10.1 An example

In earlier chapters we have defined the model linear in the parameters, commonly called the linear model, and discussed the theory of estimating parameters of the model by the method of least squares. Our applications of the theory so far have been to simple linear regression, polynomial regression and multiple regression. However, there is a very wide class of models which can be analysed on the basis of the theory we have covered. In this chapter we apply our theory to data arranged in balanced arrays or cross classifications; it is part of a very big topic usually referred to as ‘the analysis of variance’. The following example shows the type of problem we shall be dealing with.

**Example 10.1** An experimenter carried out a trial to determine the effects of four treatments, A, B, C and D, to a certain crop. A field was divided into twelve plots, and three replicates of each of the four treatments were randomly assigned to the plots. The experimental layout and results were as follows (yield/plot):

<table>
<thead>
<tr>
<th>Treatment</th>
<th>A</th>
<th>D</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>33.63</td>
<td>39.62</td>
<td>38.18</td>
<td>41.46</td>
</tr>
<tr>
<td>B</td>
<td>38.02</td>
<td>35.83</td>
<td>35.99</td>
<td>36.58</td>
</tr>
<tr>
<td>C</td>
<td>42.92</td>
<td>37.80</td>
<td>40.43</td>
<td>37.89</td>
</tr>
<tr>
<td>D</td>
<td>33.63</td>
<td>39.62</td>
<td>38.18</td>
<td>41.46</td>
</tr>
</tbody>
</table>

A suitable linear model for this set of data might be as follows:

\[
\text{Observation} = (\text{Overall average}) + \left(\text{Difference from average due to treatment}\right) + (\text{Random error})
\]

(10.1)

We can easily see some possible treatment effects if we rearrange the treatment results, as shown in Table 10.1.

The last two columns of this table show some evidence that treatment effects exist. We therefore proceed to fit a model to the data. In the rearranged...
Table 10.1

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Results 1</th>
<th>Results 2</th>
<th>Results 3</th>
<th>Total Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>33.63</td>
<td>37.80</td>
<td>36.58</td>
<td>108.01</td>
</tr>
<tr>
<td>B</td>
<td>35.83</td>
<td>38.18</td>
<td>37.89</td>
<td>111.90</td>
</tr>
<tr>
<td>C</td>
<td>42.92</td>
<td>40.43</td>
<td>41.46</td>
<td>124.81</td>
</tr>
<tr>
<td>D</td>
<td>38.02</td>
<td>39.62</td>
<td>35.99</td>
<td>113.63</td>
</tr>
</tbody>
</table>

form, we can write Equation (10.1) as

\[ Y_{ij} = \mu + \gamma_i + \varepsilon_{ij} \quad i = 1, 2, 3, 4; \quad j = 1, 2, 3 \quad (10.2) \]

where \( i = 1, 2, 3, 4 \) runs over the treatments \( A, B, C, D \) and where \( \varepsilon_{ij} \) is the random-error term, which we often take to be independently and normally distributed with expectation zero and variance \( \sigma^2 \). However, there are too many parameters here since there are four treatment means, and five parameters which we can write

\[ \theta' = (\mu, \gamma_1, \gamma_2, \gamma_3, \gamma_4) \quad (10.3) \]

This overparameterization leads to certain problems which we discuss below. If we write our model in the standard form

\[ E(Y) = a\theta \]
\[ V(Y) = I\sigma^2 \quad (10.4) \]

then the matrix \( a \) has the form:

\[
\begin{bmatrix}
1 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

(10.5)

In the context used in this chapter we call this matrix the *design matrix*. (This term is not used in the context of the multiple-regression applications discussed in Chapters 5, 6 and 7.)