Inter-laboratory Comparisons

Michael Hutchinson, B.Sc., M.Sc.

Submitted for the degree of Doctor of Philosophy at The University of Surrey, March 1999.
Abstract

A number of alloy bars were manufactured to some very precise specifications. Certain scientific institutions then performed chemical analyses and made several measurements of the content of some chemical elements of interest. The measurements made on each of the alloy bars can be considered a set of repeated measurements.

Modelling techniques for repeated measurements are now well established. Many of these techniques are based on the multivariate normal distribution with some specified mean and covariance structure. Modelling of the covariance structure is necessary so that efficient and meaningful inferences may be made about the mean structure. For the example of repeated measurements made on an alloy bar, the set of measurements is assumed to follow a multivariate normal distribution with a mean $\mu$ and a covariance structure $\Sigma$. The choice of $\mu$ and $\Sigma$ is explored.

Experiments which produce sets of repeated measurements can quite often result in a large amount of data being collected. This means that the use of statistical techniques to fit the model to the data can become computationally demanding. The use of maximum likelihood estimation is considered. Several aspects of constructing computationally efficient algorithms to maximise the likelihood function of the data are addressed.

When the proposed model has been fitted to the data the suitability of the model and its assumptions are investigated. A score test is constructed to assess the correctness of the proposed covariance structure. Normal plots of the standardised residuals are used to assess other possible defects in the model, such as an incorrect assumption of normally distributed data.

The work which has been carried out was motivated specifically by experiments where the set of repeated measurements came from a chemical analysis of an alloy
material. It is the percentage content of a number of chemical elements which is of interest and the choice of statistical models was made with this in mind. However, it is demonstrated how the statistical techniques and models for the analysis of the chemical data may be used to analyse repeated measurements which arise from other kinds of experiments.
Acknowledgements

I am indebted to my supervisor, Professor Martin Crowder, for his generous sup­port, wise guidance and encouragement throughout the course of this research programme. Without his help this work would not have been possible.

Many thanks go to the members of staff in the Department of Mathematics and Statistics at the University of Surrey for providing a great deal of assistance on both computational and statistical-related matters. I would also like to thank them, and numerous other people in the department, for providing a pleasant and sociable working environment.

I am also grateful to Engineering and Physical Sciences Research Council for the financial support over the past three years.
# Contents

Abstract i

Acknowledgements iii

Contents iv

List of Figures viii

List of Tables x

1 Introduction 1

1.1 Background ......................................................... 1

1.2 Thesis structure and overview ................................. 2

1.3 Linear models ....................................................... 4

1.3.1 The fixed effects model ....................................... 4

1.3.2 The random effects model ................................. 6

1.3.3 The mixed effects model ................................. 7

1.4 Methods of Estimation ........................................... 8

1.4.1 Analysis of variance ......................................... 8

1.4.2 Maximum likelihood estimation ......................... 11
1.5 Matrix Notation ................................................................. 12

1.5.1 Mixed Effects Model ....................................................... 13

1.6 Discussion ............................................................................. 14

2 Analysis of single chemical data ........................................ 16

2.1 Main effects model for a single chemical ................. 16

2.2 Maximum likelihood estimation of the parameters .............. 19

2.2.1 Known variance components (Henderson's Method) .......... 20

2.2.2 Fisher Scoring Algorithm ............................................... 21

2.3 Computational evaluation of the MLEs .............. 22

2.3.1 Algebraic computation of $\Sigma^{-1}$ and $\det(\Sigma)$ ............ 23

2.3.2 Properties of estimates .................................................. 26

2.4 Model Checking ................................................................. 27

2.5 Parametric tests ................................................................. 28

2.6 Example ............................................................................... 29

2.6.1 Maximum Likelihood Estimation ................................ 29

2.6.2 Analysis of Variance ..................................................... 32

3 Analysis of multi-chemical data ........................................ 33

3.1 Model with interaction for more than one chemical ........... 33

3.2 Non-singularity of the matrix $\Sigma$ ................................. 36

3.2.1 Near singularity of $\Sigma$ ............................................... 40

3.3 Example ............................................................................... 40
3.4 Inversion of $\Sigma$ for many chemicals ......................... 44
  3.4.1 Direct numerical computation of $\Sigma^{-1}$ ............... 46
  3.4.2 Algebraic computation of $\Sigma^{-1}$ and $\det(\Sigma)$ ..... 46
  3.4.3 Avoiding direct numerical computation of $\Sigma^{-1}$ ..... 48

4 Covariance structures .................................................... 51
  4.1 Some covariance structures ........................................... 52
  4.2 Equicorrelated structures .............................................. 53
  4.3 Example ............................................................................ 54
  4.4 Suitability of $\mathbf{R}$ ...................................................... 59
    4.4.1 Graphical Indications ............................................... 60
    4.4.2 Likelihood Ratio Test ............................................... 61
    4.4.3 Score-test .............................................................. 63
  4.5 Example revisited .......................................................... 66

5 Incomplete-data layouts and missing data ..................... 67
  5.1 Estimating the parameters from the incomplete-data likelihood ...... 68
    5.1.1 Over-parameterisation ........................................... 72
  5.2 Estimating the parameters and missing data from the complete-data likelihood .................................................. 74
    5.2.1 Maximising the complete-data likelihood using the EM Algorithm .................................................. 75
  5.3 Maximisation of the incomplete data likelihood vs the complete-data likelihood .................................................. 77
List of Figures

2.1 Layout of bars, laboratories and positions. ...................... 17

2.2 Content of silicon measured by the three laboratories in the eight samples of the alloy it received. Key for x-axis: 1 - Bar 1, Top; 2 - Bar 1, Bottom etc. .......................................................... 30

2.3 Normal plot of the standardised residuals. Measurements recorded by three laboratories on the content of silicon in the top and bottom of four alloy bars. ...................................................... 31

3.1 Layout of how the samples from a single bar were distributed. C1 and C2 denote the content of two chemical elements in a particular sample. ................................................................. 36

3.2 Graphical representations of the correlation matrices of the errors: a solid square denotes a value 1, an unfilled square denotes a value < 1. ................................................................. 38

3.3 Content of aluminium measured by the three laboratories in the eight samples of the alloy it received. Key for x-axis: 1 - Bar 1, Top; 2 - Bar 1, Bottom etc. ...................................................... 42

3.4 Normal plot of the standardised residuals. Measurements recorded by three laboratories on the content of 2 chemical elements in the top and bottom of four alloy bars. ......................... 45
4.1 Percentage contents of the 8 chemicals Co, Ch, Cu, Fe Mn, Mo, Ta and Ti measured by the three laboratories in the eight samples of the alloy it received. Key for x-axis: 1 - Bar 1, Top; 2 - Bar 1, Bottom etc. ................................................................. 57

4.2 Normal plot of the standardised residuals. Measurements recorded by three laboratories on the content of 10 chemical elements in the top and bottom of four alloy bars. ................................................................. 60

4.3 Graphical representations of the correlation matrices of the errors corresponding to the content of 10 chemical elements in a particular sample of the alloy. ................................................................. 61

5.1 Normal plot of the standardised residuals. Measurements recorded by three laboratories on the content of 17 chemical elements in the top and bottom of four alloy bars. ................................................................. 82

6.1 Normal plots of the standardised residuals for each of the 10 chemical elements. ................................................................. 86

6.2 Profile log-likelihood function of $\sigma$ showing how to construct the 95% confidence interval $(A, B)$ for $\sigma$. Here, $a = L(y; \hat{\sigma}, \hat{\alpha}, \hat{\theta})$ and $b = a - 1.92$. ................................................................. 92

6.3 Profile log-likelihood function of log$\sigma_L$, $L(y; \omega | \sigma_L)$. ................................................................. 93

6.4 Profile log-likelihood function of log$\sigma_{LC}$, $L(y; \omega | \sigma_{LC})$. ................................................................. 94

7.1 Normal plots of the standardised residuals. ................................................................. 103

7.2 Profile log-likelihoods for the log variance components. ................................................................. 110

7.3 Normal plot of the standardised residuals from fitting the model to two-site test data. ................................................................. 111
List of Tables

1.1 Calculations of sums of squares for complete lay-out. .................. 9
1.2 Equations for the expected mean squares for fixed, random and mixed model respectively. .............................................. 9

2.1 The content of silicon in four bars measured at two positions by three laboratories. ...................................................... 29
2.2 ANOVA table for the chemical data in Table 2.1 ......................... 32

3.1 Covariances between the errors $e_{\text{imp}}$ and $e_{\text{upc}}$. .............. 39
3.2 The percentage content of silicon and aluminium in four bars measured at two positions by three laboratories. .................. 41

4.1 A selection of different covariance structures. .............................. 53
4.2 Maximum likelihood estimates of the variance components along with their associated standard errors. .............................. 59

5.1 Maximum likelihood estimates of the variance components along with their associated standard errors. ......................... 80

6.1 Standard errors of the Fixed Effects and Variance Components using Gaussian and Non-Gaussian Estimation. ...................... 88
6.2 Standard errors of the non-zero variance components using Gaussian Estimation. The 2nd half of this table gives 95% confidence intervals for the logs of these variance components. .......................... 94

7.1 The percentage content of aluminium determined by a number of laboratories in samples of an alloy bar. ................................. 97

7.2 Maximum likelihood estimates of the variance components and their standard errors computed from the observed information matrix. 100

7.3 Maximum likelihood estimates of the variance components and their standard errors computed from the observed information matrix. 102

7.4 Results from the two-site sound test study. This is for site 1, coding system 1, and sound (item) 1 with the 32 subjects. ............... 105

7.5 Estimates of the variance components and their standard errors computed from the observed information matrix. ................. 109

A.1 The percentage content of 17 chemical elements in four bars measured at two positions by three laboratories. The 17 chemical elements under study are as follows: Aluminium (Al), Cobalt (Co), Chromium (Cr), Copper (Cu), Iron (Fe), Hafnium (Hf), Manganese (Mn), Molybdenum (Mo), Niobium (Nb), Nickel (Ni), Phosphorus (Ph), Silicon (Si), Tantalum (Ta), Titanium (Ti), Vanadium (V), Tungsten (W), Zirconium (Zr). .................................................. 117
Chapter 1

Introduction

1.1 Background

When a material has been manufactured to some well-defined specifications it is common practice for samples of that material to be sent to a number of different laboratories so that the chemical and physical properties may be examined. These analyses are usually carried out so that the quality of the material can be assessed or for regulatory purposes. A chemical analysis may be concerned with determining the percentage content of the chemical elements in a particular alloy, whereas a physical analysis may involve investigating its tensile properties.

Before an alloy is distributed to the participating laboratories it is quite often the case that it is rolled into bars so that each laboratory receives a sample from each of the bars. The set of measurements on each bar can then be thought of as a set of repeated measurements. Applying methodology which is available for longitudinal data allows conclusions to be drawn from the results of the analysis such as assessing differences between the laboratories, differences between bars or identifying other possible sources of incompatibility. Possible sources of variation, which are specific to laboratories, are attributable to many factors, including the working patterns of the scientists, the measuring equipment involved and the different methods employed by each of the establishments.

These studies usually result in a large amount of data being recorded and a distinguishing characteristic of inter-laboratory comparisons is the common occurrence
of missing data and unbalanced or incomplete-data layouts. An unbalanced layout will usually introduce numerical and computational difficulties into most statistical techniques. In the case of analysis of variance the computation of the expected mean squares can become cumbersome.

The subject of inter-laboratory comparisons is approached by deciding on a particular model for the data available. This involves deciding on which covariates should be included in the model and then determining whether they should be treated as fixed or random effects. Once the basic linear model has been chosen it is completed by introducing a covariance structure which is thought to suitably describe the data at hand. The parameters in this model must then be estimated and so some statistical technique has to be adopted. The choice of the statistical technique to be used is made with the intention that the parameter estimates possess some desirable properties and also takes into account the computational problems which may be encountered. Finally, the results of the statistical analysis have to be interpreted so that a number of important questions raised by the scientists can be answered, such as is there a significant difference between laboratories or the identification of outlying laboratories.

1.2 Thesis structure and overview

A large amount of data from a chemical analysis of some alloy has been supplied. These data consist of observations recorded on a number of chemical elements with a number of covariates. These covariates include the location at which the measurements were recorded (i.e. the laboratory) and the position on the alloy bar from which the measurements were taken.

In this chapter the topic of inter-laboratory comparisons is introduced. The choice of linear models is investigated, including the selection of fixed, random and mixed effects models. Two methods for the statistical analysis of the data are looked at. These two methods are maximum likelihood and analysis of variance.

The problem is first approached by analysing the data recorded on a single chemical element. Chapter 2 is concerned with the selection of a model that is thought to
suitably describe the data on a single chemical element. The parameters in the model are then estimated by maximum likelihood. When the model has been fitted its goodness-of-fit is assessed and parametric tests carried out.

The data available consist of measurements recorded, on not just one, but a number of different chemical elements and this problem is addressed in Chapter 3. Once again, a linear model is suggested to describe the data. However, in this case the measurements recorded will be correlated in some way and so the choice of covariance matrices is investigated. The parameters are estimated, as in Chapter 2, by maximum likelihood and some computational problems are discussed.

In Chapter 4 the covariance matrix of the data is investigated and re-parameterised. A number of different covariance structures are considered and the most suitable applied. Now that the model is complete the goodness-of-fit is assessed by the use of the likelihood ratio test and the construction of a score-test.

These types of experiments usually result in a large amount of data being collected and a distinguishing characteristic is that some data may be missing resulting in the data forming an unbalanced layout. Chapter 5 addresses this problem using two techniques. The first is to reduce the full design matrix so that it is of the same dimension as the data, and the second technique is to employ the EM algorithm to fit the model and to estimate both the parameters and missing data.

A large number of statistical techniques, including maximum likelihood as applied here, are based on the assumption that the data are normally distributed. In Chapter 6, the assumption that the data follow a normal distribution is examined. Gaussian estimation is then used to modify the standard errors of the parameter estimates in case the assumption of normality is incorrect. It is also shown how confidence intervals of the variance components may be constructed using likelihood ratios.

All of the work which has been carried out, in Chapters 1 through to 6, will be demonstrated using numerical examples. In Chapter 7 two more sets of data are examined. The first example follows on with the theme of a chemical analysis of an alloy, whereas the second example is used to show how the model introduced may be used to analyse data arising from different situations.
1.3 Linear models

Suppose that a material has been manufactured to some pre-defined specifications and that some chemical analysis has been performed on it. For example, it may be the percentage content of a chemical element in the material which is of interest. In a simple experiment this material is rolled into a number of bars and then a selection of laboratories receive a portion of each of the bars and are asked to determine the content of the chemical element under study.

In the following sections three types of linear model, fixed, random and mixed, are described which are intended to model the data taking into account the circumstances under which the chemical analysis was carried out.

1.3.1 The fixed effects model

If the $n_l$ laboratories used in this experiment are permanent scientific institutions and that inferences from a statistical analysis of the data are concerned with only the participating laboratories then the laboratories are treated as fixed. Similarly, if inferences are to be made only about the $n_b$ bars that were rolled then the bars are also treated as fixed.

If laboratory $l$ then measures the chemical content in bar $b$ several times then the $p$th measurement can be denoted by $y_{blp}$. A stochastic model which describes the measurements $y_{blp}$ is

$$y_{blp} = \mu_{bl} + \epsilon_{blp}$$

(1.1)

where $\mu_{bl}$ is defined as the 'true' underlying mean value of the $n_p$ measurements recorded on bar $b$ by laboratory $l$.

Assuming a full $n_p \times n_l \times n_p$ layout, $\mu_{bl}$ is estimated as $\bar{\mu}_l = \bar{y}_{bl}$. A dot replacing a subscript indicates the mean has been calculated over that subscript. The term $\epsilon_{blp}$ refers to the error terms which are assumed to be independently and identically distributed having a normal distribution with mean 0 and variance $\sigma^2$.

Still assuming a full layout the overall mean of the $\mu_{bl}$ terms is the average of the chemical percentage content over all the different repetitions of the experiment.
and is estimated as $\hat{\mu} = \bar{y} - \bar{\mu}$. The deviation of the chemical content in bar $b$ from the true underlying mean is defined as the main effect of bar $b$ and is estimated as

$$\hat{\alpha}_b^B = \bar{y}_b - \bar{\mu}.$$  

(1.2)

Since, in the present section, interest is only concerned with the specific bars in the study, the bar effect is a fixed effect. Similarly, the mean of the chemical content measured by laboratory $I$ is $\bar{y}_I$, and the laboratory effect is estimated as the amount by which this exceeds the overall mean

$$\hat{\alpha}_I^L = \bar{y}_I - \bar{\mu}.$$  

(1.3)

The amount by which the effect of laboratory $I$ specific to bar $b$ exceeds the effect of laboratory $I$ yields the interaction between bar $b$ and laboratory $I$ and is estimated as

$$\hat{\alpha}_{bl}^{BL} = \bar{y}_{bl} - \bar{y}_b - \bar{y}_I + \bar{y}.$$  

(1.4)

Now using the estimates of each of the effects in equations (1.2), (1.3) and (1.4) leads to

$$\hat{\mu}_{bl} = \hat{\mu} + \hat{\alpha}_b^B + \hat{\alpha}_I^L + \hat{\alpha}_{bl}^{BL}.$$  

Substituting the true bar, laboratory and ‘bar × laboratory’ effects into the model of equation (1.1) yields

$$y_{blp} = \mu + \alpha_b^B + \alpha_I^L + \alpha_{bl}^{BL} + e_{blp},$$  

(1.5)

which is the more conventional notation of the fixed effects model.

In the case $n_p = 1$, i.e. only a single measurement being recorded on bar $b$ by laboratory $I$, the model with interaction becomes indistinguishable from the model without interaction. When this is the case the interaction term and the error term are combined and so, by letting $e_{bl} = \alpha_{bl}^{BL} + e_{blp}$, the model becomes

$$y_{bl} = \mu + \alpha_b^B + \alpha_I^L + e_{bl}.$$  

and so when $n_p = 1$ only the no interaction case can be studied.
The main aspect of the fixed effects model is that no distributional assumptions are made about the effects which are included in the model. However, it is assumed that

\[ \sum_{b=1}^{n_b} \alpha_b^B = 0 = \sum_{l=1}^{n_l} \alpha_l^L. \]

and

\[ \sum_{b=1}^{n_b} \alpha_b^{BL} = 0 = \sum_{l=1}^{n_l} \alpha_l^{BL}. \]

The mean and covariance of the data may now be conveniently written as

\[ E(y_{blp}) = \mu + \alpha_b^B + \alpha_l^L + \alpha_{bl}^{BL}, \]

and

\[ \text{Cov}(y_{blp}, y_{bl'p'}) = \begin{cases} 
\sigma^2 & \text{if } b = b', l = l', p = p' \\
0 & \text{otherwise.} 
\end{cases} \]

**1.3.2 The random effects model**

Suppose now that the laboratories included in the study are just a few chosen at random from an infinite, to idealise a large, population of similar scientific establishments. Now, from a statistical analysis one would like to be able to make inferences not just about the laboratories which took part but be able to generalise and make inferences about the population of such laboratories. Similarly, the bars which were produced may be thought of as just a few chosen at random from an infinite population of such bars. When this is the case the bar and laboratory effects are called random effects.

Apart from a slight change in notation, the random effects model appears to be similar to the fixed effects model of equation (1.5):

\[ y_{blp} = \mu + \alpha_b^B + \alpha_l^L + \alpha_{bl}^{BL} + e_{blp}. \]

(1.6)

The main difference here is that assumptions are made about distributions of the random effects included in the model. As before, the error terms are assumed to be normally distributed with mean 0 and variance \( \sigma^2 \) and it is also assumed that the error terms are independent of each of the random effects \( \alpha_b^B, \alpha_l^L \) and \( \alpha_{bl}^{BL} \). The
assumptions placed on these random effects is that they are normally distributed and independent of each other and that
\[ E(a_b^B) = E(a_l^L) = E(a_{ml}^{BL}) = 0. \]
Finally, the variances of \( a_b^B, a_l^L \) and \( a_{ml}^{BL} \) are denoted by \( \sigma_B^2, \sigma_L^2 \) and \( \sigma_{BL}^2 \), respectively.
The model of equation (1.6), along with its assumptions, implies that the mean and covariance of the data satisfy
\[ E(y_{bp}) = \mu \]
and
\[ \text{Cov}(y_{bp}, y_{bp'}) = \delta_{ib} \sigma_B^2 + \delta_{ib} \sigma_L^2 + \delta_{ib} \sigma_{BL}^2 + \delta_{ib} \delta_{ip} \sigma^2 \]
where \( \delta_{ij} \) represents the Kronecker delta symbol which takes the value 1 if \( i = j \) and the value 0 otherwise. The intraclass correlation between bars is \( \sigma_B^2 / (\sigma_B^2 + \sigma_L^2 + \sigma_{BL}^2 + \sigma^2) \) and that between laboratories is \( \sigma_L^2 / (\sigma_B^2 + \sigma_L^2 + \sigma_{BL}^2 + \sigma^2) \).

### 1.3.3 The mixed effects model

In more general circumstances, especially when there are more than two effects influencing the measured content of the chemical element, it is unlikely that the model selected will consist entirely of fixed or random effects. In the present example it may seem reasonable, depending on the nature of the laboratories, that the bars should be treated as random effects and the laboratories as fixed effects. When a model contains a combination of fixed and random effects it is called a mixed effects model. The model is
\[ y_{bp} = \mu + a_t^L + a_b^B + a_{ml}^{BL} + e_{bp}. \]
Here the Greek letters have been used to represent the fixed effects while the Latin letters represent the random effects. This notation is used throughout. The sum of a particular fixed effect is again zero, and so for the laboratories
\[ \sum_{t=1}^{n_t} a_t^L = 0. \]
For the random effects the same distributional assumptions are applied as in the random effects model, which were
\[ a_{ij} \sim N(0, \sigma_b^2), \quad a_{il}^{jl} \sim N(0, \sigma_{bL}^2), \quad e_{ip} \sim N(0, \sigma^2). \]

All of this results in the following mean and covariance structure of the data:
\[ E(y_{blp}) = \mu + \alpha_i^L \]
and
\[ \text{Cov}(y_{blp}, y_{blp'}) = \delta_{bb} \sigma_b^2 + \delta_{ii} \delta_{jj} \sigma_{bL}^2 + \delta_{ii} \delta_{jj} \sigma^2. \]

1.4 Methods of Estimation

1.4.1 Analysis of variance

Analysis of variance can be thought of as a generalisation of the t-test for considering the differences between more than two groups. The idea behind analysis of variance is to increase the sensitivity of an experiment in such a way that the total variation of the variable being studied can be separated into components that are of experimental interest. Conclusions from an analysis of variance may be made by creating F-statistics to test for possible differences between groups or to test for possible interaction effects. To start an analysis of variance a model is required. Continuing with the example of the previous section, assuming fixed laboratories and random bars, the model is
\[ y_{blp} = \mu + a_{b}^R + \alpha_i^L + a_{il}^{jl} + e_{blp} \]
with the assumptions of Section 1.3 still applying.

The sums of squares are calculated in the usual way and are listed in Table 1.1 along with the corresponding degrees of freedom for each of the effects. Whichever model is being used, whether it be fixed, random or mixed, the calculation of the sums of squares remains the same. However, the computation of the expected mean squares does differ from model to model. The mean squares are calculated as \( MS = \frac{SS}{df} \). In the model where laboratories are treated as fixed and bars as
Table 1.1: Calculations of sums of squares for complete layout.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bars (B)</td>
<td>$n_b - 1$</td>
<td>$n_p n_1 \sum_{k=1}^{n_b} (\bar{y}<em>{b..} - \bar{y}</em>{..})^2$</td>
</tr>
<tr>
<td>Laboratories (L)</td>
<td>$n_l - 1$</td>
<td>$n_p n_0 \sum_{l=1}^{n_l} (\bar{y}<em>{l..} - \bar{y}</em>{..})^2$</td>
</tr>
<tr>
<td>Interaction (BL)</td>
<td>$(n_b - 1)(n_l - 1)$</td>
<td>$n_p \sum_{k=1}^{n_b} \sum_{l=1}^{n_l} (\bar{y}<em>{bl} - \bar{y}</em>{b..} - \bar{y}<em>{l..} + \bar{y}</em>{..})^2$</td>
</tr>
<tr>
<td>Residual (R)</td>
<td>$n_b n_l (n_p - 1)$</td>
<td>$\sum_{k=1}^{n_b} \sum_{l=1}^{n_l} \sum_{p=1}^{n_p} (y_{klp} - \bar{y}_{..})^2$</td>
</tr>
<tr>
<td>Total</td>
<td>$n_b n_l n_p - 1$</td>
<td>$\sum_{k=1}^{n_b} \sum_{l=1}^{n_l} \sum_{p=1}^{n_p} (y_{klp} - \bar{y}_{..})^2$</td>
</tr>
</tbody>
</table>

Table 1.2: Equations for the expected mean squares for fixed, random and mixed model respectively.

<table>
<thead>
<tr>
<th>Source</th>
<th>Bars Fixed Laboratories Fixed</th>
<th>Bars Random Laboratories Random</th>
<th>Bars Random Laboratories Fixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>$\sigma^2 + \frac{n_p n_1}{n_b - 1} \sum_{k=1}^{n_b} (\alpha_b^B)^2$</td>
<td>$\sigma^2 + n_p \sigma^2_{BL} + n_p n_1 \sigma^2_{B}$</td>
<td>$\sigma^2 + n_p n_1 \sigma^2_{B}$</td>
</tr>
<tr>
<td>L</td>
<td>$\sigma^2 + \frac{n_p n_0}{n_l - 1} \sum_{l=1}^{n_l} (\alpha_l^L)^2$</td>
<td>$\sigma^2 + n_p \sigma^2_{BL} + n_p n_0 \sigma^2_{L}$</td>
<td>$\sigma^2 + n_p \sigma^2_{BL} + \frac{n_p n_0}{n_l - 1} \sum_{l=1}^{n_l} (\alpha_l^L)^2$</td>
</tr>
<tr>
<td>BL</td>
<td>$\sigma^2 + n_p \sum_{k=1}^{n_b} \sum_{l=1}^{n_l} (\alpha_{bl}^{BL})^2$</td>
<td>$\sigma^2 + n_p \sigma^2_{BL}$</td>
<td>$\sigma^2 + n_p \sigma^2_{BL}$</td>
</tr>
<tr>
<td>R</td>
<td>$\sigma^2$</td>
<td>$\sigma^2$</td>
<td>$\sigma^2$</td>
</tr>
</tbody>
</table>

Determining which rows of Table 1.1 should be used to create the appropriate F-statistics is done by examination of Table 1.2. Consider the fixed effects model of equation (1.5),

$$y_{klp} = \mu + \alpha_b^B + \alpha_l^L + \alpha_{bl}^{BL} + e_{klp}$$
and the terms

\[ D_B = \sum_{b=1}^{n_b} (\alpha_b^R)^2, \quad D_L = \sum_{l=1}^{n_l} (\alpha_l^L)^2, \quad D_{BL} = \sum_{b=1}^{n_b} \sum_{l=1}^{n_l} (\alpha_{bl}^{BL})^2 \]

in the equations of the first column of Table 1.2. \( D_B \) and \( D_L \) are measures of the variation between the bars and laboratories respectively and are likely to be large in the presence of real bar and laboratory effects. The appropriate test statistic for testing differences between bars, i.e. for testing the null hypothesis \( \alpha_b^B = 0 \) for \( b = 1, \ldots, n_b \), would be

\[ F = \frac{SSB/(n_b - 1)}{SSR/n_p n_b (n_p - 1)} \]

where \( SSB \) and \( SSR \) are the sums of squares for bars and residual. This statistic then has an F-distribution on \( (n_b - 1) \) and \( n_p n_b (n_p - 1) \) degrees of freedom.

The second main use of the analysis of variance technique is as a tool for obtaining estimates of the various parameters or functions of the parameters. Estimates of the variance components can be calculated by equating the mean squares to their expected values. This is most clearly demonstrated by using the random effects model and the second column of Table 1.2. The estimate of \( \sigma^2 \) is just \( \frac{SSR}{df(R)} \). The estimate of \( \sigma_{BL}^2 \) can then be found by solving the equation

\[ \frac{SSBL}{df(BL)} = \sigma^2 + n_p \sigma_{BL}^2 \]  \hspace{1cm} (1.7)

where \( \sigma^2 \) is replaced by the estimate \( \hat{\sigma}^2 \). Solving equation (1.7) for \( \sigma_{BL}^2 \) yields

\[ \hat{\sigma}_{BL}^2 = \frac{1}{n_p} \left[ \frac{SSBL}{df(BL)} - \frac{SSR}{df(R)} \right] \]  \hspace{1cm} (1.8)

Estimates of the bar and laboratory variance components can then be found as

\[ \hat{\sigma}_B^2 = \frac{1}{n_p n_b} \left[ \frac{SSB}{df(B)} - \frac{SSBL}{df(BL)} \right] \]  \hspace{1cm} (1.9)

and

\[ \hat{\sigma}_L^2 = \frac{1}{n_p n_b} \left[ \frac{SSL}{df(L)} - \frac{SSBL}{df(BL)} \right] \]  \hspace{1cm} (1.10)

Under the assumptions of normality the estimates of the variance components from balanced data will be unbiased and have the minimum variance of any estimates. However, when the data is unbalanced and a mixed model is fitted these
estimates of the variance components may be biased. Another problem which is encountered with unbalanced data is that the expected mean squares become increasingly difficult to calculate.

It is quite possible that the estimates of the variance components in equations (1.8), (1.9) and (1.10) turn out to be negative. Clearly this is a problem since by definition variance components are greater than zero. One could take the occurrence of a negative estimate of a variance component to infer that the true value of the component is zero. If the variance component estimate is redefined as $\max(0, \hat{\sigma}^2)$, the mean square error of the estimate will be reduced.

The term $\frac{SS}{E(\hat{d}^2)}$ has a $\chi^2(df)$ distribution and so the distribution of $\hat{\sigma}^2$ is

$$\hat{\sigma}^2 = \frac{SSR}{df(R)} \sim \frac{\sigma^2}{n_p n_t (n_p - 1)} \chi^2_{n_p n_t (n_p - 1)}.$$

One final problem concerning the analysis of variance method is determining the distribution of the variance component estimates. This is because equating the mean squares to their expected values leads to the estimates of the variance components being linear functions of mean squares and so their distributions will be linear functions of $\chi^2$ variates. For example, using equation (1.9), the distribution of $\hat{\sigma}_B^2$ is

$$\hat{\sigma}_B^2 \sim \frac{\sigma^2 + n_p \sigma_{BL}^2 + n_p n_t \sigma_B^2}{n_p n_t (n_p - 1)} \chi^2_{n_p - 1} - \frac{\sigma^2}{n_p n_t n_b (n_p - 1)} \chi^2_{n_p - 1}(n_b - 1).$$

Unfortunately the coefficients here contain the unknown variance components and so the exact forms of these distributions cannot be found.

### 1.4.2 Maximum likelihood estimation

A method of parameter estimation which is known to possess several desirable properties is that of maximum likelihood estimation. This method will be demonstrated using the fixed effects model of Section 1.3.1. Recall that the model is

$$y_{blp} = \mu + \alpha_b^B + \alpha_l^L + \alpha_{bl}^{BL} + e_{blp}$$

where the $e_{blp}$'s come from a normal distribution with mean 0 and variance $\sigma^2$. The measurements $y_{blp}$ then form a set of normally distributed variables with mean
\[ \mu_{bl} = \mu + \alpha_i^B + \xi_j^L + \gamma_{bl}^{BL}. \]

The density function of \( y_{blp} \) is then
\[
f(y_{blp}; \mu_{bl}, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} \left[ \frac{(y_{blp} - \mu_{bl})^2}{2\sigma^2} \right]
\]
and the likelihood function is defined as
\[
\mathcal{L}(y_{11}, \ldots, y_{n_b n_l n_p}; \mu_{bl}, \sigma^2) = \prod_{b=1}^{n_b} \prod_{l=1}^{n_l} \prod_{p=1}^{n_p} f(y_{blp}; \mu_{bl}, \sigma^2), \quad (1.11)
\]

The maximum likelihood estimates of \( \mu_{bl} \) and \( \sigma^2 \) are then the values \( \hat{\mu}_{bl} \) and \( \hat{\sigma}^2 \) which maximise \( \mathcal{L} \) or, equivalently, \( \log \mathcal{L} \). It is usually more convenient to work with the log-likelihood function of the data rather than the likelihood function.

In some circumstances the estimates will be unbiased and possess the property of minimum variance. However, in this case the estimate of \( \sigma^2 \) is biased but can be made unbiased by multiplying it by the constant \( \frac{n}{n-1} \).

When either of the effects are treated as random, as in Sections 1.3.2 and 1.3.3, the possibility of negative estimates for the variance components is no longer a problem, since as long as the maximisation of the likelihood function is carried out with the correct restrictions placed on the parameters, the estimates will be non-negative.

### 1.5 Matrix Notation

For the two methods of parameter estimation discussed in Section 1.4 it is usually more convenient to write the linear models in matrix notation rather than the forms used in Section 1.3. The reason for this is that any algebraic manipulations which have to be carried out, such as calculating the expected mean squares for the variance components in the case of ANOVA or finding the set of parameters which maximise the likelihood function in maximum likelihood estimation, can be done a lot simpler. In this section it is shown how these linear models can be written in matrix notation and some of the advantages highlighted.
1.5.1 Mixed Effects Model

Of the three models described in Section 1.3 it is the mixed effects model of Section 1.3.3 which is the most appropriate for describing the data which results from the laboratory experiments and it is this model which will be used extensively in subsequent chapters. For the statistical techniques which are to be applied it is more convenient to write the linear model of equation (1.7)

\[ y_{bp} = \mu + \alpha_1^L + a_b^B + a_{bl}^{BL} + e_{bp} \]

as the two-stage linear model given by Laird and Ware (1982).

This model may be written in an expanded form as

\[ y_{bp} = \mu + \sum_{i=1}^{n_l} \delta_{il} \alpha_i^L + \sum_{i=1}^{n_u} \delta_{ib} \alpha_i^H + \sum_{i=1}^{n_b} \sum_{j=1}^{n_l} \delta_{ij} \delta_{j} a_{ij}^{BL} + e_{bp} \] (1.12)

where \( \delta_{il} = 0 \) if \( i \neq l \) and \( \delta_{il} = 1 \) if \( i = l \). Now, by letting \( y \) be the \( n \times 1 \) vector containing all the data in some specified order, and by letting \( e \) be the \( n \times 1 \) vector of the corresponding error terms, the linear model can be written as

\[ y = X\alpha + Z\alpha + e. \]

Here, the vector \( \alpha = (\mu, \alpha_1^L, \alpha_2^L, \ldots, \alpha_n^L) \) contains all the fixed effects and \( \alpha = (a_b^B, a_{b1}^B, a_{b2}^B, \ldots, a_{bn}^B, a_{b1}^{BL}, a_{b2}^{BL}, \ldots, a_{bn}^{BL}) \) contains all the random effects. The matrices \( X \) and \( Z \) are respectively called the design matrices of the fixed and random effects and are of size \( n \times p \) and \( n \times q \). The matrix \( Z \) is usually constructed to consist entirely of zeros and ones which coincide with the values of the Kronecker deltas in equation (1.12). The matrix \( X \) can be constructed in a similar fashion but usually some adjustments are made so that the constraint \( \sum_{i=1}^{n_l} \alpha_i^L = 0 \) is automatically satisfied.

The assumptions which were made about the parameters \( a_b^B, a_{b1}^{BL} \) and \( e_{bp} \) were

\[ a_b^B \sim N(0, \sigma_b^2), \quad a_{b1}^{BL} \sim N(0, \sigma_{BL}^2), \quad e_{bp} \sim N(0, \sigma^2) \]

and these can easily be carried over into this matrix notation by assuming that \( \alpha \) and \( e \) have multivariate normal distributions with

\[ E(\alpha) = 0, \quad \text{Var}(\alpha) = B = \text{diag} [\sigma_b^2 \mathbf{I}_{n_b}, \sigma_{BL}^2 \mathbf{I}_{n_{BL}}] \]
and

\[ E(e) = 0, \quad \text{Var}(e) = \mathbf{E} = \sigma^2 \mathbf{I}_n. \]

The data vector \( \mathbf{y} \) then has a multivariate normal distribution with

\[ E(\mathbf{y}) = \mathbf{X} \alpha, \quad \text{Var}(\mathbf{y}) = \Sigma = \mathbf{Z}^T \mathbf{B} \mathbf{Z} + \mathbf{E} \]

and the Gaussian log-likelihood function is

\[ L(\mathbf{y}; \alpha, \theta) = -\frac{1}{2} \log \det(2\pi \Sigma) - \frac{1}{2} (\mathbf{y} - \mathbf{X} \alpha)^T \Sigma^{-1} (\mathbf{y} - \mathbf{X} \alpha) \quad (1.13) \]

where \( \theta = (\sigma^2_B, \sigma^2_E, \sigma^2) \) and \( \Sigma = \Sigma(\theta) \). The maximum likelihood estimates of the fixed effects and variance components are then the set of parameters which maximise this function. When the model becomes more complex and the measurements recorded are not independent but correlated in some structured way the univariate likelihood function (1.11) can no longer be used and its multivariate counterpart (1.13) has to replace it.

### 1.6 Discussion

The problems caused by computing the expected mean squares in the analysis of variance approach, especially when the layouts are more complex, makes the method of maximum likelihood estimation preferable for calculating estimates of the variance components. Even though, in the case of missing data the algebraic computation of the expected mean squares becomes even more laborious, this task can still be performed quicker than iteratively maximising a complex likelihood function.

However, maximum likelihood estimation is not without its problems. In the simple example demonstrated in Sections 1.4.1 and 1.4.2 it is straightforward to algebraically find explicit solutions for the estimates of the parameters but this will not always be the case. In more complicated situations, where more complex models are employed, the univariate normal distribution has to be replaced by its multivariate counterpart because the measurements \( y_{\text{rep}} \) will be correlated in some structured way. When this is the case maximisation of the likelihood function
will almost certainly have to be carried out by some iterative numerical procedure. This may mean that inversion of the covariance matrix is required at each iteration, whereas analysis of variance only requires the inverse to be computed once. With a large amount of data, which is quite often the case with inter-laboratory studies, these matrices will also be very large. Even with the advances in the electronic computer these iterative searches for the maximum of the likelihood function can still be time consuming. This problem can be overcome to some extent by finding algebraic expressions for these inverses combined with the implementation of efficient computer algorithms.

In the chapters that follow more complicated studies are introduced where the selected model may contain many parameters. Ideas will be suggested to answer the problems discussed above and those that are encountered along the way.
The silicon (Si) content in an alloy is to be studied. An alloy was manufactured to some pre-defined specifications - one of which was that the content of silicon, as a percentage of the whole material, should be 0.5%. After the manufacturing process had been completed this alloy was rolled into four bars and then three laboratories each received a sample from the top and bottom of each bar. A layout of how the material was divided up is shown in Fig. 2. The total number of samples that were produced is then $4 \times 3 \times 2 = 24$. Each laboratory then determined, by using X-Ray fluorescence, the silicon content as a percentage in each of the samples that it received. In a more general case $n_b$ bars would be rolled and samples at $n_p$ different positions would be distributed to $n_l$ laboratories resulting in $n = n_b \times n_l \times n_p$ measurements (samples) being produced. The data on any one bar can be thought of as a set of repeated measurements.

### 2.1 Main effects model for a single chemical

The bars and laboratories used in the experiment will be regarded as just a few chosen at random from populations of such and therefore treated as random effects. Since we are only interested in a specific number of positions along the bar at which measurements were recorded the position effect is treated as fixed. A basic linear model is,

$$y_{nlp} = \mu + \alpha_p + \alpha_l + \alpha_b + e_{nlp}.$$  

(2.1)
Here, \( \mu \) is the ‘true’ underlying value of the percentage content of the chemical element under study and the \( \alpha_p^P \) term \((p = 1, \ldots, n_p)\) denotes the fixed position effects. The random laboratory and bar effects are denoted by \( a_l^L \) \((l = 1, \ldots, n_l)\) and \( a_b^B \) \((b = 1, \ldots, n_b)\) respectively, with means zero and variances \( \sigma_l^2 \) and \( \sigma_b^2 \). The term \( e_{ulp} \) represents random error and has mean 0 and variance \( \sigma^2 \). Model (2.1) is then a main-effects model, containing no interaction effects. Under model (2.1), with the assumptions described in Chapter 1,

\[
\text{Cov}(\delta_{ulp}, \delta_{ulp'}) = \delta_{ulp} \sigma_l^2 + \delta_{ulp} \sigma_l^2 + \delta_{ulp} \delta_{ulp'} \sigma^2.
\]

Let \( y \) be the vector of all the measurements with subscripts on \( y_{ulp} \) varying fastest from right to left. Then the model may be written in matrix notation as

\[
y = \mathbf{1}_n \mu + X_1 \alpha^P + Z_1 a^L + Z_2 a^B + \mathbf{e} \tag{2.2}
\]

with \( \mathbf{1}_n \) denoting the \( n \times 1 \) vector of 1's. The design matrix, \( X_1 \), of the position effects could be set up as \( X_1 = \mathbf{1}_{n_b} \otimes \mathbf{1}_{n_l} \otimes \mathbf{1}_{n_p} \) which results in a position indicator being awarded to each measurement. Here, \( \otimes \) represents the Kronecker matrix product. So, in the case of \( n_p = 2 \), a measurement taken from a top piece of material would have \( (1, 0) \) in the corresponding row of \( X_1 \), whereas a measurement made on a sample from a bottom piece of the material would have \( (0, 1) \) in that row of \( X_1 \).

A constraint which holds is that the sum of the parameters for a fixed effect is zero, e.g. \( \sum_{p=1}^{n_p} \alpha_p^P = 0 \) and so the matrix \( X_1 \) can be written as \( X_1 = \mathbf{1}_{n_b} \otimes \mathbf{1}_{n_l} \otimes C_{n_p} \). 

---

**Figure 2.1: Layout of bars, laboratories and positions.**
where \( C_{np} \) is designed to impose the above constraint. In the case of \( n_p = 2 \) this is equivalent to \( \alpha_1^P = -\alpha_1^P \) and so \( C_2 \) would simply be \( [1 \ -1]^T \). The rows of \( X_1 \) corresponding to a measurement from the top of a bar would then consist of \( (1) \) and the rows corresponding to a measurement from the bottom of a bar would contain \( (-1) \). The parameter vector \( \alpha^P \) would then contain only the single parameter \( \alpha_1^P \). In the more general case of \( n_p \) positions \( \alpha^P = (\alpha_1^P, \ldots, \alpha_{n_p-1}^P)^T \) and the corresponding rows of \( X_1 \) are \( (1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 1) \) and \( (-1, -1, \ldots, -1) \).

The design matrices of the random laboratory and bar effects, \( Z_1 \) and \( Z_2 \), are set up as

\[
Z_1 = 1_{n_l} \otimes I_{n_l} \otimes 1_{n_p}, \quad Z_2 = 1_{n_b} \otimes 1_{n_l} \otimes 1_{n_p}.
\]

The vectors \( \alpha^L = (a_1^L, \ldots, a_{n_l}^L)^T \) and \( \alpha^B = (a_1^B, \ldots, a_{n_b}^B)^T \) denote the laboratory and bar effects, and the vector \( e \) is \( n \times 1 \) and contains the random errors \( e_{lp} \) and has a covariance matrix

\[
\Sigma = \text{Cov}(e) = I_n \sigma^2.
\]

Model (2.2) is now just an expanded version of the Harville-Laird-Ware model (Laird & Ware, 1982), \( y = X\alpha + Za + e \) where \( X \) is the \( n \times p \) design matrix of the fixed effects and \( Z \) is the \( n \times q \) design matrix of the random effects. The vectors \( \alpha \) and \( a \) are then of length \( p \) and \( q \) respectively. The vector of random effects has as its covariance matrix

\[
B = \text{Var}(a) = \text{diag} \left[ \sigma_L^2 I_{n_l}, \sigma_B^2 I_{n_b} \right].
\]

The resulting expectation vector of \( y \) is then

\[
\mu = E(y) = 1_n \mu + X_1 \alpha^P = X\alpha
\]

with \( X = [1_n \ X_1] \) and \( \alpha = [\mu \ (\alpha^P)^T]^T \). Similarly, the corresponding covariance matrix is

\[
\Sigma = \text{Var}(y) = Z_1 \text{Var}(a^L)Z_1^T + Z_2 \text{Var}(a^B)Z_2^T + E
\]

\[
= \sigma_L^2 (Z_1Z_1^T) + \sigma_B^2 (Z_2Z_2^T) + I_n \sigma^2
\]

\[
= \sigma_L^2 (J_{n_l} \otimes I_{n_l} \otimes J_{n_p}) + \sigma_B^2 (I_{n_b} \otimes J_{n_l} \otimes J_{n_p}) + I_n \sigma^2. \quad (2.5)
\]

with \( J_i \) as the \( i \times i \) matrix of 1's.
2.2 Maximum likelihood estimation of the parameters

Assuming that the random terms, \(a^L, a^B, \text{ and } e\) are normally distributed, then the density function of the data is the multivariate normal density

\[
f(y; \mu, \alpha^P, \sigma_L, \sigma_B, \sigma) = \frac{1}{(2\pi)^{n/2} \det(\Sigma)} \exp \left[-\frac{1}{2}(y - X\alpha)^T \Sigma^{-1}(y - X\alpha)\right],
\]

where \(\Sigma = \Sigma(\sigma_L, \sigma_B, \sigma)\). The corresponding Gaussian log-likelihood function of the data is then given by

\[
L(y; \alpha^P, \sigma_L, \sigma_B, \sigma) = \log \left(\frac{1}{(2\pi)^{n/2} \det(\Sigma)} \exp \left[-\frac{1}{2}(y - X\alpha)^T \Sigma^{-1}(y - X\alpha)\right]\right)
= -\frac{1}{2} \log \det(2\pi \Sigma) - \frac{1}{2}(y - X\alpha)^T \Sigma^{-1}(y - X\alpha) \quad (2.6)
\]

where \(\Sigma = \Sigma(\sigma_L, \sigma_B, \sigma) = \Sigma(\theta)\). The maximum likelihood estimates of \(\alpha\) and \(\theta\) are then the set of parameters \(\hat{\alpha}\) and \(\hat{\theta}\) which maximise equation (2.6).

If the variance components of the model were known then estimating the fixed effects would be straightforward. This is demonstrated in Section 2.2.1 by applying Henderson’s method to the problem to reduce the computational demands. However, the variance components will not usually be known so both the fixed effects and variance components will have to be estimated simultaneously. Much work has been done in this area and many iterative algorithms have been suggested to jointly estimate both the fixed effects and variance components. Laird and Ware (1982) discussed in great detail how the EM algorithm can be applied to evaluate the maximum likelihood estimates and the restricted maximum likelihood estimates of the variance components of such models. Jenrich and Schluchter (1986) also consider what they termed the Hybrid EM algorithm, but restricted its use to the incomplete-data model. They also describe two further algorithms; the Newton-Raphson algorithm and the Fisher Scoring algorithm which is described in Section 2.2.2.
2.2.1 Known variance components (Henderson’s Method)

For the model,

\[ y = X\alpha + Za + e \]

with the covariance structure \( \Sigma = ZBZ^T + E \) then the normal equations are

\[ X^T\Sigma^{-1}X\hat{\alpha} = X^T\Sigma^{-1}y \]

with solution

\[ \hat{\alpha} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}y. \tag{2.7} \]

The matrices \( B \) and \( E \) are at present both diagonal and hence easily invertible. However, due to the structure which the matrix \( Z \) takes it is unlikely that the matrix \( \Sigma \) will be diagonal and so inversion of \( \Sigma \) can be computationally expensive. A method derived by Henderson which assumes that the variance components are known allows the fixed effects to be estimated with only the inversion of \( B \) and \( E \) necessary.

If the variance components are known then the term \( a \) in the model can be treated as fixed effects and so \( V(y) = E \). Treating \( a \) as fixed, the normal equations become

\[
\begin{bmatrix}
X^T \\
Z^T
\end{bmatrix} \begin{bmatrix}
E^{-1} \\
X & Z
\end{bmatrix} \begin{bmatrix}
\hat{\alpha} \\
\hat{a}
\end{bmatrix} = \begin{bmatrix}
X^T \\
Z^T
\end{bmatrix} E^{-1}y
\]

or

\[
\begin{bmatrix}
X^T E^{-1}X & X^T E^{-1}Z \\
Z^T E^{-1}X & Z^T E^{-1}Z
\end{bmatrix} \begin{bmatrix}
\hat{\alpha} \\
\hat{a}
\end{bmatrix} = \begin{bmatrix}
X^T E^{-1}y \\
Z^T E^{-1}y
\end{bmatrix}.
\]

Henderson showed that by adding the matrix \( B^{-1} \) to the lower right sub-matrix \( Z^T E^{-1}Z \) the solution for \( \alpha \) in the set of linear equations

\[
\begin{bmatrix}
X^T E^{-1}X & X^T E^{-1}Z \\
Z^T E^{-1}X & Z^T E^{-1}Z + B^{-1}
\end{bmatrix} \begin{bmatrix}
\hat{\alpha} \\
\hat{a}
\end{bmatrix} = \begin{bmatrix}
X^T E^{-1}y \\
Z^T E^{-1}y
\end{bmatrix} \tag{2.8}
\]

are identical to the solution of equation (2.7). Now, in calculating the estimates of the fixed effects, it is the diagonal matrices \( E \) and \( B \) (equations (2.3) and (2.4))
which need to be inverted rather than the more structured matrix $\Sigma$. Solving the set of linear equations in equation (2.8) requires the matrix

\[
\begin{bmatrix}
X^T E^{-1}X & X^T E^{-1}Z \\
Z^T E^{-1}X & Z^T E^{-1}Z + B^{-1}
\end{bmatrix}
\]

to be inverted. The size of this matrix is $(p + q) \times (p + q)$, i.e. the total of fixed and random effects and will usually be considerably smaller than the matrix $\Sigma$.

The vector $a$ is a vector of random variables and in some cases the estimates $\hat{a}$ may be of interest. When the variance components are known $a$ can be estimated directly from the set of linear equations given in (2.8) and are found to be

\[
\hat{a} = B Z^T (y - X\hat{\alpha}).
\]

The covariance matrix of the estimator $\hat{a}$ is

\[
V(\hat{a}) = B Z^T (\Sigma^{-1} - \Sigma^{-1} X (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1}) Z B
\]

but if the error in estimation is to be assessed then the variance of $V(\hat{a} - a) = B - V(\hat{a})$ should be used because the expression in equation (2.9) does not take into account the variation in $a$.

### 2.2.2 Fisher Scoring Algorithm

Many algorithms have been developed which are capable of maximising log-likelihood functions. One such method is the Fisher scoring algorithm (Jennrich and Schluchter, 1986) which maximises $L(y; \alpha, \theta)$ by an iterative procedure that calculates updated values of the parameters $\alpha^{i+1}$ and $\theta^{i+1}$, from the existing values $\alpha^i$ and $\theta^i$. The updated parameter values are computed as

\[
\begin{bmatrix}
\alpha^{i+1} \\
\theta^{i+1}
\end{bmatrix} =
\begin{bmatrix}
\alpha^i \\
\theta^i
\end{bmatrix} -
(E \begin{bmatrix}
H_{\alpha\alpha} & H_{\alpha\theta} \\
H_{\theta\alpha} & H_{\theta\theta}
\end{bmatrix})^{-1} \begin{bmatrix}
s_{\alpha} \\
s_{\theta}
\end{bmatrix}.
\]

Here, $s = \begin{bmatrix} s_{\alpha} \\
s_{\theta}
\end{bmatrix}$ and $H = \begin{bmatrix} H_{\alpha\alpha} & H_{\alpha\theta} \\
H_{\theta\alpha} & H_{\theta\theta}
\end{bmatrix}$ are known as the score vector and information, or Hessian, matrix respectively:

\[
s = \begin{bmatrix}
\frac{\partial L}{\partial \alpha} \\
\frac{\partial L}{\partial \theta}
\end{bmatrix} \quad \quad \phantom{\text{H = }}
\begin{bmatrix}
\frac{\partial L}{\partial \alpha} & \frac{\partial L}{\partial \alpha \theta} \\
\frac{\partial L}{\partial \theta} & \frac{\partial L}{\partial \theta \theta}
\end{bmatrix}.
\]
Expressions for the sub-vectors of $s$ are found to be

$$s_i = X^T \Sigma^{-1} (y - X \alpha);$$

$$\left(s_i\right)_l = \frac{1}{2} \text{tr} \left[ \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_l} \Sigma^{-1} \left((y - X \alpha)(y - X \alpha)^T - \Sigma\right)\right].$$

while expressions for the sub-matrices of the Hessian matrix are,

$$H_{\alpha\alpha} = -X^T \Sigma^{-1} X;$$

$$\left(H_{\alpha\theta}\right)_l = X^T \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_l} \Sigma^{-1} (y - X \alpha);$$

$$\left(H_{\theta\theta}\right)_{lm} = -\frac{1}{2} \text{tr} \left[ \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_l} \Sigma^{-1} (2(y - X \alpha)(y - X \alpha)^T - \Sigma) \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_m} \right] + \frac{1}{2} \text{tr} \left[ \Sigma^{-1} ((y - X \alpha)(y - X \alpha)^T - \Sigma) \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_l \theta_m} \right]$$

where $(s_i)_l$ denotes the $l$th element of the vector $(s_i)$, $(H_{\alpha\theta})_l$ the $l$th column of the matrix $H_{\alpha\theta}$ and $(H_{\theta\theta})_{lm}$ is the $(l, m)$th entry of $H_{\theta\theta}$. Now, because $E(H_{\alpha\theta}) = 0$, the two sets of equations can be solved separately to calculate the updated values of the parameters $\alpha$ and $\theta$.

Estimates of $\alpha$ are found by finding the solution to the set of equations $s_\alpha = 0$. The resulting estimate $\hat{\alpha}$ is then the least squares estimate (LSE), or equivalently the maximum likelihood estimate (MLE) of $\alpha$, and can be computed as

$$\hat{\alpha} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} y.$$  \hspace{1cm} (2.12)

This estimate can then be substituted for $\alpha$ in $s_\theta$ and $H_{\theta\theta}$, and then updated values of $\theta$ can be found using the equation

$$\theta^{i+1} = \theta^i - [E(H_{\theta\theta})]^{-1} s_\theta.$$

### 2.3 Computational evaluation of the MLEs

Using some estimate of the parameter vector $\theta$, estimates of the fixed effects can be computed directly from the data using the formula of equation (2.12). These estimates $\hat{\alpha}$ are now substituted into the Gaussian log-likelihood function of equation (2.6) to yield a reduced Gaussian log-likelihood function dependent
only on the variance components $\theta = (\sigma_L, \sigma_B, \sigma)$,

$$L_1(y; \theta) = \log \left( \left( \det (2\pi \Sigma) \right)^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (y - X\hat{\alpha})^T \Sigma^{-1} (y - X\hat{\alpha}) \right] \right)$$

$$= -\frac{1}{2} \log \det (2\pi \Sigma) - \frac{1}{2} (y - X\hat{\alpha})^T \Sigma^{-1} (y - X\hat{\alpha}). \quad (2.13)$$

The routine e04jaf in the FORTRAN NAG library, a Quasi-Newton minimisation method, is then employed to minimise the negative of equation (2.13). This routine only requires the value of the log-likelihood function to be calculated at some parameter vector $\theta = (\sigma_L, \sigma_B, \sigma)$ and does not need to be supplied with any of its derivatives. It does however need to be supplied with an initial estimate of the parameter vector which maximises $L_1(y; \theta)$. This method can be documented in the following steps:

1. Choose some initial value for the parameter vector $\theta = (\sigma_B, \sigma_L, \sigma)$, call it $\hat{\theta}^{(i)}$. Use this estimate to calculate the current estimate of the covariance matrix, $\hat{\Sigma}^{(i)}$.

2. Evaluate the vector of the fixed effects using the current parameter vector $\hat{\theta}^{(i)}$, using the formula $\hat{\alpha}^{(i)} = (X^T \hat{\Sigma}^{(i)} X)^{-1} X^T \hat{\Sigma}^{(i)} y$.

3. Substitute the current estimate of the fixed effects $\hat{\alpha}^{(i)}$ into the equation $L(y; \theta)$ giving $L_1^{(i)}(y; \hat{\theta}^{(i)})$.

4. Maximise $L_1^{(i)}(y; \hat{\theta}^{(i)})$ to give an updated value $\hat{\theta}^{(i+1)}$ of the parameter vector $\theta$.

5. Return to step 2 and continue until successive estimates of $\alpha, \theta$ stabilise.

### 2.3.1 Algebraic computation of $\Sigma^{-1}$ and $\det (\Sigma)$

The numerical inversion of a matrix, such as $\Sigma$, can becomes computationally demanding as its size increases. In fact the number of arithmetic operations required to invert an $n \times n$ matrix is proportional to $n^3$. It will now be demonstrated how the inverse and determinant of the matrix $\Sigma$ may be evaluated algebraically, hence reducing the amount of computation required to evaluate the MLEs of the parameters.

23
Recall from equation (2.5) that the covariance matrix is

\[ \Sigma = \sigma^2_L (J_{n_p} \otimes I_{n_t} \otimes J_{n_p}) + \sigma^2_B (I_{n_b} \otimes J_{n_t} \otimes J_{n_p}) + I_n \sigma^2. \]

This expression can be rewritten as

\[ \Sigma = I_{n_b} \otimes (\sigma^2_B J_{n_t} \otimes J_{n_p} + \sigma^2 I_{n_t} \otimes I_{n_p}) + J_{n_b} \otimes (\sigma^2_L I_{n_t} \otimes J_{n_p}). \]

Now, let

\[ C = I_i \otimes A_j + J_i \otimes B_j \]

where the matrices \( I_i \) and \( J_i \) are of size \( i \times i \) and the matrices \( A_j \) and \( B_j \) are of size \( j \times j \). Then,

\[ \det(C) = \{\det(A_j)\}^{i-1} \det(A_j + iB_j) \]

and

\[ C^{-1} = I_i \otimes A_j^{-1} + J_i \otimes \{A_j^{-1}B_j(A_j + iB_j)^{-1}\}. \]

These may be used to find algebraic expressions for the inverse and determinant of the matrix \( \Sigma \). After some routine algebraic reduction these are

\[ \Sigma^{-1} = \sigma^{-2} I_{n_b} \otimes \{I_{n_p,n_t} - c J_{n_p,n_t}\} \]

\[ -\sigma^{-2} \sigma^2_L J_{n_b} \otimes \{I_{n_t} \otimes (J_{n_p} D_1^{-1}) + J_{n_t} \otimes (\sigma^2_B \{1 - c n_p n_t\} G - c n_p D_1^{-1})\} \]

and

\[ \det(\Sigma) = \{c^{-1} \sigma^2 \sigma^2 (n_p n_t^{-1})\}^{n_t-1} \{\det(D_1)\}^{n_t-1} \det(D_2) \]

where

\[ D_1 = \sigma^2 I_{n_p} + n_b \sigma^2_L J_{n_p}, \]

\[ D_2 = D_1 + n_b \sigma^2_B J_{n_p}, \]

\[ G = J_{n_p} D_1^{-1} J_{n_p} D_1^{-1} \]

and

\[ c = \frac{\sigma^2_B}{\sigma^2 + n_p n_t \sigma^2_B}. \]
These expressions for the inverse and determinant of $\Sigma$ can now be used to help in the estimation of the fixed effects $\alpha$ through the use of equation (2.12) and the variance components $\theta$ through maximising the log-likelihood of equation (2.6). The advantage gained here is that matrices of size $n_p \times n_p$, such as $D_1$ and $D_2$, have to be inverted numerically rather than $\Sigma$ which is of size $n \times n$, where $n$ is the total number of measurements recorded ($n = n_b \times n_t \times n_p$). Substituting the expressions for $\Sigma^{-1}$ of equation (2.14), along with the matrix $X$ which has the structure as described in Section 2.3.1, into equation (2.12) results in

$$
\hat{\mu} = \frac{y}{n}; \\
\hat{\alpha}_p = \frac{y_p}{n_b \times n_t} - \hat{\mu} \quad (p = 1, \ldots, n_p - 1).
$$

Here the estimate of $\mu$, the 'true' underlying mean value, is just the average of all the measurements recorded. The estimate of the $p$th position effect is the average of all the measurements recorded at position $p$ above the underlying mean value. These estimates of $\mu$ and $\alpha_p$ can be computed without assuming values for $\sigma_L$, $\sigma_B$ and $\sigma$.

In this particular example, because of the special structures which the matrices $X$ and $\Sigma$ take, the explicit solution for $\hat{\alpha}$, given in equations (2.16) and (2.17), does not require assumed values of the variance components. The result of this is that the procedure described in Section 2.3 can be simplified by removing steps 2 and 3 because the estimate of $\alpha$ is the same at every iteration.

Maximisation of the reduced Gaussian log-likelihood function of equation (2.13) still has to be performed by some iterative procedure and at each iteration $L_1(y; \theta^{(i)})$ must be evaluated for some parameter $\theta^{(i)}$. The inverse and determinant of $\Sigma$ appear in $L_1(y; \theta)$ through the terms

$$
Q(y; \theta) = -\frac{1}{2}(y - X\hat{\alpha})^T \Sigma^{-1}(y - X\hat{\alpha})
$$

and

$$
Q_1(\theta) = -\frac{1}{2} \log \det(2\pi \Sigma),
$$

both of which may be evaluated more efficiently using the algebraic expressions derived in section 2.3.1.
In more complicated examples, such as those covered in Chapters 3 and beyond, the reduced Gaussian log-likelihood function of equation (2.13) becomes more complex. A possible consequence of this is that the function \( L_1(y; \theta) \) may not be unimodal and that the search for the parameter vector which maximises this function results in a local maximum and not a global maximum. To try to ensure that this does not happen, one possible solution is to start the searching procedure at a number of different locations in the parameter space and, if the resulting estimates of the parameter vectors are not the same, then compare the values of the likelihood function at each of the parameter estimates.

### 2.3.2 Properties of estimates

In the case where \( \Sigma \) does not contain unknown parameters, \( \hat{\alpha} \) possesses all the usual properties including those of being consistent and asymptotically normal and has a covariance matrix which can be found from equation (2.9)

\[
V_\alpha = \text{Var}(\hat{\alpha}) = -H_{xx}^{-1} = (X^T \Sigma^{-1} X)^{-1}.
\]  

Under regularity conditions, which include that the first and second derivatives of \( L_1(y; \sigma_L, \sigma_B, \sigma) \) exist, maximisation of \( L_1(y; \sigma_L, \sigma_B, \sigma) \) will lead to maximum likelihood estimates of \( \theta = (\sigma_L, \sigma_B, \sigma) \) which are consistent and asymptotically normal. The sub-vector score vector of the log-likelihood \( L_1 \), is \( s_\theta = \frac{\partial L_1}{\partial \theta} \) and it is easily shown that \( E(s_\theta) = 0 \), implying that the estimating equation \( s_\theta = 0 \) is unbiased.

Now, as long as the maximisation process is carried out with the correct restrictions placed on the parameters, this method cannot produce negative estimates of the variance components, so \( \hat{\sigma}_L^2 \geq 0 \) and \( \hat{\sigma}_B^2 \geq 0 \). It may be the case that maximisation of (2.13) leads to estimates of variance components which lie on the boundaries of the parameter space, e.g. \( \hat{\sigma}_B = 0 \). When this is the case it is quite possible that the equation \( \frac{\partial L_1}{\partial \theta} = 0 \) will no longer be satisfied. A simple way around this problem is to infer that \( \hat{\sigma}_B = 0 \) means that \( \sigma_B = 0 \), and hence to remove \( \sigma_B \) from the model and carry out the maximisation process again.

Standard errors of the variance components are found directly from the matrix
D^{-1}$, where D has as its $(i, m)$th entry

$$(D)_{im} = E\left( -\frac{\partial^2 L_1}{\partial \theta_i \partial \theta_m} \right),$$

(Patefield, 1977). From equation (2.11) this can be expressed as

$$(D)_{im} = E(-\mathbf{H}_{\hat{\theta}}) = \frac{1}{2} \text{trace} \left[ \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_m} \right]. \tag{2.19}$$

The matrix D can then be computed by inserting the estimates $\hat{\theta}$ into equation (2.19). Unfortunately, computing D by substituting the model-based estimates into this equation does not guarantee that it will be positive definite. This problem can be avoided by computing $-\frac{\partial^2 L_1}{\partial \theta_i \partial \theta_m}$ directly by second-differencing to give a numerical estimate of D. The standard errors of the variance components are then the square roots of the elements on the diagonal of $D^{-1}$.

### 2.4 Model Checking

The techniques of regression analysis are based on an assumed model which specifies the expectation of the observations in terms of a set of unknown parameters and makes further assumptions about the covariance structure. If these assumptions are valid, statistical conclusions can be based on the values of the parameter estimates. In practice however, the assumptions of the model will have to be checked and this will usually involve examining plots of the residuals.

The vector of the fitted values is calculated as $(\mathbf{X}\hat{\alpha})_{blp}$ and the raw residuals as $(y - \mathbf{X}\hat{\alpha})_{blp}$. The set of standardised residuals can then be computed as $d = S(y - \mathbf{X}\hat{\alpha})$ where S is the upper triangular matrix satisfying $SS^T = I_n$. The matrix $\hat{\Sigma}$ is the estimated covariance matrix of the measurements constructed using the estimates $\hat{\sigma}_L$, $\hat{\sigma}_B$ and $\hat{\sigma}$. Now, if the assumptions about $\mu$ and $\Sigma$ are exactly correct then the vector of standardised residuals will approximate a set of independent standard normal variates. The standardised residuals can then be plotted against the corresponding normal scores c, to produce a normal plot, and conclusions drawn. The normal plot can also be used to detect any outlying observations.
A more formal test to assess the linearity of the normal plot is to calculate the Shapiro-Wilk (SW) statistic

\[ SW = \frac{(c^T \Omega^{-1} d)^2}{(c^T \Omega^{-2} c) \sum (d_i - \bar{d})} \]

where \( \Omega \) is the covariance matrix of the normal scores \( c \). Small values of \( SW \) would then be some indication towards the lack of linearity of the normal plot and hence non-normality of the standardised residuals.

### 2.5 Parametric tests

Tests on individual regression coefficients can be made by calculating the normal deviate for that coefficient. In the case of \( r_p = 2 \), the appropriate normal deviate for testing the hypothesis \( \alpha_p^2 = 0 \), i.e. that there is no difference between positions, is calculated as \( \frac{\delta_p}{\text{se}(\delta_p)} \). This value can then be referred to normal tables to determine whether the hypothesis should be accepted. Investigating whether the overall mean parameter \( \mu \) equals the pre-defined percentage content of silicon (0.5%) results in the normal deviate \( \frac{\bar{y} - 0.5}{\text{se}(\mu)} \) which should also be referred to normal tables.

To test for differences between bars or laboratories is equivalent to testing \( \sigma_B > 0 \) or \( \sigma_L > 0 \). Since these parameters are range-restricted it means that the distributions of their estimates are not usually very 'normal'. Hence, the normal deviate \( \frac{\delta_1}{\text{se}(\delta_1)} \) is not very useful. The problem of range restriction can be avoided by working in terms of the log variance components, e.g. \( \log(\sigma_L) \). The standard error of \( \log(\delta_i) \) is \( \frac{\text{se}(\delta_i)}{\delta_i} \). Testing the hypothesis \( \sigma_L > 0 \) is the equivalent of testing the hypothesis 'log(\( \sigma_L \)) > -\infty'. A 95% confidence interval is constructed for \( \log(\sigma_L) \) as

\[ \log(\sigma_L) - 1.96 \text{se}(\log(\sigma_L)) < \log(\sigma_L) < \log(\sigma_L) + 1.96 \text{se}(\log(\sigma_L)). \]

Now, using this confidence interval it can be determined whether the hypothesis \( \sigma_L > 0 \) should be accepted or rejected. Such tests should only be used as a guide.
2.6 Example

2.6.1 Maximum Likelihood Estimation

In an experiment an alloy consisting of a number of chemical elements was manufactured. This alloy was divided up into 4 bars. A sample from the top and bottom of each bar was then sent to 3 laboratories. The percentage content of silicon measured by the laboratories in the alloy samples they received are given in Table 2.1. The percentage contents of silicon measured by each laboratory are plotted Fig 2.2. There is nothing much that can be said from this graph, apart from maybe, that the measurements recorded by laboratory 2 are smaller than the measurements recorded by the other two laboratories.

Table 2.1: The content of silicon in four bars measured at two positions by three laboratories.

<table>
<thead>
<tr>
<th></th>
<th>Bar I Top</th>
<th>Bar I Bottom</th>
<th>Bar II Top</th>
<th>Bar II Bottom</th>
<th>Bar III Top</th>
<th>Bar III Bottom</th>
<th>Bar IV Top</th>
<th>Bar IV Bottom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lab. 1</td>
<td>0.490</td>
<td>0.490</td>
<td>0.500</td>
<td>0.510</td>
<td>0.480</td>
<td>0.490</td>
<td>0.480</td>
<td>0.510</td>
</tr>
<tr>
<td>Lab. 2</td>
<td>0.467</td>
<td>0.476</td>
<td>0.459</td>
<td>0.477</td>
<td>0.450</td>
<td>0.485</td>
<td>0.454</td>
<td>0.480</td>
</tr>
<tr>
<td>Lab. 3</td>
<td>0.490</td>
<td>0.500</td>
<td>0.470</td>
<td>0.470</td>
<td>0.500</td>
<td>0.510</td>
<td>0.480</td>
<td>0.510</td>
</tr>
</tbody>
</table>

Looking at the data, it appears that the content of silicon measured in the top of a bar tends to be less than the content in the bottom of the corresponding bar and so the inclusion of a position effect seems appropriate. The linear model is that of equation (2.1),

\[ y_{ijp} = \mu + \alpha_p^L + \alpha_t^L + \alpha_b^B + \epsilon_{ijp} \]

with \( p = 1, 2 ; t = 1, 2, 3 \) and \( b = 1, 2, 3, 4 \). The covariance structure assumed for this data is

\[ \Sigma = \text{Var}(y) = \sigma_l^2 (\mathbf{J}_n \otimes \mathbf{I}_n \otimes \mathbf{J}_n) + \sigma_B^2 (\mathbf{I}_n \otimes \mathbf{J}_n \otimes \mathbf{J}_n) + \mathbf{I}_n \sigma^2. \]

The parameters to be estimated are \( (\mu, \sigma_l^L, \sigma_B, \sigma) \).

Estimates of \( \mu \) and \( \alpha_l^P \) are found directly from the data, and without having to assume values of the variance components, by using equation (2.12), which resulted
Figure 2.2: Content of silicon measured by the three laboratories in the eight samples of the alloy it received. Key for x-axis: 1 - Bar 1, Top; 2 - Bar 1, Bottom etc.

\[
\hat{\alpha} = (\hat{\mu}, \hat{\alpha}_1^P)^T = (0.4845, -7.8333 \times 10^{-3})^T
\]

and the estimated variance of \(\hat{\alpha}_1^P\) was found to be \(5.12 \times 10^{-4}\). Maximisation of \(L_1(y; \sigma_L, \sigma_B, \sigma)\) yielded,

\[
(\hat{\sigma}_L, \hat{\sigma}_B, \hat{\sigma}) = (1.0662 \times 10^{-2}, 0.0000, 1.1085 \times 10^{-2}).
\]

The estimate of \(\sigma_B\) was found to be zero and so it was decided that this parameter should be removed from the model and the log-likelihood function maximised again before the standard errors of the variance components were calculated. The standard errors of \(\hat{\sigma}_L\) and \(\hat{\sigma}\), calculated directly from the observed information matrix, are 0.0049 and 0.0017 respectively.

If the model fitted is correct then the vector of standardised residuals will approximate a vector of independent normal variates. A plot of the standardised residuals against their normal scores is shown in Fig. 2.3. There appears to be an outlying point corresponding to \(y_{2,3,2}\), the measurement made by laboratory 3 in the top of
Figure 2.3: Normal plot of the standardised residuals. Measurements recorded by three laboratories on the content of silicon in the top and bottom of four alloy bars.

bar 2. This measurement seems to contradict the trend that the content of silicon in the top of the bar is slightly less than the content in the bottom. Apart from that one observation, y_{2,3,2}, there is no reason to believe that, from Fig. 2.3, the model of equation (2.1) provides an inadequate fit to the data. This is supported by the Shapiro-Wilk test statistic which is calculated as $SW = 0.946$.

As mentioned above, since the parameters $\sigma_L$, $\sigma_B$ and $\sigma$ are range-restricted, the normal deviates $\frac{\tilde{z}}{se(\tilde{z})}$, which are used to test if the variance components are zero, may not be valid. The estimate of the log laboratory effect, $\log(\tilde{\sigma}_L)$, is $-4.5411$ and has a standard error of 0.4639. So a 95% confidence interval can be constructed for $\log(\sigma_L)$ as

$$-5.4503 < \log\sigma_L < -3.6319.$$  

This suggests that the variance component $\sigma_L$ is non-zero and therefore there exists a real difference between laboratories. The fact that the MLE of $\sigma_B$ is zero seems to suggest there are no real differences between bars.
We can test whether the content of silicon in the alloy is 0.5% by testing whether the parameter $\mu$ is 0.5. The estimate of $\mu$ is 0.4845 with a standard error of 0.0066. Testing the hypothesis that $\mu = 0.5$ results in a normal deviate of 2.35, which suggests that the content of silicon does not meet its pre-determined target of 0.5%.

2.6.2 Analysis of Variance

The example described in Section 2.6.1 will now be approached using the method of analysis of variance. The model is again that of equation (2.1)

$$y_{ip} = \mu + \alpha_i + \beta_j + \epsilon_{ip}$$

with $\alpha_i \sim N(0, \sigma_i^2)$, $\beta_j \sim N(0, \sigma_j^2)$ and $\epsilon_{ip} \sim N(0, \sigma^2)$ all assumed to be independent. The results of the analysis are given in Table 2.2. Testing the hypothesis that $\alpha_i^2 = 0$, i.e. no difference between positions, results in an $F$ value of 10.62, providing evidence that the inclusion of the position effect in the model is necessary. The variance components are estimated by equating the mean squares in

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>E(MS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bars</td>
<td>3</td>
<td>$9.83 \times 10^{-5}$</td>
<td>$3.28 \times 10^{-5}$</td>
<td>$\sigma^2 + \sum_{j=1}^{q} \sigma_j^2$</td>
</tr>
<tr>
<td>Laboratories</td>
<td>2</td>
<td>$3.10 \times 10^{-3}$</td>
<td>$1.55 \times 10^{-3}$</td>
<td>$\sigma^2 + \sum_{i=1}^{r} \sigma_i^2$</td>
</tr>
<tr>
<td>Positions</td>
<td>1</td>
<td>$1.47 \times 10^{-3}$</td>
<td>$1.47 \times 10^{-3}$</td>
<td>$\sigma^2 + \sum_{j=1}^{q} \frac{n_j}{n_i} \sum_{p=1}^{n_i} (\alpha_p^2)$</td>
</tr>
<tr>
<td>Residual</td>
<td>17</td>
<td>$2.48 \times 10^{-3}$</td>
<td>$1.46 \times 10^{-4}$</td>
<td>$\sigma^2$</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>$7.15 \times 10^{-3}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

column 4 to the expected mean squares in column 5. The computations of the expected mean squares can be found in Appendix A.1. Firstly, the estimate of $\sigma^2$ is $\hat{\sigma}^2 = 1.46 \times 10^{-4}$. The estimates of the bar and laboratory components are found by solving the equations

$$\sigma^2 + \sum_{j=1}^{q} \sigma_j^2 = 3.28 \times 10^{-5}, \quad \sigma^2 + \sum_{i=1}^{r} \sigma_i^2 = 1.55 \times 10^{-3}.$$  

This yields $\hat{\sigma}_B^2 = -1.887 \times 10^{-3}$ and $\hat{\sigma}_L^2 = 1.76 \times 10^{-4}$. The estimate of $\sigma_B^2$ is negative which was not the case in the maximum likelihood approach. This estimate could be assumed to be zero which would, again, suggest that there is no real bar effect.
Chapter 3

Analysis of multi-chemical data

In the previous section we were only interested in one chemical element in the alloy. More realistically, an alloy consists of many chemical elements, in some cases maybe as many as twenty or more. As an approach to this problem let us take the previous case but with the introduction of measurements made on aluminium.

3.1 Model with interaction for more than one chemical

As before, we have $n_b$ bars, $n_l$ laboratories and $n_p$ positions but now measurements were recorded on $n_c$ chemicals, where $n_c = 2$ in this case. The total number of measurements now recorded is $n = n_b \times n_l \times n_p \times n_c$. If we let $y_{bipc}$ denote the content of chemical $c$ at position $p$ in bar $b$ measured by laboratory $l$ then the linear model can be extended to,

$$y_{bipc} = \mu + \alpha_p^b + \alpha_c^l + \alpha_{pc} + \epsilon_{bpci}$$

This model is similar to that of equation (2.1) but with the introduction of the fixed chemical and 'position x chemical' interaction effects. Also included in the model now are the random 'bar x chemical', 'bar x position', 'laboratory x chemical' and 'bar x position x chemical' interaction effects.
Again, by letting \( y \) be the vector of all the measurements with the subscripts on \( V \) varying fastest from right to left, the model can be expressed in the Harville-Laird-Ware set up as

\[
y = 1\mu + X_1\alpha^F + X_2\alpha^C + X_3\alpha^{PC} + Z_1\alpha^B + Z_2\alpha^L + Z_3\alpha^{BC} + Z_4\alpha^{BP} + Z_5\alpha^{LC} + Z_6\alpha^{BPC} + \varepsilon. \tag{3.2}
\]

The \( X \) matrices take a similar form to those in the single chemical example, e.g.

\[
X_1 = 1_{n_b} \otimes 1_{n_r} \otimes C_{np} \otimes 1_{ne}, \quad X_2 = 1_{n_b} \otimes 1_{n_r} \otimes 1_{np} \otimes C_{nc},
\]

which all automatically impose the constraint that the sum of the parameters representing a fixed effect is zero.

The design matrices of the random effects are, again, extensions of the ones used in the single chemical case, e.g.

\[
Z_1 = 1_{n_b} \otimes 1_{n_r} \otimes 1_{np} \otimes 1_{ne}, \quad Z_2 = 1_{n_b} \otimes 1_{n_r} \otimes 1_{np} \otimes 1_{ne},
\]

So, model (3.2) is just an extension of \( y = X\alpha + Za + \varepsilon \) where \( X \) \((n \times p)\) and \( Z \) \((n \times q)\) are replaced by \([1_{n_b} \; X_1 \; X_2 \; X_3]\) and \([Z_1 \ldots Z_6]\), and the parameter vectors \( \alpha \) \((p \times 1)\) and \( a \) \((q \times 1)\) are replaced by \([\mu, (\alpha^F)^T, (\alpha^C)^T, (\alpha^{PC})^T]^T\) and \([(a^B)^T, (a^L)^T, \ldots, (a^{BPC})^T]^T\). The covariance matrix of the data is then \( \Sigma = ZBZ^T + E \), where \( E = \text{Var}(\varepsilon) \) and

\[
B = \text{Var}(a) = \text{diag}[\sigma^2_B, \sigma^2_L, \ldots, \sigma^2_{BPC}],
\]

is a diagonal matrix with \( \sigma^2_B, \sigma^2_L \) etc. along the diagonal, resulting in

\[
\Sigma = \sigma^2_B(1_{n_b} \otimes J_{n_r} \otimes J_{np} \otimes J_{ne}) + \sigma^2_C(J_{n_b} \otimes 1_{n_r} \otimes J_{np} \otimes J_{ne}) + \sigma^2_{BC}(1_{n_b} \otimes J_{n_r} \otimes J_{np} \otimes J_{ne}) + \sigma^2_{BP}(1_{n_b} \otimes J_{n_r} \otimes 1_{np} \otimes J_{ne}) + \sigma^2_{LC}(J_{n_b} \otimes 1_{n_r} \otimes J_{np} \otimes J_{ne}) + \sigma^2_{BPC}(1_{n_b} \otimes J_{n_r} \otimes 1_{np} \otimes J_{ne}) + E. \tag{3.4}
\]

The most significant difference between the model for one chemical element and two chemical elements is the structure of the matrix \( E \), which describes the covariances between the \( \varepsilon \)s. The data vector \( y \) can be partitioned into \( y = (y_{111}, y_{112}, \ldots, y_{nmnp}) \) where, for example, \( y_{111} \) is the vector consisting of the
measurements recorded at position 1, of bar 1 by laboratory 1. The vector of random errors \( e \) can similarly be partitioned into \( e = (e_{111}, e_{112}, \ldots, e_{n_c n_p}) \), where the vector \( e_{111} \) represents the errors associated with the set of measurements \( y_{111} \).

Consider any sample of the alloy bar pictured in Fig. 3.1, say the sample from the bottom of the bar which was sent to laboratory 1. This particular sample appears in the top left hand corner of the diagram. Now, let the matrix \( R \) represent the covariances between the percentage content of the chemical elements in this sample. In the simplest case of no structure being applied to \( R \), and when \( n_c = 2 \),

\[
R = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix},
\]

where \( \sigma_{11} \) and \( \sigma_{22} \) are the variances of the measurements made on the content of the chemical elements \( C_1 \) and \( C_2 \) and \( \sigma_{12} = \sigma_{21} \) is the covariance between the measurements of the contents of \( C_1 \) and \( C_2 \).

From Fig. 3.1 it seems reasonable at first sight to assume that the covariance between the content \( C_1 \) in the bottom of the bar measured by laboratory 1 and the content \( C_1 \) in the bottom of the bar measured by laboratory 2 is \( \sigma_{11} \). Similarly, the covariances of the content \( C_2 \) in the bottom of the bar measured by different laboratories are \( \sigma_{22} \), and the covariances between the contents \( C_1 \) and \( C_2 \) measured at the same position but by different laboratories are \( \sigma_{12} = \sigma_{21} \).

Now, consider the covariances of the contents of the chemical elements at different positions but measured by the same laboratories. Using the same principle, the covariances between the content \( C_1 \) in the top of the bar and \( C_1 \) in the bottom of the bar are \( \sigma_{11} \), the covariances between the contents \( C_2 \) are \( \sigma_{22} \) and the covariances between the contents \( C_1 \) and \( C_2 \) are \( \sigma_{12} = \sigma_{21} \). Going further,

\[
\text{Cov}(C_1, C_1) = \sigma_{11} \\
\text{Cov}(C_1, C_2) = \sigma_{12} \\
\text{Cov}(C_2, C_1) = \sigma_{21} \\
\text{Cov}(C_2, C_2) = \sigma_{22}
\]

regardless of the bar, laboratory or position.

The error matrix \( E \) will then have the matrix \( R \) \((n_c \times n_c)\) copied into every \( n_c \times n_c \) block of \( E \) \((n \times n)\). If this were the case then \( E \) would be set up as

\[
E = J_{n_p} \otimes J_{n} \otimes J_{n_p} \otimes R = J_{n_b \times n_b} \otimes R.
\]
Figure 3.1: Layout of how the samples from a single bar were distributed. C1 and C2 denote the content of two chemical elements in a particular sample.

or equivalently

\[ \text{Cov}(e_{bi}, e_{bi'}) = R, \quad \text{for all } b, i, p, b', i'. \]

This seems a natural set up for the matrix E but it will be shown that it is in fact unsuitable.

### 3.2 Non-singularity of the matrix Σ

By definition a covariance matrix is positive-semi-definite. However, for the existence of Σ⁻¹ the matrix Σ must be positive-definite or non-singular. The result of the matrix Σ not being invertible is a degenerate multivariate normal distribution which does not possess a density function. The matrix Σ is positive-definite if the quadratic form satisfies \( x^T \Sigma x > 0 \) for every \( x \neq 0 \). The positive-definiteness of the matrix Σ will now be investigated.

Expanding \( x^T \Sigma x \) gives

\[
x^T \Sigma x = x^T(ZBZ^T + E)x = x^T(ZBZ^T)x + x^TEx.
\]

To show that \( x^T \Sigma x = 0 \) for some \( x \), there must exist an \( x \) for which both \( x^T(ZBZ^T)x = 0 \) and \( x^TEx = 0 \). Let \( x = [x_1 \ x_2 \ \ldots \ x_n] \), then

\[
x^TEx = [x_1 \ x_2 \ \ldots \ x_n] \begin{bmatrix} R & \cdots & R \ \vdots & \ddots & \vdots \ R & \cdots & R \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.
\]
The vector \( \mathbf{x} \) can then be partitioned into \( [\mathbf{x}_1^T \mathbf{x}_2^T \ldots \mathbf{x}_n^T]^T \) where \( \mathbf{x}_1 = [x_1 \ldots x_\alpha]^T \), \( \mathbf{x}_2 = [x_{\alpha+1} \ldots x_{2\alpha}]^T \) etc., and \( \alpha = \frac{n_\alpha}{n_e} \). Now,

\[
x^T \mathbf{E} \mathbf{x} = [\mathbf{x}_1^T \mathbf{x}_2^T \ldots \mathbf{x}_n^T] \begin{bmatrix} R & \cdots & R \\ \vdots & \ddots & \vdots \\ R & \cdots & R \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} = \sum_{i=1}^{\alpha} \sum_{j=1}^{\alpha} \mathbf{x}_i^T R \mathbf{x}_j = \sum_{i=1}^{\alpha} \mathbf{x}_i^T R \mathbf{x}_i + \sum_{i=1}^{\alpha} \sum_{j \neq i} \mathbf{x}_i^T R \mathbf{x}_j. \tag{3.5}
\]

Now, if we set \( \mathbf{x}_1 = -\mathbf{x}_2 = -\mathbf{x}_{n_\alpha+1} = \mathbf{x}_{n_\alpha+2} \) and all the other \( \mathbf{x}_j \) equal to 0, then \( x^T \mathbf{E} \mathbf{x} = 0 \). Also, as a consequence of this selection for the vector \( \mathbf{x} \), \( x^T (Z B Z)^T \mathbf{x} = 0 \). This is because for each matrix in equation (3.4), for example \( \sigma_B^2 (\mathbf{I}_{n_b} \otimes \mathbf{J}_{n_i} \otimes \mathbf{J}_{n_p} \otimes \mathbf{J}_{n_e}) \),

\[
x^T (\mathbf{I}_{n_b} \otimes \mathbf{J}_{n_i} \otimes \mathbf{J}_{n_p} \otimes \mathbf{J}_{n_e}) \mathbf{x} = 0.
\]

It has been shown that \( \mathbf{E} \), \( Z B Z^T \) and hence \( \Sigma \) are all singular.

The correlation between two errors corresponding to the same chemical, but from different positions, different laboratories or different bars, is 1. For example, \( \text{Cov}(e_{blp,c}, e_{blp',c}) = \sigma_{c c} \) (for \( p \neq p' \)) and \( \text{V}(e_{blp,c}) = \text{V}(e_{blp',c}) = \sigma_{c c} \), giving \( \text{Corr}(e_{blp,c}, e_{blp',c}) = 1 \). Similarly,

\[
\text{Corr}(e_{blp,c}, e_{blp',c}) = \text{Corr}(e_{blp,c}, e_{blp',c}) = 1, \quad \text{if } b \neq b', l \neq l'.
\]

So, the reason that \( \mathbf{E} \) is singular is that the restrictions placed on the covariances between chemicals, for different bars, laboratories or positions, are too tight. This is demonstrated graphically in Fig. 3.2(a) using an example where \( n_b = 4 \), \( n_i = 3 \), \( n_p = 2 \), \( n_c = 2 \). Here the correlation matrix of the errors is \( 48 \times 48 \) and a solid square in the matrix denotes a value equal to 1 whereas an unfilled square represents a value less than 1.

One possible solution is to introduce down-weighting parameters into the \( \mathbf{J} \) matrices. If we replace the \( \mathbf{J} \)s with new matrices, for example, by \( \mathbf{L}_{n_b}^{(\phi_1)} = (1-\phi_1)\mathbf{I}_{n_b} + \phi_1 \mathbf{J}_{n_b} \), we obtain an \( n_b \times n_b \) matrix with 1s along the diagonal and \( \phi_1 \) off the diagonal. Now we take

\[
\mathbf{E} = \mathbf{L}_{n_b}^{(\phi_1)} \otimes \mathbf{L}_{n_i}^{(\phi_2)} \otimes \mathbf{L}_{n_p}^{(\phi_3)} \otimes \mathbf{R} = \mathbf{L} \otimes \mathbf{R}. \tag{3.6}
\]

37
with $0 \leq \phi_1, \phi_2, \phi_3 < 1$. The covariances between the errors $e_{\text{bpc}}$ and $e_{\text{b'l'c'}}$ are listed in Table 3.1. Figure 3.2(b) shows that the correlation matrix of the errors now has 1s along the diagonal and values less than 1 elsewhere. The non-singularity of $E$ will now be investigated. The matrix $E$ can be shown to be non-singular if it can be shown that $\det(E) \neq 0$. The error matrix of equation (3.6) can be rewritten as

$$E = \begin{bmatrix}
F & \phi_1 F & \phi_1 F & \ldots & \phi_1 F \\
\phi_1 F & F & \phi_1 F & \ldots & \phi_1 F \\
\phi_1 F & \phi_1 F & F & \ldots & \phi_1 F \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\phi_1 F & \phi_1 F & \phi_1 F & \ldots & F
\end{bmatrix}$$

where

$$F = \begin{bmatrix}
G & \phi_2 G & \phi_2 G & \ldots & \phi_2 G \\
\phi_2 G & G & \phi_2 G & \ldots & \phi_2 G \\
\phi_2 G & \phi_2 G & G & \ldots & \phi_2 G \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\phi_2 G & \phi_2 G & \phi_2 G & \ldots & G
\end{bmatrix} \quad ; G = \begin{bmatrix}
R & \phi_3 R & \phi_3 R & \ldots & \phi_3 R \\
\phi_3 R & R & \phi_3 R & \ldots & \phi_3 R \\
\phi_3 R & \phi_3 R & R & \ldots & \phi_3 R \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\phi_3 R & \phi_3 R & \phi_3 R & \ldots & R
\end{bmatrix}.$$
Table 3.1: Covariances between the errors $e_{b'c'}$ and $e_{V^p'}$.

<table>
<thead>
<tr>
<th>$\text{Cov}(e_{b'c'}, e_{V^p'})$</th>
<th>$b, l, p, b', l', p'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{b'c'}$</td>
<td>$b = b', l = l', p = p'$</td>
</tr>
<tr>
<td>$\phi_1\sigma_{b'c'}$</td>
<td>$b = b', l = l', p = p'$</td>
</tr>
<tr>
<td>$\phi_2\sigma_{b'c'}$</td>
<td>$b = b', l = l', p = p'$</td>
</tr>
<tr>
<td>$\phi_3\sigma_{b'c'}$</td>
<td>$b = b', l = l', p = p'$</td>
</tr>
<tr>
<td>$\phi_1\phi_2\sigma_{b'c'}$</td>
<td>$b = b', l = l', p = p'$</td>
</tr>
<tr>
<td>$\phi_1\phi_3\sigma_{b'c'}$</td>
<td>$b = b', l = l', p = p'$</td>
</tr>
<tr>
<td>$\phi_2\phi_3\sigma_{b'c'}$</td>
<td>$b = b', l = l', p = p'$</td>
</tr>
</tbody>
</table>

Using the rule

$$\det\begin{bmatrix} A & B & B & \ldots & B \\ B & A & B & \ldots & B \\ B & B & A & \ldots & B \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ B & B & B & \ldots & A \end{bmatrix} = (\det(A - B))^{m-1} \det[A + (m-1)B],$$

the determinant of the matrix $E$ can be evaluated as

$$\det(E) = (1 + (n_b - 1)\phi_1)^{n_b}(1 + (n_l - 1)\phi_2)^{n_l}(1 + (n_p - 1)\phi_3)^{n_p}(1 - \phi_1)^{n_b(n_b-1)}(1 - \phi_2)^{n_l(n_l-1)}(1 - \phi_3)^{n_p(n_p-1)}(\det(R))^n.$$

For the bars, laboratories and positions there are at least two of each and so $n_b > 1$, $n_l > 1$ and $n_p > 1$. The down-weighting parameters were all defined to take some value less than unity, so $\det(E) \neq 0$ if $\det(R) \neq 0$. This has proved that if $R$ is non-singular then so are the matrices $E$ and $\Sigma$.

Positive-definiteness, and therefore non-singularity, of $R$ can easily be guaranteed by setting $R$ up as $AA^T$, where the matrix $A$ is lower triangular and $A \neq 0$. For example, in the case of $n_c = 2$, $R = \begin{bmatrix} r_1 & 0 & 0 \\ r_2 & r_3 & \end{bmatrix}$ which is now dependent on $r_1$, $r_2$ and $r_3$. The number of independent parameters needed to construct $R$ is $\frac{1}{2}n_c(n_c + 1)$ and they are denoted by the vector $r = (r_1, r_2, r_3, \ldots)$.

The parameters to be estimated in our new model are now the fixed effects $\mu$, $\alpha^p$, $\alpha^c$ and $\alpha^{PC}$, the variance components of the random effects $\sigma_B$, $\sigma_L$, $\sigma_{BC}$, $\sigma_{BP}$, $\sigma_{LG}$ and $\sigma_{BPC}$, the down-weighting parameters $\phi_1$, $\phi_2$ and $\phi_3$ and the $\frac{1}{2}n_c(n_c + 1)$
parameters which govern the covariance matrix \( R, r_1, \ldots, r_{\frac{1}{2} n_0(n_0+1)} \). Maximisation of the log-likelihood is carried out as in the previous situation to yield maximum likelihood estimates of the parameters.

### 3.2.1 Near singularity of \( \Sigma \)

The iterative procedure to locate a maximum of the Gaussian log-likelihood function of equation (2.6) requires that the covariance matrix be non-singular at each iteration. This constraint is satisfied automatically with the modifications made to the covariance structure of Section 3.2. However, it is possible that the matrix \( \Sigma \) can become near-singular if any of the down-weighting parameters in the covariance structure are close to unity. If this does occur then the Gaussian log-likelihood function becomes unstable. Jennrich et al remarked that if the covariance matrix nears singularity then the iterative maximisation of the Gaussian log-likelihood can become slow.

### 3.3 Example

In the experiment of Example 2.6, even though the analysis only took into account the content of silicon in the alloy, the laboratories measured the content of a number of chemical elements. This example is a continuation of that example, now with measurements on the content of aluminium. This data are shown in Table 3.2 and consist of the measurements recorded on silicon plus measurements made on aluminium. Plotted in Fig. (3.3) are the measurements recorded on the content of aluminium in the alloy. From this plot not much can be determined, such as the possible identification of an outlying laboratory. The model to be fitted to this data is that of equation (3.2)

\[
y = \mu + X_1 \alpha^P + X_2 \alpha^C + X_3 \alpha^{PC} + Z_1 a^B + Z_2 a^L + Z_3 a^{BC} + Z_4 a^{BP} + Z_5 a^{LC} + Z_6 a^{BPC} + e
\]

with the covariance structure of the errors set up as in Section 3.2.
Table 3.2: The percentage content of silicon and aluminium in four bars measured at two positions by three laboratories.

<table>
<thead>
<tr>
<th></th>
<th>Laboratory 1</th>
<th>Bar I</th>
<th></th>
<th>Bar II</th>
<th></th>
<th>Bar III</th>
<th></th>
<th>Bar IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Top</td>
<td>Bottom</td>
<td>Top</td>
<td>Bottom</td>
<td>Top</td>
<td>Bottom</td>
<td>Top</td>
<td>Bottom</td>
</tr>
<tr>
<td>Si</td>
<td>0.490</td>
<td>0.490</td>
<td>0.500</td>
<td>0.510</td>
<td>0.480</td>
<td>0.490</td>
<td>0.480</td>
<td>0.510</td>
</tr>
<tr>
<td>Al</td>
<td>2.950</td>
<td>2.930</td>
<td>2.930</td>
<td>2.920</td>
<td>2.830</td>
<td>2.860</td>
<td>2.920</td>
<td>2.900</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Laboratory 2</th>
<th>Bar I</th>
<th></th>
<th>Bar II</th>
<th></th>
<th>Bar III</th>
<th></th>
<th>Bar IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Top</td>
<td>Bottom</td>
<td>Top</td>
<td>Bottom</td>
<td>Top</td>
<td>Bottom</td>
<td>Top</td>
<td>Bottom</td>
</tr>
<tr>
<td>Si</td>
<td>0.467</td>
<td>0.476</td>
<td>0.459</td>
<td>0.477</td>
<td>0.450</td>
<td>0.485</td>
<td>0.454</td>
<td>0.480</td>
</tr>
<tr>
<td>Al</td>
<td>2.930</td>
<td>2.930</td>
<td>2.930</td>
<td>2.930</td>
<td>2.970</td>
<td>2.910</td>
<td>2.900</td>
<td>2.910</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Laboratory 3</th>
<th>Bar I</th>
<th></th>
<th>Bar II</th>
<th></th>
<th>Bar III</th>
<th></th>
<th>Bar IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Top</td>
<td>Bottom</td>
<td>Top</td>
<td>Bottom</td>
<td>Top</td>
<td>Bottom</td>
<td>Top</td>
<td>Bottom</td>
</tr>
<tr>
<td>Si</td>
<td>0.490</td>
<td>0.500</td>
<td>0.470</td>
<td>0.470</td>
<td>0.500</td>
<td>0.480</td>
<td>0.480</td>
<td>0.510</td>
</tr>
<tr>
<td>Al</td>
<td>2.980</td>
<td>2.980</td>
<td>2.970</td>
<td>2.970</td>
<td>2.960</td>
<td>2.940</td>
<td>2.970</td>
<td>2.970</td>
</tr>
</tbody>
</table>

The matrix X is $48 \times 4$ and has as it rows one of the following combinations,

- Chemical Si, Position Top: $(1, 1, 1, 1)$;
- Chemical Si, Position Bottom: $(1, -1, 1, -1)$;
- Chemical Al, Position Top: $(1, 1, -1, -1)$;
- Chemical Al, Position Bottom: $(1, -1, -1, 1)$.

The first column is a constant term 1 and the following three columns are position, chemical and 'position x chemical' indicators with column 4 being the product of columns 2 and 3.

The matrix Z is set up as $[Z_1 \ldots Z_6]$ and consists entirely of 0s and 1s. The matrix $Z_1$ is the design matrix of the bar effects and has rows $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$ and $(0, 0, 0, 1)$ corresponding to bars 1, 2, 3 and 4 respectively. Similarly $Z_2$ is the design matrix of the laboratory effects and has rows $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$ corresponding to the three laboratories. The design matrices for the interactions $Z_3 \ldots Z_6$, are then formed by taking pairwise products of the bar, laboratory and position indicators.

First, the maximum likelihood estimates of the fixed effects are

$$(\mu, \delta^P_1, \delta^C_1, \delta^{PC}_1) = (1.7066, -4.1250 \times 10^{-3}, -1.2221, -3.7083 \times 10^{-3})$$
with standard errors \((0.0094, 0.0019, 0.0075, 0.0014)\) calculated directly from the diagonal elements of \(\hat{V}_a\). In Example 2.6, the estimates of the fixed effects were able to be evaluated without assuming values for the variance components. Now, however, due to the complex structure which the covariance matrix \(\Sigma\) takes, this is not the case.

Maximisation of the log-likelihood function yields estimates

\[
(\sigma_L, \sigma_{LC}, \sigma_{BPC}) = (7.632 \times 10^{-3}, 1.575 \times 10^{-3}, 3.660 \times 10^{-3});
\]

\[
\sigma_H = \sigma_{HP} = \sigma_{HC} = 0.000;
\]

\[
(\phi_1, \phi_2, \phi_3) = (0.000, 0.157, 0.738), \quad A = 10^{-2} \begin{bmatrix} 1.268 & 0 \\ 1.023 & 2.472 \end{bmatrix}.
\]

Once again, the parameters estimated as zero by the maximisation process are removed from the model before the analysis continues. Tests on each of the fixed
effects can be performed by calculating the normal deviate for each. The estimate of the position effect is $-4.1250 \times 10^{-3}$ with standard error 0.0019, yielding a normal deviate of 2.12, suggesting that the inclusion of the position effect is necessary. The coefficient for the chemical effect is $-1.222$ with a standard error of 0.0075, giving a normal deviate of 162.091 which is, as expected, highly significant, supporting the need for a chemical effect. As far as the 'position x chemical' interaction is concerned, estimated as $-3.708 \times 10^{-3}$ with a standard error of 0.0014, results in a normal deviate of 2.69 which supports the need for the inclusion of this interaction. This finding is surprising: there seems to be a consistent difference in the chemical content between the top and bottom of the bars.

Testing whether the laboratory and 'laboratory x chemical' interaction are necessary is done by considering the logs of these two variance components. The estimates of the the laboratory and 'laboratory x chemical' effects are $7.632 \times 10^{-3}$ and $1.575 \times 10^{-2}$ and have as their standard errors $1.371 \times 10^{-2}$ and $8.375 \times 10^{-3}$ respectively. The estimates of these log variance components are $\log(\sigma_L) = -4.875$ and $\log(\sigma_{LC}) = -4.151$, with standard errors of 1.810 and 0.530 respectively. Confidence intervals (95%) for $\log(\sigma_L)$ and $\log(\sigma_{LC})$ are computed as

$$-8.423 < \log(\sigma_L) < -1.327, \quad -5.190 < \log(\sigma_{LC}) < -3.112$$

suggesting that the 'laboratory' and 'laboratory x chemical' interaction components are non zero and should be included in the model. This is evidence that there exists a real difference between laboratories and that the chemical effect varies randomly from laboratory to laboratory.

The fact that the estimates of $\sigma_B$, $\sigma_{BP}$ and $\sigma_{BP}$ are all zero suggests that the inclusion of these variance components in the model is unnecessary. The estimate of the 'bar x position x chemical' interaction is, however, non-zero. A 95% confidence interval for the log of this component is

$$-7.321 < \log(\sigma_{BPC}) < -3.900$$

which suggests that even though there is no evidence of a difference between bars the 'position x chemical' interaction varies randomly from bar to bar.

The target percentage content of the chemicals silicon and aluminium were 0.5% and 3.0% respectively. It may be of interest to investigate whether the content of
these two chemicals measured by the laboratories meets the target specification. For silicon this can be achieved by testing the hypothesis \( H : \mu + \alpha_1^p + \alpha_1^q + \alpha_{11}^{PC} = 0.5 \) and \( \mu - \alpha_1^p + \alpha_1^q - \alpha_{11}^{PC} = 0.5 \), i.e. testing whether the content of silicon in both the top and bottom of the bar is 0.5%. The corresponding hypothesis for aluminium is \( H : \mu + \alpha_2^p - \alpha_2^q - \alpha_{12}^{PC} = 3.0 \) and \( \mu - \alpha_2^p - \alpha_2^q + \alpha_{12}^{PC} = 3.0 \). The hypothesis regarding silicon is equivalent to testing the hypothesis \( H_0 : \mu = 0.5 \). The Wald test statistic is then

\[
W = (H\hat{\alpha} - h)^T(\hat{V}_\alpha H^T)^{-1}(H\hat{\alpha} - h)
\]

which is asymptotically distributed as \( \chi^2_4 \) (\( \dim(\alpha) = 4 \)). In this case the value of \( W \) is 18.8, which when referred to \( \chi^2_4 \) suggests that the content of silicon in the alloy is not 0.5%.

To test if the percentage content of aluminium met its pre-defined target of 3.0% is done by testing the hypothesis \( H\alpha = h \), where

\[
H = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}, \quad h = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}
\]

which yields a value \( W = 26.7 \), suggesting that the aluminium content does not meet its target of 3.0%. The normal plot of the standardised residuals is shown in Fig. 3.4: the lack of linearity of this graph suggests negative skewness in the standardised error distribution.

### 3.4 Inversion of \( \Sigma \) for many chemicals

As mentioned earlier, more realistic situations will involve considerably more than two chemicals. One of the problems which arises when dealing with this amount of data is the size of the covariance matrix. In the case of four bars, three laboratories, two positions and ten chemicals there will be 240 measurements recorded, meaning that \( \Sigma \) will be a \( 240 \times 240 \) matrix of covariances. As the number of chemicals increases computation of \( \Sigma \) will become increasingly costly and so the maximisation of the log-likelihood function, which is carried out by an iterative procedure, soon becomes computationally demanding since the inversion of the covariance matrix is required at each iteration.
Figure 3.4: Normal plot of the standardised residuals. Measurements recorded by three laboratories on the content of 2 chemical elements in the top and bottom of four alloy bars.
3.4.1 Direct numerical computation of $\Sigma^{-1}$

Many numerical methods exist for finding the inverse of a matrix such as $\Sigma$. One method of matrix inversion which is particularly efficient in the field of regression is that of Cholesky Symmetric Decomposition since it requires the matrix to be symmetric and non-singular. The matrix $\Sigma$ is constructed to be positive-definite and so can be written as $\Sigma = PP^T$, where the matrix $P$ is a lower triangular matrix. The matrix $P$ can easily be evaluated by Cholesky decomposition of the matrix $\Sigma$. The elements of the matrix $P$ satisfy the equation

$$\sum_{k=1}^{\min(i,j)} P_{ki}P_{kj} = \Sigma_{ij}$$

where $\Sigma_{ij}$ is the $(i,j)$th entry in the matrix $\Sigma$.

The elements of the matrix $P$ are then given recursively by the equations

$$P_{11} = \Sigma_{11}^{\frac{1}{2}};$$
$$P_{1j} = \frac{\Sigma_{1j}}{P_{11}}, \quad \text{for} \quad j = 2, 3, \ldots, n;$$
$$P_{ij} = (\Sigma_{ij} - \sum_{k=1}^{i-1} P_{ki}^2)^{\frac{1}{2}}, \quad \text{for} \quad i = 2, 3, \ldots, n;$$
$$P_{ij} = P_{ii}^{-1}(\Sigma_{ij} - \sum_{k=1}^{i-1} P_{ki}P_{kj}), \quad \text{for} \quad j = i + 1, \ldots, n;$$

When inverting $\Sigma$ directly from the lower triangular matrix $P$ the number of arithmetic operations performed is $\frac{2}{3}n^3 + \frac{1}{2}n^2 + \frac{5}{6}n$, where $n$ is the total number of measurements recorded. So in the case of 240 measurements this results is $9.245 \times 10^6$ arithmetic operations.

3.4.2 Algebraic computation of $\Sigma^{-1}$ and det ($\Sigma$)

The matrix operations used in Section 2.3.1 to find algebraic expressions for the inverse and determinant of the covariance matrix $\Sigma$ may be applied to the covariance structure given in equation (3.4). However, due to the more complex structure of this covariance matrix finding expressions for $\Sigma^{-1}$ and det ($\Sigma$) this way can become quite laborious.

A second approach for finding algebraic expressions for the inverse and determinant of the covariance matrix $\Sigma$ is as follows. As before the covariance matrix of the
observations is \( \Sigma = ZBZ^T + E \). The inverse and determinant of \( \Sigma \) can then be written as (Rao, 1965),

\[
\Sigma^{-1} = E^{-1} - E^{-1}Z(Z^TE^{-1}Z + B^{-1})^{-1}Z^TE^{-1};
\]

Now, because the matrix \( B \) is diagonal (equation (3.3)), its inverse and determinant can be computed very simply.

The matrix \( E \) was designed to have the structure

\[
E = L_{n_b}^{(\phi_1)} \otimes L_{n_i}^{(\phi_2)} \otimes L_{n_p}^{(\phi_3)} \otimes R
\]

where the \( L \) matrices were of the form

\[
L_{n_b}^{(\phi_1)} = (1 - \phi_1)I_{n_b} + \phi_1J_{n_b}
\]

and at present the only restriction placed upon \( R \) is that it is positive-definite. Provided that \( R \) is non-singular, and using the standard formula

\[
(A \otimes B)^{-1} = A^{-1} \otimes B^{-1},
\]

the inverse of the matrix \( E \) is then found to be

\[
E^{-1} = (L_{n_b}^{(\phi_1)})^{-1} \otimes (L_{n_i}^{(\phi_2)})^{-1} \otimes (L_{n_p}^{(\phi_3)})^{-1} \otimes R^{-1}
\]

where the inverses of the \( L \) matrices satisfy

\[
(L_{n_b}^{(\phi_1)})^{-1} = \frac{1}{1 - \phi_1}I_{n_b} + \frac{\phi_1}{(1 - \phi_1)(n_b\phi_1 - \phi_1 - 1)}J_{n_b}.
\]

Thus, the matrix \( E \) can be inverted with only the matrix \( R \) having to be evaluated numerically. It will be seen in subsequent chapters that when some structure is applied to the matrix \( R \) then the inverse of \( R \), and hence the inverse of \( E \), will be able to be evaluated entirely algebraically.

In most cases, especially when there is only a small number of random higher order interaction terms in the model, and in particular when \( q \) is significantly smaller than \( n \), the matrix \( T = (Z^TE^{-1}Z + B^{-1}) \) will be considerably smaller than the matrix \( \Sigma \). When this is the case it may prove computationally effective to evaluate
\( \Sigma^{-1} \) algebraically using equation (3.7). If numeric inversion of the matrix \( T \) is to be avoided, then equation (3.7) could be re-applied to \( T = (Z^T \Sigma^{-1} Z + B^{-1}) \).

The determinant of the matrix \( \Sigma \) may be found as

\[
det(\Sigma) = \{\det(L_{n_b}^{(\phi_1)})\}^{(n_t-n_p-n_c)}\{\det(L_{n_l}^{(\phi_2)})\}^{(n_t-n_p-n_c)}
\times\{\det(L_{n_p}^{(\phi_3)})\}^{(n_t-n_m-n_c)}\{\det(R)\}^{(n_b-n_l-n_p)} \tag{3.9}
\]

with the determinants of the L matrices satisfying

\[
det(L_{n_b}^{(\phi_1)}) = (1 - \phi_1)^{n_b-1}\{1 + (n_b - 1)\phi_1\}.
\]

### 3.4.3 Avoiding direct numerical computation of \( \Sigma^{-1} \)

A method will now be described which avoids the problem of continually inverting the covariance matrix. Consider equation (2.12),

\[
\hat{\alpha} = (X^T \Sigma^{-1} X)^{-1}X^T \Sigma^{-1} y.
\]

The estimates of the fixed effects, \( \hat{\alpha} \), can be calculated following these steps:

1: Solve \( y = \Sigma g \) for \( g \).
2: Solve \( X = \Sigma G \) for \( G \).
3: Solve \( X^T g = (X^T G)\hat{\alpha} \) for \( \hat{\alpha} \).

The system of equations in step 1 can be solved efficiently using the fact that the matrix \( \Sigma \) is positive-definite. Using Cholesky decomposition of the matrix \( \Sigma \) a lower triangular matrix \( P \) can be computed such that \( \Sigma = PP^T \). Using this, the equation \( y = \Sigma g \) becomes

\[
y = Px \tag{3.10}
\]

where

\[
x = P^T g. \tag{3.11}
\]
Now, since the matrices $P$ and $P^T$ are both triangular the two sets of equations, (3.10) and (3.11), can be solved efficiently for $\mathbf{x}$ and $\mathbf{g}$ respectively. The elements of the vector $\mathbf{x}$ in equation (3.10) are evaluated using the equations

$$x_i = \frac{y_i - \sum_{k=i+1}^{n} P_{ik} x_k}{P_{ii}}, \quad \text{for } i = 1, \ldots, n$$

and the elements of $\mathbf{g}$ and hence $\Sigma^{-1}y$ are found from the set of equations

$$g_i = \frac{x_i - \sum_{k=1}^{i-1} P_{ki} x_k}{P_{ii}}, \quad \text{for } i = 1, \ldots, n.$$

A similar approach may be used to solve the equations in steps 2 and 3. Solving the equation $\mathbf{X} = \Sigma \mathbf{G}$ for $\mathbf{G}$ is achieved by partitioning the matrix $\mathbf{X}$ as $[\mathbf{x}_1 \mathbf{x}_2 \ldots \mathbf{x}_l]$ and $\mathbf{G}$ as $[\mathbf{g}_1 \mathbf{g}_2 \ldots \mathbf{g}_l]$ and then solving each of the sets of equations $\mathbf{x}_i = \Sigma \mathbf{g}_i$ for $\mathbf{g}_i$ where $i = 1, 2, \ldots, l$.

The reduced log-likelihood function of equation (2.13),

$$L_1(\mathbf{y}; \theta) = -\frac{1}{2} \log \det(2\pi \Sigma) - \frac{1}{2}(\mathbf{y} - \mathbf{X} \hat{\alpha})^T \Sigma^{-1}(\mathbf{y} - \mathbf{X} \hat{\alpha})$$

also contains the inverse of $\Sigma$. Let $\mathbf{r} = \mathbf{y} - \mathbf{X} \hat{\alpha}$, and then setting $\mathbf{g} = \Sigma^{-1} \mathbf{r}$ or $\mathbf{r} = \Sigma \mathbf{g}$ and solving for $\mathbf{g}$ as before again avoids the problem of requiring the inverse of the covariance matrix.

**Number of arithmetic operations ($+, \times, \backslash$)**

In this section it will be shown that the number of arithmetic operations performed in computing $\Sigma^{-1} \mathbf{y}$ and $\Sigma^{-1} \mathbf{X}$ using the method described in Section 3.4.3 is considerably less than if $\Sigma^{-1}$ had been computed directly.

Consider the set of equations required to solve $\mathbf{y} = \mathbf{P} \mathbf{x}$ for $\mathbf{x}$,

$$x_i = \frac{y_i - \sum_{k=i+1}^{n} P_{ik} x_k}{P_{ii}}, \quad \text{for } i = 1, \ldots, n.$$

The number of arithmetic operations required to solve this system of equations is as follows:
1. Number of divisions : \( \#(div) = n \).

2. Number of additions : \( \#(add) \).
   \[ \#(add) = \sum_{i=1}^{n} \{ \sum_{k=i+1}^{n} 1 \} = \frac{1}{2} n(n - 1). \]

3. Number of multiplications : \( \#(mul) \)
   \[ \#(mul) = \#(add) = \frac{1}{2} n(n - 1). \]

4. Total no. of operations : \( \#(tot) = n^2 \).

Since two sets of linear equations have to be solved to compute \( \Sigma^{-1} y \) the total number of operations for that task is \( 2n^2 \). In the case of 240 measurements this results in \( 1.152 \times 10^5 \) operations. In general, this method of computing \( \Sigma^{-1} y \) requires approximately \( \frac{3}{n} \) times the number operations that are required when \( \Sigma \) is inverted directly. The number of operations performed in calculating \( \Sigma^{-1} X \) is \( 2pn^2 \) where the matrix \( X \) is of size \( n \times p \). It is possible, but unlikely, that \( p > n \). If this does occur then inversion of the matrix \( \Sigma \) is best performed numerically.
Chapter 4

Covariance structures

The covariance matrix of the data is of the form

$$\Sigma = \text{Cov}(y) = ZBZ^T + E$$

where $Z$ is the design matrix for the random effects, $B$ is the covariance matrix of the random effects and $E$ is the covariance matrix of the errors set up as $E = L \otimes R$. Here, the matrix $L$ was introduced to ensure that the error matrix $E$ and hence the covariance matrix $\Sigma$ are non-singular.

Recall that the matrix $R$ is the $n_c \times n_c$ matrix which describes the covariances between the measurements recorded on a single sample of the alloy. Since the only restriction placed on the matrix $R$ is that it should be symmetric positive-definite it is therefore constructed using $\frac{1}{2}n_c(n_c+1)$ functionally independent elements, where $n_c$ is the number of chemicals in the study. Now, as the number of chemicals in the study increase, the number of parameters on which the matrix $R$, and hence the matrix $\Sigma$, depends also increases. Consider the model of equation (3.2)

$$y = 1 + X_1a^P + X_2a^C + X_3a^{PC} + Z_1a^B + Z_2a^L + Z_3a^{BC} + Z_4a^{BP} + Z_5a^{LC} + Z_6a^{BPC} + e.$$ 

The number of parameters needed to construct the covariance matrix $\Sigma$ in this model is $\frac{1}{2}n_c(n_c+1) + 9$. These are the $\frac{1}{2}n_c(n_c+1)$ parameters required for the matrix $R$, the 3 down-weighting parameters which are introduced through the matrix $L$ and the 6 variance components $\sigma^2_n$, $\sigma^2_L$ etc. which form the matrix $B$. Also included in the model are the fixed effects parameters which also have to be estimated: there are $n_p n_c$ of these.
In the example of Chapter 3 where, \( n_b = 4, n_t = 3, n_p = 2 \), and now assuming that the content of 20 chemical elements is of interest, i.e. \( n_e = 20 \), the number of parameters on which the matrix \( \Sigma \) depends is 219 and the number of fixed effects in the model is 40. Now, even if there is no missing data, and the data available form a balanced layout, there are 480 observations. In the analysis of these data it is the fixed effects which are of most interest. Therefore, letting the matrix \( R \), and hence the covariance matrix \( \Sigma \), depend on so many parameters may result in too little information being available to make meaningful inferences about the fixed effects. Another reason why it is undesirable to fit a model containing so many unknown parameters is that the maximisation of Gaussian log-likelihood functions, such as the one in equation (2.6), becomes time consuming and computationally demanding.

Apart from the data supplied, there may be some knowledge of how the experiment was carried out. This additional information may be able to be used to model the covariance structure of the errors in some realistic way so that the number of parameters on which it depends can be reduced. However, if too much structure is placed on the error matrix, or assumptions are made about it which are incorrect, then there may be model misspecification. Model misspecification may lead to apparently stronger, but biased, inferences on the regression parameters.

### 4.1 Some covariance structures

Let \( \epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_{ne})^T \) denote the vector of random errors corresponding to the \( ne \) measurements recorded on any sample of the alloy. The matrix \( R \) then has as its \((i, j)\)th entry \((R)_{ij} = \text{Cov}(\epsilon_i, \epsilon_j)\). A variety of covariance structures will be considered for the matrix \( R \) but listed in Table 4.1 are those of the more common structures.

One type of covariance structure, which is clearly not suitable in this experiment, is that of independent errors. If it were to be believed that the content of the chemicals measured on a specific experimental unit were uncorrelated then this structure could be applied. It is assumed that the matrix \( R \) is diagonal and has some value along the diagonal representing a common variance of the errors.
Table 4.1: A selection of different covariance structures.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Description</th>
<th>Number of Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstructured Errors</td>
<td>( R_{11} = r_1, \ R_{21} = r_2, \ R_{22} = r_3 \ldots )</td>
<td>( \frac{1}{2} n_e(n_e + 1) )</td>
</tr>
<tr>
<td>Independent Errors</td>
<td>( R_{ij} = 0 ) if ( i = j ) ( R_{ij} = \sigma^2 ) if ( i \neq j )</td>
<td>1</td>
</tr>
<tr>
<td>Equicorrelated Errors</td>
<td>( R_{ij} = r_i^2 ) if ( i = j ) ( R_{ij} = \rho r_i r_j ) if ( i \neq j )</td>
<td>( n_e + 1 )</td>
</tr>
</tbody>
</table>

Of the covariance structures listed in Table 4.1, that of Equicorrelated Errors which is the most realistic.

4.2 Equicorrelated structures

When the participating laboratories analysed their samples of the alloy, the measurements on a particular sample were not made in any significant time sequence, so serial correlation structure is not relevant. Bearing this in mind, the matrix \( R \) will be modelled so that the errors are equicorrelated. The variance of the \( i \)th error is assumed to be \( r_i^2 \) and each pair of errors is assumed to have equal correlation, i.e. the covariance between the two errors \( \epsilon_i \) and \( \epsilon_j \) is \( \rho r_i r_j \) where \( \rho \) is the correlation coefficient. The matrix \( R \) is set up so that it has the parameters \( r_i^2 \) along the diagonal and the elements off the diagonal are set to \( \rho r_i r_j \), where \(-1 < \rho < 1\) and \( i, j = 1, 2, \ldots, n_e \). The matrix \( R \) is

\[
R = \begin{bmatrix}
r_1^2 & \rho r_1 r_2 & \rho r_1 r_3 & \cdots & \rho r_1 r_{n_e} \\
\rho r_2 r_1 & r_2^2 & \rho r_2 r_3 & \cdots & \rho r_2 r_{n_e} \\
\rho r_3 r_1 & \rho r_3 r_2 & r_3^2 & \cdots & \rho r_3 r_{n_e} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho r_{n_e} r_1 & \rho r_{n_e} r_2 & \rho r_{n_e} r_3 & \cdots & r_{n_e}^2
\end{bmatrix}
\]

and can be more conveniently written as \( R = V W V \), where

\[
V = \text{diag}(r_1, r_2, \ldots, r_{n_e}), \quad W = (1 - \rho)I_{n_e} + \rho J_{n_e}.
\]

When the matrix \( R \) possesses an equicorrelated structure it is dependent on \( n_e + 1 \) parameters rather than \( \frac{1}{2} n_e(n_e + 1) \) parameters when no structure is applied.
The inverse and determinant of the error matrix $E$ can be computed using equations (3.8) and (3.9) in which the inverse and determinant of the matrix $R$ satisfy

$$R^{-1} = V^{-1}W^{-1}(V^{-1})$$

and

$$\det(R) = \det(W)\{\det(V)\}^2$$

with

$$V^{-1} = \text{diag}\left(\frac{1}{r_1}, \frac{1}{r_2}, \ldots, \frac{1}{r_{nc}}\right),$$

$$W^{-1} = \frac{1}{1-\rho}I_{nc} + \frac{\rho}{(1-\rho)(1-\rho+n_c\rho)}J_{nc},$$

$$\det(V) = \prod_{i=1}^{nc} r_i,$$

and

$$\det(W) = (1-\rho)^{nc-1}\{1 + (n_c - 1)\rho\}.$$ 

These equations for the inverse and determinant of $R$ can be used to evaluate the inverse and determinant of the covariance matrix $\Sigma$ algebraically.

### 4.3 Example

In this example the data used is that given in Table 3.2, for the content of Aluminium and Silicon, combined with the measurements made on eight further chemicals. The chemicals introduced in this example are Cobalt (10.0%), Chromium (5.0%), Copper (1.0%), Iron (5.0%), Manganese (1.0%), Molybdenum (1.0%), Tantalum (1.0%) and Titanium (1.0%). The percentages in parentheses are the target percentage content of these chemicals in the alloy. A full list of the measurements recorded can be found in Table A.1. The number of measurements recorded in this case is now 240. Plots of the eight chemical elements are given in Fig. 4.1. Looking at these graphs it seems that for several of them, e.g. Fig. 4.1(c), Fig.
4.1(d) and Fig. 4.1(g), corresponding to Copper, Iron and Tantalum, one of the laboratories which recorded measurements on the contents of these chemical elements does not agree with the other two. This may be an indication that the content of the chemical elements varies between laboratories.

The effects which should be included in a linear model are the same as in the case of two chemicals, see Section 3.1, and so a linear model to describe the data is, again, that of equation (3.2),

\[ y = 1 \mu + X_1 \alpha^P + X_2 \alpha^C + X_3 \alpha^{PC} + Z_1 a^B + Z_2 a^L + Z_3 a^{BC} + Z_4 a^{BP} + Z_5 a^{LC} + Z_6 a^{BPC} + e. \]

The matrix \( X \) is \( 240 \times 20 \) and has as its first column a constant term 1. The second column is the position indicator and columns 3 to 11 are the chemical indicators. Columns 12 to 20, the pairwise products of the position and chemical indicators, represents the 'position \times chemical' interaction indicators. So, for example, the rows of \( X \) corresponding to the four measurements \( y_{b11}, y_{b21}, y_{b19} \) and \( y_{b29} \), measurements made by laboratory \( l \) on Aluminium and Niobium at both the top and bottom of bar \( b \) are

\[
\begin{align*}
y_{b11} & : (1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0); \\
y_{b21} & : (1, -1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0); \\
y_{b19} & : (1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1); \\
y_{b29} & : (1, -1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, -1). \\
\end{align*}
\]

The matrix \( Z \) is set up as \( [Z_1 \ Z_2 \ \ldots \  Z_6] \), where \( Z_1, Z_2, \ldots, Z_6 \) have the same form as the \( Z \) matrices in Example 3.3. The matrix \( E = \text{Cov}(e) \) will take the form \( E = L^\phi_{nl} \otimes L^\phi_{nr} \otimes L^\phi_{nr} \otimes R \), where the matrix \( R \) has an equicorrelated structure as described in Section 4.2.

The set of parameters in the model which now have to be estimated are

\[
\begin{align*}
\alpha & = (\mu, \alpha^P_1, \alpha^C_1, \ldots) \quad \text{Fixed Effects}; \\
\sigma & = (\sigma_1, \sigma_2, \ldots) \quad \text{Variance Components}; \\
\phi & = (\phi_1, \phi_2, \ldots) \quad \text{Down-weighting parameters}; \\
r & = (\rho, r_1, \ldots, r_{nc}) \quad \text{Parameters which govern } R.
\end{align*}
\]

Fitting this model yielded estimates of zero for the parameters \( \sigma_B, \sigma_{BC} \) and \( \phi_1 \), which all lie on the boundary of the parameter space and so these parameters were
(a) Cobalt

(b) Chromium

(c) Copper

(d) Iron

(e) Manganese

(f) Molybdenum
Figure 4.1: Percentage contents of the 8 chemicals Co, Ch, Cu, Fe Mn, Mo, Ta and Ti measured by the three laboratories in the eight samples of the alloy it received. Key for x-axis: 1 - Bar 1, Top; 2 - Bar 1, Bottom etc.

removed from the model and the model re-fitted before any further analyses took place.

The estimate of the position effect was 0.0028 with a standard error of 0.0019. Testing if the position effect is zero gives a normal deviate of 1.51 suggesting that on its own the position effect is not indispensable. Testing if a ‘position × chemical’ interaction is necessary is equivalent to testing the hypothesis $\alpha_{11}^{PC} = \alpha_{12}^{PC} = \ldots = \alpha_{19}^{PC} = 0$ which can be performed by calculation of the Wald statistic. This hypothesis is equivalent to $H\hat{\alpha}^{PC} = h$ where $H = I_{(n_p-1)(n_c-1)}$ and $h = 0_{(n_p-1)(n_c-1)}$ is a vector of zeros of length $(n_p-1)(n_c-1)$. The test statistic is then

$$W = (H\hat{\alpha}^{PC} - h)^T(H\hat{V}_{PC}H^T)^{-1}(H\hat{\alpha}^{PC} - h)$$

where $\hat{\alpha}^{PC}$ are the estimates of the ‘position × chemical’ effects and $\hat{V}_{PC}$ is their estimated covariance matrix. The matrix $\hat{V}_{PC}$ is a sub-matrix of the complete estimated covariance matrix of the fixed effects $\hat{V}_a$ in equation (2.18). The test statistic is then asymptotically distributed as $\chi^2_{(n_p-1)(n_c-1)}$. The value of $W$ computed is 57.4 and is referred to $\chi^2_9$. This suggests that, even though a position effect is not necessary, a ‘position × chemical’ interaction effect is, i.e. there appear to be position effects for individual chemicals, but these average out to zero overall.
The observed chemical content of the chemical elements meets the target content is again performed using the Wald statistic. The test statistic is similar to that in Example 3.3 for the two-chemical example. The chemicals and the corresponding $W$ values are as follows; Aluminium (2.25), Copper (11.4), Chromium (40.2), Copper (1.09), Iron (8.60), Manganese (1.03), Molybdenum (6.38), Silicon (6.98), Tantalum (7.22) and Titanium (8.62). When these values are referred to tables of $\chi^2$ it suggests that the content of all the chemicals meet the pre-defined specification apart from Copper and Chromium.

The estimates of the set of parameters governing the covariance matrix $\Sigma$ are listed in Table 4.2. The second column of this table gives the maximum likelihood estimates of the parameters and the third column gives the standard errors of these estimates.

The laboratory and 'laboratory x chemical' components were estimated as $2.342 \times 10^{-2}$ and $7.825 \times 10^{-2}$, with standard errors $2.096 \times 10^{-2}$ and $1.096 \times 10^{-2}$ respectively. Confidence intervals for $\log(\sigma_L)$ and $\log(\sigma_{LC})$ are constructed as

$$-5.509 < \log(\sigma_L) < -2.000, \quad -2.822 < \log(\sigma_{LC}) < -2.273.$$

The zero estimates of the bar and 'bar x chemical' interaction effects seem to suggest that these effects are not present and are therefore not required in the model. However, the estimates of the non-zero interactions which included a bar effect are $\hat{\sigma}_{BP} = 5.951 \times 10^{-3}$ and $\hat{\sigma}_{BPC} = 2.365 \times 10^{-3}$ with standard errors $2.378 \times 10^{-3}$ and $1.283 \times 10^{-3}$ respectively. Confidence intervals for the logs of these two variance components containing a bar effect are as follows:

$$-5.907 < \log(\sigma_{BP}) < -4.340, \quad -7.110 < \log(\sigma_{BPC}) < -4.984.$$

Looking at these confidence intervals, there is nothing to indicate that these components are unnecessary and hence they suggest that position and 'position x chemical' interaction effects both vary randomly from bar to bar.

The plot of the standardised residuals is given in Fig. 4.2. and again can be used for the detection of outliers and to assess the goodness-of-fit of the model. It may be the case that the lone point in the lower left hand corner of this graph corresponds to an outlying measurement. Disregarding this point, it could be
Table 4.2: Maximum likelihood estimates of the variance components along with their associated standard errors.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLE</th>
<th>std. err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_L$</td>
<td>$2.342 \times 10^{-2}$</td>
<td>$2.095 \times 10^{-2}$</td>
</tr>
<tr>
<td>$\sigma_{LC}$</td>
<td>$7.825 \times 10^{-2}$</td>
<td>$1.096 \times 10^{-2}$</td>
</tr>
<tr>
<td>$\sigma_B$</td>
<td>0.000</td>
<td>---</td>
</tr>
<tr>
<td>$\sigma_{BP}$</td>
<td>$5.951 \times 10^{-3}$</td>
<td>$2.378 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\sigma_{BC}$</td>
<td>0.000</td>
<td>---</td>
</tr>
<tr>
<td>$\sigma_{BPC}$</td>
<td>$2.366 \times 10^{-3}$</td>
<td>$1.283 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.353</td>
<td>0.139</td>
</tr>
<tr>
<td>$r_1$</td>
<td>$6.573 \times 10^{-4}$</td>
<td>$2.159 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_2$</td>
<td>$1.595 \times 10^{-3}$</td>
<td>$4.755 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_3$</td>
<td>$6.344 \times 10^{-4}$</td>
<td>$1.945 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_4$</td>
<td>$5.822 \times 10^{-5}$</td>
<td>$2.409 \times 10^{-5}$</td>
</tr>
<tr>
<td>$r_5$</td>
<td>$2.232 \times 10^{-3}$</td>
<td>$6.970 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_6$</td>
<td>$7.709 \times 10^{-5}$</td>
<td>$4.001 \times 10^{-5}$</td>
</tr>
<tr>
<td>$r_7$</td>
<td>$4.809 \times 10^{-5}$</td>
<td>$1.792 \times 10^{-5}$</td>
</tr>
<tr>
<td>$r_8$</td>
<td>$1.211 \times 10^{-4}$</td>
<td>$4.058 \times 10^{-5}$</td>
</tr>
<tr>
<td>$r_9$</td>
<td>$1.209 \times 10^{-4}$</td>
<td>$4.094 \times 10^{-5}$</td>
</tr>
<tr>
<td>$r_{10}$</td>
<td>$2.247 \times 10^{-5}$</td>
<td>$8.643 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.000</td>
<td>---</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>$1.716 \times 10^{-2}$</td>
<td>0.111</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>0.271</td>
<td>0.116</td>
</tr>
</tbody>
</table>

assumed, that because of the near-linearity of this graph the model provides a good fit to the data. However, in the next section it will be shown that this is not the case.

### 4.4 Suitability of $R$

It is rather optimistic to believe that the error matrix for the chemical contents on a particular sample of alloy should be able to be modelled so simply by allowing the matrix $R$ to have an equicorrelated structure. In this section the suitability of this assumption is investigated. In the first approach a graphical comparison is made between the estimated correlation matrix when the matrix $R$ is restricted to be symmetric positive-definite and the case where it is equicorrelated. Secondly, two asymptotically equivalent tests, the likelihood ratio test and a score-test, are
constructed to investigate the suitability of an equicorrelated structure for $R$.

### 4.4.1 Graphical Indications

The linear model applied in Example 4.3 is fitted to the data of Table A.1 with the matrix $E$ still having the structure $E = L \otimes R$. Now, however, the assumption that the matrix $R$ has an equicorrelated structure is relaxed and so the only restriction is that it must be symmetric positive-definite. To ensure that the matrix $E$ remains non-singular the matrix $R$ is set up as $AA^T$, where $A$ is a lower triangular, non-zero matrix.

The parameters in the model are estimated in the usual way. These parameters are $\alpha$ and $\theta = (\sigma, \phi, \mathbf{r})$ where $\mathbf{r}$ is the vector of $\frac{1}{2}n_c(n_c + 1)$ parameters required to
construct the matrix $\mathbf{R}$. The estimated covariance matrix $\hat{\mathbf{R}}$ can then be constructed using the estimate of the vector $\mathbf{r}$ and the corresponding correlation matrix has as its $(i, j)$th entry $\frac{\hat{R}_{ij}}{\hat{R}_{ii}^{1/2}\hat{R}_{jj}^{1/2}}$ where $\hat{R}_{ij}$ is the $(i, j)$th element of $\hat{\mathbf{R}}$. This estimated correlation matrix, in the case where no structure has been applied to the matrix $\mathbf{R}$ is shown in Fig. 4.3(a). It is shown using different shades of grey: a black square denotes a correlation of 1 and a white square a correlation of 0; correlations between 0 and 1 are represented by different shades of grey, with darker shades corresponding to higher correlations.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{correlation_matrix.png}
\caption{(a) The correlation matrix when $\mathbf{R}$ is unstructured \hspace{1cm} (b) The correlation matrix when $\mathbf{R}$ is equicorrelated ($\rho = 0.353$)}
\end{figure}

Figure 4.3: Graphical representations of the correlation matrices of the errors corresponding to the content of 10 chemical elements in a particular sample of the alloy.

From Fig. 4.3(a) it seems to be the case that the estimated correlation matrix does not have an equicorrelated structure. Fig. 4.3(b) shows the corresponding equicorrelated case, as in Example 4.3. Comparing Fig. 4.3(a) and Fig. 4.3(b) highlights the problem that it may have been unreasonable to assume an equicorrelated structure for the matrix $\mathbf{R}$.

### 4.4.2 Likelihood Ratio Test

In the previous section some doubt was raised over the suitability of the equicorrelated structure which has been applied to the matrix $\mathbf{R}$. This was achieved by
using a graphical representation of the estimated correlation matrix of the errors, both in the case where the matrix $R$ possesses an equicorrelated structure and in the case where the matrix $R$ is unstructured. A more formal test to assess the suitability of the matrix $R$ is now described.

It has been assumed throughout that the data follows a multivariate normal distribution with the corresponding log-likelihood function given by equation (2.6),

$$L(y; \omega) = -\frac{1}{2} \log \det(2\pi \Sigma) - \frac{1}{2} (y - X\alpha)^T \Sigma^{-1} (y - X\alpha)$$

where $\omega = (\alpha, \theta)$ and $\theta = (\sigma, \phi, R)$ and $\Sigma = \Sigma(\theta)$.

Let $\hat{\omega}_R$ be the estimate of the parameter vector $\omega$ when the matrix $R$ takes the form of an equicorrelated structure. This will be referred to as the restricted maximum likelihood estimate of $\omega$. Similarly, $\hat{\omega}_U$ will denote the unrestricted maximum likelihood estimate of the parameter vector $\omega$, i.e. the estimate of $\omega$ when the matrix $R$ has no imposed structure. The likelihood ratio test statistic is then defined as

$$\lambda = -2 \cdot \log \left( \frac{L(y; \hat{\omega}_R)}{L(y; \hat{\omega}_U)} \right)$$

where $L(y; \hat{\omega}_R)$ is the value of the log-likelihood function at the restricted maximum likelihood estimate and $L(y; \hat{\omega}_U)$ is the log-likelihood function evaluated at the unrestricted maximum likelihood estimate.

Now, under mild regularity conditions, such as the log-likelihood function being differentiable, and when the assumption that the matrix $R$ is equicorrelated is true, the random variable $\lambda$ is asymptotically distributed with a $\chi^2$ distribution with the number of degrees of freedom equal to $r - s$; here $r$ is the number of component parameters in $\hat{\omega}_U$ and $s$ is the number in $\hat{\omega}_R$.

Returning to Example 4.3, the value of the log-likelihood function evaluated at the restricted maximum likelihood estimate $\hat{\omega}_R$ is $L(y; \hat{\omega}_R) = 617.28$ and the value of the log-likelihood function at the unrestricted maximum likelihood $\hat{\omega}_U$ is $L(y; \hat{\omega}_U) = 667.34$, giving a test statistic $\lambda = 100.12$. When referred to $\chi^2_d$, this supports the belief that the equicorrelated structure assumed for $R$ is incorrect.
4.4.3 Score-test

It has been shown how the likelihood ratio test may be used to test the equicorrelated structure for $R$. However, when calculating the test statistic $\lambda$, the unrestricted maximum likelihood estimates $\hat{\omega}_U = (\hat{\alpha}_U, \hat{\sigma}_U, \hat{\phi}_U, \hat{R}_U)$ are required. An alternative test for the suitability of the matrix $R$ will now be described.

The score-test is asymptotically equivalent to the corresponding likelihood ratio test, and has the advantage of not requiring the full unrestricted model to be fitted to the data.

The full Gaussian log-likelihood function of the data is $L(y; \alpha, \sigma, \phi, R)$ where $\mu = \mu(\alpha)$ and $\Sigma = \Sigma(\sigma, \phi, R)$. The matrix $R$ describes the covariances between the chemical contents in a particular sample of the material. In the case where the matrix $R$ has no imposed structure $R = VWV$, where $V = \text{diag}[r_1, r_2, \ldots, r_n]$ and

$$W = \begin{bmatrix}
1 & \rho_1 & \rho_2 & \rho_3 & \ldots \\
\rho_1 & 1 & \rho_4 & \rho_5 & \ldots \\
\rho_2 & \rho_4 & 1 & \rho_6 & \ldots \\
\rho_3 & \rho_5 & \rho_6 & 1 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}.
$$

The matrix $W$ can be re-parameterised by expressing the correlations $\rho_2, \rho_3, \ldots$ in terms of $\rho_1$. The matrix $W$ can then be rewritten as

$$W = \begin{bmatrix}
1 & \rho_1 & \rho_1 - \gamma_1 & \rho_1 - \gamma_2 & \ldots \\
\rho_1 & 1 & \rho_1 - \gamma_3 & \rho_1 - \gamma_4 & \ldots \\
\rho_1 - \gamma_1 & \rho_1 - \gamma_3 & 1 & \rho_1 - \gamma_5 & \ldots \\
\rho_1 - \gamma_2 & \rho_1 - \gamma_4 & \rho_1 - \gamma_5 & 1 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix},
$$

where the quantities $\gamma_1, \gamma_2, \ldots$ represent the departure of $\rho_2, \rho_3, \ldots$ from $\rho_1$. Using this parameterisation of the matrix $R$ the full Gaussian log-likelihood function is $L(y; \alpha, \sigma, \phi, v, \rho_1, \gamma)$ where $\gamma = (\gamma_1, \gamma_2, \ldots)$ and $v = (r_1, \ldots, r_n)$.

Now, the hypothesis $H : \rho_1 = \rho_2 = \rho_3 = \ldots$ is that the matrix $R$ is equicorrelated. In the new parameterisation this hypothesis is equivalent to $H : \gamma_1 = 0, \gamma_2 = 0, \gamma_3 = 0, \ldots$.

Let $\omega = (\alpha, \sigma, \phi, v, \rho_1)$ and let $\hat{\omega}_R = (\hat{\alpha}_R, \hat{\sigma}_R, \hat{\phi}_R, \hat{v}_R, \hat{\rho}_1 R)$ denote the maximum likelihood estimates in the restricted (equicorrelated) model. The full parameter
set in the unrestricted model is \( \pi = (\omega, \gamma) \) and the gradient vector of the full log-
likelihood function is \( \mathbf{g}_\pi = \frac{\partial \ell}{\partial \pi} \). At the restricted maximum likelihood estimates,
i.e. at \( \hat{\omega}_R \) with \( \gamma = 0 \), the gradient vector is
\[
\{g_{\gamma}\}_{\gamma=0} = (0^T, 0^T, 0^T, 0^T, 0, \{g_{\gamma}\}_{\gamma=0})
\]
where \( g_{\gamma} \) is the non-zero gradient vector of the full log-likelihood with respect to
the \( \gamma \) parameters. This assumes a calculus-type maxima for all the parameters.
Otherwise, if, for example, one or more of the \( \sigma \)-estimates are on the boundary of
the parameter space, the corresponding gradient components might not be zero.

The hypothesis, \( H : \gamma_1 = 0, \gamma_2 = 0, \gamma_3 = 0, \ldots \), is tested by assessing whether the
gradient of the full log-likelihood function, at the restricted maximum likelihood
estimate \( (\hat{\omega}_R, \hat{\sigma}_R, \hat{\phi}_R, \hat{\psi}_R, \hat{\rho}_1, \hat{\rho}_R) \), in the direction of the \( \gamma \) gamma parameters is
small, i.e. by examining the components of \( \{g_{\gamma}\}_{\gamma=0} \) in equation (4.1).

The information, or Hessian, matrix of the full log-likelihood is computed as \( \mathbf{G}_\pi =
- \frac{\partial^2 \ell}{\partial \pi^2} \). The gradient vector and the Hessian matrix can be partitioned into sub-
vectors and sub-matrices which correspond to the parameters \( \omega \) and \( \gamma \):
\[
\mathbf{g}_\pi = \begin{bmatrix} \mathbf{g}_\omega \\ \mathbf{g}_\gamma \end{bmatrix}, \quad \mathbf{G}_\pi = \begin{bmatrix} \mathbf{G}^{\omega\omega} & \mathbf{G}^{\omega\gamma} \\ \mathbf{G}^{\gamma\omega} & \mathbf{G}^{\gamma\gamma} \end{bmatrix}.
\]
The inverse of the matrix \( \mathbf{G}_\pi \) is
\[
\mathbf{G}^{-1} = \begin{bmatrix} \mathbf{G}^{\omega\omega} & \mathbf{G}^{\omega\gamma} \\ \mathbf{G}^{\gamma\omega} & \mathbf{G}^{\gamma\gamma} \end{bmatrix}
\]
where the sub-matrix \( \mathbf{G}^{\gamma\gamma} \) can be evaluated as
\[
\mathbf{G}^{\gamma\gamma} = (\mathbf{G}^{\gamma\gamma} - \mathbf{G}^{\gamma\omega} \mathbf{G}^{\omega\gamma} \mathbf{G}^{\gamma\omega})^{-1}.
\]
At the restricted maximum likelihood the gradient vector \( \mathbf{g}_\omega \) is zero and it is the
gradient \( \mathbf{g}_\gamma \) which is of interest. Under \( H \), the asymptotic distribution of \( \mathbf{g}_\gamma \) is
\[
\mathbf{g}_\gamma \sim N(0, \mathbf{G}^{\gamma\gamma});
\]
here, \( \mathbf{G}^{\gamma\gamma} \) is strictly to be calculated from \( \mathbf{E}(\mathbf{G}_\pi) \), but the observed information
\( \mathbf{G}_\pi \) can be used as an approximation.
To test the hypothesis, $H : \gamma_1 = 0, \gamma_2 = 0, \gamma_3 = 0, \ldots$, the score-test statistic is defined as

$$S = \left[ \{g_\gamma \}_{\gamma = 0} \right]^T \left[ \{G_{\gamma \gamma} \}_{\gamma = 0} \right]^{-1} \left[ \{g_\gamma \}_{\gamma = 0} \right]. \quad (4.2)$$

In the test statistic of equation (4.2) the gradient vector $g_\gamma$ and the information matrix $G_{\gamma \gamma}$ are evaluated at the restricted maximum likelihood estimate, i.e. at $\hat{\omega}_R$ with $\gamma = 0$.

Under $H$, the test statistic $S$ has a $\chi^2$ distribution with degrees of freedom $r = \frac{1}{2} n_c (n_c - 1) - 1$, the number of component parameters in $\gamma$.

The full Gaussian log-likelihood function is

$$L(y; \alpha, \sigma, \phi, \nu, \rho_1, \gamma) = -\frac{1}{2} \log \det(2\pi \Sigma) - \frac{1}{2} (y - X\alpha)^T \Sigma^{-1} (y - X\alpha)$$

and the gradient vector of $L(y; \alpha, \sigma, \phi, \nu, \rho_1, \gamma)$ with respect to $\gamma$ has as its $i$th entry

$$(g_\gamma)_i = \frac{\partial L}{\partial \gamma_i} = -\frac{1}{2} \text{tr} \left( \Sigma^{-1} \frac{\partial \Sigma}{\partial \gamma_i} \right) + \left( \frac{\partial \mu}{\partial \gamma_i} \right)^T \Sigma^{-1} (y - X\alpha) + \frac{1}{2} (y - X\alpha)^T \Sigma^{-1} \frac{\partial \Sigma}{\partial \gamma_i} \Sigma^{-1} (y - X\alpha).$$

The covariance matrix of the data is $\Sigma = ZBZ^T + E$, and since $B = B(\phi)$ and $E = E(\phi, \nu, \rho_1, \gamma)$

$$\frac{\partial \Sigma}{\partial \gamma_i} = \frac{\partial E}{\partial \gamma_i}.$$

Setting the error matrix $E$ up as $E = L \otimes [VWV]$, with $L = L(\phi), V = V(\nu)$ and $W = W(\rho_1, \gamma)$, leads to

$$\frac{\partial \Sigma}{\partial \gamma_i} = L \otimes \left[ V \frac{\partial W}{\partial \gamma_i} V \right].$$

The Hessian matrix of the full log-likelihood function is $G_\gamma = \frac{\partial^2 L}{\partial \gamma_i \partial \gamma_i}$. This matrix is best evaluated by second differencing to give a numerical estimate. The matrix $G_{\gamma \gamma}$ can then be selected from the inverse of $G_\gamma$ and the test statistic $S$ can be computed.
4.5 Example revisited

The graphical indicator and the likelihood ratio tests described in Sections 4.4.1 and 4.4.2 were both applied to Example 4.3, providing evidence that modelling the matrix $R$ with an equicorrelated structure is not appropriate. The disadvantage of these approaches was that the unrestricted maximum likelihood estimate had to be computed. The score-test statistic constructed in Section 4.4.3 will now be computed for the same example. This test has the advantage that it only requires the restricted maximum likelihood to be evaluated.

Using equation (4.2) the test statistic $S$ for the data given in Example 4.3 is calculated as 102.45. This, when referred to $\chi^2_{44}$, suggests, as expected, that restricting the matrix $R$ to have an equicorrelated structure was incorrect. This agrees with the conclusions reached in Sections 4.4.1 and 4.4.2.
Chapter 5

Incomplete-data layouts and missing data

In the examples of Chapters 3 and 4 the data used was the percentage content of a number of chemical elements, measured by three laboratories in the top and bottom of four alloy bars. In these examples the data which is available forms a balanced layout. That is to say, a measurement $y_{blpc}$ has been recorded for every combination of $b$, $l$, $p$ and $c$.

A distinguishing characteristic of laboratory studies is the failure for a complete set of data to be collected. This results in some of the data being missing, so the data which is recorded forms an unbalanced or incomplete-data layout. In the experiment which has been considered in previous chapters, when the laboratories analysed their samples of the alloy bars they were required to measure the content of 17 chemical elements. The complete set of measurements recorded by the three laboratories are given in Table A.1. Here, an entry '−' denotes that the laboratory failed to record a measurement for the content of the chemical element in that particular sample of the bar.

The occurrence of missing data means that the models which have been proposed in previous chapters are inappropriate for modelling incomplete data because they are of a greater dimension than the data. This means that the statistical techniques which have been used become more difficult to implement. In this chapter two approaches for analysing the unbalanced, or incomplete, data are considered. The method described in Section 5.1 involves reducing the full design matrix of
the complete-data model so that it is of the same dimension as the unbalanced, or incomplete, data. In the second approach, described in Section 5.2, instead of reducing the dimension of the complete-data model, the incomplete data is 'expanded' so that it forms a complete-data layout. This is achieved by replacing the missing data with some suitable estimates and then fitting the complete-data model in the usual way.

5.1 Estimating the parameters from the incomplete-data likelihood

Recall that the linear model, in matrix notation, is

\[ y = X\alpha + Z\alpha + e \]  \hspace{1cm} (5.1)

and in the case of a complete-data layout the matrices \( X \) and \( Z \), the design matrices of the fixed and random effects, are set up as in Section 3.1. This model will be called the complete-data model. The expectation and covariance structure of the data are

\[ E(y) = \mu = X\alpha \]  \hspace{1cm} (5.2)

and

\[ \text{Cov}(y) = \Sigma = ZBZ^T + \Psi \]  \hspace{1cm} (5.3)

and the corresponding Gaussian log-likelihood function of the data is

\[ L(y; \alpha, \theta) = -\frac{1}{2} \log \det(2\pi \Sigma) - \frac{1}{2} (y - X\alpha)^T \Sigma^{-1} (y - X\alpha). \]  \hspace{1cm} (5.4)

The method described in this section involves reducing the complete-data model, and hence the full Gaussian log-likelihood function, which are all of dimension \( n \), so that they are of the same dimension as the incomplete data.

Now, for both a complete and incomplete-data layout, the linear model to describe a single measurement is,

\[ y_{ibpc} = \mu + \alpha_p + \alpha_c + \alpha_{pc}^C \]
\[ + a_s + a_l + a_{pc}^B + a_{bp} + a_{lc} + a_{bpc}^C + e_{ibpc}. \]  \hspace{1cm} (5.5)
The corresponding expectation and covariance structure of the data are
\[
E(y_{bpc}) = \mu + \alpha_p + \alpha_c + \alpha_{pc} \tag{5.6}
\]
and
\[
\text{Cov}(y_{bpc}, y_{b'p'c'}) = \delta_{bb'}\sigma_{BC}^2 + \delta_{bb'}\delta_{cc'}\sigma_{BC}^2 + \delta_{bb'}\delta_{pp'}\sigma_{BP}^2
\]
\[
+ \delta_{bb'}\delta_{cc'}\sigma_{PC}^2 + \delta_{bb'}\delta_{pp'}\delta_{cc'}\sigma_{PP'}^2 + \text{Cov}(e_{bpc}, e_{b'p'c'}) \tag{5.7}
\]
where \( \delta \) is the Kronecker delta symbol as previously defined.

This is the linear model used in Chapters 3 and 4 to describe the data when it forms a complete-data layout. However, the occurrence of missing data means that the linear model of equations (5.5), (5.6) and (5.7) cannot be written in the matrix notation of the Harville-Laird-Ware model as easily as when no data is missing. This is because the sub-matrices of \( X \) and \( Z \) can no longer be constructed simply, e.g. by setting \( Z = I_n \otimes I_n \otimes I_p \otimes I_n \).

When the data forms an incomplete-data layout, the linear model of equation (5.5) can be written in matrix notation by modifying the design matrices \( X \) and \( Z \), and the error vector \( e \). Once the model for the incomplete-data has been written in matrix notation the structure of the expectation and the covariance of the data follows. It will now be described how the complete-data model of equation (5.1) may be modified so that it can be used to model incomplete data.

As before, \( n (= n_b \cdot n_c \cdot n_p \cdot n_c) \) is the total number of measurements recorded when there are no missing values and the data forms a complete-data layout. Now define \( n_1 \) to be the number of measurements which are actually recorded by the laboratories. The value of \( n_1 \) is then smaller than \( n \), and when \( n_1 = n \) it is the special case of no missing data. Previously, in the case of no missing data, \( y \) has been the vector of length \( n \) which contained all the measurements \( y_{bpc} \) recorded with the subscripts on \( y_{bpc} \) varying fastest from right to left. Now, when some of the data is missing, and the data no longer forms a complete-data layout, define \( y \) as the vector of length \( n \) which has the measurements \( y_{bpc} \) which were recorded in their usual place but the measurements which are missing are replaced by some value to indicate that fact. Since all the data are positive a natural choice for this indicator value is some value less than zero, say '9'.
Now, introduce the matrix $\mathbf{I}_{n_1}^*$. This is defined as the $n \times n$ unit matrix $\mathbf{I}_n$, with the rows deleted which correspond to missing measurements. The matrix $\mathbf{I}_{n_1}^*$ is of size $n_1 \times n$. For example, if the third measurement in the vector $\mathbf{y}$ is missing, the matrix $\mathbf{I}_{n_1}^*$ will take the form

\[
\mathbf{I}_{n_1}^* = \begin{bmatrix}
1 & 0 & 0 & 0 & \ldots & 0 \\
0 & 1 & 0 & 0 & \ldots & 0 \\
0 & 0 & 1 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & 0 & \ldots & 1
\end{bmatrix}.
\]

By setting

\[
\mathbf{y}_{n_1} = \mathbf{I}_{n_1}^* \mathbf{y}
\]

the vector $\mathbf{y}_{n_1}$ is of length $n_1$ and only contains the $n_1$ measurements which were recorded by the laboratories. The ordering of the measurements in the vector $\mathbf{y}_{n_1}$ still has the subscripts on $\mathbf{y}_{bdpc}$ varying fastest from right to left. All that has been done is that the values in the vector $\mathbf{y}$ which denote missing values have been removed.

Pre-multiplying the design matrices, and the vector of random errors, in the complete-data model of equation (5.1) by the matrix $\mathbf{I}_{n_1}^*$ gives

\[
\begin{align*}
\mathbf{X}_{n_1} &= \mathbf{I}_{n_1}^* \mathbf{X}, \\
\mathbf{Z}_{n_1} &= \mathbf{I}_{n_1}^* \mathbf{Z}
\end{align*}
\]

and

\[
\mathbf{e}_{n_1} = \mathbf{I}_{n_1}^* \mathbf{e}.
\]

What has been achieved here is that the matrices $\mathbf{X}$ and $\mathbf{Z}$ of size $n \times p$ and $n \times q$ and the vector $\mathbf{e}$ of length $n$ have been reduced to the matrices $\mathbf{X}_{n_1}$ and $\mathbf{Z}_{n_1}$, of size $n_1 \times p$ and $n_1 \times q$ respectively, and a vector $\mathbf{e}_{n_1}$ of length $n_1$. Each row of the matrices $\mathbf{X}_{n_1}$, $\mathbf{Z}_{n_1}$ and each error in the vector $\mathbf{e}_{n_1}$ corresponds to one of the $n_1$ measurements which was recorded. Using these new design matrices and the new vector of random errors the complete-data model of equation (5.1) has been reduced to model incomplete data. The linear model for incomplete data is

\[
\mathbf{y}_{n_1} = \mathbf{X}_{n_1} \alpha + \mathbf{Z}_{n_1} \mathbf{a} + \mathbf{e}_{n_1} \tag{5.8}
\]
where $\alpha$ and $\mathbf{a}$ are the fixed and random effects which are present in the complete-data model. The expectation and covariance structure of the data are

$$E(y_{n_1}) = \mu_{n_1} = X_{n_1} \alpha$$

and

$$\text{Cov}(y_{n_1}) = \Sigma_{n_1} = Z_{n_1} B Z_{n_1}^T + E_{n_1},$$

where

$$E_{n_1} = \text{Cov}(e_{n_1}) = I_{n_1} \text{E}(e_{n_1})^T.$$

The covariance matrix of the incomplete data is $\Sigma_{n_2}$, which is the covariance matrix of the complete data $\Sigma$ with the rows and columns corresponding to missing measurements deleted.

The parameters in the incomplete-data model which have to be estimated are the same as those which occur in the complete-data model. These are $\alpha$ and $\theta = (\sigma, \phi, R)$. The assumption of normality for the terms $\alpha$ and $e_{n_1}$ in the incomplete-data model results in the incomplete data having the multivariate normal density function:

$$f(y_{n_1}; \alpha, \theta) = \left\{ \det \left( 2\pi \Sigma_{n_1} \right) \right\}^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (y_{n_1} - X_{n_1} \alpha)^T \Sigma_{n_1}^{-1} (y_{n_1} - X_{n_1} \alpha) \right]$$

and the corresponding incomplete Gaussian log-likelihood function of the data is defined as

$$L(y_{n_1}; \alpha, \theta) = -\frac{1}{2} \log \det(2\pi \Sigma_{n_1}) - \frac{1}{2} (y_{n_1} - X_{n_2} \alpha)^T \Sigma_{n_1}^{-1} (y_{n_1} - X_{n_2} \alpha).$$

As in the complete-data model, when some values are assumed for the variance components, the fixed effects may be computed by solving the equation $\frac{\partial L}{\partial \alpha} = 0$. The solution to the resulting normal equations is

$$\hat{\alpha} = (X_{n_1}^T \Sigma_{n_1}^{-1} X_{n_1})^{-1} X_{n_1}^T \Sigma_{n_1}^{-1} y_{n_1}$$

and the estimated covariance matrix of these fixed effects is

$$V_{\alpha} = (X_{n_1}^T \Sigma_{n_1}^{-1} X_{n_1})^{-1}.$$
The estimates of the fixed effects from equation (5.10) may then be substituted into the incomplete log-likelihood function of equation (5.9) yielding \( L(y_n; \theta) \), a reduced log-likelihood, dependent only on the variance components \( \theta \). Maximisation of this function results in the evaluation of the maximum likelihood estimates of \( \theta \). The maximisation of the incomplete Gaussian log-likelihood can be done using a slightly modified version of the FORTRAN computer algorithm suggested in Section 2.3.2.

The standard errors of the variance components \( \theta \) can be calculated using the observed information matrix described in Section 2.3.2. The matrices \( X \) and \( \Sigma \) which appear in the relevant equations in this section are simply replaced by \( X_n \) and \( \Sigma_n \) respectively. Confidence intervals and parametric tests can then be constructed in the usual way.

5.1.1 Over-parameterisation

If the data in Table A.1 formed a complete-data layout then a total of 408 measurements would have been recorded. However only 336 measurements are available, meaning that about 20% of the data is missing. Even though only a relatively small amount of data is missing there are still some parameters which appear in the complete-data model which are not necessary in the incomplete-data model.

Consider the ‘laboratory \( \times \) chemical’ interaction terms in the incomplete-data model. Inspection of Table A.1 reveals that laboratory 1 did not record any measurements for the content of Hafnium in the samples of the alloy which it received. However, the incomplete-data model of equation (5.8) includes in the vector of random effects a ‘laboratory \( \times \) chemical’ interaction effect which corresponds to laboratory 1 and the chemical element Hafnium. Clearly, this term does not need to be included in the incomplete-data model and so the model is said to be over-parameterised. Although the incomplete-data model is over-parameterised, it can still be fitted to the data in the usual way. Two of the problems of over-parameterisation will now be addressed.

The first of these complications is a computational one. The inclusion of higher order interaction effects in the complete-data model, such as \( a^{BPC} \), significantly
increases the size of the matrix $Z$. The matrix $Z$ is of size $n \times q$, where the value of $q$ can become quite large. In the complete-data model all of the parameters in the vector $a^{BPC}$ are required and so each of the columns of $Z$ will contain a combination of 0s and 1s. When the data forms an incomplete-data layout, especially when a large amount of the data is missing, it may be the case that some of the parameters $a_{bpc}^{BPC}$ in the vector $a^{BPC}$ are not required. The result of this is that when the matrix $Z$ is reduced into the matrix $Z_{n1}$ the columns of $Z_{n1}$ which correspond to the unnecessary parameters $a_{bpc}^{BPC}$ will consist entirely of zeros. As a large amount of computational work is involved in fitting the model any unnecessary arithmetic operations are undesirable.

This problem can be easily side-stepped by simply deleting the columns of $Z_{n1}$ which consist entirely of zeros and updating the vector of random effects $a$ and its covariance matrix $B$ accordingly. The matrix $Z_{n1}$, which is of size $n_1 \times q$, can be modified as follows. Let $M$ be the $q \times q$ unit matrix with columns deleted corresponding to the parameters present in the complete-data model but not required in the incomplete-data model. The modified matrices $Z_{n1}$ and $B$ are

$$Z_{n1} = Z_{n1}M$$

and

$$B = M^TB\!M.$$  

Using these new matrices $Z_{n1}$ and $B$ the incomplete-data model will contain only the random effects which are necessary.

The second consequence of over parameterisation is somewhat more serious. In the same way that a large amount of missing data results in some of the parameters in $a$ being unnecessary, there is also the risk that some of the interaction terms in the fixed effects vector $\alpha$ may also not be needed. If this does occur then the result is that, when calculating estimates of $\alpha$ via the equation

$$\hat{\alpha} = (X_{n1}^T\Sigma_{n1}^{-1}X_{n1})^{-1}X_{n1}^T\Sigma_{n1}^{-1}\gamma_{n1},$$

an infinite number of solutions will exist for the vector $\hat{\alpha}$. When this happens interpretation of certain parameters in this vector can become difficult and confusing. The solution to this problem is similar to that in the previous case and
is overcome by deleting the columns of the matrix $X_{n_1}$ which correspond to the parameters in $\alpha$ which are not required.

5.2 Estimating the parameters and missing data from the complete-data likelihood

In the approach taken in this section, rather than reducing the complete-data model so that it is of the same dimension of the incomplete data, the reverse is done. The data available, which forms an incomplete-data layout, is ‘expanded’ so that it is of the same dimension as the complete-data model. This is achieved by substituting into the incomplete data some estimates of the measurements which were not recorded. The result of this is that a complete set of data is generated from the data which is available.

Suppose, for convenience that the data vector $y$ can be neatly partitioned into $(y_m, y_o)$ where

$y_m$ denotes the unobserved measurements;

$y_o$ denotes the observed measurements.

The complete-data Gaussian log-likelihood function is $L(y_o, y_m \mid \alpha, \theta)$. Treating the vector $y_m$ as an unknown model parameter, this may be written as $L(y_o \mid \alpha, \theta, y_m)$. To find the maximum likelihood estimates of the parameters $\alpha$ and $\theta$, and also estimates (MLEs) of the unrecorded measurements we find the set of parameters $(\alpha, \theta, y_m)$ which maximise the complete-data log-likelihood $L(y_o \mid \alpha, \theta, y_m)$. That is

$$\max_{\alpha, \theta, y_m} \{L(y_o \mid \alpha, \theta, y_m)\}.$$ 

This can be easily implemented using a slightly modified version of the computer algorithm suggested in Section 2.3. Here the algorithm is amended so that the iterative search is carried out over the new parameter space $(\alpha, \theta, y_m)$. Although this is very simple to implement, it is unlikely to be used in practice unless the number of unrecorded measurements is fairly small. This is because, when even a relatively small amount of data is missing, the number of new model parameters
representing the missing data can easily dominate the parameter vector, significantly increasing the computational work to locate a maximum. A more efficient method for performing this computation will be described in the following section.

5.2.1 Maximising the complete-data likelihood using the EM Algorithm

In Section 2.3 an algorithm, similar to the EM algorithm, was used to maximise the complete-data Gaussian log-likelihood. There, the E-step in the EM algorithm was replaced by a step to estimate the fixed effects given some value of the variance components $\theta$. The maximisation step (M-step) was to maximise the reduced Gaussian log-likelihood function $L_1(y; \theta)$.

The EM algorithm was developed to maximise the complete-data log-likelihood function in the case where some of the data is missing and is particularly easy to implement. The E-step is to calculate the expected values of the data which is missing conditioned upon the data which is available. These estimates of the missing data are then substituted into the incomplete data to form a complete-data layout. The M-step is then to maximise the complete-data log-likelihood for the data which is available and the data which has been estimated, $L(y_o, y_m; \alpha, \theta)$.

To maximise the incomplete-data likelihood function, using the EM algorithm can be achieved as follows:

1. For the vector $y = (y_m, y_o)$ replace the missing measurements with some suitable estimates of their values to yield a complete set of data $y^{(1)}$. A natural choice for $y_m^{(1)}$ is to set its entries equal to the target percentage contents of the chemical elements.

2. Fit the complete-data model using the current data vector $y^{(i)}$ as in Section 2.3. This involves maximising the log-likelihood $L(y^{(i)}; \alpha, \theta)$ to give estimates $\hat{\alpha}^{(i)}$ and $\hat{\theta}^{(i)}$.

3. Use the current estimates $\hat{\alpha}^{(i)}$ and $\hat{\theta}^{(i)}$ to give an updated estimate of the missing data $y_m^{(i+1)}$. This gives an updated complete set of data $y^{(i+1)} = (y_m^{(i+1)}, y_o)$. 

75
4 Return to step 2 and continue this process until the parameters estimates and the log-likelihood function stabilise.

This is the EM algorithm: step 2 is the maximising (M) stage and step 3 is the estimating (E) stage.

The estimates of the missing data evaluated in step 2 are the conditional expectations of the missing data given the observed data. The data has been partitioned into missing and observed data and the corresponding partitions of the mean vector \( \mu \) and the covariance matrix \( \Sigma \) are

\[
\mu = \begin{bmatrix} \mu_m \\ \mu_o \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{mm} & \Sigma_{mo} \\ \Sigma_{om} & \Sigma_{oo} \end{bmatrix}
\]

where \( \Sigma_{mm} \) and \( \Sigma_{oo} \) are symmetric positive semi-definite matrices of order \( n - n_1 \) and \( n_1 \) respectively, and \( \Sigma_{mo} = \Sigma_{om}^T \) is \( (n - n_1) \times n_1 \).

The conditional distribution of \( y_m \) given \( y_o \) is multivariate normal with

\[
E(y_m | y_o) = \mu_m + \Sigma_{mo} \Sigma_{oo}^{-1}(y_o - \mu_o)
\]

\[
\text{Var}(y_m | y_o) = \Sigma_{mm} - \Sigma_{mo} \Sigma_{oo}^{-1} \Sigma_{om}.
\]

The estimate of the missing data at the \( i \) iteration is then calculated using equation (5.11), i.e.

\[
y_m^{(i)} = \mu_m^{(i-1)} + \Sigma_{mo}^{(i-1)}(\Sigma_{oo}^{(i-1)})^{-1}(y_o - \mu_o^{(i-1)}).
\]

Convergence of the procedure can be checked by comparing successive values of the parameter estimates and the log-likelihood function. The value of the log-likelihood function at the \( i \)th iteration is denoted by \( L^{(i)}(y^{(i)}; \hat{\alpha}^{(i)}, \hat{\theta}^{(i)}) \) and \( \hat{\alpha}^{(i)} \) and \( \hat{\theta}^{(i)} \) are the parameter estimates produced at the \( i \)th iteration. The EM algorithm should then run until

\[
| L^{(i)} - L^{(i+1)} | < \delta, \quad | \hat{\alpha}^{(i)} - \hat{\alpha}^{(i+1)} | < \delta, \quad | \hat{\theta}^{(i)} - \hat{\theta}^{(i+1)} | < \delta
\]

for some suitably small \( \delta \).
5.3 Maximisation of the incomplete data likelihood vs the complete-data likelihood

Maximisation of the incomplete-data log-likelihood and the complete-data log-likelihood are both computationally demanding. For this reason the advantages of the two methods described in Sections 5.1 and 5.2 will be assessed from a computational point of view.

Because of the complex structure of the covariance matrix and the design matrices, maximisation of the log-likelihood functions has to be performed by some iterative procedure. The clearest advantage that maximising the incomplete-data log-likelihood has over the EM algorithm is that it only requires a single maximisation of a log-likelihood function. The EM algorithm requires the maximisation of a log-likelihood each time the vector of missing data is updated.

In Section 3.4.3 it was shown how the direct inversion of the covariance matrix could be avoided with a significant decrease in the number of arithmetic operations. Even so, the inversion of the covariance matrix is still the most demanding part of evaluating the log-likelihood function. Fitting the model, using the EM algorithm, requires a covariance matrix of size $n \times n$ to be inverted. In the reduced matrix method of Section 5.1 a reduced covariance matrix of size $n_1 \times n_1$ is inverted. As the number of arithmetic operations performed when inverting these matrices is proportional to $n^2$ and $n_1^2$, the occurrence of a large amount of missing data, i.e. a large value of $n/n_1$, means that the inversion of the covariance matrix, and hence evaluation of the log-likelihood function, is significantly more demanding when the EM algorithm is applied.

It appears that the method of reducing the complete-data model to the dimension of the incomplete data is best suited for fitting the model but two situations will now be described in which the application of the EM algorithm becomes competitive. In some cases it may be that the complete-data log-likelihood, as used in the EM algorithm, can be maximised non-iteratively. If explicit maximum likelihood estimates of the parameters can be found then the use of the EM algorithm should be considered. However, the presence of the parameters in the covariance structure rules this out.
Another consideration is the computation of the parameter estimate $\hat{\alpha}$ using the equation

$$\hat{\alpha} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} y.$$ 

It may be the case that the inversion of the covariance matrix $\Sigma$ in this equation can be done algebraically as in Section 3.4.2, but the matrix $\Sigma_n$ can only be inverted numerically. This would make the EM algorithm approach more efficient for this part of the computation.

### 5.4 Example

In this example the data to be used is the complete set of measurements from the experiment which was carried out. The data consist of the percentage contents of 17 chemical elements measured in the top and bottom of four alloy bars by three laboratories. These data are given in Table A.1. The chemical elements of interest are Aluminium (3.0%), Cobalt (10.0%), Chromium (5.0%), Copper (1.0%), Iron (5.0%), Hafnium (0.1%), Manganese (1.0%), Molybdenum (1.0%), Nickel (1.0%), Niobium (1.0%), Phosphorus (0.1%), Silicon (0.5%), Tantalum (1.0%), Titanium (1.0%), Vanadium (1.0%), Tungsten (5.0%) and Zirconium (0.5%). The percentages in parentheses are the target percentage contents of these chemical elements in the alloy.

If the data formed a balanced layout a total of $n = 408$ measurements would be available and a suitable linear model would be that of equation (3.2),

$$y = \mu + X_1 \alpha^P + X_2 \alpha^C + X_3 \alpha^{PC} + Z_1 a^B + Z_2 a^L + Z_3 a^{BC} + Z_4 a^{BP} + Z_5 a^{LC} + Z_6 a^{BPC} + \epsilon.$$ 

An equicorrelated structure, as described in Section 4.2, is assumed to describe the errors for the chemical contents measured in a particular sample of the alloy. The matrices $X$ and $Z$ are set up in a similar fashion to those in Example 4.3. When the complete-data model has been constructed it can be reduced, as in Section 5.1, to the dimension of the data.

Laboratory 1 did not record any measurements for the content of Hafnium, Niobium or Phosphorus. Now, because measurements for the content of Hf, Ni and Ph
were not made by laboratory 1, there is no need for the incomplete-data model to contain the parameters $a_1^{L(0)}$, $a_1^{L(10)}$ and $a_1^{L(11)}$. Similarly, laboratory 2 did not report any measurements for Niobium, Phosphorus, Vanadium, Tungsten or Zirconium and laboratory 3 did not record measurements for the content of Niobium. The parameters $a_2^{L(0)}$, $a_2^{L(10)}$, $a_2^{L(11)}$, $a_2^{L(10)}$, $a_2^{L(10)}$ and $a_2^{L(10)}$ are not necessary in the incomplete-data model. For computational efficiency these parameters should be removed from the vector $a$ and the matrices $Z_{n,m}$ and $B$ updated accordingly. In this example the exclusion of such a small number of parameters from $a$, and hence columns from $Z^{m,n}$, does not reduce the number of arithmetic operations significantly.

A measurement for the content of all 17 chemical elements was recorded in the top and bottom of each bar. Therefore, the 'position x chemical' interaction effects which are present in the complete-data model are required in the incomplete-data model. The parameter vector $\alpha$, and the matrix $X_{n,m}$, can remain unchanged.

Fitting the incomplete-data model to the data given in Table A.1 yielded a position effect estimated as $-5.981 \times 10^{-4}$ with standard error of $2.103 \times 10^{-3}$ giving a normal deviate of $-0.286$. This suggests that the content of the chemical elements are the same in the top of the bar as they are in the bottom. The hypothesis $H : \alpha_{1,1}^{PC} = \ldots, \alpha_{1,16}^{PC} = 0$ may be used to test if the 'position x chemical' interaction is required. This yields a Wald statistic of $W = 121.744$ which, when referred to tables of $\chi^2_4$, supports the inclusion of this interaction term in the model. These findings are consistent with those in Example 4.3.

If the observed content of the 17 chemical elements meets the pre-defined target content using the Wald statistic. The chemical elements with their corresponding test statistic are as follows; Aluminium (3.05), Cobalt (6.77), Chromium (48.7), Copper (11.2), Iron (5.00), Hafnium (4.64), Manganese (1.54), Molybdenum (2.2), Nickel (22.4), Niobium (7.75), Phosphorus (2.65), Silicon (7.88), Tantalum (9.43), Titanium (8.1), Vanadium (1.59), Tungsten (4.59) and Zirconium (27.4). Referring these values of $W$ to tables of $\chi^2_4$ suggests that, as found in Example 4.3, the contents of Chromium and Copper do not meet their target contents. Also, there is some doubt raised over whether the contents of Nickel and Zirconium meet their target contents.
The estimates of the set of parameters which govern the covariance matrix $\Sigma$ are given in Table 5.1. The second column of this table refers to the maximum likelihood estimates of the parameter and the third column gives the standard error of the estimate, calculated from the observed information matrix. The estimates of the bar and 'bar x chemical' interaction variance components are once again zero. This is consistent with the conclusion made earlier that the content of the chemical elements does not vary from bar to bar.

Confidence intervals (95%) for the logs of the laboratory and 'laboratory x chem-

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLE</th>
<th>Gaussian std. err</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_L$</td>
<td>$1.870 \times 10^{-2}$</td>
<td>$1.539 \times 10^{-2}$</td>
</tr>
<tr>
<td>$\sigma_{LC}$</td>
<td>$6.727 \times 10^{-2}$</td>
<td>$7.894 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\sigma_B$</td>
<td>$0.000$</td>
<td>$--$</td>
</tr>
<tr>
<td>$\sigma_{BP}$</td>
<td>$6.035 \times 10^{-4}$</td>
<td>$5.460 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\sigma_{BC}$</td>
<td>$0.000$</td>
<td>$--$</td>
</tr>
<tr>
<td>$\sigma_{BPC}$</td>
<td>$8.438 \times 10^{-4}$</td>
<td>$3.182 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\rho$</td>
<td>$0.161$</td>
<td>$6.229 \times 10^{-2}$</td>
</tr>
<tr>
<td>$r_1$</td>
<td>$5.841 \times 10^{-4}$</td>
<td>$1.752 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_2$</td>
<td>$1.651 \times 10^{-3}$</td>
<td>$5.020 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_3$</td>
<td>$6.162 \times 10^{-4}$</td>
<td>$1.834 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_4$</td>
<td>$5.183 \times 10^{-5}$</td>
<td>$1.640 \times 10^{-5}$</td>
</tr>
<tr>
<td>$r_5$</td>
<td>$2.481 \times 10^{-3}$</td>
<td>$7.993 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_6$</td>
<td>$3.697 \times 10^{-5}$</td>
<td>$1.521 \times 10^{-5}$</td>
</tr>
<tr>
<td>$r_7$</td>
<td>$4.760 \times 10^{-5}$</td>
<td>$1.519 \times 10^{-5}$</td>
</tr>
<tr>
<td>$r_8$</td>
<td>$4.851 \times 10^{-5}$</td>
<td>$1.567 \times 10^{-5}$</td>
</tr>
<tr>
<td>$r_9$</td>
<td>$1.489 \times 10^{-4}$</td>
<td>$5.559 \times 10^{-6}$</td>
</tr>
<tr>
<td>$r_{10}$</td>
<td>$1.820 \times 10^{-4}$</td>
<td>$9.503 \times 10^{-5}$</td>
</tr>
<tr>
<td>$r_{11}$</td>
<td>$3.432 \times 10^{-8}$</td>
<td>$1.961 \times 10^{-7}$</td>
</tr>
<tr>
<td>$r_{12}$</td>
<td>$1.371 \times 10^{-4}$</td>
<td>$4.341 \times 10^{-6}$</td>
</tr>
<tr>
<td>$r_{13}$</td>
<td>$1.251 \times 10^{-4}$</td>
<td>$3.854 \times 10^{-6}$</td>
</tr>
<tr>
<td>$r_{14}$</td>
<td>$4.690 \times 10^{-5}$</td>
<td>$1.408 \times 10^{-5}$</td>
</tr>
<tr>
<td>$r_{15}$</td>
<td>$8.057 \times 10^{-5}$</td>
<td>$2.872 \times 10^{-5}$</td>
</tr>
<tr>
<td>$r_{16}$</td>
<td>$1.674 \times 10^{-3}$</td>
<td>$6.709 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_{17}$</td>
<td>$5.320 \times 10^{-4}$</td>
<td>$2.073 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>$0.000$</td>
<td>$--$</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>$0.262$</td>
<td>$8.167 \times 10^{-2}$</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>$9.959 \times 10^{-2}$</td>
<td>$9.817 \times 10^{-2}$</td>
</tr>
</tbody>
</table>
icals’ interaction variance components are constructed as

$$-5.592 < \log(\sigma_L) - 2.366, \quad -2.929 < \log(\sigma_{LC}) < -2.469.$$  

suggesting that the inclusion of these terms in the model are necessary and providing evidence that there is a real difference between laboratories.

Similarly, confidence intervals for the ‘bar × position’ and ‘bar × position × chemical’ are constructed as

$$-9.186 < \log(\sigma_{BP}) - 5.640, \quad -7.818 < \log(\sigma_{BPC}) < -6.339$$

suggesting that these two random effects are necessary in the model.

Fig. 5.1 shows the normal plot of the standardised residuals. It appears that the incomplete-data model fits the data reasonably well. The point in the lower left hand corner of this plot indicates that it may be an outlying measurement. This point corresponds to the measurement which was identified in Example 4.3 as an outlier.
Figure 5.1: Normal plot of the standardised residuals. Measurements recorded by three laboratories on the content of 17 chemical elements in the top and bottom of four alloy bars.
Chapter 6

Gaussian Estimation

In Chapter 4 the problem of model misspecification was addressed. This was done by showing that the structure assumed for the covariance matrix $R$ was probably incorrect. Another factor which may be responsible for model misspecification will now be investigated. In what has gone before, to analyse the data a linear model with some covariance structure was assumed. Several further assumptions were made about this model, one of which was that the data possesses a multivariate normal distribution. In this chapter the model which has been suggested in Chapters 3 and 4 for the multi-chemical data is again fitted to the data. The assumption of normality will then be investigated and, if there is evidence to suggest a lack of normality, the analysis will be modified to account for this.

6.1 Non Normality of Data

Plots of the standardised residuals have been used in previous sections to identify outlying measurements and, to some extent, the goodness-of-fit of the model. Let us return to Example 4.3, where the content of 10 chemical elements was of interest. The plot of the entire set of standardised residuals for this example is given in Fig. 4.2. As pointed out previously, this plot could be taken to suggest that several of the observations are outlying.

Residual plots for each of the individual chemical elements are shown in Fig. 6.1. It is immediately clear that for several chemicals there is a lack of linearity in these
plots. This is particularly evident in the plots Fig. 6.1(a), Fig. 6.1(b) and Fig. 6.1(e), the plots corresponding to aluminium, cobalt and iron. The reason behind this lack of linearity may be that the measurements observed on these chemicals are in fact not normally distributed.

One possible solution to this problem may be to transform the data of those chemicals which show signs of non-normality. For non-negative data, which is what has been used so far, a log transform may improve normality.

### 6.2 Gaussian Estimation

Parametric inferences have been made using the parameter estimates $\hat{\omega} = (\hat{\alpha}, \hat{\theta})$. The tests performed on the fixed effects involved calculating Wald statistics and normal deviates, whereas the tests on the random effects and variance components used the construction of confidence intervals. All of these tests require the standard errors of the parameter estimates to be computed.

The assumed density for $f(y; \alpha, \theta)$ has been a multivariate normal density and the standard errors which have been quoted in the Examples 3.3 and 4.3 have been based on this density function. As there has been doubt raised over the correctness of the model and the assumption of normality then the density function $f(y; \alpha, \theta)$ is unlikely to be the true density function of the data. The implication of this is that the standard errors computed will need modifying to account for these misspecifications.

Gaussian estimation can be used to estimate certain quantities, such as these standard errors, to give some protection from the fact that the assumed density function is not the true density function. The approach taken in Gaussian Estimation is as follows.

Estimates of the parameters are found in the usual way by using some estimating equation. As the estimating equation used throughout has been the one which produces maximum likelihood estimates, we will continue to use that. These estimates are computed by finding the solution to the equation $\frac{\partial L}{\partial \omega} = 0$, where $\omega = (\alpha, \theta)$ and $\theta = (\sigma, \phi, R)$. The parameter estimates resulting from solving
(a) Aluminium
(b) Cobalt
(c) Chromium
(d) Copper
(e) Iron
(f) Manganese
Figure 6.1: Normal plots of the standardised residuals for each of the 10 chemical elements.

This estimating equation can then be thought of as a function of the data. The concept behind Gaussian estimation is that the computation of certain quantities, such as standard errors, should be based on this function of the data rather than the assumed density function of the data.

Now, as long as the estimating equation is asymptotically unbiased, the parameter estimates are unique and consistent with asymptotic covariance matrix

$$V = D^{-1}C(D^{-1})^T$$
where the matrices $C$ and $D$ are defined as

$$ C = \text{Cov} \left[ \frac{\partial L}{\partial \omega} \right], \quad D = E \left[ -\frac{\partial^2 L}{\partial^2 \omega} \right]. \quad (6.1) $$

Recall that the Gaussian log-likelihood function of the data is

$$ L(y; \omega) = \log \left( \{\det (2\pi \Sigma)\}^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (y - X\alpha)^T \Sigma^{-1} (y - X\alpha) \right] \right) $$

$$ = -\frac{1}{2} \log \det (2\pi \Sigma) - \frac{1}{2} (y - X\alpha)^T \Sigma^{-1} (y - X\alpha) \quad (6.2) $$

and that the corresponding Gaussian score function is $q(y; \omega) = \partial L / \partial \omega$, where the $i$th component of the vector $q(y; \omega)$ satisfies

$$ \{q(y; \omega)\}_i = -\frac{1}{2} tr \left( \Sigma^{-1} \frac{\partial \Sigma}{\partial \omega_i} \right) + X^T \Sigma^{-1} (y - X\alpha) $$

$$ + \frac{1}{2} (y - \mu)^T \Sigma^{-1} \left( \frac{\partial \Sigma}{\partial \omega_i} \right) \Sigma^{-1} (y - X\alpha). \quad (6.3) $$

Now, because in the model that has been suggested the fixed effects are distinct from those which govern the covariance structure, the Gaussian score function of equation (6.3) can be partitioned into $q^a(y; \omega) = \frac{\partial L}{\partial \alpha}$ and $q^o(y; \omega) = \frac{\partial L}{\partial \theta}$. The vector $q^a(y; \omega)$ can be evaluated using the formula

$$ q_i^a(y; \omega) = X^T \Sigma^{-1} (y - X\alpha) \quad (6.4) $$

and the $i$th component of $q^o(y; \omega)$ satisfies

$$ q_i^o(y; \omega) = -\frac{1}{2} tr \left( \Sigma^{-1} \frac{\partial \Sigma}{\partial \omega_i} \right) + \frac{1}{2} (y - X\alpha)^T \Sigma^{-1} \left( \frac{\partial \Sigma}{\partial \omega_i} \right) \Sigma^{-1} (y - X\alpha). \quad (6.5) $$

The matrices $C$ and $D$ of equation (6.1) can be partitioned into

$$ D = \begin{bmatrix} D_{\alpha\alpha} & D_{\alpha\theta} \\ D_{\theta\alpha} & D_{\theta\theta} \end{bmatrix} = E \begin{bmatrix} -\frac{\partial^2 L}{\partial^2 \alpha} & -\frac{\partial^2 L}{\partial \alpha \partial \theta} \\ -\frac{\partial^2 L}{\partial \theta \partial \alpha} & -\frac{\partial^2 L}{\partial^2 \theta} \end{bmatrix} $$

and

$$ C = \begin{bmatrix} C_{\alpha\alpha} & C_{\alpha\theta} \\ C_{\theta\alpha} & C_{\theta\theta} \end{bmatrix} = \begin{bmatrix} \text{Cov} \{q^a(y; \omega), q^a(y; \omega)\} & \text{Cov} \{q^o(y; \omega), q^a(y; \omega)\} \\ \text{Cov} \{q^o(y; \omega), q^o(y; \omega)\} & \text{Cov} \{q^o(y; \omega), q^o(y; \omega)\} \end{bmatrix}. $$

The estimating equation is asymptotically unbiased, i.e. $E(\frac{\partial L}{\partial \omega}) = 0$, and so the matrix $C$ may be evaluated as

$$ C = E \{q(y; \omega)q(y; \omega)^T\}. $$

87
Computing the matrices $C_{\alpha\alpha}$ and $D_{\alpha\alpha}$ can be carried out via the use of equation (6.4) and it can be seen that

$$C_{\alpha\alpha} = D_{\alpha\alpha} = X^T \Sigma^{-1} X.$$  

Differentiating $q_\theta^\theta(y; \omega)$ of equation (6.5) with respect to $\alpha$ and taking the expectation results in $D_{\alpha\theta} = D_{\theta\alpha}^T = 0$. Replacing $D_{\alpha\theta}$ and $D_{\theta\alpha}$ with 0 and 0$^T$ in the matrix $D$ the asymptotic covariance matrix of the parameter estimates reduces to

$$V = \begin{bmatrix} V_{\alpha\alpha} & V_{\alpha\theta} \\ V_{\theta\alpha} & V_{\theta\theta} \end{bmatrix} = \begin{bmatrix} D_{\alpha\alpha}^{-1} & D_{\alpha\theta}^{-1} C_{\alpha\theta}(D_{\alpha\alpha}^{-1})^T \\ D_{\theta\alpha}^{-1} C_{\theta\alpha}(D_{\alpha\alpha}^{-1})^T & D_{\theta\theta}^{-1} C_{\theta\theta}(D_{\theta\theta}^{-1})^T \end{bmatrix}.$$  

The covariance matrix of the fixed effects is now the same as it was in the case where Gaussian estimation was not applied and is

$$V_{\alpha} = D_{\alpha\alpha}^{-1} = (X^T \Sigma^{-1} X)^{-1}.$$  

Now, however, the covariance matrix of the variance components is no longer $V_{\theta} = D_{\theta\theta}^{-1}$ but can be found using the formula

$$V_{\theta} = D_{\theta\theta}^{-1} C_{\theta\theta}(D_{\theta\theta}^{-1})^T$$  

and the asymptotic covariances between the fixed effects and variance components found from $V_{\alpha\theta} = V_{\theta\alpha}^T$. In Table 6.1 the sub-matrices of the covariance matrix $V$ are compared in the case where both Gaussian and Non-Gaussian estimation are applied.

<table>
<thead>
<tr>
<th>Table 6.1: Standard errors of the Fixed Effects and Variance Components using Gaussian and Non-Gaussian Estimation.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed Effect</strong></td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Fixed Effects</td>
</tr>
<tr>
<td>Variance Components</td>
</tr>
<tr>
<td>Fixed effects, Variance components</td>
</tr>
</tbody>
</table>

The matrix $D_{\theta\theta}$ can be computed as in the previous examples by inserting sample based estimates into the formula

$$(D_{\theta\theta})_{lm} = E \left( -\frac{\partial^2 L}{\partial \theta_l \partial \theta_m} \right) = \frac{1}{2} \text{trace} \left[ \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_l} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_m} \right]$$
to compute the \((l, m)\)th entry of \(D_{\theta\theta}\). However, as stated earlier, computing the matrix \(D_{\theta\theta}\) this way provides no guarantee that it will be positive-definite. If the standard errors of the variance components are required, the matrix \(D_{\theta\theta}\) must be positive-definite and hence invertible.

To ensure that the matrix \(D_{\theta\theta}\) is positive-definite it can be evaluated by computing \(\frac{\partial^2 L}{\partial \theta_l \partial \theta_m}\) directly by second-differencing to give a numerical estimate of \(D_{\theta\theta}\). The \((l, m)\)th entry in the matrix \(D_{\theta\theta}\) is computed as

\[
\left\{ \frac{\partial L}{\partial \theta_l \partial \theta_m} \right\}_{\omega = \hat{\omega}} = \frac{1}{\delta^2} \left\{ L(y_1; \hat{\omega} \mid \theta_l = \theta_l + \delta) + L(y_1; \hat{\omega} \mid \theta_m = \theta_m + \delta) - 2L(y_1; \hat{\omega}) \right\}
\]

which is evaluated at the maximum likelihood estimate \(\hat{\omega}\). Here, \(\delta\) is some arbitrarily chosen small value. Now as long as the parameter estimate \(\hat{\omega}\) produces a true maximum of the Gaussian log-likelihood function then computing \(D_{\theta\theta}\) in this manner ensures that it is positive-definite.

The vector \(y\), of length \(n\), can be partitioned into \(n_b\) sub-vectors \((y_1, y_2, \ldots, y_{n_b})\) which denote the measurements recorded on the \(n_b\) bars. Now, if \(y_i \sim N(\mu_i, \Sigma_i)\) and if \(\text{Cov}(y_i, y_j) = 0\) for all \(i \neq j\), the Gaussian score vector of equation (6.3) can be written as the sum of scores from \(n_b\) independent individuals

\[
q^\theta(y; \omega) = \sum_{i=1}^{n_b} q_i^\theta(y; \omega).
\]

The \((l, m)\)th entry in the matrix \(C_{\theta\theta}\) can then be conveniently evaluated using the formula

\[
(C_{\theta\theta})_{lm} = \sum_{i=1}^{n_b} q_i^\theta(y; \hat{\omega})q_i^\theta(y; \hat{\omega}). \tag{6.6}
\]

Here, the term \(q_i^\theta(y; \hat{\omega})\) can either be computed by substituting the sample based estimates into equation (6.5) or by differencing to give a numerical estimate. To calculate the vector \(q^\theta(y; \hat{\omega})\) by differencing, the \(l\)th element is computed as

\[
\left\{ \frac{\partial L}{\partial \theta_l} \right\}_{\omega = \hat{\omega}} = \frac{1}{\delta} \left\{ L(y; \hat{\omega} \mid \theta_l = \theta_l + \delta) - L(y; \hat{\omega}) \right\}.
\]

When the matrix \(C\) is evaluated in this manner it is automatically positive semi-definite. Calculating the matrices \(C_{\alpha\alpha}\) and \(C_{\theta\alpha}\) can be done in a similar way. The
(i, j)th component of the matrix $C_{\omega\omega}$ is

$$(C_{\omega\omega})_{ij} = \sum_{i=1}^{n_o} q^\omega(y; \omega) q^\omega(y; \omega)$$

where $q^\omega(y; \omega)$ is best evaluated using equation (6.4) and the vector $q^\omega(y; \omega)$ computed by differencing using equation (6.7).

When the log-likelihood function is the true log-likelihood function, i.e. the assumption of normality is correct, the matrices $C$ and $D$ are the same as one another. In the case when $C = D$ the covariance matrix $V = D^{-1}(C)(D^{-1})^T$ reduces to $V = D^{-1}$, i.e.

$$V = \begin{bmatrix} V_{\omega\omega} & V_{\omega\theta} \\ V_{\theta\omega} & V_{\theta\theta} \end{bmatrix} = \begin{bmatrix} D_{\omega\omega}^{-1} & 0 \\ 0 & D_{\theta\theta}^{-1} \end{bmatrix}.$$

Evidence has been raised that not only is the data non-normal but also that the covariance structure suggested is not ideal. The introduction of the matrix $C$ when estimating the standard errors gives some protection to the lack of normality and model misspecification.

Unfortunately, due to the complex structure of the covariance matrix $\Sigma$ it is unlikely that a partitioning of the vector $y$, such as $(y_1, y_2, \ldots, y_n)$, will exist so that $\text{Cov}(y_i, y_j) = 0$ for all $i \neq j$. When this is the case the Gaussian score vector $q^\omega(y; \omega)$ will not be able to be written as the sum of the score vectors from independent individuals. The result of this is that the matrix $C$ cannot easily be evaluated using equation (6.6) and the standard errors of the variance components will have to be found in some other way. In Section 6.3 the use of log-likelihood ratios is described for calculating confidence intervals of parametric functions.

### 6.3 Log-likelihood ratio

In the previous section it was shown that in certain situations how the standard errors of the variance components can be evaluated using Gaussian Estimation. However, this method will not usually be able to be applied because of the structure which the covariance matrix $\Sigma$ takes. Another technique, which does not make the assumption that the data is normally distributed, is now described for constructing confidence intervals of parametric functions of the variance components.
When the data is normally distributed the log-likelihood ratio possesses a chi-square distribution. Suppose that it is some parameter in \( \theta \), say \( \sigma \), which is to be examined. If the model parameters are \((\sigma, \alpha, \theta)\), where \( \theta \) no longer contains the parameter \( \sigma \), and the likelihood model is correct, then under \( H_0 : \sigma = \sigma_0 \)

\[
2 \left\{ L(y; \hat{\sigma}, \hat{\alpha}, \hat{\theta}) - L(y; \sigma_0, \hat{\alpha}_0, \hat{\theta}_0) \right\} \sim \chi^2_i
\]

for large \( n \), where \( L \) is the log-likelihood, \((\hat{\sigma}_0, \hat{\alpha}_0)\) are the MLE's under \( H_0 \) and \((\hat{\sigma}, \hat{\alpha}, \hat{\theta})\) are the unrestricted MLE's.

If the data does not follow a normal distribution or the model which has been fitted is incorrect then the log-likelihood ratio will be approximately chi-square if there is a large amount of data. Because all the measurements which were made are correlated the data effectively gives us a sample size of 1. However, since there is a large amount of data available it is reasonable to assume the log-likelihood ratio is approximately chi-square. Even though this is only an approximation it is probably safer than relying on the assumption that the log-likelihood function is quadratic around the maximum likelihood estimate.

Using log-likelihood ratios a 95% confidence interval for \( \sigma \) can be computed by solving

\[
L(y; \sigma_0, \alpha_0, \theta_0) = L(y; \hat{\sigma}, \hat{\alpha}, \hat{\theta}) - \frac{1}{2} \chi^2_{1(95\%)}
\]

for \( \sigma_0 \). This is demonstrated graphically in Fig. 6.2 where the value of 3.84 is used for \( \chi^2_{1(95\%)} \).

### 6.4 Example revisited

The model of Example 4.3 is again fitted to the data which consists of the measurements recorded on 10 chemical elements. The data can be found in Table A.1. Now, rather than constructing confidence intervals of the log laboratory and log 'laboratory x chemical' effects by using the standard errors evaluated from the observed information matrix, the confidence intervals will be constructed using the method of log-likelihood ratios described in Section 6.3. The data can be found in Table A.1. The profile log-likelihood function of the log laboratory effect \( \log(\sigma_L) \)
Figure 6.2: Profile log-likelihood function of \( \sigma \) showing how to construct the 95% confidence interval \((A, B)\) for \( \sigma \). Here, \( a = L(y; \hat{\delta}, \hat{\alpha}, \hat{\theta}) \) and \( b = a - 1.92 \).

is shown in Fig. 6.3. Using the method described in Section 6.3 a 95% confidence interval for \( \log(\sigma_L) \) is constructed as

\[
-\infty < \log(\sigma_L) < -2.357. \tag{6.7}
\]

Since there is no lower bound for this confidence interval it suggests that the laboratory component \( \sigma_L \) is in fact zero. This contradicts the earlier finding in Example 4.3, where the conclusion was that there existed a real difference between laboratories. The reason for this disagreement is that the standard error of the estimate of \( \sigma_L \), computed in Example 4.3, was based on the assumption that the log-likelihood function is quadratic around around the maximum likelihood estimate \( \hat{\omega} \). From Fig. 6.3 this is clearly not the case, so the confidence interval of equation (6.7), which is based on an asymptotic chi-square distribution for the log-likelihood ratio, is probably more reliable.

Fig. 6.4 shows the profile log-likelihood function of the log 'laboratory \( \times \) chemical' interaction component. Using this profile log-likelihood to construct a 95%
confidence interval for \( \log(\sigma_{LC}) \) yields

\[-2.875 < \log(\sigma_{LC}) < -2.261.\]

There is agreement here between the above confidence interval and the one constructed in Example 4.3. The conclusion in both cases is that the content of the chemical elements varies randomly between laboratories. This confidence interval is comparable to the one quoted earlier because of the quadratic nature of the profile log-likelihood function of \( \log(\sigma_{LC}) \) around the maximum likelihood estimate \( \hat{\omega} \).

Table 6.2 lists the confidence intervals constructed for the log laboratory effects and log 'laboratory x chemical' interaction effects. The numbers in parentheses correspond to the 95% upper and lower bounds quoted in Example 4.3.

For the fixed effects, the standard errors computed by Gaussian Estimation are the same as the standard errors quoted earlier, so the parametric tests performed with these standard errors still apply.
Figure 6.4: Profile log-likelihood function of $\log \sigma_{LC}$, $L(y; \omega | \sigma_{LC})$.

Table 6.2: Standard errors of the non-zero variance components using Gaussian Estimation. The 2nd half of this table gives 95% confidence intervals for the logs of these variance components.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLE</th>
<th>Lower Bound</th>
<th>Log</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_L$</td>
<td>$2.342 \times 10^{-2}$</td>
<td>$-\infty$ (-5.509)</td>
<td>-3.754</td>
<td>-2.357 (-2.000)</td>
</tr>
<tr>
<td>$\sigma_{LC}$</td>
<td>$7.825 \times 10^{-2}$</td>
<td>$-2.875$ (-2.822)</td>
<td>-2.548</td>
<td>-2.261 (-2.273)</td>
</tr>
</tbody>
</table>
Chapter 7

Further Examples

A suitable model has been suggested to carry out an analysis of the original problem of inter-laboratory comparisons. The data which this model has been fitted to has consisted of a set of measurements made by a number of laboratories on the content of chemical elements in alloy bars. In this chapter the model is fitted to two new sets of data. The first of these examples continues with the original theme of a chemical analysis of an alloy bar. The second example is then used to show how the model may be applied to analyse a broader range of problems.

7.1 Another chemical example

In a new experiment two manufacturing techniques were used to produce an alloy to some pre-defined specifications. These two techniques will be referred to as BICTA1 and BICTA2. The alloy is composed of 24 chemical elements which are of experimental interest. These chemical elements are: Aluminium, Boron, Carbon, Chromium, Cobalt, Copper, Hafnium, Iron, Manganese, Molybdenum, Niobium, Nitrogen (c), Nitrogen (r), Nitrogen (s), Oxygen (c), Oxygen (s), Phosphorus, Silicon, Sulphur, Tantalum, Titanium, Tungsten, Vanadium and Zirconium.

After the manufacturing techniques BICTA1 and BICTA2 had been used to produce two alloy bars fourteen laboratories each received a sample from the two bars. The names of these 14 laboratories are abbreviated to: 'CAME', 'FIRT', 'HOWM', 'INCO', 'TREL', 'L & S.', 'LGC', 'MOD', 'MURE', 'P.I.', 'R & C.',

95
'R.R.' and 'WILL.' On receiving the samples of the alloy the laboratories were asked to determine the content of the 24 chemical elements listed above. The participating laboratories used, between them, nine different methods for carrying out the chemical analysis of the alloy. These methods are: 'A.A', 'DCPT', 'GRAV', 'LECO', 'PCS', 'STCO', 'STRO', 'VTP' and 'XRF'. For each sample of the alloy the laboratories received their task was to make four measurements for the content of each of the chemical elements. These measurements were to be recorded in pairs; two measurements recorded on two different samples of the alloy. Some of the laboratories only recorded one pair of measurements and some failed to make any measurements at all for the content of some chemical elements. A small number of laboratories recorded three pairs of measurements for the content of certain chemical elements. All of the results reported were however recorded in pairs.

The measurements resulting from the two manufacturing processes, BICTA1 and BICTA2, will be analysed separately. The reasoning behind this is that the results from the two statistical analyses can be compared for agreement. Also, as there is a large amount of data, this will make fitting the model less demanding computationally. A total of 728 measurements were recorded on the alloy manufactured using the BICTA1 process and 817 measurements were recorded on the alloy manufactured from the BICTA2 process. Hence, there are 1568 measurements in total.

The measurements made on aluminium in the alloy which was manufactured using the BICTA1 process are given in Table 7.1. Here, only eleven of the fourteen laboratories measured the content of aluminium and, in the laboratories which did make these measurements only three methods were used. Looking at this table it can be seen that two pairs of measurements were recorded in most cases but for (A.A.,INCO) and (V.T.P,INCO) only one set of measurements was made.

As usual, the linear model, which describes the content of the chemical elements, contains an overall mean, a random error term and a selection of fixed and random effects. The methods which were used to carry out the chemical analysis and the chemicals will be treated as fixed effects. Also included in the fixed effects is a 'method × chemical' interaction term. As in the previous chemical examples, where the laboratories were regarded as just a few chosen at random from a large
Table 7.1: The percentage content of aluminium determined by a number of laboratories in samples of an alloy bar.

<table>
<thead>
<tr>
<th>ELEMENT : ALUMINIUM</th>
<th>BICTA1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>Company</td>
</tr>
<tr>
<td>A.A. BAS</td>
<td>BAS</td>
</tr>
<tr>
<td>FIET</td>
<td>FIN</td>
</tr>
<tr>
<td>INCO</td>
<td>INCO</td>
</tr>
<tr>
<td>HOWM</td>
<td>HOWM</td>
</tr>
<tr>
<td>L &amp; S</td>
<td>L &amp; S</td>
</tr>
<tr>
<td>MURE</td>
<td>MURE</td>
</tr>
<tr>
<td>P.I.</td>
<td>P.I.</td>
</tr>
<tr>
<td>X.R.F.</td>
<td>X.R.F.</td>
</tr>
<tr>
<td>V.T.P.</td>
<td>V.T.P.</td>
</tr>
<tr>
<td>MOD</td>
<td>MOD</td>
</tr>
<tr>
<td>R &amp; C</td>
<td>R &amp; C</td>
</tr>
<tr>
<td>R.R.</td>
<td>R.R.</td>
</tr>
<tr>
<td>WILL</td>
<td>WILL</td>
</tr>
</tbody>
</table>

population of similar establishments, the laboratories will be treated as random effects.

In general there will be $n_l$ different laboratories, using $n_m$ different methods to measure the content of $n_c$ chemical elements. Rather than considering the case where the measurements were recorded in pairs of two, a more general model for the case where measurements are recorded in $n_p$ groups is

$$y_{mlpc} = \mu + \alpha_{mc}^M + \alpha_{mc}^C + \alpha_{mc}^{MC} + \alpha_{lp}^L + \alpha_{lp}^P + \alpha_{lp}^R + \epsilon_{mlpc}.$$  

The random effects $\alpha_{lp}^P$ and $\alpha_{lp}^R$ have variances $\sigma_P^2$ and $\sigma_R^2$ and may be used to assess between and within pair variation respectively. The usual constraints and assumptions are

$$\sum_{m=1}^{n_m} \alpha_{mc}^M = \sum_{c=1}^{n_c} \alpha_{mc}^C = \sum_{m=1}^{n_m} \sum_{c=1}^{n_c} \alpha_{mc}^{MC} = 0 \quad (7.1)$$

97
and 
\[ a^L \sim N(0, \sigma^2_L), \quad a^P \sim N(0, \sigma^2_P), \quad a^R \sim N(0, \sigma^2_R). \]

For a complete-data layout, the linear model may be written as
\[
y = \mu + X_1 \alpha^M + X_2 \alpha^C + X_3 \alpha^{MC} + Z_1 a^L + Z_2 a^P + Z_3 a^R + e
\]

where the \( X \) and \( Z \) matrices are set up in the same way as in the previous chemical examples. The matrices \( X_1, X_2 \) and \( X_3 \) are

\[
X_1 = C_{n_m} \otimes 1_{n_l} \otimes 1_{n_p} \otimes 1_{n_r} \otimes 1_{n_e}, \quad X_2 = 1_{n_m} \otimes C_{n_l} \otimes 1_{n_p} \otimes 1_{n_r} \otimes 1_{n_e}, \\
X_3 = C_{n_m} \otimes 1_{n_l} \otimes 1_{n_p} \otimes 1_{n_r} \otimes C_{n_e},
\]

where the \( C \) matrices impose the constraints of equation (7.1). The \( Z \) matrices are set up as

\[
Z_1 = 1_{n_m} \otimes 1_{n_l} \otimes 1_{n_p} \otimes 1_{n_r} \otimes 1_{n_e}, \quad Z_2 = 1_{n_m} \otimes 1_{n_l} \otimes 1_{n_p} \otimes 1_{n_r} \otimes 1_{n_e}, \\
Z_3 = 1_{n_m} \otimes 1_{n_l} \otimes 1_{n_p} \otimes 1_{n_r} \otimes C_{n_e}.
\]

The error matrix \( E = \text{Cov}(e) \), again to ensure non-singularity, contains a set of down-weighting parameters and is set up as

\[
E = L_{n_m}^{(\phi_1)} \otimes I_{n_l}^{(\phi_2)} \otimes L_{n_p}^{(\phi_2)} \otimes L_{n_r}^{(\phi_2)} \otimes R
\]

where the matrix \( R \) is set up so that it possesses an equicorrelated structure.

This model can then be reduced to the dimension of the data, as in Section 5.1.1, and the incomplete-data model fitted to the data. The results of fitting this model to the BICTA1 and BICTA2 sets of data are as follows.

**BICTA1**

The set of measurements recorded for BICTA1 did not include any measurements made for the contents of the chemical elements copper, iron, manganese, nitrogen (c) and nitrogen (s). The laboratory ‘IREL’ did not analyse the BICTA1 sample of the alloy at all, but the remaining 13 laboratories which did used all 9 of the methods available. Therefore, \( n_c = 19, n_I = 13, n_m = 9, n_p = 3 \) and \( n_r = 2 \). If the data formed a complete-data layout 13438 measurements would have been made.
As only 733 measurements are available there is a large amount of missing data and the incomplete-data model is over-parametrised. The incomplete-data model is therefore updated accordingly by modification of the design matrices $X$ and $Z$ as in Section 5.1.

The Wald test for testing for a method effect, based on $W = c^T V c$ with $c = (\alpha_1^M, \ldots, \alpha_8^M)$, yields a value $W = 24.63$ on 8 degrees of freedom. This suggests that there is a real difference between the methods of measuring the content of the chemical elements. Testing for a ‘method × chemical’ interaction is also done using the Wald test. The test statistic is $W = c^T V c$ with $c = (\alpha_{11}^{MC}, \ldots, \alpha_{6(16)}^{MC})$. This gives a value of $W = 36.92$ which, when referred to $\chi^2_{144}$, provides evidence that the chemical content does not vary between measuring methods.

Estimates of the variance components, along with their standard errors, are given in Table 7.2. The estimate of 0.000 for the effect $\sigma_L$ suggests that there are no real differences between the laboratories. Confidence intervals (95%), computed from both the standard errors and the profile log-likelihoods, for $\sigma_P$ and $\sigma_R$ are similar. The confidence intervals, based on the standard errors, for the log variance components $\log(\sigma_P)$ and $\log(\sigma_R)$ are

$$-45.195 < \log(\sigma_P) < 24.184, \quad -77.288 < \log(\sigma_R) < 53.646$$

**BICTA2**

For the sample of the alloy manufactured from the BICTA2 process no measurements were recorded for nitrogen (r). However, all of the laboratories made measurements and between them they used all but one of the measuring methods. No measurements were made using the method ‘STRO’. Hence, $n_c = 23$, $n_t = 14$, $n_m = 8$, $n_p = 3$ and $n_r = 2$.

Wald tests for testing for a method and ‘method × chemical’ interaction effects yield the same conclusions as with the BICTA1 sample of the alloy. That is, the method effect is necessary in the model but there is no need to include a ‘method × chemical’ interaction term. Estimates of the variance components and their standard errors are given in Table 7.3. The estimates of $\sigma_P$ and $\sigma_R$ are both zero, suggesting that the inclusion of the corresponding random effects are not necessary.
Table 7.2: Maximum likelihood estimates of the variance components and their standard errors computed from the observed information matrix.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_L$</td>
<td>0.000</td>
<td>---</td>
</tr>
<tr>
<td>$\sigma_P$</td>
<td>$2.745 \times 10^{-5}$</td>
<td>$4.858 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\sigma_R$</td>
<td>$7.347 \times 10^{-6}$</td>
<td>$2.454 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_1$ (Al)</td>
<td>$4.450 \times 10^{-3}$</td>
<td>$9.099 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_2$ (B)</td>
<td>$139.569$</td>
<td>$26.4$</td>
</tr>
<tr>
<td>$r_3$ (C)</td>
<td>$6.070 \times 10^{-6}$</td>
<td>$1.179 \times 10^{-5}$</td>
</tr>
<tr>
<td>$r_4$ (Cr)</td>
<td>$3.215 \times 10^{-3}$</td>
<td>$5.925 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_5$ (Co)</td>
<td>$9.736 \times 10^{-3}$</td>
<td>$1.841 \times 10^{-3}$</td>
</tr>
<tr>
<td>$r_6$ (Hf)</td>
<td>$2.466 \times 10^{-3}$</td>
<td>$8.291 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_7$ (Mo)</td>
<td>$4.017 \times 10^{-4}$</td>
<td>$7.464 \times 10^{-5}$</td>
</tr>
<tr>
<td>$r_8$ (Nb)</td>
<td>$1.772 \times 10^{-3}$</td>
<td>$4.921 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_9$ (N(r))</td>
<td>$11.864$</td>
<td>$2.59$</td>
</tr>
<tr>
<td>$r_{10}$ (O(c))</td>
<td>$0.231$</td>
<td>$0.164$</td>
</tr>
<tr>
<td>$r_{11}$ (O(s))</td>
<td>$7.308$</td>
<td>$1.950$</td>
</tr>
<tr>
<td>$r_{12}$ (P)</td>
<td>$43.379$</td>
<td>$43.100$</td>
</tr>
<tr>
<td>$r_{13}$ (Si)</td>
<td>$7.703 \times 10^{-5}$</td>
<td>$1.762 \times 10^{-5}$</td>
</tr>
<tr>
<td>$r_{14}$ (S)</td>
<td>$16.451$</td>
<td>$3.080$</td>
</tr>
<tr>
<td>$r_{15}$ (Ta)</td>
<td>$2.559 \times 10^{-3}$</td>
<td>$7.284 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_{16}$ (Ti)</td>
<td>$1.081 \times 10^{-3}$</td>
<td>$2.487 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_{17}$ (W)</td>
<td>$3.928 \times 10^{-3}$</td>
<td>$8.205 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_{18}$ (V)</td>
<td>$7.407 \times 10^{-4}$</td>
<td>$1.330 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_{19}$ (Zr)</td>
<td>$5.009 \times 10^{-5}$</td>
<td>$1.893 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\rho$</td>
<td>$-2.465 \times 10^{-2}$</td>
<td>$2.891 \times 10^{-2}$</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>$1.026 \times 10^{-4}$</td>
<td>$2.435 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>$0.000$</td>
<td>---</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>$0.428$</td>
<td>$1.639 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>$0.430$</td>
<td>$1.607 \times 10^{-3}$</td>
</tr>
</tbody>
</table>
in the model.

A 95% confidence interval for \( \log(\sigma_L) \), based on the standard errors from the observed information matrix is

\[-7.046 < \log(\sigma_L) < -3.465.\]

The corresponding confidence interval based on the profile log-likelihood function is

\[-6.354 < \log(\sigma_L) < -4.773.\]

Both of these confidence intervals suggest that the parameter \( \sigma_L \) is non-zero and that there exists a difference between laboratories.

Fig. 7.1 shows plots of the standardised residuals for the two sets of data. Fig. 7.1(a) corresponds to the standardised residuals from fitting the model to the BICTA1 set of data and Fig. 7.1(b) corresponds to the BICTA2 set of data. Fig. 7.1(b) suggests that the BICTA2 set of data exhibits non-normality and that the use of log-likelihood ratios to construct confidence intervals is necessary.

## 7.2 Two-Site Sound Test Data

In all the examples so far, the data has arisen from a chemical analysis of an alloy material. The linear models in the earlier chapters were developed specially to model this chemical data. It will now be demonstrated how this model may be used to model data which arises from other types of experiments.

In broadcasting, when a stereo sound is transmitted from one location to another it is first encoded into a suitable format for transmission. The encoded sound is then broadcast from a transmitter and picked up by a receiver in a second location. The received signal is then decoded to give a reproduced copy of the original sound.

A broadcasting company has four techniques for encoding a stereo sound into a format which is suitable for transmission. This company wishes to identify which of these methods, if any, is the best. The four coding methods will be referred to as ‘CODEC 1’, ‘CODEC 2’, ‘CODEC 3’ and ‘CODEC 4’. Experimental trials
Table 7.3: Maximum likelihood estimates of the variance components and their standard errors computed from the observed information matrix.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_L$</td>
<td>$5.22 \times 10^{-3}$</td>
<td>$4.769 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\sigma_P$</td>
<td>0.000</td>
<td>---</td>
</tr>
<tr>
<td>$\sigma_R$</td>
<td>0.000</td>
<td>---</td>
</tr>
<tr>
<td>$r_1$ (Al)</td>
<td>$9.288 \times 10^{-3}$</td>
<td>$1.269 \times 10^{-3}$</td>
</tr>
<tr>
<td>$r_2$ (B)</td>
<td>106.473</td>
<td>29.632</td>
</tr>
<tr>
<td>$r_3$ (C)</td>
<td>$1.374 \times 10^{-3}$</td>
<td>$2.233 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_4$ (Cr)</td>
<td>$6.105 \times 10^{-3}$</td>
<td>$3.426 \times 10^{-3}$</td>
</tr>
<tr>
<td>$r_5$ (Co)</td>
<td>$5.625 \times 10^{-2}$</td>
<td>$2.943 \times 10^{-3}$</td>
</tr>
<tr>
<td>$r_6$ (Cu)</td>
<td>$6.520 \times 10^{-4}$</td>
<td>$9.876 \times 10^{-5}$</td>
</tr>
<tr>
<td>$r_7$ (Hf)</td>
<td>$8.694 \times 10^{-6}$</td>
<td>$2.734 \times 10^{-8}$</td>
</tr>
<tr>
<td>$r_8$ (Fe)</td>
<td>$1.181 \times 10^{-2}$</td>
<td>$7.347 \times 10^{-3}$</td>
</tr>
<tr>
<td>$r_9$ (Mn)</td>
<td>$1.762 \times 10^{-3}$</td>
<td>$6.297 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_{10}$ (Mo)</td>
<td>$4.659 \times 10^{-3}$</td>
<td>$1.476 \times 10^{-3}$</td>
</tr>
<tr>
<td>$r_{11}$ (Nb)</td>
<td>$3.494 \times 10^{-3}$</td>
<td>$8.469 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_{12}$ (N(c))</td>
<td>9.774</td>
<td>2.109</td>
</tr>
<tr>
<td>$r_{13}$ (N(s))</td>
<td>3.816</td>
<td>0.935</td>
</tr>
<tr>
<td>$r_{14}$ (O(c))</td>
<td>0.694</td>
<td>0.297</td>
</tr>
<tr>
<td>$r_{15}$ (O(s))</td>
<td>5.815</td>
<td>1.219</td>
</tr>
<tr>
<td>$r_{16}$ (P)</td>
<td>55.603</td>
<td>47.347</td>
</tr>
<tr>
<td>$r_{17}$ (Si)</td>
<td>$4.149 \times 10^{-4}$</td>
<td>$7.694 \times 10^{-5}$</td>
</tr>
<tr>
<td>$r_{18}$ (S)</td>
<td>26.683</td>
<td>4.287</td>
</tr>
<tr>
<td>$r_{19}$ (Ta)</td>
<td>$1.162 \times 10^{-3}$</td>
<td>$7.381 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_{20}$ (Ti)</td>
<td>$7.945 \times 10^{-4}$</td>
<td>$5.829 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_{21}$ (W)</td>
<td>$7.470 \times 10^{-3}$</td>
<td>$5.324 \times 10^{-3}$</td>
</tr>
<tr>
<td>$r_{22}$ (V)</td>
<td>$1.243 \times 10^{-3}$</td>
<td>$4.921 \times 10^{-4}$</td>
</tr>
<tr>
<td>$r_{23}$ (Zr)</td>
<td>$3.632 \times 10^{-3}$</td>
<td>$6.234 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\rho$</td>
<td>$-2.046 \times 10^{-3}$</td>
<td>$2.724 \times 10^{-2}$</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.000</td>
<td>---</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>0.000</td>
<td>---</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>0.295</td>
<td>$3.493 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>0.689</td>
<td>$2.304 \times 10^{-3}$</td>
</tr>
</tbody>
</table>
(a) Measurements are the results of a chemical analysis on the BICTA1 sample of the alloy. (19 chemical elements are of interest)

(b) Measurements are the results of a chemical analysis on the BICTA2 sample of the alloy. (23 chemicals elements are of interest)

Figure 7.1: Normal plots of the standardised residuals.
were then carried out in order to assess which of the coding methods should be used.

The experiment was carried out at two separate sound studios, one in England and the other in Japan. These two studios were built to be as identical to one another as possible. The two studios will be referred to as location 1 and location 2. Fifty six volunteers were then recruited to take part in the study, thirty two at location 1 and twenty four at location 2. The experiment was then conducted as follows.

The subjects in the study were taken into one of the studios. Here, they heard a recording of a sound which will be called the reference sound. There were nine different sounds used, these were recordings of church bells and fire engines, etc. The sound was then encoded and decoded, as it would if it had been broadcast. This will be called the reproduced sound. The subject was then played the original sound and the reproduced sound without been told which was which. The original sound is identical to the reference sound. The subject was then asked to award a score, a value between 0 and 5, to each of the two sounds. If the subject felt that the sound was identical to the reference sound they would award that sound a score of 5. Whereas, if they thought there was no resemblance between the sound and the reference sound then they were to award it a score of 0. An overall score for that particular hearing was then calculated as the difference between the scores given to the reproduced sound and the original sound.

Table 7.4 lists the scores given by the 32 subjects at location 1 to sound (item) 1 which was encoded and decoded using the coding system CODEC 1. Columns 4 and 5 correspond to the scores given the original and reproduced sound and the final column gives the overall score. If a subject has a negative overall score then they felt that the original sound resembled the reference sound more than the reproduced sound. A positive overall score reflects the fact that the subject thought the reproduced sound was more like the reference sound than the original sound. An overall score of zero means that the subject thought the two sounds were identical. Therefore, for a coding system to be good an overall score close to zero is desirable.

A linear model to describe the overall score given by subject s, at location l, for
Table 7.4: Results from the two-site sound test study. This is for site 1, coding system 1, and sound (item) 1 with the 32 subjects.

<table>
<thead>
<tr>
<th>Site</th>
<th>Codec</th>
<th>Item</th>
<th>Subject</th>
<th>Reproduced</th>
<th>Original</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5.0</td>
<td>4.0</td>
<td>-1.0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>5.0</td>
<td>4.4</td>
<td>-0.6</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>5.0</td>
<td>2.9</td>
<td>-2.1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>4.3</td>
<td>5.0</td>
<td>0.7</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>5.0</td>
<td>3.7</td>
<td>-1.3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>5.0</td>
<td>4.5</td>
<td>-0.5</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>4.0</td>
<td>5.0</td>
<td>1.0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>4.8</td>
<td>5.0</td>
<td>0.2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>9</td>
<td>4.5</td>
<td>5.0</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>5.0</td>
<td>5.0</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>11</td>
<td>5.0</td>
<td>4.7</td>
<td>-0.3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>12</td>
<td>5.0</td>
<td>3.8</td>
<td>-1.2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>13</td>
<td>5.0</td>
<td>5.0</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>14</td>
<td>4.8</td>
<td>5.0</td>
<td>0.2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>15</td>
<td>5.0</td>
<td>3.8</td>
<td>-1.2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>16</td>
<td>5.0</td>
<td>2.0</td>
<td>-3.0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>17</td>
<td>5.0</td>
<td>4.9</td>
<td>-0.1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>18</td>
<td>5.0</td>
<td>4.4</td>
<td>-0.6</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>19</td>
<td>5.0</td>
<td>4.0</td>
<td>-1.0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>20</td>
<td>4.0</td>
<td>5.0</td>
<td>1.0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>21</td>
<td>5.0</td>
<td>4.0</td>
<td>-1.0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>22</td>
<td>5.0</td>
<td>5.0</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>23</td>
<td>2.0</td>
<td>5.0</td>
<td>3.0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>24</td>
<td>5.0</td>
<td>4.7</td>
<td>-0.3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>25</td>
<td>5.0</td>
<td>5.0</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>26</td>
<td>5.0</td>
<td>4.9</td>
<td>-0.1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>27</td>
<td>5.0</td>
<td>5.0</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>28</td>
<td>5.0</td>
<td>3.5</td>
<td>-1.5</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>29</td>
<td>5.0</td>
<td>3.0</td>
<td>-2.0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>30</td>
<td>4.0</td>
<td>5.0</td>
<td>1.0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>31</td>
<td>4.9</td>
<td>5.0</td>
<td>0.1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>32</td>
<td>4.0</td>
<td>5.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>
sound (item) \( i \), which has been encoded and decoded with coding system \( c \), is

\[
y_{\text{icis}} = \mu + \alpha_{ii}^L + \alpha_{ic}^C + \alpha_{i}^S + \alpha_{ci}^C + e_{\text{icis}}
\]  

(7.2)

with \( n_i = 2, n_c = 4, n_i = 9 \) and \( n_c = 56 \). This model includes a fixed location effect because it is thought that the scores awarded by the subjects at the two locations may vary. Also included in the fixed effects is an overall mean score and a coding system effect. The subjects participating in the study are just a few chosen at random from an infinite population of suitable candidates, so the model will include a random subject effect. Similarly, the sounds (items) which the subjects listened to were randomly chosen from a large number of possible sounds. As the sound (item) effect may vary randomly between the coding systems, a random ‘codec \times item’ effect is included. Other interactions are not included because it is thought that the coding systems and the sounds (items) do not differ between the two locations.

The usual constraints and assumptions are

\[
\sum_{i=1}^{n_i} \alpha_i^L = \sum_{c=1}^{n_c} \alpha_c^C = 0
\]

and

\[
\alpha_i^L \sim N(0, \sigma_i^2), \quad \alpha_i^S \sim N(0, \sigma_i^2), \quad \alpha_c^C \sim N(0, \sigma_c^2).
\]

The linear model of equation (7.2) may be written in matrix notation as

\[
y = 1\mu + X_1 \alpha^L + X_2 \alpha^C + Z_1 \alpha^I + Z_2 \alpha^S + Z_3 \alpha^CI + e,
\]

which is an expanded form of the Harville-Laird-Ware model

\[
y = X\alpha + Z\alpha + e
\]

with

\[
X = [1 \; X_1 \; X_2], \quad \alpha = (\mu \; \alpha^L \; \alpha^C)^T
\]

and

\[
Z = [Z_1 \; Z_2 \; Z_3], \quad \alpha = (\alpha^I \; \alpha^S \; \alpha^CI)^T
\]
The $X$ and $Z$ matrices are constructed the same way as in previous examples. The $X$ matrices are set up as

$$X_1 = C_{m_i} \otimes 1_{n_c} \otimes 1_{n_l} \otimes 1_{n_r}, \quad X_2 = 1_{n_i} \otimes C_{m_c} \otimes 1_{n_l} \otimes 1_{n_r},$$

where the $C$ matrices impose the constraints that the fixed effects sum to zero. The $Z$ matrices are set up as

$$Z_1 = 1_{n_i} \otimes 1_{n_c} \otimes 1_{n_l} \otimes 1_{n_r}, \quad Z_2 = 1_{n_i} \otimes 1_{n_c} \otimes 1_{n_l} \otimes 1_{n_r},$$

$$Z_3 = 1_{n_i} \otimes I_{nc} \otimes I_{nl} \otimes 1_{nr},$$

and the covariance matrix of the random effects $B$, as

$$B = \text{Cov}(a) = \text{diag}(\sigma_1^2 I_{n_f}, \sigma_2^2 I_{n_p}, \sigma_3^2 I_{nc_n}),$$

The error matrix $E = \text{Cov}(e)$ is set up as

$$E = I_{n_f}^{(\phi_1)} \otimes I_{nc}^{(\phi_2)} \otimes I_{nl}^{(\phi_3)} \otimes R,$$

to ensure that it is non-singular. It will be assumed that the overall scores awarded by the different subjects are uncorrelated. Assuming homogeneity of variance, the covariance matrix $R$ is set up as

$$R = \sigma^2 I_{n_r}.$$

Estimates of the fixed effects are

$$(\mu, \alpha_1^C, \alpha_2^C, \alpha_3^C) = (-0.267, -0.072, -0.246, -0.140, 0.115)$$

with standard errors 0.066, 0.049, 0.060, 0.060 and 0.060 respectively. To test for a location effect the normal deviate is calculated as $-1.478$. This suggests that there is no real difference between the two locations and that therefore the inclusion of a location effect in the model is unnecessary.

The Wald test for differences between coding systems, based on $W = c^TV_c$c with $c = (\alpha_1^C, \alpha_2^C, \alpha_3^C)$, gives $W = 15.1$ as $\chi^2_3$, suggesting that there is a real difference between the overall scores from different coding techniques. The estimates of the overall scores, corresponding to the four coding systems, are $-0.2916$ for CODEC 1, $-0.407$ for CODEC 2, $-0.152$ for CODEC 3 and $-0.3166$ for CODEC 4.

For a good coding system the scores for that coding system would be close to zero, i.e. the subjects cannot distinguish between the original and the reproduced
sound. To identify if any of the coding systems classed as 'good' a test can be carried by assessing whether the estimated overall scores for any of the coding systems is near zero. For CODEC 1 this is equivalent to testing the hypothesis \( H : \mu + \alpha_1^C = 0 \). The normal deviate resulting from this hypothesis is \( \frac{\mu + \alpha_1^C}{\text{SE}(\mu + \alpha_1^C)} \), which should be referred to normal tables. Computing the normal deviates for each of the coding systems yields 3.286, 4.583, 1.71 and 2.447 for CODEC 1, CODEC 2, CODEC 3 and CODEC 4 respectively. The only one of these values which is not significant, i.e. suggesting that the overall scores are zero, is 1.71 which corresponds to CODEC 3. This suggests, that for this coding system, the participants in the study felt that the reproduced sound was most like the original sound.

It is hoped that the test described above goes some way to assessing the quality of the four coding systems. However, this is not an ideal test because if a subject awarded both the original sound and the reproduced sound low scores then the overall score for that sound may be close to zero. This would mean that even if the subject thought that the reproduced sound was poor the coding system could still qualify as good under the criteria given above. Two more comprehensive tests will now be described.

Rather than using equation (7.2) to model the overall scores it might be an idea to model the scores awarded to the reproduced sounds. A more suitable test would then be to test if the scores awarded to the reproduced sounds are high, say greater than 4. For CODEC 1 this is equivalent to testing the hypothesis \( H : \mu + \alpha_1^C > 4 \), where \( \mu + \alpha_1^C \) are the scores corresponding to coding system CODEC 1.

An even more appropriate test could be constructed by adapting the linear model of equation (7.2) so that it models both the overall scores and the scores awarded to the reproduced sounds. The quality of a coding system could then be assessed by testing if the overall scores are small and the scores awarded to the reproduced sound are large.

Estimates of the variance components, along with their standard errors computed from the observed information matrix, are given in Table 7.5. The estimates of \( \log(\sigma_I) \), \( \log(\sigma_S) \), and \( \log(\sigma_C) \) are \(-2.266\), \(-1.757\) and \(-1.731\) and have standard errors \(0.509\), \(0.132\) and \(0.197\). Confidence intervals (95\%) for the logs of the
Table 7.5: Estimates of the variance components and their standard errors computed from the observed information matrix.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_I$</td>
<td>0.108</td>
<td>0.055</td>
</tr>
<tr>
<td>$\sigma_S$</td>
<td>0.319</td>
<td>0.042</td>
</tr>
<tr>
<td>$\sigma_{CI}$</td>
<td>0.173</td>
<td>0.034</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.976</td>
<td>0.015</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>$1.026 \times 10^{-4}$</td>
<td>$7.637 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>$4.729 \times 10^{-2}$</td>
<td>$1.235 \times 10^{-2}$</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>$1.032 \times 10^{-2}$</td>
<td>$8.926 \times 10^{-3}$</td>
</tr>
</tbody>
</table>

variance components, based on the standard errors from the observed information matrix, are computed as

$$-3.264 < \log(\sigma_I) < -1.268, \quad -2.016 < \log(\sigma_S) < -1.498,$$
$$-2.117 < \log(\sigma_{CI}) < -1.345.$$  \hfill (7.3)

These confidence intervals suggest that $\sigma_I$, $\sigma_S$ and $\sigma_{CI}$ are all non-zero and are necessary in the full model. However, as pointed out in Chapter 6 the profile log-likelihoods of the variance components tend to be non-quadratic, so the standard errors evaluated from the observed information matrix may be unreliable. The profile log-likelihoods of $\log(\sigma_I)$, $\log(\sigma_S)$ and $\log(\sigma_{CI})$ are shown in Fig 7.2. It is clearly evident from Fig. 7.2(a) that the profile log-likelihood of $\log(\sigma_I)$ is non-quadratic and that the confidence interval for $\log(\sigma_I)$ quoted in equation (7.3) may be incorrect.

Confidence intervals for the logs of the variance components based on the profile log-likelihoods are as follows:

$$-\infty < \log(\sigma_I) < -1.456, \quad -2.263 < \log(\sigma_S) < -1.341,$$
$$-2.172 < \log(\sigma_{CI}) < -1.143.$$  \hfill (7.3)

These confidence intervals suggest that, as before, the variance components $\sigma_S$ and $\sigma_{CI}$ are non-zero. Now, however, since there is no lower bound for the 95% confidence interval of $\log(\sigma_I)$, it appears that the component $\sigma_I$ is not actually required in the model.

A normal plot of the standardised residuals is given in Fig. 7.3. There is nothing here to suggest that any of the scores are outliers. However this plot does suggest
that the data is not normally distributed and that the use of log-likelihood ratios for the construction of confidence intervals is appropriate.
Figure 7.3: Normal plot of the standardised residuals from fitting the model to two-site test data.
Chapter 8

Conclusions and further work

In this project it has been shown how data resulting from an inter-laboratory experiment, where all the measurements are correlated, can be analysed. The four main objectives of this work can be summarised as follows:

i) Examine the data and decide upon a suitable statistical model to describe it.

ii) Fit the model to the data using some appropriate statistical method, in this case maximum likelihood estimation.

iii) Perform goodness-of-fit assessments and, if necessary, modify the model.

iv) Examine the results from fitting the model to answer a number of important questions, including assessing any significant differences between laboratories.

For the chemical examples a linear model was suggested to describe the percentage contents of the chemical elements. This model included fixed and random main effects as well as higher order interaction terms. To complete the model an error matrix $E$ was suggested. This matrix had to be carefully chosen so that the covariance matrix of the data $\Sigma$ was positive-semi definite.

Several different structures were considered for the matrix $R$ which describes the covariances between the percentage content of chemical elements measured at the same position, in the same bar and by the same laboratory. It was decided that
the most suitable structure for this matrix was to assume that the errors were equicorrelated. However, in Section 4.4 it was shown that this assumption may be incorrect. This problem was overcome by using Gaussian Estimation and log-likelihood ratios to construct confidence intervals for the model parameters. It may be worthwhile pursuing the problem of modelling the matrix $R$ more precisely.

One of the major objectives of this work was to fit a suitable statistical model to the data. Although the maximum likelihood estimates of the model parameters could have been evaluated using existing statistical computer software, the use of specially constructed computer algorithms makes the estimation of the model parameters more feasible. The two main reasons for this are as follows. The first is that the maximisation of the likelihood function is a computationally intensive task and so the use of efficient computer algorithms is necessary so that the model can be fitted to the data in a reasonable amount of time. Secondly, because of the complex structure of the model, especially the covariance matrix, it is much simpler to construct the statistical model using the specially written computer algorithms than trying to adapt existing computer software.

Although, for the examples which have been considered in this work, estimation of the model parameters was carried out in a reasonable amount of time, this will not be the case when larger amounts of data are available. If the models suggested are to be fitted to larger amounts of data then alterations have to be made to the procedure of parameter estimation. This could involve the construction of more optimal computer algorithms or finding algebraic expressions for the covariance matrix.

Parametric tests were used to assess the fixed and random effects. Ideally, for the random effects, the standard errors would be evaluated from the sandwich matrix $V_\theta = D_\theta^{-1}C_\theta(D_\theta^{-1})^T$. It was shown in Section 6.2 that this approach was not possible in many examples. This problem was overcome by using log-likelihood ratios to construct confidence intervals for the variance components. However, this is a computationally intensive task and the use of log-likelihood ratios may not be feasible for larger sets of data. One possible solution to this problem is to use modified computer algorithms and algebraic expressions for the covariance matrices, as above, to reduce the computational demands. A second
approach may be to evaluate the matrix $C$ formally, via its algebraic form, and use the resulting standard errors to construct confidence intervals. The matrix $C$ is, however, complex and would require, among other things, the evaluation of $E\{(y - \mu)(y - \mu)^T(y - \mu)(y - \mu)^T\}$.

In all of the examples looked at in Chapters 2 through to 7 normal plots of the standardised residuals were used to assess the goodness-of-fit of the model being used. Investigation of these plots was left entirely to an individual's discretion. The use of more formal methods for assessing the goodness-of-fit of the model could be explored.
Appendix A

Appendix

A.1 Calculation of Expected Mean Squares for model 2.1

Position effect is fixed, bar and laboratory effects are random

Model: $y_{blp} = \mu + \alpha_p + a_L + a_B + e_{blp}$

with $a_L \sim N(0, \sigma^2_L)$, $a_B \sim N(0, \sigma^2_B)$ and $e_{blp} \sim N(0, \sigma^2)$.

1. Total sum of squares: $TSS = \sum_{blp} y_{blp}^2 - \frac{\bar{y}_{++}^2}{n_b n_m n_p}$

   $E(y_{blp}) = \mu + \alpha_p$, $V(y_{blp}) = \sigma^2_L + \sigma^2_B + \sigma^2$, $E(y_{++}) = n_b n_m n_p \mu$

   $V(y_{++}) = n_b^2 n_m^2 n_p \sigma^2_L + n_b^2 n_m^2 n_p \sigma^2_B + n_b n_m n_p \sigma^2$

   Using the identity $E(Y^2) = \{E(Y)\}^2 + V(Y)$,

   $E(TSS) = n_b n_m \sum_p (\alpha_p^2) + n_b n_m (n_t - 1) \sigma^2_L + n_b n_m (n_b - 1) \sigma^2_B + (n_b n_m n_p) \sigma^2$

2. Sum of squares for bars: $SSB = \frac{1}{n_b n_m} \sum_b y_b^{++} - \frac{\bar{y}_{++}^2}{n_b n_m n_p}$

   $E(y_b^{++}) = n_b n_p \mu$, $V(y_b^{++}) = n_b^2 n_p \sigma^2_L + n_b^2 n_p \sigma^2_B + n_b n_p \sigma^2$

   $E(SSB) = (n_b - 1)(n_b n_p \sigma^2_B + \sigma^2)$

115
\[ E(MSB) = n_p \sigma_B^2 + \sigma^2 \]

3. Sums of squares for labs: \(SSL = \frac{1}{n_b n_p} \sum_i y_i^2 - \frac{\bar{y}_{++}^2}{n_b n_p} \)

\[ E(y_{++}) = n_b n_p \mu, \quad V(y_{++}) = n_b^2 n_p^2 \sigma_L^2 + n_b^2 n_p \sigma_B^2 + n_b n_p \sigma^2 \]

\[ E(SSL) = (n_l - 1)(n_b n_p \sigma_L^2 + \sigma^2) \]

\[ E(MSL) = n_b n_p \sigma_L^2 + \sigma^2 \]

4. Sums of squares for positions: \(SSP = \frac{1}{n_b n_l} \sum_p y_{++p}^2 - \frac{\bar{y}_{++}^2}{n_b n_l} \)

\[ E(y_{++p}) = n_b n_l \mu + n_b n_l \alpha_p^p, \quad V(y_{++p}) = n_b^2 n_l \sigma_L^2 + n_b^2 n_l \sigma_B^2 + n_b n_l \sigma^2 \]

\[ E(SSP) = (n_p - 1)\sigma^2 + n_b n_l \sum_p (\alpha_p^p)^2 \]

\[ E(MSP) = \sigma^2 + \frac{n_b n_l}{n_p - 1} \sum_p (\alpha_p^p)^2 \]

5. Residual sum of squares found by subtraction.

\[ E(SSR) = E(TSS) - E(SSB) - E(SSL) - E(SSP) \]

\[ E(RSS) = (n_b n_l n_p - n_b - n_l - n_p + 2)\sigma^2 \]

\[ E(MSR) = \sigma^2 \]
### A.2 Chemical Data

Table A.1: The percentage content of 17 chemical elements in four bars measured at two positions by three laboratories. The 17 chemical elements under study are as follows: Aluminium (Al), Cobalt (Co), Chromium (Cr), Copper (Cu), Iron (Fe), Hafnium (Hf), Manganese (Mn), Molybdenum (Mo), Niobium (Nb), Nickel (Ni), Phosphorus (Ph), Silicon (Si), Tantalum (Ta), Titanium (Ti), Vanadium (V), Tungsten (W), Zirconium (Zr).

<table>
<thead>
<tr>
<th></th>
<th>Laboratory 1</th>
<th></th>
<th>Laboratory 1</th>
<th></th>
<th>Laboratory 1</th>
<th></th>
<th>Laboratory 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Top</td>
<td>Bottom</td>
<td>Top</td>
<td>Bottom</td>
<td>Top</td>
<td>Bottom</td>
<td>Top</td>
</tr>
<tr>
<td>Al</td>
<td>2.950</td>
<td>2.930</td>
<td>2.930</td>
<td>2.920</td>
<td>2.830</td>
<td>2.920</td>
<td>2.920</td>
</tr>
<tr>
<td>Cr</td>
<td>5.270</td>
<td>5.250</td>
<td>5.260</td>
<td>5.230</td>
<td>5.200</td>
<td>5.150</td>
<td>5.240</td>
</tr>
<tr>
<td>Cu</td>
<td>0.860</td>
<td>0.860</td>
<td>0.850</td>
<td>0.860</td>
<td>0.850</td>
<td>0.850</td>
<td>0.840</td>
</tr>
<tr>
<td>Hf</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mn</td>
<td>1.050</td>
<td>1.050</td>
<td>1.050</td>
<td>1.040</td>
<td>1.050</td>
<td>1.050</td>
<td>1.030</td>
</tr>
<tr>
<td>Mo</td>
<td>1.050</td>
<td>1.070</td>
<td>1.050</td>
<td>1.050</td>
<td>1.050</td>
<td>1.050</td>
<td>1.050</td>
</tr>
<tr>
<td>Nb</td>
<td>1.000</td>
<td>1.010</td>
<td>1.000</td>
<td>1.020</td>
<td>0.990</td>
<td>1.020</td>
<td>0.960</td>
</tr>
<tr>
<td>Ni</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ph</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Si</td>
<td>0.490</td>
<td>0.490</td>
<td>0.500</td>
<td>0.510</td>
<td>0.480</td>
<td>0.490</td>
<td>0.480</td>
</tr>
<tr>
<td>Ta</td>
<td>0.810</td>
<td>0.840</td>
<td>0.820</td>
<td>0.840</td>
<td>0.810</td>
<td>0.810</td>
<td>0.800</td>
</tr>
<tr>
<td>Ti</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.990</td>
<td>1.000</td>
<td>0.970</td>
</tr>
<tr>
<td>V</td>
<td>1.010</td>
<td>1.010</td>
<td>1.000</td>
<td>1.010</td>
<td>1.000</td>
<td>0.970</td>
<td>1.000</td>
</tr>
<tr>
<td>Zr</td>
<td>0.522</td>
<td>0.545</td>
<td>0.525</td>
<td>0.528</td>
<td>0.492</td>
<td>0.571</td>
<td>0.453</td>
</tr>
<tr>
<td></td>
<td>Bar I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td>Top</td>
<td>Bottom</td>
<td>Top</td>
<td>Bottom</td>
<td>Top</td>
<td>Bottom</td>
<td>Top</td>
</tr>
<tr>
<td>Al</td>
<td>2.930</td>
<td>2.930</td>
<td>2.930</td>
<td>2.930</td>
<td>2.870</td>
<td>2.910</td>
<td>2.900</td>
</tr>
<tr>
<td>Cr</td>
<td>5.220</td>
<td>5.210</td>
<td>5.230</td>
<td>5.210</td>
<td>5.220</td>
<td>5.200</td>
<td>5.200</td>
</tr>
<tr>
<td>Cu</td>
<td>1.050</td>
<td>1.060</td>
<td>1.050</td>
<td>1.040</td>
<td>1.060</td>
<td>1.060</td>
<td>1.560</td>
</tr>
<tr>
<td>Fe</td>
<td>5.240</td>
<td>5.220</td>
<td>5.240</td>
<td>5.210</td>
<td>5.210</td>
<td>5.190</td>
<td>5.240</td>
</tr>
<tr>
<td>Hf</td>
<td>0.060</td>
<td>0.060</td>
<td>0.060</td>
<td>0.060</td>
<td>0.060</td>
<td>0.060</td>
<td>0.060</td>
</tr>
<tr>
<td>Mn</td>
<td>1.130</td>
<td>1.140</td>
<td>1.120</td>
<td>1.140</td>
<td>1.130</td>
<td>1.140</td>
<td>1.120</td>
</tr>
<tr>
<td>Mo</td>
<td>1.050</td>
<td>1.050</td>
<td>1.150</td>
<td>1.050</td>
<td>1.150</td>
<td>1.050</td>
<td>1.150</td>
</tr>
<tr>
<td>Nb</td>
<td>1.020</td>
<td>1.030</td>
<td>1.020</td>
<td>1.030</td>
<td>1.020</td>
<td>1.050</td>
<td>0.990</td>
</tr>
<tr>
<td>Ni</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ph</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Si</td>
<td>0.467</td>
<td>0.476</td>
<td>0.459</td>
<td>0.477</td>
<td>0.450</td>
<td>0.485</td>
<td>0.454</td>
</tr>
<tr>
<td>Ta</td>
<td>0.940</td>
<td>0.950</td>
<td>0.940</td>
<td>0.930</td>
<td>0.940</td>
<td>0.940</td>
<td>0.910</td>
</tr>
<tr>
<td>Ti</td>
<td>1.000</td>
<td>1.010</td>
<td>1.000</td>
<td>1.010</td>
<td>1.000</td>
<td>1.010</td>
<td>0.980</td>
</tr>
<tr>
<td>V</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>W</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Zr</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Bar I</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Top</td>
<td>Bottom</td>
<td>Top</td>
<td>Bottom</td>
<td>Top</td>
<td>Bottom</td>
<td>Top</td>
</tr>
<tr>
<td>Al</td>
<td>2.980</td>
<td>2.980</td>
<td>2.970</td>
<td>2.970</td>
<td>2.960</td>
<td>2.940</td>
<td>2.970</td>
</tr>
<tr>
<td>Cr</td>
<td>5.380</td>
<td>5.340</td>
<td>5.370</td>
<td>5.360</td>
<td>5.360</td>
<td>5.370</td>
<td>5.380</td>
</tr>
<tr>
<td>Cu</td>
<td>1.020</td>
<td>1.040</td>
<td>1.020</td>
<td>1.030</td>
<td>1.030</td>
<td>1.030</td>
<td>1.030</td>
</tr>
<tr>
<td>Fe</td>
<td>5.240</td>
<td>5.200</td>
<td>5.230</td>
<td>5.230</td>
<td>5.230</td>
<td>5.230</td>
<td>5.240</td>
</tr>
<tr>
<td>Hf</td>
<td>0.050</td>
<td>0.060</td>
<td>0.050</td>
<td>0.060</td>
<td>0.050</td>
<td>0.060</td>
<td>0.050</td>
</tr>
<tr>
<td>Mn</td>
<td>1.100</td>
<td>1.100</td>
<td>1.100</td>
<td>1.100</td>
<td>1.100</td>
<td>1.100</td>
<td>1.100</td>
</tr>
<tr>
<td>Mo</td>
<td>1.040</td>
<td>1.030</td>
<td>1.040</td>
<td>1.040</td>
<td>1.040</td>
<td>1.040</td>
<td>1.040</td>
</tr>
<tr>
<td>Nb</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ni</td>
<td>1.050</td>
<td>1.060</td>
<td>1.040</td>
<td>1.050</td>
<td>1.050</td>
<td>1.070</td>
<td>1.010</td>
</tr>
<tr>
<td>Ph</td>
<td>0.015</td>
<td>0.016</td>
<td>0.014</td>
<td>0.014</td>
<td>0.016</td>
<td>0.016</td>
<td>0.014</td>
</tr>
<tr>
<td>Si</td>
<td>0.490</td>
<td>0.500</td>
<td>0.470</td>
<td>0.470</td>
<td>0.500</td>
<td>0.510</td>
<td>0.480</td>
</tr>
<tr>
<td>Ta</td>
<td>0.960</td>
<td>0.960</td>
<td>0.960</td>
<td>0.960</td>
<td>0.970</td>
<td>0.970</td>
<td>0.950</td>
</tr>
<tr>
<td>Ti</td>
<td>1.020</td>
<td>1.030</td>
<td>1.010</td>
<td>1.020</td>
<td>1.020</td>
<td>1.030</td>
<td>1.020</td>
</tr>
<tr>
<td>V</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
<td>0.970</td>
<td>0.960</td>
<td>0.970</td>
</tr>
<tr>
<td>W</td>
<td>5.050</td>
<td>5.000</td>
<td>5.010</td>
<td>5.000</td>
<td>4.990</td>
<td>4.990</td>
<td>5.060</td>
</tr>
<tr>
<td>Zr</td>
<td>0.480</td>
<td>0.520</td>
<td>0.470</td>
<td>0.500</td>
<td>0.470</td>
<td>0.530</td>
<td>0.420</td>
</tr>
</tbody>
</table>
References


