THE DESIGN AND MIXED-MODEL ANALYSIS OF EXPERIMENTS

III. Completely Randomized Design

(ref. Myers and Milton, sec. 5.2 and 6.2; Mead, sec. 6.2)

III.A Design of a CRD III-1
III.B Models and estimation for a CRD III-2
III.C Hypothesis testing using the ANOVA method III-7
a) Expressions for the sums of squares III-9
b) Degrees of freedom III-13
c) Expected mean squares III-14
d) Distribution of the F statistic III-17
e) Analysis of variance table III-18
f) Analysis of the rat example III-18
III.D Computation in Genstat III-20
a) Regression analysis III-20
b) Mean-operator Analysis of Rat data III-22
III.E Diagnostic checking III-23
III.F Treatment differences III-30
a) Multiple comparisons procedures III-30
b) Fitting submodels III-35
c) Orthogonal contrasts III-40
III.G Summary III-44

III.A Design of a CRD

Definition III.1: An experiment is set up using a Completely Randomized Design (CRD) when each treatment is applied a specified, possibly unequal, number of times, the particular units to receive a treatment being selected completely at random. ■

To obtain a randomized layout for the experiment in Genstat set up a factor for the units and a factor for the treatments in standard (unrandomized) order. Then use *Stats > Design > Randomize* to obtain a randomized layout and PDESIGN to print the randomized layout.

Example III.1 Rat experiment

For example, consider an experiment in which the effects of three diets on rats is to be investigated. Suppose I have 6 rats to be fed one of three diets, 3 rats to be fed diet A, 2 diet B and 1 diet C. In a Genstat spreadsheet create two factors: the factor Rats with the values 1–6 and the factor Diet with the values 1, 1, 1, 2, 2, 3. Use *Spread > Update Data* to make sure that the factors are in the central store. Then select *Stats > Design > Randomize* and set *Data to Randomize:* to Diet and *Block Structure:* to Rats. Click *OK*. Finally, use PDESIGN as shown in the following output to print out the randomized layout.

```
 15  PRINT Rats,Diet

       Rats        Diet
          1           1
          2           1
          3           1
          4           2
          5           2
          6           3

 16  RANDOMIZE [BLOCKSTRUCTURE=Rats; SEED=206203] Diet
 17  PDESIGN [BLOCK=Rats; TREAT=Diet]

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


 Rats
    1   2
    2   1
    3   1
    4   3
    5   2
    6   1

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

The layout for the experiment would be:

Rat	1	2	3	4	5	6
Diet	B	A	A	C	B	A

III.B Models and estimation for a CRD

The analysis of CRD experiments uses the techniques discussed for regression. That is, least-squares or maximum likelihood estimation of the parameters of a linear model and hypothesis testing based on the ANOVA method or maximum likelihood ratio testing. We will investigate linear models and the estimation of its parameters using our rat experiment.

Example III.1 Rat experiment (continued)

Suppose, the experimenter measured the liver weight as a percentage of total body weight at the end of the experiment. The results of the experiment are as follows:

Rat	1	3	4	2	6	5
Diet	A	A	A	B	B	C
Liver wt.	3.3	3.1	2.9	3.2	3.4	2.7

The analysis of this experiment will be based on a linear model, that is

$$E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\theta} \text{ and } \text{var}[\mathbf{Y}] = \mathbf{V}_{\mathbf{Y}} = \sigma^2 \mathbf{I}_n .$$

Now, the trick here is what are $\mathbf{X}$ and $\boldsymbol{\theta}$ going to be? I suppose an intuitively obvious thing to do might be to allocate codes for diet (A = 1; B = 2; C = 3) and use these as

values to form the explanatory variable whose values make up the columns of **X**. Thus, regression equations for the example, including a term for the intercept, are as follows:

$$\begin{bmatrix} E[Y_1] \\ E[Y_2] \\ E[Y_3] \\ E[Y_4] \\ E[Y_5] \\ E[Y_6] \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \mu \\ \gamma \end{bmatrix} \text{ and } \mathbf{V} = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2 \end{bmatrix}$$

The above model can then be fitted using simple linear regression techniques. The fitted equation is:

$$\widehat{E[Y]} = 3.30 - 0.120x$$

That is, $\hat{\mu} = 3.30$, $\hat{\gamma} = -0.12$.

Now the model states that $\widehat{E[Y_i]} = \hat{\mu} + \hat{\gamma} \times x_i$

so that $\widehat{E[Y_i]} = 3.30 + (-0.12) \times x_i$

Hence $\widehat{E[Y_1]} = 3.30 + (-0.12) \times 1 = 3.18$

and $\widehat{E[Y_5]} = 3.30 + (-0.12) \times 2 = 3.06$

That is, for each unit increase in diet, liver weight decreases by 0.120. However, does this make sense? The fitted model says that, for each unit increase in diet, % liver weight decreases by 0.120. It is sensible **only if** the diets differences are based on equally spaced levels of some component; for example, if the diets represent 2, 4 and 6 mg of copper added to each 100g of food, then the unit increase in diet is 2 mg of copper. But, there is no unit increase if the diets are unequally spaced, such as 2,4, and 10 mg copper added, or the diets differ qualitatively, such as if the diets represented the addition of three different amino acids.

To overcome this problem the regression is performed on explanatory variables *indicator variables*. In this method the explanatory variables are also called *factors* and the possible values they take *levels* (remember they are limited). Thus, in our example, we have a factor Diet and it has three levels: A, B and C. Indicator variables are formed, a variable for each level of the factor; the values of a variable are either 1 or 0, 1 when the unit has the particular level and 0 otherwise. Thus, our model becomes:

$$\begin{bmatrix} E[Y_1] \\ E[Y_2] \\ E[Y_3] \\ E[Y_4] \\ E[Y_5] \\ E[Y_6] \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \text{ and } \mathbf{V} = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2 \end{bmatrix}$$

Hence,

$$\mathrm{E}[\mathrm{Y_i}] = \alpha_k \text{ and } var[\mathrm{Y_i}] = \sigma^2, \ cov[\mathrm{Y_i}, \mathrm{Y_j}] = 0, \ (i \neq j).$$

These can be written as $\boldsymbol{\psi} = E[\mathbf{Y}] = \mathbf{X}_\mathrm{T}\boldsymbol{\alpha}$ and $\mathbf{V} = \sigma^2\mathbf{I}_n$

where $\mathbf{X}_\mathrm{T}$ is a 6x3 matrix containing the indicator variables.

This model suggests that there are 3 different expected (mean) values for the diets. This contrasts with the previous model that says that liver weight increase linearly as the Diet increases.

Now, the model for the expected values is still of the general form $E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\theta}$ and on assuming $\mathbf{Y}$ is distributed N($\boldsymbol{\psi}$, $\sigma^2\mathbf{I}$), we can use standard least squares or maximum likelihood estimation. Note that N($\boldsymbol{\psi}$, $\sigma^2\mathbf{I}_n$) stands for (multivariate) normal with expected value $\boldsymbol{\psi}$ and variance-covariance $\sigma^2\mathbf{I}_n$.

In applying standard least squares estimation, the first step is to form the normal equations which are given by $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\theta}} = \mathbf{X}'\mathbf{Y}$. In the example, we obtain:

$$\begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \end{bmatrix} = \begin{bmatrix} Y_1 + Y_2 + Y_3 \\ Y_4 + Y_5 \\ Y_6 \end{bmatrix}$$

Thus, the solution of the normal equations is given by

$$\begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \end{bmatrix} = \begin{bmatrix} 1/3 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} Y_1 + Y_2 + Y_3 \\ Y_4 + Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} \dfrac{Y_1 + Y_2 + Y_3}{3} \\ \dfrac{Y_4 + Y_5}{2} \\ \dfrac{Y_6}{1} \end{bmatrix} = \begin{bmatrix} 3.10 \\ 3.30 \\ 2.70 \end{bmatrix}$$

That is, the estimates are the means of diet A, B and C

The fitted values or linear predictors are given by:

$$\mathbf{X}\hat{\boldsymbol{\theta}} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3.10 \\ 3.30 \\ 2.70 \end{bmatrix} = \begin{bmatrix} 3.10 \\ 3.10 \\ 3.10 \\ 3.30 \\ 3.30 \\ 2.70 \end{bmatrix}$$

That is, the fitted values obtained from the least squares estimators are the means for the treatments applied to the units. We prove this is generally the case for the CRD in the next theorem.

Theorem III.1: Suppose that the *k*th treatment is observed n_k times, $k = 1 \ldots t$, and, without loss of generality, that **Y** is ordered so that all the observations for the first treatment occur in the first n_1 rows, those for the second treatment occur in the next n_2 rows and so on with the last treatment occurring in the last n_t rows. Let **T** denote the estimator of the vector of fitted values for the expectation model $E[\mathbf{Y}] = \mathbf{X}_\mathrm{T}\boldsymbol{\alpha}$ where $\boldsymbol{\alpha}$ is the *t*-vector of parameters representing the population mean responses for the *t* treatments, $\mathbf{X}_\mathrm{T}$ a $n \times t$ matrix that specifies the treatment that each observation received and *n* is the total number of observations. Then,

$$\mathbf{T} = \mathbf{P}_\mathrm{T}\mathbf{Y}$$

where $\mathbf{P}_\mathrm{T} = \mathbf{X}_\mathrm{T}\left(\mathbf{X}_\mathrm{T}'\mathbf{X}_\mathrm{T}\right)^{-1}\mathbf{X}_\mathrm{T}'$ and $\mathbf{P}_\mathrm{T}$ is a partitioned matrix consisting of $t \times t$ elements whose diagonal elements are $n_k^{-1}\mathbf{J}_{n_k}$ and whose off-diagonal elements are $n_k \times n_\ell$ matrices of zeroes. $\mathbf{J}_{n_k}$ is the square matrix of order n_k whose elements are all one. That is, the vector **T** is the *n*-vector obtained from **Y** by replacing each observation with its treatment mean.

Proof:

From Definition II.10, it is clear that $\mathbf{T} = \mathbf{P}_\mathrm{T}\mathbf{Y}$ where $\mathbf{P}_\mathrm{T} = \mathbf{X}_\mathrm{T}\left(\mathbf{X}_\mathrm{T}'\mathbf{X}_\mathrm{T}\right)^{-1}\mathbf{X}_\mathrm{T}'$. It is only necessary to show that $\mathbf{P}_\mathrm{T}$ is a partitioned matrix consisting of $t \times t$ elements whose diagonal elements are $n_k^{-1}\mathbf{J}_{n_k}$ and whose off-diagonal elements are $n_k \times n_\ell$ matrices of zeroes.

Now, $\mathbf{X}_\mathrm{T} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1 \times 1} & \cdots & \mathbf{0}_{n_1 \times 1} \\ \mathbf{0}_{n_2 \times 1} & \mathbf{1}_{n_2} & \cdots & \mathbf{0}_{n_2 \times 1} \\ \vdots & \vdots & & \vdots \\ \mathbf{0}_{n_t \times 1} & \mathbf{0}_{n_t \times 1} & \cdots & \mathbf{1}_{n_t} \end{bmatrix}$ is the partitioned form of $\mathbf{X}_\mathrm{T}$.

Multiplying and transposing partitioned matrices is achieved by treating their elements as if they were single elements of a matrix, provided that the submatrices are conformable. It is reasonably straightforward to show that,

$$\mathbf{P}_{\mathrm{T}} = \mathbf{X}_{\mathrm{T}}\left(\mathbf{X}_{\mathrm{T}}'\mathbf{X}_{\mathrm{T}}\right)^{-1}\mathbf{X}_{\mathrm{T}}'$$
$$= \begin{bmatrix} \mathbf{1}_{n_1}\left(\mathbf{1}_{n_1}'\mathbf{1}_{n_1}\right)^{-1}\mathbf{1}_{n_1}' & \mathbf{0}_{n_1\times n_2} & \cdots & \mathbf{0}_{n_1\times n_t} \\ \mathbf{0}_{n_2\times n_1} & \mathbf{1}_{n_2}\left(\mathbf{1}_{n_2}'\mathbf{1}_{n_2}\right)^{-1}\mathbf{1}_{n_2}' & \cdots & \mathbf{0}_{n_2\times n_t} \\ \vdots & \vdots & & \vdots \\ \mathbf{0}_{n_t\times n_1} & \mathbf{0}_{n_t\times n_2} & \cdots & \mathbf{1}_{n_t}\left(\mathbf{1}_{n_t}'\mathbf{1}_{n_t}\right)^{-1}\mathbf{1}_{n_t}' \end{bmatrix}$$

But $\mathbf{1}_{n_k}'\mathbf{1}_{n_k} = n_k$ so that $\mathbf{1}_{n_k}\left(\mathbf{1}_{n_k}'\mathbf{1}_{n_k}\right)^{-1}\mathbf{1}_{n_k}' = n_k^{-1}\mathbf{1}_{n_k}\mathbf{1}_{n_k}' = n_k^{-1}\mathbf{J}_{n_k}$. Hence,

$$\mathbf{P}_{\mathrm{T}} = \begin{bmatrix} \mathbf{1}_{n_1}\left(\mathbf{1}_{n_1}'\mathbf{1}_{n_1}\right)^{-1}\mathbf{1}_{n_1}' & \mathbf{0}_{n_1\times n_2} & \cdots & \mathbf{0}_{n_1\times n_t} \\ \mathbf{0}_{n_2\times n_1} & \mathbf{1}_{n_2}\left(\mathbf{1}_{n_2}'\mathbf{1}_{n_2}\right)^{-1}\mathbf{1}_{n_2}' & \cdots & \mathbf{0}_{n_2\times n_t} \\ \vdots & \vdots & & \vdots \\ \mathbf{0}_{n_t\times n_1} & \mathbf{0}_{n_t\times n_2} & \cdots & \mathbf{1}_{n_t}\left(\mathbf{1}_{n_t}'\mathbf{1}_{n_t}\right)^{-1}\mathbf{1}_{n_t}' \end{bmatrix}$$
$$= \begin{bmatrix} n_1^{-1}\mathbf{J}_{n_1} & \mathbf{0}_{n_1\times n_2} & \cdots & \mathbf{0}_{n_1\times n_t} \\ \mathbf{0}_{n_2\times n_1} & n_2^{-1}\mathbf{J}_{n_2} & \cdots & \mathbf{0}_{n_2\times n_t} \\ \vdots & \vdots & & \vdots \\ \mathbf{0}_{n_t\times n_1} & \mathbf{0}_{n_t\times n_2} & \cdots & n_t^{-1}\mathbf{J}_{n_t} \end{bmatrix}$$

Now $\mathbf{T} = \mathbf{P}_{\mathrm{T}}\mathbf{Y}$ so that it is clear from the above expression that the first n_1 elements of $\mathbf{T}$ are the mean of the first treatment, the next n_2 elements are the mean of the second treatment and so on.

■

So $\mathbf{P}_{\mathrm{T}}$ could be called the treatment mean operator as it computes the treatment means from the vector to which it is applied and replaces each element of this vector with its treatment mean.

Example III.1 Rat experiment (continued)

For the liver weight experiment,

$$\mathbf{X}_{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{P}_{\mathrm{T}} = \mathbf{X}_{\mathrm{T}}\left(\mathbf{X}_{\mathrm{T}}'\mathbf{X}_{\mathrm{T}}\right)^{-1}\mathbf{X}_{\mathrm{T}}' = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

It turns out that the fitted values for orthogonal experiments, a large number of the experiments discussed in this course, are functions of means. However, nonorthogonal experiments are a practically important class of experiments and the fitted values are not so easily obtained.

III.C Hypothesis testing using the ANOVA method

Often an experimenter will be interested in determining if there are significant differences between the treatment means; in the example, whether or not there are significant differences between the diet means. This can be formulated as determining which model best describes the data by fitting a sequence of models. In this case, the sequence consists of the two models:

1. $E[y_i] = \mu$ *or* $\boldsymbol{\psi} = \mathbf{X}_G\mu$ where $\mathbf{X}_G = \mathbf{1}'$
2. $E[y_i] = \alpha_k$ *or* $\boldsymbol{\psi} = \mathbf{X}_T\boldsymbol{\alpha}$

The first model, referred to as the minimal expectation model, says that the population mean response is the same for all observations, irrespective of diet. The second model is referred to as the maximal expectation model.

Definition III.2: The **minimal expectation model** is the simplest model for the expectation that is to be considered in analysing an experiment. ■

Definition III.3: The **maximal expectation model** is the most complicated model for the expectation that is to be considered in analysing an experiment. ■

Note that, as for simple regression, there is no interest in the model $\boldsymbol{\psi} = \mathbf{0}$ and only a minimal and maximal model are under consideration for the CRD experiment.

This situation is similar to the hypothesis on subvectors considered in the context of linear regression. However, rather than just setting some parameters to zero, it has to be recognised that one model is a submodel of another. We express this by saying that one model is marginal to another.

Definition III.4: For two models, $\boldsymbol{\psi} = \mathbf{X}_1\boldsymbol{\theta}_1$ and $\boldsymbol{\psi} = \mathbf{X}_2\boldsymbol{\theta}_2$, the first model is **marginal** to the second if $\mathcal{C}(\mathbf{X}_1) \subseteq \mathcal{C}(\mathbf{X}_2)$, that is if the columns of $\mathbf{X}_1$ can be written as linear combinations of the columns of $\mathbf{X}_2$. ■

For the two models being considered in the example, $\boldsymbol{\psi} = \mathbf{X}_G\mu$ is marginal to the model $\boldsymbol{\psi} = \mathbf{X}_T\boldsymbol{\alpha}$ since $\mathcal{C}(\mathbf{X}_G) \subset \mathcal{C}(\mathbf{X}_T)$ in that an element from a row of $\mathbf{X}_G$ is the sum of the elements in the corresponding row of $\mathbf{X}_T$. Note that this property is not symmetric in that $\boldsymbol{\psi} = \mathbf{X}_T\boldsymbol{\alpha}$ is **not** marginal to $\boldsymbol{\psi} = \mathbf{X}_G\mu$ as $\mathcal{C}(\mathbf{X}_T) \not\subset \mathcal{C}(\mathbf{X}_G)$.

As we are using the ANOVA method, an analysis of variance table is formulated based on partitioning the Total corrected sum of squares into a two sums of squares, one reflecting Treatment differences and the other Residual variation. The sums of squares in the ANOVA table are obtained from the residual sums of squares (or unscaled deviance) after fitting each of the models in the sequence. These are denoted D(μ) and D(α). We then compute an F statistic to be used in deciding which model best describes the data. A summary of the hypothesis test is a follows:

Step 1: Set up hypotheses

H_0: $\alpha_1 = \alpha_2 = ... = \alpha_t = \mu$ (or $\boldsymbol{\psi} = \mathbf{X}_G\mu$)

H_1: at least one pair of population treatment means is different (or $\boldsymbol{\psi} = \mathbf{X}_T\boldsymbol{\alpha}$)

Step 2: Calculate test statistic

The analysis of variance table for a CRD is:

Source	df	SSq	MSq (s^2)	F
Units	n–1	D(μ)		
Treatments	t–1	R(α| μ)	R(α| μ)[!]/(t–1) $\left(= s_T^2\right)$	s_T^2 / s_R^2
Residual	n–t	D(α)	D(α)/(n–t) $\left(= s_R^2\right)$	

[!]Reduction in SSq for α given μ = R(α\| μ) = D(μ) – D(α)

From this table we can see that we take the total variation amongst the units and partition it into two parts: variance of difference between treatment means and the left-over (residual) unit variation.

Step 3: Decide between hypotheses

Determine probability of observed F value which has ν_1 = numerator d.f. = t-1 and ν_2 = denominator d.f. = n–t.

If $P(F \geq F_{(t\text{-}1),(n\text{-}t)}) \leq 0.05$ then the evidence suggest that the null hypothesis be rejected.

Comparison with traditional one-way ANOVA

The above analysis of variance table is essentially the same as the traditional one-way ANOVA table — the values of the F statistic from each table are exactly the same. As illustrated in the table below, the labelling differs and the Total is normally be placed at the bottom of the table, not at the top. The reason for the difference is that the table we have presented exhibits the confounding in the experiment. The indenting of Treatments under Units signifies that treatment differences are confounded or "mixed-up" with unit differences. Also, the Residual reflects differences between the units once the treatment differences have been removed or subtracted off. We will return to this after we have examined the expected mean squares for this analysis.

Source	df	Source in one-way ANOVA
Units	n–1	Total
Treatments	t–1	Between Treatments
Residual	n–t	Within Treatments

We now derive expressions for and properties of the quantities in the analysis of variance table for the CRD.

a) Expressions for the sums of squares

First, we require algebraic expressions for the sums of squares. Specifically, algebraic expressions for D(μ), D(α) and R($\alpha|\ \mu$). Note that D(μ) is the sums of squares of the deviations obtained by subtracting the fitted values under the first model from the observations; D(α) is similarly obtained using the fitted values from the second model.

Expression for D(μ)

Theorem III.2: Let **G** denote the estimator of the vector of fitted values for the expectation model $E[\mathbf{Y}] = \mathbf{X}_G\mu$ where μ is the overall population mean response, $\mathbf{X}_G = \mathbf{1}_n$ is a column vector consisting of n ones and n is the number of observations in the random sample. Then,

$$\mathbf{G} = \mathbf{P}_G\mathbf{Y} \text{ and } D(\mu) = (\mathbf{Y}-\mathbf{G})'(\mathbf{Y}-\mathbf{G}) = \mathbf{Y}'\mathbf{R}_G\mathbf{Y}$$

where $\mathbf{P}_G = \mathbf{X}_G(\mathbf{X}_G'\mathbf{X}_G)^{-1}\mathbf{X}_G' = n^{-1}\mathbf{J}_n$, $\mathbf{R}_G = \mathbf{I} - \mathbf{P}_G$ and $\mathbf{J}_n$ is the square matrix of order n whose elements are all 1.

Proof: From theorem II.9, it is clear that $\mathbf{G} = \mathbf{P}_G\mathbf{Y}$ and $D(\mu) = (\mathbf{Y}-\mathbf{G})'(\mathbf{Y}-\mathbf{G}) = \mathbf{Y}'\mathbf{R}_G\mathbf{Y}$ where $\mathbf{P}_G = \mathbf{X}_G(\mathbf{X}_G'\mathbf{X}_G)^{-1}\mathbf{X}_G'$ and $\mathbf{R}_G = \mathbf{I} - \mathbf{P}_G$. It is only necessary to show that $\mathbf{X}_G(\mathbf{X}_G'\mathbf{X}_G)^{-1}\mathbf{X}_G' = n^{-1}\mathbf{J}_n$.

Now, $\mathbf{X}_G = \mathbf{1}_n$ so that $\mathbf{X}_G'\mathbf{X}_G = \mathbf{1}_n'\mathbf{1}_n = n$ and $\mathbf{X}_G(\mathbf{X}_G'\mathbf{X}_G)^{-1}\mathbf{X}_G' = n^{-1}\mathbf{X}_G\mathbf{X}_G' = n^{-1}\mathbf{J}_n$. ■

$\mathbf{P}_G$ is called the grand mean operator; when it is applied to a vector it replaces every element of the vector by the grand mean of the elements of the vector.

Example III.1 Rat experiment (continued)

For example, for the liver weight experiment,

$$\mathbf{P}_G = \frac{1}{6}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \text{ and } \mathbf{G} = \mathbf{P}_G\mathbf{y} = \begin{bmatrix} \bar{y} \\ \bar{y} \\ \bar{y} \\ \bar{y} \\ \bar{y} \\ \bar{y} \end{bmatrix} = \begin{bmatrix} 3.1 \\ 3.1 \\ 3.1 \\ 3.1 \\ 3.1 \\ 3.1 \end{bmatrix} \text{ where } \bar{y} = \frac{\sum_{i=1}^{6} y_i}{6}$$

Expression for D(α)

Theorem III.3: Let $\mathbf{T} = \mathbf{P}_T\mathbf{Y}$ denote the estimator of the vector of fitted values for the expectation model $E[\mathbf{Y}] = \mathbf{X}_T\boldsymbol{\alpha}$ where $\boldsymbol{\alpha}$ is the t-vector of parameters representing the population mean responses for the t treatments, $\mathbf{X}_T$ a $n \times t$ matrix that specifies the treatment that each observation received, $\mathbf{P}_T = \mathbf{X}_T\left(\mathbf{X}_T'\mathbf{X}_T\right)^{-1}\mathbf{X}_T'$ and n is the total number of observations. Then,

$$D(\alpha) = (\mathbf{Y} - \mathbf{T})'(\mathbf{Y} - \mathbf{T}) = \mathbf{Y}'\mathbf{R}_T\mathbf{Y}$$

where $\mathbf{R}_T = \mathbf{I} - \mathbf{P}_T$.

Proof: The proof of this result is a straightforward application of theorem II.9. ■

Expression for R(α| μ)

Our expression for R($\alpha|\ \mu$) involves obtaining the difference between two residual sums of squares or deviances. The following theorem provides us with an expression for the reduction sum of squares in terms of a quadratic form whose matrix is symmetric and idempotent. That is, we have $R(\alpha|\ \mu) = D(\mu) - D(\alpha) = \mathbf{Y}'\mathbf{A}\mathbf{Y}$ for some symmetric idempotent $\mathbf{A}$.

Lemma III.1: Let $\mathbf{R}_G$ and $\mathbf{R}_T$ be the symmetric idempotents such that $\mathbf{R}_G\mathbf{Y}$ and $\mathbf{R}_T\mathbf{Y}$ are the estimators of the errors for the expectation models $E[\mathbf{Y}] = \mathbf{X}_G\mu$ and $E[\mathbf{Y}] = \mathbf{X}_T\boldsymbol{\alpha}$, respectively, where μ is the overall population mean response, $\boldsymbol{\alpha}$ is the t-vector of parameters representing the population mean responses for the t treatments, $\mathbf{X}_G = \mathbf{1}_n$ is a column vector consisting of n ones, $\mathbf{X}_T$ a $n \times t$ matrix that specifies the treatment that each observation received and n is the number of observations in the random sample. Then,

$$\mathbf{R}_T = \mathbf{R}_G\mathbf{R}_T = \mathbf{R}_T\mathbf{R}_G$$

where $\mathbf{R}_G = \mathbf{I} - \mathbf{P}_G$, $\mathbf{P}_G = n^{-1}\mathbf{J}_n$ and $\mathbf{J}_n$ is the square matrix of order n whose elements are all 1;
$\mathbf{R}_T = \mathbf{I} - \mathbf{P}_T$ and $\mathbf{P}_T$ is a partitioned matrix consisting of $t \times t$ elements whose diagonal elements are $n_k^{-1}\mathbf{J}_{n_k}$ and whose off-diagonal elements are $n_k \times n_\ell$ matrices of zeroes.

Proof: Now $\mathbf{R}_G\mathbf{R}_T = (\mathbf{I} - \mathbf{P}_G)(\mathbf{I} - \mathbf{P}_T) = \mathbf{I} - \mathbf{P}_T - \mathbf{P}_G + \mathbf{P}_G\mathbf{P}_T$ and $\mathbf{R}_T\mathbf{R}_G = \mathbf{I} - \mathbf{P}_G - \mathbf{P}_T + \mathbf{P}_T\mathbf{P}_G$ so that $\mathbf{R}_G\mathbf{R}_T = \mathbf{R}_T\mathbf{R}_G$ if $\mathbf{P}_G\mathbf{P}_T = \mathbf{P}_T\mathbf{P}_G$. Also, $\mathbf{R}_T = \mathbf{R}_G\mathbf{R}_T$ if $\mathbf{P}_G\mathbf{P}_T = \mathbf{P}_G$.

So we have to prove that $\mathbf{P}_G = \mathbf{P}_G\mathbf{P}_T = \mathbf{P}_T\mathbf{P}_G$.

To show that $\mathbf{P}_G\mathbf{P}_T = \mathbf{P}_T\mathbf{P}_G$, firstly note that we can write $\mathbf{X}_G$ as follows:

$$\mathbf{X}_G = \begin{bmatrix} \mathbf{1}_{n_1} \\ \mathbf{1}_{n_2} \\ \vdots \\ \mathbf{1}_{n_t} \end{bmatrix} \text{ so that } \mathbf{P}_G = n^{-1}\mathbf{J}_n = n^{-1}\begin{bmatrix} \mathbf{J}_{n_1} & \mathbf{J}_{n_1\times n_2} & \cdots & \mathbf{J}_{n_1\times n_t} \\ \mathbf{J}_{n_2\times n_1} & \mathbf{J}_{n_2} & \cdots & \mathbf{J}_{n_2\times n_t} \\ \vdots & \vdots & & \vdots \\ \mathbf{J}_{n_t\times n_1} & \mathbf{J}_{n_t\times n_2} & \vdots & \mathbf{J}_{n_t} \end{bmatrix}$$

Hence, using the expression for $\mathbf{P}_T$ from theorem III.1,

$$\mathbf{P}_G\mathbf{P}_T = n^{-1}\begin{bmatrix} \mathbf{J}_{n_1} & \mathbf{J}_{n_1\times n_2} & \cdots & \mathbf{J}_{n_1\times n_t} \\ \mathbf{J}_{n_2\times n_1} & \mathbf{J}_{n_2} & \cdots & \mathbf{J}_{n_2\times n_t} \\ \vdots & \vdots & & \vdots \\ \mathbf{J}_{n_t\times n_1} & \mathbf{J}_{n_t\times n_2} & \vdots & \mathbf{J}_{n_t} \end{bmatrix}\begin{bmatrix} n_1^{-1}\mathbf{J}_{n_1} & \mathbf{0}_{n_1\times n_2} & \cdots & \mathbf{0}_{n_1\times n_t} \\ \mathbf{0}_{n_2\times n_1} & n_2^{-1}\mathbf{J}_{n_2} & \cdots & \mathbf{0}_{n_2\times n_t} \\ \vdots & \vdots & & \vdots \\ \mathbf{0}_{n_t\times n_1} & \mathbf{0}_{n_t\times n_2} & \cdots & n_t^{-1}\mathbf{J}_{n_t} \end{bmatrix}$$

Now $\mathbf{J}_{n_i\times n_k}\mathbf{J}_{n_k\times n_j} = n_k\mathbf{J}_{n_i\times n_j}$ so that

$$\begin{aligned} \mathbf{P}_G\mathbf{P}_T &= n^{-1}\begin{bmatrix} \mathbf{J}_{n_1} & \mathbf{J}_{n_1\times n_2} & \cdots & \mathbf{J}_{n_1\times n_t} \\ \mathbf{J}_{n_2\times n_1} & \mathbf{J}_{n_2} & \cdots & \mathbf{J}_{n_2\times n_t} \\ \vdots & \vdots & & \vdots \\ \mathbf{J}_{n_t\times n_1} & \mathbf{J}_{n_t\times n_2} & \vdots & \mathbf{J}_{n_t} \end{bmatrix} \\ &= n^{-1}\mathbf{J}_n \\ &= \mathbf{P}_G \end{aligned}$$

Similarly, $\mathbf{P}_T\mathbf{P}_G = \mathbf{P}_G$ and the result is proved. ■

Note that $\mathbf{P}_T\mathbf{P}_G = \mathbf{P}_G = \mathbf{P}_G\mathbf{P}_T$ is not a surprising result when we think what each of these does when they are applied to a vector of observations. For example, $\mathbf{P}_G\mathbf{y}$ produces the vector of grand means. So, $\mathbf{P}_T\mathbf{P}_G\mathbf{y}$ can be obtained by applying $\mathbf{P}_T$ to $\mathbf{P}_G\mathbf{y}$, that is by obtaining the treatment means of the vector of grand means. Any mean of a set of grand means will be the grand mean and so $\mathbf{P}_T\mathbf{P}_G\mathbf{y} = \mathbf{P}_G\mathbf{y}$.

Theorem III.4: Let **G** and **T** denote the estimators of the vectors of fitted values for the expectation models $E[\mathbf{Y}] = \mathbf{X}_G\mu$ and $E[\mathbf{Y}] = \mathbf{X}_T\boldsymbol{\alpha}$, respectively, where μ is the overall population mean response, $\boldsymbol{\alpha}$ is the t-vector of parameters representing the population mean responses for the t treatments, $\mathbf{X}_G = \mathbf{1}_n$ is a column vector consisting of n ones, $\mathbf{X}_T$ a $n\times t$ matrix that specifies the treatment that each observation received and n is the number of observations in the random sample. Then,

$$R(\alpha|\,\mu) = D(\mu) - D(\alpha) = \mathbf{T}_e'\mathbf{T}_e$$

and $\mathbf{T}_e = \mathbf{P}_T\mathbf{R}_G\mathbf{Y}$ is the estimator of the n-vector of treatment effects with $\mathbf{P}_T\mathbf{R}_G$ symmetric and idempotent,

where D(μ) = $\mathbf{Y}'\mathbf{R}_G\mathbf{Y}$, D($\alpha$) = $\mathbf{Y}'\mathbf{R}_T\mathbf{Y}$ and $\mathbf{R}_G$ and $\mathbf{R}_T$ are as defined in lemma III.1.

Proof: $R(\alpha|\,\mu) = D(\mu) - D(\alpha) = \mathbf{Y}'\mathbf{R}_G\mathbf{Y} - \mathbf{Y}'\mathbf{R}_T\mathbf{Y} = \mathbf{Y}'(\mathbf{R}_G - \mathbf{R}_T)\mathbf{Y}$.

But, from lemma III.1, $\mathbf{R}_T = \mathbf{R}_G\mathbf{R}_T = \mathbf{R}_T\mathbf{R}_G$.

Hence, $R(\alpha|\,\mu) = \mathbf{Y}'(\mathbf{R}_G - \mathbf{R}_T)\mathbf{Y} = \mathbf{Y}'(\mathbf{R}_G - \mathbf{R}_T\mathbf{R}_G)\mathbf{Y} = \mathbf{Y}'(\mathbf{I} - \mathbf{R}_T)\mathbf{R}_G\mathbf{Y} = \mathbf{Y}'\mathbf{P}_T\mathbf{R}_G\mathbf{Y}$.

For $\mathbf{T}_e = \mathbf{P}_T\mathbf{R}_G\mathbf{Y}$, then $\mathbf{Y}'\mathbf{P}_T\mathbf{R}_G\mathbf{Y} = \mathbf{T}_e'\mathbf{T}_e$ provided $\mathbf{P}_T\mathbf{R}_G$ is symmetric and idempotent.

To establish that $\mathbf{P}_T\mathbf{R}_G$ is symmetric, note that $\mathbf{P}_T\mathbf{R}_G = \mathbf{P}_T'\mathbf{R}_G' = (\mathbf{R}_G\mathbf{P}_T)' = (\mathbf{P}_T\mathbf{R}_G)'$ since $\mathbf{P}_T\mathbf{R}_G = \mathbf{P}_T - \mathbf{P}_T\mathbf{P}_G = \mathbf{P}_T - \mathbf{P}_G\mathbf{P}_T = \mathbf{R}_G\mathbf{P}_T$

To establish that $\mathbf{P}_T\mathbf{R}_G$ is idempotent, note that $\mathbf{P}_T\mathbf{R}_G = \mathbf{P}_T\mathbf{P}_T\mathbf{R}_G\mathbf{R}_G = \mathbf{P}_T\mathbf{R}_G\mathbf{P}_T\mathbf{R}_G$. ■

The treatment effects, $\mathbf{T}_e$, are nothing more than the treatment means with the grand mean subtracted. This can be shown as follows:

$$\mathbf{T}_e = \mathbf{P}_T\mathbf{R}_G\mathbf{Y} = \mathbf{P}_T(\mathbf{I} - \mathbf{P}_G)\mathbf{Y} = (\mathbf{P}_T - \mathbf{P}_T\mathbf{P}_G)\mathbf{Y} = \mathbf{P}_T\mathbf{Y} - \mathbf{P}_G\mathbf{Y} = \mathbf{T} - \mathbf{G}.$$

Consequently, $\mathbf{T} = \mathbf{G} + \mathbf{T}_e$.

Another expression for D(α)

We have already shown that D(α) = $\mathbf{Y}'\mathbf{R}_T\mathbf{Y}$. However, we have now proved that $\mathbf{R}_T = \mathbf{R}_G\mathbf{R}_T = \mathbf{R}_T\mathbf{R}_G$. Hence, D($\alpha$) = $\mathbf{Y}'\mathbf{R}_T\mathbf{Y} = \mathbf{Y}'\mathbf{R}_T\mathbf{R}_G\mathbf{Y}$. The import of this result is that it tells us that if we already have computed $\mathbf{R}_G\mathbf{Y}$ then can apply $\mathbf{R}_T$ to $\mathbf{R}_G\mathbf{Y}$ to obtain the quantities that have to be squared and summed to obtain D(α).

Alternative methods of computing the sums of squares

In summary, the estimators for the sums of squares have been shown to be

$$\begin{aligned} D(\mu) &= (\mathbf{Y} - \mathbf{G})'(\mathbf{Y} - \mathbf{G}) = \mathbf{Y}'\mathbf{R}_G\mathbf{Y} \\ R(\alpha|\,\mu) &= \mathbf{T}_e'\mathbf{T}_e = \mathbf{Y}'\mathbf{P}_T\mathbf{R}_G\mathbf{Y} \\ D(\alpha) &= (\mathbf{Y} - \mathbf{T})'(\mathbf{Y} - \mathbf{T}) = \mathbf{Y}'\mathbf{R}_T\mathbf{Y} = \mathbf{Y}'\mathbf{R}_T\mathbf{R}_G\mathbf{Y} \end{aligned}$$

Regression analysis could be used to compute the sums of squares as follows:

1. solve the normal equations for the model $E[\mathbf{Y}] = \mathbf{X}_G\mu$ and use the fitted values for this model to obtain D(μ);

2. solve the normal equations for the model $E[\mathbf{Y}] = \mathbf{X}_T\boldsymbol{\alpha}$ and use the fitted values for this model to obtain D(α);
3. obtain $R(\alpha \mid \mu)$ by difference as it is D(μ) –D(α).

However, the above expressions also suggest that, rather than forming the normal equations and solving them by inverting matrices, the sums of squares could be computed by taking sums of squares of elements of appropriate vectors where these elements are obtained using the simple operation of taking means. Call this a **mean-operator analysis**.

The following steps could be used to form the vectors for the mean-operator analysis:

1. $\mathbf{P}_G$ is applied to $\mathbf{y}$ to form $\mathbf{P}_G\mathbf{y} = \mathbf{g}$;
   $\mathbf{g}$ subtracted from $\mathbf{y}$ to form $\mathbf{R}_G\mathbf{y} = \mathbf{y} - \mathbf{g} = \mathbf{e}_G$;
2. $\mathbf{P}_T$ is applied to $\mathbf{R}_G\mathbf{y}$ to form $\mathbf{P}_T\mathbf{R}_G\mathbf{y} = \mathbf{t}_e$;
   $\mathbf{t}_e$ is subtracted from $\mathbf{e}_G$ to form $\mathbf{R}_T\mathbf{R}_G\mathbf{y} = \mathbf{e}_T$.

Note the recursive nature of this procedure: the same operations are performed in both steps and in the second step, the operations are performed on the results of the first step. In fact, a general recursive procedure can be defined as follows.

For each step there is an input vector, a mean operator (**P**) and an output vector (**e**) — the input vector for a step is the output vector from the previous step, except for the first step where the input vector is **y**. The procedure for each step is as follows:

- a vector of means is formed by applying the **P** operator for the step to the input vector
- a sweep is performed in which the vector of means is subtracted (swept) from the input vector to form the output vector

Given the vectors formed above, the sums of squares for the analysis of variance are then the sums of squares of the following vectors:

$$\mathbf{e}_G = \mathbf{R}_G\mathbf{y},\ \mathbf{t}_e = \mathbf{P}_T\mathbf{R}_G\mathbf{y} \text{ and } \mathbf{e}_T = \mathbf{R}_T\mathbf{y} = \mathbf{R}_T\mathbf{R}_G\mathbf{y}$$

That is, the sums of squares are :

$$\begin{aligned} \mathrm{D}(\mu) &= \mathbf{e}_G'\mathbf{e}_G \\ \mathrm{D}(\alpha) &= \mathbf{e}_T'\mathbf{e}_T \\ R(\alpha \mid \mu) &= \mathbf{t}_e'\mathbf{t}_e \end{aligned}$$

b) Degrees of freedom

The following theorem establishes that the degrees of freedom of the sums of squares are as given in the analysis of variance table.

Theorem III.5: Let $D(\mu) = \mathbf{Y}'\mathbf{R}_G\mathbf{Y}$, $R(\alpha|\mu) = \mathbf{Y}'\mathbf{P}_T\mathbf{R}_G\mathbf{Y}$ and $D(\alpha) = \mathbf{Y}'\mathbf{R}_T\mathbf{Y}$ where $\mathbf{R}_G$ and $\mathbf{R}_T$ are as defined in lemma III.1. The degrees of freedom of $D(\mu)$, $R(\alpha|\mu)$ and $D(\alpha)$ are $n-1$, $t-1$ and $n-t$, respectively, where n is the number of observations and t is the number of treatments.

Proof: The matrices $\mathbf{R}_G$, $\mathbf{P}_T\mathbf{R}_G$ and $\mathbf{R}_T$ are symmetric and idempotent so that the ranks of these matrices are equal to their trace.

First, $trace(\mathbf{R}_G) = trace(\mathbf{I}_n - \mathbf{P}_G) = n - trace(n^{-1}\mathbf{J}_n) = n - n^{-1} \times n = n - 1.$

Second, $trace(\mathbf{R}_T) = trace(\mathbf{I}_n - \mathbf{P}_T) = n - trace(\mathbf{P}_T)$.

The general form of $\mathbf{P}_T$ is:

$$\mathbf{P}_T = \begin{bmatrix} n_1^{-1}\mathbf{J}_{n_1} & \mathbf{0}_{n_1 \times n_2} & \cdots & \mathbf{0}_{n_1 \times n_t} \\ \mathbf{0}_{n_2 \times n_1} & n_2^{-1}\mathbf{J}_{n_2} & \cdots & \mathbf{0}_{n_2 \times n_t} \\ \vdots & \vdots & & \vdots \\ \mathbf{0}_{n_t \times n_1} & \mathbf{0}_{n_t \times n_2} & \cdots & n_t^{-1}\mathbf{J}_{n_t} \end{bmatrix}$$

so that $trace(\mathbf{P}_T) = \sum_{k=1}^{t} trace(n_k^{-1}\mathbf{J}_{n_k}) = \sum_{k=1}^{t} 1 = t$

and $trace(\mathbf{R}_T) = n - trace(\mathbf{P}_T) = n - t$.

Finally, from the proof of theorem III.4 we know that $\mathbf{P}_T\mathbf{R}_G = \mathbf{R}_G - \mathbf{R}_T$, so that

$$trace(\mathbf{P}_T\mathbf{R}_G) = trace(\mathbf{R}_G - \mathbf{R}_T) = (n-1) - (n-t) = t - 1$$

■

c) Expected mean squares

To justify our choice of test statistic, we want to work out the expected values of the mean squares in the analysis of variance table under the maximal model:

$$E[\mathbf{Y}] = \mathbf{X}_T\alpha \text{ and } \mathbf{V} = \sigma^2\mathbf{I}$$

and the minimal model:

$$E[\mathbf{Y}] = \mathbf{X}_G\mu \text{ and } \mathbf{V}_Y = \sigma^2\mathbf{I}_n$$

In doing this, the following lemma will be found to be useful.

Lemma III.2: Let $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ be the projection matrix for the $n \times q$ matrix $\mathbf{X}$ of rank q and $\mathbf{R} = \mathbf{I} - \mathbf{P}$. Then $\mathbf{PX} = \mathbf{X}$ and $\mathbf{RX} = \mathbf{0}$.

Proof: Left as an exercise for you ■

Expected mean squares under maximal model

Theorem III.6: Let $\boldsymbol{\psi} = E[\mathbf{Y}] = \mathbf{X}_T\alpha$, $\mathbf{V}_Y = \sigma^2\mathbf{I}_n$, $R(\alpha|\,\mu) = \mathbf{Y}'\mathbf{P}_T\mathbf{R}_G\mathbf{Y}$ and $D(\alpha) = \mathbf{Y}'\mathbf{R}_T\mathbf{Y}$ where $\mathbf{R}_G$ and $\mathbf{R}_T$ are as defined in lemma III.1. Then,

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

where $f_T(\boldsymbol{\psi}) = \sum_{k=1}^{t} n_k(\alpha_k - \bar{\alpha}_.)^2 \Big/ (t-1)$, $\bar{\alpha}_. = \sum_{k=1}^{t}\alpha_k/t$, α_k is the kth element of the t-vector $\boldsymbol{\alpha}$, n_k is the number of observations for treatment k, t is the number of treatments and n is the number of observations.

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

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

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

Also,

$$\begin{aligned} (\mathbf{X}_T\boldsymbol{\alpha})'\mathbf{P}_T\mathbf{R}_G(\mathbf{X}_T\boldsymbol{\alpha}) &= (\mathbf{X}_T\boldsymbol{\alpha})'\mathbf{P}_T\mathbf{R}_G\mathbf{P}_T\mathbf{R}_G(\mathbf{X}_T\boldsymbol{\alpha}) \quad \text{as } \mathbf{P}_T\mathbf{R}_G \text{ is idempotent} \\ &= (\mathbf{P}_T\mathbf{R}_G\,\mathbf{X}_T\boldsymbol{\alpha})'\mathbf{P}_T\mathbf{R}_G(\mathbf{X}_T\boldsymbol{\alpha}) \quad \text{as } \mathbf{P}_T\mathbf{R}_G \text{ is symmetric} \end{aligned}$$

But

$$\mathbf{P}_T\mathbf{R}_G(\mathbf{X}_T\boldsymbol{\alpha}) = \mathbf{P}_T(\mathbf{I}_n - \mathbf{P}_G)(\mathbf{X}_T\boldsymbol{\alpha}) = (\mathbf{P}_T - \mathbf{P}_T\mathbf{P}_G)(\mathbf{X}_T\boldsymbol{\alpha}) = (\mathbf{P}_T - \mathbf{P}_G)(\mathbf{X}_T\boldsymbol{\alpha})$$

It can be shown that $\mathbf{P}_G\mathbf{X}_T\boldsymbol{\alpha} = \mathbf{1}\bar{\alpha}_.$ where $\bar{\alpha}_. = \left(\sum_{k=1}^{t} n_k\alpha_k\right)\Big/ n$ while, by lemma III.2, $\mathbf{P}_T\mathbf{X}_T\boldsymbol{\alpha} = \mathbf{X}_T\boldsymbol{\alpha}$

whence

$$\mathbf{P}_T\mathbf{R}_G(\mathbf{X}_T\boldsymbol{\alpha}) = \mathbf{X}_T\boldsymbol{\alpha} - \mathbf{1}\bar{\alpha}_.$$

so that

$$E\left[R(\alpha|\,\mu)/(t-1)\right] = \left\{(\text{t-1})\sigma^2 + (\mathbf{X}_T\boldsymbol{\alpha} - \mathbf{1}\bar{\alpha}_.)'(\mathbf{X}_T\boldsymbol{\alpha} - \mathbf{1}\bar{\alpha}_.)\right\}/\{t-1\}$$
$$= \sigma^2 + \sum_{k=1}^{t} n_k(\alpha_k - \bar{\alpha}_.)^2 \Big/ (t-1)$$
$$= \sigma^2 + f_T(\boldsymbol{\psi})$$

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

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

Now from theorem III.5, $trace(\mathbf{R}_T) = n-t$. Also, by lemma III.2, $\mathbf{P}_T\mathbf{X}_T\boldsymbol{\alpha} = \mathbf{X}_T\boldsymbol{\alpha}$, $\mathbf{R}_T(\mathbf{X}_T\boldsymbol{\alpha}) = \mathbf{0}$. Consequently,

$$E\left[D(\alpha)/(n-t)\right] = \left\{\sigma^2 trace(\mathbf{R}_T) + (\mathbf{X}_T\boldsymbol{\alpha})'\mathbf{R}_T(\mathbf{X}_T\boldsymbol{\alpha})\right\}/\{n-t\}$$
$$= \left\{\sigma^2(n-t) + 0\right\}/\{n-t\}$$
$$= \sigma^2$$

■

Expected mean squares under the minimal model

Theorem III.7: 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,

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

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

Proof: It is left as an exercise for you to demonstrate these results. ■

Analysis of variance incorporating expected mean squares

So the analysis is summarized in the analysis of variance table as follows:

Source	df	SSq	MSq	E[MSq]	F
Units	n–1				
Treatments	t–1	R(α|μ)	R(α|μ)/(t–1)	$\sigma^2 + f_T(\boldsymbol{\psi})$	s_T^2/s_R^2
Residual	n–t	D(α)	D(α)/(n–t)	σ^2	

That is, the Residual mean square has expectation σ^2, where σ^2 is the variation arising from uncontrolled differences (amongst units). The Treatment mean square on the other hand is influenced by uncontrolled variation and the magnitude of

treatment differences. This is not surprising if we consider the differences that are likely to contribute to differences between treatment means.

Example III.1 Rat experiment (continued)

In the liver weight example, the data and treatment means are as follows:

	Diet		
	A	B	C
	3.3	3.2	2.7
	3.1	3.4	
	2.9		
Mean	3.1	3.3	2.7

So what can potentially contribute to the observed difference 3.1 and 2.7? **Answer**: Obviously, the different diets; but what is not so obvious is that differences arising from uncontrolled variation also contribute as two different groups of rats are involved. This is then reflected in the expected means squares in that it involves σ^2 and the "variance" of the 3 effects.

Thus, the F test involves asking the question "Is the variance in the treatment means greater than can be expected from uncontrolled variation alone?".

If the variance is no greater, it is concluded that $f_{\mathrm{T}}(\boldsymbol{\psi})$ is zero and the minimal model is the correct model since the expected Treatment mean square under this model is just σ^2. Otherwise, if the variance is greater, $f_{\mathrm{T}}(\boldsymbol{\psi})$ is nonzero the maximal model is required to describe the data.

d) Distribution of the F statistic

Theorem III.8: Let $\mathbf{Y}$ be a normally-distributed random vector, $E[\mathbf{Y}] = \mathbf{X}_{\mathrm{G}}\mu$, $\mathbf{V}_Y = \sigma^2\mathbf{I}_n$, $s_T^2 = R(\alpha \mid \mu)/(t-1) = \mathbf{Y}'\mathbf{P}_{\mathrm{T}}\mathbf{R}_{\mathrm{G}}\mathbf{Y}/(t-1)$ and $s_R^2 = D(\alpha)/(n-t) = \mathbf{Y}'\mathbf{R}_{\mathrm{T}}\mathbf{Y}/(n-t)$ where $\mathbf{R}_{\mathrm{G}}$ and $\mathbf{R}_{\mathrm{T}}$ are as defined in lemma III.1. Then, the ratio of these two mean squares, given by

$$F_{(t-1),\,(n-t)} = \frac{s_{\mathrm{T}}^2}{s_{\mathrm{R}}^2}$$

is distributed as a Snedecor's F with $(t-1)$ and $(n-t)$ degrees of freedom.

Proof: We have only to show that $E[\mathbf{R}_{\mathrm{T}}\mathbf{Y}] = E[\mathbf{P}_{\mathrm{T}}\mathbf{R}_{\mathrm{G}}\mathbf{Y}] = \mathbf{0}$ and that $\mathbf{Y}'\mathbf{R}_{\mathrm{T}}\mathbf{Y}$ and $\mathbf{Y}'\mathbf{P}_{\mathrm{T}}\mathbf{R}_{\mathrm{G}}\mathbf{Y}$ are independent quadratic forms. Then theorem II.17 can be invoked to obtain the distribution of the test statistic. We leave you to show that $E[\mathbf{R}_{\mathrm{T}}\mathbf{Y}] = E[\mathbf{P}_{\mathrm{T}}\mathbf{R}_{\mathrm{G}}\mathbf{Y}] = \mathbf{0}$, a result also required for theorem III.7

Since the matrices of the two quadratic forms are idempotent, we have only to establish that $\mathbf{R}_T\mathbf{P}_T\mathbf{R}_G = \mathbf{0}$ so that two of the three conditions in theorem II.16 have been satisfied and the two quadratic forms are independent.

Now $\mathbf{R}_T\mathbf{P}_T\mathbf{R}_G = (\mathbf{I} - \mathbf{P}_T)\mathbf{P}_T\mathbf{R}_G = (\mathbf{P}_T - \mathbf{P}_T)\mathbf{R}_G = \mathbf{0}$ and the result follows. ■

e) Analysis of variance table

Gathering together the results from the previous sections, the analysis of variance table for a CRD is:

Source	df	SSq	MSq (s^2)	E[MSq]	F
Units	n–1	$D(\mu) = \mathbf{e}_G'\mathbf{e}_G$			
Treatments	t–1	$R(\alpha\|\mu) = \mathbf{t}_e'\mathbf{t}_e$	$R(\alpha\|\mu)/(t-1)$	$\sigma^2 + f_T(\boldsymbol{\psi})$	s_T^2/s_R^2
Residual	n–t	$D(\alpha) = \mathbf{e}_T'\mathbf{e}_T$	$D(\alpha)/(n-t)$	σ^2	

f) Analysis of the rat example

Example III.1 Rat experiment (continued)

We first carry out the recursive procedure described above.

Step 1: Application of grand mean operator to original data vector

The decomposition of the original data vector is achieved by applying the grand mean operator $\mathbf{P}_G$ to $\mathbf{y}$ to form $\mathbf{P}_G\mathbf{y} = \mathbf{g}$ and then subtracting $\mathbf{g}$ from $\mathbf{y}$ to form $\mathbf{R}_G\mathbf{y} = \mathbf{y} - \mathbf{g} = \mathbf{e}_G$.

Observations $\mathbf{y}$		Grand Mean $\mathbf{g}$		Rat Deviations $\mathbf{y}-\mathbf{g} = \mathbf{e}_G$
$\begin{bmatrix} 3.3 \\ 3.1 \\ 2.9 \\ 3.2 \\ 3.4 \\ 2.7 \end{bmatrix}$	=	$\begin{bmatrix} 3.1 \\ 3.1 \\ 3.1 \\ 3.1 \\ 3.1 \\ 3.1 \end{bmatrix}$	+	$\begin{bmatrix} 0.2 \\ 0.0 \\ -0.2 \\ 0.1 \\ 0.3 \\ -0.4 \end{bmatrix}$

Step 2: Application of treatment mean operator to Rat deviations vector

The decomposition of the Rats deviation vector is achieved by applying the treatment mean operator $\mathbf{P}_T$ to $\mathbf{e}_G = \mathbf{R}_G\mathbf{y}$ to form $\mathbf{P}_T\mathbf{R}_G\mathbf{y} = \mathbf{t}_e$ and then subtracting $\mathbf{t}_e$ from $\mathbf{e}_G$ to form $\mathbf{R}_T\mathbf{R}_G\mathbf{y} = \mathbf{e}_G - \mathbf{t}_e = \mathbf{e}_T$.

Rat Deviations $\mathbf{e}_G$		Diet Effects $\mathbf{t}_e$		Residual Deviations $\mathbf{e}_T$
$\begin{bmatrix} 0.2 \\ 0.0 \\ -0.2 \\ 0.1 \\ 0.3 \\ -0.4 \end{bmatrix}$	=	$\begin{bmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 0.2 \\ 0.2 \\ -0.4 \end{bmatrix}$	+	$\begin{bmatrix} 0.2 \\ 0.0 \\ -0.2 \\ -0.1 \\ 0.1 \\ 0.0 \end{bmatrix}$

In summary, we have decomposed the data vector into three component vectors as follows:

Obser-vations		Grand Mean		Diet Deviations		Residual Deviations
$\mathbf{y}$	=	$\mathbf{g}$	+	$\mathbf{t}_e$	+	$\mathbf{e}_T$
$\begin{bmatrix} 3.3 \\ 3.1 \\ 2.9 \\ 3.2 \\ 3.4 \\ 2.7 \end{bmatrix}$	=	$\begin{bmatrix} 3.1 \\ 3.1 \\ 3.1 \\ 3.1 \\ 3.1 \\ 3.1 \end{bmatrix}$	+	$\begin{bmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 0.2 \\ 0.2 \\ -0.4 \end{bmatrix}$	+	$\begin{bmatrix} 0.2 \\ 0.0 \\ -0.2 \\ -0.1 \\ 0.1 \\ 0.0 \end{bmatrix}$

Note that all vectors of deviations and effects sum to zero. Also, for the vector of Residual deviations, the sum of deviations for each treatment is zero.

The computations of the sums of squares from these vectors are outlined in the following table.

Diet	$\mathbf{y}$	$\mathbf{g}$	$\mathbf{e}_G$	$\mathbf{t}_e$	$\mathbf{e}_T$
A	3.3	3.1	0.2	0.0	0.2
A	3.1	3.1	0.0	0.0	0.0
A	2.9	3.1	–0.2	0.0	–0.2
B	3.2	3.1	0.1	0.2	–0.1
B	3.4	3.1	0.3	0.2	0.1
C	2.7	3.1	–0.4	–0.4	0.0
SSq			0.34	0.24	0.10

That is, $D(\mu) = 0.34$, $R(\alpha|\mu) = 0.24$ and $D(\alpha) = 0.10$.

We now perform the hypothesis test for the example.

Step 1: Set up hypotheses

H_0: $\alpha_A = \alpha_B = \alpha_C = \mu$ $(\boldsymbol{\psi} = \mathbf{X}_G\mu)$

H_1: at least one pair of population diet means is different $(\boldsymbol{\psi} = \mathbf{X}_T\boldsymbol{\alpha})$

Step 2: Calculate test statistic

Source	df	SSq	MSq	E[MSq]	F	Prob
Rats	5	0.34				
Diets	2	0.24	0.1200	$\sigma^2 + f_T(\boldsymbol{\psi})$	3.60	0.1595
Residual	3	0.10	0.0332	σ^2		

From this table we can see that we have taken the total variation amongst the six Rats and partitioned it into two parts: variance of difference between diet means and the left-over (residual) rat variation.

Step 3: Decide between hypotheses

As the probability of exceeding an F of 3.60 with $\nu_1 = 2$ and $\nu_2 = 3$ is 0.1595, there is not much evidence of a diet difference.

III.D Computation in Genstat

a) Regression analysis

Example III.1 Rat experiment (continued)

The data is entered into a Genstat spreadsheet with *Diet* set up as a factor with 3 levels and 6 values and *LiverWt* set up as a variate with 6 values. Then the following commands are entered to perform the regression analysis.

```
PRINT Diet,LiverWt
MODEL LiverWt
TERMS Diet
FIT [FPROB=y] Diet
```

```
17  PRINT Diet,LiverWt

        Diet     LiverWt
           1       3.300
           1       3.100
           1       2.900
           2       3.200
           2       3.400
           3       2.700

  18  MODEL LiverWt
  19  TERMS Diet
  20  FIT [FPROB=y] Diet
```

```
20.....................................................................


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

 Response variate: LiverWt
     Fitted terms: Constant, Diet


*** Summary of analysis ***

             d.f.         s.s.         m.s.      v.r.  F pr.
Regression      2       0.2400      0.12000      3.60  0.160
Residual        3       0.1000      0.03333
Total           5       0.3400      0.06800

Percentage variance accounted for 51.0
Standard error of observations is estimated to be 0.183
* MESSAGE: The following units have high leverage:
         Unit     Response    Leverage
            6        2.700        1.00


*** Estimates of parameters ***

                  estimate         s.e.      t(3)
Constant             3.100        0.105     29.41
Diet 2               0.200        0.167      1.20
Diet 3              -0.400        0.211     -1.90
```

Notice that the estimates from the Regression analysis are for the Constant, Diet 2 and Diet 3. This is because Genstat fits the following model :

$$\begin{bmatrix} E[Y_1] \\ E[Y_2] \\ E[Y_3] \\ E[Y_4] \\ E[Y_5] \\ E[Y_6] \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}$$

That is, the first column of our $\mathbf{X}_T$ is replaced with the column of $\mathbf{X}_G$. You should be able to show that for this experiment

$$\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \end{bmatrix} = \begin{bmatrix} \dfrac{Y_1 + Y_2 + Y_3}{3} \\ \dfrac{Y_4 + Y_5}{2} - \dfrac{Y_1 + Y_2 + Y_3}{3} \\ \dfrac{Y_6}{1} - \dfrac{Y_1 + Y_2 + Y_3}{3} \end{bmatrix} = \begin{bmatrix} T_A \\ T_B - T_A \\ T_C - T_A \end{bmatrix}$$ where T_A, T_B and T_C are the treatment means, and that $$\mathbf{X}\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} T_A \\ T_B - T_A \\ T_C - T_A \end{bmatrix} = \begin{bmatrix} T_A \\ T_A \\ T_A \\ T_B \\ T_B \\ T_C \end{bmatrix}$$

That is, the fitted values are precisely those obtained with our $\mathbf{X}_T$ matrix. It turns out that if we take any **X** matrix that consists of 3 linearly independent columns that are linear combinations of the columns of our $\mathbf{X}_T$ matrix, the fitted values will be the same for all of them. What will differ between different **X** matrices are the estimators which will have different interpretations.

The interpretation in this case is that the first parameter estimates the first treatment mean and that the remaining parameters estimate the differences between the other treatment means and the first treatment mean. That you cannot see this by examining the columns of the **X** matrix is because the columns of the **X** matrix are not orthogonal.

Definition III.5: Two columns of a matrix **X**, say $\mathbf{x}_i$ and $\mathbf{x}_j$, are said to be **orthogonal** if $\mathbf{x}_i'\mathbf{x}_j = 0$. ■

b) Mean-operator Analysis of Rat data

Example III.1 Rat experiment (continued)

The factor *Rats* with 6 levels and 6 values is added to the Genstat spreadsheet and then the following commands are used to perform the mean-operator analysis of the data.

```
PRINT Rats,Diet,LiverWt          "Check that factors and data correct"
BLOCKSTRUCTURE Rats               "Supply structure set for experiment"
TREATMENTSTRUCTURE Diet
ANOVA [FPROB=Y; PSE=LSD] LiverWt "Perform analysis"

15  PRINT Rats,Diet,LiverWt         "Check that factors and data correct"

        Rats        Diet     LiverWt
           1           1       3.300
           2           1       3.100
           3           1       2.900
           4           2       3.200
           5           2       3.400
           6           3       2.700

  16  BLOCKSTRUCTURE Rats             "Supply structure set for experiment"
  17  TREATMENTSTRUCTURE Diet
  18  ANOVA [FPROB=Y; PSE=LSD]LiverWt "Perform analysis"


18.....................................................................

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

Variate: LiverWt

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

Rats stratum
Diet                       2    0.24000    0.12000    3.60  0.160
Residual                   3    0.10000    0.03333
Total                      5    0.34000
```

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

Variate: LiverWt

Grand mean  3.100

     Diet        1        2        3
             3.100    3.300    2.700
     rep.        3        2        1


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

Table                Diet
rep.               unequal
d.f.                     3
l.s.d.             0.8217X min.rep
                   0.6709  max-min
                   0.4744X max.rep

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

III.E Diagnostic checking

(Box, Hunter and Hunter sec.6.5)

In performing the analysis outlined last lecture we assumed the data followed a certain model, namely, that $\mathbf{Y}$ is distributed N($\boldsymbol{\psi}$, $\sigma^2\mathbf{I}$) where $\boldsymbol{\psi} = E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\theta}$ and σ^2 is the variance of an observation. The expression for the expectation that is of most interest is $E[\mathbf{Y}] = \mathbf{X}_T\boldsymbol{\alpha}$. For the complete model to be appropriate requires that:

a) the response is operating additively, that is, that the treatments have about the same additive effect on each unit;
b) that the sets of units assigned the treatments are comparable in that the amount of uncontrolled variation exhibited by them is the same for each treatment;
c) each observation is independent of other observations; and
d) that the response of the units is normally distributed.

Now, it may be that the data do not conform to the assumptions inherent in the model and that the conclusions we have drawn based on the model are incorrect. Thus, it would seem highly desirable to check the adequacy of our model in describing the data.

Example III.2 Caffeine effects on students

To illustrate the procedures, I am going to use an experiment in which the effect of orally ingested caffeine on a physical task was investigated (Draper and Smith, 1981, sec.9.1). Thirty healthy male college students were selected and trained in finger tapping. Ten men were randomly assigned to receive one of three doses of caffeine (0, 100 or 200 mg). The number of finger taps after ingesting the caffeine was recorded for each student and the data were as follows:

Caffeine Dose (mg)		
0	100	200
242	248	246
245	246	248
244	245	250
248	247	252
247	248	248
248	250	250
242	247	246
244	246	248
246	243	245
242	244	250

The first task is to set up a spreadsheet containing the data. This spreadsheet needs to contain a factor Students with values 1–30, a factor Dose with levels 0, 100 and 200 and values depending on whether the data is entered by rows or columns, and a variate Taps with the 30 observed values of the response variable. The following steps can be used to enter the data by rows:

1. Use *Stats > Design > Generate Factors in Standard Order*, with the *Display Factors in Spreadsheet* option set, to generate the factor Students with values 1–30.
2. Use *Spread > Insert > Column after Current Column* to add the factor Dose with levels 0, 100 and 200. Then use *Spread > Column > Fill*, with *Number of Repeats* set to 1 and the *Copy down existing values over missing* option unchecked, to set the values of Dose so that the 3 levels values are recycled 10 times.
3. Use *Spread > Insert > Column after Current Column* to add the variate Taps to the spreadsheet. Enter the 30 values of Taps.

Of course, if you preferred, you could omit the first two steps and set up a spreadsheet with 3 columns by 30 rows and enter all the values manually via the keyboard. Whichever way you choose, the data in the spreadsheet should end up as illustrated in the following output.

```
Genstat 5  Release 4.1  (PC/Windows NT)              18 March 2000 23:47:11
Copyright 1998, Lawes Agricultural Trust (Rothamsted Experimental Station)

                ____________________________________________________

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

  3  FACTOR [MODIFY=yes; LEVELS=30; NVAL=30] Students
  4  GENERATE Students
  5  %WSPREAD Students
  6  DELETE [redefine=yes] Dose
  7  FACTOR [modify=yes;nvalues=30;levels=!(0,100,200)] Dose
  8  READ Dose; frepresentation=ordinal

 Identifier    Values   Missing    Levels
       Dose        30         0         3
```

```
 10
 11
 12  "Data taken from File: D:/ANALYSES/LM/ONEFAC/CRDCAFF.GSH"
 13  DELETE [redefine=yes] Taps
 14  VARIATE [nvalues=30] Taps
 15  READ Taps

  Identifier   Minimum      Mean   Maximum    Values   Missing
        Taps     242.0     246.5     252.0        30         0

 18
 19  PRINT Students,Dose,Taps

  Students        Dose        Taps
         1         0.0       242.0
         2       100.0       248.0
         3       200.0       246.0
         4         0.0       245.0
         5       100.0       246.0
         6       200.0       248.0
         7         0.0       244.0
         8       100.0       245.0
         9       200.0       250.0
        10         0.0       248.0
        11       100.0       247.0
        12       200.0       252.0
        13         0.0       247.0
        14       100.0       248.0
        15       200.0       248.0
        16         0.0       248.0
        17       100.0       250.0
        18       200.0       250.0
        19         0.0       242.0
        20       100.0       247.0
        21       200.0       246.0
        22         0.0       244.0
        23       100.0       246.0
        24       200.0       248.0
        25         0.0       246.0
        26       100.0       243.0
        27       200.0       245.0
        28         0.0       242.0
        29       100.0       244.0
        30       200.0       250.0
```

Here is the mean-operator analysis for this data:

```
 20  "
-21  **** Mean-operator Analysis ****
-22  "
 23  BLOCK Students
 24  TREAT Dose
 25  ANOVA [FPROB=Y; PSE=LSD] Taps


25.....................................................................

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

Variate: Taps

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

Students stratum
Dose                       2     61.400     30.700    6.18  0.006
Residual                  27    134.100      4.967
Total                     29    195.500
```

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

Variate: Taps

Grand mean  246.50

     Dose     0.00   100.00   200.00
            244.80   246.40   248.30


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

Table                 Dose
rep.                    10
d.f.                    27
l.s.d.               2.045
```

The analysis of this data is as follows:

Step 1: Set up hypotheses

H_0: $\alpha_0 = \alpha_{100} = \alpha_{200} = \mu$
H_1: at least one pair of population dose means is different

Step 2: Calculate test statistic

The analysis of variance table for the example is:

Source	df	SSq	MSq	F	Prob
Students	29	195.50			
Doses	2	61.40	30.70	6.18	0.006
Residual	27	134.10	4.97		

Step 3: Decide between hypotheses

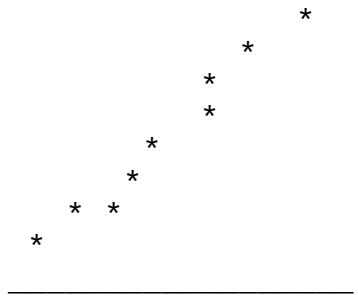
$P(F_{2,27} \geq 6.18) = 0.006$.

The evidence suggests that there is a dose difference.

To look at how well the data conform to the assumed model one examines the residuals, that is the vector $\mathbf{e}_T$. We shall use the *Residuals-versus-fitted-values plot* and the *Normal Probability plot*.

In using these plots one should take into account that the analysis of variance is robust to moderate departures from normality so that there would have to be evidence of marked nonnormality before remedial action would be required. The most common form of discrepancy revealed by such a plot is that there is an unusually large or small residual. The cause of such extreme values requires investigation. It may be a mistake in recording or entering the data or a catastrophe affecting that unit can be identified. In the absence of an explanation, serious consideration of the possibility that the result is valid and that it is the result of some unanticipated, but important effect.

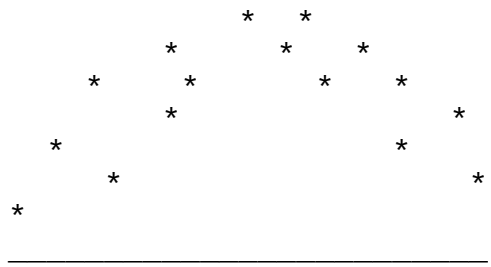
The *Normal Probability plot* should show a broadly straight-line trend.



The *Residuals-versus-fitted-values plot* involves plotting $\mathbf{e} = \mathbf{Y} - \mathbf{X}\hat{\theta}$ against $\mathbf{X}\hat{\theta}$. Generally, the points on the scatter diagram should be spread across plot evenly, that is, displaying no particular pattern.

no particular pattern

Patterns such as shown below indicate problems:



systematic trend in residuals

variance increases as level increases

variance peaks in middle

assumptions. However, for violations to be detected they must affect the treatments differently — if they are not treatment related they may not show up in the plot.

The Genstat procedure APLOT is used to produce these two plots:

```
APLOT METHOD=fit,normal
```

Example III.2 Caffeine effects on students (continued)

For this example, a violation of the assumption would occur if all the students were in the same room and the presence of other students caused anxiety to just the students that had no caffeine. That is, the response of the students is not independent. It may be that the inhibition of this group resulted in less variation in their response which would be manifest in the plot. Another situation that would lead to an unacceptable pattern in the plot is if the effect becomes more variable as the level of the response variable increases. For example, caffeine increases the tapping but at higher levels the variability of increase from student to student is greater. That is there is a lack of unit-treatment additivity.

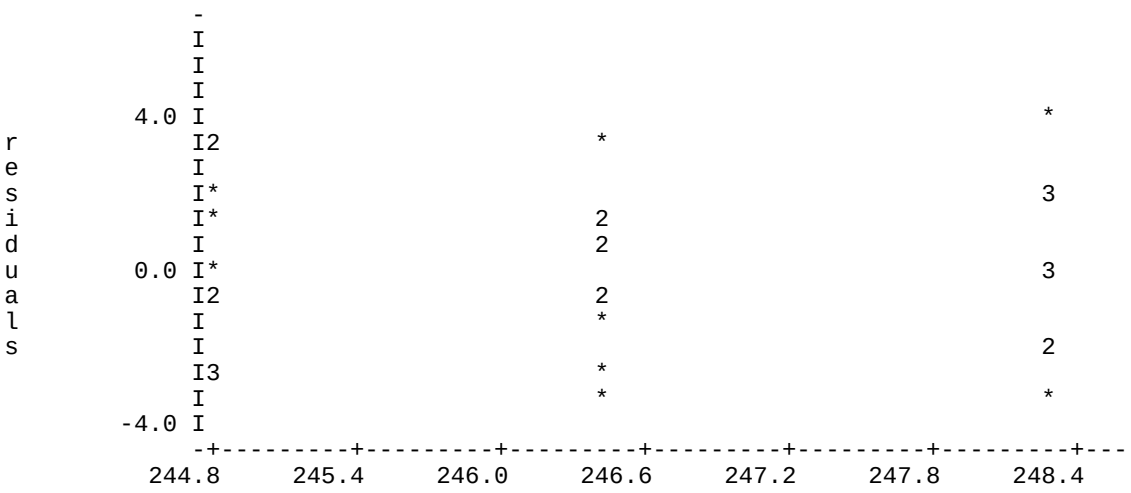
For this example, we first obtain the fitted values and residuals using *AKEEP* so that they can be printed out. In general, this is not required.

```
26  AKEEP [FIT=Fit; RES=Res]
27  PRINT Students,Dose,Taps,Fit,Res
```

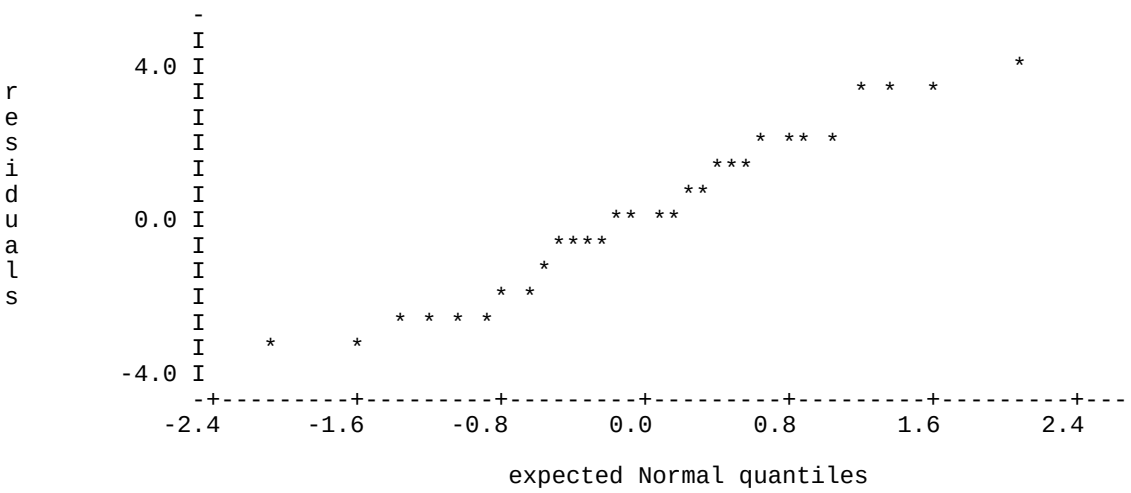
Students	Dose	Taps	Fit	Res
1	0.0	242.0	244.8	-2.800
2	100.0	248.0	246.4	1.600
3	200.0	246.0	248.3	-2.300
4	0.0	245.0	244.8	0.200
5	100.0	246.0	246.4	-0.400
6	200.0	248.0	248.3	-0.300
7	0.0	244.0	244.8	-0.800
8	100.0	245.0	246.4	-1.400
9	200.0	250.0	248.3	1.700
10	0.0	248.0	244.8	3.200
11	100.0	247.0	246.4	0.600
12	200.0	252.0	248.3	3.700
13	0.0	247.0	244.8	2.200
14	100.0	248.0	246.4	1.600
15	200.0	248.0	248.3	-0.300
16	0.0	248.0	244.8	3.200
17	100.0	250.0	246.4	3.600
18	200.0	250.0	248.3	1.700
19	0.0	242.0	244.8	-2.800
20	100.0	247.0	246.4	0.600
21	200.0	246.0	248.3	-2.300
22	0.0	244.0	244.8	-0.800
23	100.0	246.0	246.4	-0.400
24	200.0	248.0	248.3	-0.300
25	0.0	246.0	244.8	1.200
26	100.0	243.0	246.4	-3.400
27	200.0	245.0	248.3	-3.300
28	0.0	242.0	244.8	-2.800
29	100.0	244.0	246.4	-2.400
30	200.0	250.0	248.3	1.700

Here are the plots for the example:

```
  28  APLOT METHOD=fit,normal
```



Normal plot



The *Residuals-versus-fitted-values plot* appears to be fine and the *Normal Probability plot* also appears to be displaying the appropriate pattern

Genstat program to produce analysis

```
"**** Mean-operator Analysis ****"

BLOCK Students
TREAT Dose
ANOVA [FPROB=Y; PSE=LSD] Taps
APLOT METHOD=fit,normal
```

III.F Treatment differences

So far all that our analysis has accomplished is that we have decided whether or not there appears to be a difference between the population treatment means. Of greater interest to the researcher is which of the treatment means is different. We look at addressing that question now.

a) Multiple comparisons procedures
(Box, Hunter and Hunter Appendix 6C)

These procedures are used for investigating all possible pairs of differences between the means in the experiment. There are several procedures available. Probably the most commonly used is the LSD procedure — we will look at it and Tukey's procedures..

Example III.2 Caffeine effects on students (continued)

In this experiment we have already concluded that the evidence suggests that there is a dose difference. But which doses are different?

The part of the Genstat ANOVA output that is relevant is:

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

     Variate: Taps

     Grand mean  246.50

          Dose     0.00   100.00   200.00
                 244.80   246.40   248.30


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

     Table                 Dose
     rep.                    10
     d.f.                    27
     l.s.d.               2.045
```

LSD procedure

LSD stands for least significant difference and it is equivalent to determining, for every pair of treatments, whether or not the confidence limits on the differences include 0.

Each application of the procedure is based on the hypotheses:

H_o: $\alpha_A = \alpha_B$
H_1: $\alpha_A \neq \alpha_B$

One calculates the statistic

$$LSD(\alpha\%) = t_{\upsilon,\alpha}\, s_{\bar{x}_d}$$

where ν is the Residual degrees of freedom,

$t_{\nu,\alpha}$ is the value of t such that $P(t \le -t_{\nu,\alpha}) + P(t \ge t_{\nu,\alpha}) = \alpha$,

$s_{\bar{x}_d}$ is the standard error of the difference (s.e.d. in Genstat)

$$= \sqrt{\text{Residual MSq}\left(1/n_A + 1/n_B\right)}$$

n_A, n_B are the number of replicates for each of a pair of means being compared.

Note: Because of the inclusion of n_A and n_B in the formula, the LSD will depend on which means are being compared if the numbers of replicates for the treatment means is unequal.

If the treatments are all equally replicated, the formula for $s_{\bar{x}_d}$ reduces to

$$s_{\bar{x}_d} = \sqrt{\text{Residual MSq}\left(2/r\right)}$$

where r is the number of replicates in a treatment mean.

To obtain $t_{\nu,\alpha}$ use EDT (**E**quivalent **D**eviate of **T**) function of Genstat as follows:

CALC t=EDT(1 – α/2; ν)

Example III.2 Caffeine effects on students (continued)

For our caffeine example, we used $\alpha = 0.05$ so use the following instructions to obtain $t_{\nu, 0.05}$:

```
CALC t=EDT(0.975; 27)
PRINT t
```

This yields the value of t = 2.052 so that

$$LSD(5\%) = 2.052\sqrt{4.97 \times \frac{2}{10}}$$
$$= 2.05$$

The value of the LSD is also given below the table of means in the Genstat output. It is 2.045.

Next we examine all pairs of treatment mean differences. This most easily accomplished in a table with both rows and columns corresponding to the Treatments — the rows and columns are ordered in ascending order of the values for the means. The body of this table should contain the differences between row means and column means — in particular, the row mean minus the column mean for entries below the diagonal. We can then determine which pairs of means are different by comparing the differences with the LSD(100α%)

Differences between all pairs of Dose means

Dose		0	100	200
	Mean	244.8	246.4	248.3
0	244.8			
100	246.4	1.6		
200	248.3	3.5	1.9	
	LSD(5%)		2.05	

Any two means different by more than 2.05 are significantly different. Our conclusion is that the mean for 0 and 200 are different but that for 100 is somewhat intermediate.

A problem with the LSD procedure is that one has a probability, for each pair of means, of 0.05 of declaring a pair of means to be significant when they are not. This probability is independent for each pair of means and so increases as the number of pairs of means to be compared increases. This is, of course, related to the number of means. It can be shown that the following table applies:

No. means	No. Compared	Probability
5	10	0.29
10	45	0.63
15	105	0.83

Thus the procedure should only be applied when there are a few treatments to be compared.

Tukey's procedure

This procedure takes into account that all pairwise comparisons are being made with the result that the probability of a Type I error over all pairwise comparisons is 0.05. It is basically a technique for equal numbers of observations for each mean. However, it can be used if the numbers of observations in the different means are not too unequal.

One has to calculate the statistic

$$w(\alpha\%) = \frac{q_{t,\upsilon,\alpha}}{\sqrt{2}}\, s_{\bar{x}_d}$$

where $q_{t,\nu,\alpha}$ is the studentized range obtained using the Genstat procedure STUDENTRANGE with $t =$ number of means, $\nu =$ Residual df and $\alpha =$ significance level,

and $s_{\bar{x}_d}$ is the standard error of the difference for the case in which the numbers of observations for the means are equal (s.e.d. in Genstat). When the numbers of observations are not equal replace the number of replicates by the harmonic mean of the numbers of replicates. That is,

$$H = \frac{1}{\dfrac{\sum_{k=1}^{t} \dfrac{1}{n_k}}{t}}$$

Example III.2 Caffeine effects on students (continued)

For the example, $q_{3,27,0.05} = 3.508$ and so:

$$w(5\%) = \frac{3.508}{\sqrt{2}} \times 0.997$$
$$= 2.47$$

Using this value we come to the same conclusions as for the LSD. This will not always be the case as the values for Tukey's procedure is larger than that for the LSD.

Differences between all pairs of Dose means

Dose		0	100	200
	Mean	244.8	246.4	248.3
0	244.8			
100	246.4	1.6		
200	248.3	3.5	1.9	
	w(5%)		2.47	

Table of studentized range values for $\alpha = 0.05$

DF of s	Number of means (t)									
	3	4	5	6	7	8	9	10	11	12
1	27.066	32.920	37.147	40.481	43.202	45.471	47.459	49.203	50.767	52.151
2	8.342	9.813	10.890	11.744	12.444	13.039	13.551	14.003	14.407	14.761
3	5.914	6.829	7.504	8.039	8.480	8.855	9.180	9.465	9.721	9.948
4	5.044	5.760	6.290	6.709	7.055	7.349	7.605	7.829	8.031	8.212
5	4.605	5.221	5.676	6.035	6.332	6.585	6.804	6.997	7.171	7.325
6	4.342	4.898	5.307	5.630	5.897	6.124	6.321	6.495	6.651	6.791
7	4.168	4.684	5.063	5.361	5.607	5.817	5.999	6.160	6.304	6.433
8	4.043	4.531	4.888	5.169	5.400	5.598	5.769	5.920	6.055	6.177
9	3.951	4.418	4.757	5.025	5.246	5.433	5.596	5.740	5.868	5.985
10	3.879	4.328	4.655	4.913	5.126	5.306	5.462	5.600	5.723	5.835
11	3.822	4.258	4.575	4.824	5.030	5.203	5.354	5.488	5.607	5.715
12	3.775	4.200	4.509	4.751	4.951	5.120	5.266	5.396	5.512	5.616
13	3.736	4.152	4.454	4.691	4.886	5.050	5.193	5.319	5.432	5.534
14	3.703	4.112	4.408	4.640	4.832	4.992	5.131	5.255	5.365	5.464
15	3.675	4.077	4.368	4.596	4.783	4.942	5.078	5.199	5.307	5.405
16	3.651	4.047	4.334	4.558	4.742	4.898	5.032	5.151	5.257	5.353
17	3.630	4.021	4.304	4.525	4.706	4.859	4.992	5.109	5.213	5.308
18	3.611	3.998	4.277	4.496	4.675	4.826	4.956	5.071	5.175	5.268
19	3.594	3.978	4.254	4.470	4.647	4.796	4.925	5.038	5.140	5.232
20	3.579	3.960	4.233	4.446	4.621	4.769	4.897	5.009	5.109	5.200
21	3.566	3.943	4.214	4.426	4.599	4.745	4.871	4.982	5.082	5.171
22	3.554	3.928	4.197	4.407	4.578	4.723	4.848	4.958	5.056	5.145
23	3.543	3.915	4.182	4.389	4.560	4.703	4.827	4.936	5.034	5.122
24	3.533	3.902	4.168	4.374	4.543	4.685	4.808	4.916	5.013	5.100
25	3.524	3.891	4.155	4.359	4.527	4.668	4.791	4.898	4.994	5.080
26	3.515	3.881	4.143	4.346	4.513	4.653	4.774	4.881	4.976	5.062
27	3.508	3.872	4.132	4.334	4.499	4.639	4.760	4.865	4.960	5.045
28	3.500	3.863	4.121	4.323	4.487	4.626	4.746	4.851	4.945	5.030
29	3.494	3.854	4.112	4.312	4.476	4.614	4.733	4.838	4.931	5.015
30	3.488	3.847	4.103	4.303	4.465	4.603	4.721	4.825	4.918	5.002
31	3.482	3.840	4.095	4.293	4.455	4.592	4.710	4.814	4.906	4.989
32	3.476	3.833	4.087	4.285	4.446	4.582	4.700	4.803	4.895	4.977
33	3.471	3.827	4.080	4.277	4.438	4.573	4.690	4.793	4.884	4.966
34	3.467	3.821	4.073	4.269	4.429	4.564	4.681	4.783	4.874	4.956
35	3.462	3.815	4.067	4.262	4.422	4.556	4.672	4.774	4.865	4.946
36	3.458	3.810	4.061	4.256	4.415	4.548	4.664	4.765	4.856	4.937
37	3.454	3.805	4.056	4.249	4.408	4.541	4.656	4.757	4.847	4.928
38	3.450	3.801	4.050	4.244	4.401	4.534	4.649	4.750	4.840	4.920
39	3.447	3.796	4.045	4.238	4.395	4.528	4.642	4.742	4.832	4.912
40	3.443	3.792	4.040	4.233	4.389	4.522	4.635	4.736	4.825	4.905
41	3.440	3.788	4.036	4.228	4.384	4.516	4.629	4.729	4.818	4.898
42	3.437	3.784	4.031	4.223	4.379	4.510	4.623	4.723	4.812	4.891
43	3.434	3.781	4.027	4.218	4.374	4.505	4.618	4.717	4.805	4.885
44	3.431	3.777	4.023	4.214	4.369	4.500	4.612	4.711	4.800	4.879
45	3.429	3.774	4.020	4.210	4.364	4.495	4.607	4.706	4.794	4.873
50	3.417	3.760	4.003	4.191	4.345	4.474	4.585	4.682	4.769	4.848
100	3.365	3.696	3.930	4.110	4.256	4.379	4.485	4.578	4.660	4.734
150	3.315	3.633	3.858	4.030	4.170	4.286	4.387	4.474	4.552	4.622

b) Fitting submodels

In the previous section, I described the examination of mean differences. However, when the levels of a factor are quantitative, it is often better to examine the relationship between the response and the levels of the factor. This is commonly done using polynomials. Now, of course a polynomial of degree $t-1$ will fit exactly t points. So that, in our example, a quadratic will fit exactly the three means. Thus, we are unable to consider polynomials of higher order than 2. In practice, one would usually does not want to consider polynomials of order greater than 2; however, more than 3 points may be desirable so that deviations from the fitted curve can be tested.

Polynomial models

Thus, to investigate polynomial models up to order 2, the following sequence of models for the expectation would be fitted:

$$\begin{array}{ll} E[Y_i] = \mu & E[\mathbf{Y}] = \mathbf{X}_{\mathrm{G}}\mu \\ E[Y_i] = \mu + \gamma_1 x_k & E[\mathbf{Y}] = \mathbf{X}_1\boldsymbol{\theta}_1 \text{ where } \boldsymbol{\theta}_1' = [\mu \quad \gamma_1] \\ E[Y_i] = \mu + \gamma_1 x_k + \gamma_2 x_k^2 \quad \text{or} & E[\mathbf{Y}] = \mathbf{X}_2\boldsymbol{\theta}_2 \text{ where } \boldsymbol{\theta}_2' = [\mu \quad \gamma_1 \quad \gamma_2] \\ E[Y_i] = \alpha_k & E[\mathbf{Y}] = \mathbf{X}_{\mathrm{T}}\boldsymbol{\alpha} \text{ where } \boldsymbol{\alpha}' = [\alpha_0 \quad \alpha_{100} \quad \alpha_{200}] \end{array}$$

where x_k is the value of the kth level of the treatment factor,

μ is the intercept of the fitted equation and

γ_1 is the slope of the fitted equation and

γ_2 is the quadratic coefficient of the fitted equation.

To fit such models involves doing least squares fitting where, for the polynomial models, the **X** matrix is made up of columns that consist of the values of the levels of the factor and powers of those levels — not the indicator variables of before.

Note that each model in the sequence of models given above is marginal to all models before it as $\mathcal{C}(\mathbf{X}_{\mathrm{G}}) \subseteq \mathcal{C}(\mathbf{X}_1) \subseteq \mathcal{C}(\mathbf{X}_2) \subseteq \mathcal{C}(\mathbf{X}_{\mathrm{T}})$. That is, the columns of each **X** matrix in the above list are a linear combination of those of any of the **X** matrices to its right in the list. Marginality is not a symmetric relationship in that, if a model is marginal to second model, the second model is not necessarily marginal to the first; for example, $E[\mathbf{Y}] = \mathbf{X}_1\boldsymbol{\theta}_1$ is marginal to $E[\mathbf{Y}] = \mathbf{X}_2\boldsymbol{\theta}_2$ but not vice-a-versa, except when $t = 2$.

Example III.2 Caffeine effects on students (continued)

The **X** matrices for the example are:

$$\mathbf{X}_G = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \mathbf{X}_1 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 100 \\ 1 & 100 \\ \vdots & \vdots \\ 1 & 100 \\ 1 & 200 \\ 1 & 200 \\ \vdots & \vdots \\ 1 & 200 \end{bmatrix}, \quad \mathbf{X}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 1 & 100 & 10000 \\ 1 & 100 & 10000 \\ \vdots & \vdots & \vdots \\ 1 & 100 & 10000 \\ 1 & 200 & 40000 \\ 1 & 200 & 40000 \\ \vdots & \vdots & \vdots \\ 1 & 200 & 40000 \end{bmatrix}, \quad \mathbf{X}_T = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \end{bmatrix}$$

Note that for the example $\mathcal{C}(\mathbf{X}_G) \subset \mathcal{C}(\mathbf{X}_1) \subset \mathcal{C}(\mathbf{X}_2) = \mathcal{C}(\mathbf{X}_T)$. That is, the columns of each **X** matrix in the above list are a linear combination of those of any of the **X** matrices to its right in the list. In the case of $E[\mathbf{Y}] = \mathbf{X}_2\boldsymbol{\theta}_2$ and $E[\mathbf{Y}] = \mathbf{X}_T\boldsymbol{\alpha}$, these two models are marginal to each other as they are equivalent. This is because $\mathcal{C}(\mathbf{X}_2) = \mathcal{C}(\mathbf{X}_T)$ as the three columns of one matrix can be written as 3 linearly independent combinations of the columns of the other matrix; both matrices span the same space. However, while the fitted values are the same, the estimates and interpretation of the parameters are different. The parameters corresponding to $\mathbf{X}_2$ are interpreted as the intercept, slope and curvature coefficient; those corresponding to $\mathbf{X}_T$ are interpreted as the expected (mean) value for that treatment.

Also, in spite of being marginal, the estimators for the same parameter differ depending on the model that has been fitted. For example, $\hat{\mu} = \bar{Y}$ for the model $E[\mathbf{Y}] = \mathbf{X}_G\mu$, but $\hat{\mu} = \bar{Y} - \hat{\gamma}_1\bar{x}$ for $E[\mathbf{Y}] = \mathbf{X}_1\boldsymbol{\theta}_1$. Similar comments apply to $\hat{\gamma}_1$. The models might be marginal, but they are not orthogonal.

Hypothesis test incorporating submodels

The hypothesis test is based on the residual sums of squares (or unscaled deviance) after fitting each of the models. These residual sums of squares will be denoted D(μ), D(μ, γ_1), D(μ, γ_1, γ_2) and D(α). Once again the test statistics on which the hypothesis tests are based can then be conveniently computed in an analysis of variance table. The hypothesis tests are conducted as follows:

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 (Deviations from quadratic are zero)
H_1: $\alpha_k - \mu - \gamma_1 x_k - \gamma_2 x_k^2 \neq 0$ for all k

Step 2: Calculate test statistics

The analysis of variance table for a CRD is:

Source	df	SSq	MSq		F
Units	n–1	D(μ)			
Treatments	t–1	R(α| μ)†	R(α| μ)/(t–1)		
Linear	1	R(γ_1| μ)†	R(γ_1| μ)	$\left(= s_L^2\right)$	s_L^2 / s_R^2
Quadratic	1	R(γ_2| μ, γ_1)†	R(γ_2| μ, γ_1)	$\left(= s_Q^2\right)$	s_Q^2 / s_R^2
Deviations	t–3	R(α| μ, γ_1, γ_2)†	R(α| μ, γ_1, γ_2)/(t–3)	$\left(= s_D^2\right)$	s_D^2 / s_R^2
Residual	n–t	D(α)	D(μ,α)/(n–t)	$\left(= s_R^2\right)$	

†R(α\| μ) = D(μ) – D(α)
R(γ_1\| μ) = D(μ) – D(μ, γ_1)
R(γ_2\| μ, γ_1) = D(μ, γ_1) – D(μ, γ_1, γ_2)
R(α\| μ, γ_1, γ_2) = D(α) – D(μ, γ_1, γ_2)

The sequence of test statistics corresponds to the sequence of hypothesis pairs given in Step 1.

Step 3: Decide between hypotheses

We begin with the last hypothesis pair and determine its significance and continue up the sequence until a significant result is obtained.

A significant Deviations F indicates that the linear and quadratic terms provide an inadequate description of the trend in the treatment means and so a model based on them would be unsatisfactory. There is no point in continuing to test if this is the case.

If the Deviations F is not significant, then a significant Quadratic F indicates that a second degree polynomial is required to adequately describe the trend. As a linear coefficient is necessarily incorporated in a second degree polynomial there is no point to further testing in this case.

If both the Deviations and Quadratic Fs are not significant, then a significant Linear F indicates a linear relationship describes the trend in the treatment means.

Example III.2 Caffeine effects on students (continued)

The Genstat output for the fitting polynomial submodels is as follows:

```
  34  "
 -35  **** Mean-operator Analysis with Polynomials****
 -36  "
  37  BLOCK Students
  38  TREAT POL(Dose; 2)
  39  ANOVA [FPROB=Y] Taps


39.....................................................................

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

Variate: Taps

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

Students stratum
Dose                       2     61.400     30.700    6.18  0.006
  Lin                      1     61.250     61.250   12.33  0.002
  Quad                     1      0.150      0.150    0.03  0.863
Residual                  27    134.100      4.967
Total                     29    195.500


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

Variate: Taps

Grand mean  246.50

     Dose     0.00   100.00   200.00
            244.80   246.40   248.30


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

Table                  Dose
rep.                     10
d.f.                     27
s.e.d.                0.997
```

The hypothesis test for the example is given below. Note that in this case there are only three treatments so that a quadratic will follow the trend in the treatment means exactly. This is reflected in the fact that D(α) = D(μ, γ_1, γ_2) because $\mathcal{C}(\mathbf{X}_2) = \mathcal{C}(\mathbf{X}_T)$. Thus, the Deviations line is redundant in this example.

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: $\gamma_2 \neq 0$

Step 2: Calculate test statistics

Source	df	SSq	MSq	F	Prob
Students	29	195.50			
Doses	2	61.40	30.70	6.18	0.006
Linear	1	61.25	61.25	12.33	0.002
Quadratic	1	0.15	0.15	0.03	0.863
Residual	27	134.10	4.97		

Step 3: Decide between hypotheses

The Quadratic source has a probability of 0.863 and so the null hypothesis is not rejected in this case. The linear source has a probability of 0.002 and so the null hypothesis is rejected in this case. It is clear that the quadratic term is not significant but that the linear term is highly significant.

Fitted equation

The fitted equation is obtained by specifying, in the POL function of a TREATMENTSTRUCTURE, the order of the polynomial indicated by the hypothesis test as necessary for describing the pattern in the treatment means.

For the example, linear equation was adequate and so the analysis is redone with 1 for the order of the polynomial.

```
40  TREAT POL(Dose; 1)                        "Fit linear only"
  41  ANOVA [PRINT=aov] Taps                    "Print ANOVA table only"


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

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

Variate: Taps

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

Students stratum
Dose                       2     61.400     30.700    6.18
  Lin                      1     61.250     61.250   12.33
  Deviations               1      0.150      0.150    0.03
Residual                  27    134.100      4.967
Total                     29    195.500

  42  APOLYNOMIAL Dose; COEFF=Coeffs          "APOLYNOMIAL to fitted equation"


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

 244.8 + 0.0   * Dose

  43  PRINT #Coeffs


       244.8     0.01750
```

The fitted equation is $Y = 244.75 + 0.0175X$ where X is the number of taps.

The slope of this equation is 0.0175. That is, taps increase 0.0175 x 100 = 1.75 with each 100 mg of coffee. This conclusion seems a more satisfactory summary of the results than that the response at 200 is significantly greater than at 0 with 100 being intermediate. Below is a plot of the Dose means showing the linear trend in them. Note that the row of bars on the left gives the standard error of the difference between a pair of means. Multiply this by the t-value or studentized range value to determine if a pair of means is significantly different.

```
44  AGRAPH [GRAPH=line] XFACTOR=Dose; BAR=* "AGRAPH plots treatment means"

                                    Means for Dose

        -
        I                                                          *
  248.0 I
        I
        I
        I  -
        I  -
        I  -                                 *
  246.0 I  -
        I
        I
        I
        I          *
        I
  244.0 I
        -+---------+---------+---------+---------+---------+---------+---
     -40.0       0.0      40.0      80.0     120.0     160.0     200.0
```

c) Orthogonal contrasts

It is also possible to fit a set orthogonal contrasts, that is, a set of linear contrasts on the means that are orthogonal; there will be *t*–1 such contrasts, in addition to the grand mean contrast, in any one experiment. For example, with three treatments, a set of orthogonal contrasts is:

	Caffeine Dose		
Contrast	0	100	200
Mean	1	1	1
0 vs Rest	2	–1	–1
100 vs 200	0	1	–1

Note that the sum of products of the elements of any two contrasts is zero. These two terms would then be fitted as usual and sums of squares with one degree of freedom computed for each contrast. Thus, an F-test can be conducted separately for each contrast. The interpretation of the contrasts is that for the first contrast one is investigating the difference between zero and the two caffeine treatments. The second contrast investigates whether different doses of caffeine have an effect.

To fit these contrasts one merely sets up a matrix with each row of the matrix corresponding to a contrast.

Example III.2 Caffeine effects on students (continued)

The instructions to fit the contrasts for the example are:

```
  45  "
 -46  Analysis with Orthogonal Contrasts
 -47  "
  48  TEXT ConLab; !T('0 vs Rest','100 vs 200')
  49  MATRIX  [ROWS=ConLab; COLUMNS=3] Contrasts; !(2,-1,-1,  0,1,-1)
  50  BLOCK Students
  51  TREAT REG(Dose; 2; Contrasts)
  52  ANOVA [PRINT=AOV,MEANS,CONTRASTS; FPROB=Y] Taps


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

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

Variate: Taps

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

Students stratum
Dose                       2     61.400     30.700    6.18  0.006
  0 vs Rest                1     43.350     43.350    8.73  0.006
  100 vs 200               1     18.050     18.050    3.63  0.067
Residual                  27    134.100      4.967
Total                     29    195.500


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

Variate: Taps

***** Students stratum *****

*** Dose contrasts ***

0 vs Rest     -0.85   s.e. 0.288     ss.div. 60.0

100 vs 200    -0.95   s.e. 0.498     ss.div. 20.0


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

Variate: Taps

Grand mean  246.50

     Dose     0.00   100.00   200.00
            244.80   246.40   248.30


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

Table                 Dose
rep.                    10
d.f.                    27
s.e.d.               0.997
```

It can be shown that the estimate for the first contrast is in fact a third of the difference between the first treatment mean and the mean of the observations for the other two treatments. (The difference is $\left[\bar{x}+2\gamma_1\right]-\left[\bar{x}+(-1)\gamma_1\right]=3\gamma_1$) The estimate for the second contrast is half the difference of the 100 and 200 means.

Other contrasts between the means are possible; the restriction is that there can only be two and that they must be orthogonal. So the following contrasts are not valid as they are not orthogonal.

Caffeine Dose		
0	100	200
1	1	1
0	–1	–1
1	0	–1

A set of contrasts for four means is:

Treatment			
1	2	3	4
1	1	1	1
3	–1	–1	–1
0	2	–1	–1
0	0	1	–1

One point that arises from the analyses that have been performed so far is that the reduction in sums of squares for treatments can be obtained in several ways. We have investigated three for the analysis of a CRD: fitting treatment constants, polynomials and orthogonal contrasts. These amount to different bases for the treatment space with the unique elements of the **X** matrices whose columns form the basis being:

Treatment constants Polynomials Orthogonal contrasts

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 & 0 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \end{bmatrix}$$

They lead to a different parameter estimates with different interpretations and different partitions of the treatment sums of squares, but the total treatment sums of squares remains the same while the contrasts span the treatment space. That is,

$$SS_T = SS_L + SS_Q = SS_1 + SS_2$$

Example III.2 Caffeine effects on students (continued)

The Genstat instructions required to produce the complete analysis of this example are as follows:

```
PRINT Students,Dose,Taps

"**** Mean-operator Analysis ****"

BLOCK Students
TREAT Dose
ANOVA [FPROB=Y; PSE=LSD] Taps
AKEEP [FIT=Fit; RES=Res]
PRINT Students,Dose,Taps,Fit,Res
APLOT METHOD=fit,normal
"
****Compute theoretical values for multiple comparisons****
"
CALC t=EDT(0.975; 27)
PRINT t
"
**** Mean-operator Analysis with Polynomials****
"
BLOCK Students
TREAT POL(Dose; 2)
ANOVA [FPROB=Y] Taps
TREAT POL(Dose; 1)                      "Fit linear only"
ANOVA [PRINT=aov] Taps                  "Print ANOVA table only"
APOLYNOMIAL Dose; COEFF=Coeffs          "APOLYNOMIAL to fitted equation"
PRINT #Coeffs
AGRAPH [GRAPH=line] XFACTOR=Dose; BAR=* "AGRAPH plots treatment means"
"
Analysis with Orthogonal Contrasts
"
TEXT ConLab; !T('0 vs Rest','100 vs 200')
MATRIX  [ROWS=ConLab; COLUMNS=3] Contrasts; !(2,-1,-1,  0,1,-1)
BLOCK Students
TREAT REG(Dose; 2; Contrasts)
ANOVA [PRINT=AOV,MEANS,CONTRASTS; FPROB=Y] Taps
```

III.G Summary

In this chapter we have:

- described how to design an experiment using a completely randomized design;
- formulated a linear model using indicator variables to describe the results from a completely randomized design; used the results from chapter I to derive estimators for these and showed that the fitted values are just the treatment means so that the **P** matrices involved in the expressions for the estimators are just averaging operators;
- outlined an hypothesis test for choosing between two expectation models, termed the minimal and maximal models, using the ANOVA hypothesis test procedure outlined in chapter I;
    - the concept of a marginal model was introduced to describe a model that is a submodel of another model;
    - it was proved that in this case: the total sums of squares is partitioned into treatment and residual sums of squares; the sums of squares are obtained by squaring and summing effects; the effects can be obtained by a mean-operator analysis in which a sequence of mean operators, **P** matrices, is applied to an input vector and the result subtracted from the input vector to form the next input vector;
    - the expected mean squares under the minimal and marginal expectation models is derived to justify the choice of F test statistic;
    - again the sampling distribution of the test statistic is Snedecor's F distribution.
- shown how to obtain the both the regression and mean-operator analyses in Genstat;
- discussed procedures for checking the adequacy of the proposed models;
- outlined three methods for investigating in more detail the differences between the treatments: multiple comparisons procedures for investigating all possible pairs of differences when the factor is qualitative; polynomial models for describing the trend in the means when the factor is quantitative; orthogonal contrasts for investigating specific patterns of differences between the means usually when the factors are qualitative. The latter two involve extending the analysis of variance to partition the treatment sums of squares into one or more sources each with a single degree of freedom. When polynomials are fitted the partition may also include a source for deviations (from the fitted polynomial).