UNIVERSITY OF SURREY ©

B. Sc. Undergraduate Programmes in Mathematical Studies
M. Math. Undergraduate Programmes in Mathematical Studies

Level HE3 Examination

Module MS338 EXPERIMENTAL DESIGN

Time allowed – 2 hours Spring Semester 2007

Attempt THREE questions. If any candidate attempts more than THREE questions only the best THREE solutions will be taken into account.
Cambridge Statistical Tables will be provided.

Question 1

(a) A block design with t treatments and b blocks has $t \times b$ incidence matrix, $N = \{n_{ij}\}$, with $n_{ij} = 1$ if treatment i occurs in block j and $n_{ij} = 0$ otherwise. Define *balance* for a Balanced Incomplete Block Design, *BIBD*, with t treatments replicated r times in b blocks of size k. Justify, in terms of the incidence matrix, the conditions

$$tr = bk \quad \text{and} \quad \lambda(t-1) = r(k-1),$$

where λ is the number of times a pair of treatments occur together in a block. [7]

(b) Define the 'complement' $\overline{B}$ of an incomplete block design B. Let B be a *BIBD* with parameters t, r, b, k, and λ. Write down $\bar{t}$, $\bar{r}$, $\bar{b}$ and $\bar{k}$, the parameters of $\overline{B}$ in terms of t, r, b and k. Prove that $\overline{B}$ is a *BIBD* whenever $b - 2r + \lambda > 0$. [5]

(c) (i) Explain what is meant by a 'perfect difference set modulo(n)'. Illustrate your answer by consideration of $\{0, 1, 4, 6\}$ and $\{0, 1, 2, 3\}$, working modulo 13. [5]

(ii) Obtain the set of non-zero squares modulo(7). Show that this set is a perfect difference set modulo(7) and hence produce a *BIBD*, B_1 say, with parameters:

$$b = 7, \quad k = 3, \quad t = 7, \quad r = 3.$$

[6]

(iii) State how B_1 can be used to obtain a *BIBD*, B_2 say, with parameters:

$$b = 7, \quad k = 4, \quad t = 7, \quad r = 4.$$

[2]

Question 2

(a) The one-way analysis of covariance model with a single covariate, x, for rt observations y_{ij} is given by

$$y_{ij} = \mu + \alpha_i + \beta_i x_{ij} + \varepsilon_{ij} \quad i = 1, \ldots, t \quad j = 1, \ldots, r, .$$

The least-squares estimator of β_i is given by

$$\hat{\beta}_i = \frac{[S_{xy}]_i}{[S_{xx}]_i} = \frac{\sum_{j=1}^{r} x_{ij} y_{ij} - r\bar{x}_{i.}\bar{y}_{i.}}{\sum_{j=1}^{r} x_{ij}^2 - r\bar{x}_{i.}^2} \quad \text{for} \quad i = 1, \ldots, t.$$

(i) State the assumptions that are made regarding the ε_{ij}. [1]

(ii) Show that in the parallel regressions model, where $\beta_1 = \ldots = \beta_t = \beta$, the least squares estimator of β is given by

$$\hat{\beta}_w = \sum_{i=1}^{t} w_i \hat{\beta}_i.$$

Specify the weights w_i and show that $w_i \geq 0$ and $\sum_{i=1}^{t} w_i = 1$. [6]

(b) Three different processes used to manufacture a chemical are being compared. Five batches of chemical are manufactured for each process. The observation y_{ij} is the yield of chemical in grammes per 10kg of raw material and the covariate x_{ij} is the temperature (°C) inside the reactor. The data are shown in the following table:

Process 1		Process 2		Process 3	
x	y	x	y	x	y
17	98.6	16	88.7	15	84.7
12	71.5	13	76.2	15	86.7
8	55.0	14	80.4	19	104.2
10	62.4	9	56.9	11	64.5
14	85.2	20	108.2	9	56.9

These data were analysed using R with a factor A representing the processes at three levels and a variable X representing temperature. The computer output follows:

```
> fullm<-aov(y~A*x)
> fullm
Call:
   aov(formula = y ~ A * x)
Terms:
                        A          x        A:x      Residuals
Sum of Squares      146.089   4032.770     3.968      16.402
Deg. of Freedom         2          1         2           9
Residual standard error: 1.349977
```

```
> coefficients(fullm)
(Intercept)     A2         A3         x          A2:x       A3:x
 13.525000  1.9130368  -0.809868   5.001229   -0.373315   -0.168993
> fitted.values(fullm)
    1        2        3        4        5        6        7        8
 98.5459  73.5397  53.5348  63.5373  83.5422  89.4846  75.6009  80.2288
   9        10         11        12         13        14       15
 57.0892  107.9963  85.1986  85.1986  104.5276  65.869  56.2052

> parm<-aov(y~A+x)
> parm
Call:
   aov(formula = y ~ A + x)
Terms:
                    A         x        Residuals
Sum of Squares   146.089   4032.770     20.370
Deg. of Freedom     2         1          11
Residual standard error: 1.360822

> coefficients(parm)
 (Intercept)      A2         A3          x
  15.94092     -3.02705   -2.82513    4.80320
> coinm<-aov(y~x)
> coinm
Call:
   aov(formula = y ~ x)

Terms:
                       x      Residuals
Sum of Squares     4152.107    47.122
Deg. of Freedom        1         13
Residual standard error: 1.903892

> coefficients(coinm)
  (Intercept)       x
   15.34126       4.70288
> onewm<-aov(y~A)
> onewm
Call:
   aov(formula = y ~ A)

Terms:
                    A       Residuals
Sum of Squares   146.089    4053.140
Deg. of Freedom     2         12
```

SEE NEXT PAGE

```
Residual standard error: 18.37829

> coefficients(onewm)
(Intercept)        A2          A3
  74.54            7.54        4.86
```

(i) Using the output, obtain the equations of the three regression lines that are fitted by the separate regressions model. [3]

(ii) By calculation, demonstrate that the fitted value for the first observation using Process 2 is 89.4846. [1]

(iii) Obtain an analysis of covariance table, carry out all appropriate tests and state which model you would recommend. [8]

(iv) Write down the equations of the line or lines relating to the recommended model. [2]

(v) Which plots would you produce to examine the adequacy of the model? [4]

Question 3

(a) Four treatments for asthma, labelled 0, 1, 2 and 3, are being investigated. The aim is to determine if the lung capacity of a patient is affected by the treatment used. Three patients were recruited at each of four medical centres. One treatment was allocated to each patient using a Balanced Incomplete Block Design with centres as blocks. After undergoing their allocated treatment for a fixed period, the patients' lung capacities were recorded. The design and resulting lung capacities, in brackets, are:

Block(Medical Centre)			
I	II	III	IV
0 (63)	0 (81)	1 (78)	0 (60)
2 (54)	1 (61)	2 (67)	1 (44)
3 (70)	2 (62)	3 (83)	3 (63)
187	204	228	167

(i) Write down the values of the parameters t, b, k, r, and λ for the design. [1]

(ii) Show that $T_0^* = 54$ and obtain T_1^*, T_2^* and T_3^*. Hence, obtain the sum of squares for treatments adjusted for blocks, $\text{SS}_{\text{ttments(adj)}}$. [4]

(iii) Given that $\text{SS}_{\text{Total}} = 1395$ and $\text{SS}_{\text{blocks(unadj)}} = 669\frac{2}{3}$, obtain the ANOVA table and test the hypothesis that there is no difference between treatments. [4]

(iv) Treatments 0 and 3 are two steroid inhalers. Treatments 1 and 2 are two non-steroid medications. Using the method of orthogonal contrasts, decompose $\text{SS}_{\text{ttments(adj)}}$ into 3 components representing

i. A comparison between the steroid inhalers;

ii. A comparison between the non-steroid medications;

iii. A comparison between the steroid inhalers and the non-steroid medications.

For each comparison, test the null hypothesis that the corresponding linear combination of treatment parameters is zero against a suitable alternative. [7]

(b) A further investigation into the asthma treatments detailed in part (a) is to be carried out at a single medical centre. This investigation is to involve four patients (subjects). Each will experience all four treatments over the course of four time periods. The following design is suggested with ith column representing the sequence of treatments experienced by patient i and jth row representing the jth time period.

		Subject			
		S1	S2	S3	S4
Period	P1	0	1	2	3
	P2	1	2	3	0
	P3	2	3	0	1
	P4	3	0	1	2

(i) What drawbacks are there with this design? [2]

(ii) Using the method of E.J. Williams, obtain an improved Latin square design for this experiment. [3]

(iii) The trial is to be redesigned so that a fifth treatment, labelled 4, is to be included. Obtain a design consisting of a pair of Latin squares of order 5 that is appropriate for this experiment. [4]

Question 4

(a) (i) Define what is meant by a 'saturated design'. Outline the advantages and disadvantages of using a saturated design in a screening experiment for k factors each at 2 levels. [4]

(ii) Explain how a subset of the columns of the following array can be used to accommodate a factor at four levels.

Columns						
1	2	3	12	13	23	123
-1	-1	-1	1	1	1	-1
-1	-1	1	1	-1	-1	1
-1	1	-1	-1	1	-1	1
-1	1	1	-1	-1	1	-1
1	-1	-1	-1	-1	1	1
1	-1	1	-1	1	-1	-1
1	1	-1	1	-1	-1	-1
1	1	1	1	1	1	1

The array is to be used for a $2^3 \times 4$ experiment. A, B and C, the two level factors, are assigned to the first three columns. It is proposed to assign the four level factor, D to the 4th and 7th columns and these are labelled D_1 and D_2 respectively. Explain the drawback with this allocation and suggest, with justification, a better choice of column for D_2. [6]

(b) A preliminary experiment was conducted on six factors, A, B, C, D, E and F, each at two levels, which are thought to affect the thickness of a paint coating. Four measurements were taken for each treatment combination in an eighth fraction of a 2^6 full factorial experiment. The following table shows the design and summary statistics for each treatment combination, where y_{ij} is the thickness recorded for the jth observation on the ith treatment combination.

A	B	C	D	E	F	$\sum_{j=1}^{4} y_{ij}$	$\sum_{j=1}^{4} y_{ij}^2$
1	0	1	0	0	0	3.92	3.9058
0	0	1	0	1	1	6.18	9.563
1	1	0	0	0	1	4.52	5.2362
0	1	0	0	1	0	6.95	12.0931
0	0	0	1	0	1	5.96	8.8998
1	0	0	1	1	0	3.34	2.821
0	1	1	1	0	0	8.70	18.9754
1	1	1	1	1	1	5.79	8.3881

The aim of the experiment was first to identify factors that can be used to reduce the variability of the thickness of the paint layer and then to identify other factors that can be used to reduce the average thickness of the paint layer.

(i) To identify the two factors which appear to have the largest effect on the variability of the paint layer, calculate a suitable summary statistic of the four observations at each treatment combination and obtain contrast estimates for this statistic. Hence, identify the two factors and suggest which levels of these factors should be used in order to minimise the variability. [9]

(ii) Similarly, using a different summary statistic of the observations at each treatment combination, identify two factors that appear to have a fairly large effect on the average thickness of the paint layer. Which levels of these factors do you recommend in order to reduce the average thickness? [6]

Question 5

(a) (i) Define what is meant by the 'Resolution' of a fractional factorial design. Outline the difference in estimability capabilities between fractional factorial designs of Resolutions II, III and IV. [3]

(ii) In an experimental situation, factors of interest are A, B, C, D, E, F and G, each at two levels. State, with justification, the Resolution of an eighth-replicate of a 2^7 design with defining contrasts $ABCD$, $BCEF$ and $ABCF$. [2]

(iii) Suggest, again with justification, a 'better' set of defining contrasts for the experimental situation in part (ii). [3]

(b) An experiment is to be conducted with seven factors A, B, C, D, E, F and G, each at two levels, but practical considerations dictate that only 32 observations can be made.

All interactions involving the factor A can be assumed to be negligible as can all interactions involving three or more factors.

It is required to derive a quarter-replicate without confounding any main effects or estimable interactions. It is suggested that a quarter-replicate is chosen with defining contrasts $ABCDE$ and $CDEFG$. Show that, in this case:

(i) all main effects have aliases involving three or more factors; [3]

(ii) all non-negligible two-factor interactions have aliases involving three or more factors, except for BF, BG and FG; [3]

(iii) BF, BG and FG are estimable. [2]

It is decided to choose a quarter-replicate with defining contrasts $ABCDE$ and $CDEFG$ which contains the low-level treatment combination **(1)**:

(iv) show that this quarter-replicate also contains the treatment combinations **ab**, **cd**, **ce**, **fg**, **acf**, **bdg** and **bcdeg**; [3]

(v) explain clearly how the other treatment combinations which comprise the quarter-replicate can be derived; [3]

(vi) state the number of degrees of freedom which are available for estimating the residual variance. [3]

INTERNAL EXAMINER: J.D. Godolphin
EXTERNAL EXAMINER: W. Krzanowski