M3S10 / M4S10 DESIGN OF EXPERIMENTS AND SURVEYS

EXERCISES

PRINCIPLES OF EXPERIMENTAL DESIGN

Discuss possible systematic effects and any practical problems which might be encountered in:
(a) an experiment to compare the tastes of 2 different kinds of orange drink (claimed by each of the manufacturers to taste like fresh orange juice).
(b) an experiment to compare two different advertising campaigns to encourage car drivers to reduce their speed on the roads. One is a television campaign and the other uses road hoardings.
(c) an experiment to compare the tastes of 2 particular brands of margarine both of which are claimed to taste better than butter.

(d) an experiment to compare the effectiveness of two different kinds of anti dandruff shampoo
(e) an experiment to compare the effectiveness of two different kinds of rubber used for windscreen wipers.

BLOCK DESIGNS

1. Show that the complement of a BIBD with $k < t$ ($k \neq t-1$) is also a BIBD, and find the corresponding parameters.

2*. If a BIBD is resolvable, show that $b \geq t + r - 1$.(Hint: consider the incidence matrix of the BIBD)

3. The treatments of a block design D form the set $S = \{0, 1, 2, \cdots, 2n-2, \infty\}$, where addition in $S \setminus \{\infty\}$ is modulo $2n-1$. The blocks of D are the sets B_{ij}, where for $i = 0, 1, 2, \cdots, 2n-2$,

$$B_{ij} = \{i+j, i-j\} \qquad (j = 1, 2, \cdots, n-1)$$

and

$$B_{i0} = \{\infty, i\}$$

Show that, for $n > 1$, D is a resolvable BIBD, and find its parameters.

4. Show that $\{1, 2, 4, 10\}$ modulo 13 is a finite difference set. Use this set to form a symmetric BIBD with 13 blocks, 13 treatments and 4 units in each block.

Show how you would use your BIBD to construct

(i) a BIBD with 13 blocks, 13 treatments and 9 units in each block,
(ii) a BIBD with 12 blocks, 9 treatments and 3 units in each block.

In each case state the other parameters of the design.

5. By considering the sets $B = \{0, 1, 10\}$ and $B' = \{0, 5, 7\}$, show how to construct a BIBD for 13 treatments arranged in 26 blocks of 3 units each, justifying your answer. Hence show there exists a BIBD for 13 treatments arranged in 26 blocks of 10 units each. What are the other parameters of this design?

6. Let M be a $4m \times 4m$ matrix such that

(i) each entry is either +1 or -1,
(ii) $MM^T = M^T M = 4mI_{4m}$,
(iii) the first row and first column of M are all +1's.

Delete the first row and first column and change all -1's to 0. Show that the resulting matrix is the incidence matrix of a BIBD and find its parameters. (Such a design is called a Hadamard configuration).

7. Does there exist a BIBD with

(i) $t = 7, b = 21, k = 5, r = 15$?
(ii) $b = t = 46$, $k = r = 10$?

8. In a BIBD show that $t(t-1)S = bg(k,t)$,where $S = b - 3(r - \lambda)$ and $g(k,t) = 3k(k-t) + t(t-1)$.

Show that for $t > 4$, $g(k,t) > 0$ for all k, and that for $t = 3$ or $t = 4$, $g(k,t) \geq 0$ for all $k \geq 2$.

Deduce that $b \geq 3(r - \lambda)$ and find all BIBDs for which equality holds.

9. By considering the even powers of 2 modulo 19, find a symmetric BIBD with 19 blocks of 9 units each.

10. A cyclic block design with t treatments is generated from one initial block containing k distinct treatments. If k and t are co-prime, show that the design has t distinct blocks.

LATIN SQUARES

1. Construct a design for an experiment with 3 treatments A, B, C which must occur once each on each of 4 days and once each at each of 4 times of day.

2. Construct a complete orthogonal set of 5 x 5 latin squares.

3. The rows, columns and symbols of an $n \times n$ array A are labelled 0,1,2$\cdots$, $n-1$, where $n = 2m - 1$ and m is an integer. Show that if $A = (a_{ij})$ satisfies

$$a_{ij} \equiv m(i+j) \pmod{n},$$

then A is an $n \times n$ latin square.

The $n \times n$ array $B = (b_{ij})$, where b_{ij} satisfies

$$b_{ij} \equiv i - j \ (mod\ n).$$

Show that A and B form a graeco-latin square.

4. A *transversal* of an $n \times n$ latin square is a set of n cells all in different rows, in different columns and receiving different treatments. Show that each of 2 orthogonal $n \times n$ latin squares has at least n distinct transversals.

5. Construct a 12 x 12 graeco-latin square.

6. Construct BIBDs with the following parameters:

(i) $b = 30, t = 25, k = 5$
(ii) $b = t = 31, k = 6$
(iii) $b = t = 21, k = 5$

7. A design D is formed by deleting the first column of a latin square. Show that D is a Youden rectangle.

Construct two 4x3 Youden rectangles, each using the symbols 0, 1, 2 and 3, which when superimposed give every ordered pair of distinct symbols exactly once.

8. Complete this10x10 graeco-latin square.

00	47	18	76		93		34		52
86	11			70		94	45	02	
95	80	22	67	38	71		56	13	04
	96	81		07		72	60		15
73		90	82		17			35	26
68			91	83		27	12	46	
	08	75	19		84	66	23		41
14		36		51			77	88	99
	32		54	65	06	10	89		
42		64		16		31		79	87

Notice when you have finished, that you have a 10x10 array of all numbers from 00-99 such that each row and each column contains every tens digit and every units digit exactly once!

2^n FACTORIAL EXPERIMENTS

1. Arrange the treatment combinations of a 2^4 experiment in a sequence so that any treatment combination is obtained from the previous one by changing the level of just one factor. Why might this be of practical importance?

2. A 2^5 experiment with factors A, B, C, D, E is arranged in 4 blocks of 8 units each by confounding ACE and ABDE. Which other interaction is confounded? List the treatment combinations in the blocks.

If the experiment is replicated 5 times, with the above arrangement forming the first replicate, list the interactions you would confound in each of the remaining 4 replicates.

3. A 2^6 factorial experiment with factors A, B, C, D, E, F was performed in 8 blocks of 8 units each. The principal block consisted of the treatment combinations

$$1, bcde, abd, bef, adef, ace, cdf, abcf.$$

What effects are confounded with blocks? Obtain the remaining blocks.

If the experimenter can only use the contents of the principal block, is it possible to subdivide these into 4 smaller blocks of 2 units each so that no main effect is confounded with blocks?

4. Suppose that all main effects and two-factor interactions except AB and CD are of interest in a 2^4 experiment with factors A, B, C, D. Arrange the treatment combinations in a 4 x 4 array so that no effect of interest is confounded with rows or columns.

5. A $\frac{1}{4}-$replicate of a 2^6 experiment with factors A, B, C, D, E, F is required. It can be assumed that all interactions of 3 or more factors are negligible, as are any two-factor interactions involving A or C. Show that such a fractional replicate can be constructed so that no main effect or two-factor interaction of interest is aliased with any other main effect or two-factor interaction of interest. List the treatment combinations.

6. In a 2^6 experiment, the factors are A, B, C, D, E, F. Only 32 experimental units are available. In each case below how would you allocate the treatment combinations of a suitable $\frac{1}{2}-$replicate to the experimental units?

(a) The units are arranged in 4 blocks of 8 units each. All main effects and two-factor interactions with the exception of AB, CD and EF are of interest and all 4-factor and higher order interactions are known to be negligible.

(b) The units are arrangd in a 4 x 8 rectangular array. All main effects must be unconfounded with rows and unconfounded with columns of the array. All 5- and 6-factor interactions are negligible.

7. Show that it is impossible to construct a $\dfrac{1}{2^r}$ fractional replicate of a 2^m experiment in which no two main effects are aliased if $m \geq 2^{m-r}$.

8. By considering the 7 orthogonal contrasts corresponding to the effects and interactions in a 2^3 experiment, show how to form a $\frac{1}{16}$−fractional replicate of a 2^7 experiment in which no two main effects are aliased with each other.

OPTIMAL DESIGNS

You may assume that all observations are independent and have constant variance.

1. The observation Y depends on an explanatory variable $x \in [-1, 1]$ and

$$E(Y) = \beta_1 + \beta_2 x.$$

Consider the 2 exact designs:

ξ_1 :one observation at each of $x = -1, 0, +1$
ξ_2 :one observation at $x = -1$ and two at $x = +1$

Calculate the variance function for each design.

Show that ξ_2 is not G - optimal in the class of designs with 3 observations.

The design ξ puts one observation each at x_1, x_2 and $x_3 \in [-1, 1]$. Show that det $M(\xi)$ is proportional to

$$g(\underline{x}) = \sum_1^3 (x_i - \overline{x})^2$$

where $\overline{x} = \frac{1}{3} \sum x_i$ and $\underline{x} = (\ x_1, x_2,\ x_3)^T$.

By considering $g(\underline{x})$ as a quadratic in each of its 3 separate variables, show that $g(\underline{x})$ is maximised on the boundary of the region defined by $-1 \leq x_i \leq 1$ $(i = 1, 2, 3)$.

Hence show that ξ_2 is D - optimal in the class of designs with 3 observations. Comment.

2. In a particular linear model the response Y is related to 2 explanatory variables x_1 and x_2 by

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

where β_0, β_1 and β_2 are unknown parameters. The variables x_1 and x_2 are constrained to lie in the parabolic region

$$\mathcal{X} = \{(x_1, x_2) : |x_1| \leq 1, x_2 \geq 0, x_2 + x_1^2 \leq 1\}$$

Show that there exists a D-optimal design whose only points of support are $(\pm 1, 0)$ and (0,1).

3. The design ξ^* is D-optimal and it has a finite number of points of support, one of which is $\underline{x}$. Show that $d(\underline{x},\xi^*) = t$, where t is the number of unknown parameters.

4. For the model $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ the design region is

$$\mathcal{X} = \{(x_1, x_2) : x_1 \geq 0,\ x_2 \geq 0,\ x_2 + \frac{1}{2}x_1 \leq 1,\ x_1 + \frac{1}{2}x_2 \leq 1\}$$

The design measure ξ_w attaches weights $1 - 2w, w$ and w respectively to the points (0,0),(1,0) and (0,1) where $0 \leq w \leq 1$. Find the value of w which maximises det $M(\xi)$ and show that the resulting design is D-optimal.

5. If the design region $\mathcal{X}$ contains exactly t points, where t is the number of unkown parameters, show that the design attaching weight $1/t$ to each point of $\mathcal{X}$ is D-optimal.

6. The design ξ attaches weight one-quarter to each of the 4 points $(\pm 1, \pm 1)$. Find and sketch the largest region $\mathcal{X}$ in the plane, over which ξ is D-optimal for the model

$$E(Y) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

7. If $\mathcal{X}$ is a subset of the (x_1, x_2) plane and the model is

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

show that within the class of designs with 2 observations, a D-optimal design is obtained by placing one observation each at points $A, B \in \mathcal{X}$ which maximise the area of triangle OAB, where O denotes the origin.

8. A response Y depends on two explanatory variables $x_1$and x_2, and

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

where α, β_1, and β_2 are unknown parameters. The values of x_1 and x_2 are constrained by

$$x_2 \geq 0, |x_1 \pm x_2| \leq 1.$$

The design measure ξ attaches weight $\frac{1}{3}$ to each of the points $(\pm 1, 0)$ and the point (0,1). Show that ξ is D-optimal.

Suppose now that the values of $x_1$and x_2 are unconstrained, Show that ξ is D-optimal over a region R in the (x_1, x_2) plane where R is bounded by the ellipse with centre $(0, \frac{1}{3})$ and axes of length $\frac{4}{3}$ and $\frac{4}{\sqrt{3}}$.

SAMPLE SURVEYS

1. In each case below discuss practical problems you might encounter and advice you would give to an investigator who wishes to:

(a) estimate the average amount of weekly pocket money received by 8 year old children in the UK.

(b) estimate the average number of cups of coffee drunk per day in London.

2. The cost of taking a simple random sample of size n from a population of size N is $c_0 + c_1 n$, where c_0 is the overhead cost and $c_1$is the cost per sampled unit. The additional cost of estimating the population mean of some characteristic Y by the sample mean $\overline{y}$ is $\lambda\ var(\ \overline{y})$, where λ is a known constant. Show that the total cost is minimised when n is approximately

$$\left(\frac{\lambda}{c_1}\right)^{\frac{1}{2}} \sigma$$

where σ is the population standard deviation.

How would you find the value of n which minimises the total cost if the cost per sampled unit is now c_1 for the first m units sampled and $c_1/2$ for the rest, where m is fixed?

3. The proportion P of units with a particular characteristic in a population of known size N is to be estimated. Show that the population variance is

$$\frac{NP(1-P)}{N-1}$$

A simple random sample of size n has proportion p with the characteristic. Show that p is unbiased for P and that

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

Find an unbiased estimator of *var* p.

If we replace P by p in the expression for *var* p, we get a biased estimator for *var* p. Find an expression for the bias in terms of P, N and n.

4. Suppose that we wish to conduct a survey on a sensitive issue e.g. to estimate the proportion P of women in a particular country who have had an abortion. A simple random sample of size n is taken from the population of N women and each is asked to answer one of the following two questions.

Question 1: Have you ever had an abortion?
Question 2: Is your birthday in January?

So that the interviewer is not aware of the question being asked each woman independently selects a card from a shuffled pack. The pack consists of a proportion θ which have question 1 and a proportion $1-\theta$ which have question 2 written on them.

Of the sampled women, a proportion p answer "yes". Assuming that one-twelfth of the population have birthdays in January, show how you might estimate P. Discuss the factors that would influence your choice of θ.

5. With a certain population, it is known that the measurements Y_i are zero for a portion qN of the N units $(0 < q < 1)$. Sometimes these units can be found and listed so that they need not be sampled. If s^2 and s^2_{NZ} are the population variances for the whole population and the "non-zero" population, show that neglecting terms of order $1/N$,

$$s^2_{NZ} = \frac{s^2}{p} - \frac{q}{p}\overline{Y}^2,$$

where $p = 1 - q$.

If the population total is estimated from a simple random sample of size n, from the Np non-zero units, show that, with the exclusion of the "zero" units, the fractional reduction in variance of the estimator is,neglecting terms of order $1/N$, equal to

$$\frac{q(V^2+1)}{V^2},$$

where $V^2 = s^2/\overline{Y}^2$.

6. With the aim of estimating $\overline{Y}$ a sampler divides a population of size N into 2 strata, of sizes N_1 and N_2. The population standard deviations within each stratum are assumed to be equal but the costs of sampling (per unit) in the 2 strata are c_1 and c_2 respectively. If n_i is the sample size for the i^{th}stratum, show

that the relative precision of the optimal allocation (o) to that of proportional allocation (p), for fixed total cost, is given by

$$\frac{var(\overline{y}_{st}(p))}{var(\overline{y}_{st}(o))} = \frac{N(N_1c_1 + N_2c_2)}{(N_1\sqrt{c_1} + N_2\sqrt{c_2})^2},$$

where the finite population correction for each stratum has been ignored.

The sampler thinks that $2c_1 < c_2 < 4c_1$. Find the maximum relative increase in variance from using proportional allocation and show that it cannot exceed 1/8.