Qualification Accredited

OCR
Oxford Cambridge and RSA

AS/A LEVEL GCE

Examiners' report

MATHEMATICS (MEI)

3895-3898, 7895-7898

4766/01 Summer 2018 series

Version 1

Contents

Introduction

Our examiners' reports are produced to offer constructive feedback on candidates' performance in the examinations. They provide useful guidance for future candidates. The reports will include a general commentary on candidates' performance, identify technical aspects examined in the questions and highlight good performance and where performance could be improved. The reports will also explain aspects that caused difficulty and why the difficulties arose, whether through a lack of knowledge, poor examination technique, or any other identifiable and explainable reason.

Where overall performance on a question/question part was considered good, with no particular areas to highlight, these questions have not been included in the report. A full copy of the question paper can be downloaded from OCR.

Paper 4766/01 series overview

Statistics 1 (4766/01) can be taken as one of three components for AS Level Mathematics or six components for A Level Mathematics. It is an optional component, consisting entirely of statistics. This component develops the statistics studied at GCSE and introduces hypothesis testing based on the binomial distribution. To do well on this paper, candidates need to be very confident with the material studied at GCSE and be able to apply their understanding of new areas, such as the binomial distribution and more advanced probability.

Overall candidates found this an accessible paper, with lower ability candidates able to access most of the questions. It was however rigorous enough to challenge and differentiate the most able candidates, as very few gained full, or very close to full, marks.

The paper gave most candidates sufficient time to fully answer all the questions, although there was evidence of possibly rushed responses at the end of the paper from a small number of candidates.

Over-specification was less of an issue than in previous sessions, however some marks were lost because of premature rounding in calculations (for example in Q4(ii)) or truncating final answers (such as the first frequency density in Q7(v)). Candidates should be advised to ensure that they only round the final answer and that they round correctly.

The use of correct notation was also secure from almost all candidates. However, poor interpretative skills and a lack of accuracy in terminology let down many candidates. They have, in most cases, been well prepared for calculations required in the paper, but less so for analysing or explaining their findings.

Section A

Question 1(i)

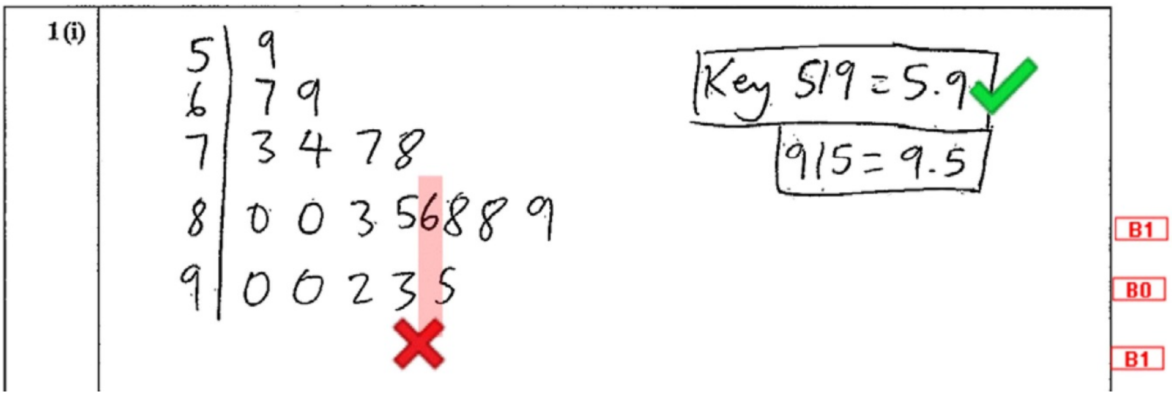
1 During a storm in the English Channel, the heights in metres of a random sample of 20 waves were measured. The heights are given below.

9.5	8.8	8.9	8.0	7.3	8.5	7.4	8.6	9.0	7.7
6.9	8.0	6.7	8.8	7.8	9.0	9.2	5.9	8.3	9.3

(i) Construct a sorted stem and leaf diagram to represent these data, taking stem values of 5, 6, 7, 8, 9. **[3]**

Although almost all candidates were credited at least 2 marks out of the 3 available, many did not align their leaves properly and thus did not gain the final mark. If a candidate misses out one of the data values in their diagram, the diagram needs to be redrawn, rather than inserting the omitted value in between two other values. One way of avoiding this error is for candidates to look at the list of data values and number them in order from lowest to highest, before adding them to the stem and leaf diagram.

Exemplar 1



This exemplar illustrates a situation where a candidate realised that one of the values has been omitted. However, instead of redrawing the diagram, the candidate adds the missing value of 8.6 between 8.5 and 8.8. The diagram is therefore not properly aligned and so the relevant mark cannot be credited.

Question 1(iii)

(iii) Write down the median and midrange of the data. **[2]**

Candidates almost always found the median correctly, although occasionally responses of 8.3 were given instead of 8.4. Many did not know how to find the midrange however, with a wide variety of wrong answers seen (including the range, amongst others).

Question 1(iv)

(iv) Give one reason why the median is a better measure of central tendency for these data than the midrange. **[1]**

The vast majority of candidates did not take into account the instruction 'for these data' and instead gave a general comment that could have related to any data.

Question 2(i)(*A*)

2 Each morning, Peter either cycles or drives to work. For any day, the probability that he drives is 0.25. If he drives, the probability that he arrives late for work is 0.2. The overall probability that he is late for work on any day is 0.08.

For a randomly chosen day,

- D is the event that Peter drives to work,
- L is the event that Peter arrives late for work.

(i) (*A*) Find $\mathrm{P}(D \cap L)$. **[2]**

This question required candidates to rearrange the conditional probability formula to $\mathrm{P}(D \cap L) = \mathrm{P}(L|D) \times \mathrm{P}(D)$ and then insert the relevant probabilities. Many were successful, but many others were unable to translate the words in the question into symbols, or used the wrong formula. Common wrong answers were $0.2 \times 0.08 = 0.016$ and $0.25 \times 0.08 = 0.02$.

Question 2(i)(*B*)

(*B*) Draw a Venn diagram showing the events D and L, and fill in the probability corresponding to each of the four regions of your diagram. **[3]**

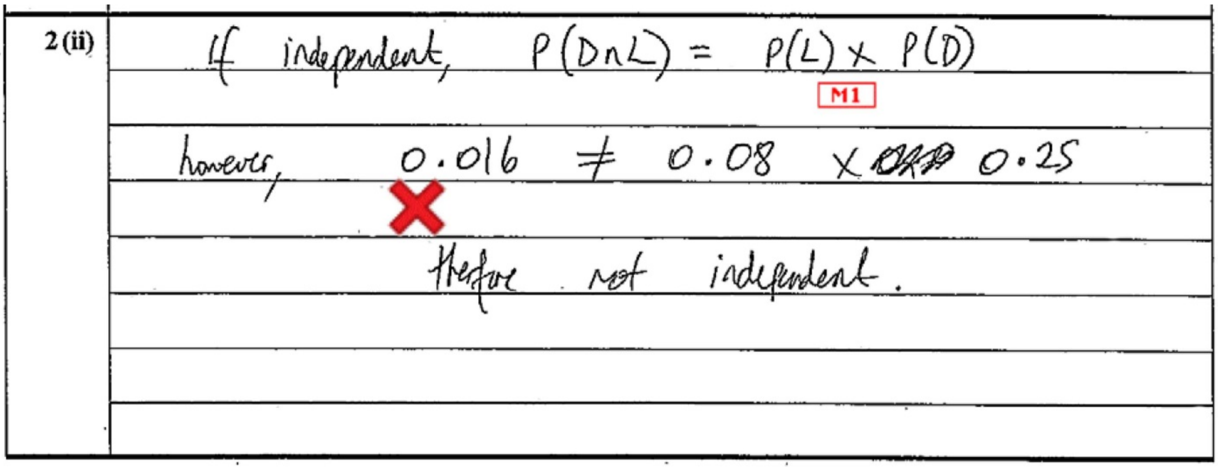
Most candidates answered this question successfully, with many gaining full marks for following through their answer to part (i). Follow through was allowed only if their answer to part (i) was less than 0.08, or in other words less than the value of $\mathrm{P}(L|D)$.

Question 2(ii)

(ii) Determine whether or not the events D and L are independent, justifying your answer. **[2]**

This question could be answered very easily by simply stating that $\mathrm{P}(L|D)$ is not equal to $\mathrm{P}(L)$ and so the events are not independent. Candidates who used this method were only credited 2 marks if they had the correct value of $\mathrm{P}(L|D)$ in part (i). However most candidates instead compared $\mathrm{P}(D \cap L)$ with $\mathrm{P}(D) \times \mathrm{P}(L)$. To be credited both marks for this method, candidates had to have the correct value of $\mathrm{P}(D \cap L)$ in part (i), which of course many did not have. Some chose to use $\mathrm{P}(D|L)$ instead of $\mathrm{P}(L|D)$, but this approach was very rarely successful.

Exemplar 2


2 (ii)
If independent, P(D∩L) = P(L) × P(D)
M1
however, 0.016 ≠ 0.08 × ~~0.2~~ 0.25
therefore not independent.

This candidate's response would be correct if they had found the correct value of 0.05 in part (i). However only 1 mark was credited here since the value of P($L\cap D$) is in fact 0.05. Follow through is sometimes allowed in this type of situation, but not in this particular question. This is because of the fact that, as mentioned above, all that is required is to state that P($L|D$) is not equal to P(L) and so the events are not independent.

Question 3(ii)(*A*)

(ii) The team coach decides that the squad must consist of equal numbers of women and men.

(*A*) How many different squads are possible now? **[2]**

Most candidates gained both marks, but a few added rather than multiplied the two correct combinations and so scored zero.

Question 3(ii)(*B*)

(*B*) There are 4 players from the squad on the court at any time. Assuming that all possibilities are equally likely, find the probability that all of the players from the squad who are on the court are women. **[3]**

More able candidates usually had no problem with this question. However many did not read the question correctly, often not realising that 4 players were needed or from how many players the choice was made. Most candidates attempted a fraction method rather than using combinations, however, many gave an answer of 0.5^4, or multiplied their correct answer by a combination. Others thought that this was a binomial situation. Finally some would have been credited all 3 marks, but then gave their answer either as 0.03 or 0.033. The first of these lost the final mark for lack of accuracy and the second because they missed out a zero.

Question 4(i)

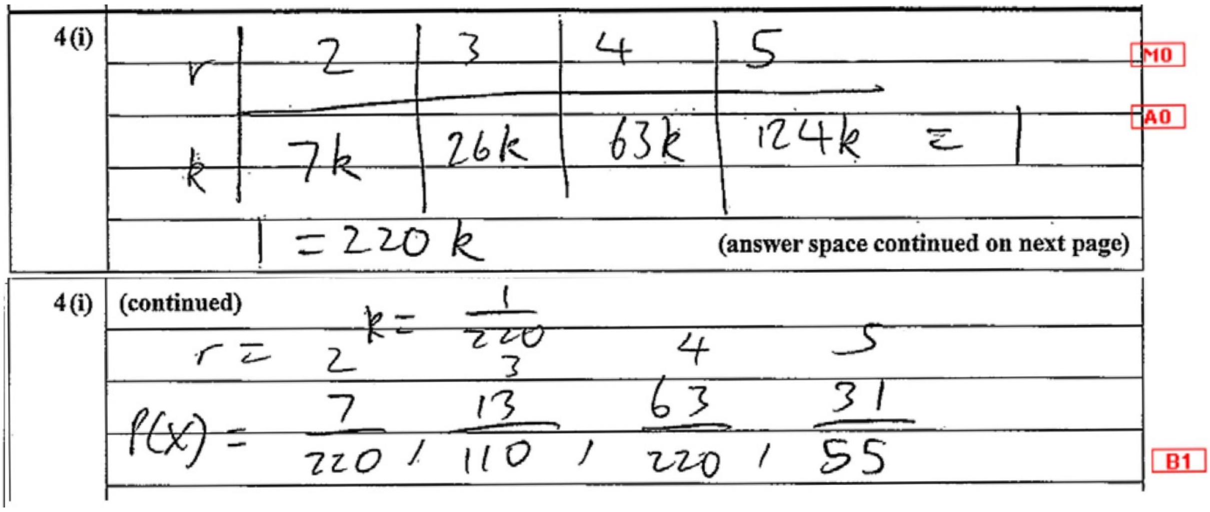
4 The probability distribution of the random variable X is given by the formula

$$P(X = r) = k(r^3 - 1) \text{ for } r = 2, 3, 4, 5.$$

(i) Show that the value of k is $\frac{1}{220}$ and, using this value of k, show the probability distribution of X in a table, giving the probabilities as exact fractions. **[3]**

The majority of candidates correctly showed that $k = \frac{1}{220}$ by forming an equation summing the probabilities and making it equal to 1, before stating that $220k = 1$ and then $k = \frac{1}{220}$. Some candidates who tried to use this method did not show that they were adding the probabilities together and so could not gain the two marks for this part of the question. Some candidates instead substituted $k = \frac{1}{220}$ into the probabilities, which was an acceptable method provided that they then showed that the probabilities added to a total of one, but some did not do this convincingly. The table of probabilities gained most candidates the final mark, but a few candidates gave the probabilities as decimals rather than the fractions asked for in the question.

Exemplar 3


4(i)

r	2	3	4	5	
k	7k	26k	63k	124k	= 1

= 220 k

M0 A0

(answer space continued on next page)

4(i) (continued)

k = 1/220

r =	2	3	4	5
P(X) =	7/220 ,	13/110 ,	63/220 ,	31/55

B1

Although at first sight this response seems correct, there is no indication that the four probabilities are being added and so neither of the first two marks was credited. The final table was given the mark available, despite the commas and the lack of a box around the table.

Question 4(ii)

(ii) Find E(X) and Var(X). **[5]**

This question was well answered with most candidates being credited all of the 5 marks available. Candidates who lost marks usually lost the final accuracy mark for Var(X) due to using a rounded value for the mean in their calculation of the variance. It was pleasing to see that very few candidates lost a mark for over-specification of their answer. Many of those who did over-specify gave their answer as a fraction before a decimal answer and so still gained full credit.

Question 5(i)

5 The probability of someone who lives in a particular city being a car owner is 0.3. The probability of someone who lives in the countryside surrounding the city being a car owner is 0.75. Two people who live in the city and two people who live in the surrounding countryside are selected at random.

(i) Find the probability that exactly one of these four people is a car owner. **[4]**

There was a very mixed response to this question with approximately half of candidates being credited all 4 marks, some with just a few lines of working out. Others drew tree diagrams and then tried to pick out the correct branches to follow. Others mixed up the probabilities and used 0.7 when they should have used 0.3 and likewise with 0.75 and 0.25. The majority of candidates gained one mark for a quadruple with at least two correct decimals, but many did not get the correct four quadruples. Some got the two different correct quadruples, but either didn't realise that they needed two of both quadruples or multiplied the quadruples by a number other than 2.

Question 5(ii)

(ii) Given that exactly one of the four people is a car owner, find the probability that this person lives in the city. **[3]**

Marks credited for this part were almost equally divided between 0, 1 and 3, with very few candidates scoring 2. Those who scored 1 mark did so because they realised that they had to divide a probability by their answer to part (i), but they used the wrong probability as their numerator.

Section B

Question 6(ii)

A student thinks that if a survey were to be carried out now, the figure would be lower than 50%. She selects a random sample of 20 males in this age group and asks each of them whether they have drunk alcohol in the last week. The number of them who say that they have drunk alcohol in the last week is 6.

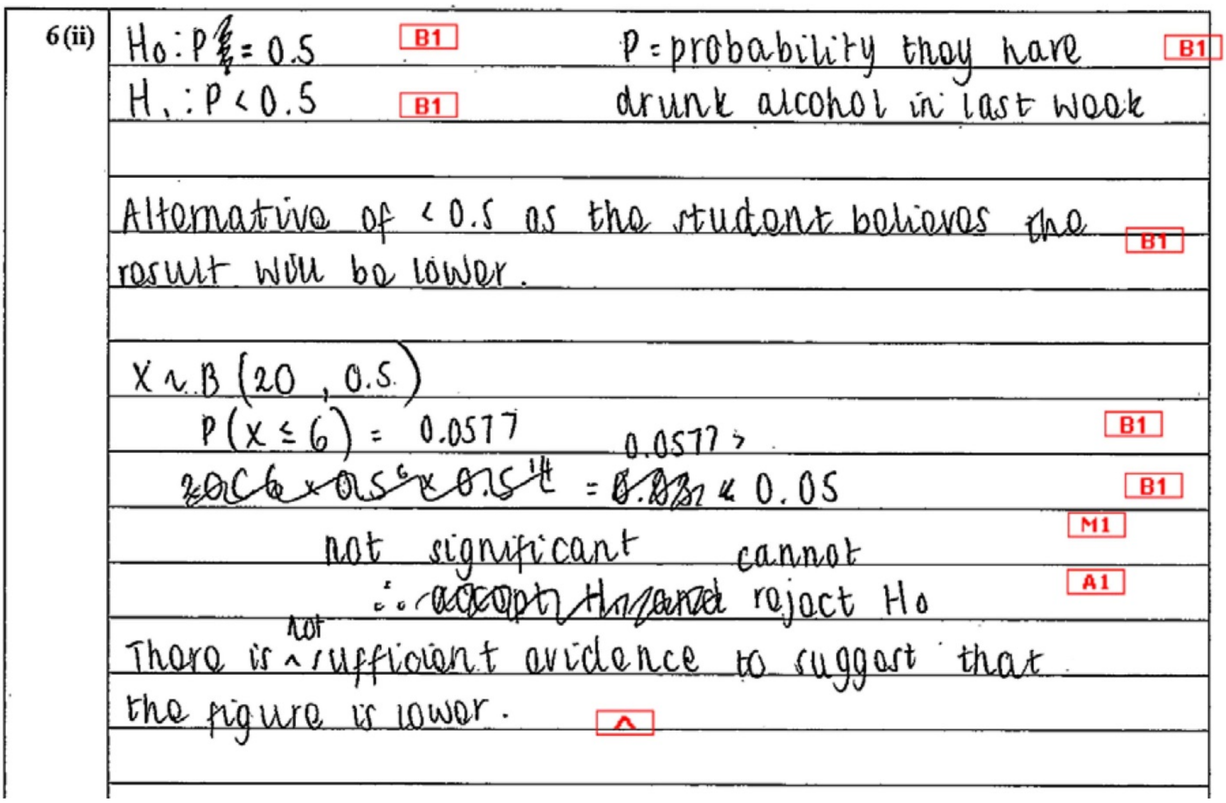
(ii) Carry out a hypothesis test at the 5% significance level to investigate the student's belief. Give a reason for your choice of alternative hypothesis. **[9]**

There were many very pleasing responses to this question, with two thirds of candidates being credited at least 8 marks out of the 9 available. However the first mark for defining *p* still causes difficulty for some candidates; a single letter must be used and the definition must both include the word probability or similar and be in the context of the question. The explanation of the reason for choice of alternative hypothesis should be just that, i.e. not a repeat of the statement of the alternative hypothesis in words rather than symbols. Most candidates correctly found the probability $P(X \leq 6) = 0.0577$, but then some did not explicitly compare it to the significance level. Candidates must actually write $0.0577 > 0.05$.; if they do not do this then they cannot score any of the last 3 marks for the question. A few did write a comparison, but then thought that this meant that the null hypothesis should be rejected.

A small number of candidates used the critical region method. Some of those who tried to find the critical region again missed out comparison with 5% and so again lost the final marks.

Whichever method is used, the final answer should be in the context of the question and should also include an element of doubt. Many candidates, having done everything correctly up to this stage were denied the final mark due to one of these two reasons, or despite putting their answer in context they discussed the 'number' rather than the 'proportion' of young males who had drunk alcohol in the last week.

Exemplar 4


6(ii)
H0: p = 0.5 B1 p = probability they have B1
H1: p < 0.5 B1 drunk alcohol in last week
Alternative of < 0.5 as the student believes the result will be lower. B1
X ~ B(20, 0.5)
P(X ≤ 6) = 0.0577 B1
0.0577 > 0.05 B1
not significant M1
∴ cannot reject H0 A1
There is not sufficient evidence to suggest that the figure is lower. ^

This response is almost perfect and gained 8 marks out of the 9 available. The hypotheses are stated in symbols and the meaning of *p* is defined. The reason for choice of alternative hypothesis is given clearly and concisely. A random variable *X* is defined and then P($X \le 6$) is found correctly and compared to 0.05. The correct result of 'not significant' is then reached. The candidate then states as part of the final conclusion 'not sufficient evidence to suggest that'. However the final conclusion is not in context and so the final mark was not scored.

Question 6(iii)

(iii) A teacher at the school attended by the student suggests that she should have used a larger sample. A new random sample of 100 males between the ages of 16 and 24 is selected. The number of them who say that they have drunk alcohol in the previous week is 41. Using the same hypotheses as in part **(ii)**, carry out another test at the 5% significance level. You may use the information that for $X \sim \mathrm{B}(100, 0.5)$,

$\mathrm{P}(X = 40) = 0.0108$, $\mathrm{P}(X = 41) = 0.0159$, $\mathrm{P}(X < 41) = 0.0284$, $\mathrm{P}(X \leqslant 41) = 0.0443$. **[4]**

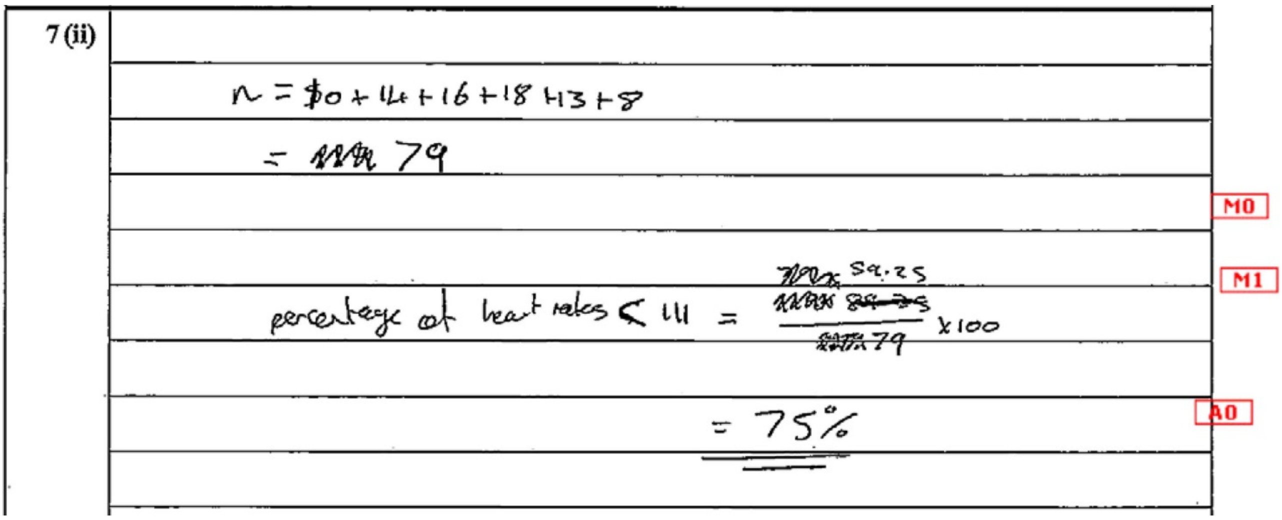
This final part was again generally well done, although again approximately one third of candidates did not give their answer in context. A few chose the wrong probability to compare with 5% or compared all of them with 5% and thus marks were not credited.

Question 7(ii)

(ii) Previously collected data suggest that the 75th percentile of the resting heart rates of all 4-year-old children is 111. Calculate an estimate of the percentage of children in the sample whose resting heart rate is 111 or below. **[3]**

There was a mixed response to this question with just under half of candidates being credited with all 3 marks. Many did not gain the first mark for linear interpolation - several used half way through the group and others divided by 14 and multiplied by 10 rather than the other way around. Others gave no indication of where their numerator came from, although they were dividing by 79 and so scored a method mark. After getting their value some candidates did not work out the percentage (or used 100 instead of 79 as their total). A few candidates tried to subtract from the top end, but those that did often gained all 3 marks

Exemplar 5



This candidate correctly divides by 79 and so gained a method mark. There is however no indication of where their numerator of 59.25 comes from and so they did not gain any further credit.

Question 7(iii)

The table below shows the resting heart rates, y BPM, of a sample of 18-year-olds.

Resting heart rate	$35 \leqslant y < 50$	$50 \leqslant y < 60$	$60 \leqslant y < 70$	$70 \leqslant y < 75$	$75 \leqslant y < 85$	$85 \leqslant y < 105$
Frequency	4	12	30	16	22	16

(iii) Calculate estimates of the mean and standard deviation of these data. **[4]**

The majority of candidates were credited with all 4 marks here. Almost all used correct midpoints, although a few made an error with 42.5 or 80. When calculating S_{xx} some candidates squared the frequencies instead of the midpoint. Candidates who rely on the built in functions of their calculator need to be aware that if they show no method they will not get any marks if they put an incorrect value into their calculator. If candidates do use their calculator for this type of question, they should be encouraged to write down the midpoints that they are using; this will both allow them to score at least one mark if they make an error with one midpoint, and also help them to put them correctly into their calculator. A pleasingly small number of candidates over-specified the value of the standard deviation thus denying themselves the final mark.

Question 7(iv)

(iv) Use your answers to part **(iii)** to investigate whether there may be any outliers. **[4]**

Most candidates used the correct formula for outliers, but some came to an incorrect conclusion – either saying that there were no outliers or that there were definitely outliers. Very few candidates tried to use the median and interquartile range, but some got confused and used 1.5 or 2.5 as the multiplier for the standard deviation.

Question 7(v)

(v) Add a histogram for these data on the copy of the diagram in the answer book. **[4]**

This was well answered although quite a few candidates gave the first frequency density as 0.26, thus denying themselves a mark. A very few lower ability candidates calculated group width divided by frequency. The histogram was drawn more accurately than charts in previous years due to the scales being given on the axes. The first bar was sometimes incorrectly drawn at 2.7 instead of 0.27. Some candidates had the lower boundary of the first bar at 30 instead of 35 or the upper boundary of the final bar at 110 instead of 105.

Question 7(vi)

(vi) Use the two histograms to compare the central tendency and variation of the resting heart rates of 4-year-old children and 18-year-olds. **[3]**

Only a few candidates gained full marks on this question with the most common omission being a comment about the modal class for each group. Many thought there was a difference in the variation but did not back it up with a comment about the range (or other measure of variation). Some candidates showed that they did not understand what central tendency means, seeming to think that it is something to do with variation.

Copyright acknowledgement

Section B, Question 6

Drinking Habits Amongst Adults, 2012', Office for National Statistics, www.ons.gov.uk, 2012. Reproduced under the terms of the Open Government Licence v3.0.

Supporting you

For further details of this qualification please visit the subject webpage.

Review of results

If any of your students' results are not as expected, you may wish to consider one of our review of results services. For full information about the options available visit the OCR website. If university places are at stake you may wish to consider priority service 2 reviews of marking which have an earlier deadline to ensure your reviews are processed in time for university applications.

activeresults

Active Results offers a unique perspective on results data and greater opportunities to understand students' performance.

It allows you to:

- Review reports on the **performance of individual candidates**, cohorts of students and whole centres
- **Analyse results** at question and/or topic level
- **Compare your centre** with OCR national averages or similar OCR centres.
- Identify areas of the curriculum where students excel or struggle and help **pinpoint strengths and weaknesses** of students and teaching departments.

http://www.ocr.org.uk/administration/support-and-tools/active-results/

CPD Hub
Your route to OCR's teacher training

Attend one of our popular CPD courses to hear exam feedback directly from a senior assessor or drop in to an online Q&A session.

https://www.cpdhub.ocr.org.uk

The small print

We'd like to know your view on the resources we produce. By clicking on the 'Like' or 'Dislike' button you can help us to ensure that our resources work for you. When the email template pops up please add additional comments if you wish and then just click 'Send'. Thank you.

Whether you already offer OCR qualifications, are new to OCR, or are considering switching from your current provider/awarding organisation, you can request more information by completing the Expression of Interest form which can be found here: www.ocr.org.uk/expression-of-interest

Our documents are updated over time. Whilst every effort is made to check all documents, there may be contradictions between published support and the specification, therefore please use the information on the latest specification at all times. Where changes are made to specifications these will be indicated within the document, there will be a new version number indicated, and a summary of the changes. If you do notice a discrepancy between the specification and a resource please contact us at: resources.feedback@ocr.org.uk.

OCR acknowledges the use of the following content:
Square down and Square up: alexwhite/Shutterstock.com

Please get in touch if you want to discuss the accessibility of resources we offer to support delivery of our qualifications: resources.feedback@ocr.org.uk

Looking for a resource?

There is now a quick and easy search tool to help find **free** resources for your qualification:

www.ocr.org.uk/i-want-to/find-resources/

www.ocr.org.uk

OCR Customer Contact Centre

General qualifications

Telephone 01223 553998
Facsimile 01223 552627
Email general.qualifications@ocr.org.uk

OCR is part of Cambridge Assessment, a department of the University of Cambridge. *For staff training purposes and as part of our quality assurance programme your call may be recorded or monitored.*

Cambridge Assessment

CERTIFIED ISO 9001

UKAS MANAGEMENT SYSTEMS 001