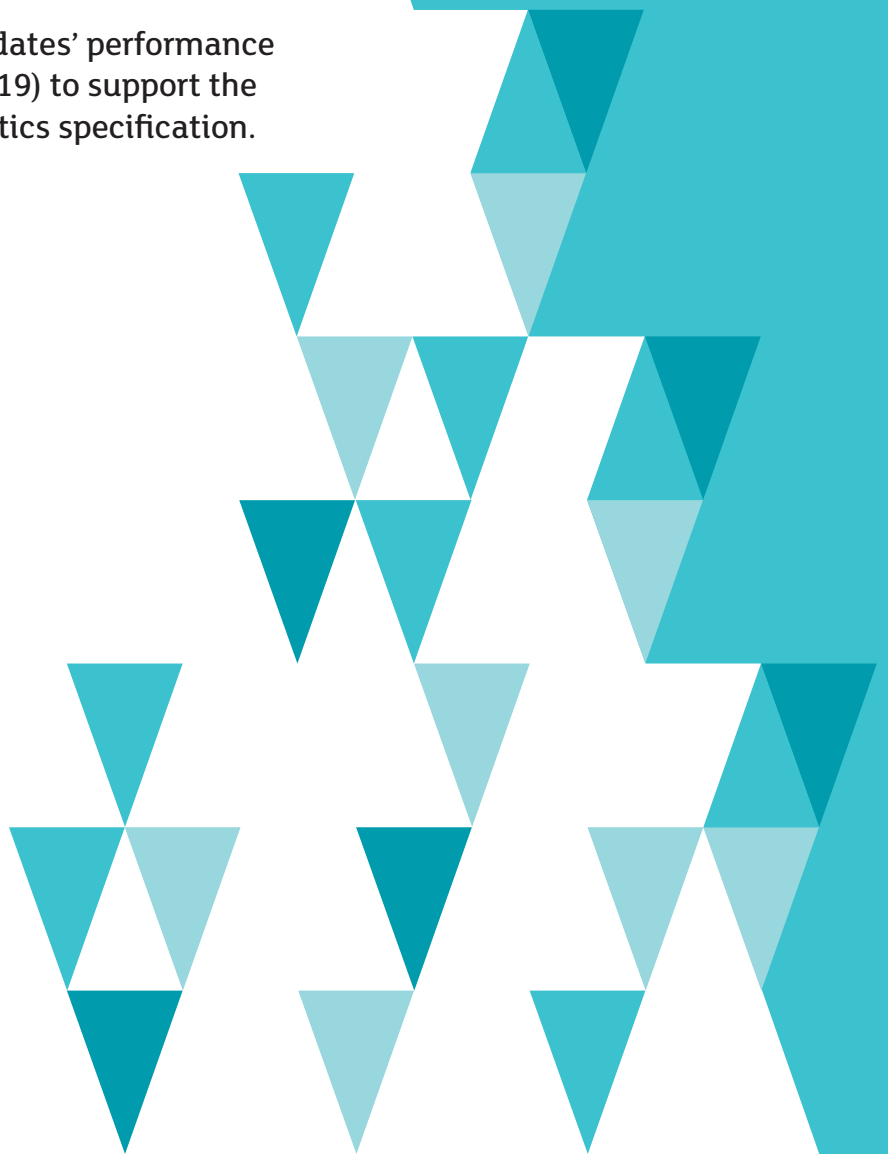


GCSE



CCEA GCSE Exemplifying Examination Performance **Statistics**

This is an exemplification of candidates' performance in GCSE examinations (Summer 2019) to support the teaching and learning of the Statistics specification.



Permission to reproduce all copyright material has been applied for. In some cases, efforts to contact copyright holders may have been unsuccessful and CCEA will be happy to rectify any omissions of acknowledgement in future if notified.

EXEMPLIFYING EXAMINATION PERFORMANCE

GCSE Statistics

Introduction

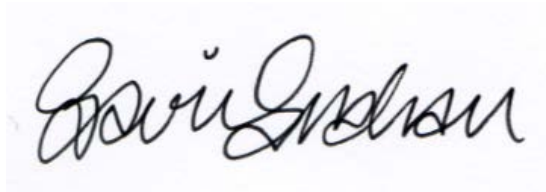
These materials illustrate aspects of performance from the 2019 summer GCSE examination series of CCEA's revised GCSE Specification in 2017.

Students' grade A responses are reproduced verbatim and accompanied by commentaries written by senior examiners. The commentaries draw attention to the strengths of the students' responses and indicate, where appropriate, deficiencies and how improvements could be made.

It is intended that the materials should provide a benchmark of candidate performance and help teachers and students to raise standards.

For further details of our support package, please visit our website at www.ccea.org.uk

Best wishes

A handwritten signature in black ink, reading 'Gavin Graham', is centered on the page. The signature is written in a cursive style with a small 'i' over the 'v' in 'Gavin'.

Gavin Graham

Education Manager, Statistics

Email: ggraham@ccea.org.uk

Telephone: 028 9026 1200 ext. 2658

GCSE: Statistics

Unit 1: Higher

Grade: A Exemplar

HIGHER TIER FORMULA SHEET

$$\text{Standard deviation} = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$$

Spearman's Rank Correlation Coefficient

$$r_s = 1 - \left(\frac{6 \sum d^2}{n(n^2 - 1)}\right)$$

Answer **all** questions

Q1 A researcher wants to get the opinions of a town's residents about proposed changes to the traffic system.

The council decides to carry out a sample survey to gather residents' opinions instead of a census.

Q1a Give two advantages to carrying out a sample survey rather than a census.

Student's response

1. *A Sample costs less* [1]
2. *A Sample takes less time* [1]

The researcher will use a questionnaire for the survey.

One of the questions on the questionnaire is:

<i>There is currently too much road traffic in the town. Don't you agree?</i>	
Yes <input type="checkbox"/>	No <input type="checkbox"/>

Q1b Give one reason why this question is not suitable. [1]

Student's response

This is a leading question. Leading questions are biased.

Before using the questionnaire with the sample of residents, the researcher decides to conduct a pilot survey.

Q1c Give two reasons why this is a good idea.

Student's response

1. *It makes sure the wording of the question is correct.* [1]
2. *It makes sure all possible responses are covered.* [1]

A sample of 300 people is chosen in the town centre for the survey.

Q1d Suggest a reason why this would not be a suitable sampling method. [1]

Student's response

No guarantee all 300 people drive. The sample must ask people who are relevant.

Examiner's comments

The candidate's responses in Question 1 were concise, accurate and relevant, indicating a detailed understanding of data collection processes. The candidate avoided a commonly observed misconception in (b) by correctly identifying that the problem with the survey question was due to it being a leading question rather than a closed question. In (c), the candidate accurately identified two independent benefits from conducting a pilot survey.

Marks: 5/6

Q2 The table below gives information about the employment rate and median starting salary for students graduating from university in 2017.

Graduates 21–30	Employment rate	Unemployment rate	Continuing education	Median salary (nearest £500)
Gender				
Male	78.6%	4.7%	17.0%	£26,500
Female	78.8%	3.5%	18.0%	£23,000
Area of England				
East Midlands	79.0%	3.8%	17.3%	£22,000
East of England	77.9%	5.2%	17.2%	£26,000
London	78.8%	5.4%	16.0%	£29,500
North East	74.9%	12.1%	12.9%	£22,000
North West	79.1%	2.4%	18.5%	£23,500
South East	80.3%	2.9%	16.8%	£26,500
South West	77.7%	2.0%	20.3%	£24,000
West Midlands	77.7%	3.5%	18.8%	£22,000
Yorkshire and the Humber	79.8%	2.7%	17.5%	£23,000

Source: <http://www.gov.uk>

For one area of England, the unemployment rate is twice that of another.

Q2a Name these two areas. [1]

Student's response

Yorkshire and the humber and London

Q2b Does the area of England with the highest employment rate have the lowest unemployment rate? [1]

Student's response

Yes No

The median starting salary for graduates in London was £29,500

Q2c Give an interpretation, in context, of this value. [2]

Student's response

The median salary was £29,500. This means that 50% of graduates in London have a salary less than £29,500 and 50% have a salary more than £29,500

James is in his final year at university and he is looking at the table to see where he should look for a job next year.

Q2d Which area of England would you recommend for James? [2]

Student's response

South east

Give a reason for your answer

It was the highest employment rate. for graduates.

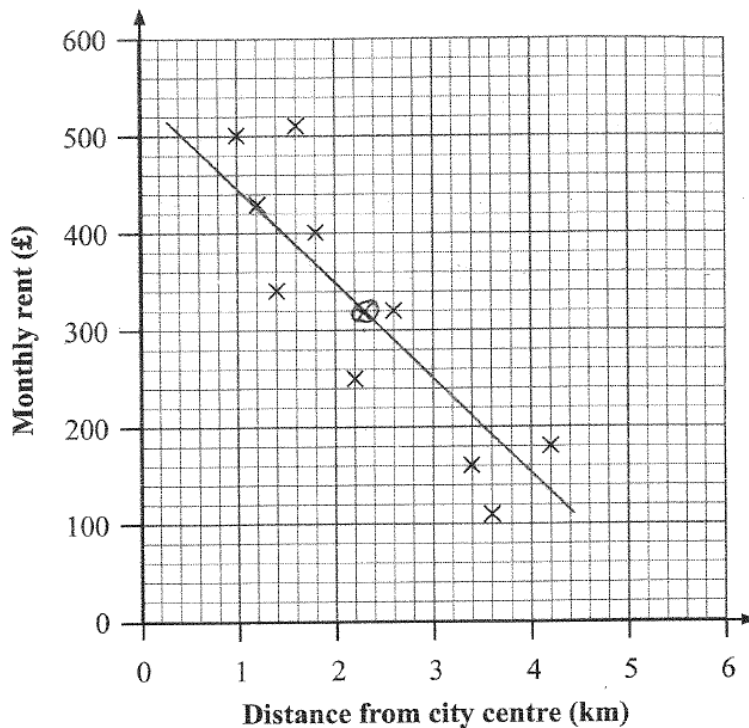
Examiner's comments

Of particular note in Question 2 was the candidate's ability in (c) to give an accurate and relevant interpretation of the median in context while avoiding unnecessary/irrelevant reference to other data in the table.

Marks: 6/6

Q3 The scatter diagram below shows some information about the monthly rent and distance from the city centre for 10 apartments in a city. [2]

Student's response



Q3a Show that the mean distance from the city centre is 2.3 km. [2]

Student's response

$$4.2 + 3.6 + 3.4 + 2.2 + 2.6 + 1.8$$

$$+ 1.4 + 1.2 + 1.6 + 1 = 23$$

$$23 \div 10 = 2.3$$

The mean monthly rent for the 10 apartments is £320

Q3b Plot the double mean point on the scatter diagram and draw a line of best fit. [2]

Q3c(i) What type of correlation does the scatter diagram show? [1]

Student's response

Positive Correlation

Negative Correlation

Q3c(ii) What conclusion can you draw about the relationship between the distance from the city centre and the monthly rent of an apartment? [1]

Student's response

As the distance from the city centre increases the monthly rent decreases.

Q3d Ella lives 3.0 km from the city centre.

Q3d(i) Could Ella use the line of best fit on the scatter diagram to estimate a price for her monthly rent? [1]

Student's response

Yes

No

Q3d(ii) Give a reason for your answer. [1]

Student's response

The point 3.0 is on the best fit line. It is between the range the line was drawn for.

Examiner's comments

The responses in this question were exactly what were required in each case. The candidate accurately read the distances from the scatter graph and demonstrated the correct mean calculation in (a) and then used the double mean point to plot the line of best fit in (b). The scatter graph was accurately interpreted in (c) and the correct reason was offered in (d) for why it was suitable to use the line of best fit to estimate the required price, i.e., that the point in question was within the range of the previously plotted data.

Marks: 8/8

Q4 Peter and Helen want to collect information about how much time pupils at their school spend on homework every week. They attend a large post-primary school with 1400 pupils.

Peter is planning to use systematic sampling to select 50 pupils.

Q4a Describe how Peter could select his sample. [3]

Student's response

Peter would collect all possible units. Peter would divide 1400 by 50 to get 28. Peter would pick a Start point at random and select every 28th pupil.

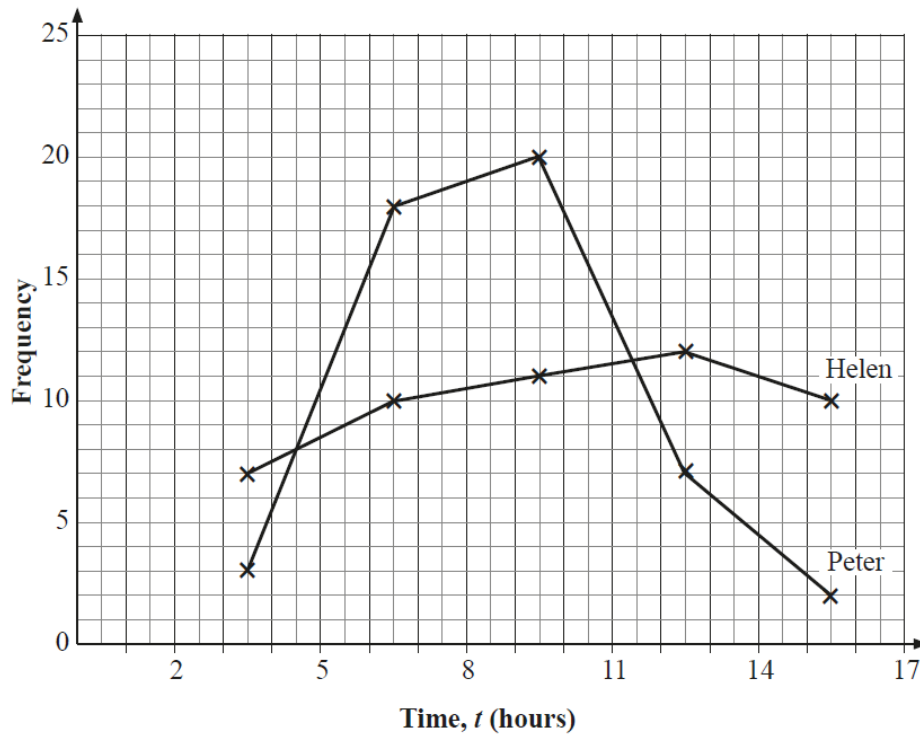
Helen is planning to select 50 pupils from outside the canteen.

Q4b Name the method of sampling which Helen is planning to use. [1]

Student's response

Opportunity sampling

Peter and Helen summarise their results in a frequency polygon.



Q4c Use the frequency polygon to write down:

Q4c(i) the median class for Peter's data; [2]

Student's response

Answer 8 – 11

Q4c(ii) the modal class for Helen's data. [1]

Student's response

Answer 11 – 14

Q4d Compare the results from both sample surveys. [2]

Student's response

Peter's sample shows more than Helen's sample

Q4e(i) Who has collected a more reliable set of data? [1]

Student's response

Peter

Q4e(ii) Give a reason for your answer. [1]

Student's response

In a systematic sample every person in the population has the opportunity to be chosen so it is more representative of the population.

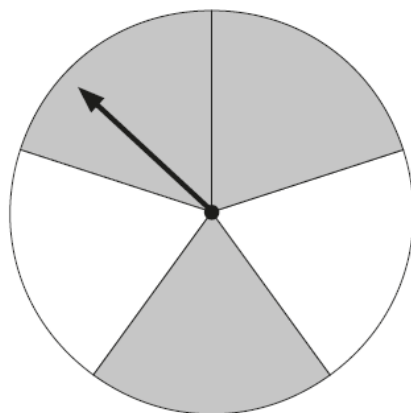
Examiner's comments

The candidate's description of the systematic sampling method was outstanding for its clarity and precision. This again demonstrated an excellent working knowledge of data collection processes. The candidate also showed an excellent working knowledge of frequency polygons and the ability to calculate relevant averages from them.

Marks: 9/11

Q5 Cathy and John are doing an investigation with a spinner.

The spinner has equally-sized sectors which are either shaded (S) or not shaded (N), as shown in the diagram below.



John spins the spinner twice.

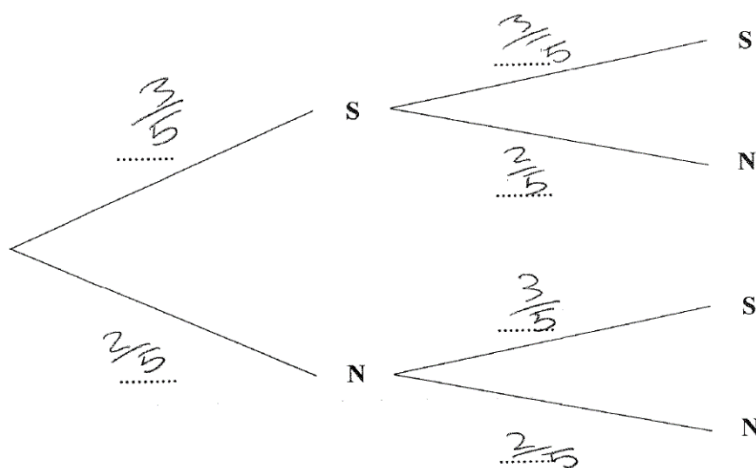
Q5a Write down the probability that the arrow lands on a shaded sector on John's first spin. [1]

Student's response

Answer $\frac{3}{5}$

Q5b Complete the probability tree diagram for John's two spins. [3]

Student's response



Q5c Calculate the probability that John gets

Q5c(i) a shaded sector on both spins; [2]

Student's response

$$\frac{3}{5} \times \frac{3}{5} \qquad \text{Answer } \frac{9}{25}$$

Q5c(ii) a different outcome on each spin. [3]

Student's response

$$\left(\frac{3}{5} \times \frac{2}{5}\right) + \left(\frac{2}{5} \times \frac{3}{5}\right) = \frac{12}{25}$$

$$\text{Answer } \frac{12}{25}$$

Cathy thinks that the spinner might be biased, so she spins it 200 times and records the outcome each time.

Q5d How many times should Cathy expect the spinner to land on a shaded sector? [1]

Student's response

$$\frac{3}{5} \times 200$$

$$\text{Answer } 120$$

In the 200 spins, the spinner landed on a shaded sector 118 times.

Q5e How might Cathy interpret this result? [2]

Student's response

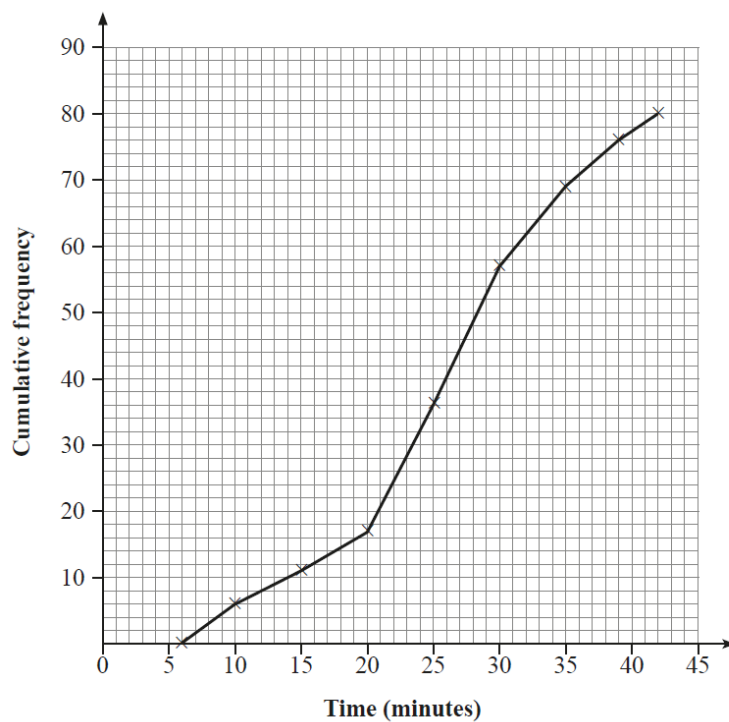
The number of times the spinner landed on a shaded sector was close to the estimate.

Examiner's comments

In this question, the candidate demonstrated a thorough grasp of the theory of probability and the computational skills associated with this topic, such as knowing when to add probabilities and when to multiply probabilities. All answers were calculated accurately with clear methodology shown.

Marks: 12/12

Q6 Mark surveyed the teachers in his school to find out how long it took them to travel to work. He plotted his results in the cumulative frequency diagram below.



Q6a What was the median time taken by the teachers to travel to school? [1]

Student's response

Answer 26 minutes

Q6b Find the interquartile range of the times taken by the teachers to travel to school. [2]

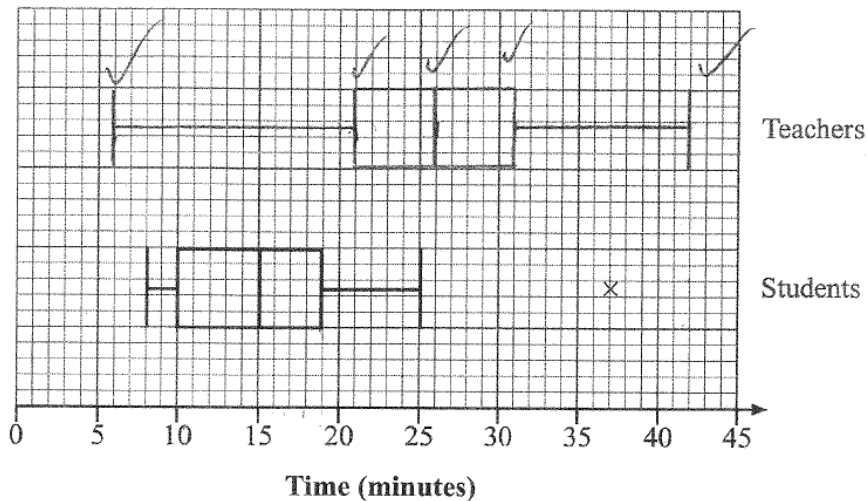
Student's response

$$31 - 21 = 10$$

Answer 10 minutes

On the same day, Mark surveyed the students in the school about how long it took them to travel to school. His results are summarised in the box plot below.

Student's response



The point marked \times on the diagram represents an outlier.

Q6c Use a suitable calculation to verify that the value represented by \times is an outlier. [3]

Student's response

$UQ + 1.5 (IQR) \quad \therefore 37 \text{ is an outlier}$

$$19 + 1.5 (19 - 10)$$

$$19 + (1.5)(9) = 32.5$$

Q6d In the space indicated on the diagram, draw a box plot for the teachers. [3]

Student's response

Can be seen in diagram above.

Q6e Compare the travel times for teachers with the travel times for students. [4]

Student's response

The median time for the teachers to get to school was higher than the median time for students to get to school. This shows that on average teachers take longer to get to school. The range in time for the students is smaller than the range in time for teachers. This shows there is less variation among the student times.

Due to roadworks, the travel time for every student in the school will increase by exactly 5 minutes.

Q6f Comment on the effect this will have on

Q6f(i) the median of the journey times for students; [1]

Student's response

The median time will increase by 5.

Q6f(ii) the range of the journey times for students. [1]

Student's response

The range will stay the same.

Examiner's comments

Question 6 tested a range of skills and the candidate's responses were outstanding throughout. Responses to (a) and (b) demonstrated the candidate's ability to analyse a cumulative frequency diagram and use it to work out statistics relating to both average and spread. The response in (c) demonstrated a clear and concise demonstration of the methodology to verify an outlier. The quality of written communication in (e) was excellent: the response was well-structured, correct terminology was employed, repetition was avoided and no irrelevant comparisons of the box plots were made. Both responses in (f) were correct, including the extra detail of the median increasing by 5 minutes.

Marks: 15/15

Q7 Table 1 below shows the estimated quarterly expenditure on cycling equipment in Northern Ireland between 2014 and 2016

Table 1

		Expenditure (nearest £ thousand)			
		Q1	Q2	Q3	Q4
Year	2014	127	188	240	160
	2015	145	199	254	147
	2016	145	202	254	163

The estimated expenditure on cycling equipment was highest in Quarter 3 each year.

Q7a Suggest a reason for this. [1]

Student's response

It was the best weather for cycling.

Q7b The first seven 4-point moving averages have been calculated, to the nearest thousand, as follows:

179 183 186 190 186 186 187

Q7b(i) Explain briefly why a 4-point moving average has been used. [1]

Student's response

There are 4 quarters in a year.

Q7b(ii) Calculate the last two 4-point moving averages. [2]

Student's response

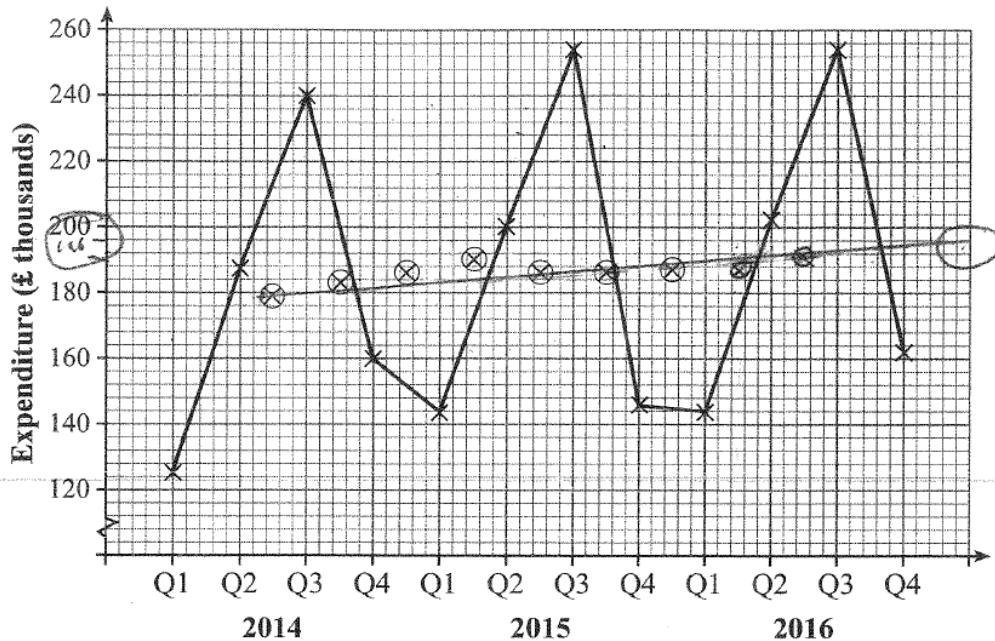
Answer 187 and 191

The time series graph below shows the data in **Table 1**.

In addition, the first seven moving averages have been plotted.

Q7c Plot the remaining two moving averages, calculated in part **(b)(ii)**, on the graph and draw a trend line. [2]

Student's response



Q7d Use your trend line to estimate the expenditure, to the nearest £ thousand, for Quarter 1 of 2017 [3]

Student's response

$$\frac{202+254+163+x}{4} = 196 \qquad 169 + x = 196 \times 4 = 780$$

$$780 - 619 = x$$

$$\frac{619+x}{4} = 196 \qquad x = 165$$

Answer £ 165 thousand

Examiner's comments

The candidate's responses in this question demonstrated a thorough knowledge of the theory behind moving averages/time series graphs and the associated methodology and calculations.

Marks: 7/9

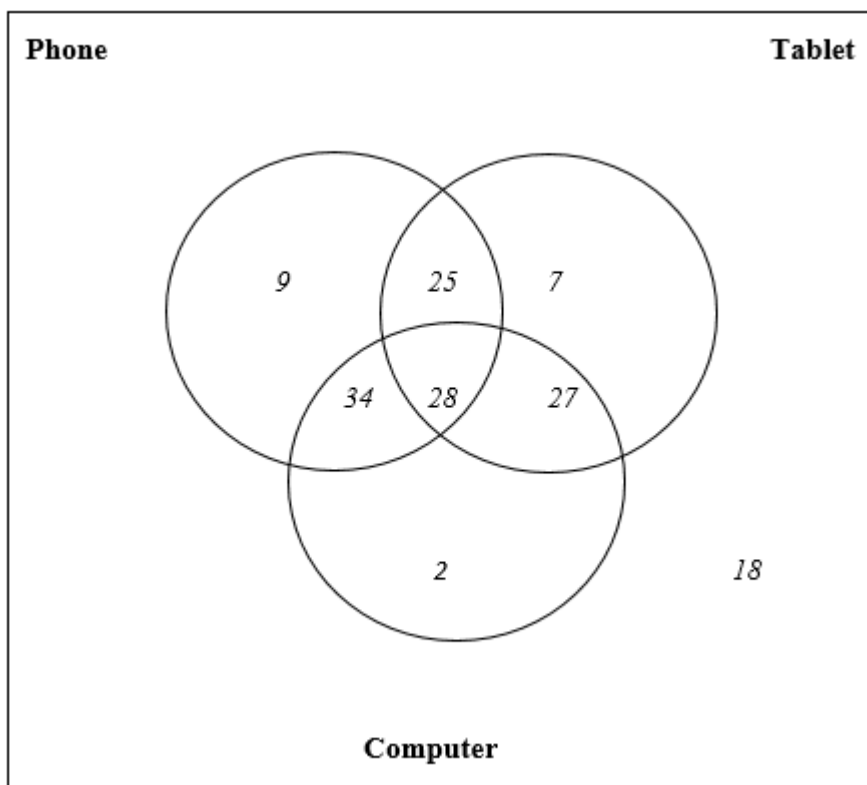
Q8 A group of 150 people were surveyed about electronic devices they owned.

The results were as follows:

- 96 people owned a phone,
- 87 people owned a tablet,
- 91 people owned a computer,
- 53 people owned a phone and a tablet,
- 62 people owned a phone and a computer,
- 55 people owned a tablet and a computer,
- 28 people owned all three devices

Q8a Complete the Venn diagram below. [4]

Student's response



Q8b How many people did not own a phone or a computer? [1]

Student's response

Answer 22

One of the 150 people surveyed was chosen at random.

Q8c Find the probability that this person owned a tablet only. [1]

Student's response

Answer $\frac{7}{150}$

One person who owned a tablet was chosen at random.

Q8d Find the probability that they owned a computer too. [2]

Student's response

$$\frac{28+27}{7+25+28+27} = \frac{55}{87}$$

Answer $\frac{55}{87}$

Examiner's comments

The candidate correctly filled in every section of the 3-circle Venn diagram and was able to use the diagram to correctly work out both a basic probability and a conditional probability.

Marks: 7/8

Q9 On a given day, the probability that Michael is on time for work is 0.6

Q9a Calculate the probability that Michael is not on time for work on two days in a row. [2]

Student's response

$$0.6 \times 0.6 = 0.36$$

Answer 0.36

Q9b(i) Write down the name of the most appropriate distribution to model the number of times Michael will be on time for work in a five-day period. [1]

Student's response

Answer *Binomial*

Q9b(ii) For this model, write down the number of trials, n , and the probability of a success, p .

Student's response

$$n = 5 \quad [1]$$

$$p = 0.6 \quad [1]$$

Q9c Calculate the probability that Michael will be on time for work twice in a five-day period. [3]

$$\text{You may use } (p + q)^5 = p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5$$

Student's response

$$10p^2q^3$$

$$10(0.6)^2(0.4)^3$$

$$= 0.2304$$

Answer 0.23

Q9d Calculate the probability that Michael will be on time for work at least four times in a five-day period. [3]

You may use $(p + q)^5 = p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5$

Student's response

$$p^5 + 5p^4q$$

$$p^5 = 0.6^5 = 0.07776$$

$$5p^4q = 5(0.6)^4(0.4)$$

$$= 0.2592$$

$$0.2592$$

$$+ 0.07776$$

$$= 0.33696$$

$$= 0.337$$

Answer 0.34

Examiner's comments

Question 9 described a specific context for which the candidate correctly identified the binomial distribution as the most appropriate model to apply; the candidate also correctly identified the model parameters for the number of trials and the probability of a success. For the required probability in (c), the candidate identified the appropriate term in the provided binomial expansion, substituted in the correct values and accurately calculated the result. For the required probability in (d), the candidate identified the two relevant terms in the provided binomial expansion, correctly calculated the probability associated with each term and then added together the results.

Marks: 9/11

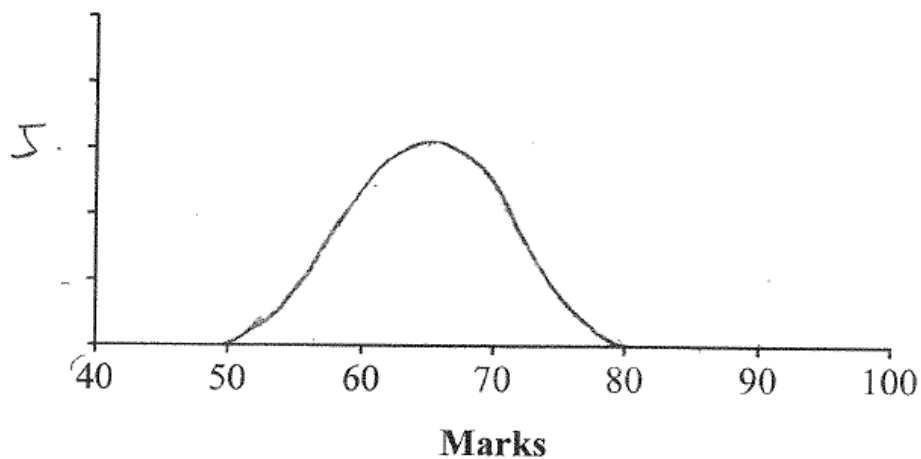
Q10 Year 12 students completed tests in Digital Technology, Home Economics and Religious Studies.

The results in each test were normally distributed.

The mean of the Digital Technology results was 65 and the standard deviation was 5

Q10a On the grid below sketch the distribution of the Digital Technology results. [2]

Student's response



Niamh got 71 marks in the Digital Technology test.

Q10b Calculate Niamh's standardised score in Digital Technology. [2]

Student's response

$$Z = \frac{x - \mu}{\sigma}$$

$$Z = \frac{71 - 65}{5}$$

Answer 1.2

Niamh's standardised score in the Home Economics test was 0.3

Q10c(i) Which test did Niamh do better in? [1]

Student's response

Digital Technology

Home Economics

Q10c(ii) Give a reason for your answer. [1]

Student's response

The standardised score for digital technology was higher so she did better. on the Digital Technology test.

The marks (x) of 30 students in the Religious Studies test are summarised as follows:

$$\sum x = 2124 \quad \sum x^2 = 151\,887$$

Q10d(i) Calculate the mean of the Religious Studies marks. [2]

Student's response

$$\frac{\sum fx}{\sum f} = \frac{\sum x}{\sum n} = \frac{2134}{30} = 70.8$$

Answer 71

Q10d(ii) Calculate the standard deviation of the Religious Studies marks. [3]

Student's response

$$\begin{aligned} \text{Stand. dev} &= \sqrt{\frac{\Sigma f x^2}{\Sigma f} - \left(\frac{\Sigma f x}{\Sigma f}\right)^2} \\ &= \sqrt{\frac{151887}{30} - \left(\frac{2124}{30}\right)^2} = 7.089 \end{aligned}$$

Answer 7.09

Niamh's standardised score in the Religious Studies test was -0.92

Q10e Using your answers to parts **(d)(i)** and **(d)(ii)**, calculate Niamh's test result in Religious Studies. [3]

Student's response

$$\begin{aligned} Z &= \frac{x - \mu}{\sigma} && -6.5228 + 71 \\ -0.92 &= \frac{x - 71}{7.09} && = 64.4772 \\ &&& \text{test round} \end{aligned}$$

$$7.09 \times -0.92 = x - 71$$

$$-6.5228 = x - 71$$

Answer 64

Examiner's comments

The candidate achieved full marks for the last question of the paper which tested understanding of various aspects of the normal distribution. The candidate correctly sketched a graph of the normal distribution for the provided parameters (a), calculated a standardised score for a given test result (b), compared and interpreted a pair of standardised scores (c), calculated a mean and standard deviation from a set of summary statistics (d) and calculated a test result from a given standardised score (e).

Marks: 14/14

The candidate's responses throughout the paper (especially in Questions 9 and 10) were very impressive and confirmed the candidate's thorough understanding and knowledge of the entire CCEA GCSE Statistics course.

GCSE: Statistics

Unit 2: Higher

Grade: A Exemplar

HIGHER TIER FORMULAE SHEET

$$\text{Standard deviation} = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$$

Spearman's Rank Correlation Coefficient

$$r_s = 1 - \left[\frac{6 \sum d^2}{n(n^2 - 1)} \right]$$

Answer **all** questions

Q1 The Visitor Attraction Survey is a census of all visitor attractions in Northern Ireland.

Q1a Explain what is meant by the term census. [1]

Student's response

Collecting every data from all source and combining them together.

Q1b Give one advantage and one disadvantage of taking a census.

Student's response

Advantage *The data will be more accurate* [1]

Disadvantage *It costs money to collect all data* [1]

Q1c Suggest a reason why the Visitor Attraction Survey could be incomplete. [1]

Student's response

There may be some visitor attractions that are not considered visitor attractions and are not collected for the survey.

Electronic counters were used by 23% of attractions to record the numbers of visitors to them.

Q1d(i) Explain how an electronic counter could over-count the number of visitors.
[1]

Student's response

Person may going in and out the attraction repeatedly.

Q1d(ii) Explain how an electronic counter could under-count the number of visitors.
[1]

Student's response

Visitors may walk in a parallel fashion and only counted as one person.

Examiner's comments

This question referred to information contained in the pre-release document and the candidate was clearly familiar with it. Parts (a) and (b) were generic and well-answered by this candidate. Parts (c) and (d) made specific reference to the pre-release and this candidate demonstrated familiarity through the clear answers supplied.

Marks: 6/6

Q2 The table below shows the number of visitors to some visitor attractions in Northern Ireland between 2013 and 2017

**Top ten participating visitor attractions
(Country Parks/Parks/Forests/Gardens)**

		Number of visitors (thousands)				
		2013	2014	2015	2016	2017
1	Lagan Valley Regional Park (inc. Lagan Towpath)	1397	1257	1296	1375	1427
2	Peace Bridge	1243	1071	981	981	926
3	Lurgan Park	770	784	n/a	634	702
4	Crawfordsburn Country Park	800	748	585	603	625
5	Kilbroney Park	115	95	122	124	534
6	Slieve Gullion Forest Park & The Giant's Liar	n/a	311	361	388	484
7	Antrim Castle Gardens	168	221	343	451	444
8	Dungannon Park	224	332	291	368	348
9	Sir Thomas & Lady Dixon Park	300	300	270	300	300
10	Roe Valley Country Park	250	250	250	248	273

Q2a How many visitors were there to Lurgan Park in 2014? [1]

Student's response

Answer 784000

Q2b What was the total number of visitors to Antrim Castle Gardens between 2013 and 2017? [1]

Student's response

Answer 1627000

Q2c Suggest a suitable diagram which could be used to illustrate the trend in the number of visitors to Dungannon Park over this five-year period. [1]

Student's response

Line Graph

Between 2013 and 2016, one of the attractions counted the number of cars coming to the site, whereas in 2017 the number of people visiting was counted.

Q2d(i) Which attraction do you think this could have been? [1]

Student's response

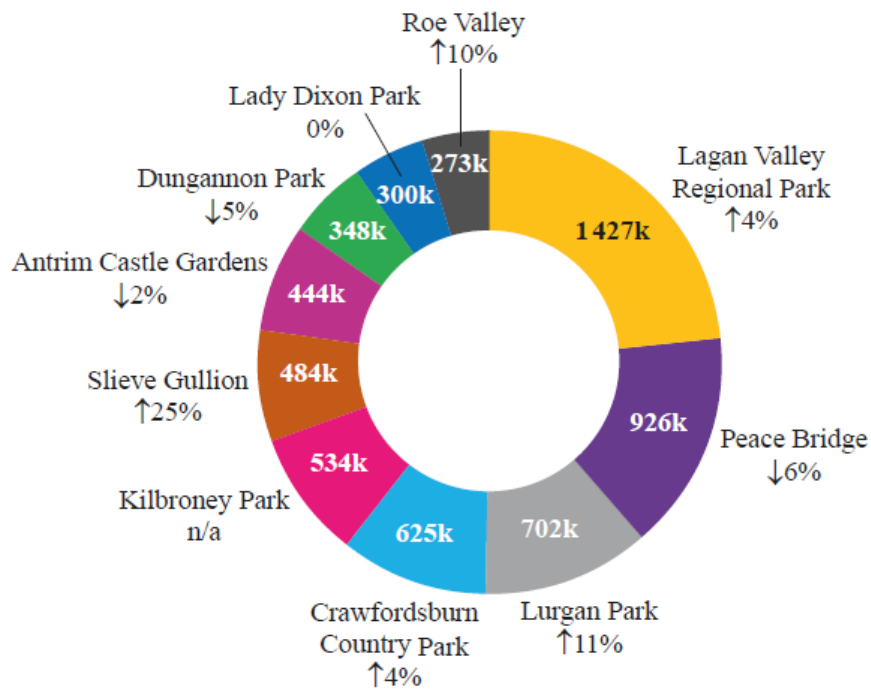
Kilbroney Park

Q2d(ii) Give a reason for your answer. [1]

Student's response

Number suddenly increased by 4.5 times because before the change, car can carry more than one person and some people may not be coming in by car.

Top Ten Country Parks/Parks/Forests/Gardens 2017



Q2e Explain why this type of diagram was used instead of a bar chart in the ‘Northern Ireland visitor attraction survey 2017’ report. [2]

Student’s response

It is more aesthetically pleasing; can be used to easily compare the competitiveness of the top 10 parks/forests/gardens.

The population of Northern Ireland is approximately 1.852 million.

Kerrie says that half of the population of Northern Ireland visited the Peace Bridge in 2017

Q2f(i) Do you think Kerrie is correct? [1]

Student's response

Yes No

Q2f(ii) Give two reasons for your answer.

Student's response

1. *Not all visits are from the population of Northern Ireland.* [1]
2. *People may need to travel across the bridge instead of visiting and it is counted as visiting.* [1]

Examiner's comments

This was a straightforward question requiring interpretation of a table of values taken directly from the pre-release document. The values in the table are expressed in thousands which this candidate has noticed as their answers to parts (a) and (b) are expressed in this form. For part (d) the answer is correct, although there is some irrelevant / hypothetical detail in part (d)(ii) but this is condoned since the correct answer appears within the response. Part (e) made some useful observations but was only worth one mark for the comment about comparisons. For the second mark, the candidate would need to have been specific about why the chart presented was better than a bar chart, such as referring to the proportions of visitors which are given in this chart but not a bar chart. The reference to 'aesthetically pleasing' is irrelevant. For part (f) the number of visitors to the Peace bridge was 926,000 which is indeed half of the figure quoted by Kerrie but her conclusion was wrong for two reasons: the visitors were not necessarily from Northern Ireland and some may have visited the bridge more than once, or used the bridge but not necessarily as a tourist. This candidate's answer makes reference to both of these points, so gets full marks.

Marks: 9/10

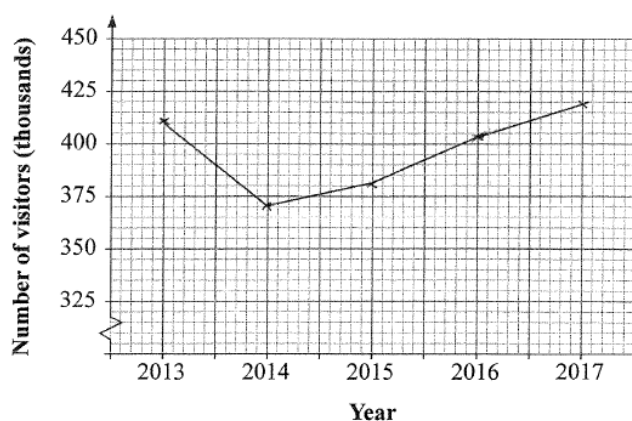
Q3 The table below shows the number of visitors, in thousands, to some visitor attractions in Northern Ireland between 2013 and 2017

**Top ten participating visitor attractions
(excluding Country Parks/Parks/Forests/Gardens)**

		Number of visitors (thousands)				
		2013	2014	2015	2016	2017
1	Giant's Causeway World Heritage Site	754	788	851	944	1012
2	Titanic Belfast	604	634	622	667	760
3	Dundonald International Ice Bowl	n/a	n/a	n/a	500	536
4	Ulster Museum	416	466	465	460	533
5	Carrick-a-Rede Rope Bridge	263	324	354	440	434
6	Kinnego Marina	326	209	428	430	430
7	Derry's Walls	411	370	381	403	419
8	The Guildhall	269	299	313	334	350
9	W5	270	324	372	370	323
10	SS Nomadic	n/a	n/a	n/a	n/a	269

Q3a Draw a line graph to show the number of visitors to Derry's Walls between 2013 and 2017. [3]

Student's response



Q3b(i) In which year did the fewest visitors come to Derry's Walls? [1]

Student's response

Answer 2014

Q3b(ii) Explain how this is shown on the graph. [1]

Student's response

The plot is the lowest of all plots.

Q3c Describe briefly the trend in the number of visitors to Derry's Walls over the five-year period. [2]

Student's response

The number of visitors decreases rapidly from 2013 to 2014, after that, it increases gradually.

Taking 2013 as the base year, Barry calculated index numbers for the number of visitors to Derry's Walls. His results were:

90.0 100 101.9 98.1 92.7

Barry has written these down in the wrong order.

Q3d Complete the table below using Barry's values. [3]

Student's response

Year	2013	2014	2015	2016	2017
No. of visitors (thousands)	411	370	381	403	419
Index number	<i>100</i>	<i>90.0</i>	<i>92.7</i>	<i>98.1</i>	<i>101.9</i>

Examiner's comments

This candidate has correctly drawn the line graph in part (a). A common error in this part was candidates drawing vertical lines at the years instead of plotting points. The interpretation and explanation of the methodology in part (b) is straightforward and correct, as is part (c) where only a simple explanation was required. In part (d), the index numbers had been worked out already so this was just a matching up exercise. This candidate has correctly placed the 100 under the base year (2013) and the other numbers have been placed in ascending order to match the ascending order of the visitor numbers.

Marks: 10/10

Q4 Keith reads a report about visits to visitor attractions. **The report says:**

- there were an estimated 15.0 million visits to visitor attractions in 2016; and
- there were a reported 19.8 million visits to visitor attractions in 2017 by respondents to a survey.

Keith decides to use the statistical enquiry cycle to investigate if there was an increase in the number of visits to all visitor attractions between 2016 and 2017

Q4a Explain how Keith could plan his investigation. [6]

Your answer should include:

- a suitable hypothesis for Keith's investigation
- what data Keith would need to collect
- a possible source for the data
- one problem Keith might encounter with his data collection and how this could be overcome.

Student's response

Keith can give a hypothesis that not all visitor attractions has an increase in numbers. Keith then need to collect that data for each attractions, he can go through different attraction websites for the number of visitors in 2016 and 2017. Keith might find out that an attraction does not count visitors number or does not publish them, then he can estimate by researching other attractions that has a similar size and scale.

Keith wants to analyse the data using diagrams and calculations.

Q4b State one diagram and one calculation he could use and explain the reason for choosing each. [4]

Student's response

Diagram

Two way bar graph

Explanation

It can represent increase in visitor number but can also represent decrease

Calculation

Percentage increase / decrease

Explanation

To normalise the result, giving a better representation.

Examiner's comments

The simplest route to six marks in part (a) was to follow the order of the bullet points and address each one carefully which is exactly what this candidate has done. In part (b) no credit was available for the first part as the named 'two-way bar graph' could imply a couple of different charts, but the second part was fine.

Marks: 7/10

Q5 The manager of a gift shop recorded the number of orders she received over a period of 15 days.

The results are shown in the stem and leaf diagram below.

KEY

0		2		means 2 orders
0		2	3	
1		1	2 4 4 6 8 9	
2		0	1 2 3 7	
3		2		

Q5a For the orders received over the 15-day period,

Q5a(i) find the median; [1]

Student's response

Answer 18

Q5a(ii) calculate the interquartile range. [2]

Student's response

$$\frac{15}{4} = 3.75 \qquad LQ = 12$$

$$VQ = 22$$

$$3.75 \times 3 = 11.25 \qquad IOR = 22 - 12$$
$$= 10$$

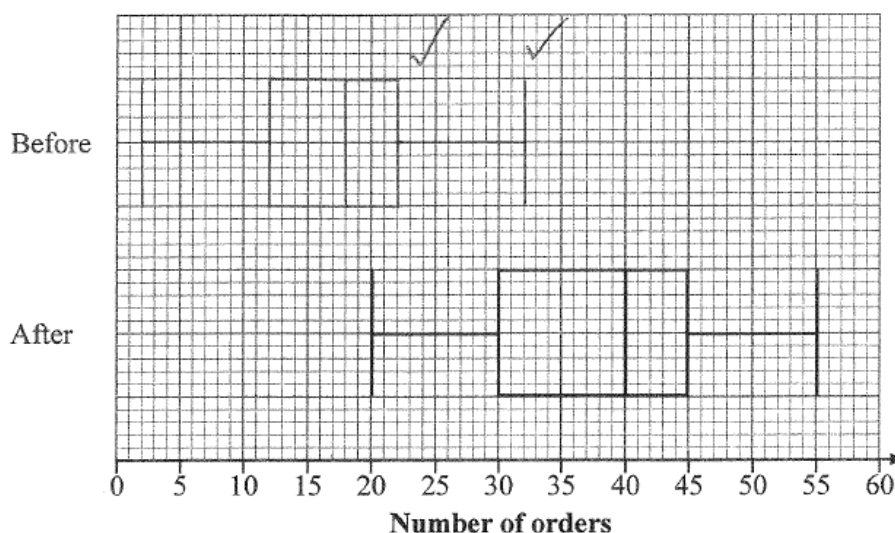
Answer 10

The manager decided to advertise her shop on a social media site.

She recorded the number of orders she received over the next 15 days and drew a box plot to display her results.

Q5b On the same diagram, draw a box plot to show the number of orders before advertising her shop. [2]

Student's response



Q5c Comment on the effectiveness of advertising on the number of orders received. [4]

Student's response

Although after advertising, it created a slightly more negative skewness, the number of orders each day does not drop below the original median, which means it is very effective. Upper quartile, lower quartile, median, lowest, and highest number all increased. But the range increases giving more unpredictability, the inter-quartile range has also increased. Overall, it is very effective.

Examiner's comments

The median in part (a) was fairly easy to find as the stem and leaf diagram had an odd number of entries. The candidate has shown their thinking for how to locate the quartiles: finding one-quarter of 15 (3.75) and three-quarters of 15 (11.25) then locating, respectively, the 4th and 12th items from the list. This is an excellent approach and will work for a set of discrete data of any size. The box plot is correctly drawn in part (b) and the candidate has drawn the lines with a ruler which was not always the case in many other scripts seen. An easy way to gain marks in comparison questions like part (c) is to give an account of the differences between the medians and IQRs and give an account for these, i.e. explain them in the context of the question. This candidate's approach is somewhat haphazard and lacks any real structure, but they do comment on the increases in the median and range which are individually worth a mark, but they only indicate what the increase in range means in context – 'more unpredictability'. This answer only just gets three marks.

Marks: 8/9

Q6 Rob is investigating the relationship between the admission charge and the annual number of visitors to tourist attractions in Northern Ireland.

Rob downloads some data from the internet.

Q6a Has Rob collected primary data or secondary data? [1]

Student's response

Primary data

Secondary data

Q6b State a hypothesis Rob could use for his investigation. [1]

Student's response

As the admission charge increases, the annual number of visitors to tourist attraction decreases.

Rob tries to calculate the product moment correlation coefficient for his data.

He says, ' $r = -1.615$ '

Q6c How can you tell Rob has made an error in his calculation? [1]

Student's response

The correlation coefficient cannot be lower than -1.000

Rob's data is shown in the table below.

Admission charge	£4.00	£5.30	£2.50	£7.80	£6.20	£9.40	£11.00	£8.30
Number of visitors (thousands)	148	206	251	445	214	197	75	223

Q6d Use the statistical functions on your calculator to calculate the product moment correlation coefficient for Rob's data.

Give your answer correct to three decimal places. [2]

Student's response

Answer -0.164

Q6e Give an interpretation, in relation to the hypothesis stated in part (b), to the value you found in part (d). [2]

Student's response

It represents a very weak negative correlation, the hypothesis is not true. The correlation is weak enough to disregard and can be seen as no correlation.

Examiner's comments

The majority of candidates got the mark for part (a) and most got part (b) too. This candidate has correctly stated that the reason for the error is that r cannot be less than -1 , but a very common incorrect answer was that r could not be negative. The value of r in part (d) has been correctly calculated as -0.164 and the candidate has not written anything else down so has used their calculator, as is required by the specification. For both marks in part (e), the candidates needed to state that the value of r indicated a weak negative correlation (or even that it was close enough to 0 to suggest no correlation), and then make reference to *their* hypothesis in (b) which this candidate does. Many candidates did not get this last mark as there was no reference to part (b), even though this was made explicit in the question.

Marks: 7/7

Q7 The frequency table below shows the masses, in kilograms, of all suitcases weighed at an airport security section before one flight.

Mass, m (kg)	Frequency	<i>Medium of mass</i>	<i>Fx (freq \times mass)</i>	
$0 \leq m < 5$	1	2.5	2.5	
$5 \leq m < 10$	14	7.5	105	
$10 \leq m < 15$	37	12.5	462.5	
$15 \leq m < 20$	54	17.5	945	
$20 \leq m < 25$	17	22.5	382.5	
$25 \leq m < 35$	2	30	60	

Q7a Give one advantage and one disadvantage of using a grouped frequency table.

Student's response

Advantage *Can easily observe the distribution of data across groups* [1]

Disadvantage *Loss of exact data* [1]

Q7b Show that an estimate of the mean mass of a suitcase is 15.66 kg.

You may use the blank columns in the table to help you. [4]

Student's response

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f}$$

$$\begin{aligned}\Sigma f \\ &= 1 + 14 + 37 + 54 + 17 + 2 \\ &= 125\end{aligned}$$

$$\begin{aligned}\Sigma fx \\ &= 2.5 + 105 + 462.5 + 945 + 382.5 + 60 \\ &= 1957.5\end{aligned}$$

Answer 15.66

$$\begin{aligned}\text{Mean} \\ &= \frac{1957.5}{125} \\ &\approx 15.66 \text{ kg}\end{aligned}$$

Carla estimates the standard deviation for the masses of the suitcases to be 4.79 kg.

She says that two-thirds of the suitcases should have masses between 10.87 kg and 20.45 kg.

Q7c Show how Carla arrived at these numbers. [2]

Student's response

$$\frac{2}{3} = 66\% \approx 68\%$$

68% of masses lies 1 Standard deviation away from the mean.

$$\text{Upper : } 15.66 + 4.79 = 20.45 \text{ kg}$$

$$\text{Lower : } 15.66 - 4.79 = 10.87 \text{ kg}$$

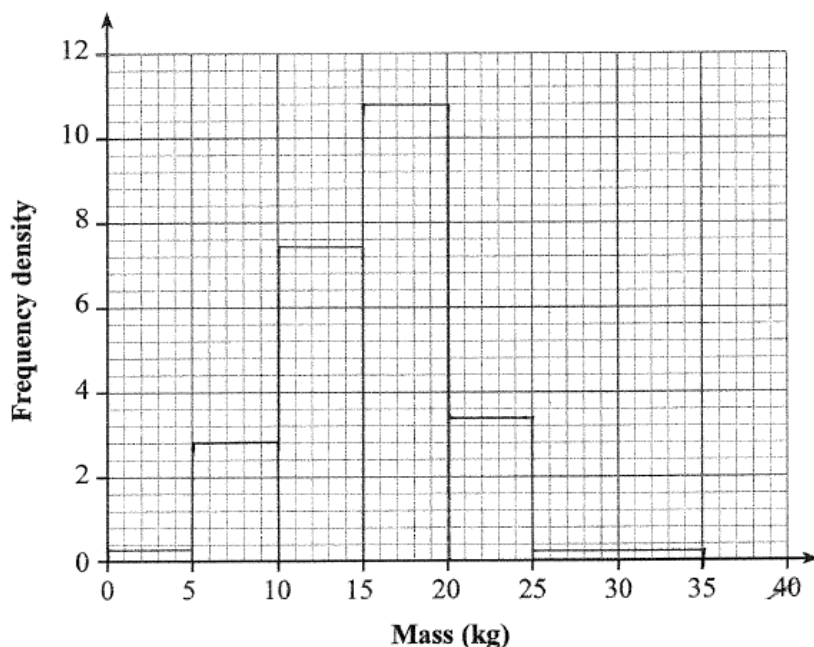
Q7d What assumption did Carla make about the distribution of masses of the suitcases?
[1]

Student's response

The masses in each class is normal distributed

Q7e Draw a histogram to display the data from the frequency table. [4]

Student's response



Q7f Use your histogram to decide whether the assumption made by Carla in part (d) is justified. [2]

Student's response

I do not think the assumption is justified as the shape of the graph does not form a smooth bell shape curve and it is not symmetrical.

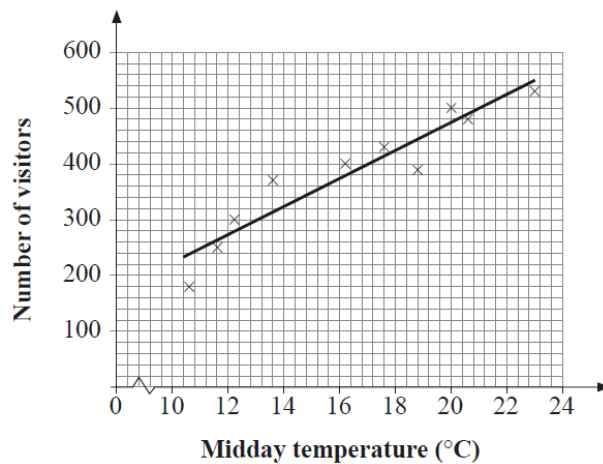
Examiner's comments

This candidate has noticed that the width of the final class is different from the others so avoids a mistake in part (b) and part (e). The calculation in part (b) is well laid out and clear understanding is evident. Unlike many candidates, this candidate knew that two-thirds of the masses of the suitcases would be within one standard deviation of the mean, which they have demonstrate clearly in part (c) and the correct assumption of normality was given in (d). The histogram in (e) is also correct, and this candidate avoided the common error of assuming all class widths were the same. No marks were available for the response in (e) as the candidate did not grasp that the normal distribution was being used as a model so an exactly symmetrical histogram was not required.

Marks: 12/15

Q8 A travel agent recorded the midday temperature, x , and the number of visitors, y , to a park over a period of ten Saturdays between April and June.

His results are shown on the scatter diagram below.



Q8a What type of data is shown in the scatter diagram?

Circle the two correct words below. [2]

Student's response

Qualitative

Univariate

Bivariate

Ordinal

Quantitative

Categorical

Q8b What was the midday temperature on the coldest Saturday? [1]

Student's response

Answer 10.6 °C

Q8c Explain why it is appropriate to draw a line of best fit on this scatter diagram. [1]

Student's response

The data 'shows' a linear relationship.

The equation of the line of best fit is $y = -25.5 + 24.9x$

Q8d Give an interpretation, in context, of the number 24.9 in the equation. [2]

Student's response

The number of visitors rises on each degree higher. For example, There will be 24.9 visitors more if the temperature is to increase from 13° C to 14° C.

Q8e Use the equation of the line of best fit to estimate the number of visitors to the park when the midday temperature is 16° C [2]

Student's response

$$\begin{aligned} Y &= -25.5 + 24.9 \times 16 \\ &= 372.9 \\ &\approx 373 \end{aligned}$$

Answer 373

One Saturday in January, the midday temperature was 0° C

Kieran used the equation of the line of best fit to estimate the number of visitors to the park.

Q8f Write down Kieran's result. [1]

Student's response

Answer -25.5

Q8g Give two reasons why Kieran's use of the line of best fit was inappropriate. [1]

Student's response

1. *0°C is an outlier (out of range) from the data shown on the graph* [1]
2. *The line of best fit only predicts the number of visitors from April to June* [1]

Examiner's comments

This candidate demonstrated a clear understanding of how scatter diagrams and lines of best fit, and their equations, should be interpreted. In particular, the interpretation of the gradient of the line was very clear in (d) as were the substitutions in (e) and (f). Unlike many candidates, this candidate gave two different reasons for the inappropriate use of the line of best fit in (g): the fact that zero fell outside the range of values for which the equation of the line had been calculated was noted, acceptably, by the candidate as an outlier. In addition, this candidate noticed that the months were different so had clearly read the question.

Marks: 11/11

- Q9** The partially completed table below shows the mean children's admission price and chain base index number for some tourist attractions in Northern Ireland.

Year	2013	2014	2015	2016	2017
Mean price	£2.54	£2.88	£3.89	£4.14	£3.78
Chain base index number	100	113.4	135.1	106.4	91.3

The chain base index number for 2014 is 113.4

- Q9a** Give an interpretation of this number. [2]

Student's response

The mean price for 2014 increases by 13.4% compared to 2013's mean price.

- Q9b** Calculate the mean price for 2016 [2]

Student's response

Mean Price

$$= 3.89 \div 100 \times 106.4$$

$$= £4.14$$

Answer £ 4.14

Q9c Calculate the chain base index number for 2015 [2]

Student's response

Chain base index number for 2015

$$= 3.89 \div 2.88 \times 100$$

$$= 135.1$$

Answer 135.1

Q9d Give one reason for using chain base index numbers instead of simple index numbers. [1]

Student's response

The value of money subjects to inflation and deflation and will change over time. It is unfair to compare to the first price after 4 years.

A country park has three levels of membership: bronze, silver and gold.

50% of members have bronze membership

35% of members have silver membership

15% of members have gold membership.

The prices for each level of membership for 2016 and 2017 are given in the table below.

	Bronze	Silver	Gold
2016	£15.99	£17.99	£19.49
2017	£16.49	£18.29	£20.99

Q9e Taking 2016 as the base year, calculate the weighted index number for the income from membership fees for 2017

Give your answer correct to one decimal place. [5]

Student's response

Index number for Bronze

$$= 16.49 \div 15.99 \times 100$$

$$= 103.1$$

Silver

$$= 18.29 \div 17.99 \times 100$$

$$= 101.7$$

Gold

$$= 20.99 \div 19.49 \times 100$$

$$= 107.7$$

Answer 103.3

Weighted index number

$$= 103.1 \times 50\% + 101.7 \times 35\% + 107.7 \times 15\%$$

$$= 103.3$$

Q9f Explain why a weighted mean is a more appropriate measure of average than the geometric mean in part (e). [1]

Student's response

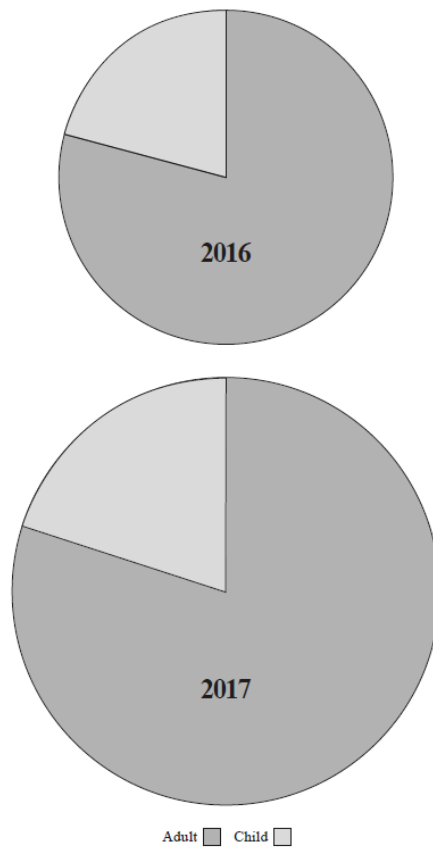
Not all index number affect the same proportion of the income from membership. Geometric mean assume so.

Examiner's comments

This was a testing question on index numbers. Parts (a) to (c) were fairly standard and well-answered. The candidate has given good detail in part (d) to account for why chain base index numbers would be used over simple index numbers. The calculation in (e) was testing and this candidate has given the detail for each level of membership separately before weighting them, showing a clear understanding of how these numbers are constructed. Again, the answer to (f) indicates understanding of the influence of proportions on weighted index numbers.

Marks 13/13

Q10 The comparative pie charts below show the proportions of adults and children who visited a church in 2016 and 2017



Q10a Explain how the pie charts show that the church had more visitors in 2017 than 2016 [1]

Student's response

Its area is larger.

In 2016, the number of children who visited the church was 1575

Q10b Calculate the number of adults who visited the church in 2016 [3]

Student's response

Angle of children Sector

$$= 75^\circ$$

Number of adults

$$= 1575 \div \frac{75^\circ}{360^\circ} \times \frac{360^\circ - 75^\circ}{360^\circ}$$

$$= 1575 \times \frac{\cancel{360^\circ}}{75^\circ} \times \frac{\cancel{360^\circ}}{285^\circ}$$

$$= 1575 \times 3.8$$

$$= 5985$$

Answer 5985

Q10c Given that the pie charts are drawn to scale, calculate the number of adults who visited the church in 2017 [5]

Student's response

Total number of 2017

$$= \frac{\pi r^2}{\pi R^2} \times N_1$$

$$= \frac{6.4^2}{5^2} \times (5985 + 1575)$$

$$= 1.6384 \times 7560$$

$$= 12386$$

Angle of adult sector

$$= 360^\circ - 72^\circ$$

$$= 288^\circ$$

Number of adults in 2017

$$= 12386 \times \frac{288^\circ}{360^\circ}$$

$$\approx 9909$$

Answer 9909

Examiner's comments

In all parts requiring calculation, this candidate spells out their working and does not rely on a formula or procedure at any time. This is again evident in parts (b) and (c). The thinking demonstrated in the candidate's line of writing in (b) shows an impressive understanding of this topic. Similarly, in part (c), which was rarely answered correctly, the candidate has methodically calculated the total number of visitors to the church in 2017 by comparing the areas of the circles, not just R^2 with r^2 , and then worked out the number of adult visitors to the church using the angle they correctly measured.

Marks: 9/9



INVESTORS
IN PEOPLE

