Imperial College London

UNIVERSITY OF LONDON

BSc and MSci EXAMINATIONS (MATHEMATICS)

May-June 2007

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

M3S10/M4S10

Design of Experiments & Surveys

Date: Thursday, 31st May 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.

1. (i) Define a *balanced incomplete block design* (BIBD).

A cyclic block design D_n $(n \geq 3)$ is generated from the initial blocks $B_j = \{0, j\}$ for $j = 1, 2, \ldots, m$, where $n = 2m + 1$ and all arithmetic is modulo n.

(ii) Construct D_7.

(iii) Show that D_n has $\binom{n}{2}$ distinct blocks.

(iv) Show that D_n is a BIBD and determine its parameters.

(v) Show that there exists an $\binom{n}{2} \times 2$ rectangular array of n symbols such that the rows form a BIBD and each column contains each symbol exactly m times.

(vi) Describe an experimental situation where the array found in part (v) might be useful.

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

(ii) If $L_1, L_2, \ldots, L_m$ are $n \times n$ latin squares such that each distinct pair forms a graeco-latin square, prove that $m \leq n - 1$.

(iii) Find four 5×5 latin squares such that each pair forms a graeco-latin square.

(iv) State (without proof) MacNeish's Theorem.

(v) Prove that if n is not congruent to 2 modulo 4, then there exists an $n \times n$ graeco-latin square.

3. (i) In the context of a 2^n factorial experiment explain briefly what is meant by (a) *an effect is aliased with the mean* and (b) *an effect is confounded with the block effects.*

In a 2^6 factorial experiment with factors $A, B, C, D, E,$ and F, only 16 experimental units are available and these are arranged in a 4×4 square array with the following allocation of treatment combinations:

$$\begin{array}{cccc} bcd & adef & ab & cef \\ abcef & 1 & bdef & acd \\ ae & bcf & cde & abdf \\ df & abcde & acf & be \end{array}$$

(ii) Which effects are aliased with the mean?

(iii) Show that no two main effects are aliased with each other but that each of 3 main effects is aliased with a two-factor interaction.

(iv) What advice about the labelling of the 6 factors would you give to the experimenter?

(v) Which effects are confounded with the row effects?

(vi) Explain why no main effect is confounded with either the row or the column effects.

(vii) Show that the 16 treatment combinations cannot be arranged in a 2×8 rectangular array such that no main effect is confounded with either the row or the column effects.

4. For $x \in [0,1]$, it is possible to observe $Y(x)$ which satisfies

$$E(Y(x)) = \beta_1 + \beta_2 x \quad \text{and} \quad var(Y(x)) = \sigma^2 \ (> 0).$$

All observations are independent. The design measure ξ attaches weight one-third each to the 3 points x_1, x_2 and $x_3 \in [0,1]$. Not all of x_1, x_2 and x_3 are equal.

(i) Show that the information matrix $M(\xi)$ has determinant

$$\frac{1}{9}(3S_2 - S_1^2)$$

where $S_k = \sum_{i=1}^{3} x_i^k \quad (k = 1, 2)$.

(ii) Show that the variance function of ξ is

$$\frac{3(S_2 - 2S_1 x + 3x^2)}{3S_2 - S_1^2}.$$

(iii) Prove that for fixed x_2 and x_3, the determinant of $M(\xi)$ is maximised with respect to $x_1 \in [0,1]$ when $x_1 = 0$ or $x_1 = 1$. Deduce that the maximum value of $\det M(\xi)$ with respect to x_1, x_2 and x_3 occurs when each $x_i = 0$ or 1 $(i = 1, 2, 3)$.

(iv) Hence show that the maximum value of $\det M(\xi)$ with respect to x_1, x_2 and $x_3 \in [0,1]$ is $\frac{2}{9}$.

(v) Suppose the design measure ξ_1 has $x_1 = x_2 = 0$ and $x_3 = 1$ and the design measure ξ_2 has $x_1 = 0$, $x_2 = \frac{1}{2}$ and $x_3 = 1$. Show that ξ_1 is D-optimal, but not G-optimal, within the class of design measures of the type ξ above.

(vi) Comment on this result in relation to the General Equivalence Theorem.

5. (i) Define *a simple random sample* of size n without replacement from a population of size N.

(ii) If $Y_1, Y_2, \ldots, Y_N$ denote the population values and $y_1, y_2, \ldots, y_n$ denote the sample values show that $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is unbiased for $\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$.

(iii) Show that the sample variance, s^2, is unbiased for the population variance, σ^2.

In a particular survey an estimate of the proportion of individuals in the population who possess a certain characteristic, C, is required.

(iv) By setting $Y_i = 1$ if the i^{th} population member has characteristic C, and $Y_i = 0$ otherwise, show that the proportion, π, of the sample who have C is unbiased for the proportion of the population who have C.

(v) Prove that

$$\frac{N-n}{N(n-1)}\pi(1-\pi)$$

is unbiased for $var(\pi)$.

(vi) Discuss briefly the problem of obtaining accurate responses from sampled individuals when they are asked "Do you possess characteristic C?" in a face-to-face interview situation.

[You may assume that $var(\overline{y}) = (1 - \frac{n}{N})\frac{\sigma^2}{n}$ and that for any real numbers $x_1, x_2, \ldots, x_m$, $\sum_{i=1}^{m}(x_i - \overline{x})^2 = \sum_{i=1}^{m} x_i^2 - m\overline{x}^2$, where $\overline{x} = \frac{1}{m}\sum_{i=1}^{m} x_i$]