}\right) = e^{-\frac{1}{2.8} \times 0} - e^{-\frac{1}{2.8} \times \frac{6}{12}}$

$= 0.164$

The wind farm has six wind turbines. During the past five years, there have been 84 wind turbine failures at the wind farm.

(g) Show how Petra used this information to calculate the value of λ (2.8)

(1)

$6 \times 5 = 30$ $\quad 84 \div 30 = 2.8$

(h) Make **three** criticisms, in context, of Petra's use of a Poisson model with $\lambda = 2.8$ for the number of failures per year for each wind turbine.

You may consider both the use of a Poisson model and the value of λ

(3)

Random events.
↳ The events of the number of failures per year for each wind turbine is random.

Independant events.
↳ One wind turbine failing is independant of another wind turbine failing.

Constant average rate.
↳ the number of failures per year for each wind turbine has a constant average rate of 2.8

Q4

A
B
C

9 /14

Part (a)
B1: 1.67 clearly stated.

Part (b)
B1: Correct probability calculated.

Part (c)
M1: Candidate has clearly restated the Poisson distribution for $\lambda = 8.4$
A0: Incorrect probability as candidate has interpreted 'more than seven' as $\mathrm{P}(X \geq 7)$

Part (d)
B1: Correctly identified exponential.
B0: A common misconception was using the mean of the exponential as the parameter, as shown here.

Part (e)
B1: $\frac{1}{2.8}$ was given credit but would be more suitable as a fraction without decimal or as a rounded decimal.

(f)
M1: Candidate has recognised that the question requires use of the memoryless property.
M1ft: Although candidate has the incorrect exponential distribution from (d) they have calculated the probability correctly.
A0: Did not obtain correct probability due to incorrect parameter.

Part (g)
B1: A common approach to show where the value of 2.8 was derived from.

Part (h)
E1: Conditions required for a Poisson model to be valid have been recognised by the candidate.
E0: However, the three comments do not give an example in context for these conditions. Therefore, no further marks can be given.
E0: However, the three comments do not give an example in context for these conditions. Therefore, no further marks can be given.

Question 5

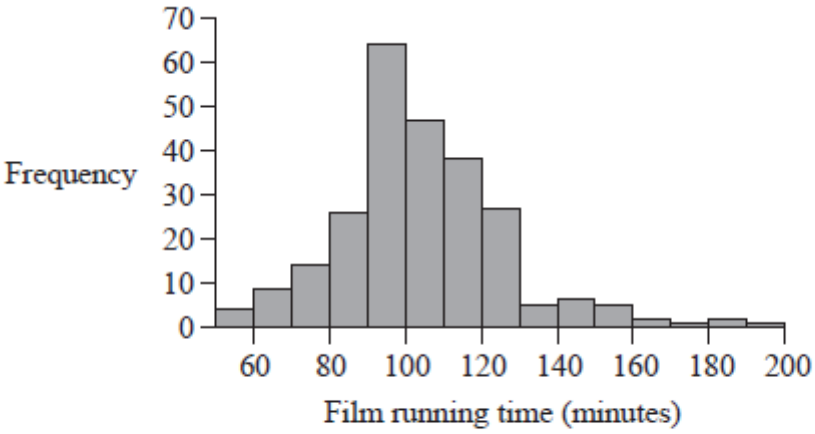
Question | Mark Scheme | Examiner Comments | Performance | Response A | Response B | Response C

Question 5 - Question

5 Dennis and Viola are doing a course in film studies. They want to conduct a study into the factors affecting film running times. They start by investigating the distribution of film running times.

Dennis and Viola extract a random sample of 250 films from a database containing the names and dates of all films ever made and look up the running time of each film in their sample on the internet.

They use graphing software to produce a histogram of their results, shown in **Figure 5**.



[Data sources: www.cinemetrics.lv, www.imdb.com]

Figure 5

Viola believes that a normal distribution might be a suitable model for film running times.

(a) Explain **one** feature of **Figure 5** that supports Viola's belief.

(1)

(b) Explain **one** feature of **Figure 5** that does not support Viola's belief.

(1)

Dennis and Viola agree that they should perform a goodness-of-fit test to investigate whether a normal distribution would be a suitable model for film running times.

For their sample of 250 films, the mean running time is 105.26 minutes and the standard deviation of the running times is 22.18 minutes.

Part of their working is shown in **Figure 6**.

Film running time (minutes)	Observed frequency	Probability (4 d.p.)	Expected frequency (2 d.p.)
$x \leq 60$	4	0.0206	5.16
$60 < x \leq 80$	22	0.1067	26.68
$80 < x \leq 100$	90	s	u
$100 < x \leq 120$	85	t	v
$120 < x \leq 140$	32	0.1945	48.63
$140 < x \leq 160$	11	0.0518	12.96
$x > 160$	6	0.0068	1.70
Total	250	1	250

Figure 6

(c) Find the values of s, t, u and v missing from **Figure 6**.

(4)

(d) Explain, with a reason, what adjustment Viola and Dennis must make to the data in **Figure 6** before they can carry out the χ^2 goodness-of-fit test.

(1)

Viola and Dennis make the necessary adjustment to their data.

They calculate the χ^2 test statistic to be 13.04 (to 2 d.p.)

(e) Use this information to complete their χ^2 goodness-of-fit test.

You must state appropriate hypotheses.

(4)

(Total for Question 5 is 11 marks)

Question 5 - Mark Scheme

Question	Scheme	Marks	AO	Notes
5(a)	(The distribution has an approximate) bell shape	E1	2.1a	oe Accept "(Distribution is) monomodal" **or** "unimodal" **or** "One clear peak (in the distribution)"
5(b)	The distribution has a (positive) skew.			Skew **or** not symmetrical
	Too much of the distribution is **in** the **tails** (for it to be a normal distribution).			oe Accept "(distribution/shape) **too triangular**" or (distribution has) **tails** (that are) **too large**" or "(distribution has) **high kurtosis**"
		E1	2.1a	**Either** May use calculations that show an equivalent argument but the point being made must be clear.
5(c)	$s = P(80 < X < 100)$ or $t = P(100 < X < 120)$	M1	1.2	PI oe Clear attempt to find either probability using correct normal distribution.
	$s = 0.2789$ or $t = 0.3406$	A1	1.2	**Either** s or t correct s awfw 0.278~0.280 t awfw 0.340~0.342
	$250 \times s$ or $250 \times t$	M1ft	1.2	Either PI ft candidate's s or t
	$u = 69.73$ and $v = 85.15$ or 85.14	A1	1.2	**Both** u and v correct u awfw 69.5~70 v awfw 85~85.5 Correct to 1 or more dp

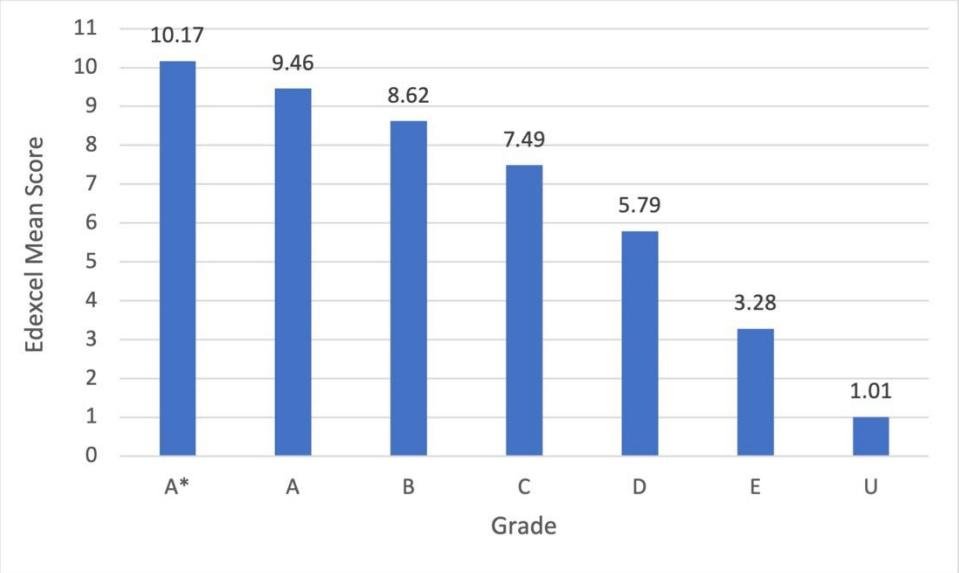
5(d)	They should pool $140 < x \leq 160$ with $160 < x$ because the expected frequency is less than 5.	E1	1.3	oe Accept "pooling because 1.7 < 5"
5(e)	H_0: The normal distribution is a suitable model H_1: The normal distribution is not a suitable model	B1	1.3	oe Hypotheses, both
	df = 3 cv = 7.815 or *p*-value = 0.00455	B1	1.3	cv or awfw *p*-value = 0.004~0.005
	(13.04 > 7.815 or 0.00455 < 0.05) Reject H_0	M1dep	2.1b	Comparison dep B1
	There is sufficient evidence to conclude that the normal distribution is **not** a suitable model for film running times.	E1dep	2.1a	Correct conclusion in context dep B1M1
	Total	**11**		

Question 5 - Examiner Comments

This question was answered well, although, in part (d), some candidates missed that two degrees of freedom were lost due to estimating the mean and standard deviation, and hence they were unable to complete the hypothesis test and gain full marks.

Question 5 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade:							
			ALL	A*	A	B	C	D	E	U
7.17	11	65	7.17	10.17	9.46	8.62	7.49	5.79	3.28	1.01

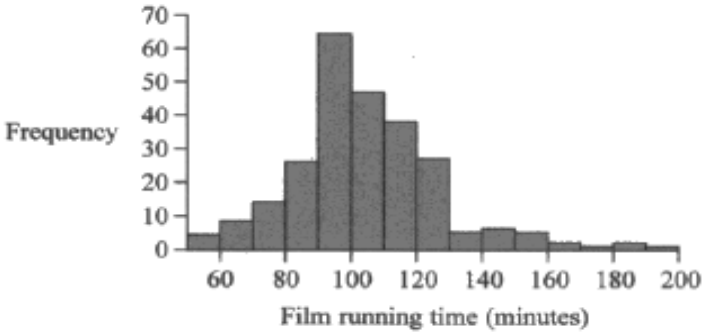


Question 5 - Response A

5 Dennis and Viola are doing a course in film studies. They want to conduct a study into the factors affecting film running times. They start by investigating the distribution of film running times.

Dennis and Viola extract a random sample of 250 films from a database containing the names and dates of all films ever made and look up the running time of each film in their sample on the internet.

They use graphing software to produce a histogram of their results, shown in **Figure 5**.



[Data sources: www.cinemetrics.lv, www.imdb.com]

Figure 5

Viola believes that a normal distribution might be a suitable model for film running times.

(a) Explain **one** feature of **Figure 5** that supports Viola's belief.

(1)

The mean, median and mode are
in the suitable bounds
The data is continuous.

(b) Explain **one** feature of **Figure 5** that does not support Viola's belief.

(1)

The data is positively skewed.

(c) Find the values of s, t, u and v missing from **Figure 6**.

(4)

$s = 0.2789$
$t = 0.3406$
$u = 69.725$
$v = 85.15$

(d) Explain, with a reason, what adjustment Viola and Dennis must make to the data in **Figure 6** before they can carry out the χ^2 goodness-of-fit test.

(1)

They must give the data
an upper and lower limit.

(e) Use this information to complete their χ^2 goodness-of-fit test.

You must state appropriate hypotheses.

(4)

H_0: The normal distribution would not be a suitable model for film running times.

H_1: The normal distribution would be a suitable model for film running times.

df = 6

Ts: 13.04

Crit: 13.04 < 14.449

Not significant

Accept H_0

There is insufficient evidence to suggest the normal distribution would be a suitable model for film running times at the 5% significance level.

5 /11

Q5

Part (a)

E0: Although it is correct that the data is continuous that does not support the use of a normal distribution, for instance it could be exponential.

Part (b)

E1: The candidate has recognised that the distribution is skewed, type of skew was not required for credit.

Part (c)

M1: Attempt to find *s* or *t*.

A1: *s* or *t* correct, only one required to be correct for this mark.

M1: Attempt to find *u* or *v*.

A1: Both *u* and *v* correct.

Part (d)

E0: A common misconception that the film times must have a lower or upper limit

Part (e)

B0: Hypotheses are reversed.

B0: A common error where candidates had forgotten the information given in part (c) stating 'For their 250 films' therefore a degree of freedom should have been taken off for the mean and standard deviation as well.

M0: Although the candidate has correctly concluded for their critical value, only certain values for the critical value were accepted 7.815, 9.488 or 11.07

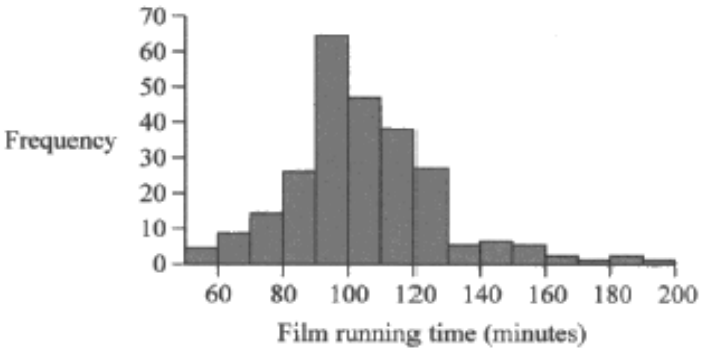
E0: Although conclusion was correct and in context, this mark was dependent on the correct critical value and comparison.

Question 5 - Response B

5 Dennis and Viola are doing a course in film studies. They want to conduct a study into the factors affecting film running times. They start by investigating the distribution of film running times.

Dennis and Viola extract a random sample of 250 films from a database containing the names and dates of all films ever made and look up the running time of each film in their sample on the internet.

They use graphing software to produce a histogram of their results, shown in **Figure 5**.



[Data sources: www.cinemetrics.lv, www.imdb.com]

Figure 5

Viola believes that a normal distribution might be a suitable model for film running times.

(a) Explain **one** feature of **Figure 5** that supports Viola's belief.

(1)

The data size is large enough to support a normal distribution.

(b) Explain **one** feature of **Figure 5** that does not support Viola's belief.

(1)

The films times don't range from 0 to 200, so a mean of roughly 100 isn't halfway, so the histogram is skewed to the left.

Q5

A B C

(c) Find the values of s, t, u and v missing from **Figure 6**.

(4)

$S = 0.2789$

$t = 0.3406$

$U = 69.72$

$V = 85.14$

(d) Explain, with a reason, what adjustment Viola and Dennis must make to the data in **Figure 6** before they can carry out the χ^2 goodness-of-fit test.

(1)

They have to combine the last 2 groups

$140 < x \leq 160$
$x > 160$ → $x > 140$

(e) Use this information to complete their χ^2 goodness-of-fit test.

You must state appropriate hypotheses.

(4)

H_0: Ratio 4:22:90:85:32:17
H_1: No ratio

$TS = 13.04$

$df = n - 1 - x$
$= 6 - 1 - 0$
$= 5$

$CV = 11.070$

$TS > CV$ [Reject H_0]

There is sufficient evidence at the 5% significance level to suggest that there is no ratio, so a normal distribution wouldn't be suitable.

Q5

A
B
C

7 /11

Part (a)

E0: Misconception that a large sample implies a population is normally distributed. This could possibly be a confusion with central limit theorem.

Part (b)

E1: The candidate has recognised that the distribution is skewed.

Part (c)

M1: Attempt to find *s* or *t*.

A1: *s* or *t* correct, only one required to be correct for this mark.

M1: Attempt to find *u* or *v*.

A1: Both *u* and *v* correct.

Part (d)

E1: Although candidate has not explicitly said the reason why, they have identified the correct categories to pool implying they have recognised the expected value less than 5.

Part (e)

B0: Hypotheses are incorrect.

B0: A common error where candidates had forgotten the information given in part (c) stating 'For their 250 films' therefore a degree of freedom should have been taken off for the mean and standard deviation as well.

M1: However, the candidate has correctly concluded to reject H_0 for their critical value.

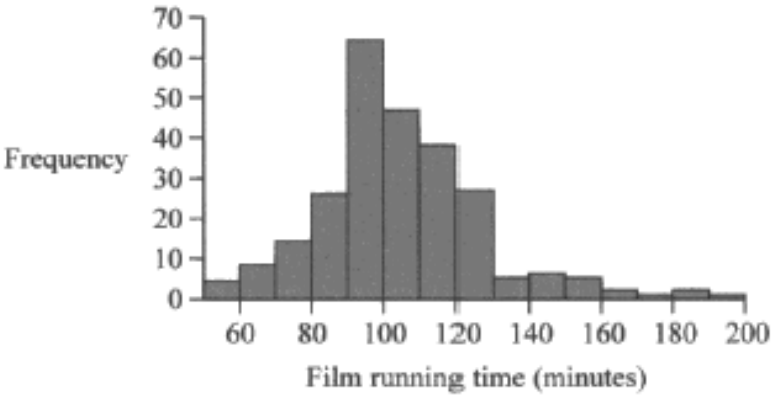
E0: Although conclusion was correct and in context, this mark was dependent on the correct critical value and comparison.

Question 5 - Response C

5 Dennis and Viola are doing a course in film studies. They want to conduct a study into the factors affecting film running times. They start by investigating the distribution of film running times.

Dennis and Viola extract a random sample of 250 films from a database containing the names and dates of all films ever made and look up the running time of each film in their sample on the internet.

They use graphing software to produce a histogram of their results, shown in **Figure 5**.



[Data sources: www.cinemetrics.lv, www.imdb.com]

Figure 5

Viola believes that a normal distribution might be a suitable model for film running times.

(a) Explain **one** feature of **Figure 5** that supports Viola's belief.

(1)

The data is bell shaped

(b) Explain **one** feature of **Figure 5** that does not support Viola's belief.

(1)

Data is positively skewed & not symmetrical

Q5

A B C

(c) Find the values of s, t, u and v missing from **Figure 6**.

(4)

$X \sim N(105.26, 27.18^2)$

$s = 0.2789$ $\quad t = 0.3406$

$u = 0.2789 \times 250 = 69.73$

$v = 0.3406 \times 250 = 85.15$

(d) Explain, with a reason, what adjustment Viola and Dennis must make to the data in **Figure 6** before they can carry out the χ^2 goodness-of-fit test.

(1)

Combine cell $140 < x \leq 160$ and $x > 160$ to $x > 140$

Since expected value for $x > 160$ is less than 5

(e) Use this information to complete their χ^2 goodness-of-fit test.

You must state appropriate hypotheses.

(4)

H_0: Film running time can be modelled by the normal distribution $X \sim N(105.26, 27.18^2)$ (or a random variable)

H_1: Film running time can not be modelled by the normal distribution.

$\chi^2 = 13.04$

$DF = 6 - 2 - 1 = 3$

C.V = 7.815

$13.04 > 7.815$ reject null

evidence to at 5% significance level to suggest film running time can not be modelled by a normal distribution.

11 /11

Q5

A B C

Part (a)

E1: Candidate has recognised that the distribution is roughly bell shaped.

Part (b)

E1: The candidate has recognised that the distribution is skewed, type of skew was not required for credit.

Part (c)
M1: Attempt to find *s* or *t*.
A1: *s* or *t* correct, only one required to be correct for this mark.
M1: Attempt to find *u* or *v*.
A1: Both *u* and *v* correct.

Part (d)
E1: An excellent response clearly identifying the correct cells with a specific reason why.

Part (e)
B1: Hypotheses are correct, a specific normal distribution being stated was not required.
B1: Candidate has correctly considered the number of rows after pooling as well as the parameter estimates from the sample to obtain the critical value 0f 7.815
M1: Correct comparison of test statistic with critical value.
E1: Conclusion is correct and in context with the correct critical value used to conclude.

Question 6

Question | Mark Scheme | Examiner Comments

Performance | Response A | Response B | Response C

Question 6 - Question

6. Russell and Robert are interested in sport. They believe that wearing red may affect the success of a person or team in a sporting competition.

Russell finds the following information about the 2004 Olympic games.

- Male combatants in combat sports were randomly assigned either a red or a blue outfit.
- There was a total of 21 rounds of fighting for these combat sports.
- One round had an equal number of red and blue winners.
- Four rounds had **more blue** than red winners
- Sixteen rounds had **more red** than blue winners.

[Source: doi:10.1038/435293a]

(a) Making any necessary assumptions, use Russell's data to investigate their belief.

Make a full conclusion, in context, about any effect seen, or otherwise.

(7)

Robert collected data about the performance of five teams that sometimes wore a red kit at the Euro 2004 men's football competition.

He awarded one point for a win, zero for a draw and minus one point for a loss (disregarding penalty shoot-outs). He then found the average points scored when each team wore red and when they did not wear red.

Opposing football teams never wear the same colour kit as each other.

Robert's results are shown in **Figure 7**.

	Total points		Number of games		Average points	
Team	**Wearing red**	**Not wearing red**	**Wearing red**	**Not wearing red**	**Wearing red**	**Not wearing red**
Croatia	0	–1	1	2	0	–0.5
Czechia	1	2	1	4	1	0.5
England	1	0	1	3	1	0
Latvia	–1	–1	2	1	–0.5	–1
Spain	1	–1	2	1	0.5	–1

[Data source: doi.org/10.1038/435293a]

Figure 7

Q6

(b) Making any necessary assumptions, use the **average points** data in **Figure 7** and a **paired *t*-test** to investigate Russell and Robert's belief that wearing red may affect the success of a team.

Make a full conclusion, in context, about any effect seen, or otherwise.

(8)

(c) Considering the data and tests used in (*a*) and (*b*), make **four** distinct criticisms.

In your answer you might consider

- the size of sample provided,
- the sampling design,
- the test used and its assumptions.

(4)

Charlotte reads Russell and Robert's research and tells her women's netball team that they should change to wearing red to improve their results.

(d) Make **two** comments about Charlotte's statement in light of the data given and the conclusions to the tests in (a) and (b).

(2)

(Total for Question 6 is 21 marks)

Question 6 - Mark Scheme

Question	Scheme	Marks	AO	Notes
6(a)	(Sign test)	B1	2.1b	PI sign test clearly used or stated Implied by use of binomial e.g. P($X \leq 4$)
	[X = number of rounds with more blue winners than red winners] $X \sim B(20,0.5)$	B1	2.1a	PI use of B(20,0.5) Condone n=21
	H_0: p = 0.5 H_1: p ≠ 0.5	B1	1.3	oe Condone 1-tail
	P($X \leq 4$)	M1	1.3	oe PI Attempt to calculate p-value Accept P($X \geq 16$) Or Attempt to find critical region
	= 0.00591 or Critical region is $X \leq 5$ as p=0.021	A1	1.3	awrt 0.0059
	0.00591 < 0.025 or 4<5) Reject H_0	M1dep	2.1b	Comparison Condone compared to 0.05 if 1-tailed hypotheses Dep M1
	There is significant evidence that wearing red affects the success of (male) combatants.	E1dep	2.1a	Dep A1M1 Condone "There appears to be an advantage of wearing red (over blue) for the combatants."

Question	Scheme	Marks	AO	Notes
6(b)	H_0: $\mu_d = 0$ H_1: $\mu_d \neq 0$	B1	1.3	oe Hypotheses re: population means e.g. $\mu_{red} = \mu_{not\ red}$ Condone 1-tail
	d = 0.5, 0.5, 1, 0.5, 1.5	M1	1.3	PI Attempt at differences (signs may be all negative)
	$\bar{d} = 0.8 \quad s_d = 0.4472$	A1ft	1.3	PI mean and sd of their differences
	$ts = \dfrac{0.8}{\left(\dfrac{0.4472}{\sqrt{5}}\right)}$	M1	1.3	PI calculation of ts may be negative Allow their mean/sd or $0 \pm 2.776 \times \frac{0.4472}{\sqrt{5}}$ which scores the B1 for cv
	t = 4.00	A1	1.3	awrt 4.00 or $\frac{\bar{d}}{\frac{0.4472}{\sqrt{5}}} = 2.776$
	(df = 4) cv = 2.776	B1	1.3	Either correct cv (ignore sign) or p-value or p-value = 0.0161 awfw 0.0159~0.0162 or $\bar{d} = \pm 0.555$ **Condone** 1-tail cv = 2.132 if 1-tailed hypotheses
	4 > 2.776 Reject H_0.	M1dep	2.1b	PI or -4 < -2.776 or 0.0161 < 0.025 or 0.8>0.555 dep A1B1 Correct comparison
	There is **evidence** at the 5% significance level that the **success** of a football team is **affected** by wearing	E1dep	2.1a	oe full explanation in context required for E1 mark dep A1B1

Q6

A B C

	red; (teams appear to do **better** when **wearing red**.)			
SC Two sample t test max B1M0A0M1A0B1M1E0 $\bar{d}_{red} = 0.4, \bar{d}_{not\ red} = -0.4$ $s_p = 0.65$ t=1.94				

Question	Scheme	Marks	AO	Notes
6(c)	**Possible criticisms (not exhaustive)**			
	Both (a) and (b) have very small samples			oe Or either sample small
	There is no evidence of randomisation in (b).			oe Or neither sample was selected at random. "Football teams weren't randomly assigned to wear red or not red"
	Fights in (a) not independent as same fighter may fight in red and in blue			
	Additional factors not taken into account e.g. relative strength of teams, ability of combatant			Blocking factors could be mentioned here
	(a) only considers 2004			
	(b) only considers Europe			
	All combat rounds not equally weighted			
	(a) only tests red against blue…			
	…whereas (b) tests against many different colours, so the two tests are not providing consistent conclusions.			
	The t-test [in (b)] may not be appropriate because the (differences) may not be normally distributed.			oe
		E1, E1, E1, E1	3.1a, 3.1a, 3.1a, 3.1a,	Any four **distinct** correct comments.

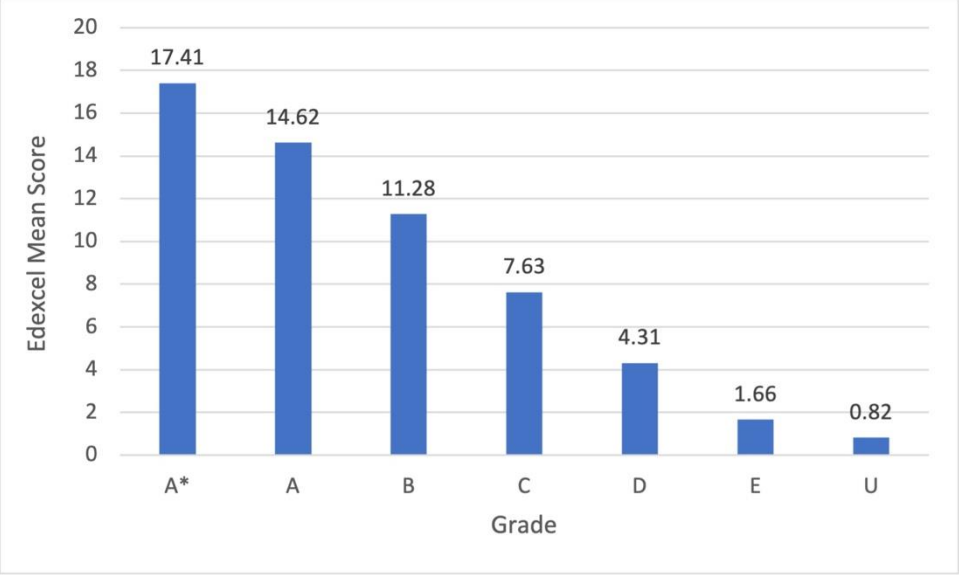
Question	Scheme	Marks	AO	Notes
6(d)	The data in (a) and (b) is **only about men's** sport (so the results may not be applicable to women's sport).			oe Accept "all data only about men"
	The tests were about **football and combat** sports so results might **not** extend to **netball**			
	In both (a) and (b) it was found that the players **wearing red did better** (so it may be helpful to Charlottes team also).			oe Accept "tests show playing in red helps"
	Both of the studies in (a) and (b) were **very small** (so it might not be worth paying for red kit without further evidence).			oe reference to limited size of evidence Accept "sample sizes are too small"
	Neither study was conclusive that the effect found was due to wearing red. **Effects could have been due to colours other than red** (so it might not be worth paying for new kit without further evidence).			oe reference to interpretation of evidence Accept "no evidence found of causation/causal effect of red on success"
	The **assumptions** of the tests used in this research **may not be true**(so Charlotte shouldn't spend money on new kit without better evidence).			oe
	The tests in (a) and (b) **were two tailed**, so we only have evidence that red is a difference not an improvement			
		E1, E1	3.1b, 3.1b	Any two distinct comments
SC if concluded that there was no difference in (a) or (b) may earn one mark for a relevant comment				
	Total	**21**		

Question 6 - Examiner Comments

This question was quite tricky, due to the unfamiliar way the data was presented in the table. Candidates who realised they were required to complete a sign test generally did very well in part (a), and, likewise, those able to identify that a test of differences was required in part (b) answered it well and many gained full marks. Candidates were able to comment reasonably well in part (c), and part (d) was the best answered part of this question in general.

Question 6 - Performance

Mean score	Max score	Mean %	Edexcel averages: mean scored by candidates achieving grade:							
			ALL	A*	A	B	C	D	E	U
8.33	21	40	8.33	17.41	14.62	11.28	7.63	4.31	1.66	0.82



Question 6 - Response A

6 Russell and Robert are interested in sport. They believe that wearing red may affect the success of a person or team in a sporting competition.

Russell finds the following information about the 2004 Olympic games.

- Male combatants in combat sports were randomly assigned either a red or a blue outfit.
- There was a total of 21 rounds of fighting for these combat sports.
- One round had an equal number of red and blue winners.
- Four rounds had **more blue** than red winners.
- Sixteen rounds had **more red** than blue winners.

prob = red team

X ~ B(21, 0·5)

X ~ B(4, 0·25)

X ~ B(16, 0·75)

n = 21

[Source: doi:10.1038/435293a]

(a) Making any necessary assumptions, use Russell's data to investigate their belief.

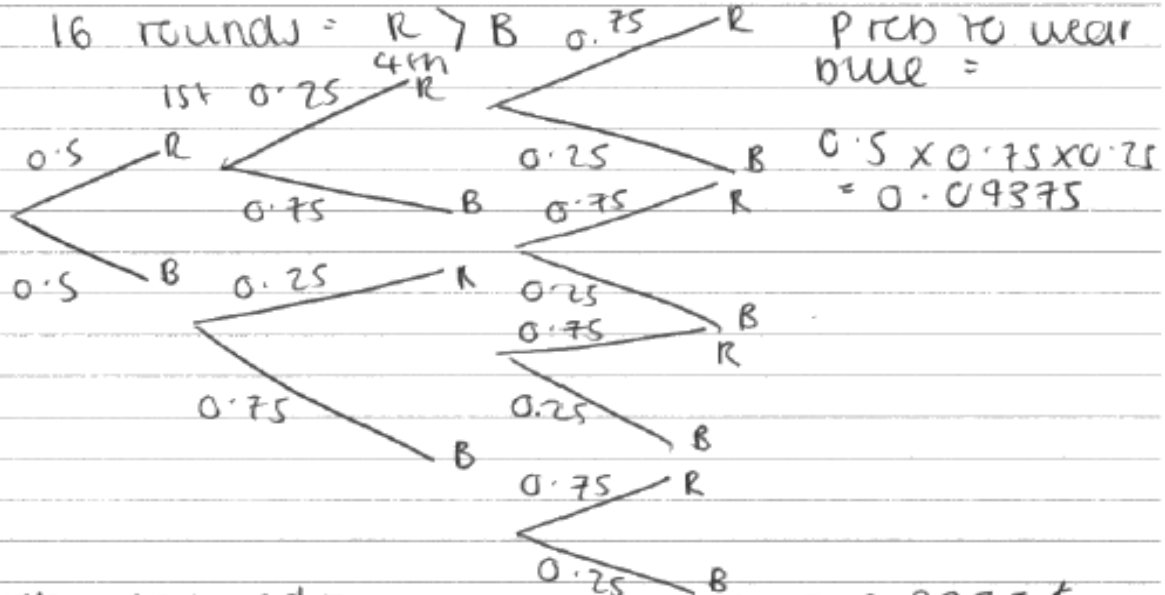
Make a full conclusion, in context, about any effect seen, or otherwise.

(7)

1 round = 0·5 21 × 0·5 = 10·5

4 rounds = B 7 R 16 th

16 rounds = R > B



Prob to wear blue = 0·5 × 0·75 × 0·25 = 0·09375

Prob wear red = 0·5 × 0·25 × 0·75 = 0·09375

$\frac{1}{21} \times \frac{4}{21} \times \frac{16}{21} = 0.006917$

Q6

A B C

• Those who wear red have greate endurance & they meane it past the first few rounds as they demonan in the last few rounds as theres more red in the 16th round so shows if you wear red and have endurence more likely have a better success.

So more who are in blue seem to do better with in the first few rounds as theres more blue in 4 rounds so wearing red may have effect on success in later rounds rather than in the begging as started off with equal amounts of red & blue.

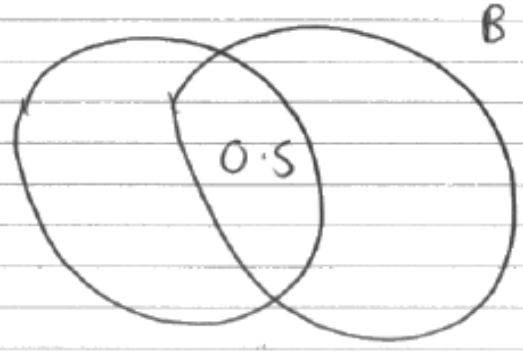
The probably of wearing red or blue & getting to the 16 rounds is equal as 0.09375 so may be no effect of colour wom & sucess or team.

Binomial dist

* $P(X=1) = 1.001358 \times 10^{-5}$ 1st

* $P(X=4) = 0.17574$

* $P(X=16) = 0.1492$



Robert collected data about the performance of five teams that sometimes wore a red kit at the Euro 2004 men's football competition.

He awarded one point for a win, zero for a draw and minus one point for a loss (disregarding penalty shoot-outs). He then found the average points scored when each team wore red and when they did not wear red.

Opposing football teams never wear the same colour kit as each other.

Robert's results are shown in **Figure 7**.

	Total points		Number of games		Average points	
Team	**Wearing red**	**Not wearing red**	**Wearing red**	**Not wearing red**	**Wearing red**	**Not wearing red**
Croatia	0	–1	1	2	0	–0.5
Czechia	1	2	1	4	1	0.5
England	1	0	1	3	1	0
Latvia	–1	–1	2	1	–0.5	–1
Spain	1	–1	2	1	0.5	–1

A B

[Data source: doi.org/10.1038/435293a]

Figure 7

(b) Making any necessary assumptions, use the **average points** data in **Figure 7** and a **paired t-test** to investigate Russell and Robert's belief that wearing red may affect the success of a team.

Make a full conclusion, in context, about any effect seen, or otherwise.

(8)

mean no - ndiff = 0
ni - ndiff > 0

diff A - B = diff
wear red = A
not wear red = B

diffs = -0.5, 0.5, 1, 0.5, 1.5

$\bar{x} = 0.6$
SD = 0.74162
n = 5

$TS = \frac{0.6 - (0)}{0.74162 / \sqrt{5}}$

$= 1.8091$

I rejected
5% sig

Q6

TS = 1.8091

CV = table 5 CV = 2.132 > TS = 1.8091

V = 5 - 1 = 4 not a significant result so accept Ho that there's no difference in average points so no effect of weather wearing red or not will increase the success of a team so Russel & Roberts belief aren't correct as wearing red doesn't effect success.

Therefore, there's no effect seen of wearing red & the average score so wearing red doesn't increase success.

But it was seen when Mexico & england were red they won their games had an average score of 1 so may have individual effects but then the effect of wearing red wasn't the same for the other countries.

when doing a paired t-test we make the assumption differences are normally distributed.

Q6

A
B
C

(c) Considering the data and tests used in (a) and (b), make **four** distinct criticisms.

In your answer you might consider

- the size of sample provided,
- the sampling design,
- the test used and its assumptions.

(4)

· sample size where small as only sampled 5 teams at Euro so may not be representative of the effects of wearing red to results played by these teams.

· they used a random design but randomly assigning either blue or red to home company but may be individual differences so instead should use randomised block design to make sure company tested for both red & blue outfits as used independent group design but would make findings more accurate if a was matched pairs or repeated measures.

· used a paired t-test which assumes differences are normally distributed which may not be true as can't apply central limit theory as more a small sample size

Charlotte reads Russell and Robert's research and tells her women's netball team that they should change to wearing red to improve their results.

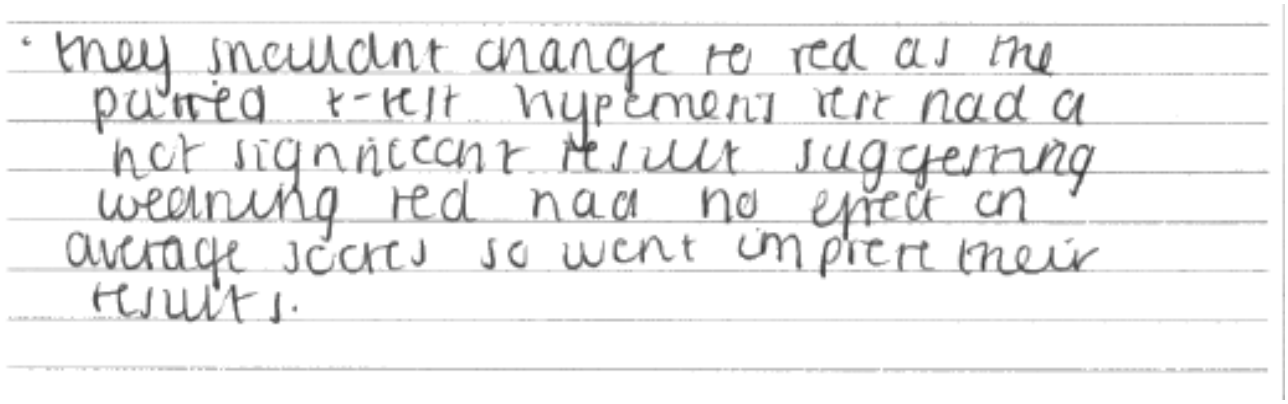
(d) Make **two** comments about Charlotte's statement in light of the data given and the conclusions to the tests in (a) and (b).

(2)

· russel & roberts data may not be representative as was done on males so may not be able to generalise findings to womans netball team.

Q6

A
B
C



11 /21

Part (a)

B0: This was a common method used by candidates who failed to recognise the use of the sign test in (a)

B1: Although the incorrect binomial distribution was used, credit was given for this mark with $n = 21$

B0: No hypotheses stated.

M0: Although a binomial distribution was stated it was not used.

A0: Therefore candidate did not obtain the correct *p*-value.

M0: No comparison made between *p*-value and significance level.

E0: No conclusion to hypothesis test attempted.

Part (b)

B1: Correct hypotheses (1 tail tests were given credit).

M1: There is an attempt at finding the differences.

A1ft: However, candidate has miscalculated the first difference leading to the incorrect mean and standard deviation. This is given credit as a follow through mark.

M1: The candidate has used their mean and standard deviation correctly to find 'their' test statistic.

A0: Due to the incorrect difference earlier, the candidate has not calculated the correct test statistic, no credit given.

B1: As the hypotheses are one tailed the candidate selected the correct critical value for their hypotheses.

M0: Although they have correctly compared, this mark was dependent on obtaining the correct test statistic.

E0: Although conclusion was correct and in context for their values, this mark was dependent on the correct critical value and comparison.

Part (c)

E1: First mark was awarded for recognising the use of a small sample.

E1: Second mark was given for additional factors not taken into account.

E1: Third mark awarded for stating that the differences may not be normally distributed.

E0: Further comments that were made did not gain credit.

Part (d)

E1: Candidate has recognised that the data for males may not be representative for females.

E1: The special case here as the conclusion the candidate has made agrees with their conclusion for (b)

Question 6 - Response B

6 Russell and Robert are interested in sport. They believe that wearing red may affect the success of a person or team in a sporting competition.

Russell finds the following information about the 2004 Olympic games.

- Male combatants in combat sports were randomly assigned either a red or a blue outfit
- There was a total of 21 rounds of fighting for these combat sports. $n = 21$
- One round had an equal number of red and blue winners. $R\backslash = 0$
- Four rounds had **more blue** than red winners $4R = b > r$
- Sixteen rounds had **more red** than blue winners. $16R = r > b$

[Source: doi:10.1038/435293a]

(a) Making any necessary assumptions, use Russell's data to investigate their belief.

Make a full conclusion, in context, about any effect seen, or otherwise.

(7)

$H_0: \eta_d = 0$
$H_1: \eta_d \neq 0$ $\quad X \sim B(20, 0.5)$

$T^+ = 16$
$T^- = 4$
$T = 4$

$P(x \leq 4) = 0.00591$

↓

~~0.0325~~

$0.00591 < 0.0325$

reject H_0

There is significant evidence to suggest that the population median wins of a person or term in a sporting competition was affected by whether they wore blue or red.

↳ This means that wearing red did mean a person or a team had a better chance of winning.

Q6

A B C

Robert collected data about the performance of five teams that sometimes wore a red kit at the Euro 2004 men's football competition.

He awarded one point for a win, zero for a draw and minus one point for a loss (disregarding penalty shoot-outs). He then found the average points scored when each team wore red and when they did not wear red.

Opposing football teams never wear the same colour kit as each other.

Robert's results are shown in **Figure 7**.

Team	Total points		Number of games		Average points		
	Wearing red	Not wearing red	Wearing red	Not wearing red	Wearing red	Not wearing red	
Croatia	0	−1	1	2	0	−0.5	= 0.5
Czechia	1	2	1	4	1	0.5	= 0.5
England	1	0	1	3	1	0	= 1
Latvia	−1	−1	2	1	−0.5	−1	= 0.5
Spain	1	−1	2	1	0.5	−1	= 1.5

[Data source: doi.org/10.1038/435293a]

Figure 7

(b) Making any necessary assumptions, use the **average points** data in **Figure 7** and a **paired t-test** to investigate Russell and Robert's belief that wearing red may affect the success of a team.

Make a full conclusion, in context, about any effect seen, or otherwise.

(8)

$H_0 : \mu_d = 0$

$H_1 : \mu_d \neq 0$

0.5 (1.5) 0.5 (1.5) 1 (4) 0.5 (1.5) 1.5 (5)

$\bar{x} = 2.7$

$sd = 1.681$

$n = 5$

$\nu = 4$

$t = 2.776$

$= \mu \pm t \times \frac{sd}{\sqrt{n}}$

$= 0 \pm 2.776 \times \frac{1.681}{\sqrt{5}}$

$= \pm 2.0869.$

Q6

A B C

-2.0869 2.0869

2.0869 < 2.7
reject H_0

There is significant evidence to suggest that the population mean differences that wearing red may affect the success of a team in the 2004 Euro men's football competition.

↳ This means that there is an effect that wearing red in the 2004 Euro's mens football competition.

(c) Considering the data and tests used in (a) and (b), make **four distinct** criticisms.

In your answer you might consider

- the size of sample provided,
- the sampling design,
- the test used and its assumptions.

(4)

- ~~In part (a) our sample size was 21 and in part (b) our sample size was 5 this means that~~

- In part (b) our sample size was 5 this means that our result may contain experimental error / a sample size of 5 is not enough to come to a valid / reliable conclusion therefore the result in part (b) about the 2004 Euro's mens football competition isn't reliable.

- In part (a) the sampling design doesn't take into account the size of the players this means that when a team was wearing red they might of had better players that the team wearing blue, this presents bias and means my result is unreliable. this is also applicable in part (b).

Q6

A
B
C

- In part (b) for a paired t test I have assumed that the population distribution for the differences follow a normal distribution when they may be incorrect as players might of gotten injured ~~which effects~~
- In part (b) the football teams may be in the same group meaning if England win Croatia ~~no~~ will lose therefore the assumption isn't valid.

Charlotte reads Russell and Robert's research and tells her women's netball team that they should change to wearing red to improve their results.

(d) Make **two** comments about Charlotte's statement in light of the data given and the conclusions to the tests in (a) and (b).

(2)

- In part (a) and (b) it was only about men's combatants in combant sport and men's 2004 Euro's football competition, therefore women's performance of wearing red may mean that they lose more netball games.
- Also netball is a different sport to football and combat sports therefore wearing red may mean Charlottes team loses more games.

Q6

A
B
C

15 /21

Part (a)

B1: Although not stated, the method of using a binomial and calculating $\mathrm{P}(X \leq 4)$ was enough to achieve this mark.

B1: A correct binomial distribution is stated

B1: Candidate has treated this as a paired design

M1: $\mathrm{P}(X \leq 4)$ has been attempted

A1: Correctly calculated the *p*-value as 0.00591

M1: Correct comparison using the split significance level as the candidate stated a two-tail test in hypotheses.

E1: Conclusion is correct and in context with the correct *p*-value and significance level.

Part (b)

B1: Correct hypotheses.

M1: There is an attempt at finding the differences.

A0: However, the candidate has now ranked these differences and calculated the mean and standard deviation for their ranked data. *Candidate uses an alternative valid method than that shown in the mark scheme.*

M1: The candidate has calculated the critical value using the formula and their mean and standard deviation.

A0: Due to the incorrect method previously, the correct critical value was not found, no credit given.

B1: The candidate did select the correct value of t to use in their critical value formula.

M0: Although they have correctly compared, this mark was dependent on obtaining the correct test statistic.

E0: Although conclusion was correct and in context for their values, this mark was dependent on the correct critical value and comparison.

Part (c)

E1: First mark was awarded for recognising the use of a small sample.

E1: Second mark was given for additional factors not taken into account 'the size of the players'.

E1: Third mark awarded for stating that the differences may not be normally distributed.

E0: Further comment about being in the same group was not awarded credit as not relevant to the question being asked.

Part (d)

E1: Candidate has recognised that the data for males may not be representative for females.

E1: Candidate has recognised that the data for football and combat sports may not be representative for netball.

Question 6 - Response C

6 Russell and Robert are interested in sport. They believe that wearing red may affect the success of a person or team in a sporting competition.

Russell finds the following information about the 2004 Olympic games.

- Male combatants in combat sports were randomly assigned either a red or a blue outfit.
- There was a total of 21 rounds of fighting for these combat sports.
- One round had an equal number of red and blue winners.
- Four rounds had **more blue** than red winners
- Sixteen rounds had **more red** than blue winners.

[Source: doi:10.1038/435293a]

(a) Making any necessary assumptions, use Russell's data to investigate their belief.

Make a full conclusion, in context, about any effect seen, or otherwise.

(7)

16+ 4-

$H_0 - p = 0.5$
$H_1 - p \neq 0.5$

2-tail
5% sig

$B(20, 0.5)$

$P(X \geq 16) = 0.00591$

Reject H_0 if
$P(X \geq 16) < 0.025$
$0.00591 < 0.025$
$\therefore$ reject H_0

There is sufficient evidence at the 5% level to suggest that wearing red does affect the success of male combatants, and given the data provided we may suggest it improves their success.

Q6

A
B
C

Robert collected data about the performance of five teams that sometimes wore a red kit at the Euro 2004 men's football competition.

He awarded one point for a win, zero for a draw and minus one point for a loss (disregarding penalty shoot-outs). He then found the average points scored when each team wore red and when they did not wear red.

Opposing football teams never wear the same colour kit as each other.

Robert's results are shown in **Figure 7**.

	Total points		Number of games		Average points	
Team	**Wearing red**	**Not wearing red**	**Wearing red**	**Not wearing red**	**Wearing red**	**Not wearing red**
Croatia	0	−1	1	2	0	−0.5
Czechia	1	2	1	4	1	0.5
England	1	0	1	3	1	0
Latvia	−1	−1	2	1	−0.5	−1
Spain	1	−1	2	1	0.5	−1

[Data source: doi.org/10.1038/435293a]

Figure 7

Q6

(b) Making any necessary assumptions, use the **average points** data in **Figure 7** and a **paired t-test** to investigate Russell and Robert's belief that wearing red may affect the success of a team.

Make a full conclusion, in context, about any effect seen, or otherwise.

(8)

~~$H_0: \mu_1 = \mu_2$~~

$H_0: \mu_d = 0$

$H_1: \mu_d \neq 0$

(d = red − not red)

5% sig°

2-tail

$T.S = \dfrac{\bar{d} - \mu_d}{s_d / \sqrt{n}}$

diff:
+0.5
0.5
1
0.5
1.5

$T.S = \dfrac{0.8 - 0}{0.447/\sqrt{5}} = \dfrac{0.8}{0.199} = 4.02$

C.V = 2.77

Reject H_0 5%°

T.S > C.V

4.02 > 2.77

∴ Reject H_0

There is sufficient evidence at the 5% level to suggest wearing red does effect a teams success. in this case it appears to improve success of football teams at Euro 2004.

(c) Considering the data and tests used in (a) and (b), make **four** distinct criticisms.

In your answer you might consider

- the size of sample provided,
- the sampling design,
- the test used and its assumptions.

(4)

The sample sizes in each instant were very small 20 and 5 meaning a distinct lack of reliability in the data of athletes and teams wearing red.

Through using random sampling it allows for other variables to impact upon the outcome for example individual differences such as ability of the opponent.

It is difficult to suggest the difference in the paired t test are normally distributed due to a lack of data and no evidence of symetry.

Furthermore, it is difficult to say test a used a random sample as all events were used from one thing so may impact each other, for example they should've looked at male combatants from a range of years.

21 /21

Part (a)

B1: Although not stated, the method of using a binomial and calculating $P(X \leq 4)$ was enough to achieve this mark.

B1: A correct binomial distribution is stated.

B1: Correct hypotheses using p in this case.

M1: $P(X \geq 16)$ has been attempted.

A1: Correctly calculated the p-value as 0.00591

M1: Correct comparison using the split significance level as the candidate stated a two-tail test in hypotheses.

E1: Conclusion is correct and in context with the correct p-value and significance level. The second comment was ignored but it is worth noting that a two-tailed test will not indicate that wearing red shows an improvement in the success.

Part (b)

B1: Correct hypotheses.

M1: There is an attempt at finding the differences.

A1: Correct values for mean and standard deviation have been calculated from the differences.

M1: The candidate has calculated the test statistic using the formula.

A1: The correct value of the test statistic was calculated.

B1: The candidate did select the correct value of t to use in their critical value formula.

M1: Correctly compared the correct test statistic and critical value.

E1: Conclusion is correct and in context. The second comment was again ignored but it is worth noting that a two-tailed test will not indicate that wearing red shows an improvement in the success.

Part (c)

E1: First mark was awarded for recognising the use of a small sample.

E1: Second mark was given for additional factors not taken into account 'ability of opponent'.

E1: Third mark awarded for stating that the differences may not be normally distributed.

E1: Forth mark awarded for recognising the samples are from particular events.

Part (d)

E1: Candidate has recognised that the data for males may not be representative for females.

E1: Candidate has recognised that the sample being small will provide limited size of evidence.