UNIVERSITY OF LONDON

IMPERIAL COLLEGE LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)
MAY–JUNE 2003

This paper is also taken for the relevant examination for the Associateship.

M3S10/M4S10 DESIGN OF EXPERIMENTS AND SURVEYS

DATE: Friday, 30th May 2003 TIME: 10 am – 12 noon

Credit will be given for all questions attempted but extra credit will be given for complete or nearly complete answers.

Calculators may not be used. Statistical tables will not be available.

1. *a)* Explain the terms *systematic effect* and *blocking* in connection with experimental design.

b) Three different insecticides A, B and C are to be tested on 36 closely spaced plum trees arranged in a 6 x 6 square array. Two possible layouts are suggested:

Layout 1

A	B	C	A	B	C
B	C	A	B	C	A
C	A	B	C	A	B
A	B	C	A	B	C
B	C	A	B	C	A
C	A	B	C	A	B

Layout 2

A	A	C	B	C	B
A	B	A	B	C	C
C	A	C	A	B	B
B	B	A	C	A	C
C	C	B	A	B	A
B	C	B	C	A	A

i) Discuss the advantages and disadvantages of these two layouts and state with reasons which one you consider to be more appropriate.

ii) What practical problems might there be in carrying out an experiment of this kind?

2. *a)* Define a *balanced incomplete block design* (BIBD).

b) Show that in a symmetric BIBD any two blocks have the same number of treatments in common. You may assume that the incidence matrix of a symmetric BIBD is invertible.

c) Define a *finite difference set.*

d) By considering the set $\{0, 1, 6, 8, 18\}$ with arithmetic modulo 21, find a BIBD for 16 treatments, with 20 blocks of 4 units each. Justify your construction in detail and find the other parameters of your BIBD.

3. *a)* In the context of a 2^n factorial experiment explain what is meant by saying that two main effects are aliased with each other.

b) In a particular case the factors are A, B, C, D and E. Only 16 experimental units are available and these are arranged in a 4×4 square array. The 5-factor interaction is aliased with the mean, the effects ABC and CD are confounded with the row effects and the effects ADE and AC are confounded with the column effects.

i) Suggest a set of 16 treatment combinations to be used in the experiment.

ii) List all the other effects which are confounded with

(I) the row effects.

(II) the column effects.

iii) Construct a suitable design and discuss its properties, making any assumptions clear.

iv) Show that your set of 16 treatment combinations cannot be allocated to the cells of a 2×8 rectangular array so that no main effects are confounded with either row or column effects.

4. In a linear model, the expected value of the observation $Y(\underline{x})$ at the point $\underline{x}$ in a design region is $\underline{f}(\underline{x})^T\underline{\beta}$, where $\underline{f}(\underline{x})$ is a known vector of continuous functions of $\underline{x}$, and $\underline{\beta} = (\beta_1, \ldots \beta_t)^T$ is a $t \times 1$ vector of unknown parameters. All observations are independent and have the same variance.

a) In this context, explain what is meant by a *D-optimal design measure* and a *G-optimal design measure.*

b) State, without proof, the General Equivalence Theorem for linear models.

c) In a particular case, $E(Y(x)) = \beta_1 x + \beta_2 x^2$, where $Y(x)$ is the observation at x, and x is restricted to lie in the closed interval [1,2].

i) Calculate the information matrix, $M(\xi)$, of the design measure ξ which attaches weight one-half to each of the points $x = 1$ and $x = 2$.

ii) Show that the variance function of ξ is

$$d(x, \xi) = \frac{x^2}{2}(17 - 18x + 5x^2).$$

iii) Show that $x - 1$ and $x - 2$ are factors of $h(x) = d(x, \xi) - 2$.

iv) By factorising $h(x)$ show that $d(x, \xi) \leq 2$ for all $x \in [1, 2]$.

v) Show that for any design measure ξ' over [1,2], the determinant of $M(\xi')$ cannot exceed 1.

5. *a)* Describe the advantages of a sample survey over a complete census.

b) Explain what is meant by a *simple random sample* from a finite population.

c) Show that the sample mean $\overline{y}$ (of some observation Y) is an unbiased estimator of the population mean $\overline{Y}$, and state $var(\overline{y})$ in terms of the population variance, the sampling fraction and the sample size.

d) In a survey to examine smoking habits a simple random sample of n individuals is taken from a fixed population of size N. Each person is asked to classify himself or herself as one of:

1. a smoker.
2. a non-smoker.

The (unknown) proportion of people in the population who fall in the first category is P and the observed proportion of people in the sample who fall in this category is p.

It is desired to estimate D, the difference between the proportions in the population of those who smoke and those who do not.

i) Show that $d = 2p - 1$ is unbiased for D and that

$$var(d) = \frac{4(N-n)P(1-P)}{n(N-1)}.$$

ii) It is a requirement that $var(d) \leq v$, where $v > 0$ is a given constant. Show that in order to satisfy this requirement, whatever the value of P, n should be at least

$$\frac{N}{1 + v(N-1)}.$$

iii) Discuss any difficulties which might arise from asking people to classify themselves.