SOLUTIONS

UNIVERSIDADE DE SÃO PAULO
Campus "Luiz de Queiroz"
ESCOLA SUPERIOR DE AGRICULTURA "LUIZ DE QUEIROZ"

DEPARTMENTO DE CIÊNCIAS EXATAS

Estatística e Experimentação Agronômica

LCE802 Planejamento de Experimentos e Análise de Dados

FINAL PAPER

PART A — Practical

General Instructions to Candidates **Time allowed: 4 hours**

1. This examination is an open book examination. You may need a calculator and statistical tables.

2. Attempt **ALL** questions. The marks for each question are shown with the question and in the following table.

PART	Question	Max. Mark
A — Practical	1	14
	2	34
	3	15
B — Theory	4	7
	5	6
	6	7
	7	12
	Total	95

3. In carrying out statistical procedures, use a 5% level of significance and a 95% confidence level, unless otherwise specified. Also, you should carry out the complete procedure even if the assumptions underlying it are not met.

Q1 An experiment was conducted to investigate the effects of four diets on the growth of young chicks. There were 16 cages of 15 chicks with the cages arranged in four rows (layers) by four columns (stacks). The four diets were assigned to the cages using a Latin square design. The four diets were:

C: Control diet that includes lysine;
L: Lysine enriched diet that is C plus 0.2% artificial lysine;
W: Wheat-added diet that is C, except that some starch and sugar of C is replaced with sufficient wheat to add an extra 0.2% lysine — note that, because wheat naturally contains amino acids, this diet also has additional amino acids to C and L;
A: Amino enriched diet that is L, plus the amino acids added naturally by the wheat to diet W.

These 4 treatments form natural groupings as follows:

1. (C) versus (L, W, A) compares no lysine with lysine;
2. within the lysine treatments, (L) versus (W, A) compares no amino acid to amino acid;
3. within the amino acid treatments, (W) versus (A) compares the two sources of amino acid.

The mean weights of the 15 chicks in each cage are given in the following table:

	Stack							
	1		2		3		4	
Layer	Diet	Weight	Diet	Weight	Diet	Weight	Diet	Weight
1	C	161	L	228	A	224	W	190
2	L	246	W	199	C	169	A	210
3	W	186	A	251	L	241	C	175
4	A	229	C	170	W	183	L	247

a) Enter the data into Genstat and obtain a printout out the factors and variates that have been entered.

```
18  PRINT Layer,Stack,Diet,Weight; DEC=0

      Layer        Stack         Diet       Weight
          1            1            C          161
          1            2            L          246
          1            3            W          186
          1            4            A          229
          2            1            L          228
          2            2            W          199
          2            3            A          251
          2            4            C          170
          3            1            A          224
          3            2            C          169
          3            3            L          241
          3            4            W          183
          4            1            W          190
          4            2            A          210
          4            3            C          175
          4            4            L          247
```

[6 marks]

b) Perform the analysis of variance for a Latin square that includes a source for Diets with 3 degrees of freedom.

```
BLOCK Layer*Stack
TREAT Diet
ANOVA [FPROB=y] Weight
```

```
21.................................................................................

***** Analysis of variance *****

Variate: Weight

Source of variation     d.f.       s.s.       m.s.    v.r.  F pr.

Layer stratum              3      147.7       49.2    0.34

Stack stratum              3      316.2      105.4    0.72

Layer.Stack stratum
Diet                       3    13414.7     4471.6   30.49  <.001
Residual                   6      879.9      146.6

Total                     15    14758.4


* MESSAGE: the following units have large residuals.

Layer 4    Stack 2           -16.4   s.e. 7.4


***** Tables of means *****

Variate: Weight

Grand mean  206.8

     Diet        C        L        W        A
             168.8    240.5    189.5    228.5


*** Standard errors of differences of means ***

Table                Diet
rep.                    4
d.f.                    6
s.e.d.               8.56
```

[3 marks]

c) Using either orthogonal contrasts or nested factors, perform a second analysis of variance in which the Diets sums of squares has been partitioned into 3 single-degree-of-freedom sources corresponding to the three natural groupings of the four diets.

```
FACTOR [LEV=2] Lysine,Amino,Source
CALC Lysine=NEWLEVELS(Diet; !v(1,2,2,2))
&    Amino=NEWLEVELS(Diet; !v(1,1,2,2))
&    Source=NEWLEVELS(Diet; !v(1,1,1,2))
TREAT Lysine/Amino/Source
ANOVA [FPROB=y] Weight
```

```
27...........................................................................

***** Analysis of variance *****

Variate: Weight

Source of variation     d.f.       s.s.       m.s.    v.r.  F pr.

Layer stratum              3      147.7       49.2    0.34

Stack stratum              3      316.2      105.4    0.72

Layer.Stack stratum
Lysine                     1     7726.7     7726.7   52.69  <.001
Lysine.Amino               1     2646.0     2646.0   18.04  0.005
Lysine.Amino.Source        1     3042.0     3042.0   20.74  0.004
Residual                   6      879.9      146.6

Total                     15    14758.4


* MESSAGE: the following units have large residuals.

Layer 4    Stack 2           -16.4   s.e. 7.4


***** Tables of means *****

Variate: Weight

Grand mean  206.8

   Lysine        1        2
             168.8    219.5
     rep.        4       12

   Lysine    Amino        1        2
        1             168.8
              rep.        4
        2             240.5    209.0
              rep.        4        8

             Amino        1                 2
   Lysine   Source        1        2        1        2
        1             168.8
        2             240.5             189.5    228.5


*** Standard errors of differences of means ***

Table               Lysine      Lysine      Lysine
                                 Amino       Amino
                                            Source
rep.               unequal     unequal           4
d.f.                     6           6           6
s.e.d.                            8.56             min.rep
                      6.99        7.42        8.56  max-min
                                  6.05X             max.rep

(No comparisons in categories where s.e.d. marked with an X)
```

[5 marks]

[Total: 14 marks = 6 + 3 + 5 marks]

Q2 An experiment was conducted to compare the effects of 5 row spacings on the yields of 2 varieties of soya beans. The 2 varieties were assigned to plots using a randomized complete block design of 6 blocks. Each plot was subdivided into 5 subplots and the 5 row spacings randomized to the subplots within each plot. The yields, in unrandomized order, are given in the table below.

Data for a soya bean experiment

		Block					
Variety	Spacing	1	2	3	4	5	6
OM	18	33.6	37.1	34.1	34.6	35.4	36.1
	24	31.1	34.5	30.5	32.7	30.7	30.3
	30	33.0	29.5	29.2	30.7	30.7	27.9
	36	28.4	29.9	31.6	32.3	28.1	26.9
	42	31.4	28.3	28.9	28.6	*	33.4
B	18	28.0	25.5	28.3	29.4	27.3	28.3
	24	23.7	26.2	27.0	25.8	26.8	23.8
	30	23.5	26.8	24.9	23.3	21.4	22.0
	36	25.0	25.3	25.6	26.4	24.6	24.5
	42	25.7	23.2	23.4	25.6	24.5	22.9

An initial analysis of variance for the data is appended to this exam paper. The data is in the Genstat spreadsheet file *SplSoyMV.gsh* in *G:\Disciplina\Genstat*.

a) Assuming that the unrandomized factors are random and the randomized factors are fixed, what are the maximal expectation and variation models, the variance components and their multipliers and the expected mean squares for the appended analysis?

The maximal expectation and variation models are:

E[Y] = Variety.Spacing
Var[Y] = Blocks + Blocks.Plots + Blocks.Plots.Subplots

The variance components and their multipliers are σ^2_{BPS} *,* $5\sigma^2_{\text{BP}}$ *and* $10\sigma^2_{\text{B}}$.

Source	df	E[MSq]
Blocks	5	$\sigma^2_{\text{BPS}} + 5\sigma^2_{\text{BP}} + 10\sigma^2_{\text{B}}$
Blocks.Plots	6	
Variety	1	$\sigma^2_{\text{BPS}} + 5\sigma^2_{\text{BP}} + f_{\text{V}}(\psi)$
Residual	5	$\sigma^2_{\text{BPS}} + 5\sigma^2_{\text{BP}}$
Blocks.Plots.Subplots	48	
Spacing	4	$\sigma^2_{\text{BPS}} + f_{\text{S}}(\psi)$
Variety.Spacing	4	$\sigma^2_{\text{BPS}} + f_{\text{SF}}(\psi)$
Residual	40	σ^2_{BPS}

[9 marks]

b) Based on the appended analysis, what conclusions do you draw about the effects of the factors Variety and Spacing on the yield of soya beans? Make sure you indicate how you reached your conclusions.

The interaction of Variety and Spacing is not significant but both main effect are significant. This is because the p-value of the interaction is less than 0.05 but the p-

values for the main effects are less than 0.05. Consequently both Spacing and Variety affects the yield of soya bean, but they do so independently.

[5 marks]

c) Perform diagnostic checking and write a summary of your conclusions from each graph or analysis.

The output containing the diagnostic checking for this example is given below.

```
38  APLOT METHOD=fit,normal
                                                                              *
             -
             I                             *
             I
             I     *
         2.5 I                          *    *       *
r            I                              *
e            I      *    *     *
s            I                               *                 *   *
i            I     * *      *         *    *                   *
d            I       *  *  **      *                *
u        0.0 I    *     ** *      *  *      *  *
a            I   *     *     *             *     *             *
l            I *      *                  **      2 *               *
s            I       ***                   *   *
             I            *                                      *
             I   *                       * *  *
        -2.5 I                      *         *
             -+---------+---------+---------+---------+---------+---------+---
           22.5      25.0      27.5      30.0      32.5      35.0      37.5

                                         fitted values

                                          Normal plot
                                                                              *
             -
             I
         4.0 I                                                           *
r            I                                                      *
e            I                                                 * *
s            I                                              * *
i            I                                         2****
d            I                                  *2*2*2*
u        0.0 I                            22*222*
a            I                      *2*222
l            I                 **2*2
s            I      *  * ** ***
             I *
             I
        -4.0 I
             -+---------+---------+---------+---------+---------+---------+---
           -2.4      -1.6      -0.8       0.0       0.8       1.6       2.4

                                   expected Normal quantiles

  39  "
 -40  **** Tukey''s one-degree-of-freedom-for-non-additivity.
 -41  **** It is the term designated covariate in the following analysis
 -42  "
  43  AKEEP [FIT=Fit]
  44  CALC ResSq=Fit*Fit
  45  ANOVA [PRINT=*]  ResSq; RES=ResSq
  46  COVAR ResSq                         "A computational trick"
  47  ANOVA [PRINT=A; FPROB=Y] Yields


47................................................................................

***** Analysis of variance (adjusted for covariate) *****
```

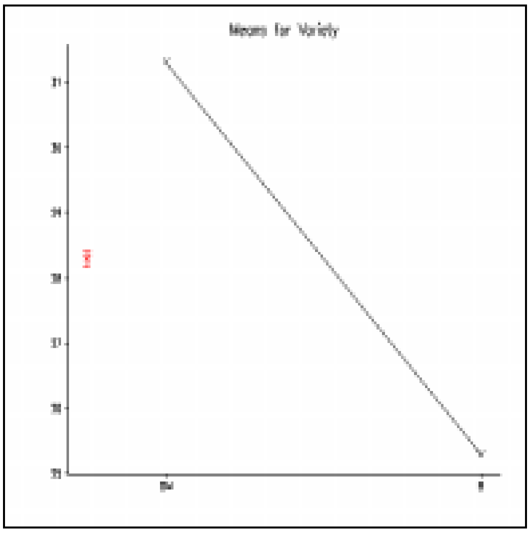
```
Variate: Yields
Covariate: ResSq

Source of variation        d.f.(m.v.)       s.s.        m.s.     v.r. cov.ef.   F pr.

Blocks stratum
Covariate                     1            1.760       1.760     0.72            0.443
Residual                      4            9.713       2.428     2.21     0.94

Blocks.Plots stratum
Variety                       1          453.005     453.005   412.01     0.83  <.001
Covariate                     1            0.002       0.002     0.00            0.968
Residual                      4            4.398       1.100     0.37     0.80

Blocks.Plots.SubPlots stratum
Spacing                       4          171.896      42.974    14.59     1.00  <.001
Variety.Spacing               4           15.060       3.765     1.28     1.00  0.296
Covariate                     1            0.289       0.289     0.10            0.756
Residual                     38(1)       111.941       2.946              0.98

Total                        58(1)       855.752

  48  COVAR
```

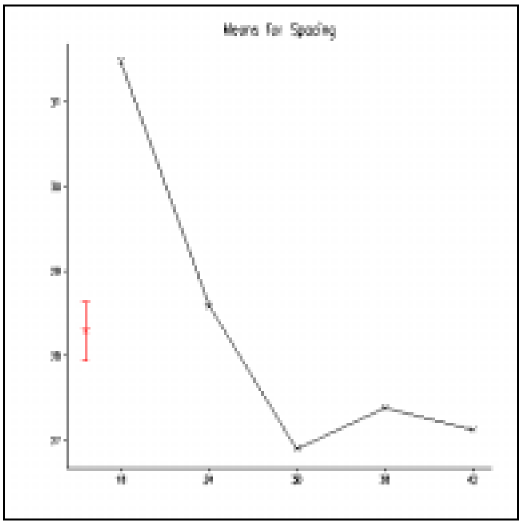
The residual-versus-fitted-values plot is showing no particular pattern and so is satisfactory. The normal probability plot appears to be displaying some curvature and so the assumption of normality is in doubt. Tukey's one-degree-of-freedom-for-nonadditivity is not significant and so the additivity assumption appears to be met.

[7 marks]

d) Given your conclusions in b), obtain plots of appropriate tables of means. Using either multiple comparisons or polynomial submodels, examine the differences between the treatments. Give reasons for your choice. If you use multiple comparisons, summarize what you conclude about the differences. If you use polynomial submodels, summarize what you can say about the fitted trend and give the fitted equation(s).

The plots of the Variety and Spacing means are given in the following diagrams:





In this case multiple comparisons are appropriate for Varieties as these are qualitative, although they are not required because there are only two means. As Spacing is quantitative orthogonal polynomials are appropriate. So the analysis is repeated to include the orthogonal polynomials and the output is shown below.

```
  49  BLOCK Blocks/Plots/SubPlots
  50  TREAT Variety*POL(Spacing; 2)
  51  ANOVA [FPROB=Y] Yields


51...................................................................................

***** Analysis of variance *****

Variate: Yields

Source of variation     d.f.(m.v.)       s.s.        m.s.    v.r.  F pr.

Blocks stratum              5          11.473       2.295    2.61

Blocks.Plots stratum
Variety                     1         542.673     542.673  616.67  <.001
Residual                    5           4.400       0.880    0.31

Blocks.Plots.SubPlots stratum
Spacing                     4         171.774      42.944   14.92  <.001
  Lin                       1         117.632     117.632   40.88  <.001
  Quad                      1          47.138      47.138   16.38  <.001
  Deviations                2           7.004       3.502    1.22  0.307
Variety.Spacing             4          15.049       3.762    1.31  0.284
  Variety.Lin               1           7.013       7.013    2.44  0.127
  Variety.Quad              1           1.842       1.842    0.64  0.429
  Deviations                2           6.194       3.097    1.08  0.351
Residual                  39(1)       112.230       2.878

Total                     58(1)       855.752


* MESSAGE: the following units have large residuals.

Blocks 6    Plots 1    SubPlots 5            3.74   s.e. 1.37
```

From this output there is no significant deviations from the quadratic equation fitted to the overall spacing means. However, the significant quadratic terms means that a quadratic equation is required. The Genstat regression output to achieve this is given below. The fitted equations is:

$$Yield = 44.92 - 1.006\,Spacing + 0.01390\,Spacing^2$$

Also from the Variety mean plot or table of means it is concluded that OM has a greater yield than B.

However, the significance of the Varieties means that the fitted equation for spacings has different intercepts for each variety. So one could obtain these equations, which would then give the fitted values. The Genstat regression output below produces them and they are:

For Variety OM, $Yield = 48.62 - 1.062\,Spacing + 0.01498\,Spacing^2$

For Variety B, $Yield = 42.58 - 1.062\,Spacing - 0.01498\,Spacing^2$

```
  52  "
 -53  **** Use regression to obtain the fitted equation
 -54  "
  55  VARI [60] Spacings
  56  CALC Spacings=Spacing & SpaceSq=Spacings*Spacings
  57  MODEL Yields
  58  TERMS Spacings+SpaceSq
  59  FIT Spacings+SpaceSq

59................................................................................


***** Regression Analysis *****

 Response variate: Yields
     Fitted terms: Constant + Spacings + SpaceSq

*** Summary of analysis ***

              d.f.         s.s.         m.s.      v.r.
Regression       2        168.5        84.24      6.86
Residual        56        687.3        12.27
Total           58        855.8        14.75

Percentage variance accounted for 16.8
Standard error of observations is estimated to be 3.50
* MESSAGE: The residuals do not appear to be random;
           for example, fitted values in the range 26.73 to 27.20
           are consistently larger  than observed values
           and fitted values in the range 27.20 to 27.26
           are consistently smaller than observed values


*** Estimates of parameters ***

                              estimate         s.e.     t(56)
Constant                         44.92         6.48      6.93
Spacings                        -1.006        0.458     -2.20
SpaceSq                        0.01390      0.00760      1.83

  60  TERMS Variety+Spacings+SpaceSq
  61  FIT Variety+Spacings+SpaceSq

61................................................................................


***** Regression Analysis *****

 Response variate: Yields
     Fitted terms: Constant + Variety + Spacings + SpaceSq


*** Summary of analysis ***

              d.f.         s.s.         m.s.      v.r.
Regression       3        706.2      235.399     86.57
Residual        55        149.6        2.719
Total           58        855.8       14.754

Percentage variance accounted for 81.6
Standard error of observations is estimated to be 1.65


*** Estimates of parameters ***
                              estimate         s.e.     t(55)
Constant                         48.62         3.06     15.88
Variety B                       -6.042        0.430    -14.06
Spacings                        -1.062        0.216     -4.93
SpaceSq                        0.01498      0.00358      4.19
```

[9 marks]

e) From the appended analysis, do you think that the precision of the Variety main effects is less than it might have been because of the use of a split-plot design? Provide evidence for your conclusion.

No, because the variability of main plot is no greater than subplots as evidenced by the nonsignificant v.r. for the Residual from Blocks.Plots (p = 0.1581).

```
35  CALC pB=1-FPROB(2.295 / 0.88; 5; 5)
36  &    pBP=1-FPROB(0.88 / 2.878; 5; 39)
37  PRINT pB,pBP

        pB         pBP
    0.1581      0.9065
```

[4 marks]

[Total: 34 marks = 9 + 5 + 7 + 9 + 4 marks]

Q3 An experiment is to be conducted to investigate suspected cobalt and copper deficiency of pasture. A flock of sheep is to be run on a paddock suspected to be deficient in these trace elements. The experiment will be a 2^2 factorial experiment made up of the two factors cobalt (administered or not administered) and copper (administered or not administered). The treatments will be randomized to the sheep using a completely randomized design and all sheep will be grazed on the suspect pasture. It is believed that the variance $\left(\sigma^2\right)$ of sheep within a treatment will be about 30 kg. It is desired to be able to detect a change of 4 kg in the cobalt difference between the two copper treatments with power of at least 80% and with significance 0.05.

a) How many replicates should there be of each treatment combination to achieve the desired power?

Sample Size (r)	alpha	DF numer -ator	DF denomin- ator	central F	no. values in a mean (m)	delta	standard deviation	lambda	power
30	0.05	1	116	3.9229	30	4	5.477226	8	0.8009

To achieve a power of at least 80% requires 30 replicates.

[9 marks]

b) Using Genstat to generate a layout for the experiment with the number of replicates you have computed in a). Use a seed of 443322. If you have not got a satisfactory answer to a), generate a design for 15 replicates. Obtain a printout of the layout.

```
Genstat 5  Release 4.1  (PC/Windows NT)              11 May 2000 21:21:58
Copyright 1998, Lawes Agricultural Trust (Rothamsted Experimental Station)

                  ______________________________________________

                  Genstat 5 Fourth Edition - (for Windows)
                  Genstat 5 Procedure Library Release PL11
                  ______________________________________________

   3  DESIGN
   4  PDESIGN [BLOCK=Sheep; TREAT=Cobalt,Copper]

 *** Treatment combinations on each unit of the design ***
```

Sheep			31	2	2	62	2	2	93	2	2
1	1	1	32	1	1	63	1	1	94	1	2
2	1	1	33	2	1	64	1	2	95	1	1
3	1	2	34	2	1	65	1	2	96	1	1
4	2	1	35	1	2	66	2	2	97	1	1
5	1	1	36	2	1	67	2	1	98	1	2
6	2	2	37	2	1	68	2	2	99	1	1
7	1	1	38	2	2	69	2	2	100	2	2
8	2	1	39	1	2	70	2	2	101	2	1
9	1	1	40	1	2	71	1	2	102	1	2
10	2	1	41	1	2	72	2	1	103	1	1
11	2	1	42	2	1	73	1	2	104	1	2
12	2	2	43	1	2	74	1	2	105	2	1
13	2	1	44	2	2	75	2	2	106	1	1
14	2	1	45	1	1	76	2	1	107	1	1
15	1	2	46	2	1	77	2	2	108	2	2
16	1	2	47	2	2	78	1	1	109	1	2
17	1	2	48	2	1	79	1	2	110	2	1
18	2	1	49	1	1	80	1	2	111	1	1
19	2	2	50	1	1	81	2	2	112	2	1
20	1	1	51	2	1	82	1	2	113	2	2
21	2	2	52	2	1	83	2	1	114	1	2
22	2	1	53	2	2	84	2	2	115	2	2
23	1	2	54	2	2	85	1	1	116	2	1
24	1	1	55	2	2	86	2	2	117	1	1
25	2	2	56	1	1	87	1	1	118	1	1
26	1	2	57	1	2	88	1	2	119	1	1
27	2	1	58	1	2	89	2	1	120	2	2
28	2	1	59	1	1	90	2	2			
29	1	1	60	2	2	91	1	2			
30	1	2	61	2	1	92	1	1			

Treatment factors are listed in the order: Cobalt Copper

[6 marks]

[Total: 15 marks = 9 + 6 marks]

PART B — Theory

General Instructions to Candidates **Time allowed: 2 hours**

Q4 An experiment is to be conducted on the effects of air and soil heating on the yield of glasshouse tomatoes. There are 8 compartments of a glasshouse to be used in the experiment with these compartments arranged in four rows each containing two compartments. Two air temperatures (55 and 60 C) were randomized to the two compartments in each row. In each compartment there were two long troughs divided in half with the soil in one half of each trough to be heated and the other half not heated. For each trough the half to be unheated was randomly selected. Also, the two temperatures (65 and 75 C) for the heated half was randomly assigned to the two troughs in a compartment. In each half trough 4 tomato plants were planted and the total yield of the four plants was measured.

What are the components of this study?

1. *the observational unit* – *a half-trough*
2. *response variable* – *Yield*
3. *unrandomized factors* – *Rows, Compartments, Troughs, Half-troughs*
4. *randomized factors* – *Air temperature, Soil temperature*
5. *type of study* – *Split-plot design with main plots in an RCBD*

[Total: 7 marks]

Q5 An experiment is conducted to investigate the differences between eight cultivars of corn. The experimental area is divided into four rows by four columns. Each row-column area is divided into eight plot, each plot consisting of 3 lines of 20 plants. The eight cultivars are randomized to the eight plots in each row-column area. The centre line of each plot is harvested and the total yield of the 20 plants is recorded.

The components of this experiment are:

1.	the observational unit	a plot
2.	response variable	Yield
3.	unrandomized factor	Rows, Columns, Plots
4.	randomized factors	Cultivar
5.	type of study	?

a) What is the experimental structure, including the numbers of levels of each of the factors, for the experiment?

Structure	*Formula*
unrandomized	*(4 Rows*4 Columns)/8 Plots*
randomized	*8 Cultivars*

[3 marks]

b) What type of study is this? Give reasons for your answer.

An RCBD because the row-column combinations form the blocks and the plots are randomized within these.

[1 marks]

c) Which of the factors are likely to be random and which fixed? Give reasons for your answer.

Plots is likely to be random and Rows, Columns and Cultivars are likely to be fixed. It is likely that it is intended that the plots are representative of all a large population of plots and it would seem likely that their effects could be represented as effects with a common mean and displaying some variance. On the other hand, Rows, Columns and Cultivars are likely to differ systematically and a distribution would not be a suitable model for their effects.

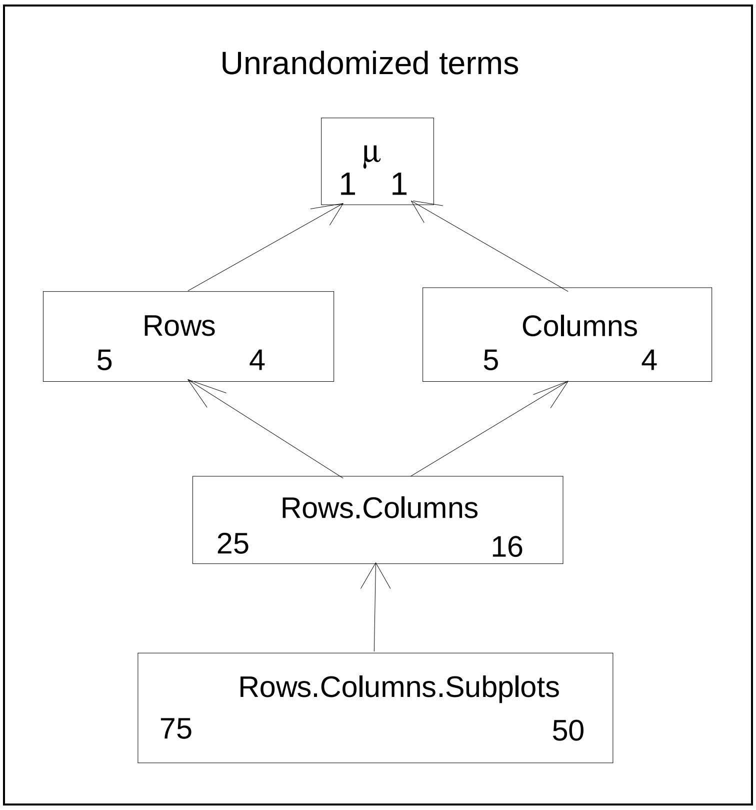
[2 marks]

[Total: 6 marks = 3 + 1 + 2 marks]

Q6 An experiment that employs a split-plot design with five main-plot treatments assigned using a Latin square and three subplot treatments assigned completely at random would have the following unrandomized structure:

(*5* Rows**5* Columns)/*3* Subplots
= Rows + Columns + Rows.Columns + Rows.Columns.Subplots

What is the Hasse diagram of term marginalities, including degrees of freedom, for these unrandomized terms?



[Total: 7 marks]

Q7 Suppose the data from a Latin square experiment with t treatments is to be analysed. Let $\psi = E[\mathbf{Y}] = \mathbf{X}_{\mathrm{T}}\tau$, $\mathbf{V} = \sigma^2\mathbf{I}_n$, and $R(\tau \mid \mu) = \mathbf{Y}'\mathbf{P}_{\mathrm{T}}\mathbf{R}_{\mathrm{G}}\mathbf{Y}$ where $\mathbf{R}_{\mathrm{G}} = \mathbf{I} - \mathbf{P}_{\mathrm{G}}$, $\mathbf{P}_{\mathrm{G}} = t^{-2}\mathbf{J}_t \otimes \mathbf{J}_t$ and $\mathbf{P}_{\mathrm{T}} = \mathbf{X}_{\mathrm{T}}\left(\mathbf{X}'_{\mathrm{T}}\mathbf{X}_{\mathrm{T}}\right)^{-1}\mathbf{X}'_{\mathrm{T}}$. Also, it is known that $\mathbf{P}_{\mathrm{T}}\mathbf{P}_{\mathrm{G}} = \mathbf{P}_{\mathrm{G}}\mathbf{P}_{\mathrm{T}} = \mathbf{P}_{\mathrm{G}}$ and $trace\left(\mathbf{P}_{\mathrm{T}}\mathbf{R}_{\mathrm{G}}\right) = t - 1$.

a) Prove that $\mathbf{P}_{\mathrm{T}}\mathbf{R}_{\mathrm{G}}$ is symmetric and idempotent, assuming that $\mathbf{P}_{\mathrm{T}}$ and $\mathbf{P}_{\mathrm{G}}$ are symmetric and idempotent.

Firstly note that $\mathbf{P}_{\mathrm{T}}\mathbf{R}_{\mathrm{G}} = \mathbf{P}_{\mathrm{T}}\left(\mathbf{I} - \mathbf{P}_{\mathrm{G}}\right) = \mathbf{P}_{\mathrm{T}} - \mathbf{P}_{\mathrm{G}}$.

Since $\mathbf{P}_{\mathrm{T}}$ *and* $\mathbf{P}_{\mathrm{G}}$ *are symmetric* $\left(\mathbf{P}_{\mathrm{T}} - \mathbf{P}_{\mathrm{G}}\right)' = \mathbf{P}'_{\mathrm{T}} - \mathbf{P}'_{\mathrm{G}} = \mathbf{P}_{\mathrm{T}} - \mathbf{P}_{\mathrm{G}}$ *so that* $\mathbf{P}_{\mathrm{T}}\mathbf{R}_{\mathrm{G}}$ *is symmetric.*

To show that $\mathbf{P}_{\mathrm{T}}\mathbf{R}_{\mathrm{G}}$ *is idempotent, we have that, since* $\mathbf{P}_{\mathrm{T}}$ *and* $\mathbf{P}_{\mathrm{G}}$ *are idempotent and* $\mathbf{P}_{\mathrm{T}}\mathbf{P}_{\mathrm{G}} = \mathbf{P}_{\mathrm{G}}\mathbf{P}_{\mathrm{T}} = \mathbf{P}_{\mathrm{G}}$,

$$\begin{aligned}\left(\mathbf{P}_{\mathrm{T}} - \mathbf{P}_{\mathrm{G}}\right)\left(\mathbf{P}_{\mathrm{T}} - \mathbf{P}_{\mathrm{G}}\right) &= \mathbf{P}_{\mathrm{T}}\mathbf{P}_{\mathrm{T}} - \mathbf{P}_{\mathrm{T}}\mathbf{P}_{\mathrm{G}} - \mathbf{P}_{\mathrm{G}}\mathbf{P}_{\mathrm{T}} + \mathbf{P}_{\mathrm{G}}\mathbf{P}_{\mathrm{G}} \\ &= \mathbf{P}_{\mathrm{T}} - \mathbf{P}_{\mathrm{G}} - \mathbf{P}_{\mathrm{G}} + \mathbf{P}_{\mathrm{G}} \\ &= \mathbf{P}_{\mathrm{T}} - \mathbf{P}_{\mathrm{G}}\end{aligned}$$

[4 marks]

b) Prove that

$$E\left[R(\tau \mid \mu)/(t-1)\right] = \sigma^2 + f_{\mathrm{T}}(\psi)$$

where $f_{\mathrm{T}}(\psi) = \sum_{j=1}^{t} t\left(\tau_j - \overline{\tau}_.\right)^2 \Big/ (t-1)$, $\overline{\tau}_. = \sum_{j=1}^{t} \tau_j \big/ t$, τ_j is the jth element of the t-vector τ.

First use theorem II.11 to show that

$$\begin{aligned}
E\left[R(\tau \mid \mu)/(t-1)\right] &= E\left[\mathbf{Y}'\mathbf{P}_{\mathrm{T}}\mathbf{R}_{\mathrm{G}}\mathbf{Y}\right]/(t-1) \\
&= \left\{trace\left(\mathbf{P}_{\mathrm{T}}\mathbf{R}_{\mathrm{G}}\sigma^2\mathbf{I}_n\right) + \left(\mathbf{X}_{\mathrm{T}}\tau\right)'\mathbf{P}_{\mathrm{T}}\mathbf{R}_{\mathrm{G}}\left(\mathbf{X}_{\mathrm{T}}\tau\right)\right\} \Big/ \{t-1\} \\
&= \left\{\sigma^2 trace\left(\mathbf{P}_{\mathrm{T}}\mathbf{R}_{\mathrm{G}}\right) + \left(\mathbf{X}_{\mathrm{T}}\tau\right)'\mathbf{P}_{\mathrm{T}}\mathbf{R}_{\mathrm{G}}\left(\mathbf{X}_{\mathrm{T}}\tau\right)\right\} \Big/ \{t-1\}
\end{aligned}$$

Now, $trace\left(\mathbf{P}_{\mathrm{T}}\mathbf{R}_{\mathrm{G}}\right) = t-1$.

Also,

$$\mathbf{P}_{\mathrm{T}}\mathbf{R}_{\mathrm{G}}\left(\mathbf{X}_{\mathrm{T}}\tau\right) = \left(\mathbf{P}_{\mathrm{T}} - \mathbf{P}_{\mathrm{G}}\right)\mathbf{X}_{\mathrm{T}}\tau = \mathbf{P}_{\mathrm{T}}\mathbf{X}_{\mathrm{T}}\tau - \mathbf{P}_{\mathrm{G}}\mathbf{X}_{\mathrm{T}}\tau$$

Now, from lemma III.2, $\mathbf{P}_{\mathrm{T}}\mathbf{X}_{\mathrm{T}}\tau = \mathbf{X}_{\mathrm{T}}\tau$. *Also,* $\mathbf{P}_{\mathrm{G}}\mathbf{X}_{\mathrm{T}}\tau = t^{-2}\mathbf{J}_t \otimes \mathbf{J}_t\mathbf{X}_{\mathrm{T}}\tau = \overline{\tau}_.\mathbf{1}_{t^2}$

Consequently,

$$\left(\mathbf{P}_{\mathrm{T}} - \mathbf{P}_{\mathrm{G}}\right)\mathbf{X}_{\mathrm{T}}\tau = \mathbf{X}_{\mathrm{T}}\tau - \overline{\tau}_.\mathbf{1}_{t^2}$$

Hence the expected mean square is

$$\begin{aligned}
E\left[R(\tau \mid \mu)/(t-1)\right] &= \left\{\sigma^2 trace\left(\mathbf{P}_{\mathrm{T}}\mathbf{R}_{\mathrm{G}}\right) + \left(\mathbf{X}_{\mathrm{T}}\tau\right)'\mathbf{P}_{\mathrm{T}}\mathbf{R}_{\mathrm{G}}\left(\mathbf{X}_{\mathrm{T}}\tau\right)\right\} \Big/ \{t-1\} \\
&= \left\{\sigma^2(t-1) + \left(\mathbf{X}_{\mathrm{T}}\tau - \overline{\tau}_.\mathbf{1}_{t^2}\right)'\left(\mathbf{X}_{\mathrm{T}}\tau - \overline{\tau}_.\mathbf{1}_{t^2}\right)\right\} \Big/ \{t-1\} \\
&= \sigma^2 + \sum_{j=1}^{t} t\left(\tau_j - \overline{\tau}_.\right)^2 \Big/ (t-1) \\
&= \sigma^2 + f_{\mathrm{T}}(\psi)
\end{aligned}$$

[8 marks]

Note that in answering this question you can use the results of any theorems given in the lecture notes, other than the theorem stating the results you are asked to prove, but you must cite the theorem when it is used.

[Total: 12 marks = 4 + 8 marks]