OCR
RECOGNISING ACHIEVEMENT

ADVANCED SUBSIDIARY 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

Monday 24 January 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 **8** 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 The stem and leaf diagram shows the weights, rounded to the nearest 10 grams, of 25 female iguanas.

8	3	9						
9	3	5	6	6	6	8	9	9
10	0	2	2	3	4	6	9	
11	2	4	7	8				
12	3	4	5					
13	2							

Key: 11 | 2 represents 1120 grams

(i) Find the mode and the median of the data. **[2]**

(ii) Identify the type of skewness of the distribution. **[1]**

2 The table shows all the possible products of the scores on two fair four-sided dice.

		Score on second die			
		1	2	3	4
Score on first die	1	1	2	3	4
	2	2	4	6	8
	3	3	6	9	12
	4	4	8	12	16

(i) Find the probability that the product of the two scores is less than 10. **[1]**

(ii) Show that the events 'the score on the first die is even' and 'the product of the scores on the two dice is less than 10' are not independent. **[3]**

3 There are 13 men and 10 women in a running club. A committee of 3 men and 3 women is to be selected.

(i) In how many different ways can the three men be selected? **[2]**

(ii) In how many different ways can the whole committee be selected? **[2]**

(iii) A random sample of 6 people is selected from the running club. Find the probability that this sample consists of 3 men and 3 women. **[2]**

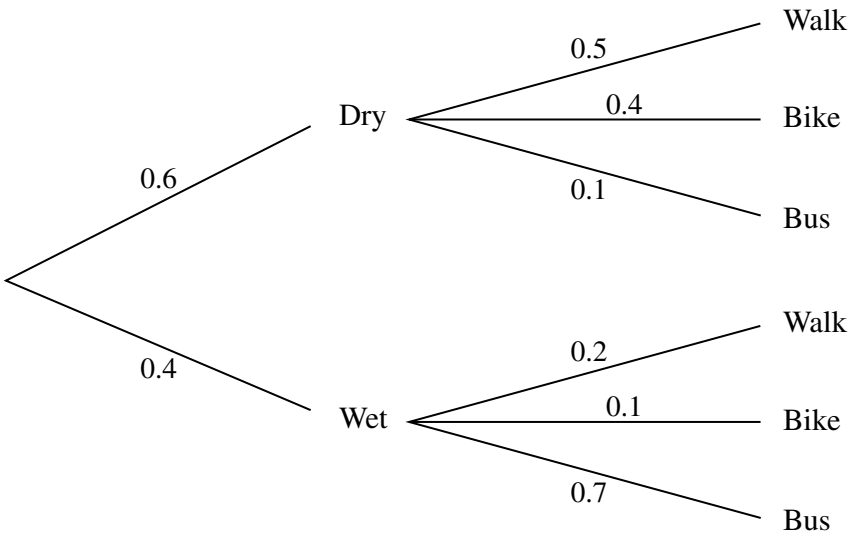
4 The probability distribution of the random variable X is given by the formula

$$P(X = r) = kr(r + 1) \quad \text{for } r = 1, 2, 3, 4, 5.$$

(i) Show that $k = \frac{1}{70}$. **[2]**

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

5 Andy can walk to work, travel by bike or travel by bus. The tree diagram shows the probabilities of any day being dry or wet and the corresponding probabilities for each of Andy's methods of travel.



A day is selected at random. Find the probability that

(i) the weather is wet and Andy travels by bus, **[2]**

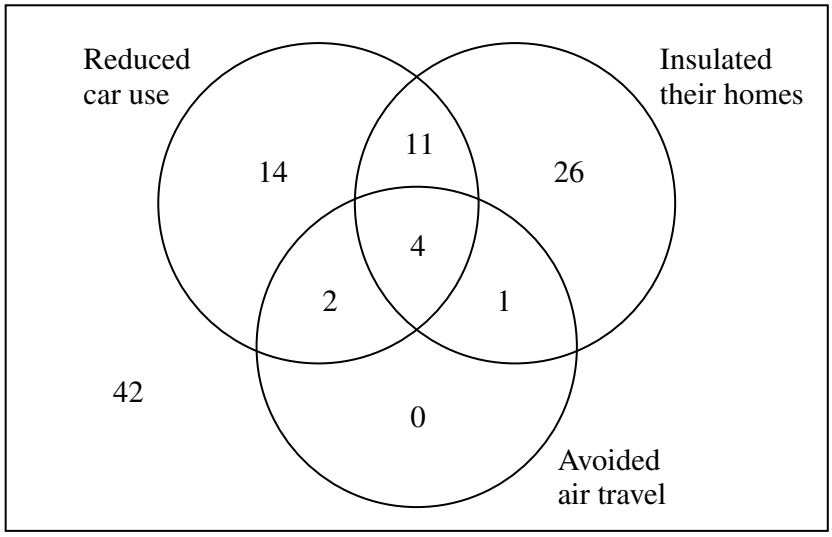
(ii) Andy walks or travels by bike, **[3]**

(iii) the weather is dry given that Andy walks or travels by bike. **[3]**

6 A survey is being carried out into the carbon footprint of individual citizens. As part of the survey, 100 citizens are asked whether they have attempted to reduce their carbon footprint by any of the following methods.

- Reducing car use
- Insulating their homes
- Avoiding air travel

The numbers of citizens who have used each of these methods are shown in the Venn diagram.



One of the citizens is selected at random.

(i) Find the probability that this citizen

(*A*) has avoided air travel, **[1]**

(*B*) has used at least two of the three methods. **[2]**

(ii) Given that the citizen has avoided air travel, find the probability that this citizen has reduced car use. **[2]**

Three of the citizens are selected at random.

(iii) Find the probability that none of them have avoided air travel. **[3]**

Section B (36 marks)

7 The incomes of a sample of 918 households on an island are given in the table below.

Income (x thousand pounds)	$0 \leqslant x \leqslant 20$	$20 < x \leqslant 40$	$40 < x \leqslant 60$	$60 < x \leqslant 100$	$100 < x \leqslant 200$
Frequency	238	365	142	128	45

(i) Draw a histogram to illustrate the data. **[5]**

(ii) Calculate an estimate of the mean income. **[3]**

(iii) Calculate an estimate of the standard deviation of the incomes. **[4]**

(iv) Use your answers to parts **(ii)** and **(iii)** to show there are almost certainly some outliers in the sample. Explain whether or not it would be appropriate to exclude the outliers from the calculation of the mean and the standard deviation. **[4]**

(v) The incomes were converted into another currency using the formula $y = 1.15x$. Calculate estimates of the mean and variance of the incomes in the new currency. **[3]**

8 Mark is playing solitaire on his computer. The probability that he wins a game is 0.2, independently of all other games that he plays.

(i) Find the expected number of wins in 12 games. **[2]**

(ii) Find the probability that

(*A*) he wins exactly 2 out of the next 12 games that he plays, **[3]**

(*B*) he wins at least 2 out of the next 12 games that he plays. **[3]**

(iii) Mark's friend Ali also plays solitaire. Ali claims that he is better at winning games than Mark. In a random sample of 20 games played by Ali, he wins 7 of them. Write down suitable hypotheses for a test at the 5% level to investigate whether Ali is correct. Give a reason for your choice of alternative hypothesis. Carry out the test. **[9]**

BLANK PAGE

BLANK PAGE

OCR
RECOGNISING ACHIEVEMENT

ADVANCED SUBSIDIARY 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

Monday 24 January 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 **8** pages. Any blank pages are indicated.

Section A (36 marks)

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

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

4 (i)	
4 (ii)	

5 (i)

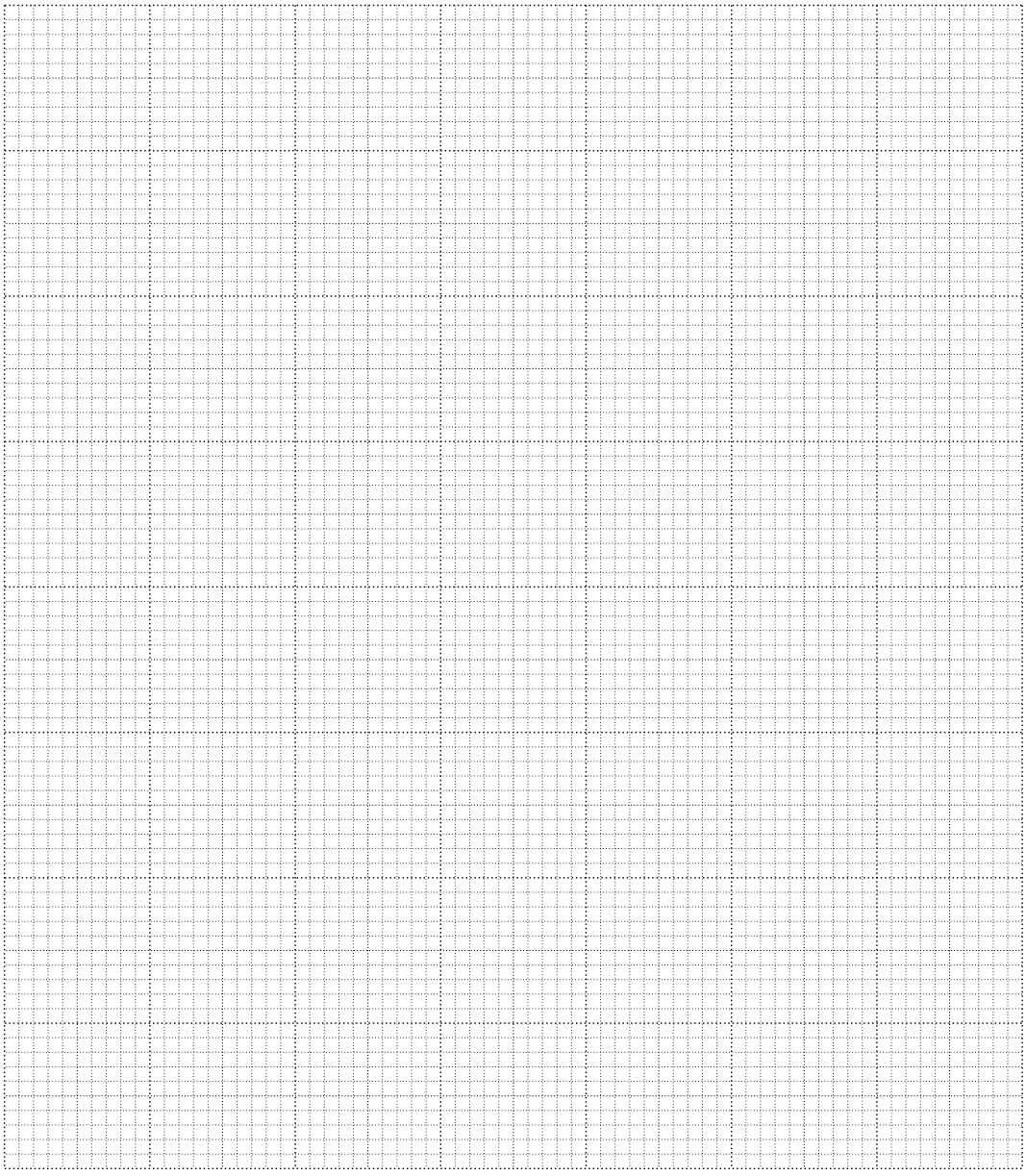
5 (ii)

5 (iii)

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

Section B (36 marks)

7 (i)



7 (ii)

7 (iii)

7 (iv)	
7 (v)	

8 (i)

8 (ii) (*A*)

8 (ii) (*B*)

8 (iii)

8 (iii)	(continued)

OCR
RECOGNISING ACHIEVEMENT

GCE

Mathematics (MEI)

Advanced Subsidiary GCE

Unit **4766:** Statistics 1

Mark Scheme for January 2011

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 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.

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

Marking instructions for GCE Mathematics (MEI): Statistics strand

1. You are advised to work through the paper yourself first. Ensure you familiarise yourself with the mark scheme before you tackle the practice scripts.

2. You will be required to mark ten practice scripts. This will help you to understand the mark scheme and will not be used to assess the quality of your marking. Mark the scripts yourself first, using the annotations. Turn on the comments box and make sure you understand the comments. You must also look at the definitive marks to check your marking. If you are unsure why the marks for the practice scripts have been awarded in the way they have, please contact your Team Leader.

3. When you are confident with the mark scheme, mark the ten standardisation scripts. Your Team Leader will give you feedback on these scripts and approve you for marking. (If your marking is not of an acceptable standard your Team Leader will give you advice and you will be required to do further work. You will only be approved for marking if your Team Leader is confident that you will be able to mark candidate scripts to an acceptable standard.)

4. Mark strictly to the mark scheme. If in doubt, consult your Team Leader using the messaging system within *scoris*, by email or by telephone. Your Team Leader will be monitoring your marking and giving you feedback throughout the marking period.

   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.

5. 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.

6. 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.

7. 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.

8. 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. 3 significant figures may often be the norm for this, but this always needs to be considered in the context of the problem in hand. For example, in quoting probabilities from Normal tables, we generally expect *some* evidence of interpolation and so quotation to 4 decimal places will often be appropriate. But even this does not always apply – quotations of the standard critical points for significance tests such as 1.96, 1.645, 2.576 (maybe even 2.58 – but not 2.57) will commonly suffice, especially if the calculated value of a test statistic is nowhere near any of these values. Sensible discretion *must* be exercised in such cases.

Discretion must also be exercised in the case of small variations in the degree of accuracy to which an answer is given. For example, if 3 significant figures are expected (either because of an explicit instruction or because the general context of a problem demands it) but only 2 are given, loss of an accuracy ("A") mark is likely to be appropriate; but if 4 significant figures are given, this should not normally be penalised. Likewise, answers which are slightly deviant from what is expected in a very minor manner (for example a Normal probability given, after an attempt at interpolation, as 0.6418 whereas 0.6417 was expected) should not be penalised.

However, answers which are *grossly* over- or under-specified should normally result in the loss of a mark. This includes cases such as, for example, insistence that the value of a test statistic is (say) 2.128888446667 merely because that is the value that happened to come off the candidate's calculator. Note that this applies to answers that are given as final stages of calculations; intermediate working should usually be carried out, and quoted, to a greater degree of accuracy to avoid the danger of premature approximation.

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.

9. **Rules for crossed out and/or replaced work**

If work is crossed out and not replaced, examiners should mark the crossed out work if it is legible.

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 two or more attempts are made at a question, and just one is not crossed out, examiners should ignore the crossed out work and mark the work that is not crossed out.

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.

10. Genuine misreading (of numbers or symbols, occasionally even of text) occurs. If this results in the object and/or difficulty of the question being considerably changed, it is likely that all the marks for that question, or section of the question, will be lost. However, misreads are often such that the object and/or difficulty remain substantially unaltered; these cases are considered below.

The simple rule is that *all* method ("M") marks [and of course all independent ("B") marks] remain accessible but at least some accuracy ("A") marks do not. It is difficult to legislate in an overall sense beyond this global statement because misreads, even when the object and/or difficulty remains unchanged, can vary greatly in their effects. For example, a misread of 1.02 as 10.2 (perhaps as a quoted value of a sample mean) may well be catastrophic; whereas a misread of 1.6748 as 1.6746 may have so slight an effect as to be almost unnoticeable in the candidate's work.

A misread should normally attract *some* penalty, though this would often be only 1 mark and should rarely if ever be more than 2. Commonly in sections of questions where there is a numerical answer either at the end of the section or to be obtained and commented on (eg the value of a test statistic), this answer will have an "A" mark that may actually be designated as "cao" [correct answer only]. This should be interpreted *strictly* – if the misread has led to failure to obtain this value, then this "A" mark must be withheld even if all method marks have been earned. It will also often be the case that such a mark is implicitly "cao" even if not explicitly designated as such.

On the other hand, we commonly allow "fresh starts" within a question or part of question. For example, a follow-through of the candidate's value of a test statistic is generally allowed (and

often explicitly stated as such within the marking scheme), so that the candidate may exhibit knowledge of how to compare it with a critical value and draw conclusions. Such "fresh starts" are not affected by any earlier misreads.

A misread may be of a symbol rather than a number – for example, an algebraic symbol in a mathematical expression. Such misreads are more likely to bring about a considerable change in the object and/or difficulty of the question; but, if they do not, they should be treated as far as possible in the same way as numerical misreads, *mutatis mutandis*. This also applied to misreads of text, which are fairly rare but can cause major problems in fair marking.

The situation regarding any particular cases that arise while you are marking for which you feel you need detailed guidance should be discussed with your Team Leader.

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

11. 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.

12. For answers scoring no marks, you must either award NR (no response) or 0, as follows:

Award NR (no response) if:

- Nothing is written at all in the answer space
- There is a comment which does not in any way relate to the question being asked ("can't do", "don't know", etc.)
- There is any sort of mark that is not an attempt at the question (a dash, a question mark, etc.)

The hash key [#] on your keyboard will enter NR.

Award 0 if:

- There is an attempt that earns no credit. This could, for example, include the candidate copying all or some of the question, or any working that does not earn any marks, whether crossed out or not.

13. The following abbreviations may be used in this mark scheme.

M1	method mark (M2, etc, is also used)
A1	accuracy mark
B1	independent mark
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
ft	follow through
isw	ignore subsequent working

oe	or equivalent
rot	rounded or truncated
sc	special case
soi	seen or implied
www	without wrong working

14. Annotating scripts. The following annotations are available:

✓and ✖	
BOD	Benefit of doubt
FT	Follow through
ISW	Ignore subsequent working (after correct answer obtained)
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 is also available to highlight any particular points on a script.

15. The comments box will be used by the Principal Examiner to explain his or her marking of the practice scripts for your information. Please refer to these comments when checking your practice scripts.

Please do not type in the comments box yourself. Any questions or comments you have for your Team Leader should be communicated by the *scoris* messaging system, e-mail or by telephone.

16. Write a brief report on the performance of the candidates. Your Team Leader will tell you when this is required. The Assistant Examiner's Report Form (AERF) can be found on the Cambridge Assessment Support Portal. This should contain notes on particular strengths displayed, as well as common errors or weaknesses. Constructive criticisms of the question paper/mark scheme are also appreciated.

17. Link Additional Objects with work relating to a question to those questions (a chain link appears by the relevant question number) – see scoris assessor Quick Reference Guide page 19-20 for instructions as to how to do this – this guide is on the Cambridge Assessment Support Portal and new users may like to download it with a shortcut on your desktop so you can open it easily! For AOs containing just formulae or rough working not attributed to a question, tick at the top to indicate seen but not linked. When you submit the script, *scoris* asks you to confirm that you have looked at all the additional objects. Please ensure that you have checked all Additional Objects thoroughly.

18. The schedule of dates for the marking of this paper is displayed under 'OCR Subject Specific Details' on the Cambridge Assessment Support Portal. It is vitally important that you meet these requirements. If you experience problems that mean you may not be able to meet the deadline then you must contact your Team Leader without delay.

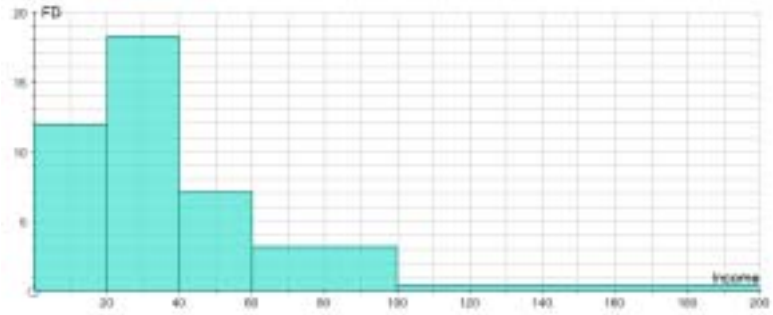
	SECTION A			
Q1 (i)	Mode = 960 (grams) Median = 1020 (grams) N.B. 96 and 102 gets SC1	B1 CAO B1 CAO	**2**	Ignore units and working
(ii)	Positive	E1	**1**	Not right skewed Not positive correlation
		TOTAL	**3**	
Q2 (i)	P(product of two scores < 10) = $\frac{13}{16}$ = 0.8125	B1	**1**	Allow 0.813 or 0.812
(ii)	P(*even*) × P(< 10) = 0.5 × $\frac{13}{16}$ = $\frac{13}{32}$ = 0.40625 P(*even* ∩ < 10) = $\frac{6}{16}$ = 0.375 So not independent.	M1 for 0.5 × $\frac{13}{16}$ or $\frac{13}{32}$ FT their answer to (i) M1 for $\frac{6}{16}$ A1	**3**	Do not allow these embedded in probability formulae Also allow P(even|<10) = 6/13≠ P(even) = 1/2 Or P(<10|even) = 6/8≠ P(<10) = 13/16 Or P(even|<10) = 6/13≠ P(even|<10') = 2/3 Or P(<10|even) = 6/8≠ P(<10|even') = 7/8 For all of these alternatives allow M2 for both probabilities. (M1 not available except if they correctly state both probabilities EG P(even|<10) and P(even) and get one correct) If they do not state what probabilities they are finding, give M2 for one of the above pairs of probabilities with ≠ symbol
		TOTAL	**4**	

Q3 (i)	$\binom{13}{3}$ ways of choosing the men = 286	M1 for $\binom{13}{3}$ seen A1	**2**	Accept $^{13}C_3$ or $^{13!}/_{(3!10!)}$ or equivalent for M1 No marks for permutations
(ii)	$\binom{13}{3}\times\binom{10}{3}$ = 286 × 120 = 34320	M1 for product A1 FT their 286	**2**	For permutations 1716 × 720 = 1235520 allow SC1 406 (from 286 + 120) scores SC1 (without further working)
(iii)	$\binom{23}{6}$=100947 34320/100947 = 1040/3059 = 0.340 (allow 0.34)	M1 for denominator of $\binom{23}{6}$ A1 FT	**2**	FT their 34320 Or $^6C_3\times13/23\times12/22\times11/21\times10/20\times9/19\times8/18$ = 0.340 scores M1 for product of fractions and A1 for $^6C_3\times$ and correct evaluation For permutations 1235520/72681840=0.017 scores SC1 Allow full marks for fractional answers, even if unsimplified 406/100947 = 0.00402 gets M1A1 with or without working
		TOTAL	**6**	

Q4 (i)	$2k + 6k + 12k + 20k + 30k = 1,\ 70k = 1$ $k = \frac{1}{70}$	M1 A1 NB **ANSWER GIVEN**	**2**	For five multiples of k (at least four correct multiples) Do not need to sum or =1 for M1 Condone omission of either $70k = 1$ or $k = 1/70$ but not both Condone omission of k: 2+6+12+20+30=70 Allow substitution of $k = 1/70$ into formula and getting at least four of 2/70, 6/70, 12/70, 20/70, 30/70 for M1 and 2/70+6/70+12/70+20/70+30/70 = 1 for A1
(ii)	$E(X) = 1\times\frac{2}{70}+2\times\frac{6}{70}+3\times\frac{12}{70}+4\times\frac{20}{70}+5\times\frac{30}{70}=4$ $E(X^2) =$ $1\times\frac{2}{70}+4\times\frac{6}{70}+9\times\frac{12}{70}+16\times\frac{20}{70}+25\times\frac{30}{70}=\frac{1204}{70}=17.2$ $Var(X) = 17.2 - 4^2 = 1.2$	M1 for Σrp (at least 3 terms correct) A1 CAO M1 for $\Sigma r^2 p$ (at least 3 terms correct) M1dep for - their $E(X)^2$ A1 FT their E(X) but not an error in $E(X^2)$ provided $Var(X) > 0$	**5**	280/70 scores M1A0 USE of $E(X-\mu)^2$ gets M1 for attempt at $(x-\mu)^2$ should see $(-3)^2$, $(-2)^2$, $(-1)^2$, 0^2, 1^2 (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 with their probabilities) Allow all M marks with their probabilities, (unless not between 0 and 1, see below for all probs 1/70). 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. SC2 for use of 1/70 for all probabilities leading to $E(X) = 3/14$ and $Var(X) = 145/196 = 0.74$
		TOTAL	**7**	

Q5 (i)	P(Wet and bus) = 0.4×0.7 $= 0.28$	M1 for multiplying probabilities A1 CAO	**2**	Fractional answer = 7/25 (Allow 28/100)
(ii)	P(Walk or bike) = $0.6 \times 0.5 + 0.6 \times 0.4 + 0.4 \times 0.2 + 0.4 \times 0.1$ or 0.3+0.24+0.08+0.04 = 0.66	M1 for any two correct pairs M1 for sum of all four correct terms With no extra terms for second M1 A1 CAO	**3**	Or = $0.6 \times 0.9 + 0.4 \times 0.3$ gets M1 for either term = 0.54 + 0.12 gets M1 for sum of both A1 CAO Or = 1 – 0.6×0.1 – 0.4×0.7 = 0.66. M1 for 1 – one correct term, M1 for complete correct expression and A1 for correct evaluation.
(iii)	P(Dry given walk or bike) $= \dfrac{\text{P(Dry and walk or bike)}}{\text{P(Walk or bike)}}$ $= \dfrac{0.6 \times 0.9}{0.66} = \dfrac{0.54}{0.66} = \dfrac{9}{11} = 0.818$	M1 for numerator leading to 0.54 M1 for denominator Ft their P(Walk or bike) from (ii) provided between 0 and 1 A1 FT	**3**	Allow 0.82, not 0.819 More accurate answer =0.81818 Fractional answer = 54/66 = 27/33 = 9/11 Condone answer of 0.8181 Do not give final A1 if ans ≥ 1
		TOTAL	**8**	

Q6 (i)	(A) P(Avoided air travel) $= \frac{7}{100} = 0.07$ (B) P(At least two) $= \frac{11+2+1+4}{100} = \frac{18}{100} = \frac{9}{50} = 0.18$	B1 aef isw M1 for (11+2+1+4)/100 A1 aef isw	**1** **2**	For M1 terms must be added must be as above or better with no extra terms (added or subtracted) for M1 Must simplify to 18/100 or 9/50 or 0.18 for A1 SC1 for 18/58 Or 1 – (14+26+0+42)/100 = 0.18 gets M1A1
(ii)	P(Reduced car use | Avoided air travel) $= \frac{6}{7} = 0.857$	M1 for denominator 7 or 7/100 or 0.07 FT their (i)A A1 CAO	**2**	Allow 0.86
(iii)	P(None have avoided air travel) $= \frac{93}{100} \times \frac{92}{99} \times \frac{91}{98} = 0.8025$	M1 for 93/100× (triple product) M1 for product of remaining fractions A1	**3**	Fuller answer 0.802511, so allow 0.803 without working, but 0.80 or 0.8 only with working . $(93/100)^3$ scores M1M0A0 which gives answer 0.804357 so watch for this. M0M0A0 for binomial probability including 0.93^{100} but ${}^3C_0 \times 0.07^0 \times 0.93^3$ still scores M1 $(k/100)^3$ for values of k other than 93 scores M0M0A0 $\frac{k}{100} \times \frac{(k-1)}{99} \times \frac{(k-2)}{98}$ for values of k other than 93 scores M1M0A0 Correct working but then multiplied or divided by some factor scores M1M0A0 ${}^{93}P_3 / {}^{100}P_3 = 0.803$ ${}^{93}P_3$ seen M1 divided by ${}^{100}P_3$ M1 0.803 A1 ${}^{93}C_3 / {}^{100}C_3 = 0.803$ Allow unsimplified fractional answer 778596/970200 =9269/11550
		TOTAL	**8**	

	SECTION B			
Q7 (i)	<table><tr><th>Income</th><th>Frequency</th><th>Width</th><th>FD</th></tr><tr><td>$0 \le x \le 20$</td><td>238</td><td>20</td><td>11.9</td></tr><tr><td>$20 < x \le 40$</td><td>365</td><td>20</td><td>18.25</td></tr><tr><td>$40 < x \le 60$</td><td>142</td><td>20</td><td>7.1</td></tr><tr><td>$60 < x \le 100$</td><td>128</td><td>40</td><td>3.2</td></tr><tr><td>$100 < x \le 200$</td><td>45</td><td>100</td><td>0.□5</td></tr></table>  <u>INCORRECT DIAGRAMS:</u> Frequency diagrams can get M0, A0, L0, W1, H0 MAXIMUM Thus frequency density = frequency × width, frequency/midpoint etc gets MAX M0A0L0W1H0	M1 for fds A1 CAO Accept any suitable unit for fd such as eg freq per £1000. L1 linear scale and label on vertical axis W1 linear scale on horizontal axis and correct width of bars H1 height of bars	**5**	At least 4 fds correct for M1 M1 can be also be gained from freq per 10K - 119, 182.5, 71, 32, 4.5 (at least 4 correct) and A1 for all correct Accept any suitable unit for fd, eg freq per £10K, BUT NOT FD per £1000 Allow fds correct to at least one dp If fd not explicitly given, M1 A1 can be gained from all heights correct (within one square) on histogram (and M1A0 if at least 4 correct) Allow restart although given fd wrong For L1, label required on vert axis in relation to first M1 mark ie fd or frequency density or if relevant freq/£10K, freq/£k etc (NOT fd/£10K) Accept f/w or f/cw (freq/width or freq/class width) Ignore horizontal label L1 can also be gained from an accurate key – may see 1 square = 36.5 or 23.8 or 14.2 For W1, must be drawn at 0, 20, 40 etc NOT 19.5 or 20.5 etc NO GAPS ALLOWED Must have linear scale. No inequality labels on their own such as 0≤I<20, 20≤I<40 etc but allow if a clear horizontal linear scale is also given. FT of heights *dep* on M1 all must agree with their fds If fds not given and one height is wrong then max M1A0L1W1H0 – visual check only (within one square) –no need to measure precisely

(ii)	Mean $= \dfrac{10\times 238 + 30\times 365 + 50\times 142 + 80\times 128 + 150\times 45}{918}$ $= \dfrac{37420}{918} = 40.8$	M1 for midpoints M1 for midpoints ×frequencies with divisor 918 A1 CAO	**3**	At least three midpoints correct for M1 (seen in (ii) or in table in (i)) No marks if not using midpoints Second M1 for sight of at least 3 double pairs seen out of 10 × 238 + 30 × 365 + 50 × 142 + 80 × 128 + 150 × 45 with divisor 918 Numerator = 2380+10950+7100+10240+6750 Use of LCB or UCB for midpoints here scores 0 For answer 40.76 or 40.8 or 41 mark as B3 37420/918 o.e. scores M1M1A0 NB Accept answers seen without working in part (ii) or (iii) (from calculator) Use of 'not quite right' midpoints such as 10.5, 30.5, etc can get M0M1A0 here and SC3 in (iii) Watch for incorrect method 238/10+365/30+142/50+128/80+45/150=40.71 Allow max 4 sf in final answer Also accept £40760, £40800 etc
(iii)	$\sum fx^2 = 238\times 10^2 + 365\times 30^2 + 142\times 50^2 + 128\times 80^2 + 45\times 150^2$ $= 2539000$ Or $238\times 100 + 365\times 900 + 142\times 2500 + 128\times 6400 +$ $45\times 22500 = 2539000$ Or $2380\times 10 + 10950\times 300 + 7100\times 50 + 10240\times 80 +$ $13500\times 150 = 2539000$ $S_{xx} = 2539000 - \dfrac{37420^2}{918} = 1013666$ $s = \sqrt{\dfrac{1013666}{917}} = 33.2$	M1 for at least 3 multiples fx^2 A1 for Σfx^2 M1 for attempt at S_{xx} Dep on first M1 BUT NOTE M1M0 if their $S_{xx} < 0$ A1 CAO If using LCB or UCB	**4**	For A1, all midpoints and frequencies correct Or $Sxx = 2539000 - 918 \times 40.76^2 = 1013855$, s=33.25. Using mean 40.8 leads to 1010861 , s= 33.20, Using mean = 41 leads to $Sxx = 995844$ and s = 32.95 M1M1 for $\sum f(x\text{-}x\text{bar})^2$ M1 for first three terms, M1 for all 5 terms $238 \times (10\text{-}40.76)^2 + 365 \times (30\text{-}40.76)^2 + 142 \times (50\text{-}40.76)^2 + 128 \times (80\text{-}40.76)^2 + 45 \times (150\text{-}40.76)^2$ (= 1013666) A1 for $S_{xx} = 1013666$ A1 for final answer

		consistently then allow SC2 if working is fully correct but SC0 otherwise but no marks in part (ii)		For answer 33.25 or 33.3 or 33.2 (www) can just mark as B4 - these may be from calculator without working Allow 33 with correct working rmsd = √(1013666/918) (=33.23) gets M1A1M1A0 (if seen) **WATCH FOR DIVISOR OF 918** Allow max 4 sf in final answer Allow £33200 etc
(iv)	($\bar{x} - 2s = 40.76 - 2\times 33.25 = -25.74$) $\bar{x} + 2s = 40.76 + 2\times 33.25 = 107.26$ Comment that there are almost certainly some outliers. Appropriate comment such as 'No, since there is nothing to indicate that these high earners represent a separate population.'	M1 for $\bar{x} + 2s$ or $\bar{x} - 2s$ A1 for 107.26 (FT) E1 E1 Dep on upper limit in range 106 - 108	**4**	FT any positive mean and positive sd for M1 Only follow through numerical values, not variables such as *s*, so if a candidate does not find *s* but then writes here 'limit is 40.76+ 2 × standard deviation', do NOT award M1 (This rule of not following through variables applies in all situations) Award E0E0 if their upper limit > 200 Allow 'Must be some outliers' Allow any comments that implies that there are outliers No marks in (iv) unless using $\bar{x} + 2s$ or $\bar{x} - 2s$
(v)	New mean = 1.15 × 40.76 = 46.87 New variance = $1.15^2 \times 33.25^2 = 1462$ For misread 1.5 in place of 1.15 For 1.5 × 40.76 = 61.1 and $1.5^2 \times 33.25^2 = 2490$ allow SC2 if all present but SC0 otherwise	B1 FT M1A1 FT	**3**	FT their mean (if given to ≥ 2 s.f.) FT their s (if given to ≥ 2 s.f.) provided their s>0 If RMSD found in part (i) rather than s, then FT their RMSD For new SD = 38.24 found instead of variance give M1A0 even if called variance (and FT their *s*) M0A0 for 1.15 x 33.25^2= 1271 Allow max 4 sf in final answers Min 2 sf If candidate 'starts again' only award marks for CAO
		TOTAL	**19**	

Q8 (i)	$E(X) = np = 12 \times 0.2 = 2.4$ Do not allow subsequent rounding.	M1 for product A1 CAO	**2**	If wrong n used consistently throughout, allow M marks only. NB If they round to 2, even if they have obtained 2.4 first they get M1A0. For answer of '2.4 or 2 if rounded up' allow M1A0 Answer of 2 without working gets M0A0. If they attempt E(X) by summing products xp give no marks unless answer is fully correct.
(ii)	X ~ B(12, 0.2) (***A***) P(Wins exactly 2) $= \binom{12}{2} \times 0.2^2 \times 0.8^{10} = 0.2835$ OR from tables $0.5583 - 0.2749 = 0.2834$	M1 $0.2^2 \times 0.8^{10}$ M1 $\binom{12}{2} \times p^2 q^{10}$ A1 CAO OR: M2 for 0.5583 – 0.2749 A1 CAO	**3**	With $p + q = 1$ Also for 66×0.004295 Allow answers within the range 0.283 to 0.284 with or without working or 0.28 to 0.283 if working shown See tables at the website http://www.mei.org.uk/files/pdf/formula_book_mf2.pdf
	(***B***) P(Wins at least 2) $= 1 - 0.2749 = 0.7251$	M1 P($X \leq 1$) M1 1-P(X≤1) A1 CAO	**3**	M1 0.2749 seen M1 1 – 0.2749 seen Allow 0.725 to 0.73 but not 0.72. Point probability method: $P(1) = 12 \times 0.2 \times 0.8^{11} = 0.2062$, $P(0) = 0.8^{12} = 0.0687$ So P($X \leq 1$) = 0.2749 gets M1 then mark as per scheme SC1 for 1 – P($X \leq 2$) = 1 – 0.5583 = 0.4417 For misread of tables value of 0.2749 , allow 0 in (*A*) but MAX M1M1 in (*B*) For P(X>1) = P(X=2) + P(X=3) + P(X=4) +.... allow M1 for 0.2835+0.2362+0.1329+0.0532+0.0155 and second M1 for 0.0033+0.0005+0.0001 and A1 for 0.725 or better M0M0A0 for 1 - P(X=1) = 1 – 0.2062 = 0.7938

(iii)	Let p = probability that Ali wins a game H_0: $p = 0.2$ H_1: $p > 0.2$ H_1 has this form as Ali claims that he is better at winning games than Mark is. *EITHER Probability method:* $P(X \geq 7) = 1 - P(X \leq 6)$ $= 1 - 0.9133 = 0.0867 > 5\%$ 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 Ali is better at winning games than Mark. Must include 'not enough evidence' or something similar for E1. 'Not enough evidence' can be seen in the either for the A mark or the E mark. Do not allow final conclusions for E1 such as : 'there is evidence to suggest that Ali is no better at winning games than Mark' or 'Mark and Ali have equal probabilities of winning games'	B1 for definition of p in context B1 for H_0 B1 for H_1 E1 B1 for $P(X \geq 7)$ B1 for 0.0867 Or 1 – 0.9133 seen M1 for comparison with 5% dep on B1 for 0.0867 A1 for not significant or 'accept H_0' or 'cannot reject H_0' or 'reject H_1' E1 dep on M1A1 Do not award first B1 for poor symbolic notation such as $P(X = 7) = 0.0867$ This comment applies to all methods	**4** **5**	Minimum needed for B1 is p = probability that Ali wins. Allow p = P(Ali wins) 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 Ali wins a game, NOT just a sentence 'probability is 0.2' H_0 : p(Ali wins) = 0.2, H_1 : p(Ali wins) > 0.2 gets B0B1B1Allow p=20%, allow θ or π and ρ but not x. However allow any single symbol <u>if defined</u> Allow $H_0 = p$=0.2, Allow $H_0 : p = {}^2/_{10}$ Do not allow $H_0 : P(X=x) = 0.2$, $H_1 : P(X=x) > 0.2$ Do not allow H_0: =0.2, =20%, P(0.2), p(0.2), p(x)=0.2, x=0.2 (unless x correctly defined as a probability) Do not allow $H_1 : p \geq 0.2$, 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 0.2 oe. Zero for use of point prob - $P(X = 7) = 0.0546$

	OR Critical region method: Let $X \sim B(20, 0.2)$ $P(X \geq 7) = 1 - P(X \leq 6) = 1 - 0.9133 = 0.0867 > 5\%$ $P(X \geq 8) = 1 - P(X \leq 7) = 1 - 0.9679 = 0.0321 < 5\%$ So critical region is {8,9,10,11,12,13,14,15,16,17,18,19,20} 7 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 Ali is better at winning games than Mark.	B1 for 0.0867 B1 for 0.0321 M1 for at least one comparison with 5% A1 CAO for critical region and not significant or 'accept H_0' or 'cannot reject H_0' or 'reject H_1' *dep* on M1 and at least one B1 E1 dep on M1A1		Allow any form of statement of CR eg $X \geq 8$, 8 to 20, 8 or above, $X > 8$, {8, …}, annotated number line, etc but not $P(X \geq 8)$ {8,9,10,11,12} gets max B2M1A0 – tables stop at 8. NB USE OF POINT PROBABILITIES gets B0B0M0A0 Use of complementary probabilities Providing there is sight of 95%, allow B1 for 0.9133, B1 for 0.9679, M1 for comparison with 95% A1CAO for correct CR See additional notes below the scheme for other possibilities **PLEASE CHECK THAT THERE IS NO EXTRA WORKING ON THE SECOND PAGE IN THE ANSWER BOOKLET**
		TOTAL	**17**	

NOTE RE OVER-SPECIFICATION OF ANSWERS

If answers are grossly over-specified (see instruction 8), deduct the final answer mark in every case, except where there are more than two over-specified answers in a single question (only likely in question 7) in which case deduct a mark in only the first two cases of over-specification in that question. Probabilities should also be rounded to a sensible degree of accuracy.

ADDITIONAL NOTES RE Q8 PART iii

Use of n = 12

$P(X \geq 7) = 1 - P(X \leq 6) = 1 - 0.9961 = 0.0039 < 5\%$

So significant or reject H_0 etc, so there evidence to suggest that Ali is better at winning games than Mark.

Gets B1 for $P(X \geq 7)$ B1 for 0.0039 M1 for comparison with 5% dep on B1 for 0.0039 A1 for significant E1 for evidence to suggest that Ali is better at winning games than Mark. Then award MR -1 so maximum of 4 possible

Comparison with 95% method

B1 for $P(X \leq 6)$

B1 for 0.9133

M1 for comparison with 95% dep on B1
A1 for not significant or 'accept H_0' or 'cannot reject H_0'
E1

Smallest critical region method:

Either:
Smallest critical region that 7 could fall into gets B1 and has size 0.0867 gets B1, This is > 5% gets M1, A1, E1 as per scheme
NB These marks only awarded if 7 used, not other values.

Use of *k* method with no probabilities quoted:

$P(X \geq 7) = 1 - P(X \leq 6) > 5\%$
$P(X \geq 8) = 1 - P(X \leq 7) < 5\%$
These may be seen in terms of *k* or *n*.
Either $k = 8$ or $k - 1 = 7$ so $k = 8$ gets SC1
so CR is {8,9,10,11,12,13,14,15, 16, 17, 18, 19, 20} gets another SC1and conclusion gets another SC1

Use of *k* method with one probability quoted:

1 - 0.9679 < 5% or 0.0321 < 5% gets B0B1M1
$P(X \leq k - 1) = P(X \leq 7)$
so $k - 1 = 7$ so $k = 8$ (or just $k = 8$)
so CR is {8,9,10,11,12,13,14,15, 16, 17, 18, 19, 20} and conclusion gets A1E1

Two tailed test with H_1: $p \neq 0.2$
Hyp gets max B1B1B0E0
$P(X \geq 7) = 0.0867$ gets B1B1comparison with 2.5% gets M1 (must be 2.5%)
Final marks A0E0

Two tailed test done but with correct H_1: $p > 0.2$
Hyp gets max B1B1B1E1

if compare with 5% ignore work on lower tail and mark upper tail as per scheme so can score full marks
if compare with 2.5% no marks B0B0M0A0E0

One tailed test with H_1: $p<0.2$
Hyp gets max B1B1B0E0
no further marks B0B0M0A0E0

Lower tailed test with H_1: $p>0.2$
Hyp gets max B1B1B0E0
no further marks B0B0M0A0E0

Line diagram method
B1 for squiggly line between 7 and 8 or on 8 exclusively (ie just one line), B1*dep* for arrow pointing to right, M1 0.0321 seen on diagram from squiggly line or from 8, A1E1 for correct conclusion

Bar chart method
B1 for line clearly on boundary between 7 and 8 or within 8 block exclusively (ie just one line),, B1*dep* for arrow pointing to right, M1 0.0321 seen on diagram from boundary line or from 8, A1E1 for correct conclusion

Using P(Not faulty) method
H_0: $p=0.8$, H_1: $p<0.8$, where p represents the prob that Ali loses a game Ali claims that the proportion of games that he loses is less than 80% gets B1B1B1E1
$P(X\leq 13)=0.0867>5\%$ 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 Ali is better at winning games than Mark. Gets B1B1M1A1E1

Chief Examiner's Introduction to Statistics Reports

Two general matters, to which some attention is also drawn in the individual subject reports, are worthy of mention in a general introduction, as they apply to all the Statistics modules.

First, advice was circulated several months ago concerning the issue of numerical accuracy of final answers, in particular to the practice of some candidates of gross over-specification in this regard. As an example, this would refer to the quotation of the value of a test statistic as, say, 2.18735693762 merely because this is the number that happened to appear on the candidate's calculator. This shows a complete lack of understanding of statistical practice and, indeed, of basic concepts of numeracy. In the current round of examinations, accuracy marks (but not method marks) were normally withheld in such cases. The earlier advice had explicitly stated that this would occur, and it will continue in future rounds. This is of course different from the desirable practice of retaining sufficient accuracy in intermediate calculations to avoid problems resulting from premature rounding.

Secondly, there are many references in the individual subject reports to the importance of securely stating hypotheses when conducting statistical tests. In future rounds of examinations, candidates will be expected to state their null and alternative hypotheses even if this is not explicitly asked for in the question. In many cases, this can sensibly and compactly be done in the usual notation of the subject, for example "H_0: $\mu = 25$; H_1: $\mu > 25$", but it would be expected that any parameters appearing in those statements are themselves briefly but adequately defined verbally. In the example, this might be achieved by adding "where μ is the population mean". There is no objection to hypotheses being stated verbally (for example "the null hypothesis is that the population mean is 25 (cm) and the alternative hypothesis is that it is greater"), but candidates must be careful to be precise in their wording (notably, explicit use of the word "population" will often be necessary for full marks to be awarded).

4766 Statistics 1

General Comments

The level of difficulty of the paper appeared to be appropriate for the candidates. There were fewer cases than in previous sessions where candidates scored most of their marks on one or two topics only and all the questions seemed accessible to any reasonably prepared candidate. The more able candidates scored well throughout the paper with the exception of question 6(iii) (the answer 0.93^3 was common), and 7(v) (the new variance often being given as 1.15×33.25^2). The less successful candidates often gained the majority of their marks from question 4 and the more straightforward parts of the probability tasks in questions 5, 6 and 8. No candidates appeared to have problems in completing the paper in the time available.

Most candidates supported their numerical answers with appropriate explanations and working. However there were a small number who left out the essential steps in a solution. For example there were a few who, in the hypothesis test, reached a correct probability of 0.0867 or 0.9133 and then failed to compare explicitly with 0.05 or 0.95, thus losing three marks. Some candidates did not state any probabilities to justify their critical region. In the past this omission has been treated generously, but in future candidates who fail to do so can expect to receive very few if any marks. A similar loss of three marks was incurred by those who just wrote an answer of 0.8 without any working in question 6(iii). Arithmetic accuracy was generally good although there is still evidence of candidates not being proficient or sensible in their use of calculators. In particular many candidates did not check their answers to question 7(ii) and 7(iii) with their calculator. The use of point probabilities still occurs in the hypothesis test, not only by the weakest candidates, although rather less frequently than in previous series. There was also an apparent lack of understanding of independent events, with only the more able candidates being able to give a convincing argument as to why the two events were dependent in question 2.

The over-specification of answers was prevalent in question 7. Inevitably the average and more able candidates were penalized more heavily by the loss of an accuracy mark as often the weaker candidates had already made some error in their working. It did appear that few candidates were aware of the necessity to consider how many figures should be specified. Often the answer to an estimated mean was given to 5 or more significant figures. On this occasion only, a maximum of 2 marks were deducted in a question for over-specification. However, in future series, over-specification will be penalised every time it occurs. (Please see the 'Note on accuracy in Statistics modules' contained in the Chief Examiners' report for June 2010).

Presentation was generally good. Fortunately only a small minority of candidates attempted parts of questions in answer sections intended for a different question/part! On a practical note most students are more aware than previously on how to use the answer booklet effectively i.e. using heavier pens, starting at the top of the box, clearly identifying final answers etc. Few candidates needed to use an additional answer book although for those that did, the use of additional sheets of graph paper caused problems with seeing the graphs. Candidates should draw their graphs boldly so that they can be viewed on screen clearly.

Comments on Individual Questions

1) (i) Most candidates achieved full marks with the most common mistake being to leave the answers as 96 and 102.

1) (ii) Almost all candidates could identify a positive skew correctly.

2) (i) Most candidates answered correctly. Occasionally a wrong answer of 3/16 was seen.

2) (ii) In contrast this part was only correctly completed by about one third of the candidature. Although many could quote a correct test for independence eg P(A∩B) = P(A)×P(B) a substantial minority attempted vague arguments about being "connected". For those using probabilities, calculating P(even ∩ <10) from the table seemed difficult, and many, having focussed on the "even" rows, calculated P(<10|even) instead. Those who attempted some form of conditional probability, e.g. comparing P(<10|even) with P(<10), usually failed to produce a correct argument, even if they managed to state the two probabilities which they were comparing. Rarely did they find both of them correctly. Some good solutions lost a mark due to the lack of a conclusion.

3) (i) Almost all candidates answered correctly but a very small number used permutations rather than combinations.

3) (ii) The majority of candidates answered correctly but a number added rather than multiplying or evaluated $\binom{23}{6}$.

3) (iii) Relatively few correct answers were seen in this part. A product of fractions leading to 0.017 was quite often seen as was a final answer of 100947 rather than 34320/100947.

4) (i) This was well answered. Some candidates lost a mark for failing to state that the sum of the probabilities is equal to 1.

4) (ii) Again most candidates answered correctly but a number made things difficult for themselves by using decimals rather than fractions, often losing 1 or 2 marks due to inaccuracy. When calculating the variance some candidates subtracted E(X) rather than $E(X)^2$.

5) (i) This was almost always answered correctly.

5) (ii) Again most candidates gave the correct answer of 0.66 although a number had difficulty selecting the appropriate pairs, despite the fully labelled tree diagram being provided. A few candidates insisted on multiplying all their otherwise correct values together hence gaining only the first M1.

5) (iii) Whilst approximately half of the candidature scored full marks, many gained just one mark for the correct denominator 0.66. It was far less frequent to see 0.54 as the numerator, 0.396 (0.6 x 0.66) or just 0.6 being two common errors. Some candidates calculated 0.54 and gave that as their answer, failing to recognise the conditional nature of the probability.

6) (i)(A) This was almost always answered correctly.

6) (i)(B) Whilst most candidates answered correctly, a number found the probability of exactly rather than at least two of the methods being used.

6) (ii) Many candidates had the correct denominator but rather fewer had the correct numerator. The most common errors were 2/7, again not including the 4 where all 3 methods had been used, and 0.31 as the denominator showing candidates' lack of understanding of conditional probability.

6) (iii) An answer of 0.93^3 was more common than the correct answer. A significant minority of candidates attempted to use a binomial expression, some of whom gained a method mark if 0.93^3 was the only value of significance (although it often was not). Several candidates gained from our acceptance of un-simplified fractional answers in this part.

7) (i) Most candidates scored well on the histogram although a very small number plotted frequencies rather than frequency density. The usual errors of lack of label or incorrect label (frequency or even cumulative frequency when frequency density was being used) on the vertical axis, a non linear scale or inequality label on the horizontal axis and the occasional mis-plotting (10.9 instead of 11.9 for example) did occur but with less regularity than in the past. A small number of candidates omitted the working for the frequency densities.

7) (ii) The estimation of the mean was answered well with the vast majority of candidates using the midpoints of the intervals and multiplying by the appropriate frequencies. A very small number used class boundaries or slightly incorrect midpoint values such as 10.5 instead of 10. A few replaced the midpoints by the interval widths, possibly because they had been used in the first part of the question. A relatively large number of candidates gave their answers to 5 or more significant figures, thus losing a mark for over-specification.

7) (iii) Candidates were less successful in finding the standard deviation, due either to the use of an incorrect formula or to the omission of the frequencies. The use of

$\sum f(x-\bar{x})^2$ was seen but rarely did it produce an accurate answer. Few candidates seem to use their calculator functions to check the accuracy of their answer. Very few incorrectly found the root mean square deviation rather than the standard deviation. Some candidates lost a mark in this part due to over-specification.

7) (iv) Most candidates knew that they needed to use the mean +/- 2 standard deviations to establish the possibility of outliers although a few tried to use the quartile method or produced a written argument unsupported by any figures. A number lost a mark because they failed to specifically state that there almost certainly some outliers. A significant number of candidates wished wrongly to exclude the outliers because 'it was grouped data' or because 'we do not know how many there are' or because 'there are too many outliers as 107.26 is much less that 200'.

7) (v) Very few candidates scored full marks in this part. Most found the new mean correctly but it was very often over-specified, even when the original mean in part (ii) was not. It was extremely rare to see the correct $1.15^2 \times 33.25^2$ or $(1.15 \times 33.25)^2$. The most common wrong answer was 1.15×33.25 (called the 'new standard deviation' or 'new variance' seemingly at random) followed closely by 1.15×33.25^2. Even the very few who wrote $1.15^2 \times 33.25^2$ often spoilt their final answer by over-specification.

8) (i) Nearly all candidates knew that the expected number of wins was calculated using *np*. However some rounded to 2 or occasionally 3, losing a mark in both cases.

8) (ii)(A) This part was also well answered. Most candidates used the formula successfully, though occasionally the binomial coefficient was forgotten or the power of 0.8 was given as 8 rather than 10. The binomial coefficient was rarely omitted. A few candidates used tables, usually correctly.

8) (ii)(B) Most candidates used tables for this part but a significant number chose to calculate the P(X=0) and P(X=1) and subtract. The majority of candidates arrived at the correct answer by one of these methods. There was some confusion between inequality statements and many used the wrong notation even if they selected the correct value from tables. Common wrong answers included 1 - 0.2749 - 0.0687, 1 - P(X≤ 2) = 0.4417 or just P(X≤ 2) = 0.5583.

8) (iii) In this final part most candidates either scored three marks or fewer (for the hypotheses) or at least 7. The hypotheses, on the whole, were well defined. The vast majority correctly used *p* rather than some other letter, though X appeared occasionally and sometimes no letter was used at all. Once again the definition of *p* was often absent though more made an attempt than in the past - the mistake this year was to miss out the fact that it was the probability that **Ali** won the game. The reason for the choice of alternative hypothesis was not always clearly defined; sometimes it could have been mistaken for the hypothesis given in words rather than a separate explanation. Again confusion over use of inequalities was often seen, candidates writing P(X=7) even when they meant P(X≥7). Point probabilities continue to be the preferred wrong method in this question, but they were seen less than in the past. Another common wrong method was 1 – P(X ≤ 7) = 0.0321, rather than P(X ≥ 7), leading to the opposite conclusion. Of those using the C.R. method, many also made this mistake, resulting in a C.R. of (7,8,9,… 20), again with the consequent opposite conclusion. Some candidates still failed to compare their probabilities with the significance level, though this was seen less than in previous series. Although it is given in the mark scheme, it is worth repeating here the recommended method for comparing the probabilities with the significance level: Candidates should find the two in this case upper tail cumulative probabilities which straddle the significance level and compare them both with the significance level.
$P(X \geq 7) = 1 - P(X \leq 6) = 1 - 0.9133 = 0.0867 > 5\%$
$P(X \geq 8) = 1 - P(X \leq 7) = 1 - 0.9679 = 0.0321 < 5\%$
The decision whether or not the value was significant was usually correct for the candidates who got this far. Explanations were, on the whole, very pleasing. Most candidates indicated in some way that there was 'not enough evidence' though a few still fail to put their conclusion in context. However the answers to this question showed a marked improvement from those in the past. Thank you to the Centres for taking note of the comments made in previous sessions. It is worth reiterating here the point made in the General Comments above 'Some candidates did not state any probabilities to justify their critical region. In the past this omission has been treated generously, but in future candidates who fail to do so can expect to receive very few if any marks.'

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