OCR
RECOGNISING ACHIEVEMENT

Monday 23 January 2012 – Morning

AS GCE MATHEMATICS (MEI)

4766 Statistics 1

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4766
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

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. HB 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 recycled. Please contact OCR Copyright should you wish to re-use this document.

Section A (36 marks)

1 The mean daily maximum temperatures at a research station over a 12-month period, measured to the nearest degree Celsius, are given below.

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
8	15	25	29	31	31	34	36	34	26	15	8

(i) Construct a sorted stem and leaf diagram to represent these data, taking stem values of 0, 10, ... **[4]**

(ii) Write down the median of these data. **[1]**

(iii) The mean of these data is 24.3. Would the mean or the median be a better measure of central tendency of the data? Briefly explain your answer. **[2]**

2 The hourly wages, £x, of a random sample of 60 employees working for a company are summarised as follows.

$$n = 60 \qquad \Sigma x = 759.00 \qquad \Sigma x^2 = 11\,736.59$$

(i) Calculate the mean and standard deviation of x. **[3]**

(ii) The workers are offered a wage increase of 2%. Use your answers to part **(i)** to deduce the new mean and standard deviation of the hourly wages after this increase. **[2]**

(iii) As an alternative the workers are offered a wage increase of 25p per hour. Write down the new mean and standard deviation of the hourly wages after this 25p increase. **[2]**

3 Jimmy and Alan are playing a tennis match against each other. The winner of the match is the first player to win three sets. Jimmy won the first set and Alan won the second set. For each of the remaining sets, the probability that Jimmy wins a set is

- 0.7 if he won the previous set,
- 0.4 if Alan won the previous set.

It is not possible to draw a set.

(i) Draw a probability tree diagram to illustrate the possible outcomes for each of the remaining sets. **[3]**

(ii) Find the probability that Alan wins the match. **[3]**

(iii) Find the probability that the match ends after exactly four sets have been played. **[2]**

4 In a food survey, a large number of people are asked whether they like tomato soup, mushroom soup, both or neither. One of these people is selected at random.

- T is the event that this person likes tomato soup.
- M is the event that this person likes mushroom soup.

You are given that $P(T) = 0.55$, $P(M) = 0.33$ and $P(T \mid M) = 0.80$.

(i) Use this information to show that the events T and M are not independent. **[1]**

(ii) Find $P(T \cap M)$. **[2]**

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

5 A couple plan to have at least one child of each sex, after which they will have no more children. However, if they have four children of one sex, they will have no more children. You should assume that each child is equally likely to be of either sex, and that the sexes of the children are independent. The random variable X represents the total number of girls the couple have.

(i) Show that $P(X = 1) = \frac{11}{16}$. **[3]**

The table shows the probability distribution of X.

r	0	1	2	3	4
$P(X = r)$	$\frac{1}{16}$	$\frac{11}{16}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$

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

Section B (36 marks)

6 It is known that 25% of students in a particular city are smokers. A random sample of 20 of the students is selected.

(i) (A) Find the probability that there are exactly 4 smokers in the sample. **[3]**

(B) Find the probability that there are at least 3 but no more than 6 smokers in the sample. **[3]**

(C) Write down the expected number of smokers in the sample. **[1]**

A new health education programme is introduced. This programme aims to reduce the percentage of students in this city who are smokers. After the programme has been running for a year, it is decided to carry out a hypothesis test to assess the effectiveness of the programme. A random sample of 20 students is selected.

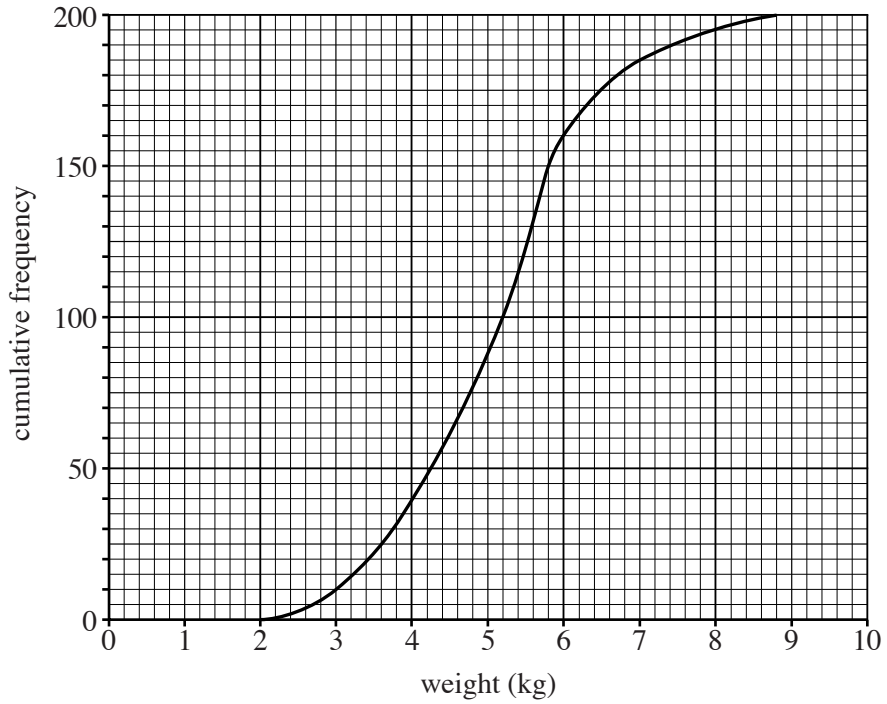
(ii) (A) Write down suitable null and alternative hypotheses for the test. **[3]**

(B) Explain why the alternative hypothesis has the form that it does. **[1]**

(iii) Find the critical region for the test at the 5% level, showing all of your calculations. **[4]**

(iv) In fact there are 3 smokers in the sample. Complete the test, stating your conclusion clearly. **[2]**

7 The birth weights of 200 lambs from crossbred sheep are illustrated by the cumulative frequency diagram below.

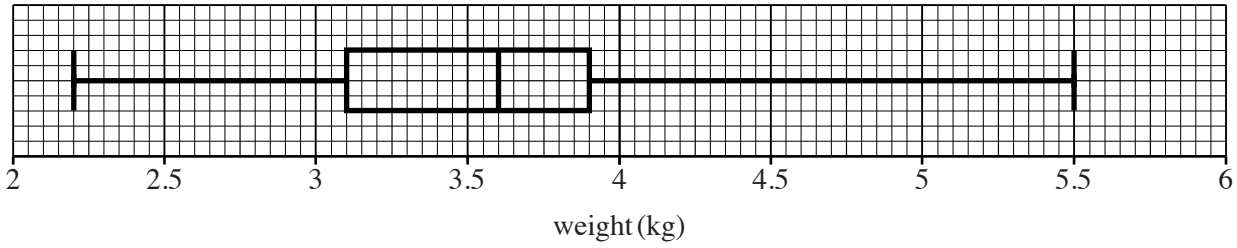


(i) Estimate the percentage of lambs with birth weight over 6 kg. **[2]**

(ii) Estimate the median and interquartile range of the data. **[3]**

(iii) Use your answers to part **(ii)** to show that there are very few, if any, outliers. Comment briefly on whether any outliers should be disregarded in analysing these data. **[4]**

The box and whisker plot shows the birth weights of 100 lambs from Welsh Mountain sheep.



(iv) Use appropriate measures to compare briefly the central tendencies and variations of the weights of the two types of lamb. **[4]**

(v) The weight of the largest Welsh Mountain lamb was originally recorded as 6.5 kg, but then corrected. If this error had not been corrected, how would this have affected your answers to part **(iv)**? Briefly explain your answer. **[2]**

(vi) One lamb of each type is selected at random. Estimate the probability that the birth weight of both lambs is at least 3.9 kg. **[4]**

OCR
RECOGNISING ACHIEVEMENT

Monday 23 January 2012 – Morning

AS GCE MATHEMATICS (MEI)

4766 Statistics 1

PRINTED ANSWER BOOK

Candidates answer on this Printed Answer Book.

OCR supplied materials:
- Question Paper 4766 (inserted)
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

4732160112

4766

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. HB 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)	
2 (iii)	

3 (i)	
3 (ii)	
3 (iii)	

4 (i)	
4 (ii)	
4 (iii)	

5 (i)

5 (ii)

Section B (36 marks)

6 (i) (*A*)	
6 (i) (*B*)	
6 (i) (*C*)	

6 (ii) (*A*)	
6 (ii) (*B*)	
6 (iii)	

6 (iv)	

7 (i)	
7 (ii)	
7 (iii)	

7 (iv)

7 (v)

7 (vi)	

OCR
RECOGNISING ACHIEVEMENT

GCE

Mathematics (MEI)

Advanced Subsidiary GCE

Unit **4766:** Statistics 1

Mark Scheme for January 2012

Oxford Cambridge and RSA Examinations

OCR (Oxford Cambridge and RSA) is a leading UK awarding body, providing a wide range of qualifications to meet the needs of candidates 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.

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

Annotations

Annotation in scoris	**Meaning**
✓and ✗	
BOD	Benefit of doubt
FT	Follow through
ISW	Ignore subsequent working
M0, M1	Method mark awarded 0, 1
A0, A1	Accuracy mark awarded 0, 1
B0, B1	Independent mark awarded 0, 1
SC	Special case
^	Omission sign
MR	Misread
Highlighting	
Other abbreviations in mark scheme	**Meaning**
E1	Mark for explaining
U1	Mark for correct units
G1	Mark for a correct feature on a graph
M1 dep*	Method mark dependent on a previous mark, indicated by *
cao	Correct answer only
oe	Or equivalent
rot	Rounded or truncated
soi	Seen or implied
www	Without wrong working

Subject-specific Marking Instructions

a Annotations should be used whenever appropriate during your marking.

The A, M and B annotations must be used on your standardisation scripts for responses that are not awarded either 0 or full marks. It is vital that you annotate standardisation scripts fully to show how the marks have been awarded.

For subsequent marking you must make it clear how you have arrived at the mark you have awarded.

b An element of professional judgement is required in the marking of any written paper. Remember that the mark scheme is designed to assist in marking incorrect solutions. Correct *solutions* leading to correct answers are awarded full marks but work must not be judged on the answer alone, and answers that are given in the question, especially, must be validly obtained; key steps in the working must always be looked at and anything unfamiliar must be investigated thoroughly.

Correct but unfamiliar or unexpected methods are often signalled by a correct result following an *apparently* incorrect method. Such work must be carefully assessed. When a candidate adopts a method which does not correspond to the mark scheme, award marks according to the spirit of the basic scheme; if you are in any doubt whatsoever (especially if several marks or candidates are involved) you should contact your Team Leader.

c The following types of marks are available.

M
A suitable method has been selected and *applied* in a manner which shows that the method is essentially understood. Method marks are not usually lost for numerical errors, algebraic slips or errors in units. However, it is not usually sufficient for a candidate just to indicate an intention of using some method or just to quote a formula; the formula or idea must be applied to the specific problem in hand, eg by substituting the relevant quantities into the formula. In some cases the nature of the errors allowed for the award of an M mark may be specified.

A
Accuracy mark, awarded for a correct answer or intermediate step correctly obtained. Accuracy marks cannot be given unless the associated Method mark is earned (or implied). Therefore M0 A1 cannot ever be awarded.

B
Mark for a correct result or statement independent of Method marks.

E

A given result is to be established or a result has to be explained. This usually requires more working or explanation than the establishment of an unknown result.

Unless otherwise indicated, marks once gained cannot subsequently be lost, eg wrong working following a correct form of answer is ignored. Sometimes this is reinforced in the mark scheme by the abbreviation isw. However, this would not apply to a case where a candidate passes through the correct answer as part of a wrong argument.

d When a part of a question has two or more 'method' steps, the M marks are in principle independent unless the scheme specifically says otherwise; and similarly where there are several B marks allocated. (The notation 'dep *' is used to indicate that a particular mark is dependent on an earlier, asterisked, mark in the scheme.) Of course, in practice it may happen that when a candidate has once gone wrong in a part of a question, the work from there on is worthless so that no more marks can sensibly be given. On the other hand, when two or more steps are successfully run together by the candidate, the earlier marks are implied and full credit must be given.

e The abbreviation ft implies that the A or B mark indicated is allowed for work correctly following on from previously incorrect results. Otherwise, A and B marks are given for correct work only — differences in notation are of course permitted. A (accuracy) marks are not given for answers obtained from incorrect working. When A or B marks are awarded for work at an intermediate stage of a solution, there may be various alternatives that are equally acceptable. In such cases, exactly what is acceptable will be detailed in the mark scheme rationale. If this is not the case please consult your Team Leader.

Sometimes the answer to one part of a question is used in a later part of the same question. In this case, A marks will often be 'follow through'. In such cases you must ensure that you refer back to the answer of the previous part question even if this is not shown within the image zone. You may find it easier to mark follow through questions candidate-by-candidate rather than question-by-question.

f Wrong or missing units in an answer should not lead to the loss of a mark unless the scheme specifically indicates otherwise. Candidates are expected to give numerical answers to an appropriate degree of accuracy, with 3 significant figures often being the norm. Small variations in the degree of accuracy to which an answer is given (e.g. 2 or 4 significant figures where 3 is expected) should not normally be penalised, while answers which are grossly over- or under-specified should normally result in the loss of a mark. The situation regarding any particular cases where the accuracy of the answer may be a marking issue should be detailed in the mark scheme rationale. If in doubt, contact your Team Leader.

g Rules for replaced work

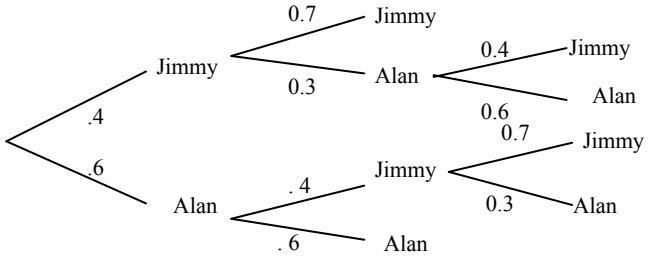
If a candidate attempts a question more than once, and indicates which attempt he/she wishes to be marked, then examiners should do as the candidate requests.
If there are two or more attempts at a question which have not been crossed out, examiners should mark what appears to be the last (complete) attempt and ignore the others.

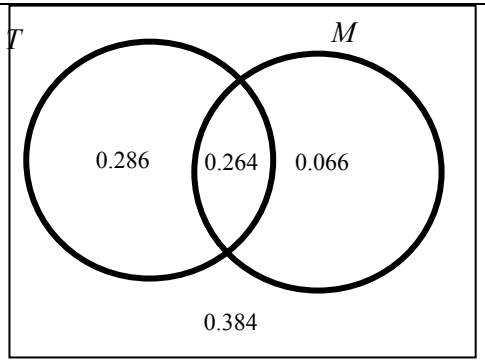
NB Follow these maths-specific instructions rather than those in the assessor handbook.

h For a *genuine* misreading (of numbers or symbols) which is such that the object and the difficulty of the question remain unaltered, mark according to the scheme but following through from the candidate's data. A penalty is then applied; 1 mark is generally appropriate, though this may differ for some units. This is achieved by withholding one A mark in the question.

Note that a miscopy of the candidate's own working is not a misread but an accuracy error.

Question			Answer	Marks	Guidance	Additional Guidance
1	(i)		0 | 8 8 10 | 5 5 20 | 5 6 9 30 | 1 1 4 4 6 Key 20 | 9 represents 29 degrees Celsius	G1 G1 G1 G1 **[4]**	Stem **(in either order)** Leaves Sorted and aligned **(use paper test if unsure)** Key	Do not allow leaves 25 ,26, 29 etc Ignore commas between leaves (indep). Condone 1 error or omission Allow errors in leaves if sorted Condone missing units (Celsius) Allow stem 0, 1, 2, 3
	(ii)		Median = 27.5	B1 **[1]**		**CAO**
1	(iii)		The median since the mean is affected by the skewness of the distribution	B1 E1 **[2]**	For median Allow **E2 for** mean if supported by very convincing reason **EG takes all values into account and no extreme values**	Do not allow 'less affected by extremes or outliers' unless also mention (**positive or negative)** skewness. Condone 'bottom half more spread' or similar
2	(i)		Mean $= \frac{759.00}{60} = £12.65$ Sxx $= 11736.59 - \frac{759^2}{60} = 2135.24$ s $= \sqrt{\frac{2135.24}{59}} = £6.02$	B1 M1 A1 **[3]**	Ignore units For Sxx CAO ignore units Allow more accurate answers	**CAO** Do not allow 759/60 as final answer but allow $12\ ^{13}/_{20}$ M1 for 11736.59 - 60 × their mean^2 BUT NOTE M0 if their $S_{xx} < 0$ For s^2 of 36.2 (or better) allow M1A0 with or without working For RMSD of 5.97 or 5.96 (or better) allow M1A0 provided working seen For RMSD^2 of 35.6 (or better) allow M1A0 provided working seen
2	(ii)		New mean = 12.65 × 1.02 = £12.90 New sd = 6.02 × 1.02 = £6.14	B1 B1 **[2]**	FT their mean Awrt 12.90 Allow 12.9 FT their sd	If candidate 'starts again' only award marks for CAO **Deduct at most 1 mark overall in whole question for overspecification of Mean and 1mark overall for SD**

Question			Answer	Marks	Guidance	Additional Guidance
2	(iii)		New mean = 12.65 + 0.25 = £12.90 New sd = £6.02	B1 B1 **[2]**	FT their mean Awrt 12.90 FT their sd (unless negative) Awrt 6.02	If candidate 'starts again' only award marks for CAO Allow sd unchanged (or similar)
3	(i)			 G1 G1 G1 **[3]**	Do a vertical scan and give: First column Second column Final column	All indep All probs must be correct Without extra branches in final column Ignore anything before third set Allow labels 'win' and 'lose' in place of Jimmy and Alan respectively **but if no labels, no marks**
3	(ii)		P(Alan wins) = (0.4×0.3×0.6) + (0.6×0.4×0.3) + (0.6×0.6) = 0.504	M1 M1 A1 **[3]**	For any one '**correct**' product For all three '**correct**' products **and no extras** CAO	FT their tree for both M marks Provided **correct number of terms in product(s) for both M1's**
3	(iii)		P(Ends after 4) = (0.4×0.7) + (0.6×0.6) = 0.28 + 0.36 = 0.64	M1 A1 **[2]**	For both products CAO	FT their tree for M mark but not for A mark Provided **two terms in each product**
4	(i)		Because P($T \mid M$) $\neq$ P(T)	E1 **[1]**	Or 0.8 ≠ 0.55	Or P($T \cap M$) (= 0.264) ≠ P(T) × P(M), **provided 0.264 in (ii)** Or 0.264 ≠ 0.55 × 0.33 (=0.1815) Look out for complement methods, etc
4	(ii)		P($T \cap M$) = P($T \mid M$) × P(M) = 0.80 × 0.33 = 0.264	M1 A1 **[2]**	For product CAO	**A0 for 0.26**

Question			Answer	Marks	Guidance	Additional Guidance
4	(iii)			G1 G1 G1 [3]	For two labelled intersecting circles For at least 2 correct probabilities. FT their $P(T \cap M)$ For remaining probabilities. FT their $P(T \cap M)$, **providing probabilities between 0 and 1**	Allow labels such as P(T) etc Allow other shapes in place of circles **No need for 'box'** **FT from 0.1815 in (ii) gives 0.3685, 0.1815, 0.1485, 0.3015**
5	(i)		$P(X=1) = P(g,b)+P(b,g)+P(b,b,g)+P(b,b,b,g)$ $= \frac{1}{4}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}=\frac{11}{16}$ OR $P(X=1) = 1- P(X\neq 1) = 1 - (P(bbbb)+P(ggb)+P(gggb)+P(gggg))$ $=1-\left(\frac{1}{16}+\frac{1}{8}+\frac{1}{16}+\frac{1}{16}\right)=\frac{11}{16}$	M1 M1 A1 [3]	For any two correct fractions For all four correct fractions *NB Answer given*	Must have correct ref to numbers of boys and girls, not just fractions With no extras Accept 0.6875, not 0.688. Watch for use of B(4, 0.5) $P(X\leq 2)=0.6875$ which gets M0M0A0.

Question			Answer	Marks	Guidance	Additional Guidance
5	(ii)		$E(X) = (0 \times \frac{1}{16}) + (1 \times \frac{11}{16}) + (2 \times \frac{1}{8}) + (3 \times \frac{1}{16}) + (4 \times \frac{1}{16})$ $= 1\frac{3}{8} = 1.375$ $E(X^2) = (0 \times \frac{1}{16}) + (1 \times \frac{11}{16}) + (4 \times \frac{1}{8}) + (9 \times \frac{1}{16}) + (16 \times \frac{1}{16})$ $= 2\frac{3}{4} = 2.75$ Var $(X) = 2\frac{3}{4} - \left(1\frac{3}{8}\right)^2 = \frac{55}{64} = 0.859$	M1 A1 M1 M1 A1 [5]	For Σrp (at least 3 terms correct) A1 CAO **Allow 1.38, not 1.4** For $\Sigma r^2 p$ (at least 3 terms correct) M1dep for – their E(X)² A1 FT their E(X) provided Var(X) > 0 **0.86, not 0.9**	Allow 22/16 Use of $E(X-\mu)^2$ gets M1 for attempt at $(x-\mu)^2$ should see $(-1.375)^2$, $(-0.375)^2$, $(0.625)^2$, 1.625^2, 2.625^2 (if E(X) correct but FT their E(X)) (all 5 correct for M1), then M1 for $\Sigma p(x-\mu)^2$ (at least 3 terms correct) Division by 5 or other spurious value at end gives max M1A1M1M1A0, or M1A0M1M1A0 if E(X) also divided by 5. Unsupported correct answers get 5 marks. **Using 1.38 gets Var of 0.8456 gets A1**
6	(i)	(*A*)	X ~ B(20, 0.25) P(4 smokers) = $\binom{20}{4} \times 0.25^4 \times 0.75^{16} = 0.1897$ **OR** Or from tables = 0.4148 – 0.2252 = 0.1896	M1 M1 A1 M2 A1 [3]	For $0.25^4 \times 0.75^{16}$ For $\binom{20}{4} \times p^4 \times q^{16}$ CAO For 0.4148 – 0.2252 CAO	With $p + q = 1$ Also for 4845×0.00003915 Allow 0.19 or better See tables at the website http://www.mei.org.uk/files/pdf/formula_book_mf2.pdf **0.189 gets A0**
6	(i)	(*B*)	P($3 \leq X \leq 6$) = 0.7858 – 0.0913 = 0.6945	M1 M1 A1 [3]	For (P($X \leq 6$) =) 0.7858 seen For **their** 0.7858 – 0.0913 CAO	Or P(X=3) + P(X=4) + P(X=5) + P(X=6) = 0.1339 + 0.1897 + 0.2023 + 0.1686 = 0.6945. M1 for three correct terms (to 2sf). Accept 0.69 or better **P($X \geq 3$) - P($X > 6$) = 0.9087-0.2142 = 0.6945 Gets M1 M1 A1**

Question			Answer	Marks	Guidance	Additional Guidance
6	(i)	(*C*)	E(X) = np = 20 × 0.25 = 5	B1 **[1]**	CAO	
6	(ii)	(*A*)	Let p = probability that a randomly selected student is a smoker H_0: $p = 0.25$ H_1: $p < 0.25$	B1 B1 B1 **[3]**	For definition of p in context For H_0 For H_1 **Allow complementary probabilities. Mark as per scheme.** **ie H_0:$p = 0.75$ etc**	Minimum needed for B1 is p = probability that student is a smoker. Allow p = P(student smokes) 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 that student is a smoker.,NOT just a sentence 'probability is 0.25' H_0 : p(student is a smoker) = 0.25, H_1 : p(student is a smoker) < 0.25 gets B0B1B1 Allow p=25%, allow θ or π and ρ but not x. However allow any single symbol if defined Allow H_0 = p=0.25, Do not allow H_0 : P(X=x) = 0.25, H_1 : P(X=x) < 0.25 Do not allow H_0: =0.25, =25%, P(0.25), p(0.25), p(x)=0.25, x=0.25 (unless x correctly defined as a probability) Do not allow H_1:$p \leq 0.25$, Do not allow H_0 and H_1 reversed for B marks but can still get E1 below Allow NH and AH in place of H_0 and H_1 For hypotheses given in words allow Maximum B0B1B1 and E1 below. Hypotheses in words must include probability (or chance or proportion or percentage) and the figure 0.25 oe.

Question			Answer	Marks	Guidance	Additional Guidance
6	(ii)	(*B*)	H_1 has this form as the programme aims to reduce the **proportion** of smokers.	E1 **[1]**	**Allow 'number'** **Allow 'aims for a reduction' or similar**	E0 if H_1 **upper tail or two tailed**
6	(iii)		$P(X \leq 1) = 0.0243 < 5\%$ $P(X \leq 2) = 0.0913 > 5\%$ So critical region is {0,1}	B1 B1 M1 A1 **[4]**	For **$P(X \leq 1) =$** 0.0243 For **$P(X \leq 2) =$** 0.0913 For at least one comparison with 5% CAO for critical region *dep* on M1 and at least one B1	**With full correct notation.** **Penalise once for eg $P(X=1)$, $P(X=2)$** Allow any form of statement of CR eg $X \leq 1$, $X < 2$, annotated number line, etc but not $P(X \leq 1)$ NB USE OF POINT PROBABILITIES gets B0B0M0A0 If no working but correct CR, no marks See additional notes below the scheme for other possibilities
6	(iv)		3 does not lie in the critical region, so not significant, So there is not enough evidence to reject the null hypothesis and we conclude that there is not enough evidence to suggest that the percentage of smokers has decreased.	E1**dep** E1**dep** **[2]**	For 3 not in CR or for not significant **or reject H_1** For conclusion in context **Condone omission of 'not enough evidence' in this case**	Dep on correct CR, (**correctly obtained)** E0E0 for $P(X=3)$ not in CR **E0E0 if wrong working after 3 not in CR** Alternative scheme $P(X \leq 3) = 0.2252 > 5\%$ so not sig etc. gets E2 for complete method but E0 otherwise.
7	(i)		Percentage = $\frac{40}{200} \times 100 = 20$	M1 A1 **[2]**	For 40 **seen or implied** CAO	
7	(ii)		Median = 5.2 kg Q1 = 4.2 Q3 = 5.8 Inter-quartile range = 5.8 – 4.2 = 1.6	B1 B1 B1 **[3]**	 For Q1 or Q3 For IQR	 Allow 4.2 to **4.3** for Q1 **Dep on both quartiles correct**

Question			Answer	Marks	Guidance	Additional Guidance
7	(iii)		Lower limit $4.2 - (1.5 \times 1.6) = 1.8$ Upper limit $5.8 + (1.5 \times 1.6) = 8.2$ So there are one or more outliers (if any lamb weighs more than 8.2 kg) Should not be disregarded because: **'Nothing to suggest they are not genuine items of data'** **Allow other convincing reasons such as very few so will not make much difference**	B1 B1 E1 E1 **[4]**	For 1.8 ft For 8.2 ft Dep on their 1.8 and 8.2 **Allow any number of outliers ≤ 5** Indep **Must give reason.**	Any use of median ± 1.5 IQR scores B0 B0 E0 **E0 if say some outliers at bottom end, unless lower limit > 2.0** If FT leads to limits above 9.0 and below 2.0 then E0 No marks for ± 2 or 3 IQR **With 4.3 and 5.8 lower = 2.05 and upper = 8.05** In this part FT their values from (ii) if sensibly obtained but not from location ie 12.5, 37.5 No marks for use of mean ± 2s
7	(iv)		Median for Welsh Mountain = 3.6 IQR for Welsh Mountain = 0.8 Welsh Mountain lambs have lower average weight than crossbred Welsh Mountain lambs also have lower variation in weight than crossbred	B1 B1 E1 **indep** E1 **indep** **[4]**	Must imply average or CT, not just median. **Allow generally lighter** **Must imply spread or variation, not just IQR or range** **Allow correct comment on consistency**	FT their medians FT their IQRs Can get max B1B0E1E1 for use of range
7	(v)		**Median unchanged** **IQR unchanged OR range or spread increased**	E1 E1 **[2]**	**even if used IQR in (iv)**	**E2 for 'Both comparisons remain the same'** **E1 for 'the range remains smaller'**

Question			Answer	Marks	Guidance	Additional Guidance
7	(vi)		$P(\text{Crossbred} > 3.9) = \frac{165}{200}$ $P(\text{Welsh Mountain} > 3.9) = \frac{1}{4}$ $P(\text{Both} > 3.9) = \frac{165}{200} \times \frac{1}{4} = \frac{165}{800} = \frac{33}{160} = 0.206$	B1 B1 M1 A1 **[4]**	For product of their probabilities , provided one is correct CAO	Allow 162 to 165 out of 200 **Allow answers in range 0.2025 to 0.20625 with correct working**

NOTE RE OVER-SPECIFICATION OF ANSWERS
If answers are grossly over-specified , 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.

Additional notes re Q6 parts iii, iv:

Smallest critical region method for part (iii):

Smallest critical region that 1 could fall into has size 0.0243 gets B1,
Smallest critical region that 2 could fall has size 0.0913 gets B1, This is > 5% or above < 5% gets M1, A1 as per scheme

Use of *k* method with no probabilities quoted:

$P(X \leq k) > 5\%$ and $P(X \leq k - 1) < 5\%$ followed by $k = 2$ gets SC1
so CR is {0, 1} gets another SC1 dep on first SC1

Use of *k* method with one probability quoted:

Mark as per scheme – max B0B1M1A1

Two tailed test with H_1: $p \neq 0.25$
Gets SC2 for fully correct FT with working as follows $P(X \leq 1) = 0.0243 < 0.025$ and $P(X \geq 10) = 0.0139 > 0.025$ B1 CR is {0,1, 10, 11, ..., 20}
(iv) Final 2 marks Max M1A1.

Two tailed test done but with correct H_1: $p<0.25$
(ii) gets max B1B1B1E1
(iii) if compare with 5% ignore work on upper tail and mark lower tail as per scheme but if include upper tail in CR then A0
if compare with 2.5% no marks B0B0M0A0
(iv) Final 2 marks can get M1A1 if correct CR, or SC2 if they start again, provided that they compare with 5%, not 2.5%.

Lower **or upper** tailed test with H_1: $p>0.25$ **and 6(ii)B wrong way around**
(ii) gets max B1B1B0E0
(iii) no marks B0B0M0A0
(iv) Final 2 marks get M0A0

Lower tailed test with H_1: $p>0.25$ and 6(ii)B right way around
(ii) gets max B1B1B0E0 , note E0, not E1
(iii) and (iv) Mark as per scheme, so full marks possible

Line diagram method for (iii)
No marks unless some 0.0243 shown on diagram, then B1 for squiggly line between 1and 2 or on 1, B1*dep* for arrow pointing to left, M1 0.0243 seen on diagram from squiggly line or from 1, A1 for CR written down in words/symbols. . **If 0.0243 and 0.0913 both seen and no other marks earned give B1.**

(iv) M1A1 as per scheme

Bar chart method for (iii)
No marks unless 0.0243 shown on diagram, then B1 for line clearly on boundary between 1 and 2 or within 1 block, B1*dep* for arrow pointing to left, M1 0.0243 seen on diagram from boundary line or from 1, A1 for CR written down in words/ symbols. **If 0.0243 and 0.0913 both seen and no other marks earned give B1.**
(iv) M1A1 as per scheme.

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

OCR Customer Contact Centre

Education and Learning
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

4766/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. Most candidates were well prepared for the paper and lower-scoring candidates scored marks throughout the paper, rather than on just a few questions. In general candidates supported their numerical answers with appropriate explanations and working, although in the more discursive questions, such as question 7 parts (iv) and (v), many candidates found it difficult to produce succinct answers and instead produced a rather 'rambling' solution. Presentation was generally satisfactory. Most candidates had adequate space in the answer booklet without having to use additional sheets, and very few candidates attempted parts of questions in answer sections intended for a different question/part. Once again many candidates over-specified some of their answers, despite recent Examiner's reports warning against this. This was particularly the case in Question 2, where for instance many candidates gave the new mean in part (ii) as £12.903, thus losing a mark. It is pleasing to report that there was much less over-specification of probabilities than in previous sessions. It is also pleasing to report that there was less use of point probabilities in the hypothesis testing in question 6. Please note that in future papers from January 2013 onwards, the definition of *p* will be expected to include 'in the population' – see question 6(ii)A.

Comments on Individual Questions

1 (i) Many candidates scored full marks in this easy start to the paper. A number of candidates omitted a key and rather more did not align the numbers well enough – particularly the last line '1 1 4 4 6'.

1 (ii) This was another well answered question. The most common error was an answer of 7.5, as a result of forgetting to add on the stem value of 20.

1 (iii) The majority of candidates stated that median was preferable and mentioned outliers or extremes. However, as these data have no clear outliers, candidates were expected to comment on the skewness of the data to gain full credit. Some candidates suggested that the mean was a better measure for these data, but credit was only given for a very convincing reason for this.

2 (i) Almost all candidates found the mean correctly and the majority also found the standard deviation, although this was sometimes over specified, thus losing a mark. The main difficulty found by candidates was in using the formula for *Sxx*. Very few found the root mean square deviation rather than the standard deviation.

2 (ii) Although most candidates used the correct method by multiplying their values from part (i) by 1.02, a good number lost marks for over-specification, very often giving the value of the mean as 12.903. Some candidates multiplied by 1.2 rather than 1.02. A number of candidates 'started again' to work out the new mean and standard deviation from scratch.

2 (iii) Most candidates scored both marks by adding 0.25 to the mean and saying that the standard deviation was unaltered (from part (i)). A few candidates incorrectly added 0.25 to their original standard deviation and also occasionally 25p became £25 to give 37.65 for the mean. A few candidates used their new values from (ii) instead of their original values found in part (i).

3 (i) Many candidates produced a fully correct tree diagram although a significant number added additional branches in the fifth round. Some were confused by the results in rounds 1 and 2 and tried to incorporate those, leading in most cases to significant loss of marks in parts (ii) and (iii). The probabilities 0.4/0.6 were often reversed, but 0.7/0.3 were nearly always correct. Labelling was often correct but some candidates either gave no labels or labelled the first set of branches only.

3 (ii) For those with the correct tree this probability was almost universally then found correctly. Credit was given for an attempt at follow through probabilities that matched the correct form of tree diagram.

3 (iii) Candidates with a correct tree almost always gained full marks and many others gained a follow through mark.

4 (i) Very few candidates realised that all they had to say was, "because $P(T|M) \neq P(T)$". Many candidates attempted $P(T) \times P(M) \neq P(T \cap M)$ often with success, but there was much confusion among some candidates with many incorrect statements.

4 (ii) There were many correct responses, but a significant number of candidates assumed independence (despite the question stating a lack of independence) and calculated $P(T) \times P(M) = 0.1815$.

4 (iii) This was done well by most candidates although often as a follow through from an incorrect $P(T \cap M)$.

5 (i) Marks could only be scored if candidates wrote down the four correct alternatives, *GB*, *BG*, *BBG, BBBG* in some form, and unfortunately the majority of candidates failed to do this. Those who did so almost universally scored full marks and almost all of the rest scored zero. A disappointing number of candidates showed that 11/16 was (1 – the sum of the remaining probabilities), which of course gained no credit.

5 (ii) This was very well answered, with many candidates scoring the full 5 marks. A few found $E(X^2)$ and stated that that was $VAR(X)$. It was pleasing to see that far fewer candidates than in previous years divided the mean and/or the variance by 5, or by other spurious factors.

6 (i) (*A*) The majority of candidates used the binomial formula rather than tables, but most answers were correct by either method, except for occasional over specification.

6 (i) (*B*) Many correct answers were seen from tables, but $P(X \leq 6) – P(X \leq 3)$ was a fairly common error, gaining just one mark and $P(X \leq 6) \times P(X \geq 3)$ was occasionally seen. Some candidates added individual probabilities often successfully.

6 (i) (*C*) Generally very well answered.

6 (ii) (*A*) The hypotheses were correctly stated in most cases although a few candidates gave a two tailed alternative hypothesis. More candidates than in the past are now giving an acceptable definition of *p*. However, please note that in future papers from January 2013 onwards, the definition of *p* will be expected to include 'in the population'. For example, in this paper, a suitable definition of *p* would have been: 'Let *p* = probability that a randomly selected student in the population is a smoker'.

6 (ii) (*B*) (ii)(*B*) The reasons for H_1 being $p<0.25$ were correct in most cases although some candidates simply stated the meaning of H_1.

6 (iii) This part was generally well done, and rather better done than in previous sessions. However, for those candidates not gaining full credit, a common error was the use of poor notation – the notation P(X = 1) was often seen instead of P($X \leq 1$), despite candidates then writing down the correct P($X \leq 1$) = 0.0213. The comparison with 5% was often not shown, losing the final 2 marks. It is pleasing to report that point probabilities are being used rather less than previously. A small number of candidates, having correctly found the probabilities and carried out the comparison, then stated the wrong critical region, thus demonstrating an insecure understanding.

6 (iv) Only the best candidates seemed to attempt this part in the way expected by the structure of the question. Many candidates did not realise that they could use their answer to part (iii) and started again. Some who had part (iii) correct now used point probabilities, thus losing 2 marks. A small number of candidates failed to make a conclusion in context, thus losing the final mark. On this occasion, as there were only 2 marks available, a statement of the form 'there is insufficient evidence to reject the null hypothesis' was not insisted on, but instead a statement such as 'accept the null hypothesis' was condoned, provided that this was followed by a conclusion in context. However, in future sessions such statements may not be condoned.

7 (i) This was very well answered. A few candidates lost marks due to either leaving the answer as 40 without giving the percentage, or working out that 80% had a birth weight of under 6kg rather than 20% over 6kg.

7 (ii) Again this was very well answered, with only a small number of the weakest candidates giving the location as the median and quartiles.

7 (iii) Most candidates found the upper and lower limits and correctly stated that there were a few outliers at the top end of the distribution. Some candidates used the median instead of the quartiles to work out the limits, whilst others multiplied the IQR by 2 instead of 1.5. A few candidates wrongly suggested that there were some outliers at the lower end of the distribution. Candidates often gave vague reasons for including or not including the outliers in the calculations, and few simply stated that there was nothing to suggest that these outliers were not genuine items of data.

7 (iv) There were some very good answers to this question, which were precise and concise. However not all candidates quoted the figures of 3.6 and 0.8, which were needed to gain full marks. Some candidates found it hard to refer to the 'central tendency' or 'average' and 'variation' and simply referred to the median (mean in some cases) and the range, thus losing 2 marks.

7 (v) This question was poorly answered. Many candidates did not realise that the median and IQR would remain unchanged. Several candidates were awarded 1 mark for stating that the range would have been increased. There were many candidates who 'waffled' and gave no substantive comments.

7 (vi) (vi) Many candidates struggled with this question. Some used probabilities of 'less than 3.9' rather than 'greater than 3.9'. Other candidates found both probabilities but did not know what to do with them. Some gave the probability for crossbred as $^{170}/_{200}$ rather than $^{165}/_{200}$. However, roughly one third of candidates produced a fully correct solution.