OCR
RECOGNISING ACHIEVEMENT

ADVANCED SUBSIDIARY GCE

MATHEMATICS (MEI)

4766

Statistics 1

Candidates answer on the Answer Booklet

OCR Supplied Materials:
- 8 page Answer Booklet
- Graph paper
- MEI Examination Formulae and Tables (MF2)

Other Materials Required:
None

Monday 25 January 2010
Morning

Duration: 1 hour 30 minutes

INSTRUCTIONS TO CANDIDATES

- Write your name clearly in capital letters, your Centre Number and Candidate Number in the spaces provided on the Answer Booklet.
- Use black ink. Pencil may be used for graphs and diagrams only.
- Read each question carefully and make sure that 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 graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

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.
- The total number of marks for this paper is **72**.
- This document consists of **8** pages. Any blank pages are indicated.

© OCR 2010 [H/102/2650]
OCR is an exempt Charity

Section A (36 marks)

1 A camera records the speeds in miles per hour of 15 vehicles on a motorway. The speeds are given below.

73 67 75 64 52 63 75 81 77 72 68 74 79 72 71

(i) Construct a sorted stem and leaf diagram to represent these data, taking stem values of 50, 60, … . **[4]**

(ii) Write down the median and midrange of the data. **[2]**

(iii) Which of the median and midrange would you recommend to measure the central tendency of the data? Briefly explain your answer. **[2]**

2 In her purse, Katharine has two £5 notes, two £10 notes and one £20 note. She decides to select two of these notes at random to donate to a charity. The total value of these two notes is denoted by the random variable £X.

(i) (*A*) Show that P(X = 10) = 0.1. **[1]**

(*B*) Show that P(X = 30) = 0.2. **[2]**

The table shows the probability distribution of X.

r	10	15	20	25	30
P($X = r$)	0.1	0.4	0.1	0.2	0.2

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

3 In a survey, a large number of young people are asked about their exercise habits. One of these people is selected at random.

- G is the event that this person goes to the gym.
- R is the event that this person goes running.

You are given that P(G) = 0.24, P(R) = 0.13 and P($G \cap R$) = 0.06.

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

(ii) Determine whether the events G and R are independent. **[2]**

(iii) Find P($R \mid G$). **[3]**

© OCR 2010

4 In a multiple-choice test there are 30 questions. For each question, there is a 60% chance that a randomly selected student answers correctly, independently of all other questions.

(i) Find the probability that a randomly selected student gets a total of exactly 20 questions correct. **[3]**

(ii) If 100 randomly selected students take the test, find the expected number of students who get exactly 20 questions correct. **[2]**

5 My credit card has a 4-digit code called a PIN. You should assume that any 4-digit number from 0000 to 9999 can be a PIN.

(i) If I cannot remember any digits and guess my number, find the probability that I guess it correctly. **[1]**

In fact my PIN consists of four different digits. I can remember all four digits, but cannot remember the correct order.

(ii) If I now guess my number, find the probability that I guess it correctly. **[2]**

6 Three prizes, one for English, one for French and one for Spanish, are to be awarded in a class of 20 students.

Find the number of different ways in which the three prizes can be awarded if

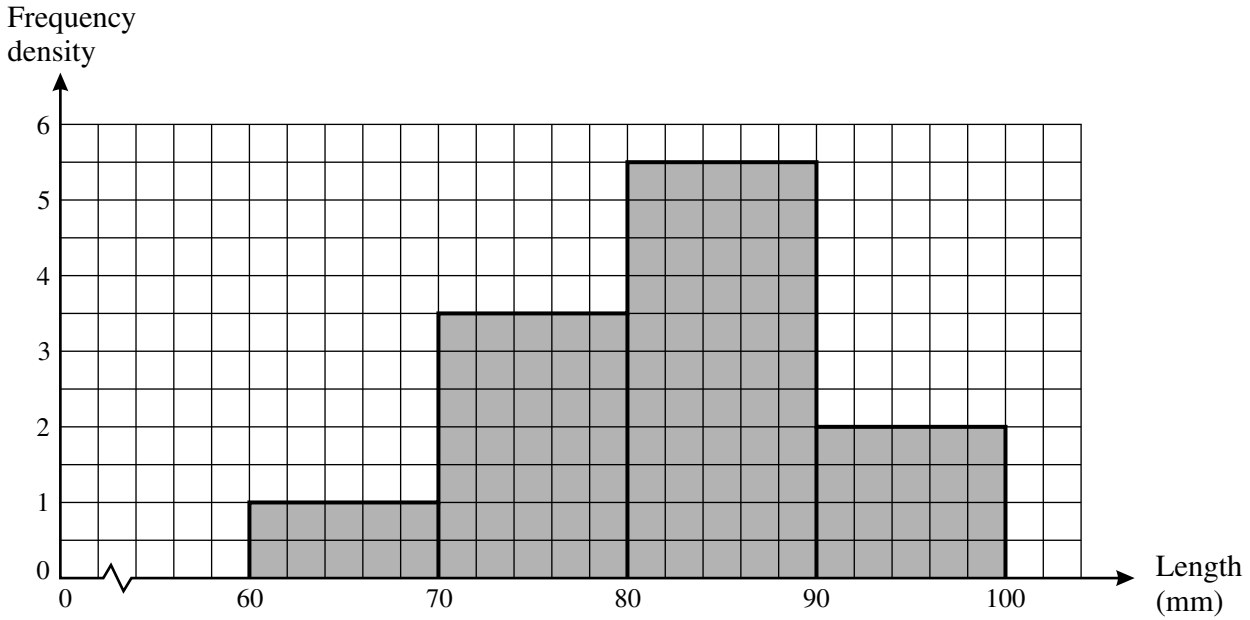
(i) no student may win more than 1 prize, **[2]**

(ii) no student may win all 3 prizes. **[2]**

© OCR 2010

Section B (36 marks)

7 A pear grower collects a random sample of 120 pears from his orchard. The histogram below shows the lengths, in mm, of these pears.



(i) Calculate the number of pears which are between 90 and 100 mm long. **[2]**

(ii) Calculate an estimate of the mean length of the pears. Explain why your answer is only an estimate. **[4]**

(iii) Calculate an estimate of the standard deviation. **[3]**

(iv) Use your answers to parts **(ii)** and **(iii)** to investigate whether there are any outliers. **[4]**

(v) Name the type of skewness of the distribution. **[1]**

(vi) Illustrate the data using a cumulative frequency diagram. **[5]**

© OCR 2010

8 An environmental health officer monitors the air pollution level in a city street. Each day the level of pollution is classified as low, medium or high. The probabilities of each level of pollution on a randomly chosen day are as given in the table.

Pollution level	Low	Medium	High
Probability	0.5	0.35	0.15

(i) Three days are chosen at random. Find the probability that the pollution level is

(*A*) low on all 3 days, **[2]**

(*B*) low on at least one day, **[2]**

(*C*) low on one day, medium on another day, and high on the other day. **[3]**

(ii) Ten days are chosen at random. Find the probability that

(*A*) there are no days when the pollution level is high, **[2]**

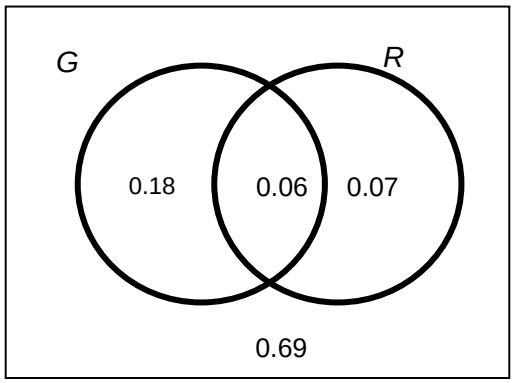
(*B*) there is exactly one day when the pollution level is high. **[3]**

The environmental health officer believes that pollution levels will be low more frequently in a different street. On 20 randomly selected days she monitors the pollution level in this street and finds that it is low on 15 occasions.

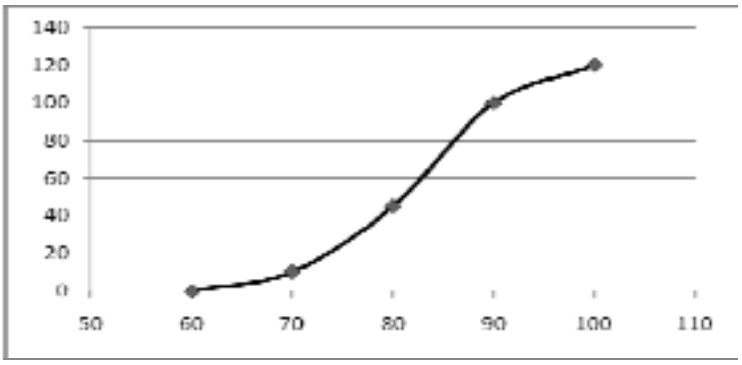
(iii) Carry out a test at the 5% level to determine if there is evidence to suggest that she is correct. Use hypotheses H_0: $p = 0.5$, H_1: $p > 0.5$, where p represents the probability that the pollution level in this street is low. Explain why H_1 has this form. **[5]**

© OCR 2010

4766 Statistics 1

1	(i)	5 | 2 6 | 3 4 7 8 7 | 1 2 2 3 4 5 5 7 9 8 | 1 Key 6 | 3 represents 63 mph	G1 stem G1 leaves CAO G1 sorted G1 key	[4]
	(ii)	Median = 72 Midrange = 66.5	B1 FT B1 CAO	[2]
	(iii)	*EITHER:* Median since midrange is affected by outlier (52) *OR:* Median since the lack of symmetry renders the midrange less representative	E1 for median E1 for explanation **TOTAL**	[2] [8]
2	(i)	(*A*) P(X = 10) = P(5 then 5) = 0.4×0.25 = 0.1 (*B*) P(X = 30) = P(10 and 20) = 0.4×0.25 + 0.2×0.5 = 0.2	B1 **ANSWER GIVEN** M1 for full calculation A1 **ANSWER GIVEN**	[1] [2]
	(ii)	$E(X) = 10\times0.1+15\times0.4+20\times0.1+25\times0.2+30\times0.2 = 20$ $E(X^2) = 100\times0.1+225\times0.4+400\times0.1+625\times0.2+900\times0.2 = 445$ Var(X) = 445 – 20^2 = 45	M1 for Σrp (at least 3 terms correct) A1 CAO M1 for Σr^2p (at least 3 terms correct) M1 dep for – their $E(X)^2$ A1 FT their E(X) provided Var(X) > 0 **TOTAL**	[5] [8]
3	(i)		G1 for two labelled intersecting circles G1 for at least 2 correct probabilities G1 for remaining probabilities	[3]
	(ii)	P(G) × P(R) = 0.24 × 0.13 = 0.0312 ≠ P($G\cap R$) or ≠ 0.06 So not independent.	M1 for 0.24 × 0.13 A1	[2]

	(iii)	$P(R \mid G) = \dfrac{P(R \cap G)}{P(G)} = \dfrac{0.06}{0.24} = \dfrac{1}{4} = 0.25$	M1 for numerator M1 for denominator A1 CAO **TOTAL**	**[3]** **[8]**
4	**(i)**	P(20 correct) = $\binom{30}{20} \times 0.6^{20} \times 0.4^{10} = 0.1152$	M1 $0.6^{20} \times 0.4^{10}$ M1 $\binom{30}{20} \times p^{20} q^{10}$ A1 CAO	**[3]**
	(ii)	Expected number = 100 × 0.1152 = 11.52	M1 A1 FT (Must not round to whole number) **TOTAL**	**[2]** **[5]**
5	**(i)**	P(Guess correctly) = $0.1^4 = 0.0001$	B1 CAO	**[1]**
	(ii)	P(Guess correctly) = $\dfrac{1}{4!} = \dfrac{1}{24}$	M1 A1 CAO **TOTAL**	**[2]** **[3]**
6	**(i)**	20 × 19 × 18 = 6840	M1 A1	**[2]**
	(ii)	$20^3 - 20 = 7980$	M1 for figures – 20 A1 **TOTAL**	**[2]** **[4]**

7	**(i)**	$10 \times 2 = 20$.	M1 for 10×2 A1 CAO	**[2]**
	(ii)	Mean $= \dfrac{10\times65+35\times75+55\times85+20\times95}{120} = \dfrac{9850}{120} = 82.08$ It is an estimate because the data are grouped.	M1 for midpoints M1 for double pairs A1 CAO E1 indep	**[4]**
	(iii)	$10 \times 65^2 + 35 \times 75^2 + 55 \times 85^2 + 20 \times 95^2$ (= 817000) $S_{xx} = 817000 - \dfrac{9850^2}{120}$ (= 8479.17) $s = \sqrt{\dfrac{8479.17}{119}} = 8.44$	M1 for Σfx^2 M1 for valid attempt at S_{xx} A1 CAO	**[3]**
	(iv)	$\bar{x} - 2s = 82.08 - 2\times8.44 = 65.2$ $\bar{x} + 2s = 82.08 + 2\times8.44 = 98.96$ So there are probably some outliers.	M1 FT for $\bar{x} - 2s$ M1 FT for $\bar{x} + 2s$ A1 for both E1 dep on A1	**[4]**
	(v)	Negative.	E1	**[1]**
	(vi)	Upper bound 60 70 80 90 100 Cumulative frequency 0 10 45 100 120  	C1 for cumulative frequencies S1 for scales L1 for labels 'Length and CF' P1 for points J1 for joining points dep on P1 All dep on attempt at cumulative frequency. **TOTAL**	**[5]** **[19]**

8	**(i)**	(*A*) P(Low on all 3 days) = $0.5^3 = 0.125$ or $^1/_8$	M1 for 0.5^3 A1 CAO	**[2]**
		(*B*) P(Low on at least 1 day) = $1 - 0.5^3 = 1 - 0.125 = 0.875$	M1 for $1 - 0.5^3$ A1 CAO	**[2]**
		(*C*) P(One low, one medium, one high) $= 6 \times 0.5 \times 0.35 \times 0.15 = 0.1575$	M1 for product of probabilities $0.5 \times 0.35 \times 0.15$ or $^{21}/_{800}$ M1 × 6 or × 3! or $^3P_3$ A1 CAO	**[3]**
	(ii)	X ~ B(10, 0.15) (*A*) P(No days) = $0.85^{10} = 0.1969$ Or from tables P(No days) = 0.1969	M1 A1	**[2]**
		(*B*) *Either* P(1 day) $= \binom{10}{1} \times 0.15^1 \times 0.85^9 = 0.3474$ *or* from tables P(1 day) $= P(X \leq 1) - P(X \leq 0)$ $= 0.5443 - 0.1969 = 0.3474$	M1 $0.15^1 \times 0.85^9$ M1 $\binom{10}{1} \times p^1 q^9$ A1 CAO OR: M2 for 0.5443 – 0.1969 A1 CAO	**[3]**
	(iii)	Let $X \sim B(20, 0.5)$ *Either:* $P(X \geq 15) = 1 - 0.9793 = 0.0207 < 5\%$	*Either:* B1 for correct probability of 0.0207 M1 for comparison	
		Or: Critical region is {15,16,17,18,19,20} 15 lies in the critical region.	*Or:* B1 for CR, M1 for comparison	
		So there is sufficient evidence to reject H_0	A1 CAO dep on B1M1	
		Conclude that there is enough evidence to indicate that the probability of low pollution levels is higher on the new street.	E1 for conclusion in context	**[5]**
		H_1 has this form as she believes that the probability of a low pollution level is greater in this street.	E1 indep	
			TOTAL	**[17]**