THE DESIGN AND MIXED-MODEL ANALYSIS OF EXPERIMENTS

PRACTICAL VI SOLUTIONS

VI.1 It is desired to run a wine-tasting experiment in which the differences between six wines are to be evaluated by scoring them on a 20 point scale. It is decided to have 6 expert judges evaluate the wines by evaluating a glass of wine on each of six consecutive occasions. It is desired to be able to isolate judge differences in scoring and differences between occasions so that a Latin square is to be employed. Use Genstat, with 224533 as the seed, to obtain a suitable randomized layout for the experiment.

The following information was supplied in response to questions from Genstat:

How many rows and columns are there in the Latin square?	6
How many treatment factors (or mutually orthogonal Latin squares) do you want to generate? (up to 3 available)	1
What would you like to call the treatment factor?	*Wines*
Give the identifier to be used for the row factor?	*Occasion*
Give the identifier to be used for the column factor?	*Judge*
Seed for randomization (0 for none)?	*224533*
What would you like to print?	design
Do you want to check the design by ANOVA	yes

The generated design is gven in the following output. For example on the first occasion, wines 1, 6, 4, 2, 5, 3 will be tasted by jusges 1–6, respectively.

```
Genstat 5  Release 4.1  (PC/Windows NT)              28 March 2000 16:01:25
Copyright 1998, Lawes Agricultural Trust (Rothamsted Experimental Station)

                 ______________________________________________

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

   3  DESIGN

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


    Judge   1   2   3   4   5   6
 Occasion
        1   1   6   4   2   5   3
        2   3   5   6   1   4   2
        3   2   4   5   3   6   1
        4   6   2   3   4   1   5
        5   4   3   1   5   2   6
        6   5   1   2   6   3   4

 Treatment factors are listed in the order: Wines
```

```
3.............................................................................

***** Analysis of variance *****
Source of variation     d.f.

Occasion stratum           5

Judge stratum              5

Occasion.Judge stratum
Wines                      5
Residual                  20

Total                     35
```

VI.2 In the lecture, a Latin square example was discussed that involved 4 areas and 4 intervals in investigating the bias of samplers in selecting wheat samples. Discuss the circumstances in which the factors Area and Interval are likely to be regarded as fixed and those in which they are likely to be regarded as random.

Areas would be fixed if the 4 areas were not representative of a larger population of areas and one expects a systematic difference between the areas. For example, if the areas were of 4 specific types (for example sandy, marshy, stoney and shaded) and these were the only 4 types of area of interest. One expects the pattern in the deviations of the area means from the grand mean to be quite irregular and their distribution is likely to uninformative. The best model would appear to be that each brand has a different mean value.

Intervals might be random if it was felt that there was unlikely to be systematic effects and that the observed intervals were representative of a larger population. For example, one might argue that practice and fatigue would not affect the response because the samplers are experienced. Also, perhaps the intervals occur randomly throughout a short period of time and so there are breaks in between. In this case one expects the deviations to vary randomly above and below the grand mean and to be described using a probability distribution with some variance.

In this example, we went for the most likely classification of Areas as random and Intervals as fixed.

VI.3 The following data are from a Latin Square experiment designed to investigate the moisture content of turnip greens. The experiment involved the measurement of the percent moisture content of five leaves of different sizes from each of five plants. The treatments were time of measurement in days since the beginning of the experiment.

		Plant				
		1	2	3	4	5
	A	5 6.67	2 5.40	3 7.32	1 4.92	4 4.88
	B	4 7.15	5 4.77	2 8.53	3 5.00	1 6.16
Leaf Size	C	1 8.29	4 5.40	5 8.50	2 7.29	3 7.83
(A = smallest,	D	3 8.95	1 7.54	4 9.99	5 7.85	2 5.83
E = largest)	E	2 9.62	3 6.93	1 9.68	4 7.08	5 8.51

Classify the factors Leaf Size and Plant as either fixed or random.

It is most likely that Leaf Size will be fixed whereas Plants will be random. There may well be systematic differences between leaves of different sizes and so this is best modelled using different means for each size. Plants on the other hand are most likely just 5 plants selected from many plants of this type and a probability distribution with some variance is likely to be an appropriate model.

The Moisture contents have been saved in the Genstat spreadsheet file *LSTurn.gsh* in the directory *G:\Disciplina\Genstat*. Add the factors Size, Plant and Time to this spreadsheet.

Analyze the data using Genstat, including diagnostic checking and the examination of mean differences. You will find that there is an outlier. Set the value for this observation missing by replacing the value with Genstat's missing value indicator, an asterisk ('*'). What efect does this have on the analysis?

```
Genstat 5  Release 4.1  (PC/Windows NT)              28 March 2000 17:46:40
Copyright 1998, Lawes Agricultural Trust (Rothamsted Experimental Station)

                 ______________________________________________

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

   3  "Data taken from File: D:/ANALYSES/LM/ONEFAC/LSTURNALL.GSH"
   4  DELETE [redefine=yes] Size,Plant,Time,Moisture
   5  FACTOR [modify=yes;nvalues=25;levels=5] Size
   6  READ Size; frepresentation=ordinal

    Identifier    Values   Missing    Levels
          Size        25         0         5

   8  FACTOR [modify=yes;nvalues=25;levels=5] Plant
   9  READ Plant; frepresentation=ordinal

    Identifier    Values   Missing    Levels
         Plant        25         0         5

  11  FACTOR [modify=yes;nvalues=25;levels=5] Time
  12  READ Time; frepresentation=ordinal

    Identifier    Values   Missing    Levels
          Time        25         0         5

  14  VARIATE [nvalues=25] Moisture
  15  READ Moisture

    Identifier   Minimum      Mean   Maximum    Values   Missing
      Moisture     4.770     7.204     9.990        25         0

  18
  19  PRINT Size,Plant,Time,Moisture

        Size       Plant        Time    Moisture
           1           1           5       6.670
           1           2           2       5.400
           1           3           3       7.320
           1           4           1       4.920
           1           5           4       4.880
           2           1           4       7.150
           2           2           5       4.770
```

```
          2           3           2       8.530
          2           4           3       5.000
          2           5           1       6.160
          3           1           1       8.290
          3           2           4       5.400
          3           3           5       8.500
          3           4           2       7.290
          3           5           3       7.830
          4           1           3       8.950
          4           2           1       7.540
          4           3           4       9.990
          4           4           5       7.850
          4           5           2       5.830
          5           1           2       9.620
          5           2           3       6.930
          5           3           1       9.680
          5           4           4       7.080
          5           5           5       8.510

  20  BLOCK Plant*Size
  21  TREAT POL(Time; 2)
  22  ANOVA [FPROB=Y; PSE=LSD] Moisture


22..............................................................................

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

Variate: Moisture

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

Plant stratum              4    28.8853     7.2213   10.71

Size stratum               4    23.7081     5.9270    8.79

Plant.Size stratum
Time                       4     0.6273     0.1568    0.23  0.915
  Lin                      1     0.1512     0.1512    0.22  0.644
  Quad                     1     0.0929     0.0929    0.14  0.717
  Deviations               2     0.3831     0.1916    0.28  0.758
Residual                  12     8.0879     0.6740

Total                     24    61.3086


* MESSAGE: the following units have large residuals.

Plant 5    Size 4           -1.77   s.e. 0.57


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

Variate: Moisture

Grand mean  7.20

     Time        1        2        3        4        5
              7.32     7.33     7.21     6.90     7.26


*** Least significant differences of means (5% level) ***

Table                 Time
rep.                     5
d.f.                    12
l.s.d.               1.131

  23  CALC FP=7.2213/0.6740 & pP=1-FPROB(FP; 4; 12) : PRINT FP,pP

          FP           pP
       10.71   0.0006232
```

```
  24  CALC FS=5.9270/0.6740 & pS=1-FPROB(FS; 4; 12) : PRINT FS,pS

          FS          pS
       8.794    0.001483

  25  AGRAPH [GRAPH=line] XFACTOR=Time; BAR=*

                                         Means for Time

             -
             I
         7.5 I
             I      -
             I      -
             I      -            *
             I      -   *
             I      -                                           *
         7.2 I      -                   *
             I      -
             I      -
             I      -
             I      -
             I      -
         6.9 I                                        *
             -+---------+---------+---------+---------+---------+---------+---
            0.0       1.0       2.0       3.0       4.0       5.0       6.0

  26  APLOT METHOD=fit,normal

             -
             I
             I
             I
         2.0 I
r            I
e            I
s            I                             *
i            I        *                     *  *     *             *
d            I                   *        * *          *            *
u        0.0 I          *                 *     *               *
a            I            2                   **           *               *
l            I              *    *                              *
s            I
             I
             I
             I                                    *
        -2.0 I
             -+---------+---------+---------+---------+---------+---------+---
            4.0       5.0       6.0       7.0       8.0       9.0      10.0

                                         fitted values

                                          Normal plot
             -
             I
         2.0 I
r            I
e            I
s            I                                                       *
i            I                                        ** *  *  *
d            I                                * *** *
u        0.0 I                            ****
a            I                   ** * ***
l            I           *  *  *
s            I
             I
             I     *
        -2.0 I
             -+---------+---------+---------+---------+---------+---------+---
           -2.4      -1.6      -0.8       0.0       0.8       1.6       2.4

                                   expected Normal quantiles
```

```
 27  "
-28  **** Tukey''s one-degree-of-freedom-for-non-additivity.
-29  **** It is the term designated covariate in the following analysis
-30  "
 31  AKEEP [FIT=Fit]
 32  CALC ResSq=Fit*Fit
 33  ANOVA [PRINT=*]  ResSq; RES=ResSq
 34  COVAR ResSq                          "A computational trick"
 35  ANOVA [PRINT=A; FPROB=Y] Moisture
```

```
35.................................................................................

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

Variate: Moisture
Covariate: ResSq
```

Source of variation	d.f.	s.s.	m.s.	v.r.	cov.ef.	F pr.
Plant stratum	4	28.8853	7.2213	9.93		
Size stratum	4	23.7081	5.9270	8.15		
Plant.Size stratum						
Time	4	0.6273	0.1568	0.22	1.00	0.924
Lin	1	0.1512	0.1512	0.21	1.00	0.657
Quad	1	0.0929	0.0929	0.13	1.00	0.728
Deviations	2	0.3831	0.1916	0.26	1.00	0.773
Covariate	1	0.0853	0.0853	0.12		0.738
Residual	11	8.0026	0.7275		0.93	
Total	24	61.3086				

Step 1: Set up hypotheses

a) H_0: $\tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5$
H_1: at least one pair of population time means is different

b) H_0: $\sigma_P^2 = 0$
H_1: $\sigma_P^2 > 0$

c) H_0: $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5$
H_1: at least one pair of population size means is different

Step 2: Calculate test statistics

The analysis of variance table for a Latin Square is:

Source	df	SSq	MSq	E[MSq]	F	Prob
Plant	4	28.8853	7.2213	$\sigma^2_{\mathrm{PS}} + 5\sigma^2_{\mathrm{P}}$	10.71	<0.001
Size	4	23.7081	5.9270	$\sigma^2_{\mathrm{PS}} + f_{\mathrm{S}}(\psi)$	8.79	0.002
Plant.Size	16					
Times	4	0.6273	0.1568	$\sigma^2_{\mathrm{PS}} + f_T(\psi)$	0.23	0.915
Linear	1	0.1512	0.1512		0.22	0.644
Quadratic	1	0.0929	0.0929		0.14	0.717
Deviations	2	0.3831	0.1916		0.28	0.758
Residual	12	8.0879	0.6740	σ^2_{PS}		
Non-additivity	1	0.0853	0.0853		0.12	0.738
Deviations	11	8.0026	0.7275			
Total	24	61.3806				

Step 3: Decide between hypotheses

There analysis indicates that there are no significant differences between the treatments, in spite of a very effective Latin Square (both Plant and Size are significant). Further, there is no evidence of transformable non-additivity. However, the plot of residuals-versus-fitted-values indicates that there is an observation for which has an extremely low residual. The Normal Probability plot is also exhibiting the same problem. The warning message unde rthe analysis of variance table indicates that this observation is Plant 5 and Size 4 or the 20th observation. This observation is set to missing and the analysis repeated to see what effect the outlier has on the analysis.

```
"Analysis with Plant 5 Size 4 missing"
CALC Moisture$[20]=!(*)
COVAR
ANOVA [FPROB=Y; PSE=LSD] Moisture
```

```
39.............................................................................

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

Variate: Moisture

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

Plant stratum              4         26.9192     6.7298   47.60

Size stratum               4         31.9976     7.9994   56.58

Plant.Size stratum
Time                       4          3.7668     0.9417    6.66  0.006
  Lin                      1          0.8292     0.8292    5.87  0.034
  Quad                     1          0.0185     0.0185    0.13  0.724
  Deviations               2          2.9190     1.4595   10.32  0.003
Residual                 11(1)        1.5551     0.1414

Total                    23(1)       59.3432


* MESSAGE: the following units have large residuals.

Plant 5    Size 1          -0.535   s.e. 0.249


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

Variate: Moisture

Grand mean  7.351

     Time        1        2        3        4        5
             7.318    8.072    7.206    6.900    7.260


*** Least significant differences of means (5% level) ***

Table                 Time
rep.                     5
d.f.                    11
l.s.d.              0.5234

(Not adjusted for missing values)


***** Missing values *****

Variate: Moisture

 Unit  estimate
   20     9.519

Max. no. iterations 4

  40  CALC FP=6.7298/0.1414 & pP=1-FPROB(FP; 4; 11) : PRINT FP,pP

          FP          pP
       47.59 0.000000705

  41  CALC FS=7.9994/0.1414 & pS=1-FPROB(FS; 4; 11) : PRINT FS,pS

          FS          pS
       56.57 0.000000288

  42  AGRAPH [GRAPH=line] XFACTOR=Time; BAR=*
```

```
                                        Means for Time

             -
             I                    *
         8.0 I
             I
             I
             I
             I      -
             I      -   *
         7.2 I      -                       *                   *
             I
             I                                        *
             I
             I
             I
         6.4 I
             -+---------+---------+---------+---------+---------+---------+---
            0.0       1.0       2.0       3.0       4.0       5.0       6.0
```

43 APLOT METHOD=fit,normal

```
                                                                              *
             -
             I
             I
             I
         0.6 I
r            I                           *     *
e            I
s            I
i            I            *                  *           *              *
d            I                         *   **   *   *   *
u        0.0 I         *                       *           *
a            I         *                                                *
l            I                       *                                   *
s            I             *   *                              *
             I
             I              *                                       *
        -0.6 I
             -+---------+---------+---------+---------+---------+---------+---
            4.0       5.0       6.0       7.0       8.0       9.0      10.0

                                        fitted values


                                         Normal plot
                                                                              *
             -
             I
         0.6 I
r            I                                                 *    *
e            I
s            I
i            I                                       * * * *
d            I                               *** ** *
u        0.0 I                           ***
a            I                        **
l            I                     **
s            I               * * *
             I
             I      *    *
        -0.6 I
             -+---------+---------+---------+---------+---------+---------+---
          -2.4      -1.6      -0.8       0.0       0.8       1.6       2.4

                                  expected Normal quantiles
```

```
  44  "
 -45  **** Tukey''s one-degree-of-freedom-for-non-additivity.
 -46  **** It is the term designated covariate in the following analysis
 -47  "
  48  AKEEP [FIT=Fit]
  49  CALC ResSq=Fit*Fit
  50  ANOVA [PRINT=*]  ResSq; RES=ResSq
  51  COVAR ResSq                            "A computational trick"
  52  ANOVA [PRINT=A; FPROB=Y] Moisture


52.................................................................................

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

Variate: Moisture
Covariate: ResSq

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

Plant stratum
Covariate                  1          0.0051     0.0051    0.00           0.982
Residual                   3         26.9141     8.9714   76.89    0.75

Size stratum
Covariate                  1         12.5785    12.5785    1.94           0.258
Residual                   3         19.4191     6.4730   55.48    1.24

Plant.Size stratum
Time                       4          3.5191     0.8798    7.54    1.00  0.005
  Lin                      1          0.7804     0.7804    6.69    1.00  0.027
  Quad                     1          0.0129     0.0129    0.11    1.00  0.746
  Deviations               2          2.7258     1.3629   11.68    1.00  0.002
Covariate                  1          0.3883     0.3883    3.33          0.098
Residual                  10(1)       1.1668     0.1167            1.21

Total                     23(1)      59.3432
```

	ORIGINAL ANALYSIS				MISSING VALUE ANALYSIS			
Source	df	MSq	F	Prob	df	MSq	F	Prob
Plant	4	7.2213	10.71	<0.001	4	6.7298	47.60	<0.001
Size	4	5.9270	8.79	0.002	4	7.9994	56.58	<0.001
Plant.Size	16				15			
Times	4	0.1568	0.23	0.915	4	0.9417	6.66	0.006
Linear	1	0.1512	0.22	0.644	1	0.8292	5.87	0.034
Quadratic	1	0.0929	0.14	0.717	1	0.0185	0.13	0.724
Deviations	2	0.1916	0.28	0.758	2	1.4595	10.32	0.003
Residual	12	0.6740			11†	0.1414		
Non-additivity	1	0.0853	0.12	0.738	1	0.3883	3.33	0.098
Deviations	11	0.7275			10	0.1167		
Total	24				23			

†The Residual degrees of freedom have been reduced by one to take account of the missing value.

The analysis now indicates that there are differences between the Times and that the Times means exhibit a linear trend from which there are significant deviations. Consequently we can only examine treatment differences using a multiple comparisons procedure.

The Times means (adjusted for the missing value) and accompanying plot are:

Time					
1	2	3	4	5	s.e.d
7.318	8.072	7.206	6.900	7.260	0.2378

```
25  AGRAPH [GRAPH=line] XFACTOR=Time; BAR=*

                                Means for Time

    -
    I
7.5 I
    I      -
    I      -
    I      -             *
    I      -   *
    I      -                                           *
7.2 I      -                       *
    I      -
    I      -
    I      -
    I      -
    I      -
6.9 I                                        *
    -+---------+---------+---------+---------+---------+---------+---
   0.0       1.0       2.0       3.0       4.0       5.0       6.0
```

It is clear that the trend is not described by either a straight line or a quadratic curve.

VI.4 An experimenter wants to investigate the effects of four different rations on the apparent consumption of total carbohydrates (as a percentage) by calves. He has available four calves of around 280 kg. He plans to use a Latin square for two reasons. Firstly, so that each calf receives the four rations, one in each of four periods. Secondly, so that differences, such as climatic differences, between the periods are eliminated from treatment differences. The experimenter is willing to run a 5% chance of making a type I error and would like to have a 95% chance of detecting any difference of 7.5% or more in the apparent consumption between rations. A variance of 10% for the animal-period combinations is expected in the experiment. Will the Latin square have the desired power?

ANOVAPower.xls is used to compute the power and the values in the cells below the headings are as follows:

sample size (r)	**alpha**	**DF**		**central F**	**no. values in a mean (m)**	**delta**	**standard deviation**	**lambda**	**power**
		numer-ator	**denomin-ator**						
4	0.05	3	6	4.7571	4	7.5	3.162278	11.25	0.5155

The power is well below the desired 95%.