OCR
RECOGNISING ACHIEVEMENT

Oxford, Cambridge and RSA Examinations

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MEI STRUCTURED MATHEMATICS 2613/1

Statistics 1

Tuesday **12 JUNE 2001** Afternoon 1 hour 20 minutes

Additional materials:
Answer paper
Graph paper
MEI Examination Formulae and Tables (MF12)

TIME 1 hour 20 minutes

INSTRUCTIONS TO CANDIDATES

Write your Name, Centre Number and Candidate Number in the spaces provided on the answer booklet.

Answer **all** questions.

You are permitted to use a graphical calculator in this paper.

INFORMATION FOR CANDIDATES

The approximate allocation 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 60.

This question paper consists of 4 printed pages.

1 The cumulative frequency graph in Fig. 1 illustrates the distribution of the weights, x kg, of 150 newly-born babies in a hospital's maternity unit.

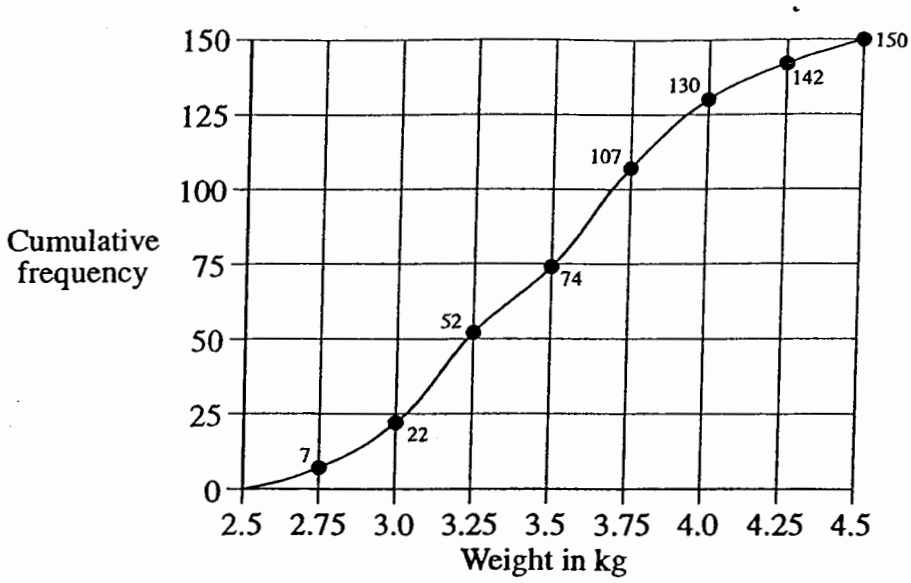


Fig. 1

(i) Use the cumulative frequency graph to estimate the median and interquartile range. [3]

(ii) Draw a box and whisker plot for the data. [2]

(iii) Copy and complete the following frequency table. Hence calculate an estimate of the mean weight of the 150 babies. [4]

Weight of baby (kg)	Frequency
$2.50 < x \leqslant 2.75$	7
$2.75 < x \leqslant 3.00$	15
$3.00 < x \leqslant 3.25$	
$3.25 < x \leqslant 3.50$	
$3.50 < x \leqslant 3.75$	
$3.75 < x \leqslant 4.00$	
$4.00 < x \leqslant 4.25$	
$4.25 < x \leqslant 4.50$	

During the same period of time, 18 babies were born in the special care baby unit of the hospital. The mean weight of all 168 babies was found to be 3.35 kg (to 3 significant figures).

(iv) Calculate an estimate of the mean weight of the 18 babies in the special care unit. [3]

(v) What effect will the addition of the 18 extra data items have on the median and on the interquartile range? Explain your reasoning. [3]

2 A market researcher has been commissioned to find out how many times per month members of the public visit an out-of-town shopping centre. She intends to take a quota sample of 80 adults in the 18 to 65 age range.

(i) Suggest suitable strata and describe how a quota sample may be taken. Give one advantage and one disadvantage of quota sampling. [5]

The results of her survey are as follows.

Number of visits	0	1	2	3	4	>4
Frequency	15	25	21	12	7	0

(ii) Determine the modal and median number of visits. [2]

(iii) Calculate the mean and standard deviation of the number of visits. [3]

Suppose that the probabilities for numbers of visits in the population being surveyed are the same as the proportions in the sample. Five adults are chosen at random from the population.

(iv) Find the probability that

(*A*) all five make at least one visit per month, [2]

(*B*) just two of the five people chosen make more than 2 visits per month. [3]

3 In each round of a quiz a contestant is asked a sequence of up to 5 questions. The round stops when the contestant has answered *two* questions wrongly or has attempted all five questions. One point is awarded for each correct answer.

As an example, Gerald gets question 1 wrong, questions 2 and 3 right, and question 4 wrong; the round then stops and he has scored 2 points.

(i) Give both ways in which Gerald would score exactly 1 point. [1]

Claire's probability of getting any particular question correct is 0.8, independently of other questions. She takes part in one round of the quiz.

(ii) Find the probability that Claire

(*A*) gets all 5 questions correct, [2]

(*B*) scores no points. [2]

(iii) Show that the probability that Claire scores exactly 4 points is 0.4096. [3]

(iv) Find the probability that Claire

(*A*) scores exactly 3 points, [3]

(*B*) is asked fewer than 5 questions. [4]

4 The probability that a certain type of cactus seed will germinate is p. In a long-term study, 1200 of these seeds were planted, of which 420 germinated.

(i) Write down an estimate of p. [1]

Malcolm plants 30 such seeds at the beginning of the growing season in two batches, each of 15 seeds.

(ii) Using the value of p found in part **(i)**, find the probability that

(*A*) exactly 10 of the 30 seeds germinate, [3]

(*B*) each batch produces at least 5 germinating seeds. [4]

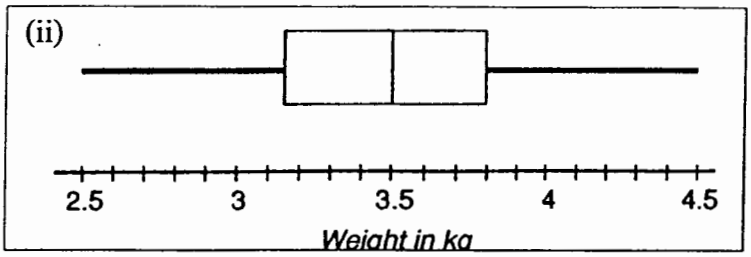
Charmaine develops a new variety of cactus seed which she claims has a better chance of germinating than the variety used by Malcolm. She plants 18 of her seeds and intends to carry out a hypothesis test, at the 5% level, to examine her claim.

(iii) Find the critical region for the test. Draw a diagram, in the form of a number line, showing which numbers of germinating seeds lead to acceptance of the null hypothesis and which do not. State a condition for this test to be valid. [7]

Mark Scheme

June 2001 2613 MEI Statistics 1

Question 1

(i)	Median = 3.50 to 3.51 kg IQR = (3.80 to 3.85) – (3.10 to 3.15) = (0.65 to 0.75) kg	B1 for median M1 for boundaries and subtraction A1	3
(ii)		G1 for box G1 for whiskers (allow whiskers from $D_1 = 2.9$ to $D_9 = 4.1$)	2
(iii)	Frequencies: 7, 15, 30, 22, 33, 23, 12, 8 Mid-interval points: 2.625, 2.875, 3.125, 3.375, ..., 4.375 (s.o.i.) Mean $= \frac{7\times 2.625+15\times 2.875+\ldots+8\times 4.375}{150}$ $= \frac{522.75}{150} = 3.49$ kg (3 s.f.)	B1 for frequencies (allow 1 error) B1 for mid-interval points (allow 2 errors) M1 for their sensible $\frac{\text{"522.75"}}{150}$ A1 **cao**	4
(iv)	Total mass of all babies = 168 x 3.35 = 562.8 kg Total mass of special care babies = 562.8 – 522.75 = 40.05 Mean mass of special care babies $= \frac{40.05}{18} = 2.23$ kg (3 s.f.)	M1 for attempt at total using their "522.75" M1 for a positive mean A1 **cao**	3
(v)	Median will decrease IQR will increase Extra data items *at lower end* of range	B1 B1 B1 (dependant on two previous B1 marks)	3
			15

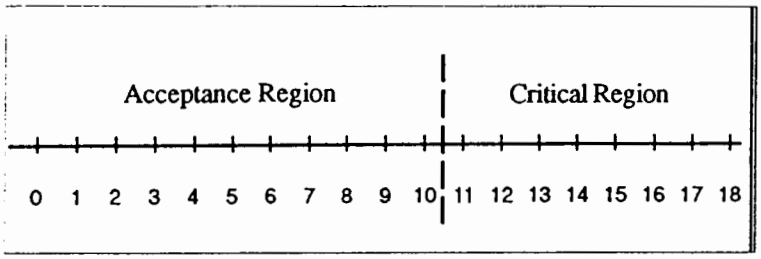
Question 2

(i)	Suitable strata: divide population into age groups Listing of age strata Idea of taking a fixed number from each strata (reference to equal numbers of males/females, etc.) Advantage: Any comment suggesting convenience, etc. e.g. no sampling frame required, ease of operation sample is representative/unrepresentative Disadvantage: e.g. sample non-random, biased difficulty in obtaining quota	E1 for idea of age strata E1 for listing age strata (could imply first E1) E1 for conduct E1 for advantage (dep. on any of 1st 3 E1) E1 for disadvantage (dep. on any of 1st 3 E1)	**5**
(ii)	Modal number of visits = 1 Median number of visits = 1.5	B1 for modal value B1 for median	**2**
(iii)	Mean $= \frac{131}{80} = 1.6$ (2 s.f.) S.d. $= \sqrt{\frac{329}{80} - 1.6375^2}$ = (1.19 to 1.20) (3 s.f.)	B1 for mean M1 for variance A1 **cao**	**3**
(iv)	(*A*) P(all 5 make at least one visit per month) = $\left(\frac{65}{80}\right)^5 = 0.35$ (2 s.f.) (*B*) P(just two make more than 2 visits per month) = ${}^5C_2 \times \left(\frac{19}{80}\right)^2 \times \left(\frac{61}{80}\right)^3 = 0.250$ (3 s.f.)	M1 for realising $p = \frac{65}{80}$ A1 **cao** B1 for ${}^5C_2$ M1 for $\frac{19}{80}$ *or* $\frac{61}{80}$ A1 **cao**	**2** **3**
			15

Question 3

(i)	Q1 right, Q2 wrong, Q3 wrong, *or* Q1 wrong, Q2 right, Q3 wrong	B1 for *both* answers (condone extra answer: Q1 W, Q2 W, Q3 R)	**1**
(ii)	(*A*) P(all questions correct) = $0.8^5 = 0.328$ (3 s.f.)	M1 for probability A1 **cao**	**2**
	(*B*) P(scores no points) = $0.2^2 = 0.04$	M1 for probability A1 **cao**	**2**
(iii)	P(Claire scores exactly 4 points) = $^5C_1 \times 0.8^4 \times 0.2 = 0.4096$ *or* $0.08192 \times 5 = 0.4096$ *or* $0.6723 - 0.2627 = 0.4096$ (from tables)	B1 for $^5C_1$ *or* $^5C_4$ B1 for $0.8^4 \times 0.2$ (indep.) *or* B1 for 0.08192 B1 for "x 5" *or* B1 for 0.6723 B1 for "– 0.2627" A1 for accuracy (dep.)	**3**
(iv)	(*A*) P(Claire scores exactly 3 points) = $^4C_1 \times 0.8^3 \times 0.2^2 = 0.082$ (2 s.f.)	B1 for $^4C_1$ *or* 4 B1 for $0.8^3 \times 0.2^2$ (indep.) A1 **cao**	**3**
	(*B*) P(Claire is asked fewer than 5 questions) = *Either* 1 – P(scores more than 2 points) $1 - (0.32768 + 0.4096 + 0.08192)$ $= 0.18$ (2 s.f.) *Or* P(asked 2 Q) + P(asked 3 Q) + P(asked 4 Q) $= 0.2^2 + 2 \times 0.8 \times 0.2^2 + 3 \times 0.8^2 \times 0.2^2$ $= 0.04 + 0.064 + 0.0768$ $= 0.18$ (2 s.f.)	M1 for idea/structure M1 for P(scores 3 pts) M1 for P(scores 5 pts) A1 **cao** *or* M1 for idea/structure M1 for 1 correct prob. * M1 for 2nd correct prob. * A1 **cao** * *dependent on 1st M1*	**4**
			15

Question 4

(i)	Estimate for $p = \frac{420}{1200} = \frac{7}{20} = 0.35$	B1 for probability as fraction, decimal or %	**1**
(ii)	*(A)* ${}^{30}C_{10} \times 0.35^{10} \times 0.65^{20} = 0.15$ (2 s.f.)	B1 for ${}^{30}C_{10}$ B1 for $0.35^{10} \times 0.65^{20}$ A1 **cao**	**3**
	(B) P(1 batch produces at least 5 germinating seeds) $= 1 - P(X < 5)$ $= 1 - 0.3519 = 0.6481$	M1 for use of tables with correct probabilities A1 **cao**	
	hence P(each batch produces at least 5 germinating seeds) $= 0.6481^2 = 0.420$ (3 s.f.)	M1 for their "0.6481^2" A1	**4**
(iii)	H_0: $p = 0.35$; H_1: $p > 0.35$	B1 for null hypothesis B1 for alternative hypothesis	
	$P(X \geq 11) = 1 - P(X \leq 10) = 1 - 0.9788$ $= 0.0212 < 0.05$ (5%) *and*	M1 for sight of 0.0212 or 0.9788	
	$P(X \geq 10) = 1 - P(X \leq 9) = 1 - 0.9403$ $= 0.0597 > 0.05$ (5%)	M1 for sight of 0.0597 or 0.9403	
	hence Critical Region = {11, 12, 13, 14, 15, 16, 17, 18} *or* {11, …. , 18} *or* $11 \leq X \leq 18$	B1 for critical region listed or on diagram	
		B1 for seeing the acceptance region on the diagram	
	Charmaine's seeds must be grown under the same conditions as Malcolm's *or* "seeds selected randomly"	E1 for explanation	**7**
			15

Examiner’s Report

Statistics 1 (2613 and 5513)

Whatever is going on in the examination questions, it seems clear that students understand the principles of sampling. However, candidates often do not define their population clearly enough for a reader to understand what is being investigated. A typical problem concerns the refinement. If the population is the students of one College which is sampled effectively then "taking students from another college" is not a refinement but a different task.

Another longstanding problem is a confusion between quality and precision. Care to ensure that the data collected are precisely recorded does not improve the quality which is to do with how representative the sample is.

The presentation of diagrams is quite variable. Students should note that computer software packages do not always produce what is required and so they should ask themselves whether the diagram is appropriate or even correct! They should certainly not conclude that because the computer did it it must be correct and relevant!