THE DESIGN AND MIXED-MODEL ANALYSIS OF EXPERIMENTS

PRACTICAL III
SOLUTIONS

III.1 Prove lemma III.2

We have that $\mathbf{P} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ *so that* $\mathbf{PX} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X} = \mathbf{X}$.

Also, $\mathbf{RX} = \left(\mathbf{I} - \mathbf{P}\right)\mathbf{X} = \mathbf{X} - \mathbf{PX} = \mathbf{0}$.

III.2 Show that

$$\mathbf{P}_{\mathrm{T}} = \mathbf{X}_{\mathrm{T}}\left(\mathbf{X}_{\mathrm{T}}'\mathbf{X}_{\mathrm{T}}\right)^{-1}\mathbf{X}_{\mathrm{T}}'$$

$$= \begin{bmatrix} \mathbf{1}_{n_1}\left(\mathbf{1}_{n_1}'\mathbf{1}_{n_1}\right)^{-1}\mathbf{1}_{n_1}' & \mathbf{0}_{n_1\times n_2} & \cdots & \mathbf{0}_{n_1\times n_t} \\ \mathbf{0}_{n_2\times n_1} & \mathbf{1}_{n_2}\left(\mathbf{1}_{n_2}'\mathbf{1}_{n_2}\right)^{-1}\mathbf{1}_{n_2}' & \cdots & \mathbf{0}_{n_2\times n_t} \\ \vdots & \vdots & & \vdots \\ \mathbf{0}_{n_t\times n_1} & \mathbf{0}_{n_t\times n_2} & \cdots & \mathbf{1}_{n_t}\left(\mathbf{1}_{n_t}'\mathbf{1}_{n_t}\right)^{-1}\mathbf{1}_{n_t}' \end{bmatrix}$$

where

$$\mathbf{X}_{\mathrm{T}} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1\times 1} & \cdots & \mathbf{0}_{n_1\times 1} \\ \mathbf{0}_{n_2\times 1} & \mathbf{1}_{n_2} & \cdots & \mathbf{0}_{n_2\times 1} \\ \vdots & \vdots & & \vdots \\ \mathbf{0}_{n_t\times 1} & \mathbf{0}_{n_t\times 1} & \cdots & \mathbf{1}_{n_t} \end{bmatrix}$$

First

$$\mathbf{X}_{\mathrm{T}}'\mathbf{X}_{\mathrm{T}} = \begin{bmatrix} \mathbf{1}_{n_1}' & \mathbf{0}_{1\times n_2} & \cdots & \mathbf{0}_{1\times n_t} \\ \mathbf{0}_{1\times n_1} & \mathbf{1}_{n_2}' & \cdots & \mathbf{0}_{1\times n_t} \\ \vdots & \vdots & & \vdots \\ \mathbf{0}_{1\times n_1} & \mathbf{0}_{1\times n_2} & \cdots & \mathbf{1}_{n_t}' \end{bmatrix}\begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1\times 1} & \cdots & \mathbf{0}_{n_1\times 1} \\ \mathbf{0}_{n_2\times 1} & \mathbf{1}_{n_2} & \cdots & \mathbf{0}_{n_2\times 1} \\ \vdots & \vdots & & \vdots \\ \mathbf{0}_{n_t\times 1} & \mathbf{0}_{n_t\times 1} & \cdots & \mathbf{1}_{n_t} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{n_1}'\mathbf{1}_{n_1} & 0 & \cdots & 0 \\ 0 & \mathbf{1}_{n_2}'\mathbf{1}_{n_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \mathbf{1}_{n_t}'\mathbf{1}_{n_t} \end{bmatrix}$$

Since $\mathbf{1}_{n_k}'\mathbf{1}_{n_k}$ *is a scalar* $\left(\mathbf{1}_{n_k}'\mathbf{1}_{n_k}\right)^{-1}$ *exists and*

$$\mathbf{P}_T = \mathbf{X}_T\left(\mathbf{X}_T'\mathbf{X}_T\right)^{-1}\mathbf{X}_T'$$

$$= \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1\times 1} & \cdots & \mathbf{0}_{n_1\times 1} \\ \mathbf{0}_{n_2\times 1} & \mathbf{1}_{n_2} & \cdots & \mathbf{0}_{n_2\times 1} \\ \vdots & \vdots & & \vdots \\ \mathbf{0}_{n_t\times 1} & \mathbf{0}_{n_t\times 1} & \cdots & \mathbf{1}_{n_t} \end{bmatrix} \begin{bmatrix} \left(\mathbf{1}_{n_1}'\mathbf{1}_{n_1}\right)^{-1} & 0 & \cdots & 0 \\ 0 & \left(\mathbf{1}_{n_2}'\mathbf{1}_{n_2}\right)^{-1} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \left(\mathbf{1}_{n_t}'\mathbf{1}_{n_t}\right)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{1}_{n_1}' & \mathbf{0}_{1\times n_2} & \cdots & \mathbf{0}_{1\times n_t} \\ \mathbf{0}_{1\times n_1} & \mathbf{1}_{n_2}' & \cdots & \mathbf{0}_{1\times n_t} \\ \vdots & \vdots & & \vdots \\ \mathbf{0}_{1\times n_1} & \mathbf{0}_{1\times n_2} & \cdots & \mathbf{1}_{n_t}' \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{1}_{n_1}\left(\mathbf{1}_{n_1}'\mathbf{1}_{n_1}\right)^{-1} & \mathbf{0}_{n_1\times 1} & \cdots & \mathbf{0}_{n_1\times 1} \\ \mathbf{0}_{n_2\times 1} & \mathbf{1}_{n_2}\left(\mathbf{1}_{n_2}'\mathbf{1}_{n_2}\right)^{-1} & \cdots & \mathbf{0}_{n_2\times 1} \\ \vdots & \vdots & & \vdots \\ \mathbf{0}_{n_t\times 1} & \mathbf{0}_{n_t\times 1} & \cdots & \mathbf{1}_{n_t}\left(\mathbf{1}_{n_t}'\mathbf{1}_{n_t}\right)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{1}_{n_1}' & \mathbf{0}_{1\times n_2} & \cdots & \mathbf{0}_{1\times n_t} \\ \mathbf{0}_{1\times n_1} & \mathbf{1}_{n_2}' & \cdots & \mathbf{0}_{1\times n_t} \\ \vdots & \vdots & & \vdots \\ \mathbf{0}_{1\times n_1} & \mathbf{0}_{1\times n_2} & \cdots & \mathbf{1}_{n_t}' \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{1}_{n_1}\left(\mathbf{1}_{n_1}'\mathbf{1}_{n_1}\right)^{-1}\mathbf{1}_{n_1}' & \mathbf{0}_{n_1\times n_2} & \cdots & \mathbf{0}_{n_1\times n_t} \\ \mathbf{0}_{n_2\times n_1} & \mathbf{1}_{n_2}\left(\mathbf{1}_{n_2}'\mathbf{1}_{n_2}\right)^{-1}\mathbf{1}_{n_2}' & \cdots & \mathbf{0}_{n_2\times n_t} \\ \vdots & \vdots & & \vdots \\ \mathbf{0}_{n_t\times n_1} & \mathbf{0}_{n_t\times n_2} & \cdots & \mathbf{1}_{n_t}\left(\mathbf{1}_{n_t}'\mathbf{1}_{n_t}\right)^{-1}\mathbf{1}_{n_t}' \end{bmatrix}$$

III.3 Show that $\mathbf{P}_T\mathbf{P}_G\mathbf{X}_T\boldsymbol{\alpha} = \mathbf{1}\bar{\alpha}_.$ where $\bar{\alpha}_. = \left(\sum_{k=1}^{t} n_k\alpha_k\right)\Big/ n$ and $\mathbf{P}_T\mathbf{X}_T\boldsymbol{\alpha} = \mathbf{X}_T\boldsymbol{\alpha}$

First since $\boldsymbol{P}_T\boldsymbol{P}_G = \boldsymbol{P}_G$, $\mathbf{P}_T\mathbf{P}_G\mathbf{X}_T\boldsymbol{\alpha} = \mathbf{P}_G\mathbf{X}_T\boldsymbol{\alpha}$ *and*

$$\mathbf{P}_G\mathbf{X}_T\boldsymbol{\alpha} = \mathbf{P}_G\begin{bmatrix} \alpha_1\mathbf{1}_{n_1} \\ \alpha_2\mathbf{1}_{n_2} \\ \vdots \\ \alpha_t\mathbf{1}_{n_t} \end{bmatrix} = \frac{1}{n}\left(n_1\alpha_1 + n_2\alpha_2 + \ldots n_t\alpha_t\right)\mathbf{1}_n = \frac{\sum_{i=1}^{t} n_i\alpha_i}{n}\mathbf{1}_n = \bar{\alpha}_.\mathbf{1}_n .$$

III.4 Verify that $\mathbf{e}_G - \mathbf{t}_e = \mathbf{R}_T\mathbf{R}_G\mathbf{y}$.

Now $\mathbf{e}_G = \mathbf{R}_G\mathbf{y}$ *and* $\mathbf{t}_e = \mathbf{P}_T\mathbf{R}_G\mathbf{y}$

so that $\mathbf{e}_G - \mathbf{t}_e = \mathbf{R}_G\mathbf{y} - \mathbf{P}_T\mathbf{R}_G\mathbf{y} = (\mathbf{I} - \mathbf{P}_T)\mathbf{R}_G\mathbf{y} = \mathbf{R}_T\mathbf{R}_G\mathbf{y}$.

III.5 Let $E[\mathbf{Y}] = \mathbf{X}_G\mu$, $\mathbf{V}_Y = \sigma^2\mathbf{I}_n$, $R(\alpha|\,\mu) = \mathbf{Y}'\mathbf{P}_T\mathbf{R}_G\mathbf{Y}$ and D(α) = $\mathbf{Y}'\mathbf{R}_T\mathbf{Y}$ where $\mathbf{R}_G$ and $\mathbf{R}_T$ are as defined in lemma III.1. Then, show that

$$E\left[R(\alpha|\,\mu)/(t-1)\right] = \sigma^2 \quad \text{and} \quad E\left[D(\alpha)/(n-t)\right] = \sigma^2$$

where *t* is the number of treatments and n is the number of observations.

For $E\left[R(\alpha|\,\mu)/(t-1)\right]$*, we first use theorem II.11 to show that*

$$\begin{aligned} E\left[R(\alpha|\,\mu)/(t-1)\right] &= E[\mathbf{Y}'\mathbf{P}_T\mathbf{R}_G\mathbf{Y}]/(t-1) \\ &= \left\{trace\left(\mathbf{P}_T\mathbf{R}_G\sigma^2\mathbf{I}_n\right) + (\,\mathbf{X}_G\mu)'\,\mathbf{P}_T\mathbf{R}_G(\mathbf{X}_G\mu)\right\}/\{t-1\} \\ &= \left\{\sigma^2 trace(\mathbf{P}_T\mathbf{R}_G) + (\,\mathbf{X}_G\mu)'\,\mathbf{P}_T\mathbf{R}_G(\mathbf{X}_G\mu)\right\}/\{t-1\} \end{aligned}$$

Now from theorem III.5, $trace(\mathbf{P}_T\mathbf{R}_G) = t-1$.

Also, by lemma III.2, $\mathbf{R}_G\mathbf{X}_G = \mathbf{0}$ *and so* $\mathbf{P}_T\mathbf{R}_G(\mathbf{X}_G\mu) = \mathbf{0}$.

Hence,

$$\begin{aligned} E\left[R(\alpha|\,\mu)/(t-1)\right] &= \left\{\sigma^2 trace(\mathbf{P}_T\mathbf{R}_G) + (\,\mathbf{X}_G\mu)'\,\mathbf{P}_T\mathbf{R}_G(\mathbf{X}_G\mu)\right\}/\{t-1\} \\ &= \left\{\sigma^2(t-1) + 0\right\}/\{t-1\} \\ &= \sigma^2 \end{aligned}$$

For $E\left[D(\alpha)/(n-t)\right]$,

$$E\left[D(\alpha)/(n-t)\right] = E[\mathbf{y}'\mathbf{R}_T\mathbf{y}]/(n-t) = \left\{\sigma^2 trace(\mathbf{R}_T) + (\mathbf{X}_G\mu)'\,\mathbf{R}_T(\mathbf{X}_G\mu)\right\}/\{n-t\}$$

Now from theorem III.5, $trace(\mathbf{R}_T) = n-t$.

From lemma III.1 $\mathbf{R}_T = \mathbf{R}_T\mathbf{R}_G$ *and from lemma III.2* $\mathbf{R}_G\mathbf{X}_G = \mathbf{0}$ *so that* $\mathbf{R}_T(\mathbf{X}_G\mu) = \mathbf{R}_T\mathbf{R}_G\mathbf{X}_G\mu = \mathbf{0}$. *Consequently,*

$$\begin{aligned} E\left[D(\alpha)/(n-t)\right] &= \left\{\sigma^2 trace(\mathbf{R}_T) + (\,\mathbf{X}_G\mu)'\,\mathbf{R}_T(\mathbf{X}_G\mu)\right\}/\{n-t\} \\ &= \left\{\sigma^2(n-t) + 0\right\}/\{n-t\} \\ &= \sigma^2 \end{aligned}$$

III.6 An investigation was conducted to examine differences between 3 brands of tyre in their braking distances (ft.) from a speed of 30 miles per hour. Altogether 9 tests were conducted with the particular brand tested at each test being chosen at random. The results are as follows:

Brand		
A	B	C
28	27	27
26	25	32
30	26	31

Perform the one-way analysis manually using mean operators (**P**).

Test	Brand	Stopping Distance **y**	Grand mean **g**	Total Test Deviations $\mathbf{e}_G$	Brand effects $\mathbf{t}_e$	Residual Test Deviations $\mathbf{e}_T$
1	1	28	28	0	0	0
2	1	26	28	-2	0	-2
3	1	30	28	2	0	2
4	2	27	28	-1	-2	1
5	2	25	28	-3	-2	-1
6	2	26	28	-2	-2	0
7	3	27	28	-1	2	-3
8	3	32	28	4	2	2
9	3	31	28	3	2	1
Sums of squares of deviations				48	24	24
Degrees of freedom				8	2	6
Variance				6	12	4

Step 1: Set up hypotheses

H_0: $\alpha_1 = \alpha_2 = \alpha_3 = \mu$

H_1: at least one pair of population brand means is different

Step 2: Calculate test statistic

The analysis of variance table for the example is:

Source	df	SSq	MSq	F
Tests	8	48		
Brands	2	24	12	3.0
Residual	6	24	4	

Step 3: Decide between hypotheses

The probability of exceeding an F of 3.0 with $\nu_1 = 2$ and $\nu_2 = 6$ is $P(F \geq 3.0) = 0.1250$. Little evidence of a difference between the brands.

III.7 Use Genstat to produce a randomized layout for an experiment involving 21 plots to which 7 treatments are to be allocated so that each is replicated 3 times. Use the seed 413634 in producing the layout.

The layout is given at the end of the following Genstat output.

```
Genstat 5  Release 4.1  (PC/Windows NT)              22 March 2000 16:27:47
Copyright 1998, Lawes Agricultural Trust (Rothamsted Experimental Station)

                ____________________________________________

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

  3  DELETE [redefine=yes] Plots,Treat
  4  FACTOR [modify=yes;nvalues=21;levels=21] Plots
  5  READ Plots; frepresentation=ordinal

   Identifier    Values   Missing    Levels
        Plots        21         0        21

  7  FACTOR [modify=yes;nvalues=21;levels=7] Treat
  8  READ Treat; frepresentation=ordinal

   Identifier    Values   Missing    Levels
        Treat        21         0         7

 10
 11  PRINT Plots,Treat

      Plots       Treat
          1           1
          2           1
          3           1
          4           2
          5           2
          6           2
          7           3
          8           3
          9           3
         10           4
         11           4
         12           4
         13           5
         14           5
         15           5
         16           6
         17           6
         18           6
         19           7
         20           7
         21           7

 12  RANDOMIZE [BLOCKSTRUCTURE=Plots; SEED=413634] Treat
 13  PDESIGN [BLOCK=Plots; TREAT=Treat]
```

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


Plots
    1   6
    2   1
    3   3
    4   6
    5   7
    6   1
    7   3
    8   7
    9   2
   10   7
   11   4
   12   5
   13   1
   14   2
   15   3
   16   5
   17   6
   18   4
   19   2
   20   4
   21   5

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

III.8 in an experiment to compare melon varieties, six plots of each of four varieties were grown. The varieties were allocated to the 24 plots in the experiment in a completely random manner. The melon yields for each plot are as follows:

Variety			
A	B	C	D
25.12	40.25	18.30	28.05
17.25	35.25	22.60	28.55
26.42	31.98	25.90	33.20
16.08	36.52	15.05	31.68
22.15	43.32	11.42	30.32
15.92	37.10	23.68	27.58

This data is available in *CRDMelon.gsh* in the directory *G:\Disciplina\Genstat*. Add the 24-level factor Plots and the factor Variety with levels A–D to this Genstat spreadsheet and make sure that the factors and variate are available in the central store.

Conduct a regression analysis to determine if there is any measurable effect of variety on the yield.

What are the treatment means?

Use the Genstat BLOCKSTRUCTURE, TREATMENTSTRUCTURE and ANOVA commands to perform an analysis of variance on this data and confirm your answers from the regression analysis.

Use Genstat to obtain a residuals-versus-fitted-values plot and a normal probability plot as a check on the assumptions underlying the analysis.

Also, carry out the LSD and Tukey's procedures to determine exactly which Varieties differed.

```
38  PRINT Plots,Variety,Yield

       Plots     Variety       Yield
           1           A       25.12
           2           A       17.25
           3           A       26.42
           4           A       16.08
           5           A       22.15
           6           A       15.92
           7           B       40.25
           8           B       35.25
           9           B       31.98
          10           B       36.52
          11           B       43.32
          12           B       37.10
          13           C       18.30
          14           C       22.60
          15           C       25.90
          16           C       15.05
          17           C       11.42
          18           C       23.68
          19           D       28.05
          20           D       28.55
          21           D       33.20
          22           D       31.68
          23           D       30.32
          24           D       27.58

  39  MODEL Yield
  40  TERMS Variety
  41  FIT [FPROB=y] Variety


41.....................................................................


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

 Response variate: Yield
     Fitted terms: Constant, Variety


*** Summary of analysis ***

              d.f.         s.s.         m.s.      v.r.  F pr.
Regression       3       1291.5       430.49     23.42  <.001
Residual        20        367.7        18.38
Total           23       1659.1        72.14

Percentage variance accounted for 74.5
Standard error of observations is estimated to be 4.29
* MESSAGE: The following units have large standardized residuals:
         Unit     Response    Residual
           17        11.42       -2.06
* MESSAGE: The error variance does not appear to be constant:
           intermediate responses are less variable than small or large
responses


*** Estimates of parameters ***

                  estimate         s.e.     t(20)
Constant             20.49         1.75     11.71
Variety B            16.91         2.48      6.83
Variety C            -1.00         2.48     -0.40
Variety D             9.41         2.48      3.80
```

Step 1: Set up hypotheses

H_0: $\alpha_A = \alpha_B = \alpha_C = \alpha_D = \mu$
H_1: at least one pair of population Variety means is different

Step 2: Calculate test statistic

The analysis of variance table for the example is:

Source	df	SSq	MSq	F	p
Plots	23	1291.5			
Treatments	3	367.7	430.49	23.49	<0.001
Residual	20	1659.1	72.14		

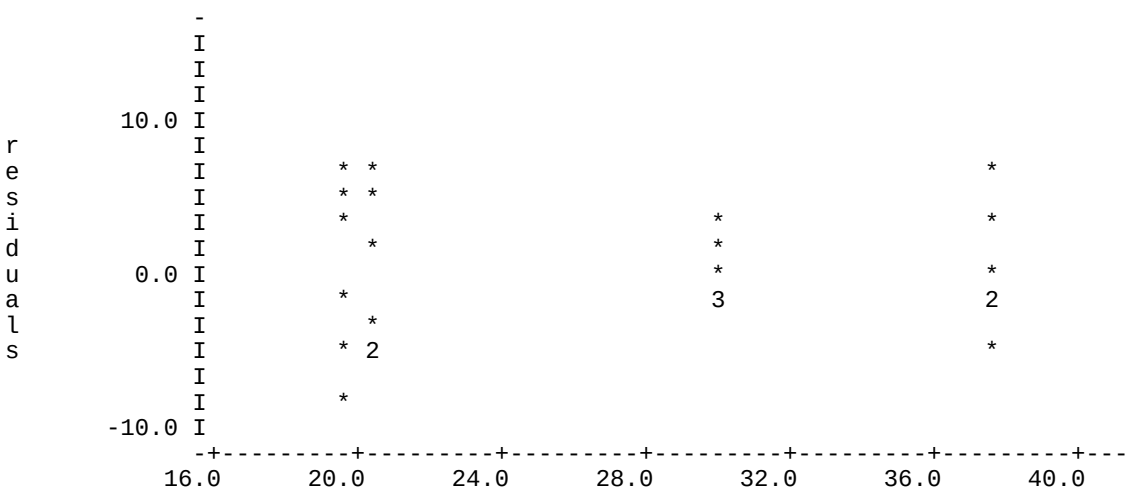
Step 3: Decide between hypotheses

The probability of exceeding an F of 23.49 with $\nu_1 = 3$ and $\nu_2 = 20$ is $P(F \geq 23.42) < 0.001$. The evidence suggests that there is a Variety difference.

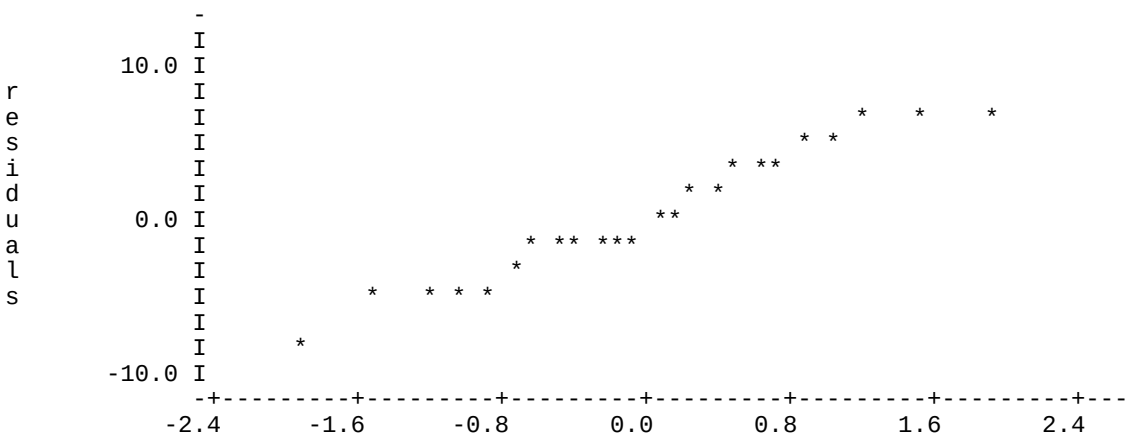
The treatment means are 20.49, 20.49+16.91 = 37.40, 20.49–1.00 = 19.49 and 20.49+9.41 = 29.90.

```
  42  BLOCK Plots
  43  TREAT Variety
  44  ANOVA [FPROB=Y; PSE=LSD] Yield

44.....................................................................

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

Variate: Yield

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

Plots stratum
Variety                    3    1291.48     430.49   23.42  <.001
Residual                  20     367.65      18.38
Total                     23    1659.13


* MESSAGE: the following units have large residuals.

Plots 17          -8.07   s.e. 3.91


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

Variate: Yield

Grand mean  26.82

  Variety        A        B        C        D
            20.49    37.40    19.49    29.90
```

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

Table              Variety
rep.                     6
d.f.                    20
l.s.d.                5.164


  45  APLOT METHOD=fit,normal
```



Normal plot



The analysis of variance and means from the mean operator analysis are as obtained from the regression analysis.

It appears that the homogeneity of variance assumption may not be met as the size of the band for the second highest-yielding variety has less variance than the other varieties. It would be useful to use a hypothesis test to confirm this conclusion but we will not cover these in this subject. From the normal probability plot it appears that the normality assumption is met except for some outliers (an extremely small

and an extremely large residual). As we cannot check the observations in question we will do nothing further about them.

Finally, we examine the difference between the pairs of means using the LSD and Tukey's procedure. Strictly, speaking this is not appropriate as the assumptions have not been met. We do so as an exercise realizing that we cannot rely on the conclusions made.

Differences between all pairs of Variety means

Variety		C	A	D	B
	Mean	19.49	20.49	29.90	37.40
C	19.49				
A	20.49	1.00			
D	29.90	10.41	9.41		
B	37.40	17.91	16.91	7.50	
	LSD(5%)		5.164		

Variety

C	A	D	B
19.49	20.49	29.90	37.40

Varieties C and A yielded lower than D which was in turn lower than B.

Differences between all pairs of Variety means

Variety		C	A	D	B
	Mean	19.49	20.49	29.90	37.40
C	19.49				
A	20.49	1.00			
D	29.90	10.41	9.41		
B	37.40	17.91	16.91	7.50	
	w(5%)		6.93		

$$w(5\%) = \frac{3.960}{\sqrt{2}}\sqrt{18.38\frac{2}{6}}$$
$$= 2.6879 \times 1.392$$
$$= 6.93$$

Variety

C	A	D	B
19.49	20.49	29.90	37.40

Varieties C and A yielded lower than D which was in turn lower than B.

III.9 In an experiment to investigate the effect of different rates of injection on the lethal dose of ouabain, animals were injected at one of four different rates with ouabain and the lethal dose recorded. The data given in the following table:

	1	5	9	11	13	14	16	17	20	22	28	31	31
Rate of	2	3	6	22	27	27	28	28	37	40	42	50	
Injection	4	34	34	38	40	46	58	60	60	65			
(mg/kg/min)	8	51	56	62	63	70	73	76	89	92			
/1045.75													

Input the data into Genstat. Note that the replication of the rates is not equal. To set up a factor for Rates you will either have to manually enter the values of the levels or use *Stats > Design > Generate Factors in Standard Order* or *Spread > Column > Fill* to generate an equal number for each rate, the number generated being between the minimum and maximum replication. Then extras can be deleted and/or deficiencies rectified by adding extra cells.

Use Genstat to perform a one-way analysis of variance on this data in which a quadratic is fitted to determine if there is a relationship between the lethal dose and the rate of injection. Perform the analysis using the averaging operator method. Produce residuals-versus-fitted-values and normal probability plots to check the assumptions underlying the analysis. Obtain the equation of the fitted line and a plot of the Rate means.

```
Genstat 5  Release 4.1  (PC/Windows NT)              19 March 2000 22:52:34
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/CRDCATS.GSH"
  4  DELETE [redefine=yes] Animals,Rate,Dose
  5  FACTOR [modify=yes;nvalues=41;levels=41] Animals
  6  READ Animals; frepresentation=ordinal

   Identifier    Values   Missing    Levels
      Animals        41         0        41

  9  FACTOR [modify=yes;nvalues=41;levels=!(1,2,4,8)] Rate
 10  READ Rate; frepresentation=ordinal

   Identifier    Values   Missing    Levels
         Rate        41         0         4

 13  VARIATE [nvalues=41] Dose
 14  READ Dose

   Identifier   Minimum      Mean   Maximum    Values   Missing
         Dose      3.00     38.88     92.00        41         0

 17
 18  PRINT Animals,Rate,Dose

    Animals        Rate        Dose
          1       1.000        5.00
          2       1.000        9.00
          3       1.000       11.00
          4       1.000       13.00
```

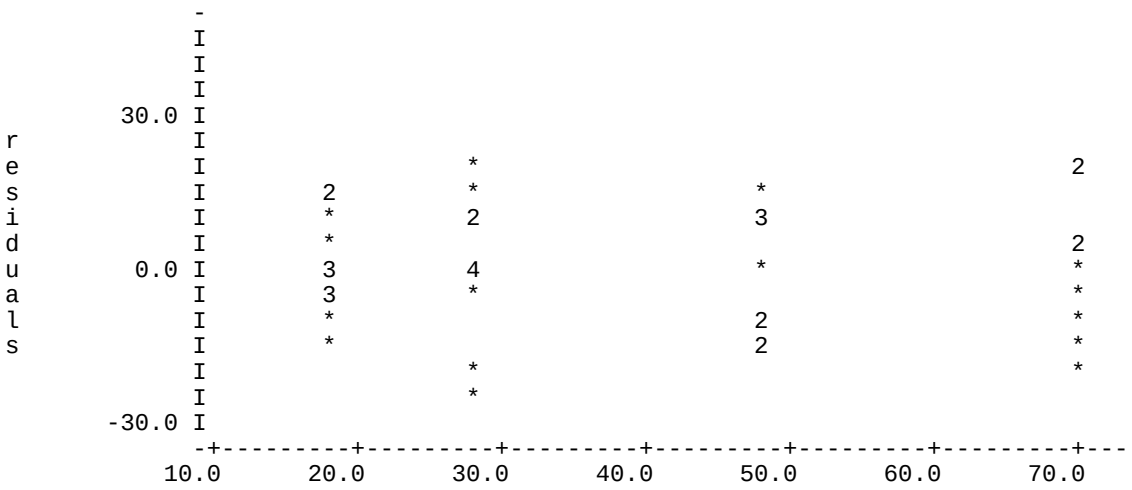
```
           5       1.000       14.00
           6       1.000       16.00
           7       1.000       17.00
           8       1.000       20.00
           9       1.000       22.00
          10       1.000       28.00
          11       1.000       31.00
          12       1.000       31.00
          13       2.000        3.00
          14       2.000        6.00
          15       2.000       22.00
          16       2.000       27.00
          17       2.000       27.00
          18       2.000       28.00
          19       2.000       28.00
          20       2.000       37.00
          21       2.000       40.00
          22       2.000       42.00
          23       2.000       50.00
          24       4.000       34.00
          25       4.000       34.00
          26       4.000       38.00
          27       4.000       40.00
          28       4.000       46.00
          29       4.000       58.00
          30       4.000       60.00
          31       4.000       60.00
          32       4.000       65.00
          33       8.000       51.00
          34       8.000       56.00
          35       8.000       62.00
          36       8.000       63.00
          37       8.000       70.00
          38       8.000       73.00
          39       8.000       76.00
          40       8.000       89.00
          41       8.000       92.00

  19  BLOCK Animals
  20  TREAT POL(Rate; 2)
  21  ANOVA [FPROB=Y] Dose


21.....................................................................................

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

Variate: Dose

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

Animals stratum
Rate                       3    16094.3     5364.8   35.13  <.001
  Lin                      1    15700.1    15700.1  102.81  <.001
  Quad                     1      382.6      382.6    2.51  0.122
  Deviations               1       11.6       11.6    0.08  0.785
Residual                  37     5650.1      152.7
Total                     40    21744.4


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

Variate: Dose

Grand mean  38.9

     Rate     1.00     2.00     4.00     8.00
              18.1     28.2     48.3     70.2
     rep.       12       11        9        9
```

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

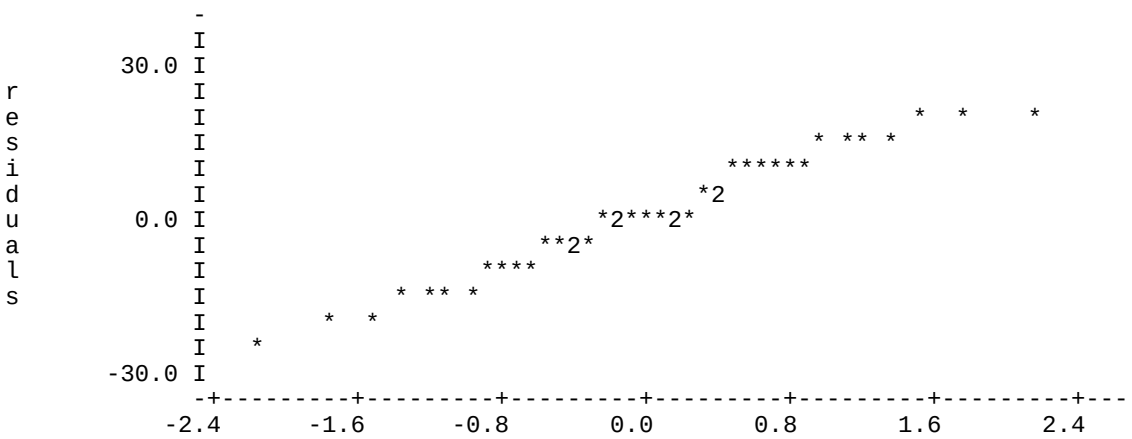
Table	Rate	
rep.	unequal	
d.f.	37	
s.e.d.	5.83	min.rep
	5.45	max-min
	5.04X	max.rep

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

```
  22  AKEEP [FIT=Fit; RES=Res]
  23  APLOT METHOD=fit,normal
```



Normal plot



Step 1: Set up hypotheses

a) H_0: $\gamma_1 = 0$
   H_1: $\gamma_1 \neq 0$

b) H_0: $\gamma_2 = 0$
   H_1: $\gamma_2 \neq 0$

c) H_0: $\alpha_k - \mu - \gamma_1 x_k - \gamma_2 x_k^2 = 0$ for all k
   H_1: $\alpha_k - \mu - \gamma_1 x_k - \gamma_2 x_k^2 \neq 0$ for all k

Step 2: Calculate test statistics

Source	df	SSQ	MSQ	F	Prob
Animals	40	21744.4			
Rates	3	16094.3	5364.8	35.13	<0.001
Linear	1	15700.1	15700.1	102.81	<0.001
Quadratic	1	382.6	382.6	2.51	0.122
Deviations	1	11.6	11.6	0.08	0.785
Residual	37	5650.1	152.7		

Step 3: Decide between hypotheses

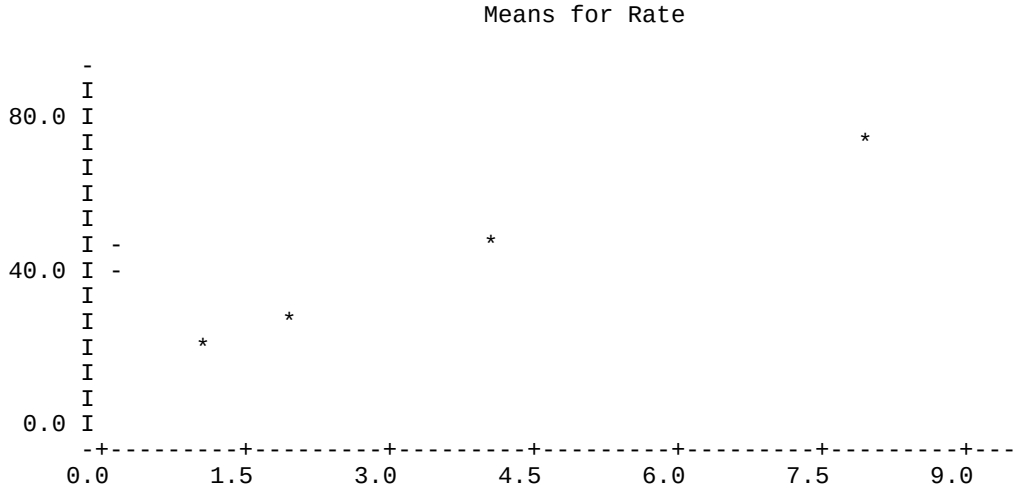
The Deviations and Quadratic terms are not significant. However, the Linear term is highly significant.

Note that the residual-versus-fitted-values plot looks satisfactory so that the homogeneity of variance assumption appears to be met. Also, the normal probability plot appears to be basically linear and so the normality assumption also appears to be met.

We now fit a linear equation and obtain its formula.

```
  24  "
 -25  **** Fit linear only ****
 -26  "
  27  TREAT POL(Rate; 1)
  28  ANOVA [PRINT=aov] Dose


28.....................................................................

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

Variate: Dose

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

Animals stratum
Rate                       3    16094.3     5364.8   35.13
  Lin                      1    15700.1    15700.1  102.81
  Deviations               2      394.2      197.1    1.29
Residual                  37     5650.1      152.7
Total                     40    21744.4
```

```
 29  APOLYNOMIAL Rate; COEFF=Coeffs

***** Equation of the polynomial *****

13.14 + 7.43  * Rate

 30  PRINT #Coeffs

      13.14       7.433

 31  AGRAPH [GRAPH=line] XFACTOR=Rate; BAR=*
```



Examination of the output reveals that the fitted equation is

$$y_i = 13.14 + 7.433\, x_k.$$

III.10 An experiment was conducted to compare the pain relief afforded by three drugs. Altogether there were 39 patients that were randomly assigned one of three drug treatments to receive so that 13 patients received each drug. Each patient was given the drug assigned to them when they declared that their level of pain required relief. The number of hours relief afforded by the first administration of the drug was recorded and is given in the following table.

	D	2	6	4	13	5	8	4	6	7	6	8	12	4
Drug	T1	2	0	3	3	0	0	8	1	4	2	2	1	3
	T2	6	4	4	0	1	8	2	8	12	1	5	2	4

This data is contained in *CRDDrug.gsh* in the directory *G:**Disciplina**Genstat*. Add the 39-level factor Plots and the factor Drug with levels D, T1 and T2 to this Genstat spreadsheet and make sure that the factors and variate are available in the central store.

The drug treatment D was a form of control so it is of interest to compare this treatment with the other two. This could be achieved by using the following orthogonal contrasts:

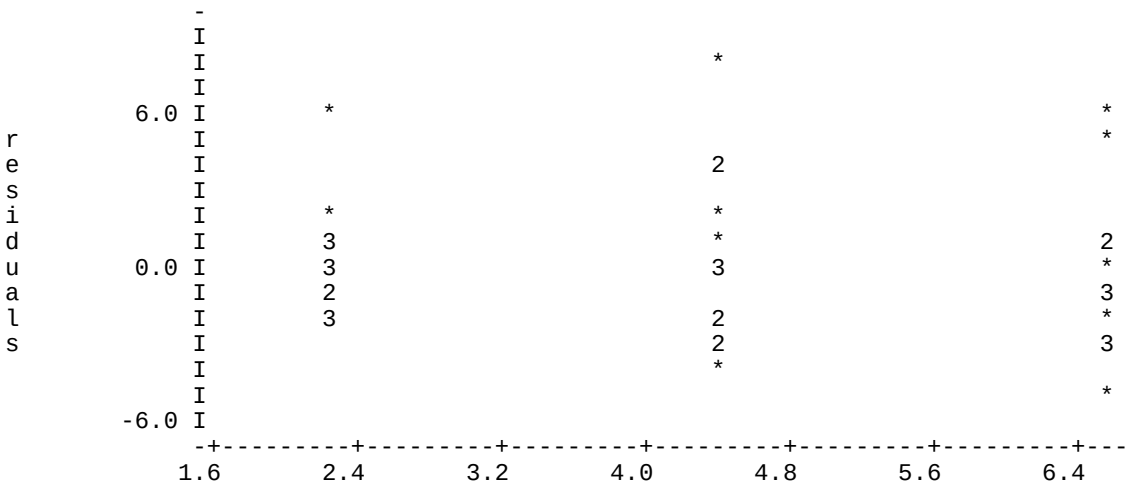
	Drug	
D	T1	T2
2	-1	-1
0	1	-1

Use Genstat to perform a one-way analysis of variance on this data, including the fitting the orthogonal contrasts and checking the assumptions.

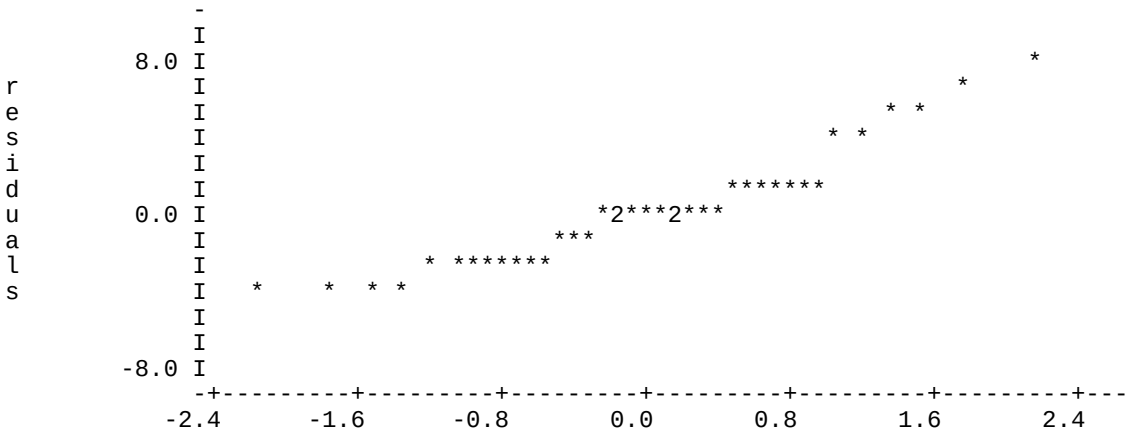
```
Genstat 5  Release 4.1  (PC/Windows NT)              20 March 2000 08:14:46
Copyright 1998, Lawes Agricultural Trust (Rothamsted Experimental Station)

                ____________________________________________

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

   3  "Data taken from File: D:/ANALYSES/LM/ONEFAC/CRDDRUGALL.GSH"
   4  DELETE [redefine=yes] Patients,Drug,Relief
   5  FACTOR [modify=yes;nvalues=39;levels=39] Patients
   6  READ Patients; frepresentation=ordinal

    Identifier    Values   Missing    Levels
      Patients        39         0        39

   9  FACTOR [modify=yes;nvalues=39;levels=3;labels=!t('D','T1','T2')] Drug
  10  READ Drug; frepresentation=ordinal

    Identifier    Values   Missing    Levels
          Drug        39         0         3

  13  VARIATE [nvalues=39] Relief
  14  READ Relief

    Identifier   Minimum      Mean   Maximum    Values   Missing
        Relief     0.000     4.385    13.000        39         0

  17
  18  PRINT Patients,Drug,Relief

    Patients        Drug      Relief
           1           D       2.000
           2           D       6.000
           3           D       4.000
           4           D      13.000
           5           D       5.000
           6           D       8.000
           7           D       4.000
           8           D       6.000
           9           D       7.000
          10           D       6.000
          11           D       8.000
          12           D      12.000
          13           D       4.000
          14          T1       2.000
          15          T1       0.000
          16          T1       3.000
          17          T1       3.000
          18          T1       0.000
          19          T1       0.000
          20          T1       8.000
          21          T1       1.000
          22          T1       4.000
          23          T1       2.000
          24          T1       2.000
          25          T1       1.000
          26          T1       3.000
```

```
          27          T2       6.000
          28          T2       4.000
          29          T2       4.000
          30          T2       0.000
          31          T2       1.000
          32          T2       8.000
          33          T2       2.000
          34          T2       8.000
          35          T2      12.000
          36          T2       1.000
          37          T2       5.000
          38          T2       2.000
          39          T2       4.000

  19  "
 -20  ****Orthogonal Contrasts ****
 -21  "
  22  TEXT ConDrug; !T('D vs rest','T1 vs T2')
  23  MATRIX [ROW=ConDrug; COL=3] Contrasts;   \
  24                      !(2,2(-1), 0,1,-1)
  25  BLOCK Patients
  26  TREAT REG (Drug; 2; Contrasts)
  27  ANOVA [FPROB=Y; PRINT=AOV,MEANS,CONTRAST; PSE=LSD] Relief


27.....................................................................

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

Variate: Relief

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

Patients stratum
Drug                       2    120.615     60.308    6.86  0.003
  D vs rest                1     90.462     90.462   10.29  0.003
  T1 vs T2                 1     30.154     30.154    3.43  0.072
Residual                  36    316.615      8.795
Total                     38    437.231


***** Tables of contrasts *****

Variate: Relief

***** Patients stratum *****

*** Drug contrasts ***

D vs rest     1.08   s.e. 0.336     ss.div. 78.0

T1 vs T2     -1.08   s.e. 0.582     ss.div. 26.0


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

Variate: Relief

Grand mean  4.38

     Drug        D       T1       T2
              6.54     2.23     4.38


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

Table                  Drug
rep.                     13
d.f.                     36
l.s.d.                2.359
```

```
  28  AKEEP [FIT=Fit; RES=Res]
  29  APLOT METHOD=fit,normal
```



Normal plot



Step 1: Set up hypotheses

For each contrast:

H_0: $\mathbf{x}'\boldsymbol{\alpha} = 0$ where $\mathbf{x}$ is a vector of contrast coefficients
H_1: $\mathbf{x}'\boldsymbol{\alpha} \neq 0$

Step 2: Calculate test statistics

Source	df	SSQ	MSQ	F	Prob
Patients	38	437.231			
Drug	2	120.615	60.308	6.86	0.003
D vs rest	1	90.462	90.462	10.29	0.003
T1 vs T2	1	30.154	30.154	3.43	0.072
Residual	36	316.615	8.795		

Step 3: Decide between hypotheses

The relief provided by T1 and T2 is less than that provided by D; there is not a significant difference between T1 and T2.

Note that the residual-versus-fitted-values plot looks satisfactory so that the homogeneity of variance assumption appears to be met. However, the normal probability plot has some evidence of curvature in it and so it seems that the normality assumption may not be met. As ANOVA is robust to departures from normality we will continue with the analysis.

The means for the different drugs are given in the following table.

	Drug		
	D	T1	T2
Mean	6.54	2.23	4.38

Note that the estimate for the first orthogonal contrast is 1.08 and that

$$6.54 - \frac{2.23 + 4.38}{2} = 3.235 \text{ and } \frac{3.235}{3} = 1.08.$$

So the difference between D and the mean of T1 and T2 is 3.24.