

OXFORD CAMBRIDGE AND RSA EXAMINATIONS

**Advanced Subsidiary General Certificate of Education
Advanced General Certificate of Education**

MEI STRUCTURED MATHEMATICS

4768

Statistics 3

Wednesday **24 MAY 2006** Afternoon 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.
- The total number of marks for this paper is 72.

This question paper consists of 5 printed pages and 3 blank pages.

2

- 1 Design engineers are simulating the load on a particular part of a complex structure. They intend that the simulated load, measured in a convenient unit, should be given by the random variable X having probability density function

$$f(x) = 12x^3 - 24x^2 + 12x, \quad 0 \leq x \leq 1.$$

- (i) Find the mean and the mode of X . [6]

- (ii) Find the cumulative distribution function $F(x)$ of X .

Verify that $F\left(\frac{1}{4}\right) = \frac{67}{256}$, $F\left(\frac{1}{2}\right) = \frac{11}{16}$ and $F\left(\frac{3}{4}\right) = \frac{243}{256}$. [3]

The engineers suspect that the process for generating simulated loads might not be working as intended. To investigate this, they generate a random sample of 512 loads. These are recorded in a frequency distribution as follows.

Load x	$0 \leq x \leq \frac{1}{4}$	$\frac{1}{4} < x \leq \frac{1}{2}$	$\frac{1}{2} < x \leq \frac{3}{4}$	$\frac{3}{4} < x \leq 1$
Frequency	126	209	131	46

- (iii) Use a suitable statistical procedure to assess the goodness of fit of X to these data. Discuss your conclusions briefly. [9]

3

- 2 A bus route runs from the centre of town A through the town's urban area to a point B on its boundary and then through the country to a small town C. Because of traffic congestion and general road conditions, delays occur on both the urban and the country sections. All delays may be considered independent.

The scheduled time for the journey from A to B is 24 minutes. In fact, journey times over this section are given by the Normally distributed random variable X with mean 26 minutes and standard deviation 3 minutes.

The scheduled time for the journey from B to C is 18 minutes. In fact, journey times over this section are given by the Normally distributed random variable Y with mean 15 minutes and standard deviation 2 minutes.

Journey times on the two sections of route may be considered independent. The timetable published to the public does not show details of times at intermediate points; thus, if a bus is running early, it merely continues on its journey and is not required to wait.

- (i) Find the probability that a journey from A to B is completed in less than the scheduled time of 24 minutes. [3]
- (ii) Find the probability that a journey from A to C is completed in less than the scheduled time of 42 minutes. [3]
- (iii) It is proposed to introduce a system of bus lanes in the urban area. It is believed that this would mean that the journey time from A to B would be given by the random variable $0.85X$. Assuming this to be the case, find the probability that a journey from A to B would be completed in less than the currently scheduled time of 24 minutes. [3]
- (iv) An alternative proposal is to introduce an express service. This would leave out some bus stops on both sections of the route and its overall journey time from A to C would be given by the random variable $0.9X + 0.8Y$. The scheduled time from A to C is to be given as a whole number of minutes. Find the least possible scheduled time such that, with probability 0.75, buses would complete the journey on time or early. [6]
- (v) A programme of minor road improvements is undertaken on the country section. After their completion, it is thought that the random variable giving the journey time from B to C is still Normally distributed with standard deviation 2 minutes. A random sample of 15 journeys is found to have a sample mean journey time from B to C of 13.4 minutes. Provide a two-sided 95% confidence interval for the population mean journey time from B to C. [3]

3 An employer has commissioned an opinion polling organisation to undertake a survey of the attitudes of staff to proposed changes in the pension scheme. The staff are categorised as management, professional and administrative, and it is thought that there might be considerable differences of opinion between the categories. There are 60, 140 and 300 staff respectively in the categories. The budget for the survey allows for a sample of 40 members of staff to be selected for in-depth interviews.

- (i) Explain why it would be unwise to select a simple random sample from all the staff. [2]
- (ii) Discuss whether it would be sensible to consider systematic sampling. [3]
- (iii) What are the advantages of stratified sampling in this situation? [2]
- (iv) State the sample sizes in each category if stratified sampling with as nearly as possible proportional allocation is used. [1]

The opinion polling organisation needs to estimate the average wealth of staff in the categories, in terms of property, savings, investments and so on. In a random sample of 11 professional staff, the sample mean is £345 818 and the sample standard deviation is £69 241.

- (v) Assuming the underlying population is Normally distributed, test at the 5% level of significance the null hypothesis that the population mean is £300 000 against the alternative hypothesis that it is greater than £300 000. Provide also a two-sided 95% confidence interval for the population mean. [10]

4 A company has many factories. It is concerned about incidents of trespassing and, in the hope of reducing if not eliminating these, has embarked on a programme of installing new fencing.

(i) Records for a random sample of 9 factories of the numbers of trespass incidents in typical weeks before and after installation of the new fencing are as follows.

Factory	A	B	C	D	E	F	G	H	I
Number before installation	8	12	6	4	14	22	4	13	14
Number after installation	6	11	0	1	18	10	11	5	4

Use a Wilcoxon test to examine at the 5% level of significance whether it appears that, on the whole, the number of trespass incidents per week is lower after the installation of the new fencing than before. [9]

(ii) Records are also available of the costs of damage from typical trespass incidents before and after the introduction of the new fencing for a random sample of 7 factories, as follows (in £).

Factory	T	U	V	W	X	Y	Z
Cost before installation	1215	95	546	467	2356	236	550
Cost after installation	1268	110	578	480	2417	318	620

Stating carefully the required distributional assumption, provide a two-sided 99% confidence interval based on a t distribution for the population mean difference between costs of damage before and after installation of the new fencing.

Explain why this confidence interval should not be based on the Normal distribution. [9]