OCR
RECOGNISING ACHIEVEMENT

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education

MEI STRUCTURED MATHEMATICS 2614/1

Statistics 2

Tuesday **18 JANUARY 2005** Afternoon 1 hour 20 minutes

Additional materials:
Answer booklet
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 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 circumference of the trunk of a tree depends on its age.

The data in the table show the ages in years (x) and trunk circumferences in centimetres (y) of a random sample of seven trees of a particular species.

Age (x)	11	13	21	28	34	45	51
Circumference (y)	24.4	32.1	52.8	78.3	79.2	102.7	121.2

(i) Draw a scatter diagram on graph paper to illustrate these data. [2]

(ii) For these data,

$$n = 7,\ \sum x = 203,\ \sum y = 490.7,\ \sum x^2 = 7297,\ \sum y^2 = 42\,053.87,\ \sum xy = 17\,482.4.$$

Calculate the equation of the least squares regression line of y on x and plot this line on your scatter diagram. [6]

(iii) Use your equation to estimate the circumference of the trunk of a tree which is

(A) 25 years old,

(B) 100 years old.

Comment on the reliability of each of these estimates. [4]

(iv) Mark the residuals on your diagram.

Calculate the residual which has the largest absolute value. [3]

2 Jam is packed into jars at a food processing plant. The amount of jam, in grams, packed into a jar is Normally distributed with mean 460 and variance 25. Any jar containing less than the 454 grams stated on the label is described as 'underweight'.

(i) Find the probability that a randomly selected jar will be underweight. [3]

(ii) It is found that 95% of jars contain at least k grams of jam. Find the value of k. [3]

(iii) The jars, selected randomly, are packed in trays of 12. Find the probability that a randomly selected tray will have no jars that are underweight. [2]

(iv) 1000 trays are supplied to a supermarket.

(A) State the exact distribution of the number of trays which have no jars that are underweight.

(B) Use a suitable approximating distribution to find the probability that at least 200 of the 1000 trays will have no jars that are underweight. [7]

3 A manufacturer produces computer monitor screens in which the picture is composed of 780 000 pixels. On average 1 in 500 000 of these pixels is faulty. Let X represent the number of faulty pixels in one monitor screen.

(i) State the condition required for X to be binomially distributed. [1]

(ii) Explain why a Poisson distribution provides a good approximation for X. State the mean of this Poisson distribution. [2]

(iii) Hence find the probability that a randomly selected monitor screen has

(A) exactly one faulty pixel,

(B) at least two faulty pixels. [5]

(iv) A retailer orders a batch of five monitor screens from the manufacturer. (The batch may be regarded as a random sample.)

(A) Find the probability that exactly one of the five monitor screens has at least two faulty pixels.

(B) Find the probability that there are at most 10 faulty pixels in total in the batch. [4]

(v) The manufacturer wishes to improve quality so that 90% of the monitor screens have no faulty pixels at all. To what value must the manufacturer reduce the probability of a pixel being faulty in order to achieve this? [3]

4 In a promotional scheme for packets of a new breakfast cereal, free models of famous tennis players are given away. There are four different tennis players and each packet contains just one model. It is equally likely that any of the four models is in a given packet.

Simon opens a number of packets at random and stops when he has got the same model twice. Let X represent the number of packets that Simon opens.

(i) Explain why $\mathrm{P}(X = 1) = 0$ and why $\mathrm{P}(X = 2) = \frac{1}{4}$. [2]

The probability distribution of X is shown in the table.

r	1	2	3	4	5
$\mathrm{P}(X = r)$	0	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{9}{32}$	p

(ii) Calculate the value of p. [1]

(iii) Illustrate the probability distribution in a suitable diagram. [2]

(iv) Find the mean and variance of X. [4]

Mary's favourite tennis player is one of the four who is modelled. She decides to keep opening packets until she has found a model of this player. Y represents the number of packets that Mary needs to open.

(v) Show that the probability distribution of Y is given by

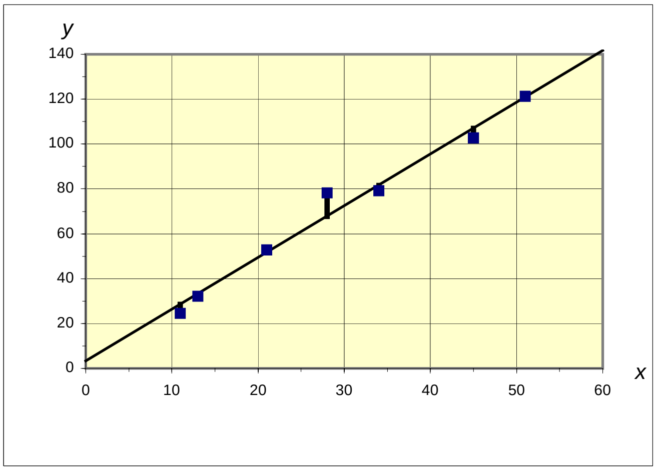
$$\mathrm{P}(Y = r) = \left(\tfrac{3}{4}\right)^{r-1} \times \tfrac{1}{4} \qquad \text{for } r = 1, 2, 3, \ldots\ .$$ [2]

(vi) Find the probability that Simon opens at least twice as many packets as Mary. [4]

Mark Scheme

January 2005 2614 MEI Statistics 2

Question 1

(i)		G1 for scales *dep on some plotted points* G1 for all points plotted (*dep on scales)*	**2**
(ii)	$\bar{x} = 29,\ \ \bar{y} = 70.1$ $b = \frac{Sxy}{Sxx} = \frac{17482.4 - 203 \times 490.7/7}{7297 - 203^2/7} = \frac{3252.1}{1410} = 2.3064$ OR $b = \frac{17482.4/7 - 29 \times 70.1}{7297/7 - 29^2} = \frac{464.6}{201.4} = 2.3064$ hence least squares regression line is: $y - \bar{y} = b(x - \bar{x})$ $\Rightarrow\ y - 70.1 = 2.306(x - 29)$ $\Rightarrow\ y = 2.306x + 3.21$ NB Use of calculator function can score SC4 Line plotted on graph	M1 for use of $\bar{x}$ and $\bar{y}$ M1 for correct structure of gradient b M1 indep for equation of line using their b and $\bar{x}$, $\bar{y}$ A1 (explicit) **cao** G1 for line through (29,70.1) or (0,3.2) G1 CAO for fully correct line	**6**
(iii)	(A) $x = 25 \Rightarrow y = 2.306{\times}25 + 3.21 = 60.9$ (B) $x = 100 \Rightarrow y = 2.306{\times}100 + 3.21 = 233.8$ First prediction is likely to be fairly accurate since, it is within the data range, and the fit looks good. Second prediction is well outside the data range, so may be unreliable. SC1 if no reasons given, or if don't say which is which	M1 for prediction(s) A1FT for both answers E1 for first comment E1 for second comment	**4**
(iv)	Residuals on scatter diagram Largest residual is at $x = 28$ Largest residual = $78.3 - (2.306{\times}28 + 3.21)$ $= 78.3 - 67.8 = 10.5$	G1 for at least 3 residuals M1 for finding any residual A1 CAO (A0 for -10.5)	**3**
			15

January 2005 2614 MEI Statistics 2

Question 2

(i)	$\mathrm{P}(X < 454) = \mathrm{P}\left(Z < \frac{454-460}{5}\right)$ $= \mathrm{P}(Z < -1.2)$ $= \Phi(-1.2) = 1 - \Phi(1.2)$ $= 1 - 0.8849 = 0.1151$	M1 for standardizing M1 for correct tail and use of tables A1 (to at least 2 s.f.)	**3**
(ii)	From tables $\Phi^{-1}(0.95) = 1.645$ $\frac{k-460}{5} = -1.645$ $k = 460 + 5 \times (-1.645)$ $k = 451.8$	B1 for 1.645 seen M1 for equation in k with sensible negative z-value A1 CAO (to at least 1 d.p., ISW if subsequently rounded)	**3**
(iii)	P(No jars underweight) $= 0.8849^{12}$ $= 0.2305$	M1 for 0.8849^{12} A1FT (to at least 2 s.f.)	**2**
(iv)	(A) Binomial (1000, 0.2305)	B1 for binomial B1FT for n, p	**2**
	(B) Use N (230.5, 177.4) $\mathrm{P}(X \geq 200) = \mathrm{P}\left(Z \geq \frac{199.5-230.5}{13.32}\right)$ $= \mathrm{P}(Z \geq -2.327)$ $= 1 - \Phi(-2.327) = \Phi(2.327)$ $= 0.9900$	B1 indep for Normal approx. B1FT for μ and σ^2 B1 CAO for continuity correction M1 for calculation with correct tail, using their μ and σ A1 CAO	**5**
			15

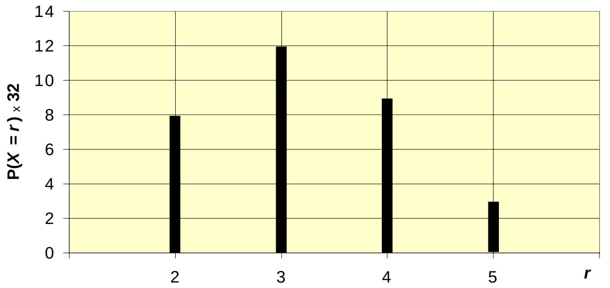
January 2005 2614 MEI Statistics 2

Question 3

(i)	Independence	E1 for independence seen	**1**
(ii)	n is large and p is small $\lambda = 780000 \times 0.000002 = 1.56$	E1 B1 for value of λ	**2**
(iii)	(A) $\mathrm{P}(X = 1) = e^{-1.56}\dfrac{1.56^1}{1!} = 0.3278 = 0.328$ (3 s.f.) (B) $\mathrm{P}(X = 0) = e^{-1.56}\dfrac{1.56^0}{0!} = 0.2101$ $\mathrm{P}(X \geq 2) = 1 - 0.3278 - 0.2101 = 1 - 0.5379 = 0.4621$	M1 for probability calc. A1 FT min 3SF M1 for P(X=0) using their <u>unrounded</u> λ M1 for 1 – (P(X=0)+ P(X=1)) A1 CAO min 3SF	**5**
(iv)	(A) P(just one monitor has at least two faulty) $= 5 \times 0.4621 \times 0.5379^4 = 0.1934$ (B) New mean $= 5 \times 1.56 = 7.8$ From tables $\mathrm{P}(X \leq 10) = 0.8352$	M1 for binomial prob: A1 FT M1 for 5λ A1 CAO	**4**
(v)	P(None faulty) = 0.9 *EITHER* $(1-p)^{780000} = 0.9$ $(1-p) = 0.9^{1/780000} = 0.999999864$ $p = 1 - 0.999999864 = 0.000000136 = 1.36\times10^{-7}$ *OR* $e^{-\lambda} = 0.9$ $\lambda = -\ln 0.9 = 0.1054$ $780000\,p = 0.1054$ $p = 0.1054 \div 780000 = 1.35\times10^{-7}$	M1 for equation for p M1 for $0.9^{1/780000}$ A1 CAO min 2sf M1 for equation for p or $e^{-\lambda}$ M1 for equation in p A1 CAO min 2sf	**3**
			15

January 2005 2614 MEI Statistics 2

Question 4

(i)	P(X=1) = 0 since at least two models are needed to get a repeat P(X=2) = $\frac{1}{4}$, since once the first model is known, there are four outcomes for the second model, of which only one is the same as the first.	E1 E1 allow alternatives	**2**
(ii)	$\frac{1}{4}+\frac{3}{8}+\frac{9}{32}+p=1,\quad \frac{29}{32}+p=1,$ $p=\frac{3}{32}=0.09375$	B1 CAO for p	**1**
(iii)		G1 for scaled axes dependent on attempt at representing data G1FT for lines in proportion	**2**
(iv)	$E(X) = \Sigma\, r\, P(X=r)$ $= 2\times\frac{1}{4}+3\times\frac{3}{8}+4\times\frac{9}{32}+5\times\frac{3}{32}=\frac{103}{32}=3\frac{7}{32}=3.22$ $\text{Var}(X) = E(X^2)-[E(X)]^2$ $= 4\times\frac{1}{4}+9\times\frac{3}{8}+16\times\frac{9}{32}+25\times\frac{3}{32}-\left(\frac{103}{32}\right)^2$ $=\frac{359}{32}-\left(\frac{103}{32}\right)^2=\frac{879}{1024}=0.858$ (to 3 s.f.)	M1 for E(X) if $0<p<1$ A1 FT if from $\Sigma\, p = 1$ M1 for $\Sigma\, x^2p$ if $0\le p\le 1$ A1 CAO	**4**
(v)	If Mary opens r packets, the first r-1 must be not her favourite, hence $\left(\frac{3}{4}\right)^{r-1}$ The final packet must contain her favourite, so multiply by $\frac{1}{4}$ to get $P(Y=r)=\left(\frac{3}{4}\right)^{r-1}\times\frac{1}{4}$	E1 E1	**2**
(vi)	Simon opens at least twice as many as Mary if either: (A) Mary opens one in which case Simon can open any number (≥2) *or* (B) Mary opens two and Simon opens 4 or 5 $P(A) = \frac{1}{4}=0.25$ $P(B)=\frac{3}{4}\times\frac{1}{4}\times\left(\frac{9}{32}+\frac{3}{32}\right)$ $=\frac{3}{16}\times\frac{12}{32}=\frac{9}{128}=0.0703$ $P(A)+P(B)=\frac{1}{4}+\frac{9}{128}=\frac{41}{128}=0.320$ (3 s.f.)	M1 for any correct pair M1 for all six pairs correct and no extras M1 for sum of at least two pairs A1 CAO min 2SF NB Allow alternative valid methods	**4**
			15

Examiner's Report