## OXFORD CAMBRIDGE AND RSA EXAMINATIONS

## Advanced Subsidiary General Certificate of Education Advanced General Certificate of Education

## MEI STRUCTURED MATHEMATICS

## 4767

Statistics 2
Monday 22 MAY $2006 \quad$ Morning 1 hour 30 minutes

Additional materials:
8 page answer booklet
Graph paper
MEI Examination Formulae and Tables (MF2)

## TIME 1 hour 30 minutes

## 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.
- 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.
- $\quad$ The total number of marks for this paper is 72.

1 A low-cost airline charges for breakfasts on its early morning flights. On average, $10 \%$ of passengers order breakfast.
(i) Find the probability that, out of 8 randomly selected passengers, exactly 1 orders breakfast.
(ii) Use a suitable Poisson approximating distribution to find the probability that the number of breakfasts ordered by 30 randomly selected passengers is
(A) exactly 6 ,
(B) at least 8 .
(iii) State the conditions under which the use of a Poisson distribution is appropriate as an approximation to a binomial distribution.
(iv) The aircraft carries 120 passengers and the flight is always full. Find the mean $\mu$ and variance $\sigma^{2}$ of a Normal approximating distribution suitable for modelling the total number of passengers on the flight who order breakfast.
(v) Use your Normal approximating distribution to calculate the probability that more than 15 breakfasts are ordered on a particular flight.
(vi) The airline wishes to be at least $99 \%$ certain that the plane will have sufficient breakfasts for all passengers who order them. Find the minimum number of breakfasts which should be carried on each flight.

2 The head circumference of 3-year-old boys is known to be Normally distributed with mean 49.7 cm and standard deviation 1.6 cm .
(i) Find the probability that the head circumference of a randomly selected 3-year-old boy will be
(A) over 51.5 cm ,
(B) between 48.0 and 51.5 cm .
(ii) Four 3-year-old boys are selected at random. Find the probability that exactly one of them has head circumference between 48.0 and 51.5 cm .
(iii) The head circumference of 3-year-old girls is known to be Normally distributed with mean $\mu$ and standard deviation $\sigma$. Given that $60 \%$ of 3 -year-old girls have head circumference below 49.0 cm and $30 \%$ have head circumference below 47.5 cm , find the values of $\mu$ and $\sigma$.

A nutritionist claims that boys who have been fed on a special organic diet will have a larger mean head circumference than other boys. A random sample of ten 3 -year-old boys who have been fed on this organic diet is selected. It is found that their mean head circumference is 50.45 cm .
(iv) Using the null and alternative hypotheses $\mathrm{H}_{0}: \mu=49.7 \mathrm{~cm}, \mathrm{H}_{1}: \mu>49.7 \mathrm{~cm}$, carry out a test at the $10 \%$ significance level to examine the nutritionist's claim. Explain the meaning of $\mu$ in these hypotheses. You may assume that the standard deviation of the head circumference of organically fed 3-year-old boys is 1.6 cm .

3 A student is investigating the relationship between the length $x \mathrm{~mm}$ and circumference $y \mathrm{~mm}$ of plums from a large crop. The student measures the dimensions of a random sample of 10 plums from this crop. Summary statistics for these dimensions are as follows.

$$
\begin{aligned}
& \sum x=4715 \quad \sum y=13175 \quad \sum x^{2}=2237725 \\
& \sum y^{2}=17455825 \quad \sum x y=6235575 \quad n=10
\end{aligned}
$$

(i) Calculate the sample product moment correlation coefficient.
(ii) Carry out a hypothesis test at the $5 \%$ significance level to determine whether there is any correlation between length and circumference of plums from this crop. State your hypotheses clearly, defining any symbols which you use.
(iii) (A) Explain the meaning of a 5\% significance level.
(B) State one advantage and one disadvantage of using a $1 \%$ significance level rather than a $5 \%$ significance level in a hypothesis test.

The student decides to take another random sample of 10 plums. Using the same hypotheses as in part (ii), the correlation coefficient for this second sample is significant at the $5 \%$ level. The student decides to ignore the first result and concludes that there is correlation between the length and circumference of plums in the crop.
(iv) Comment on the student's decision to ignore the first result. Suggest a better way in which the student could proceed.

4 A survey of a random sample of 250 people is carried out. Their musical preferences are categorized as pop, classical or jazz. Their ages are categorized as under 25,25 to 50 , or over 50 . The results are as follows.

|  |  | Musical preference |  |  | Row <br> totals |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Classical | Jazz |  |  |
| Age group | Under 25 | 57 | 15 | 12 | $\mathbf{8 4}$ |
|  | $25-50$ | 43 | 21 | 21 | $\mathbf{8 5}$ |
|  | Over 50 | 22 | 32 | 27 | $\mathbf{8 1}$ |
| Column totals |  |  | $\mathbf{1 2 2}$ | $\mathbf{6 8}$ | $\mathbf{6 0}$ |
| $\mathbf{2 5 0}$ |  |  |  |  |  |

(i) Carry out a test at the $5 \%$ significance level to examine whether there is any association between musical preference and age group. State carefully your null and alternative hypotheses. Your working should include a table showing the contributions of each cell to the test statistic.
(ii) Discuss briefly how musical preferences vary between the age groups, as shown by the contributions to the test statistic.

| (i) | $\begin{aligned} \mathrm{P}(X=1) & =8 \times 0.1^{1} \times 0.9^{7} \\ & =0.383 \end{aligned}$ | M1 for binomial probability $\mathrm{P}(X=1)$ A1 (at least 2sf) CAO | 2 |
| :---: | :---: | :---: | :---: |
| (ii) | $\lambda=30 \times 0.1=3$ <br> (A) $\mathrm{P}(X=6)=\mathrm{e}^{-3} \frac{3^{6}}{6!}=0.0504$ (3 s.f.) or from tables $=0.9665-0.9161=0.0504$ <br> (B) Using tables: $\mathrm{P}(X \geq 8)=1-\mathrm{P}(X \leq 7)$ $=1-0.9881=0.0119$ | B1 for mean SOI <br> M1 for calculation or use of tables to obtain $\mathrm{P}(X=6)$ <br> A1 (at least 2sf) CAO <br> M1 for correct <br> probability calc' <br> A1 (at least 2sf) CAO | 1 2 2 |
| (iii) | $n$ is large and $p$ is small | B1, B1 <br> Allow appropriate numerical ranges | 2 |
| (iv) | $\begin{aligned} & \mu=n p=120 \times 0.1=12 \\ & \sigma^{2}=n p q=120 \times 0.1 \times 0.9=10.8 \end{aligned}$ | $\begin{aligned} & \text { B1 } \\ & \text { B1 } \end{aligned}$ | 2 |
| (v) | $\begin{aligned} & \mathrm{P}(X>15.5)=\mathrm{P}\left(Z>\frac{15.5-12}{\sqrt{10.8}}\right) \\ & =\mathrm{P}(Z>1.065)=1-\Phi(1.065)=1-0.8566 \\ & =0.1434 \end{aligned}$ <br> NB Allow full marks for use of $\mathrm{N}(12,12)$ as an approximation to Poisson(12) leading to $1-\Phi(1.010)=1$ $-0.8438=0.1562$ | B1 for correct continuity correction. <br> M1 for probability using correct tail A1 cao, (but FT wrong or omitted CC) | 3 |
| (vi) | From tables $\Phi^{-1}(0.99)=2.326$ $\begin{aligned} & \frac{x+0.5-12}{\sqrt{10.8}} \geq 2.326 \\ & x=11.5+2.326 \times \sqrt{10.8} \geq 19.14 \end{aligned}$ <br> So 20 breakfasts should be carried <br> NB Allow full marks for use of $N(12,12)$ leading to $x \geq 11.5+2.326 \times \sqrt{12}=19.56$ | B1 for 2.326 seen <br> M1 for equation in $x$ and positive $z$-value <br> A1 CAO (condone 19.64) <br> A1FT for rounding appropriately (i.e. round up if c.c. used o/w rounding should be to nearest integer) | 4 |
|  |  |  | 18 |

## Question 2

| (i) | $\begin{aligned} & X \sim \mathrm{~N}\left(49.7,1.6^{2}\right) \\ & (A) \quad \mathrm{P}(X>51.5)=\mathrm{P}\left(Z>\frac{51.5-49.7}{1.6}\right) \\ & \quad=\mathrm{P}(Z>1.125) \\ & \quad=1-\Phi(1.125)=1-0.8696=0.1304 \end{aligned}$ $\begin{aligned} & \text { (B) } \quad \begin{aligned} \mathrm{P} & (X<48.0)=\mathrm{P}\left(Z<\frac{48.0-49.7}{1.6}\right) \\ \quad & =\mathrm{P}(Z<-1.0625)=1-\Phi(1.0625) \\ & =1-0.8560=0.1440 \\ \mathrm{P}(48.0 & <X<51.5)=1-0.1304-0.1440=0.7256 \end{aligned} \end{aligned}$ | M1 for standardizing <br> M1 for prob. calc. <br> A1 (at least 2 s.f.) <br> M1 for appropriate prob' calc. <br> A1 (0.725-0.726) | 5 |
| :---: | :---: | :---: | :---: |
| (ii) | P (one over 51.5, three between 48.0 and 51.5) $=\binom{4}{1} \times 0.7256 \times 0.2744^{3}=0.0600$ | M1 for coefficient <br> M1 for $0.7256 \times$ $0.2744^{3}$ <br> A1 FT (at least 2 sf ) | 3 |
| (iii) | From tables, $\begin{aligned} & \Phi^{-1}(0.60)=0.2533, \Phi^{-1}(0.30)=-0.5244 \\ & 49.0=\mu+0.2533 \sigma \\ & 47.5=\mu-0.5244 \sigma \\ & 1.5=0.7777 \sigma \\ & \sigma=1.929, \mu=48.51 \end{aligned}$ | B1 for 0.2533 or 0.5244 seen M1 for at least one correct equation $\mu \& \sigma$ <br> M1 for attempt to solve two correct equations A1 CAO for both | 4 |
| (iv) | Where $\mu$ denotes the mean circumference of the entire population of organically fed 3 -year-old boys. $n=10,$ <br> Test statistic $Z=\frac{50.45-49.7}{1.6 / \sqrt{10}}=\frac{0.75}{0.5060}=1.482$ <br> $10 \%$ level 1 tailed critical value of $z$ is 1.282 <br> $1.482>1.282$ so significant. <br> There is sufficient evidence to reject $\mathrm{H}_{0}$ and conclude that organically fed 3 -year-old boys have a higher mean head circumference. | E1 <br> M1 <br> A1(at least 3sf) <br> B1 for 1.282 <br> M1 for comparison leading to a conclusion <br> A1 for conclusion in context | 6 |
|  |  |  | 18 |

Question 3

| (i) | EITHER: <br> OR: | M1 for method for $S_{x y}$ <br> M1 for method for at least one of $S_{x x}$ or $S_{y y}$ <br> A1 for at least one of $\mathrm{S}_{x y}, \mathrm{~S}_{x x}$ or $\mathrm{S}_{y y}$ correct <br> M1 for structure of $r$ A1 (0.62 to 0.63) <br> M1 for method for cov ( $x, y$ ) <br> M1 for method for at least one msd <br> A1 for at least one of $\mathrm{S}_{x y}, \mathrm{~S}_{x x}$ or $\mathrm{S}_{y y}$ correct <br> M1 for structure of $r$ A1 (0.62 to 0.63) | 5 |
| :---: | :---: | :---: | :---: |
| (ii) | $\mathrm{H}_{0}: \rho=0$ <br> $\mathrm{H}_{1}: \rho \neq 0$ (two-tailed test) <br> where $\rho$ is the population correlation coefficient <br> For $n=10,5 \%$ critical value $=0.6319$ <br> Since $0.624<0.6319$ we cannot reject $\mathrm{H}_{0}$ : <br> There is not sufficient evidence at the $5 \%$ level to suggest that there is any correlation between length and circumference. | B 1 for $\mathrm{H}_{0}, \mathrm{H}_{1}$ in symbols B1 for defining $\rho$ <br> B1FT for critical value <br> M1 for sensible comparison leading to a conclusion <br> A1 FT for result <br> B1 FT for conclusion in context | 6 |
| (iii) | (A) This is the probability of rejecting $\mathrm{H}_{0}$ when it is in fact true. <br> (B) Advantage of $1 \%$ level - less likely to reject $\mathrm{H}_{0}$ when it is true. <br> Disadvantage of $1 \%$ level - less likely to accept $\mathrm{H}_{1}$ when $\mathrm{H}_{0}$ is false. | B1 for ' $\mathrm{P}\left(\right.$ reject $\mathrm{H}_{0}$ )' B1 for 'when true' <br> B1, B1 Accept answers in context | 2 |


|  |  |  |  |
| :--- | :--- | :--- | :---: |
| (iv) | The student's approach is not valid. <br> If a statistical procedure is repeated with a new <br> sample, we should not simply ignore one of the two <br> outcomes. <br> The student could combine the two sets of data into a <br> single set of twenty measurements. | E1 - allow suitable <br> alternatives. <br> E1 for combining <br> samples. | $\mathbf{3}$ |
|  |  |  | $\mathbf{1 8}$ |

## Question 4



| (ii) | The values of 6.25 and 7.77 show that under 25's <br> have a strong positive association with pop whereas <br> over 50's have a strong negative association with <br> pop. | B1, B1 <br> for specific reference <br> to a value from the <br> table of contributions <br> The values of 4.51 and 2.94 show that over 50's haved by an <br> appropriate comment <br> a reasonably strong positive association with both <br> classical and jazz. <br> The values of 2.70 and 3.30 show that under 25's <br> have a reasonably strong negative associations with <br> second value for <br> both classical and jazz. <br> The $25-50$ group's preferences differ very little from <br> the overall preferences. | B1, B1 (as above for <br> third value) |
| :--- | :--- | :--- | :--- |

