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

Tuesday **18 JANUARY 2005** Afternoon 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.
- 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 6 printed pages and 2 blank pages.

Registered Charity 1066969
[Turn over

Section A (36 marks)

1 The number of minutes of recorded music on a sample of 100 CDs is summarised below.

Time (t minutes)	$40 \leqslant t < 45$	$45 \leqslant t < 50$	$50 \leqslant t < 60$	$60 \leqslant t < 70$	$70 \leqslant t < 90$
Number of CDs	26	18	31	16	9

(i) Illustrate the data by means of a histogram. [5]

(ii) Identify two features of the distribution. [2]

2 A sprinter runs many 100-metre trials, and the time, x seconds, for each is recorded. A sample of eight of these times is taken, as follows.

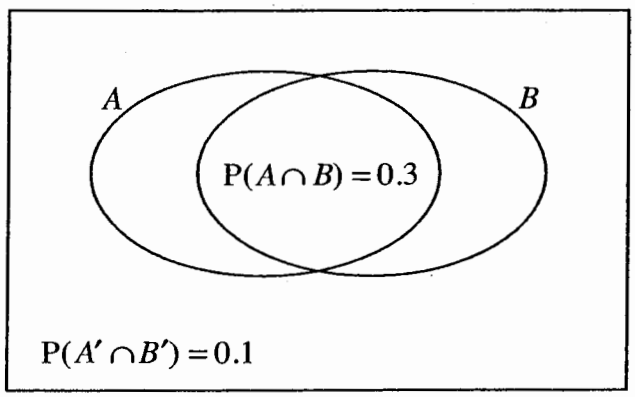
10.53 10.61 10.04 10.49 10.63 10.55 10.47 10.63

(i) Calculate the sample mean, $\bar{x}$, and sample standard deviation, s, of these times. [3]

(ii) Show that the time of 10.04 seconds may be regarded as an outlier. [2]

(iii) Discuss briefly whether or not the time of 10.04 seconds should be discarded. [2]

3 The Venn diagram illustrates the occurrence of two events A and B.



You are given that $P(A \cap B) = 0.3$ and that the probability that neither A nor B occurs is 0.1. You are also given that $P(A) = 2P(B)$.

Find $P(B)$. [3]

4 The number, X, of children per family in a certain city is modelled by the probability distribution $P(X = r) = k(6 - r)(1 + r)$ for $r = 0, 1, 2, 3, 4$.

(i) Copy and complete the following table and hence show that the value of k is $\frac{1}{50}$. [3]

r	0	1	2	3	4
$P(X = r)$	$6k$	$10k$			

(ii) Calculate $E(X)$. [2]

(iii) Hence write down the probability that a randomly selected family in this city has more than the mean number of children. [1]

5 A rugby union team consists of 15 players made up of 8 forwards and 7 backs. A manager has to select his team from a squad of 12 forwards and 11 backs.

(i) In how many ways can the manager select the forwards? [2]

(ii) In how many ways can the manager select the team? [3]

6 An amateur weather forecaster describes each day as either sunny, cloudy or wet. He keeps a record each day of his forecast and of the actual weather. His results for one particular year are given in the table.

		Weather Forecast			Total
		Sunny	Cloudy	Wet	
Actual Weather	Sunny	55	12	7	74
	Cloudy	17	128	29	174
	Wet	3	33	81	117
Total		75	173	117	365

A day is selected at random from that year.

(i) Show that the probability that the forecast is correct is $\frac{264}{365}$. [2]

Find the probability that

(ii) the forecast is correct, given that the forecast is sunny, [2]

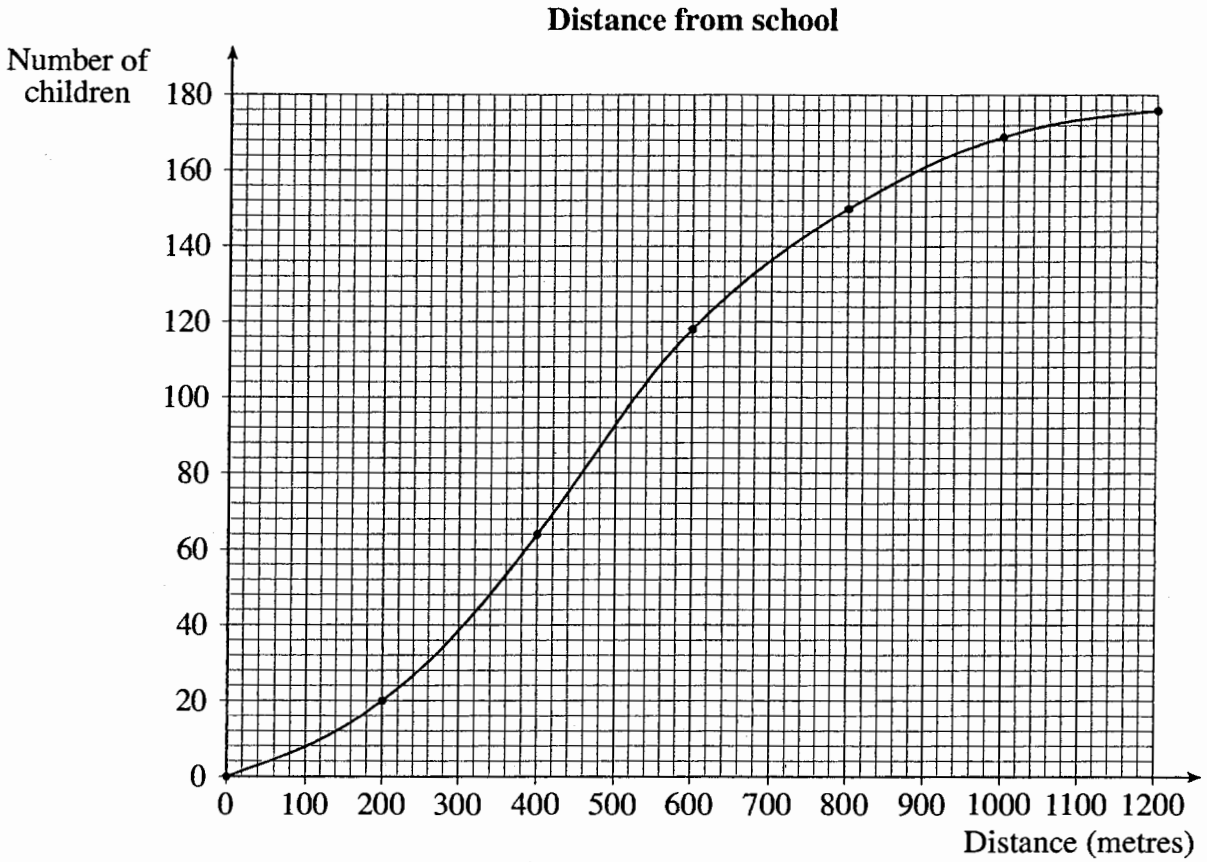
(iii) the forecast is correct, given that the weather is wet, [2]

(iv) the weather is cloudy, given that the forecast is correct. [2]

[Turn over

Section B (36 marks)

7 The cumulative frequency graph below illustrates the distances that 176 children live from their primary school.



(i) Use the graph to estimate, to the nearest 10 metres,

(*A*) the median distance from school, [2]

(*B*) the lower quartile, upper quartile and interquartile range. [3]

(ii) Draw a box and whisker plot to illustrate the data. [3]

[Question 7 continues on the next page.]

The graph on page 4 used the following grouped data.

Distance (metres)	200	400	600	800	1000	1200
Cumulative frequency	20	64	118	150	169	176

(iii) Copy and complete the grouped frequency table below describing the same data. [2]

Distance (d metres)	Frequency
$0 < d \leqslant 200$	20
$200 < d \leqslant 400$	

(iv) Hence estimate the mean distance these children live from school. [3]

It is subsequently found that none of the 176 children lives within 100 metres of the school.

(v) Calculate the revised estimate of the mean distance. [2]

(vi) Describe what change needs to be made to the cumulative frequency graph. [2]

[Question 8 is printed overleaf.]

8 At a doctor's surgery, records show that 20% of patients who make an appointment fail to turn up. During afternoon surgery the doctor has time to see 16 patients.

There are 16 appointments to see the doctor one afternoon.

(i) Find the probability that all 16 patients turn up. [2]

(ii) Find the probability that more than 3 patients do not turn up. [3]

To improve efficiency, the doctor decides to make more than 16 appointments for afternoon surgery, although there will still only be enough time to see 16 patients. There must be a probability of at least 0.9 that the doctor will have enough time to see all the patients who turn up.

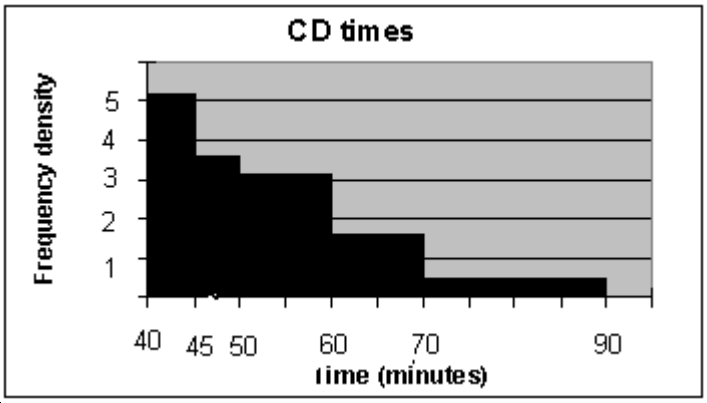
(iii) The doctor makes 17 appointments for afternoon surgery. Find the probability that at least one patient does not turn up. Hence show that making 17 appointments is satisfactory. [3]

(iv) Now find the greatest number of appointments the doctor can make for afternoon surgery and still have a probability of at least 0.9 of having time to see all patients who turn up. [4]

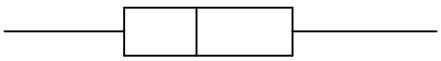
A computerised appointment system is introduced at the surgery. It is decided to test, at the 5% level, whether the proportion of patients failing to turn up for their appointments has changed. There are always 20 appointments to see the doctor at morning surgery. On a randomly chosen morning, 1 patient does not turn up.

(v) Write down suitable hypotheses and carry out the test. [7]

Mark Scheme

Qn	Answer	Mk	Comment
1 (i)	Time freq width freq density 40- 26 5 5.2 45- 18 5 3.6 50- 31 10 3.1 60- 16 10 1.6 70- 9 20 0.45	M1 A1	Calculation of fd's (accept values in proportion)
		G1 G1 G1	Linear scales Widths of bars Heights of bars
(ii)	e.g. The distribution is positively skewed The mode is at the extreme left of the distribution. Accept range = 50 or median = 52	E1 E1	
2 (i)	Mean = 83.95/8 = 10.49	B1	
	Variance = $\dfrac{881.2119 - \frac{83.95^2}{8}}{7}$ = 0.03737	M1	
	Standard deviation = 0.193	A1	
(ii)	2 standard deviations below mean = 10.49 – 2(0.193) = 10.104 but 10.04 < 10.104 so 10.04 is an outlier.	M1 A1	Follow through if divisor n has been used above.
(iii)	This time is much faster than the others. This may be the result of wind assistance, faulty timing, false start and should be discarded. Opposite conclusion such as this could be a genuinely fast time, can also receive full credit.	E1 E1	Appreciating need for investigation Comment in context

Qn	Answer	Mk	Comment
3	Let P(B) = x		
	Using P(AUB) = P(A) + P(B) – P(A ∩ B)	M1	Correct set of equations
	0.9 = 2x + x – 0.3	M1	Correct solution
	x = 0.4		
	P(B) = 0.4	A1	
4	r 0 1 2 3 4		
(i)	P($X = r$) $6k$ $10k$ **$12k$** **$12k$** **$10k$**	B1	1 value correct
		B1	all 3 correct
	$50k = 1 \rightarrow k = 1/50$	M1	sum of 1
(ii)	$E(X) = 110k = 2.2$	M1	sum of rp
		A1	cao
(iii)	$P(X > 2.2) = 22k = 0.44$	B1	
5			
(i)	$\binom{12}{8}$ ways of choosing forwards = 495	M1	
		A1	
(ii)			
	$\binom{12}{8} \times \binom{11}{7}$ ways of choosing team	M1	Product with (i)
		M1	backs
	= 495x330 = 163350	A1	cao
6			
(i)	P(Correct forecast) = $\frac{55+128+81}{365} = \frac{264}{365}$	M1	Numerator
		A1	
(ii)	P(Correct forecast given sunny forecast)		
	$= \frac{55}{75} = 0.733$	M1	Denominator
		A1	
(iii)	P(Correct forecast given wet weather)		
	$= \frac{81}{117} = 0.692$	M1	Denominator
		A1	
(iv)	P(Cloudy weather given correct forecast)		
	$= \frac{128}{264} = 0.485$	M1	Denominator
		A1	
Qn	**Answer**	**Mk**	**Comment**

7			
(i)	Median distance = 88^{th} value = 480	M1	Within 5
A		A1	cao
B	Lower Quartile = 44^{th} value = 320	B1	
	Upper Quartile = 132^{nd} value = 680	B1	
	Interquartile range = 680 – 320 = 360	M1	ft
(ii)		G1	Basic idea
		G1	Linear 0 - 1200
	0 320 480 680 1200	G1	Box including median (accurate)
(iii)	Distance Frequency		
	$0 < d \leq 200$ 20		
	$200 < d \leq 400$ 44	M1	Correct classes
	$400 < d \leq 600$ 54		Correct
	$600 < d \leq 800$ 32	M1	frequencies
	$800 < d \leq 1000$ 19		
	$1000 < d \leq 1200$ 7		
(iv)	Mid (x) f fx		
	100 20 2000	M1	mid points
	300 44 13200	M1	fx
	500 54 27000		
	700 32 22400		
	900 19 17100		
	1100 7 7700		
	176 89400		
	Estimate of mean = 507.95	A1	
(v)	Mid point of first class now 150	M1	150
	Total increase of 1000		
	New estimate of mean = 513.6	A1	
(vi)	The point (0,0) would move to (100,0)	E1	point (0,0)
		E1	point (100,0)

Qn	**Answer**	**Mk**	**Comment**

8	Number not turning up X~B(16,0.2)		
(i)	P(X = 0) = 0.8^{16} = 0.0281	M1 A1	0.8^{16} or tables
(ii)	P(X > 3) = 1 – P($X \leq 3$) or P($X \leq 12$) = 1 – 0.5981 = 0.4019	M1 M1 A1	Manipulation Use of tables
(iii)	X~B(17,0.2) → P($X \geq 1$) = 0.9775 Greater than 0.9 so acceptable	M1 A1 E1	B(17,0.2) 0.9775
(iv)	X~B(18,0.2) → P($X \geq 2$) = 0.9009 Can make 18 appointments X~B(19,0.2) → P($X \geq 3$) = 0.7631	M1 A1 A1 M1	18 and ≥2 0.9009 18 ok 19 and ≥3
(v)	Now X ~ B(20,p) Let p be probability of not turning up. H_0: p = 0.2 H_1: p ≠ 0.2 P($X \leq 1$) = 0.0692 > 2.5% cannot reject H_0 conclude that the proportion of patients not turning up is unchanged.	B1 B1 B1 M1 M1 A1 E1	0.0692 correct comparison cannot reject H_0

Examiner’s Report