OCR RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MEI STRUCTURED MATHEMATICS 4766

Statistics 1

Thursday **9 JUNE 2005** Morning 1 hour 30 minutes

Additional materials:
Answer booklet
Graph paper
MEI Examination Formulae and Tables (MF2)

TIME 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

- Write your name, centre number and candidate number in the spaces provided on the answer booklet.
- Answer **all** the questions.
- There is an **insert** for use in Questions **2**, **5** and **6.**
- You are permitted to use a graphical calculator in this paper.

INFORMATION FOR CANDIDATES

- The number of marks is given in brackets [ ] at the end of each question or part question.
- 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.
- Final answers should be given to a degree of accuracy appropriate to the context.
- The total number of marks for this paper is 72.

This question paper consists of 5 printed pages, 3 blank pages and an insert.

Section A (36 marks)

1 At a certain stage of a football league season, the numbers of goals scored by a sample of 20 teams in the league were as follows.

22 23 23 23 26 28 28 30 31 33 33 34 35 35 36 36 37 46 49 49

(i) Calculate the sample mean and sample variance, s^2, of these data. [3]

(ii) The three teams with the most goals appear to be well ahead of the other teams. Determine whether or not any of these three pieces of data may be considered outliers. [2]

2 **Answer part (i) of this question on the insert provided.**

A taxi driver operates from a taxi rank at a main railway station in London. During one particular week he makes 120 journeys, the lengths of which are summarised in the table.

Length (x miles)	$0 < x \leqslant 1$	$1 < x \leqslant 2$	$2 < x \leqslant 3$	$3 < x \leqslant 4$	$4 < x \leqslant 6$	$6 < x \leqslant 10$
Number of journeys	38	30	21	14	9	8

(i) On the insert, draw a cumulative frequency diagram to illustrate the data. [3]

(ii) Use your graph to estimate the median length of journey and the quartiles.

Hence find the interquartile range. [4]

(iii) State the type of skewness of the distribution of the data. [1]

3 Jeremy is a computing consultant who sometimes works at home. The number, X, of days that Jeremy works at home in any given week is modelled by the probability distribution

$$P(X = r) = \frac{1}{40}r(r+1) \quad \text{for } r = 1, 2, 3, 4.$$

(i) Verify that $P(X = 4) = \frac{1}{2}$. [1]

(ii) Calculate $E(X)$ and $\text{Var}(X)$. [5]

(iii) Jeremy works for 45 weeks each year. Find the expected number of weeks during which he works at home for exactly 2 days. [2]

4 An examination paper consists of three sections.

- Section A contains 6 questions of which the candidate must answer 3
- Section B contains 7 questions of which the candidate must answer 4
- Section C contains 8 questions of which the candidate must answer 5

(i) In how many ways can a candidate choose 3 questions from Section A? [2]

(ii) In how many ways can a candidate choose 3 questions from Section A, 4 from Section B and 5 from Section C? [3]

A candidate does not read the instructions and selects 12 questions at random.

(iii) Find the probability that they happen to be 3 from Section A, 4 from Section B and 5 from Section C. [3]

5 **Answer part (i) of this question on the insert provided.**

The lowest common multiple of two integers, x and y, is the smallest positive integer which is a multiple of both x and y. So, for example, the lowest common multiple of 4 and 6 is 12.

(i) On the insert, complete the table giving the lowest common multiples of all pairs of integers between 1 and 6. [1]

		Second integer					
		1	2	3	4	5	6
First integer	1	1	2	3	4	5	6
	2	2	2	6	4	10	6
	3	3	6	3	12	15	6
	4	4	4	12			12
	5	5	10	15			
	6	6	6	6	12		

Two fair dice are thrown and the lowest common multiple of the two scores is found.

(ii) Use the table to find the probabilities of the following events.

(A) The lowest common multiple is greater than 6. [1]

(B) The lowest common multiple is a multiple of 5. [1]

(C) The lowest common multiple is both greater than 6 and a multiple of 5. [2]

(iii) Use your answers to part **(ii)** to show that the events "the lowest common multiple is greater than 6" and "the lowest common multiple is a multiple of 5" are not independent. [2]

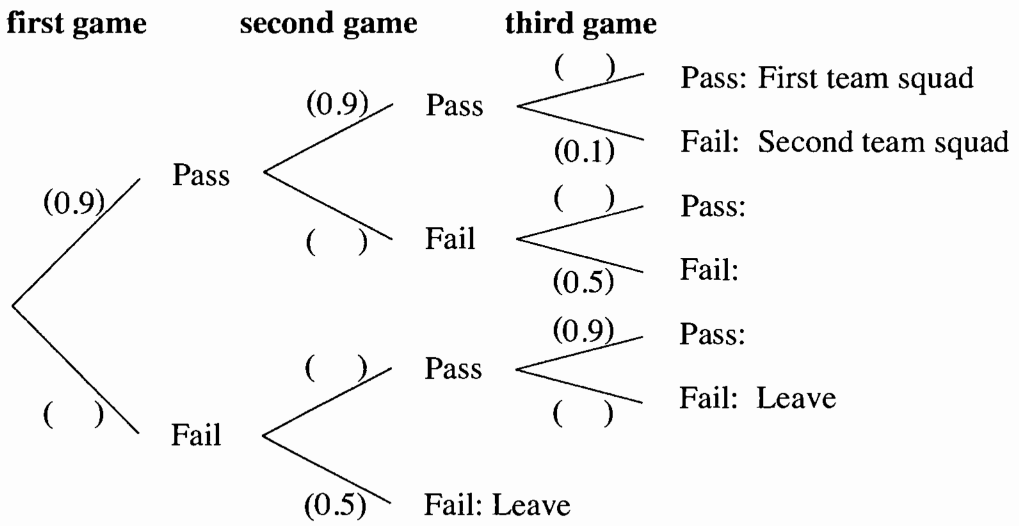
Section B (36 marks)

6 Answer part (i) of this question on the insert provided.

Mancaster Hockey Club invite prospective new players to take part in a series of three trial games. At the end of each game the performance of each player is assessed as pass or fail. Players who achieve a pass in all three games are invited to join the first team squad. Players who achieve a pass in two games are invited to join the second team squad. Players who fail in two games are asked to leave. This may happen after two games.

- The probability of passing the first game is 0.9
- Players who pass any game have probability 0.9 of passing the next game
- Players who fail any game have probability 0.5 of failing the next game

(i) On the insert, complete the tree diagram which illustrates the information above. [2]



(ii) Find the probability that a randomly selected player

(*A*) is invited to join the first team squad, [2]

(*B*) is invited to join the second team squad. [3]

(iii) Hence write down the probability that a randomly selected player is asked to leave. [1]

(iv) Find the probability that a randomly selected player is asked to leave after two games, given that the player is asked to leave. [2]

Angela, Bryony and Shareen attend the trials at the same time. Assuming their performances are independent, find the probability that

(v) at least one of the three is asked to leave, [3]

(vi) they pass a total of 7 games between them. [5]

7 A game requires 15 identical ordinary dice to be thrown in each turn.

Assuming the dice to be fair, find the following probabilities for any given turn.

(i) No sixes are thrown. [2]

(ii) Exactly four sixes are thrown. [3]

(iii) More than three sixes are thrown. [2]

David and Esme are two players who are not convinced that the dice are fair. David believes that the dice are biased against sixes, while Esme believes the dice to be biased in favour of sixes.

In his next turn, David throws no sixes. In her next turn, Esme throws 5 sixes.

(iv) Writing down your hypotheses carefully in each case, decide whether

(*A*) David's turn provides sufficient evidence at the 10% level that the dice are biased against sixes, [5]

(*B*) Esme's turn provides sufficient evidence at the 10% level that the dice are biased in favour of sixes. [4]

(v) Comment on your conclusions from part **(iv)**. [2]

Candidate Name	Centre Number	Candidate Number

RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MEI STRUCTURED MATHEMATICS 4766

Statistics 1

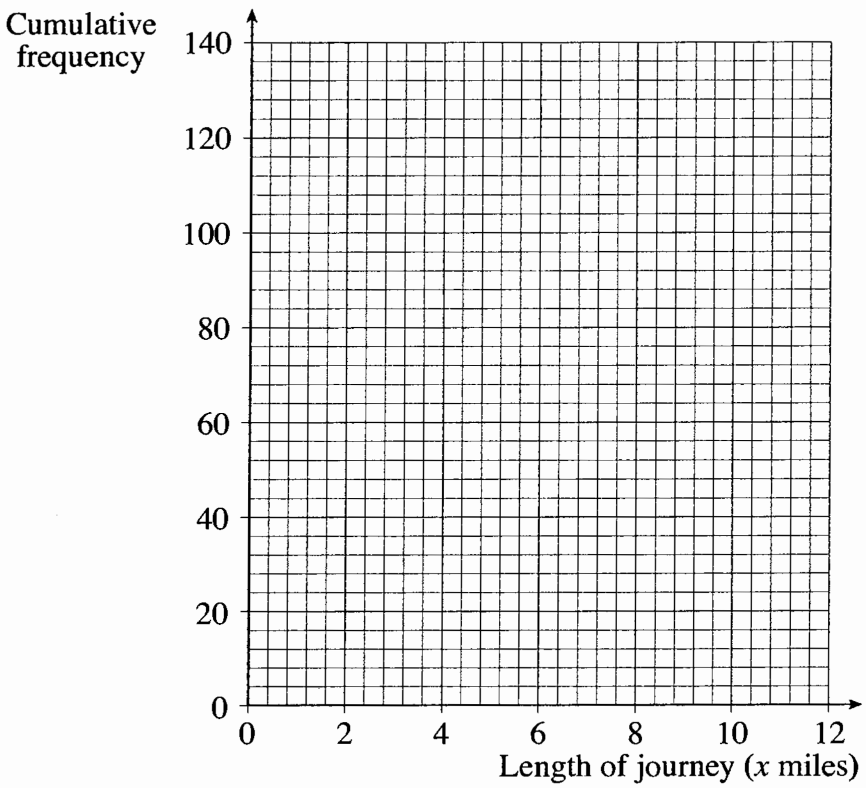
INSERT

Thursday **9 JUNE 2005** Morning 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

- This insert should be used in Questions **2** part **(i)**, **5** part **(i)** and **6** part **(i)**.
- Write your name, centre number and candidate number in the spaces provided at the top of this page and attach it to your answer booklet.

2 (i)



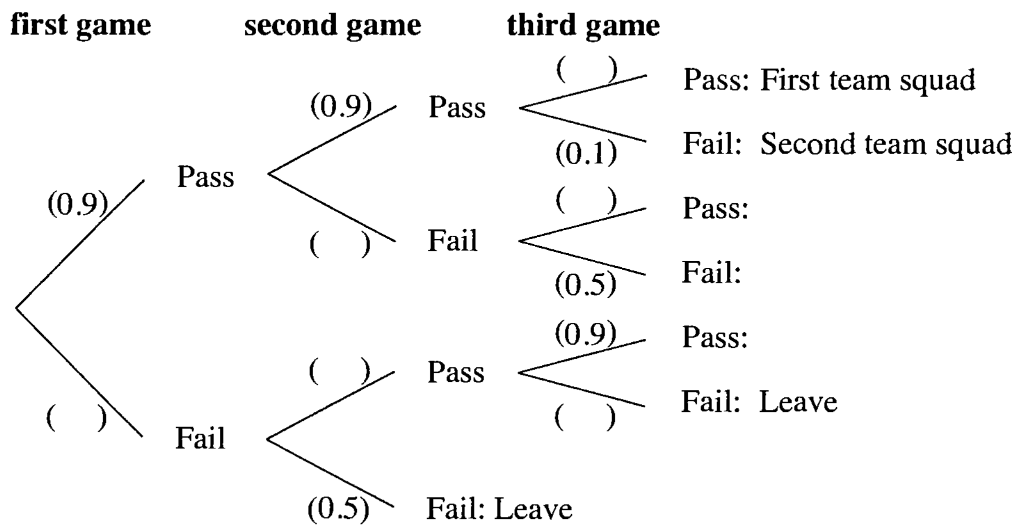
This insert consists of 2 printed pages.

[Turn over

5 (i)

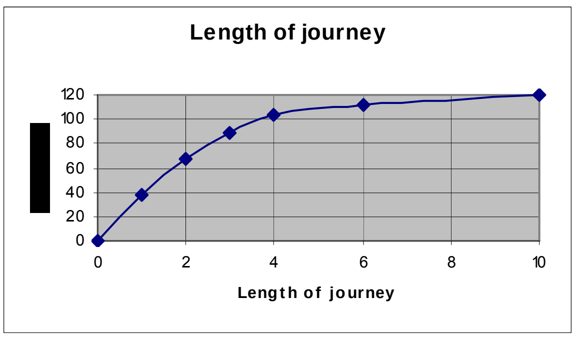
		Second integer					
		1	2	3	4	5	6
First integer	1	1	2	3	4	5	6
	2	2	2	6	4	10	6
	3	3	6	3	12	15	6
	4	4	4	12			12
	5	5	10	15			
	6	6	6	6	12		

6 (i)



Mark Scheme 4766
June 2005

Statistics 1 (4766)

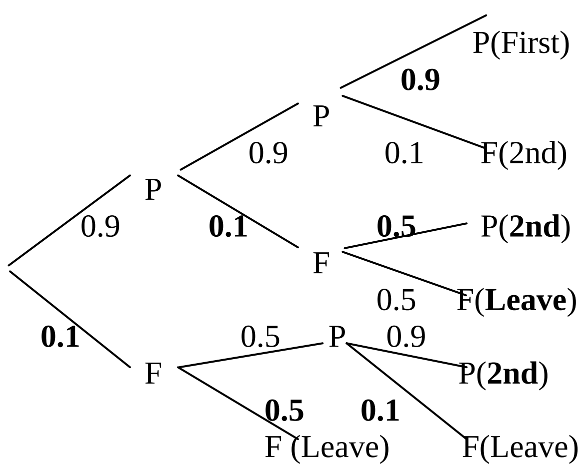
Qn	Answer	Mk	Comment
1 (i)	Mean = 657/20 = 32.85	B1 cao	
	Variance = $\frac{1}{19}(22839 - \frac{657^2}{20}) = 66.13$	M1 A1 cao	
(ii)	Standard deviation = 8.13		
	32.85 + 2(8.13) = 49.11	M1 ft	Calculation of 49.11
	none of the 3 values exceed this so no outliers	A1 ft	
2 (i)		G1 G1 G1	For calculating 38,68,89,103,112,120 Plotting end points Heights inc (0,0)
(ii)	Median = 1.7 miles	B1	
	Lower quartile = 0.8 miles	M1	
	Upper quartile = 3 miles	M1	
	Interquartile range = 2.2 miles	A1 ft	
(iii)	The graph exhibits positive skewness	E1	

3 (i)	$P(X=4) = \frac{1}{40}(4)(5) = \frac{1}{2}$ (Answer given)	B1	Calculation must be seen
(ii)	$E(X) = (2+12+36+80)\frac{1}{40}$ So $E(X) = 3.25$	M1 A1 cao	Sum of rp
	$\text{Var}(X) = (2+24+108+320)\frac{1}{40} - 3.25^2$ $= 11.35 - 10.5625$ $= 0.7875$	M1 M1 dep A1 cao	Sum of r^2p -3.25^2
(iii)	Expected number of weeks $= \frac{6}{40} \times 45$ $= 6.75$ weeks	M1 A1	Use of np
4 (i)	Number of choices $= \binom{6}{3} = 20$	M1 A1	For $\binom{6}{3}$
(ii)	Number of ways $= \binom{6}{3}\times\binom{7}{4}\times\binom{8}{5}$ $= 20\times35\times56$ $= 39200$	M1 M1 A1 cao	Correct 3 terms Multiplied
(iii)	Number of ways of choosing 12 questions $= \binom{21}{12} = 293930$ Probability of choosing correct number from each section = 39200/293930 = 0.133	M1 M1 ft A1 cao	For $\binom{21}{12}$

5 (i)	*(see table below)*	B1	All correct

	1	**2**	**3**	**4**	**5**	**6**
1	1	2	3	4	5	6
2	2	2	6	4	10	6
3	3	6	3	12	15	6
4	4	4	12	**4**	**20**	12
5	5	10	15	**20**	**5**	**30**
6	6	6	6	12	**30**	**6**

(ii)	(*A*) P(LCM > 6) = 1/3	B1	
	(*B*) P(LCM = 5n) = 11/36	B1	
	(*C*) P(LCM > 6 ∩ LCM = 5n) = 2/9	M1 A1 cao	Use of diagram
(iii)	$\frac{1}{3} \times \frac{11}{36} \neq \frac{2}{9}$	M1	Use of definition
	Hence events are not independent	E1	

6 (i)		G1 G1	Probabilities Outcomes
(ii)		M1	

(*A*) (*B*) (iii) (iv) (v) (vi)	P(First team) = 0.9^3 = 0.729 P(Second team) = $0.9 \times 0.9 \times 0.1 + 0.9 \times 0.1 \times 0.5 + 0.1 \times 0.9 \times 0.5$ = 0.081 + 0.045 + 0.045 = 0.171 P(asked to leave) = 1 -0.729 – 0.171 = 0.1 P(Leave after two games given leaves) $= \dfrac{0.1 \times 0.5}{0.1} = \dfrac{1}{2}$ P(at least one is asked to leave) $= 1 - 0.9^3$ = 0.271 P(Pass a total of 7 games) =P(First, Second, Second) + P(First, First, Leave after three games) = $3 \times 0.729 \times 0.171^2 + 3 \times 0.729^2 \times 0.05$ = 0.064 + 0.080 = 0.144	A1 M1 M1 A1 B1 M1 ft A1 cao M1 ft M1 A1 cao M1 M1 ft M1 ft M1 A1 cao	 1 correct triple 3 correct triples added Denominator Calc'n of 0.9 $1 - (\ )^3$ Attempts both $0.729(0.171)^2$ $0.05(0.729)^2$ multiply by 3

7 (i) (ii)	$X \sim B\left(15, \dfrac{1}{6}\right)$ $P(X=0) = \left(\dfrac{5}{6}\right)^{15}$ = 0.065 $P(X=4) = \dbinom{15}{4} \times \left(\dfrac{1}{6}\right)^4 \times \left(\dfrac{5}{6}\right)^{11}$ = 0.142 (or 0.9102-0.7685)	 M1 A1 cao M1 M1 A1 cao	 $\left(\dfrac{5}{6}\right)^{15}$ $\left(\dfrac{1}{6}\right)^4 \left(\dfrac{5}{6}\right)^{11}$ multiply by $\dbinom{15}{4}$

(iii)	$P(X > 3) = 1 - P(X \le 3)$ $= 1 - 0.7685 = 0.232$	M1 A1	
(iv) (*A*)	Let p = probability of a six on any throw $H_0 : p = \frac{1}{6}$ $H_1 : p < \frac{1}{6}$ $X \sim B\left(15, \frac{1}{6}\right)$ $P(X = 0) = 0.065$ 0.065 < 0.1 and so reject H_0 Conclude that there is sufficient evidence at the 10% level that the dice are biased against sixes.	B1 B1 M1 M1 dep E1 dep	Definition of p Both hypotheses 0.065 Comparison
(*B*)	Let p = probability of a six on any throw $H_0 : p = \frac{1}{6}$ $H_1 : p > \frac{1}{6}$ $X \sim B\left(15, \frac{1}{6}\right)$ $P(X \ge 5) = 1 - P(X \le 4) = 1 - 0.910 = 0.09$ 0.09 < 0.1 and so reject H_0 Conclude that there is sufficient evidence at the 10% level that the dice are biased in favour of sixes.	B1 M1 M1 dep E1 dep	Both hypotheses 0.09 Comparison
(v)	Conclusions contradictory. Even if null hypothesis is true, it will be rejected 10% of the time purely by chance. Or other sensible comments.	E1 E1	Contradictory By chance

4766 - Statistics 1

General Comments

The majority of candidates appeared to be well prepared for this paper and were able to have a good attempt at all the questions. However a significant number of candidates struggled with even the very straightforward material in questions 1, 2 and 5. The work of these candidates was also characterised by poor explanations, a lack of clear working and general carelessness. There was little evidence that candidates did not have sufficient time to complete the paper.

Comments on Individual Questions

1) (i) Most candidates were able to calculate the mean correctly. Any errors tended to be pure carelessness. The sample variance proved to be a greater challenge, with candidates confusing variance with standard deviation, divisor 20 with 19, $\sum x^2$ with $(\sum x)^2$ and $\sum (x-\bar{x})^2$.

(ii) Most candidates used the two standard deviation definition method and did so successfully. A minority of candidates used the 1.5 interquartile range method and received full credit.

2) (i) A majority of candidates did not show their calculated values of the cumulative frequencies. This was not a problem unless the points were plotted incorrectly, in which case no method marks could be gained. A significant number of candidates plotted points in the middle of class intervals rather than at the end.

(ii) Most candidates knew how to obtain values for the median and the quartiles from their graph, and almost without exception were ale to calculate the interquartile range.

(iii) The majority of candidates correctly described the skewness as positive, but a significant number, possibly confused by the shape of the cumulative frequency graph, gave the opposite response.

3) Most candidates did well on this question.

(i) This part of the question was almost always answered correctly.

(ii) Most candidates were able to calculate the mean, although a few calculated $\sum p$ rather than $\sum rp$. There were more errors in the calculation of the variance, including forgetting to subtract $(\mathrm{E}[X])^2$, or getting lost in a method based on $\sum (x-\bar{x})^2$. A small number of candidates did not attempt this part of the question.

(iii) This part of the question proved to be accessible even to those candidates who were unable to attempt part (ii). A significant number of candidates felt that the answer needed to be an integer, and so gave the answer 7 weeks. A smaller number of candidates converted the answer to days.

4) (i) Almost always answered correctly.

(ii) Although most candidates correctly obtained the correct three values of 20, 35 and 56, a considerable number of candidates then proceeded to add them, rather than multiply.

(iii) Despite being led by the previous part, most candidates were unable to make much progress with this part. Those attempting a solution using a product of fractions were, virtually without exception, doomed to failure. Often seen was $\frac{3}{6} \times \frac{4}{7} \times \frac{5}{8}$, and even those candidates who successfully obtained a string of 12 correct fractions failed to include a combination term.

5) (i) Virtually all candidates were able to complete the table correctly.

(ii) Parts *A* and *B* were often done correctly, but in part *C,* the majority of candidates assumed independence and simply multiplied their answers to parts *A* and *B.* Naturally, this gave them a problem in part (iii). Many other candidates also simply gave an answer with no supporting working. Simple annotation of the table could have earned these candidates marks for method.

(iii) Of those candidates who had not assumed independence earlier in the question, a significant number confused independence with mutual exclusivity and stated that the events could not be independent because some values were both greater than 6 and multiples of 5. Finally, some candidates who knew the definition for independence gave insufficiently clear answers such as $\frac{1}{3} \times \frac{11}{36} = \frac{11}{108}$ so independent.

6) This question proved a good source of marks for most candidates and also gave the opportunity for the very best candidates to shine in the final part.

(i) Virtually all candidates were able to complete the insert correctly.

(ii) Almost always correct.

(iii) Almost always correct.

(iv) Usually well done, but a significant minority of candidates failed to realise that conditional probability was involved and simply gave the answer of 0.05.

(v) Those candidates who took the approach of 1 – "the probability that no-one is asked to leave" were by far the most successful. Those who took an additive approach often omitted the required factors of 3.

(vi) This was probably the most difficult part of the paper and it prompted some very good solutions from a small number of candidates. A pleasing number of candidates were also able to gain some credit for being able to show that they had some understanding of the structure of the situation. Many candidates, however, based their answer on $B(9, 0.7)$.

7) The response to this question was not as good as in previous sessions, particularly in terms of hypothesis testing. The use of point probabilities was seen extensively.

(i) Almost always correct.

(ii) Usually correct, but some candidates omitted the combination term.

(iii) Often correct, but a significant number of candidates gave $P(X > 3)$ to be equal to either $1 - P(X \leq 2)$ or $1 - P(X = 3)$. Some also took an additive approach which rarely succeeded.

(iv)*A* Most candidates failed to define p in the hypotheses. Most candidates were able to calculate the correct probability, compare this with 10% and then reject the null hypothesis. However, only a minority then went on to explain this rejection in the context of the situation, i.e. Conclude that there is sufficient evidence at the 10% level that the dice are biased against sixes.

(iv)*B* This part was done much less well than the previous part. Many candidates calculated $P(X = 5)$. Many others were unable to calculate $P(X \geq 5)$ correctly.

(v) There were some good answers here which mentioned the fact that the results were contradictory, that different decisions would have been made at the 5% level and that these events could have occurred purely by chance.