RECOGNISING ACHIEVEMENT

ADVANCED SUBSIDIARY GCE UNIT

4766/01

MATHEMATICS (MEI)

Statistics 1

FRIDAY 12 JANUARY 2007

Morning

Time: 1 hour 30 minutes

Additional Materials:
- Answer booklet (8 pages)
- Graph paper
- MEI Examination Formulae and Tables (MF2)

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.
- 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.
- The total number of marks for this paper is 72.

ADVICE TO CANDIDATES

- Read each question carefully and make sure you know what you have to do before starting your answer.
- 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.

This document consists of **6** printed pages and **2** blank pages.

Section A (36 marks)

1 The total annual emissions of carbon dioxide, x tonnes per person, for 13 European countries are given below.

6.2 6.7 6.8 8.1 8.1 8.5 8.6 9.0 9.9 10.1 11.0 11.8 22.8

(i) Find the mean, median and midrange of these data. [4]

(ii) Comment on how useful each of these is as a measure of central tendency for these data, giving a brief reason for each of your answers. [3]

2 The numbers of absentees per day from Mrs Smith's reception class over a period of 50 days are summarised below.

Number of absentees	0	1	2	3	4	5	6	>6
Frequency	8	15	11	8	3	4	1	0

(i) Illustrate these data by means of a vertical line chart. [2]

(ii) Calculate the mean and root mean square deviation of these data. [3]

(iii) There are 30 children in Mrs Smith's class altogether. Find the mean and root mean square deviation of the number of children who are present during the 50 days. [2]

3 The times taken for 480 university students to travel from their accommodation to lectures are summarised below.

Time (t minutes)	$0 \leqslant t < 5$	$5 \leqslant t < 10$	$10 \leqslant t < 20$	$20 \leqslant t < 30$	$30 \leqslant t < 40$	$40 \leqslant t < 60$
Frequency	34	153	188	73	27	5

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

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

4 A fair six-sided die is rolled twice. The random variable X represents the higher of the two scores. The probability distribution of X is given by the formula

$$P(X = r) = k(2r - 1) \text{ for } r = 1, 2, 3, 4, 5, 6.$$

(i) Copy and complete the following probability table and hence find the exact value of k, giving your answer as a fraction in its simplest form. [3]

r	1	2	3	4	5	6
$P(X = r)$	k					$11k$

(ii) Find the mean of X. [2]

A fair six-sided die is rolled three times.

(iii) Find the probability that the total score is 16. [3]

5 Each day the probability that Ashwin wears a tie is 0.2. The probability that he wears a jacket is 0.4. If he wears a jacket, the probability that he wears a tie is 0.3.

(i) Find the probability that, on a randomly selected day, Ashwin wears a jacket and a tie. [2]

(ii) Draw a Venn diagram, using one circle for the event 'wears a jacket' and one circle for the event 'wears a tie'. Your diagram should include the probability for each region. [3]

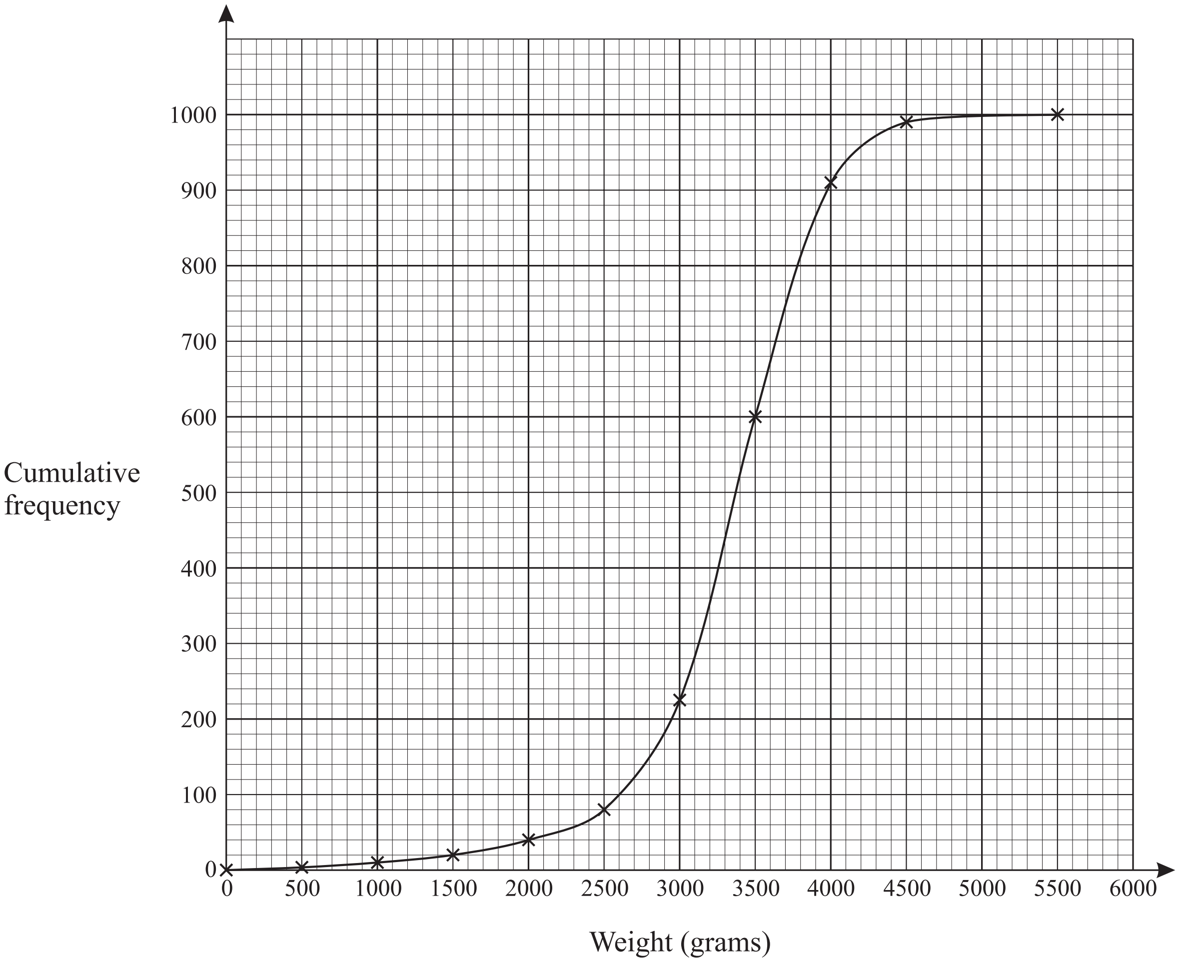
(iii) Using your Venn diagram, or otherwise, find the probability that, on a randomly selected day, Ashwin

(A) wears either a jacket or a tie (or both),

(B) wears no tie or no jacket (or wears neither). [3]

Section B (36 marks)

6 The birth weights in grams of a random sample of 1000 babies are displayed in the cumulative frequency diagram below.



(i) Use the diagram to estimate the median and interquartile range of the data. [3]

(ii) Use your answers to part **(i)** to estimate the number of outliers in the sample. [4]

(iii) Should these outliers be excluded from any further analysis? Briefly explain your answer. [2]

(iv) Any baby whose weight is below the 10th percentile is selected for careful monitoring. Use the diagram to determine the range of weights of the babies who are selected. [2]

12% of new-born babies require some form of special care. A maternity unit has 17 new-born babies. You may assume that these 17 babies form an independent random sample.

(v) Find the probability that

(A) exactly 2 of these 17 babies require special care, [3]

(B) more than 2 of the 17 babies require special care. [3]

(vi) On 100 independent occasions the unit has 17 babies. Find the expected number of occasions on which there would be more than 2 babies who require special care. [1]

7 When onion seeds are sown outdoors, on average two-thirds of them germinate. A gardener sows seeds in pairs, in the hope that at least one will germinate.

(i) Assuming that germination of one of the seeds in a pair is independent of germination of the other seed, find the probability that, if a pair of seeds is selected at random,

(A) both seeds germinate,

(B) just one seed germinates,

(C) neither seed germinates. [3]

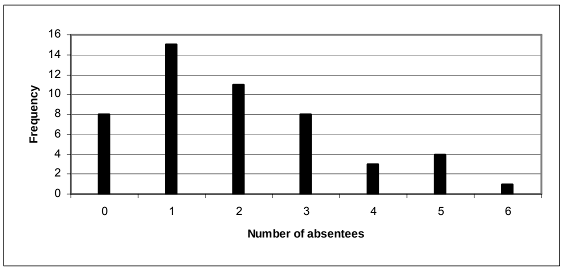
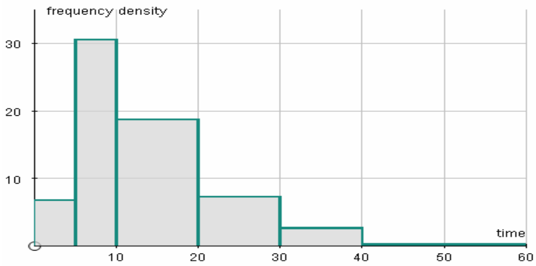
(ii) Explain why the assumption of independence is necessary in order to calculate the above probabilities. Comment on whether the assumption is likely to be valid. [2]

(iii) A pair of seeds is sown. Find the expectation and variance of the number of seeds in the pair which germinate. [3]

(iv) The gardener plants 200 pairs of seeds. If both seeds in a pair germinate, the gardener destroys one of the two plants so that only one is left to grow. Of the plants that remain after this, only 85% successfully grow to form an onion. Find the expected number of onions grown from the 200 pairs of seeds. [3]

If the seeds are sown in a greenhouse, the germination rate is higher. The seed manufacturing company claims that the germination rate is 90%. The gardener suspects that the rate will not be as high as this, and carries out a trial to investigate. 18 randomly selected seeds are sown in the greenhouse and it is found that 14 germinate.

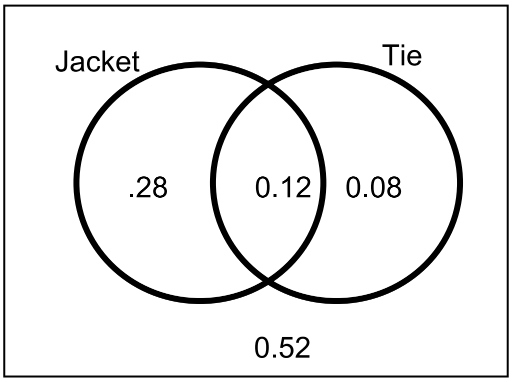
(v) Write down suitable hypotheses and carry out a test at the 5% level to determine whether there is any evidence to support the gardener's suspicions. [7]

Q 1 (i)	Mean = 127.6/13 = 9.8 Median = 8.6 Midrange = 14.5	M1 for 127.6/13 soi A1 CAO B1 CAO B1 CAO	**4**
(ii)	Mean slightly inflated due to the outlier Median good since it is not affected by the outlier Midrange poor as it is highly inflated due to the outlier	B1 B1 B1	**3**
		TOTAL	**7**
Q 2 (i)		G1 labelled linear scales on both axes G1 heights	**2**
(ii)	Mean = $\frac{99}{50}$ = 1.98 S_{xx} = $315 - \frac{99^2}{50}$ (= 118.98) *rmsd* = $\sqrt{\frac{118.98}{50}}$ = 1.54 *NB full marks for correct results from recommended method which is use of calculator functions*	B1 for mean M1 for attempt at S_{xx} A1 CAO	**3**
(iii)	New mean = 30 – 1.98 = 28.02 New rmsd = 1.54 (unchanged)	B1 FT their mean B1 FT their rmsd	**2**
		TOTAL	**7**
Q 3 (i)	time / freq / width / f dens 0- / 34 / 5 / 6.8 5- / 153 / 5 / 30.6 10- / 188 / 10 / 18.8 20- / 73 / 10 / 7.3 30- / 27 / 10 / 2.7 40- / 5 / 20 / 0.25 	M1 for fds A1 CAO Accept any suitable unit for fd such as eg freq per 5 mins. G1 linear scales on both axes and label G1 width of bars G1 height of bars	**5**
(ii)	Positive skewness	B1 CAO (indep)	**1**
		TOTAL	**6**

<table>
<tr><td>Q 4(i)</td><td><table><tr><td>r</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr><tr><td>P(X = r)</td><td>k</td><td>$3k$</td><td>$5k$</td><td>$7k$</td><td>$9k$</td><td>$11k$</td></tr></table>
$36k = 1$, so $k = \frac{1}{36}$</td><td>B1 for $3k$, $5k$, $7k$, $9k$

M1 for sum of six multiples of k = 1

A1 CAO MUST BE FRACTION IN SIMPLEST FORM</td><td>3</td></tr>
<tr><td>(ii)</td><td>E(X) =
$1\times\frac{1}{36}+2\times\frac{3}{36}+3\times\frac{5}{36}+4\times\frac{7}{36}+5\times\frac{9}{36}+6\times\frac{11}{36}=\frac{161}{36}=4.47$</td><td>M1 for $\Sigma\, rp$
A1 CAO</td><td>2</td></tr>
<tr><td>(iii)</td><td>P(X=16) = $6\times\left(\frac{1}{6}\right)^3$
$=\frac{6}{216}=\frac{1}{36}$</td><td>M1 for 6 ×
M1 indep for $\left(\frac{1}{6}\right)^3$

A1 CAO</td><td>3</td></tr>
<tr><td></td><td></td><td>TOTAL</td><td>8</td></tr>
<tr><td>Q 5(i)</td><td>P(jacket and tie) = 0.4 × 0.3 = 0.12</td><td>M1 for multiplying
A1 CAO</td><td>2</td></tr>
<tr><td>(ii)</td><td>
</td><td>G1 for two intersecting circles labelled

G1 for 0.12 and either 0.28 or 0.08

G1 for remaining probabilities

<u>Note</u> FT their 0.12 provided < 0.2</td><td>3</td></tr>
<tr><td>(iii)</td><td>(A) P(jacket or tie) =P(J) + P(T) – P(J∩T)
= 0.4 + 0.2 – 0.12 = 0.48
OR = 0.28 + 0.12 + 0.08 = 0.48

(B) P(no jacket or no tie) = 0.52 + 0.28 + 0.08 = 0.88
OR 0.6 + 0.8 – 0.52 = 0.88
OR 1 – 0.12 = 0.88</td><td>B1 FT

B2 FT
<u>Note</u> FT their 0.12 provided < 0.2</td><td>3</td></tr>
<tr><td></td><td></td><td>TOTAL</td><td>8</td></tr>
</table>

Q 6 (i)	Median = 3370 Q_1 = 3050 Q_3 = 3700 Inter-quartile range = 3700 – 3050 = 650	B1 B1 for Q_3 or Q_1 B1 for IQR	**3**
(ii)	Lower limit 3050 – 1.5 × 650 = 2075 Upper limit 3700 + 1.5 × 650 = 4675 Approx 40 babies below 2075 and 5 above 4675 so total 45	B1 B1 M1 (for either) A1	**4**
(iii)	Decision based on convincing argument: eg 'no, because there is nothing to suggest that they are not genuine data items and these data may influence health care provision'	E2 for convincing argument	**2**
(iv)	All babies below 2600 grams in weight	B2 CAO	**2**
(v)	(*A*) $X \sim \mathrm{B}(17, 0.12)$ $\mathrm{P}(X = 2) = \binom{17}{2} \times 0.12^2 \times 0.88^{15} = 0.2878$ (*B*) $\mathrm{P}(X > 2)$ $= 1 - (0.2878 + \binom{17}{1} \times 0.12 \times 0.88^{16} + 0.88^{17})$ $= 1 - (0.2878 + 0.2638 + 0.1138) = 0.335$	M1 $\binom{17}{2} \times p^2 \times q^{15}$ M1 indep $0.12^2 \times 0.88^{15}$ A1 CAO M1 for $\mathrm{P}(X{=}1) + \mathrm{P}(X{=}0)$ M1 for $1 - \mathrm{P}(X \le 2)$ A1 CAO	**3** **3**
(vi)	Expected number of occasions is 33.5	B1 FT	**1**
		TOTAL	**18**

Q 7 (i)	(*A*) P(both) = $\left(\frac{2}{3}\right)^2 = \frac{4}{9}$ (*B*) P(one) = $2 \times \frac{2}{3} \times \frac{1}{3} = \frac{4}{9}$ (*C*) P(neither) = $\left(\frac{1}{3}\right)^2 = \frac{1}{9}$	B1 CAO B1 CAO B1 CAO	**3**
(ii)	Independence necessary because otherwise, the probability of one seed germinating would change according to whether or not the other one germinates. May not be valid as the two seeds would have similar growing conditions eg temperature, moisture, etc. *NB Allow valid alternatives*	E1 E1	**2**
(iii)	Expected number = 2 × $\frac{2}{3} = \frac{4}{3}$ (= 1.33) E(X^2) = $0 \times \frac{1}{9} + 1 \times \frac{4}{9} + 4 \times \frac{4}{9} = \frac{20}{9}$ Var(X) = $\frac{20}{9} - \left(\frac{4}{3}\right)^2 = \frac{4}{9}$ = 0.444 *NB use of npq scores M1 for product, A1CAO*	B1 FT M1 for E(X^2) A1 CAO	**3**
(iv)	Expect $200 \times \frac{8}{9} = 177.8$ plants So expect 0.85 × 177.8 = 151 onions	M1 for 200 × $\frac{8}{9}$ M1 dep for × 0.85 A1 CAO	**3**
(v)	Let $X \sim B(18, p)$ Let p = probability of germination (for population) H_0: $p = 0.90$ H_1: $p < 0.90$ P($X \le 14$) = 0.0982 > 5% So not enough evidence to reject H_0 Conclude that there is not enough evidence to indicate that the germination rate is below 90%. Note: use of critical region method scores M1 for region {0,1,2,…, 13} M1 for 14 does not lie in critical region then A1 E1 as per scheme	B1 for definition of p B1 for H_0 B1 for H_1 M1 for probability M1 dep for comparison A1 E1 for conclusion in context	**7**
		TOTAL	**18**