Imperial College London

UNIVERSITY OF LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)

May-June 2005

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

M3S10/M4S10

Design of Experiments & Surveys

Date: Wednesday, 1st June 2005 Time: 2 pm – 4 pm

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. (i) Explain briefly the purpose of randomisation in experimental design.

   (ii) Define a *balanced incomplete block design* (BIBD) for t treatments arranged in blocks of k experimental units each and explain how you would randomise such a design.

   (iii) Define an *unreduced BIBD* and state its parameters in terms of k and t.

   (iv) Show that any BIBD with $k = 2$ is unreduced.

   (v) Show that the complement, D', of any BIBD D with $k < t - 1$ is also a BIBD.

   (vi) Hence show that a BIBD with $k = t - 2$ is unreduced.

2. (i) Define (a) an $n \times n$ *latin square* and (b) an $n \times n$ *graeco-latin square*.

   (ii) Construct a 3×3 graeco-latin square, and a 4×4 graeco-latin square.

   (iii) Explain carefully how you would use these graeco-latin squares to construct a 12×12 graeco-latin square. There is no need to write out the whole design.

   (iv) The treatment combinations of a complete $3^2 \times 4^2$ factorial experiment are to be allocated, one each, to the 144 cells of a 12 x 12 array so that no main effect is confounded with the rows or with the columns of the array. By considering a suitable graeco-latin square show how this can be done and justify your answer. There is no need to write out the whole design.

3. (i) Explain the purpose of *fractional replication* in a 2^n factorial experiment.

(ii) In a 2^5 factorial experiment the following treatment combinations are used:

$$\{1, ab, bde, ade, bce, ace, cd, abcd\}.$$

All three-factor and higher order interactions are negligible.

Which effects are aliased with the mean?

Show that no main effect is aliased with any other main effect.

(iii) It is now desired to split these 8 treatment combinations into 2 blocks of 4 each without confounding any main effects with blocks. By examining the alias structure of the design explain why you would not want to use a 4- or 5- factor interaction as a defining contrast.

(iv) By considering a suitable 3-factor interaction construct the two blocks.

4. In a linear model the observation Y made at $(x_1, x_2) \in \mathbb{R}^2$ satisfies

$$E(Y) = \beta_1 x_1 + \beta_2 x_2,$$

where β_1 and β_2 are unknown parameters. All observations are independent and have common unknown variance. Suppose that (x_1, x_2) is restricted to the design region

$$\mathcal{X} = \left\{(x_1, x_2) : 0 \le x_1 \le 1;\ 0 \le x_2 \le 1;\ x_1 + x_2 \le \frac{3}{2}\right\},$$

(i) Let ξ_p be the design measure which attaches weight p to each of the points $(0, 1)$ and $(1, 0)$, and weight q to each of the points $(1, \frac{1}{2})$ and $(\frac{1}{2}, 1)$, where $p + q = \frac{1}{2}$. Calculate the information matrix, $M(\xi_p)$ and show that $\det M(\xi_p)$ is maximised with respect to p when $p = 11/30 (= p^*$, say).

(ii) Find the variance function, $d\left((x_1, x_2), \xi_{p^*}\right)$, of ξ_{p^*} and show that it is proportional to

$$3(x_1 + x_2)^2 + 5(x_1 - x_2)^2.$$

(iii) Hence show that $d\left((x_1, x_2), \xi_{p^*}\right) \le 2$ for all $(x_1, x_2) \in \mathcal{X}$ such that both $x_1 \in [\frac{1}{2}, 1]$ and $x_2 \in [\frac{1}{2}, 1]$.

(iv) By considering $d\left((x_1, x_2), \xi_{p^*}\right)$ for $x_1 \in [0, 1]$ and $x_2 \in [0, \frac{1}{2}]$, or otherwise, deduce that ξ_{p^*} is G-optimal. You should state clearly any theorems of which you make use.

5. A bank is interested in its customers' views on a certain aspect, S, of its service. They aim to contact a simple random sample (without replacement) of n customers from their total population of N customers. Each will be asked "Are you satisfied with S?" and will be required to reply with a score of either 1 or 2, where 1 = not satisfied and 2 = satisfied.

The bank would like to estimate the average score, $\overline{Y}$, of the whole population of its customers. First however it must decide to conduct the survey either (a) by interviewing the n sampled customers over the telephone or (b) by writing to them and requesting a postal reply.

(i) Discuss from a statistical viewpoint the advantages and disadvantages of (a) and (b).

(ii) Suppose that of the n sampled customers only m respond ($m < n$) and that the average score for these m customers is $\overline{y}_{(m)}$. Although the bank is unaware of it, all the $n - m$ customers who did not respond are in fact satisfied with S. Show that, conditionally on m

$$E(\overline{y}_{(m)}) = \frac{\overline{Y} - 2(1-\alpha)}{\alpha}$$

where $\alpha = \frac{m}{n}$.

(iii) The bank erroneously assumes that the m customers who responded formed a simple random sample of size m of its total population of customers and they calculate var$(\overline{y}_{(m)})$ accordingly. Show that the ratio of the bank's wrongly calculated var$\big(\overline{y}_{(m)}\big)$ to the true var$\big(\overline{y}_{(m)}\big)$ is

$$\frac{\alpha(N - \alpha n)}{N - n}.$$

Hence show that the bank will obtain a value of var$(\overline{y}_{(m)})$ which is lower than the true value if and only if

$$\alpha < \frac{N}{n} - 1$$

and that this will always be the case when $n < \frac{1}{2}N$.

[You may make use without proof standard results proved in lectures about the sample mean of observations obtained from a simple random sample from a finite population.]