RECOGNISING ACHIEVEMENT

ADVANCED SUBSIDIARY GCE **4766/01**

MATHEMATICS (MEI)

Statistics 1

TUESDAY 15 JANUARY 2008 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 in capital letters, your Centre Number and Candidate Number in the spaces provided on the Answer Booklet.
- Read each question carefully and make sure you know what you have to do before starting your answer.
- 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.
- 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 **4** printed pages.

Section A (36 marks)

1 Alice carries out a survey of the 28 students in her class to find how many text messages each sent on the previous day. Her results are shown in the stem and leaf diagram.

0	0	0	1	1	3	5	7	7	7	8	8
1	0	1	2	3	3	4	4	6	9		
2	0	1	3	3	7						
3	5	7									
4											
5	8										

Key: 2 | 3 represents 23

(i) Find the mode and median of the number of text messages. [2]

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

(iii) Alice is considering whether to use the mean or the median as a measure of central tendency for these data.

(A) In view of the skewness of the distribution, state whether Alice should choose the mean or the median. [1]

(B) What other feature of the distribution confirms Alice's choice? [1]

(iv) The mean number of text messages is 14.75. If each message costs 10 pence, find the total cost of all of these messages. [2]

2 Codes of three letters are made up using only the letters A, C, T, G. Find how many different codes are possible

(i) if all three letters used must be different, [3]

(ii) if letters may be repeated. [2]

3 Steve is going on holiday. The probability that he is delayed on his outward flight is 0.3. The probability that he is delayed on his return flight is 0.2, independently of whether or not he is delayed on the outward flight.

(i) Find the probability that Steve is delayed on his outward flight but not on his return flight. [2]

(ii) Find the probability that he is delayed on at least one of the two flights. [3]

(iii) Given that he is delayed on at least one flight, find the probability that he is delayed on both flights. [3]

4 A company is searching for oil reserves. The company has purchased the rights to make test drillings at four sites. It investigates these sites one at a time but, if oil is found, it does not proceed to any further sites. At each site, there is probability 0.2 of finding oil, independently of all other sites.

The random variable X represents the number of sites investigated. The probability distribution of X is shown below.

r	1	2	3	4
$\mathrm{P}(X = r)$	0.2	0.16	0.128	0.512

(i) Find the expectation and variance of X. [5]

(ii) It costs £45 000 to investigate each site. Find the expected total cost of the investigation. [1]

(iii) Draw a suitable diagram to illustrate the distribution of X. [2]

5 Sophie and James are having a tennis competition. The winner of the competition is the first to win 2 matches in a row. If the competition has not been decided after 5 matches, then the player who has won more matches is declared the winner of the competition.

For example, the following sequences are two ways in which Sophie could win the competition. (**S** represents a match won by Sophie; **J** represents a match won by James.)

SJSS **SJSJS**

(i) Explain why the sequence **SSJ** is not possible. [1]

(ii) Write down the other three possible sequences in which Sophie wins the competition. [3]

(iii) The probability that Sophie wins a match is 0.7. Find the probability that she wins the competition in no more than 4 matches. [4]

Section B (36 marks)

6 The maximum temperatures x degrees Celsius recorded during each month of 2005 in Cambridge are given in the table below.

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
9.2	7.1	10.7	14.2	16.6	21.8	22.0	22.6	21.1	17.4	10.1	7.8

These data are summarised by $n = 12$, $\Sigma x = 180.6$, $\Sigma x^2 = 3107.56$.

(i) Calculate the mean and standard deviation of the data. [3]

(ii) Determine whether there are any outliers. [3]

(iii) The formula $y = 1.8x + 32$ is used to convert degrees Celsius to degrees Fahrenheit. Find the mean and standard deviation of the 2005 maximum temperatures in degrees Fahrenheit. [3]

(iv) In New York, the monthly maximum temperatures are recorded in degrees Fahrenheit. In 2005 the mean was 63.7 and the standard deviation was 16.0. Briefly compare the maximum monthly temperatures in Cambridge and New York in 2005. [2]

The total numbers of hours of sunshine recorded in Cambridge during the month of January for each of the last 48 years are summarised below.

Hours h	$70 \leqslant h < 100$	$100 \leqslant h < 110$	$110 \leqslant h < 120$	$120 \leqslant h < 150$	$150 \leqslant h < 170$	$170 \leqslant h < 190$
Number of years	6	8	10	11	10	3

(v) Draw a cumulative frequency graph for these data. [5]

(vi) Use your graph to estimate the 90th percentile. [2]

7 A particular product is made from human blood given by donors. The product is stored in bags. The production process is such that, on average, 5% of bags are faulty. Each bag is carefully tested before use.

(i) 12 bags are selected at random.

(A) Find the probability that exactly one bag is faulty. [3]

(B) Find the probability that at least two bags are faulty. [2]

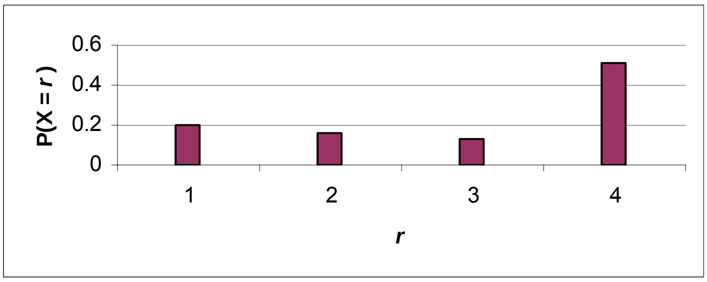
(C) Find the expected number of faulty bags in the sample. [2]

(ii) A random sample of n bags is selected. The production manager wishes there to be a probability of one third or less of finding any faulty bags in the sample. Find the maximum possible value of n, showing your working clearly. [3]

(iii) A scientist believes that a new production process will reduce the proportion of faulty bags. A random sample of 60 bags made using the new process is checked and one bag is found to be faulty. Write down suitable hypotheses and carry out a hypothesis test at the 10% level to determine whether there is evidence to suggest that the scientist is correct. [8]

4766 Statistics 1

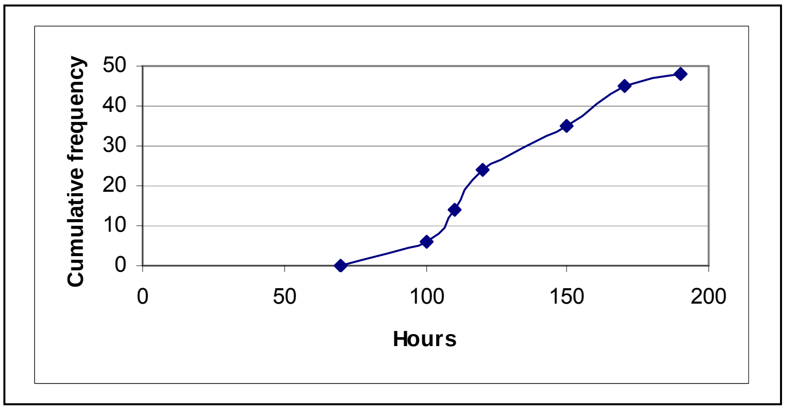
Q1 (i)	Mode = 7 Median = 12.5	B1 **cao** B1 **cao**	**2**
(ii)	Positive or positively skewed	E1	**1**
(iii)	(*A*) Median (*B*) There is a large outlier or possible outlier of 58 / figure of 58. Just 'outlier' on its own without reference to either 58 or large scores E0 Accept the large outlier affects the mean (more) E1	E1 cao E1indep	**2**
(iv)	There are 14.75 × 28 = 413 messages So total cost = 413 × 10 pence = £41.30	M1 for 14.75 × 28 but 413 can also imply the mark A1**cao**	**2**
		TOTAL	**7**
Q2 (i)	$\binom{4}{3} \times 3! = 4 \times 6 = 24$ codes or $^4P_3 = 24$ (M2 for $^4P_3$) *Or* 4 × 3 × 2 = 24	M1 for 4 M1 for ×6 A1	**3**
(ii)	$4^3 = 64$ codes	M1 for 4^3 A1 **cao**	**2**
		TOTAL	**5**
Q3 (i)	Probability = 0.3 × 0.8 = 0.24	M1 for 0.8 from (1-0.2) A1	**2**
(ii)	*Either:* P(*A*U*B*) = P(*A*) + P(*B*) – P(*A*∩*B*) = 0.3 + 0.2 – 0.3 × 0.2 = 0.5 – 0.06 = 0.44 *Or:* P(*A*U*B*) = 0.7 × 0.2 + 0.3 × 0.8 + 0.3 × 0.2 = 0.14 + 0.24 + 0.06 = 0.44 *Or:* P(*A*U*B*) = 1 – P(*A′* ∩*B′*) = 1 – 0.7 × 0.8 = 1 – 0.56 = 0.44	M1 for adding 0.3 and 0.2 M1 for **subtraction** of (0.3 × 0.2) A1 **cao** M1 either of first terms M1 for last term A1 M1 for 0.7 × 0.8 or 0.56 M1 for complete method as seen A1	**3**
(iii)	P(*A*|*B*) = $\frac{P(A \cap B)}{P(B)} = \frac{0.06}{0.44} = \frac{6}{44} = 0.136$	M1 for numerator of their 0.06 only M1 for 'their 0.44' in denominator A1 FT (must be valid p)	**3**
		TOTAL	**8**

Q4 (i)	E(X) = $1\times0.2+2\times0.16+3\times0.128+4\times0.512=2.952$ Division by 4 or other spurious value at end loses A mark E(X^2) = $1\times0.2+4\times0.16+9\times0.128+16\times0.512=10.184$ Var(X) = 10.184 – 2.952^2 = 1.47 (to 3 s.f.)	M1 for Σ *rp* (at least 3 terms correct) A1 cao M1 for Σ x^2p *at least 3 terms correct* M1 for E(X^2) – E(X)^2 Provided ans > 0 A1 FT their E(X) but not a wrong E(X^2)	**5**
(ii)	Expected cost = 2.952 × £45000 = £133000 (3sf)	B1 FT (no extra multiples / divisors introduced at this stage)	**1**
(iii)		G1 labelled linear scales G1 height of lines	**2**
		TOTAL	**8**
Q5 (i)	Impossible because the competition would have finished as soon as Sophie had won the first 2 matches	E1	**1**
(ii)	**SS, JSS, JSJSS**	B1, B1, B1 (-1 each error or omission)	**3**
(iii)	$0.7^2 + 0.3\times0.7^2 + 0.7\times0.3\times0.7^2 = 0.7399$ or 0.74(0) { 0.49 + 0.147 + 0.1029 = 0.7399}	M1 for any correct term M1 for any other correct term M1 for sum of all three correct terms A1 cao	**4**
		TOTAL	**8**

	Section B		
Q6 (i)	Mean = $\frac{180.6}{12}$ = 15.05 or 15.1 S_{xx} = $3107.56 - \frac{180.6^2}{12}$ or 3107.56 – 12(their 15.05)^2 = (389.53) s = $\sqrt{\frac{389.53}{11}}$ = 5.95 or better NB Accept answers seen without working (from calculator)	B1 for mean M1 for attempt at S_{xx} A1 cao	**3**
(ii)	$\bar{x}$ + 2s = 15.05 + 2 × 5.95 = 26.95 $\bar{x}$ – 2s = 15.05 – 2 × 5.95 = 3.15 So no outliers	M1 for attempt at either M1 for both A1 for limits and conclusion FT their mean and sd	**3**
(iii)	New mean = 1.8 × 15.05 + 32 = 59.1 New s = 1.8 × 5.95 = 10.7	B1FT M1 A1FT	**3**
(iv)	New York has a higher mean or ' is on average' higher (oe) New York has greater spread /range /variation or SD (oe)	E1FT using ^{0}F ($\bar{x}$ dep) E1FT using ^{0}F (σ dep)	**2**
(v)	See table and graph below. NB all G marks dep on attempt at cumulative frequencies.	B1 for all **correct** cumulative frequencies (may be implied from graph). **<u>Ignore cf of 0 at this stage</u>** G1 for linear scales (linear from 70 to 190) ignore x < 70 vertical: 0 to 50 but not beyond 100 (no inequality scales) G1 for labels G1 for points plotted as (UCB, their cf). <u>Ignore (70,0) at this stage</u>. **No** mid – point or LCB plots**.**	**5**
(vi)	NB All G marks dep on attempt at cumulative frequencies Line on graph at cf = 43.2(soi) or used 90th percentile = 166	G1 for joining all of 'their points'(line or smooth curve) **AND now including (70,0)** M1 for use of 43.2 A1FT but **dep on 3rd G mark earned**	**2**
		TOTAL	**18**

Q6 (v):

Upper bound	(70)	100	110	120	150	170	190
Cumulative frequency	(0)	6	14	24	35	45	48



Q7 **(i)**	$X \sim B(12, 0.05)$ (**A**) $P(X = 1) = \binom{12}{1} \times 0.05 \times 0.95^{11} = 0.3413$ OR from tables $0.8816 - 0.5404 = 0.3412$ (**B**) $P(X \geq 2) = 1 - 0.8816 = 0.1184$ (**C**) Expected number $E(X) = np = 12 \times 0.05 = 0.6$	M1 0.05×0.95^{11} M1 $\binom{12}{1} \times pq^{11}$ (p+q) = 1 A1 cao OR: M1 for 0.8816 seen and M1 for subtraction of 0.5404 A1 cao M1 for 1 – P(X ≤ 1) A1 cao M1 for 12×0.05 A1 cao (= 0.6 seen)	3 2 2
(ii)	*Either*: $1 - 0.95^n \leq ⅓$ $0.95^n \geq ⅔$ $n \leq \log ⅔ / \log 0.95$, so $n \leq 7.90$ Maximum $n = 7$ *Or:* (using tables with $p = 0.05$): $n = 7$ leads to $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.6983 = 0.3017$ (< ⅓) or 0.6983 (> 2/3) $n = 8$ leads to $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.6634 = 0.3366$ (> ⅓) or 0.6634 (< 2/3) Maximum $n = 7$ (total accuracy needed for tables) *Or:* (using trial and improvement): $1 - 0.95^7 = 0.3017$ (< ⅓) or $0.95^7 = 0.6983$ (> 2/3) $1 - 0.95^8 = 0.3366$ (> ⅓) or $0.96^8 = 0.6634$ (< 2/3) Maximum $n = 7$ (3 sf accuracy for calculations) NOTE: $n = 7$ unsupported scores SC1 only	M1 for equation in *n* M1 for use of logs A1 cao M1indep M1indep A1 cao dep on both M's M1indep (as above) M1indep (as above) A1 cao dep on both M's	 3
(iii)	Let $X \sim B(60, p)$ Let p = probability of a bag being faulty H_0: $p = 0.05$ H_1: $p < 0.05$ $P(X \leq 1) = 0.95^{60} + 60 \times 0.05 \times 0.95^{59} = 0.1916 > 10\%$ So not enough evidence to reject H_0 Conclude that there is not enough evidence to indicate that the new process reduces the failure rate or scientist incorrect/ wrong.	B1 for definition of *p* B1 for H_0 B1 for H_1 M1 A1 for probability M1 for comparison A1 E1	 8
		TOTAL	**18**