University of Strathclyde Glasgow

B.Sc. FINAL EXAMINATION

53.483 DATA ANALYSIS I

Tuesday 13th January 2009
10.00a.m. - 12noon

All questions may be attempted.
Credit will be given for the best THREE answers only.

1. The following data were generated from a randomised block design with 4 treatments and 3 blocks:

	Treat 1	Treat 2	Treat 3	Treat 4	Total
Block 1	3.4	10.1	11.4	3.6	28.5
Block 2	5.9	10.2	12.6	7.2	35.9
Block 3	6.3	11.3	12.3	8.7	38.6
Total	15.6	31.6	36.3	19.5	103.0

The GLIM mathematical model for the design is

$$X_{ij} = \tilde{\mu} + \tilde{\tau}_j + \tilde{\beta}_i + \varepsilon_{ij}, \quad 1 \leq i \leq 3, \ 1 \leq j \leq 4,$$

where $\tilde{\tau}_j$ is the effect of treatment j, $\tilde{\beta}_i$ is the effect of block i and $\varepsilon_{ij} \sim N(0, \sigma_e^2)$ independently while GLIM set $\tilde{\tau}_1 = \tilde{\beta}_1 = 0$. Assume both treatment and block are fixed effects.

(a) State the GLIM commands you would use to input the data in order to produce the following results: **(4 marks)**

```
[i] ? $ fit $
[o]    deviance = 116.22
[o] residual df = 11
[o]
[i] ? $ fit + block $
[o]    deviance = 102.55  (change = -13.67)
[o] residual df = 9       (change = -2)
[o]
[i] ? $ fit + treat $
[o]    deviance = 6.6750  (change = -95.87)
[o] residual df = 6       (change = -3)
[i] ? $ disp e $
[o]                estimate      s.e.    parameter
[o] 1              3.742         0.7458  1
[o] 2              1.850         0.7458  BLOCK(2)
[o] 3              2.525         0.7458  BLOCK(3)
[o] 4              5.333         0.8612  TREAT(2)
[o] 5              6.900         0.8612  TREAT(3)
[o] 6              1.300         0.8612  TREAT(4)
[o] scale parameter 1.112
```

(b) Based on the GLIM output, carry out the ANOVA (hypothesis tests for blocks and treatments using $\alpha = 0.05$) and compute the 95% confidence intervals for $\tilde{\mu}$, $\tilde{\tau}_2$ and $\tilde{\beta}_2$. **(7 marks)**

(c) Carry out the Newman-Keuls MRT for the 4 treatments (using $\alpha = 0.05$). **(7 marks)**

(d) State the least-squares estimator for the GLIM parameter $\tilde{\mu}$ and show its variance is $\sigma_e^2/2$. **(7 marks)**

2.(a) Assume that a randomized block design has k treatments and n blocks. The mathematical model can be described by

$$X_{ij} = \mu + \beta_i + \tau_j + \epsilon_{ij}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq k$$

where μ is the grand mean, β_i the effect of block i, τ_j the effect of treatment j and ϵ_{ij} the experimental error. Assume that ϵ_{ij} are independent and follow a $N(0, \sigma_e^2)$ distribution. Assume also that blocks and treatments are both fixed effects, $\sum_{i=1}^{n} \beta_i = 0$ and $\sum_{j=1}^{k} \tau_j = 0$. Let B_i $(1 \leq i \leq n)$ and T_j $(1 \leq j \leq k)$ be the block and treatment totals, and G be the grand total. Let m, b_i and t_j be the least squares estimators for μ, β_i and τ_j, respectively. Imposing the constraints $\sum_{i=1}^{n} b_i = 0$ and $\sum_{j=1}^{k} t_j = 0$, obtain expressions for μ, β_i and τ_j in terms of B_i, T_j and G. **(13 marks)**

(b) In a 2^4 design with 2 blocks and 2 replicates, the design is arranged so that ABC and ABD are confounded with block effects in replication 1 and 2 respectively. Specify the units to go into each block for 2 replications and identify the composition of the ANOVA table.

(12 marks)

3.(a) An antibiotic can be prepared using formulations A, B or C. To study the bioequivalence of the compound, two subjects were administered each formulation and the activity of the antibiotic measured in blood samples taken from the subjects 1h, 12h, 24h and 36h after administration. The data were as follows:

		Formulation					
		1		2		3	
	Subject	1	2	1	2	1	2
Time	1h	37.5	34.8	26.9	20.4	40.1	36.1
	12h	30.1	22.1	11.3	8.4	30.3	26.3
	24h	5.4	1.6	5.5	3.9	2.1	2.3
	36h	5.2	1.4	5.8	3.3	2.2	2.1

If different subjects were used for each formulation, state a linear model that may be used to represent the data and state the assumptions of the model. State the detailed GLIM commands you would use to undertake the analysis (you need to include the code for checking assumptions).

(13 marks)

(b) The partially completed ANOVA table for a $p \times q$ design is shown below

Source	SS	df	MS	F
A	120	4		
B	16			
AB		8	6.5	
Error				
Total	290	44		

Assume A is a fixed effect and B is a random effect so this is a mixed model. Using a suitable notation, give a clear description of the model, including any assumptions necessary to undertake the analysis of variance. Fill in the blanks in the ANOVA table. Carry out the F-tests for interations AB and main effects A and B. Test using $\alpha = 0.05$.

(12 marks)

4.(a) Assume that in a randomized block design, the experimental errors are independent and follow a $N(0, \sigma_e^2)$ distribution. Assume also that $MS_{error} = 0.6$ and the d.f. for error is 20. Compute the 90% confidence interval for σ_e^2.

(6 marks)

(b) In a 2^4 design with treatments A, B, C, D, 3 replications are carried out for each of the following half of the 16 treatment combinations:

(1) a bc abc bd abd cd acd

Identify the defining contrasts and indicate aliases. Also identify the composition of the ANOVA table.

(13 marks)

(c) In a 2^5-design with treatments A, B, C, D, E, how many 2-factor interactions are there? If the 3-, 4-, 5-factor interactions are used as error, identify the composition of the ANOVA table.

(6 marks)

1(a)

```
[i] ? $units 12
[i] ? $data yield$
[i] ? $read
[i] $REA? 3.4 10.1 11.4 3.6
[i] $REA? 5.9 10.2 12.6 7.2
[i] $REA? 6.3 11.3 12.3 8.7
[i] ? $cal block=%gl(3,4)$
[i] ? $cal treat=%gl(4,1)$
[i] ? $factor block 3 treat 4$
[i] ? $look yield block treat$
[o]        YIELD     BLOCK     TREAT
[o]   1     3.400    1.000     1.000
[o]   2    10.100    1.000     2.000
[o]   3    11.400    1.000     3.000
[o]   4     3.600    1.000     4.000
[o]   5     5.900    2.000     1.000
[o]   6    10.200    2.000     2.000
[o]   7    12.600    2.000     3.000
[o]   8     7.200    2.000     4.000
[o]   9     6.300    3.000     1.000
[o]  10    11.300    3.000     2.000
[o]  11    12.300    3.000     3.000
[o]  12     8.700    3.000     4.000
[i] ? $yvar yield$
```

~ bkwk

(b)

Anova Table

Source	SS	df	MS	F
Block	13.67	2	6.68	6.1
Treat	95.87	3	31.96	28.8
Error	6.68	6	1.11	
Total	116.22	11		

~ bkwk

Hypothesis tests.

Block: H_0: Block means are the same

H_a: Block means differ

$F = 6.1 > F_{0.05}(2, 6) = 5.14$

$\therefore$ reject H_0 and accept H_a

Treat: H_0: treat means are the same

H_a: - - - differ

$F = 28.8 > F_{0.05}(3, 6) = 4.76$

$\therefore$ reject H_0 and accept H_a.

95% C.I.

$\tilde{\mu}$: $3.742 \pm t_{0.025}(6) \times 0.7458$

$= 3.742 \pm 2.447 \times 0.7458$

$= 3.742 \pm 1.825$

$\tilde{\tau}_2$: $5.333 \pm t_{0.025}(6) \times 0.8610$
$= 5.333 \pm 2.107$

$\tilde{\beta}_2$: $1.850 \pm t_{0.025}(6) \times 0.7458$
$= 1.850 \pm 1.825$

(c) Newman-Keuls MRT

			~ bkwk	
$k = 4$ means	Treat 1	Treat 4	Treat 2	Treat 3
	5.20	6.50	10.53	12.10

$s_e^2 = 1.11$, $s_e = 1.053$ s.e. of mean $= \frac{s}{\sqrt{3}} = 0.61$

Studentized Range table with df 6 and $\alpha = 5\%$.

P	2	3	4
Range	3.46	4.34	4.90

Least Significant Range (range × se of mean)

P	2	3	4
	2.11	2.65	2.99

	difference		l.s.r.	
Treat 3 v Treat 1	6.9	>	2.99	⇒ Sign.
Treat 3 v Treat 4	5.6	>	2.65	⇒ Sign.
Treat 3 v Treat 2	1.57	<	2.11	⇒ Not sign.
Treat 2 v Treat 1	5.33	>	2.65	⇒ Sign
Treat 2 v Treat 4	4.03	>	2.11	⇒ Sign
Treat 4 v Treat 1	1.30	<	2.11	⇒ Not sign

(d) The least-squares estimator for $\tilde{\mu}$ is

$$\bar{\bar{T}}_1 + \bar{B}_1 - \bar{G}$$

new

where $\bar{\bar{T}}_1 = \frac{1}{3}\sum_{i=1}^{3} X_{i1}$

$$\bar{B}_1 = \frac{1}{4}\sum_{j=1}^{4} X_{1j}$$

$$\bar{G} = \frac{1}{12}\sum_{i=1}^{3}\sum_{j=1}^{4} X_{ij}$$

Hence $\bar{\bar{T}}_1 + \bar{B}_1 - \bar{G}$

$$= \left(\frac{1}{3} + \frac{1}{4} - \frac{1}{12}\right) X_{11} + \left(\frac{1}{3} - \frac{1}{12}\right)\sum_{i=2}^{3} X_{i1}$$

$$+ \left(\frac{1}{4} - \frac{1}{12}\right)\sum_{j=2}^{4} X_{1j} - \frac{1}{12}\sum_{i=2}^{3}\sum_{j=2}^{4} X_{ij}$$

Note that X_{ij}'s are independent and $\mathrm{Var}(X_{ij\cdot}) = \sigma_e^2$.

Thus $\mathrm{Var}(\bar{\bar{T}}_1 + \bar{B}_1 - \bar{G})$

$$= \frac{1}{4}\sigma_e^2 + \frac{2}{16}\sigma_e^2 + \frac{3}{36}\sigma_e^2 + \frac{1}{144}\times 6\sigma_e^2$$

$$= \left(\frac{1}{4} + \frac{1}{8} + \frac{1}{12} + \frac{1}{24}\right)\sigma_e^2$$

$$= \frac{1}{2}\sigma_e^2$$

as required.

2(a) The l.s. estimators m, b_i, t_j minimize the sum of squares for error
new t stats

$$S(m, b_i, t_j) = \sum_{i=1}^{n} \sum_{j=1}^{k} \varepsilon_{ij}^2 = \sum_{i=1}^{n} \sum_{j=1}^{k} (X_{ij} - m - b_i - t_j)^2$$

Compute

$$\frac{\partial S}{\partial m} = \sum_{i=1}^{n} \sum_{j=1}^{k} (X_{ij} - m - b_i - t_j) \cdot 2 \cdot (-1)$$

$$\frac{\partial S}{\partial b_i} = \sum_{j=1}^{k} (X_{ij} - m - b_i - t_j) \cdot 2 \cdot (-1), \quad 1 \le i \le n$$

$$\frac{\partial S}{\partial t_j} = \sum_{i=1}^{n} (X_{ij} - m - b_i - t_j) \cdot 2 \cdot (-1), \quad 1 \le j \le k$$

Setting

$$\frac{\partial S}{\partial m} = 0, \quad \frac{\partial S}{\partial b_i} = 0, \quad \frac{\partial S}{\partial t_j} = 0$$

gives the normal equations

$$G - knm - k \sum_{i=1}^{n} b_i - n \sum_{j=1}^{k} t_j = 0$$

$$B_i - km - kb_i - \sum_{j=1}^{k} t_j = 0, \quad 1 \le i \le n$$

$$T_j - nm - \sum_{i=1}^{n} b_i - n t_j = 0, \quad 1 \le j \le k$$

Recalling $\sum_{i=1}^{n} b_i = 0$, $\sum_{j=1}^{k} t_j = 0$

we get

$$m = \frac{G}{kn},$$

$$b_i = \frac{B_i}{k} - \frac{G}{kn}, \quad 1 \le i \le n$$

$$t_j = \frac{T_j}{n} - \frac{G}{kn}, \quad 1 \le j \le k$$

2(b) 2^4 treat. combinations:

(1) a b ab c ac bc abc

d ad bd abd cd acd bcd abcd ~ bkwk

are assigned in 2 blocks as follows:

Replication 1:

ABC Confounded ⇒

Block 1:	(1)	ab	ac	bc
	d	abd	acd	bcd
Block 2:	a	b	c	abc
	ad	bd	cd	abcd

Replication 2:

ABD Confounded ⇒

Block 1:	(1)	ab	c	abc
	ad	bd	acd	bcd
Block 2:	a	b	ac	bc
	d	abd	cd	abcd

~~Replication 3:~~

~~ACD Confounded ⇒~~

~~Block 1~~	~~(1)~~	~~b~~	~~ac~~	~~abc~~
	~~ad~~	~~abd~~	~~cd~~	~~bcd~~
~~Block 2.~~	~~a~~	~~ab~~	~~c~~	~~bc~~
	~~d~~	~~bd~~	~~acd~~	~~abcd~~

~~Replication 4:~~

~~BCD Confounded ⇒~~

~~Block 1~~	~~(1)~~	~~a~~	~~bc~~	~~abc~~
	~~bd~~	~~abd~~	~~cd~~	~~acd~~
~~Block 2:~~	~~b~~	~~ab~~	~~c~~	~~ac~~
	~~d~~	~~ad~~	~~bcd~~	~~abcd~~

ANOVA

Source	d.f
Replicates	~~2~~ 1
Blocks	1
Repl. × Blocks	~~2~~ 1
A	1
B	1
C	1
D	1
AB	1
AC	1
AD	1
BC	1
BD	1
CD	1
ABC*	1
ABD*	1
ACD	1
BCD	1
ABCD	1
Error	~~9~~ 13
Total	~~31~~ 31

* Come from the replicates not confounded

3(a) If different subjects were used, the linear model for the design is $p \times q$ factorial with repeated measurements ($p = 3$ and $q = 4$, and $n = 2$ subjects).

bkwk

$$X_{ijk} = \mu + \alpha_i + \pi_{k(i)} + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$$

$$1 \le i \le 3, \quad 1 \le j \le 4, \quad 1 \le k \le 2$$

where μ — grand mean

α_i — effect of formulation i ($i = 1$ for A, 2 for B & 3 for C)

$\pi_{k(i)}$ — effect of subject k nested within formulation i

β_j — effect of time level j ($j = 1$ for 1hr, 2 for 12 hr etc)

$\alpha\beta_{ij}$ — effect of interaction between formulation i & time level j

ε_{ijk} — experimental error

Assumptions:

~~fo~~formulation & time are fixed-effects while subjects are a random effect

α_i, β_j $\alpha\beta_{ij}$ are all constant

$$\sum_{i=1}^{3} \alpha_i = 0, \quad \sum_{j=1}^{4} \beta_j = 0$$

$$\sum_{i=1}^{3} \alpha\beta_{ij} = 0 \quad \forall j = 1, 2, 3, 4$$

$$\sum_{j=1}^{4} \alpha\beta_{ij} = 0 \quad \forall i = 1, 2, 3$$

$$\left.\begin{array}{l} \pi_{k(i)} \sim N(0, \sigma_\pi^2) \\ \varepsilon_{ijk} \sim N(0, \sigma_e^2) \end{array}\right\} \text{ independently}$$

GLIM Commands:

~~bkwk~~

```
$ unit 24 $
$ data anti $
$ Read
   37.5  34.8  26.9 - - - -        36.1
   30.1  22.1  - - - - - -         26.3
   5.4   1.6   - - - -             2.3
   5.2   1.4   - - - -             2.1
$ ca form = %gl(3, 2) $
$ ca subj = %gl(2, 1) $
$ ca time = %gl(4, 6) $
$ look anti form subj time $
$ factor form 3 subj 2 time 4 $
$ yvar anti $
$ fit $
$ fit + form $
$ fit + subj. form $
$ fit + time $
$ fit + form. time $
$ disp e s v r $
$ cal res = anti - %fv $
$ plot res form $
$ plot res time $
$ plot res subj $
$ sort res $
$ ca pos = %cu(1): pos = (pos-0.5)/%nu : pos = %nd(pos) $
$ plot res pos $
$ stop
```

3(b) For a $p \times q$ design we know new

d.f. of A $= p-1 = 4 \Rightarrow p = 5$

d.f. of AB $= (p-1)(q-1) = 4(q-1) = 8 \Rightarrow q = 3$

d.f. of total $= pqn - 1 = 15n - 1 = 44 \Rightarrow n = 3$

So this is a 5×3 design with $n=3$ replications.

The math. model is

$$X_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$$

$$1 \le i \le 3, \quad 1 \le j \le 3, \quad 1 \le k \le 3$$

where

μ is the grand mean

α_i is the effect of factor A level i

β_j is - - - - B level j

$\alpha\beta_{ij}$ is the interaction effect of factor A level i and factor B level j

ε_{ijk} is the experimental error

Given that this is a mixed model, the assumptions are:

α_i's are constant, $\sum_{i=1}^{5} \alpha_i = 0$

$$\left.\begin{array}{l} \beta_i \sim N(0, \sigma_B^2) \\ \alpha\beta_{ij} \sim N(0, \sigma_{AB}^2) \\ \varepsilon_{ijk} \sim N(0, \sigma_e^2) \end{array}\right\} \text{independently}$$

ANOVA

Source	SS	d.f.	MS	F
A	120	4	30	$MS_A / MS_{AB} = \frac{30}{6.5} = 4.615$
B	16	2	8	$MS_B / MS_{Error} = \frac{8}{3.4} = 2.35$
AB	52	8	6.5	$MS_{AB} / MS_{Error} = \frac{6.5}{3.4} = 1.911$
Error	102	30	3.4	
Total	290	44		

Hypothesis tests:

AB. $H_0: \sigma_{AB}^2 = 0$

$H_a: \sigma_{AB}^2 > 0$

$F = \frac{MS_{AB}}{MS_{error}} = \frac{6.5}{3.4} = 1.911$

$F_{0.05}(8, 30) = 2.27$

$\because F < F_{0.05}(8, 30)$

$\therefore$ do not reject H_0, i.e. AB is not sign

A: $H_0: \alpha_1 = \cdots = \alpha_5 = 0$

$H_a: \alpha_i$'s differ

$F = \frac{MS_A}{MS_{AB}} = \frac{30}{6.5} = 4.615$

$F_{0.05}(4, 8) = 3.84$

$\because F > F_{0.05}(4, 8)$

$\therefore$ reject H_0 and accept H_a, i.e. A is sign.

B: $H_0: \sigma_B^2 = 0$

$H_a: \sigma_B^2 > 0$

$F = \frac{MS_B}{MS_{error}} = \frac{8}{3.4} = 2.35$

$F_{0.05}(2, 30) = 3.32$

$\because F < F_{0.05}(2, 30)$

$\therefore$ don't reject H_0, i.e. B is not sign.

4@ Given $MS_{error} = 0.6$ with the d.f. = 20.
the 90% C.I. for σ_e^2 is ~bkwk

$$\frac{20\, MS_{error}}{\chi^2_{20}(0.05)} < \sigma_e^2 < \frac{20\, MS_{error}}{\chi^2_{20}(0.95)}$$

From the χ^2-table.

$\chi^2_{20}(0.05) = 31.41$

$\chi^2_{20}(0.95) = 10.85$

Hence
$$\frac{20 \times 0.6}{31.41} < \sigma_e^2 < \frac{20 \times 0.6}{10.85}$$

$$0.382 < \sigma_e^2 < 1.11$$

~bkwk

4(b) the $\frac{1}{2}$-replicate is
(1) a bc abc bd abd cd acd

$A \times$ (the $\frac{1}{2}$-repl.) (mod. 2) yields
a (1) abc bc abd bd acd cd
= the same $\frac{1}{2}$-repl.
so $A \notin$ the defining contrast

$B \times$ (the $\frac{1}{2}$-repl.) (mod. 2) yields
b ab c ac d ad bcd abcd
= the other $\frac{1}{2}$-repl.
so $B \in$ the defining contrast

$C \times$ (the $\frac{1}{2}$-repl.) (mod. 2) yields
c ac b ab bcd abcd d ad
= the other $\frac{1}{2}$-repl.
so $C \in$ the defining contrast

D × (the ½ -repl.) (mod. 2) yields

d ad bcd abcd is ab c ac

= the other ½ -repl

so D ∈ the defining contrast

Here the defining contrast is BCD

Factor/interaction	Aliase
A	ABCD
B	CD
C	BD
D	BC
AB	ACD
AC	ABD
AD	ABC

AMOVA Table

Source	d.f
Repl.	2
A or ABCD	1
B or CD	1
C or BD	1
D or BC	1
AB or ACD	1
AC or ABD	1
AD or ABC	1
Error	14
Total	23

(c) In the 2^5-design, there are

$$\binom{5}{2} = \frac{5\times 4}{2} = 10$$

new

2-factor interactions. The ANOVA Table is as follows

Source	d.f
A	1
B	1
C	1
D	1
E	1
AB	1
AC	1
AD	1
AE	1
BC	1
BD	1
BE	1
CD	1
CE	1
DE	1
Error	16
Total	31