



**ADVANCED GCE UNIT
MATHEMATICS (MEI)**

Statistics 2

MONDAY 21 MAY 2007

4767/01

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, 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.
- The total number of marks for this paper is 72.

ADVICE TO CANDIDATES

- Read each question carefully and make sure you know what you have to do before starting your answer.
- 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.

- 1 The random variable X represents the time taken in minutes for a haircut at a barber's shop. X is Normally distributed with mean 11 and standard deviation 3.

(i) Find $P(X < 10)$. [4]

(ii) Find the probability that exactly 3 out of 8 randomly selected haircuts take less than 10 minutes. [3]

(iii) Use a suitable approximating distribution to find the probability that at least 50 out of 100 randomly selected haircuts take less than 10 minutes. [4]

A new hairdresser joins the shop. The shop manager suspects that she takes longer on average than the other staff to do a haircut. In order to test this, the manager records the time taken for 25 randomly selected cuts by the new hairdresser. The mean time for these cuts is 12.34 minutes. You should assume that the time taken by the new hairdresser is Normally distributed with standard deviation 3 minutes.

(iv) Write down suitable null and alternative hypotheses for the test. [3]

(v) Carry out the test at the 5% level. [5]

- 2 A medical student is trying to estimate the birth weight of babies using pre-natal scan images. The actual weights, x kg, and the estimated weights, y kg, of ten randomly selected babies are given in the table below.

x	2.61	2.73	2.87	2.96	3.05	3.14	3.17	3.24	3.76	4.10
y	3.2	2.6	3.5	3.1	2.8	2.7	3.4	3.3	4.4	4.1

(i) Calculate the value of Spearman's rank correlation coefficient. [5]

(ii) Carry out a hypothesis test at the 5% level to determine whether there is positive association between the student's estimates and the actual birth weights of babies in the underlying population. [5]

(iii) Calculate the value of the product moment correlation coefficient of the sample. You may use the following summary statistics in your calculations:

$$\Sigma x = 31.63, \quad \Sigma y = 33.1, \quad \Sigma x^2 = 101.92, \quad \Sigma y^2 = 112.61, \quad \Sigma xy = 106.51. \quad [5]$$

(iv) Explain why, if the underlying population has a bivariate Normal distribution, it would be preferable to carry out a hypothesis test based on the product moment correlation coefficient.

Comment briefly on the significance of the product moment correlation coefficient in relation to that of Spearman's rank correlation coefficient. [4]

3

3 The number of calls received at an office per 5 minutes is modelled by a Poisson distribution with mean 3.2.

(i) Find the probability of

(A) exactly one call in a 5-minute period,

(B) at least 6 calls in a 5-minute period.

[4]

(ii) Find the probability of

(A) exactly one call in a 1-minute period,

(B) exactly one call in each of five successive 1-minute periods.

[4]

(iii) Use a suitable approximating distribution to find the probability of at most 45 calls in a period of 1 hour. [4]

Two assumptions required for a Poisson distribution to be a suitable model are that calls arrive

- at a uniform average rate,
- independently of each other.

(iv) Comment briefly on the validity of each of these assumptions if the office is

(A) the enquiry department of a bank,

(B) a police emergency control room.

[4]

4 The sexes and ages of a random sample of 300 runners taking part in marathons are classified as follows.

Observed		Sex		Row totals
		Male	Female	
Age group	Under 40	70	54	124
	40–49	76	36	112
	50 and over	52	12	64
Column totals		198	102	300

(i) Carry out a test at the 5% significance level to examine whether there is any association between age group and sex. State carefully your null and alternative hypotheses. Your working should include a table showing the contributions of each cell to the test statistic. [10]

(ii) Does your analysis support the suggestion that women are less likely than men to enter marathons as they get older? Justify your answer. [3]

For marathons in general, on average 3% of runners are 'Female, 50 and over'. The random variable X represents the number of 'Female, 50 and over' runners in a random sample of size 300.

(iii) Use a suitable approximating distribution to find $P(X \geq 12)$. [5]