

ADVANCED SUBSIDIARY GCE

MEI STATISTICS

Statistics 1 (Z1)

G241

QUESTION PAPER

Candidates answer on the printed answer book.

OCR supplied materials:

- Printed answer book G241
- MEI Examination Formulae and Tables (MF2)

Other materials required:

- Scientific or graphical calculator

Thursday 26 May 2011
Morning

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the printed answer book and the question paper.

- The question paper will be found in the centre of the printed answer book.
- Write your name, centre number and candidate number in the spaces provided on the printed answer book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the printed answer book.** Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

This information is the same on the printed answer book and the question paper.

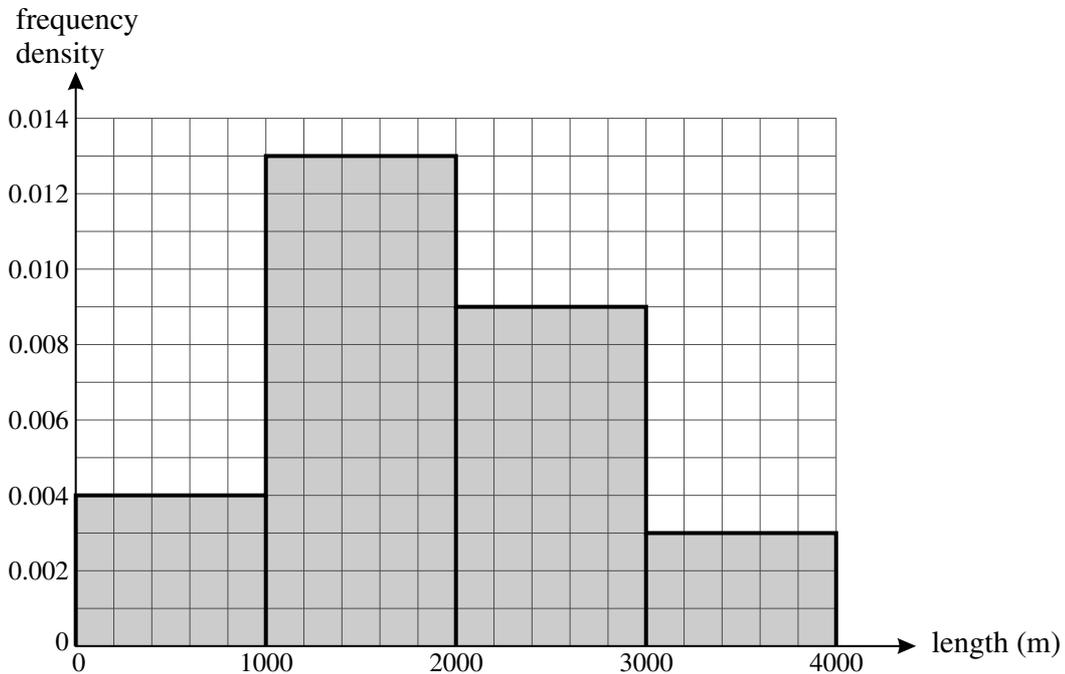
- The number of marks is given in brackets [] at the end of each question or part question on the question paper.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- The total number of marks for this paper is **72**.
- The printed answer book consists of **12** pages. The question paper consists of **4** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER / INVIGILATOR

- Do not send this question paper for marking; it should be retained in the centre or destroyed.

Section A (36 marks)

- 1 In the Paris-Roubaix cycling race, there are a number of sections of cobbled road. The lengths of these sections, measured in metres, are illustrated in the histogram.



- (i) Find the number of sections which are between 1000 and 2000 metres in length. [2]
- (ii) Name the type of skewness suggested by the histogram. [1]
- (iii) State the minimum and maximum possible values of the midrange. [2]
- 2 I have 5 books, each by a different author. The authors are Austen, Brontë, Clarke, Dickens and Eliot.
- (i) If I arrange the books in a random order on my bookshelf, find the probability that the authors are in alphabetical order with Austen on the left. [2]
- (ii) If I choose two of the books at random, find the probability that I choose the books written by Austen and Brontë. [3]
- 3 25% of the plants of a particular species have red flowers. A random sample of 6 plants is selected.
- (i) Find the probability that there are no plants with red flowers in the sample. [2]
- (ii) If 50 random samples of 6 plants are selected, find the expected number of samples in which there are no plants with red flowers. [2]

- 4 Two fair six-sided dice are thrown. The random variable X denotes the difference between the scores on the two dice. The table shows the probability distribution of X .

r	0	1	2	3	4	5
$P(X = r)$	$\frac{1}{6}$	$\frac{5}{18}$	$\frac{2}{9}$	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{1}{18}$

- (i) Draw a vertical line chart to illustrate the probability distribution. [2]
- (ii) Use a probability argument to show that
- (A) $P(X = 1) = \frac{5}{18}$, [2]
- (B) $P(X = 0) = \frac{1}{6}$. [1]
- (iii) Find the mean value of X . [2]
- 5 In a recent survey, a large number of working people were asked whether they worked full-time or part-time, with part-time being defined as less than 25 hours per week. One of the respondents is selected at random.
- W is the event that this person works part-time.
 - F is the event that this person is female.

You are given that $P(W) = 0.14$, $P(F) = 0.41$ and $P(W \cap F) = 0.11$.

- (i) Draw a Venn diagram showing the events W and F , and fill in the probability corresponding to each of the four regions of your diagram. [3]
- (ii) Determine whether the events W and F are independent. [2]
- (iii) Find $P(W | F)$ and explain what this probability represents. [3]
- 6 The numbers of eggs laid by a sample of 70 female herring gulls are shown in the table.

Number of eggs	1	2	3	4
Frequency	10	40	15	5

- (i) Find the mean and standard deviation of the number of eggs laid per gull. [5]
- (ii) The sample did not include female herring gulls that laid no eggs. How would the mean and standard deviation change if these gulls were included? [2]

Section B (36 marks)

7 Any patient who fails to turn up for an outpatient appointment at a hospital is described as a ‘no-show’. At a particular hospital, on average 15% of patients are no-shows. A random sample of 20 patients who have outpatient appointments is selected.

(i) Find the probability that

(A) there is exactly 1 no-show in the sample, [3]

(B) there are at least 2 no-shows in the sample. [2]

The hospital management introduces a policy of telephoning patients before appointments. It is hoped that this will reduce the proportion of no-shows. In order to check this, a random sample of n patients is selected. The number of no-shows in the sample is recorded and a hypothesis test is carried out at the 5% level.

(ii) Write down suitable null and alternative hypotheses for the test. Give a reason for your choice of alternative hypothesis. [4]

(iii) In the case that $n = 20$ and the number of no-shows in the sample is 1, carry out the test. [4]

(iv) In another case, where n is large, the number of no-shows in the sample is 6 and the critical value for the test is 8. Complete the test. [3]

(v) In the case that $n \leq 18$, explain why there is no point in carrying out the test at the 5% level. [2]

8 The heating quality of the coal in a sample of 50 sacks is measured in suitable units. The data are summarised below.

Heating quality (x)	$9.1 \leq x \leq 9.3$	$9.3 < x \leq 9.5$	$9.5 < x \leq 9.7$	$9.7 < x \leq 9.9$	$9.9 < x \leq 10.1$
Frequency	5	7	15	16	7

(i) Draw a cumulative frequency diagram to illustrate these data. [5]

(ii) Use the diagram to estimate the median and interquartile range of the data. [3]

(iii) Show that there are no outliers in the sample. [3]

(iv) Three of these 50 sacks are selected at random. Find the probability that

(A) in all three, the heating quality x is more than 9.5, [3]

(B) in at least two, the heating quality x is more than 9.5. [4]

Copyright Information

OCR is committed to seeking permission to reproduce all third-party content that it uses in its assessment materials. OCR has attempted to identify and contact all copyright holders whose work is used in this paper. To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced in the OCR Copyright Acknowledgements Booklet. This is produced for each series of examinations and is freely available to download from our public website (www.ocr.org.uk) after the live examination series. If OCR has unwittingly failed to correctly acknowledge or clear any third-party content in this assessment material, OCR will be happy to correct its mistake at the earliest possible opportunity. For queries or further information please contact the Copyright Team, First Floor, 9 Hills Road, Cambridge CB2 1GE.

OCR is part of the Cambridge Assessment Group; Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.

ADVANCED SUBSIDIARY GCE

MEI STATISTICS

Statistics 1 (Z1)

G241

PRINTED ANSWER BOOK

Candidates answer on this printed answer book.

OCR supplied materials:

- Question paper G241 (inserted)
- MEI Examination Formulae and Tables (MF2)

Other materials required:

- Scientific or graphical calculator

Thursday 26 May 2011
Morning

Duration: 1 hour 30 minutes



Candidate forename		Candidate surname	
--------------------	--	-------------------	--

Centre number						Candidate number				
---------------	--	--	--	--	--	------------------	--	--	--	--

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the printed answer book and the question paper.

- The question paper will be found in the centre of the printed answer book.
- Write your name, centre number and candidate number in the spaces provided on the printed answer book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the printed answer book.** Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

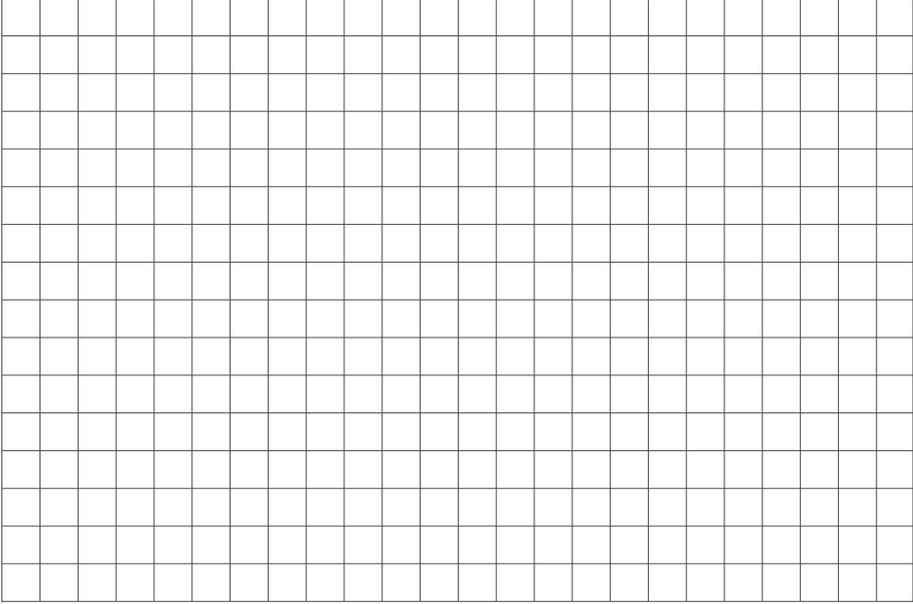
INFORMATION FOR CANDIDATES

This information is the same on the printed answer book and the question paper.

- The number of marks is given in brackets [] at the end of each question or part question on the question paper.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- The total number of marks for this paper is **72**.
- The printed answer book consists of **12** pages. The question paper consists of **4** pages. Any blank pages are indicated.

Section A (36 marks)

1 (i)	
1 (ii)	
1 (iii)	
2 (i)	
2 (ii)	

4 (i)	
4(ii)(A)	<hr/> <hr/> <hr/> <hr/> <hr/> <hr/>
4(ii)(B)	<hr/> <hr/> <hr/> <hr/> <hr/>
4 (iii)	<hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>

5 (i)	
5 (ii)	<hr/>
5 (iii)	<hr/>

6 (i)	
6 (ii)	

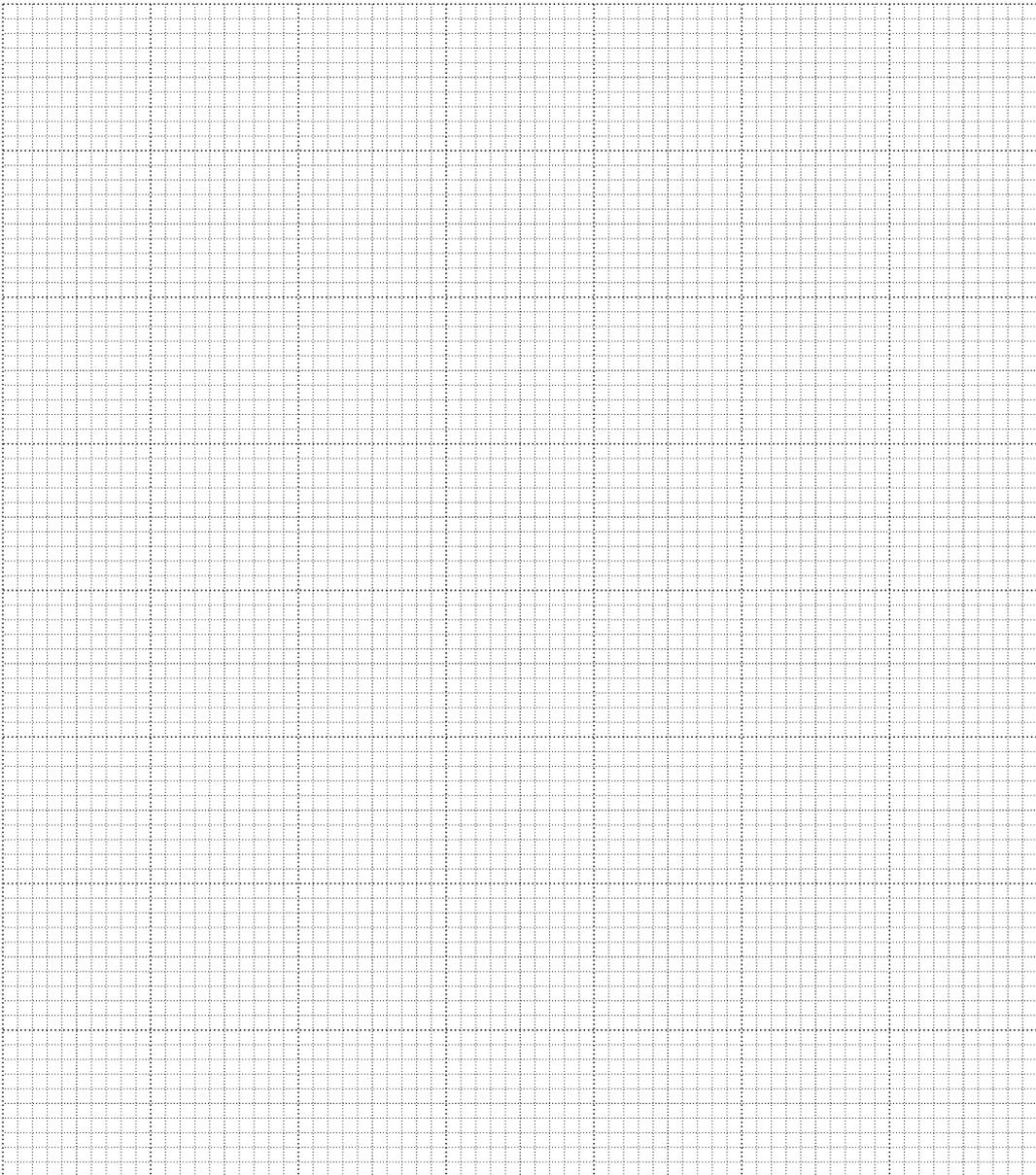
Section B (36 marks)

7(i)(A)	

7(i)(B)	

7 (ii)	
7 (ii)	

8 (i)



8(iv)(A)	
8(iv)(B)	

Statistics (MEI)

Advanced Subsidiary GCE

Unit G241: Statistics 1 (Z1)

Mark Scheme for June 2011

OCR (Oxford Cambridge and RSA) is a leading UK awarding body, providing a wide range of qualifications to meet the needs of pupils of all ages and abilities. OCR qualifications include AS/A Levels, Diplomas, GCSEs, OCR Nationals, Functional Skills, Key Skills, Entry Level qualifications, NVQs and vocational qualifications in areas such as IT, business, languages, teaching/training, administration and secretarial skills.

It is also responsible for developing new specifications to meet national requirements and the needs of students and teachers. OCR is a not-for-profit organisation; any surplus made is invested back into the establishment to help towards the development of qualifications and support which keep pace with the changing needs of today's society.

This mark scheme is published as an aid to teachers and students, to indicate the requirements of the examination. It shows the basis on which marks were awarded by Examiners. It does not indicate the details of the discussions which took place at an Examiners' meeting before marking commenced.

All Examiners are instructed that alternative correct answers and unexpected approaches in candidates' scripts must be given marks that fairly reflect the relevant knowledge and skills demonstrated.

Mark schemes should be read in conjunction with the published question papers and the Report on the Examination.

OCR will not enter into any discussion or correspondence in connection with this mark scheme.

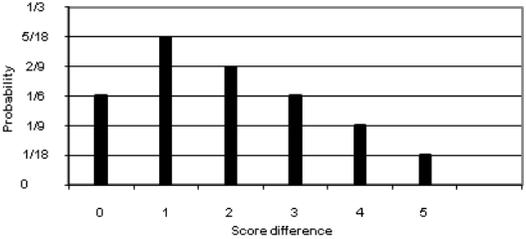
© OCR 2011

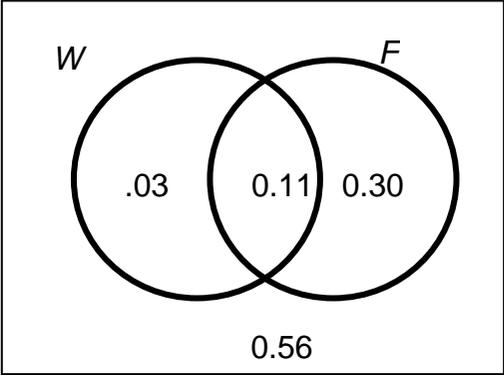
Any enquiries about publications should be addressed to:

OCR Publications
PO Box 5050
Annesley
NOTTINGHAM
NG15 0DL

Telephone: 0870 770 6622
Facsimile: 01223 552610
E-mail: publications@ocr.org.uk

SECTION A				
Q1 (i)	$1000 \times 0.013 = 13$ Or $0.2 \times 65 = 13$ Or $0.2 \times 5 \times 13 = 13$	M1 A1 M1 for 0.2×65	2	Allow with or without working For MR $1000 \times 0.13 = 130$ Allow M1A0 Allow M1A0 if extra terms added eg 1000×0.004 SC1 for $1000 \times 0.014 = 14$ For whole calculation
(ii)	Positive	B1	1	Allow +ve but NOT skewed to the right Do not allow 'positive correlation'
(iii)	Minimum value = 1500 Maximum value = 2500	B1 Without wrong working B1 Without wrong working	2	Exact answers only unless good explanation such as eg no road has length zero so min is eg 1501 SC1 for lower answer between 1499 and 1501 and upper between 2499 and 2501 Allow answer given as inequality
		TOTAL	5	
Q2 (i)	Either $P(\text{alphabetic order}) = \frac{1}{5} \times \frac{1}{4} \times \frac{1}{3} \times \frac{1}{2} \times \frac{1}{1} = \frac{1}{120}$ or $P(\text{alphabetic order}) = \frac{1}{5!} = \frac{1}{120} = 0.00833$	M1 for 5! or 120 or 5P_5 seen or product of correct fractions A1 CAO	2	Allow 0.0083 or 1/120 but not 0.008
(ii)	Either $P(\text{picks Austen and Bronte}) = \frac{2}{5} \times \frac{1}{4} = \frac{1}{10}$ or $P(\text{picks Austen and Bronte}) = \frac{1}{5} \times \frac{1}{4} \times 2 = \frac{1}{10}$ or $P(\text{picks Austen and Bronte}) = \frac{1}{\binom{5}{2}} = \frac{1}{10}$	M1 for denominators M1 for $2 \times$ <i>dep on correct denominators</i> A1 CAO Or M1 for $\binom{5}{2}$ or 10 M1 for $1/\binom{5}{2}$ A1 CAO	3	$1/5P_2$ scores M1 also 1/20 oe scores M1 even if followed by further incorrect working $\binom{5}{2}$ seen as part of a binomial expression gets M0M0A0
		TOTAL	5	

Q3 (i)	$P(X = 0) = 0.75^6 = 0.178$	M1 for 0.75^6 A1 CAO	2	Or from tables 0.1780 Or 729/4096 Allow 0.18 with working
(ii)	$E(X) = np = 50 \times 0.178 = 8.9$	M1 for product A1 FT	2	FT their answer to (i) providing it's a probability NB A0 if subsequently rounded
		TOTAL	4	
Q4 (i)		G1 labelled linear scales on both axes G1 heights	2	Accept r or x for horizontal label and p or better for vertical including probability distribution Visual check only Allow G1G0 for points rather than lines Bars must not be wider than gaps for second G1 Condone vertical scale 1, 2, 3, 4, 5 and Probability (\times) 1/18 as label BOD for height of $r = 0$ on vertical axis
(ii)	<p>(A) If $X = 1$, possible scores are (1,2), (2,3), (3,4), (4,5), (5,6) and (2,1), (3,2), (4,3), (5,4), (6,5)</p> <p>(All are equally likely) so probability = $\frac{10}{36} = \frac{5}{18}$</p> <p>(B) If $X = 0$, possible scores are (1,1), (2,2), (3,3), (4,4), (5,5), (6,6) so probability = $\frac{6}{36} = \frac{1}{6}$</p>	M1 A1 B1	2 1	Also M1 for a clear correct sample space seen with the ten 1's identified by means of circles or ticks oe soi. Must be convincing. No additional values such as 0,1 and 1,0 Do not allow 'just 10 ways you can have a difference of 1 so 10/36' or equivalent SC1 for possible scores are (1,2), (2,3), (3,4), (4,5), (5,6) so probability = $2 \times 5 \times 1/36$ with no explanation for $2 \times$ Also B1 for a clear correct sample space seen with the six 0's identified by means of circles or ticks oe soi. Must be convincing. No additional values. Allow both dice must be the same so probability = $6/36 = 1/6$. Allow $1 \times 1/6 = 1/6$ BOD
(iii)	Mean value of $X =$ $0 \times \frac{1}{6} + 1 \times \frac{5}{18} + 2 \times \frac{2}{9} + 3 \times \frac{1}{6} + 4 \times \frac{1}{9} + 5 \times \frac{1}{18} = 1 \frac{17}{18} = 1.94$	M1 for $\sum rp$ (at least 3 terms correct) A1 CAO	2	Or 35/18 Division by 6 or other spurious factor gets MAX M1A0
		TOTAL	7	

<p>Q5 (i)</p>		<p>G1 for two labelled intersecting circles</p> <p>G1 for at least 2 correct probabilities.</p> <p>G1 for remaining correct probabilities</p>	<p>3</p>	<p>Allow labels such as P(W) and P(F)</p> <p>Allow other sensible shapes in place of circles</p>
<p>(ii)</p>	<p>$P(W) \times P(F) = 0.14 \times 0.41 = 0.0574 \neq P(W \cap F) = 0.11$ So not independent.</p>	<p>M1 for 0.41×0.14</p> <p>A1 Condone dependent</p> <p>Must have full method www</p> <p>Must have either $P(W \cap F)$ or 0.11</p>	<p>2</p>	<p>Answer of 0.574 gets Max M1A0</p> <p>Omission of 0.0574 gets M1A0 Max</p> <p>Or: $P(W F) = 0.11/0.41 = 0.268 \neq P(W) (= 0.14)$ M1 for full working</p> <p>$P(F W) = 0.11/0.14 = 0.786 \neq P(F) (= 0.41)$ M1 for full working</p> <p>No marks without correct working</p>
<p>(iii)</p>	$P(W F) = \frac{P(W \cap F)}{P(F)} = \frac{0.11}{0.41} = \frac{11}{41} = 0.268$ <p>This is the probability that a randomly selected respondent works (part time), given that the respondent is female.</p>	<p>M1 for correct fraction</p> <p>A1</p> <p>E1</p> <p>For E1 must be in context – not just talking about events F and W</p>	<p>3</p>	<p>Allow 0.27 with working</p> <p>Allow 11/41 as final answer</p> <p>Condone ‘if’ or ‘when’ for ‘given that’ but not the words ‘and’ or ‘because’ or ‘due to’ for E1.</p> <p>E1 (independent of M1): the order/structure must be correct i.e. no reverse statement</p> <p>Allow ‘The probability that a randomly selected female respondent works part time’ oe</p>
<p style="text-align: right;">TOTAL</p>			<p>8</p>	

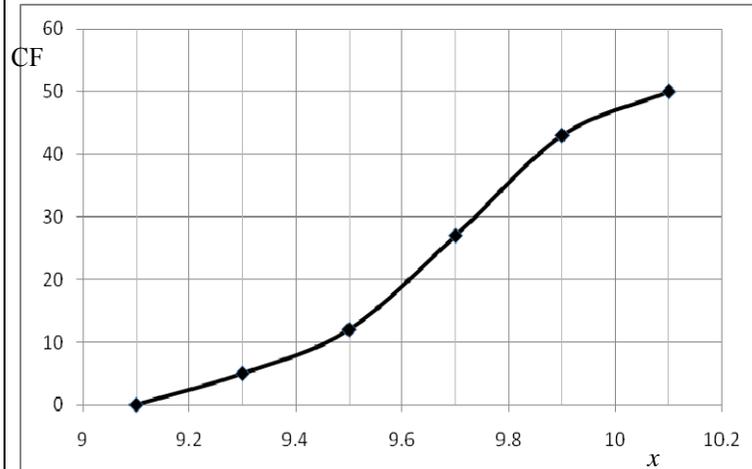
<p>Q6 (i)</p>	$\text{Mean} = \frac{1 \times 10 + 2 \times 40 + 3 \times 15 + 4 \times 5}{70} = \frac{155}{70} = 2.214$ $S_{xx} =$ $1^2 \times 10 + 2^2 \times 40 + 3^2 \times 15 + 4^2 \times 5 - \frac{155^2}{70} = 385 - 343.21 = 41.79$ $s = \sqrt{\frac{41.79}{69}} = 0.778$	<p>M1 A1 CAO</p> <p>M1 for Σfx^2 s.o.i.</p> <p>M1 for attempt at S_{xx} Dep on first M1</p> <p>A1 CAO If 0.778 or better seen ignore previous incorrect working (calculator answer) Allow final answer to 2 sig fig (www)</p>	<p>5</p>	<p>For M1 allow sight of at least 3 double pairs seen from $1 \times 10 + 2 \times 40 + 3 \times 15 + 4 \times 5$ with divisor 70. Allow answer of 155/70 or 2.2 or 2.21 or 31/14 oe For 155/70 = eg 2.3 , allow A1 isw</p> <p>M1 for $1^2 \times 10 + 2^2 \times 40 + 3^2 \times 15 + 4^2 \times 5$ with at least three correct terms Using exact mean leads to $S_{xx} = 41.79$, $s=0.778$, Using mean 2.214 leads to $S_{xx} = 41.87$, $s=0.779$, Using mean 2.21 leads to $S_{xx} = 43.11$ and $s = 0.790$ Using mean 2.2 leads to $S_{xx} = 46.2$ and $s = 0.818$ Using mean 2 leads to $S_{xx} = 105$ and $s = 1.233$ All the above get M1M1A1 except the last one which gets M1M1A0 RMSD(divisor n rather than $n - 1$) = $\sqrt{(41.79/70)} = 0.772$ gets M1M1A0 Alternative method, award M1 for at least 3 terms of and second M1 for all 4 terms of $(1 - 2.214)^2 \times 10 + (2 - 2.214)^2 \times 40 + (3 - 2.214)^2 \times 15 + (4 - 2.214)^2 \times 5 (= 41.79)$ NB Allow full credit for correct answers without working (calculator used)</p>
<p>(ii)</p>	<p>Mean would decrease</p> <p>Standard deviation would increase</p>	<p>B1</p> <p>B1</p>	<p>2</p>	<p>Do not accept increase/decrease seen on their own – must be linked to mean and SD. Allow eg ‘It would skew the mean towards zero’ And eg ‘It would stretch the SD’ SC1 for justified argument that standard deviation might either increase or decrease according to number with no eggs ($n \leq 496$ increase, $n \geq 497$ decrease)</p>
		<p>TOTAL</p>	<p>7</p>	

(ii)	<p>Let $X \sim B(n, p)$ Let $p =$ probability of a 'no-show' (for population) $H_0: p = 0.15$ $H_1: p < 0.15$</p> <p>H_1 has this form because the hospital management hopes to reduce the proportion of no-shows.</p>	<p>B1 for definition of p B1 for H_0 B1 for H_1</p> <p>E1 Allow correct answer even if H_1 wrong</p>	<p>4</p> <p>Allow $p = P(\text{no-show})$ for B1 Definition of p must include word probability (or chance or proportion or percentage or likelihood but NOT possibility). Preferably as a separate comment. However can be at end of H_0 as long as it is a clear definition '$p =$ the probability of no-show, NOT just a sentence 'probability is 0.15' $H_0 : p(\text{no-show}) = 0.15, H_1 : p(\text{no-show}) < 0.15$ gets B0B1B1 Allow $p=15\%$, allow θ or π and ρ but not x. However allow any single symbol <u>if defined</u> Allow $H_0 = p=0.15,$ Do not allow $H_0 : P(X=x) = 0.15, H_1 : P(X=x) < 0.15$ Do not allow $H_0: =0.15, =15\%, P(0.15), p(0.15), p(x)=0.15, x=0.15$ (unless x correctly defined as a probability) Do not allow $H_1:p \leq 0.15,$ Do not allow H_0 and H_1 reversed for B marks but can still get E1 Allow NH and AH in place of H_0 and H_1 For hypotheses given in words allow Maximum B0B1B1E1 Hypotheses in words must include probability (or chance or proportion or percentage) and the figure 0.15 oe.</p>
(iii)	<p>$P(X \leq 1) = 0.1756 > 5\%$</p> <p>So not enough evidence to reject H_0. Not significant. Conclude that there is not enough evidence to indicate that the proportion of no-shows has decreased.</p> <p>Note: use of critical region method scores M1 for region $\{0\}$ M1 for 1 does not lie in critical region, then A1 E1 as per scheme</p>	<p>M1 for probability seen, but not in calculation for point probability M1 dep for comparison A1</p> <p>E1 dep for conclusion in context.</p>	<p>4</p> <p>Zero for use of point prob - $P(X = 1) = 0.1368$ Do <u>NOT</u> FT wrong H_1 Allow accept H_0, or reject H_1 Full marks only available if 'not enough evidence to...' oe mentioned somewhere Do not allow 'enough evidence to reject H_1' for final mark but can still get 3/4 Upper end comparison: $1 - 0.1756 = 0.8244 < 95\%$ gets M2 then A1E1 as per scheme</p>

				<p><u>Line diagram method</u> M1 for squiggly line between 0 and 1 with arrow pointing to left, M1 0.0388 seen on diagram from squiggly line or from 0, A1E1 for correct conclusion</p> <p><u>Bar chart method</u> M1 for line clearly on boundary between 0 and 1 and arrow pointing to left, M1 0.0388 seen on diagram from boundary line or from 0, A1E1 for correct conclusion</p>
(iv)	<p>$6 < 8$</p> <p>So there is sufficient evidence to reject H_0 Conclude that there is enough evidence to indicate that the proportion of no-shows appears to have decreased.</p>	<p>M1 for comparison seen A1 E1 for conclusion in context</p>	3	<p>Allow '6 lies in the CR' Do NOT insist on 'not enough evidence' here Do not FT wrong $H_1: p > 0.15$ but may get M1 In part (iv) ignore any interchanged H_0 and H_1 seen in part (ii)</p>
(v)	<p>For $n \leq 18$, $P(X \leq 0) > 0.05$ so the critical region is empty.</p>	<p>E1 for $P(X \leq 0) > 0.05$</p> <p>E1 indep for critical region is empty</p>	2	<p>E1 also for sight of 0.0536 Condone $P(X = 0) > 0.05$ or all probabilities or values, (but not outcomes) in table (for $n \leq 18$) > 0.05</p> <p>Or 'There is no critical region' For second E1 accept 'H_0 would always be accepted'</p> <p>Do NOT FT wrong H_1 Use professional judgement - allow other convincing answers</p>
		TOTAL	18	

Q8
(i)

Upper Bound	9.1	9.3	9.5	9.7	9.9	10.1
Cumulative frequency	0	5	12	27	43	50



B1 for cumulative frequencies

G1 for scales

G1 for labels

G1 for points
(Provided plotted at correct UCB positions)

G1 for joining points

All G's dep on attempt at cumulative frequency but not cumulative fx's or other spurious values.

5

May be implied from graph. Condone omission of 0 at this stage.

Linear horizontal scale.

Linear vertical scale: 0 to 50

(no inequality scales - Not even <9.1 , <9.3 , <9.5 ...)Heating quality or x and Cumulative frequency or just CF or similar but not just frequency or fd nor cumulative fd

Plotted as (UCB, their cf). Ignore (9.1,0) at this stage. No midpoint or LCB plots.

Plotted within $\frac{1}{2}$ small squareFor joining all of 'their points' (line or smooth curve)
AND now including (9.1,0) dep on previous G1

Mid point or LCB plots may score first three marks

Can get up to 3/5 for cum freq bars

Allow full credit if axes reversed correctly

Lines of best fit could attract max 4 out of 5.

(ii)	<p>Median = 9.67</p> <p>Q1 = 9.51 Q3 = 9.83 Inter-quartile range = $9.83 - 9.51 = 0.32$</p>	<p>B1 FT Allow answers between 9.66 and 9.68 without checking curve. Otherwise check curve.</p> <p>B1 FT for Q3 or Q1 B1 FT for IQR providing both Q1 and Q3 are correct Allow answers between 9.50 and 9.52 and between 9.82 and 9.84 without checking curve. Otherwise check curve.</p>	3	<p>Based on 25th to 26th value on a cumulative frequency graph ft their mid-point plot (not LCB's) approx 9.57 for m.p. plot Allow 9.56 to 9.58 without checking B0 for interpolation</p> <p>Based on 12th to 13th and 37th to 38th values on a cumulative frequency graph ft their mid -point plot (not LCB's) approx Q1 = 9.42; Q3 = 9.73 Allow 9.41 to 9.43 and 9.72 to 9.74 without checking B0 for interpolation Allow correct IQR from graph if quartiles not stated</p> <p>Lines of best fit: B0 B0 B0 here.</p>
(iii)	<p>Lower limit $9.51 - 1.5 \times 0.32 = 9.03$ Upper limit $9.83 + 1.5 \times 0.32 = 10.31$ Thus there are no outliers in the sample.</p>	<p>B1 FT their Q₁, IQR B1 FT their Q₃, IQR E1 NB E mark dep on both B marks</p>	3	<p>Any use of <u>median</u> ± 1.5 IQR scores B0 B0 E0</p> <p>If FT leads to limits above 9.1 or below 10.1 then E0 No marks for ± 2 or 3 IQR</p> <p>In this part FT their values from (ii) if sensibly obtained (eg from LCB plot) or lines of best fit, but not from location ie 12.5, 37.5 or cumulative fx's or similar. For use of mean $\pm 2s$, Mean = 9.652, s = 0.235, Limits 9.182, 10.122 gets M1 for correct lower limit, M1 for correct upper limit, zero otherwise, but E0 since there could be outliers using this definition</p>

<p>(iv)</p>	<p>(A) $P(\text{All 3 more than 9.5}) = \frac{38}{50} \times \frac{37}{49} \times \frac{36}{48} = 0.4304$ $(=50616/117600 = 2109/4900)$</p> <p>(B) $P(\text{At least 2 more than 9.5}) = 3 \times \frac{38}{50} \times \frac{37}{49} \times \frac{12}{48} + 0.4304$ $= 3 \times 0.1435 + 0.4304$ $= 0.4304 + 0.4304$ $= 0.861$ $(=101232/117600 = 4218/4900 = 2109/2450)$</p> <p>OR</p> <p>$P(\text{At least 2 more than 9.5}) = 1 - (P(0) + P(1))$ $= 1 - \left[\left(\frac{12}{50} \times \frac{11}{49} \times \frac{10}{48} \right) + \left(3 \times \frac{12}{50} \times \frac{11}{49} \times \frac{38}{48} \right) \right]$ $= 1 - [0.01122 + 0.12796] = 1 - 0.13918 = 0.861$</p>	<p>M1 for $38/50 \times$ (triple product) M1 for product of remaining fractions A1 CAO</p> <p>M1 for product of 3 correct fractions seen M1 for $3 \times$ a sensible triple or sum of 3 sensible triples M1 indep for $+ 0.4304$ FT (providing it is a probability) A1 CAO</p> <p>M1 for $12/50 \times 11/49 \times 38/48$ M1 for $3 \times$ a sensible triple or sum of 3 sensible triples M1 dep on both previous M1's for $1 - [0.01122 + 0.12796]$ A1 CAO</p>	<p>3</p> <p>$(38/50)^3$ which gives answer 0.4389 scores M1M0A0 so watch for this. M0M0A0 for binomial probability including 0.76^{100} but ${}^3C_0 \times 0.24^0 \times 0.76^3$ still scores M1 $(k/50)^3$ for values of k other than 38 scores M0M0A0 $\frac{k}{50} \times \frac{(k-1)}{49} \times \frac{(k-2)}{48}$ for values of k other than 38 scores M1M0A0 Correct working but then multiplied or divided by some factor scores M1M0A0 Accept 0.43 with working and 0.430 without working Or $\binom{38}{3} / \binom{50}{3} = 2109/4900 = 0.4304$</p> <p>Allow unsimplified fraction as final answer 50616/117600</p> <p>4</p> <p>Or $\binom{38}{2} \binom{12}{1} / \binom{50}{3} = 0.4304$ gets first two M1M1's</p> <p>SC1 for $3 \times \frac{38}{50} \times \frac{38}{50} \times \frac{12}{50}$ or other sensible triple and SC2 if this + their 0.4304 (= 0.8549) Allow 0.86 or 2109/2450 or 4218/4900, but only M3A0 for other unsimplified fractions</p> <p>Use of 1 – method ‘with replacement’ SC1 for $3 \times \frac{12}{50} \times \frac{12}{50} \times \frac{38}{50}$ SC2 for whole of $1 - 3 \times \frac{12}{50} \times \frac{12}{50} \times \frac{38}{50} + \frac{12}{50} \times \frac{12}{50} \times \frac{12}{50}$ $(= 1 - (0.1313 + 0.0138) = 1 - 0.1451 = 0.8549)$</p>
		TOTAL	18

NOTE RE OVER-SPECIFICATION OF ANSWERS

If answers are grossly over-specified (see instruction 8), deduct the final answer mark in every case. Probabilities should also be rounded to a sensible degree of accuracy. In general final non probability answers should not be given to more than 4 significant figures. Allow probabilities given to 5 sig fig. In general accept answers which are correct to 3 significant figures when given to 4 or 5 significant figures.

OCR (Oxford Cambridge and RSA Examinations)
1 Hills Road
Cambridge
CB1 2EU

OCR Customer Contact Centre

14 – 19 Qualifications (General)

Telephone: 01223 553998

Facsimile: 01223 552627

Email: general.qualifications@ocr.org.uk

www.ocr.org.uk

For staff training purposes and as part of our quality assurance programme your call may be recorded or monitored

Oxford Cambridge and RSA Examinations
is a Company Limited by Guarantee
Registered in England
Registered Office; 1 Hills Road, Cambridge, CB1 2EU
Registered Company Number: 3484466
OCR is an exempt Charity



OCR (Oxford Cambridge and RSA Examinations)
Head office
Telephone: 01223 552552
Facsimile: 01223 552553

G241 Statistics 1

General Comments

The level of difficulty of the paper appeared to be appropriate for the candidates and there was no evidence of candidates being unable to complete the paper in the allocated time. On the whole, candidates appeared to be well prepared for the paper with relatively few unable to gain many marks. Most candidates supported their numerical answers with appropriate explanations and working. Most candidates had adequate space in the answer booklet without having to use additional sheets. Once again a few candidates over-specified some of their answers, despite recent Examiner's reports warning against this. Please see the comments about this in the Chief Examiner's report.

Many candidates struggled with the hypothesis testing in question 7 and there was still quite a lot of use of using point probabilities in their arguments. Candidates found the unfamiliar style of parts (iv) and (v) particularly challenging. Question 6 provided a valuable source of marks for most candidates, but a rather surprisingly large number of candidates did not score highly on the first three parts of question 8. However question 8(iv)A and B proved too difficult almost all candidates. There were few correct answers to question 1(iii), suggesting that 'midrange' is a measure of average which receives little attention. In question 4(ii) many candidates failed to understand what is required for a probability argument, despite this phrase having been used in past papers.

Comments on Individual Questions

- 1) In part (i) candidates were able to make a successful start to the paper by realising that the frequency was equal to the frequency density \times class width. Most gained the expected answer of 13 but occasionally the examiners saw 14 (due to a misread of the vertical scale) or 130 (due to not being able to multiply by 1000 correctly).

In part (ii), the vast majority of candidates recognised that the distribution was positively skewed but some still insist on using the unacceptable terms of 'right skew' or 'symmetrical skew'.

Part (iii) defeated many candidates. Whilst many understood the idea of the mid-range, few were able to apply it in the context of the question. Very few appreciated that the maximum mid-range could only be found by averaging the highest value in the last class with highest value in the first class to give $(4000 + 1000)/2 = 2500$. Many wrote $(4000 + 0)/2 = 2000$ as their response here. Similarly, the minimum mid-range could only be found by averaging the lowest value in the last class with lowest value of the first class to yield $(3000 + 0)/2 = 1500$.

- 2) Part (i) was successfully answered by most candidates.

Many candidates gained only 1 mark out of 3 in part (ii), giving an answer of $(1/5) \times (1/4) = 1/20$, failing to realise that Austen could be picked first followed by Brontë, or vice versa, hence requiring their answer to be multiplied by 2.

- 3) Part (i) was usually answered correctly, the standard correct response being $(0.75)^6 = 0.178$. However a surprising number of candidates thought that the probability was just 0.75.

In part (ii) the expectation was usually found correctly by multiplication of 0.178 by 50 but occasionally some used 300 instead. Candidates should be reminded not to round their final answer in an expectation calculation. There were too many cases of 8.9 being rounded to 9 which lost the final mark. Candidates who had got the wrong answer to part (i) were allowed a full follow through in part (ii).

- 4) Part (i) was generally well answered. Candidates who used fractions (in multiples of $1/18$) on their probability scale usually scored full marks. Candidates who used decimals made the question more difficult, which often led to inaccurate heights and a loss of one mark. Some candidates lost the first mark due to failure to label both axes.

In part (ii) parts A and B, a significant number of candidates failed to understand the questions by thinking that they had to use the probability distribution given, subtracting the other probabilities from 1, but there was no actual probability argument evident. Those who did begin to identify combinations with a difference of one often did not recognize that the order mattered and then claimed that there were only 18 possible outcomes in order to make the numbers fit the given answer. Most candidates who were successful compiled a two way table of all of the possibilities. A correct numerical method which lacked the essential explanation of where it had come from was fairly commonly seen.

In part (iii), a large majority used a correct method, but a surprising number did not realise that expectation and mean are interchangeable in this context and consequently they divided by 6 or some other number.

- 5) Many candidates got full marks for their Venn diagram in part (i). A minority failed to subtract 0.11 from 0.41 and 0.14 but even these usually produced two intersecting circles labelled correctly to get the first mark. A few candidates did not work out the probability for the fourth region (0.56).

Part (ii) was answered fairly well and showed that many candidates know how to test for independence, although surprisingly candidates often used the probabilities from their Venn diagram rather than those from the question. Some candidates failed to evaluate 0.14×0.41 and consequently lost the accuracy mark. A minority of candidates, having correctly completed the working, then got the conclusion the wrong way round. A small number of candidates used a conditional probability method, not always correctly.

Part (iii) was also answered fairly well, but again a significant number of candidates used the wrong figures from their Venn diagram. An impressively large proportion of candidates did get the correct explanation of what this probability represented but several missed out this mark because they did not explain the conditional probability in the context of the question.

- 6) In part (i) most candidates found the mean correctly and whilst decimal answers were frequently over-specified candidates gained the mark for giving the answer in fractional form, even if subsequently over-specified. Most candidates made a good attempt at the standard deviation; the main errors were the usual ones - calculating the rmsd instead of the standard deviation, incorrect squaring when calculating $\sum fx^2$ or using $n=4$ instead of $n=70$. The number of candidates who simply used the statistical functions on their calculator was fairly small, despite this being the easiest way to do the question.

In part (ii) many candidates found the mean correctly but thought that the standard deviation remained unchanged. Those candidates who understood that standard deviation is a measure of deviation were usually able to see clearly that the deviations would (tend to) be increased. Those who tried to reason their way through the formula usually came to the wrong conclusion. A very few very able candidates correctly said that if the number of gulls laying no eggs was very large (over around 500) then the standard deviation would decrease.

- 7) Part (i)A was generally answered correctly, although when using the binomial formula, a few candidates forgot to round off sensibly.

Part (i)B was found to be slightly more difficult. Most candidates used tables but some went wrong by calculating $1 - P(X < 1)$ or $1 - P(X \leq 2)$. A reasonable number of candidates first found $P(X = 0)$ and then usually went on to finish off the question correctly.

In part (ii) most candidates correctly stated their hypotheses in terms of p , but then often lost the available mark for defining p . Most were able to give an explanation of the reason for the nature of the alternative hypothesis.

In parts (iii), (iv) and (v), too many candidates forgot to state their conclusions in context. This is required in every exam and so teachers should be careful to instruct their students to do this.

Part (iii) was a relatively easy hypothesis test, since it was a lower tail test. However, many candidates (almost half of the candidature) used point probabilities and thus gained no marks. Of those who gained some credit, most either got full marks, or lost the final mark for conclusion in context.

In part (iv) some candidates wasted a lot of time for these 3 marks, testing out trial distributions for large n . Candidates should appreciate that, with only 3 marks at stake, there must be a more tractable solution. In fact all that was required was a comparison of the test statistic with the critical value, followed by a conclusion in context. In fact, only one third of the candidature gained any marks at all.

Part (v) was expecting candidates to give a valid reason for the critical region being empty. A number of fully correct solutions were seen, and the question was generously marked, so that candidates who got some way to an explanation gained one mark.

- 8) In part (i) most candidates scored at least three marks. Many did not explicitly write down their calculations for the cumulative frequency, preferring to plot their points directly on the graph, but still gained the mark. Scales were usually correct and sensible but some candidates used a linear horizontal scale starting from zero, which made a very cramped graph. Labels were often forgotten altogether and the vertical scale was often seen as just 'frequency', losing the mark. Points were usually plotted correctly at the right height but far too often were plotted at the mid-points i.e. 9.2, 9.4 etc. losing the final 2 marks. Only a few candidates used the lower class boundaries. Many candidates lost the final mark by not joining (9.3, 5) to (9.1, 0). Cumulative frequency bars were sometimes seen as were lines of best fit. Occasionally no attempt was made at a cumulative frequency graph at all, with some candidates just plotting frequency against midpoints or attempting to find frequency density.

Part (ii) was generally well answered often from a follow through from a 'sensible' graph. Some of the scales used in part (i) meant that it was very difficult to read the figures if they fell outside the allowed ranges. The 12th value was often used instead of the 12.5th value, perhaps because it was easier to read as there was a point plotted there. A few candidates failed to calculate the IQR even though both quartiles were found.

There were many correct answers to part (iii) (or correct fit answers) but many candidates tried to use the median or twice the IQR. A few candidates reverted to calculating and using the mean and standard deviation, gaining up to 2 marks out of 3 although they could not gain the last mark with this method because outliers could exist.

In part (iv)(A) only the better candidates obtained the correct answer. Many used $(38/50)^3$ scoring one mark only. Others candidates had more complicated incorrect versions of binomial probabilities. Occasionally the numerators decreased but the denominators didn't. Some candidates did not find the correct value of 38 from the table.

Part (iv)(B) was found very challenging and only about ten percent of candidates gained full credit. Many candidates scored one mark for adding their answer from part (A), but otherwise a common incorrect answer of 0.8549 was often seen, which scored SC2. Some candidates thought that they had to multiply only two probabilities when finding the probability of two being more than 9.5. Many candidates did not realise that there were three different ways of getting two more than 9.5. Those candidates who drew tree diagrams fared better here, and in realising that the probabilities diminished. Those candidates attempting $1 - (P(0) + P(1))$ were on the whole not as successful, sometimes not including both probabilities or failing to include the factor of 3.

GCE Science (AS only)								
		Max Mark	a	b	c	d	e	u
G641/01 Remote Sensing and the Natural Environment	Raw	60	45	40	35	31	27	0
	UMS	90	72	63	54	45	36	0
G642/01 Science and Human Activity	Raw	100	70	62	54	46	38	0
	UMS	150	120	105	90	75	60	0
G643/01 Practical Skills in Science	Raw	40	33	30	27	25	23	0
	UMS	60	48	42	36	30	24	0

GCE Sociology								
		Max Mark	a	b	c	d	e	u
G671/01 Exploring Socialisation, Culture and Identity	Raw	100	73	66	59	53	47	0
	UMS	100	80	70	60	50	40	0
G672/01 Topics in Socialisation, Culture and Identity	Raw	100	71	63	55	47	39	0
	UMS	100	80	70	60	50	40	0
G673/01 Power and Control (A2)	Raw	100	70	62	54	47	40	0
	UMS	100	80	70	60	50	40	0
G674/01 Exploring Social Inequality and Difference (A2)	Raw	100	70	62	54	47	40	0
	UMS	100	80	70	60	50	40	0

GCE Spanish								
		Max Mark	a	b	c	d	e	u
F721/01 Spanish: Speaking (AS): Externally Marked (OCR Repository)	Raw	60	46	41	36	31	26	0
F721/02 Spanish: Speaking (AS): Externally Marked (CD)	Raw	60	46	41	36	31	26	0
F721/03 Spanish: Speaking (AS): Visiting Examiner	Raw	60	46	41	36	31	26	0
F721 Spanish: Speaking	UMS	60	48	42	36	30	24	0
F722/01 Spanish: Listening, Reading and Writing 1	Raw	140	106	94	82	71	60	0
	UMS	140	112	98	84	70	56	0
F723/01 Spanish: Speaking (A2): Externally Marked (OCR Repository)	Raw	60	47	41	35	30	25	0
F723/02 Spanish: Speaking (A2): Externally Marked (CD)	Raw	60	47	41	35	30	25	0
F723/03 Spanish: Speaking (A2): Visiting Examiner	Raw	60	47	41	35	30	25	0
F723 Spanish: Speaking	UMS	60	48	42	36	30	24	0
F724/01 Spanish: Listening, Reading and Writing 2 (A2)	Raw	140	101	89	77	65	54	0
	UMS	140	112	98	84	70	56	0

GCE Statistics (MEI)								
		Max Mark	a	b	c	d	e	u
G241/01 (Z1) Statistics 1	Raw	72	53	45	38	31	24	0
	UMS	100	80	70	60	50	40	0
G242/01 (Z2) Statistics 2	Raw	72	55	48	41	34	27	0
	UMS	100	80	70	60	50	40	0
G243/01 (Z3) Statistics 3	Raw	72	56	48	41	34	27	0
	UMS	100	80	70	60	50	40	0