

**DigiSem**  
Wir beschaffen und  
digitalisieren



---

b  
UNIVERSITÄT  
BERN

Universitätsbibliothek Bern

Dieses Dokument steht Ihnen online zur Verfügung  
dank DigiSem, einer Dienstleistung der  
Universitätsbibliothek Bern.

Kontakt: Gabriela Scherrer  
Koordinatorin digitale Semesterapparate  
E-Mail [digisem@ub.unibe.ch](mailto:digisem@ub.unibe.ch), Telefon 031 631 93 26

José C. Pinheiro  
Douglas M. Bates

# Mixed-Effects Models in S and S-PLUS

With 172 Illustrations



Springer



José C. Pinheiro  
Department of Biostatistics  
Novartis Pharmaceuticals  
One Health Plaza  
East Hanover, NJ 07936-1080  
USA  
[jose.pinheiro@pharma.novartis.com](mailto:jose.pinheiro@pharma.novartis.com)

Douglas M. Bates  
Department of Statistics  
University of Wisconsin  
Madison, WI 53706-1685  
USA  
[bates@stat.wisc.edu](mailto:bates@stat.wisc.edu)

*Series Editors:*

J. Chambers  
Bell Labs, Lucent  
Technologies  
600 Mountain Ave.  
Murray Hill, NJ 07974  
USA

W. Eddy  
Department of Statistics  
Carnegie Mellon University  
Pittsburgh, PA 15213  
USA

W. Härdle  
Institut für Statistik und  
Ökonometrie  
Humboldt-Universität zu Berlin  
Spandauer Str. 1  
D-10178 Berlin  
Germany

S. Sheather  
Australian Graduate School  
of Management  
University of New South  
Wales  
Sydney NSW 2052  
Australia

L. Tierney  
School of Statistics  
University of Minnesota  
Vincent Hall  
Minneapolis, MN 55455  
USA

Library of Congress Cataloging-in-Publication Data  
Pinheiro, José C.

Mixed-effects models in S and S-PLUS / José C. Pinheiro, Douglas M. Bates  
p. cm. — (Statistics and computing)  
Includes bibliographical references and index.  
ISBN 0-387-98957-9 (alk. paper)  
I. Bates, Douglas M. II. Title. III. Series.  
QA76.73.S15P56 2000  
005.13'3—dc21

99-053566

Printed on acid-free paper.

© 2000 Springer Verlag New York, LLC

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag New York, LLC, 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Printed in the United States of America. (HAM)

9 8 7 6 5 SPIN 10995662

Springer Verlag is a part of *Springer Science+Business Media*

[springeronline.com](http://springeronline.com)

# 1

## Linear Mixed-Effects Models: Basic Concepts and Examples

Many common statistical models can be expressed as linear models that incorporate both *fixed effects*, which are parameters associated with an entire population or with certain repeatable levels of experimental factors, and *random effects*, which are associated with individual experimental units drawn at random from a population. A model with both fixed effects and random effects is called a *mixed-effects* model.

Mixed-effects models are primarily used to describe relationships between a response variable and some covariates in data that are grouped according to one or more classification factors. Examples of such *grouped data* include *longitudinal data*, *repeated measures data*, *multilevel data*, and *block designs*. By associating common random effects to observations sharing the same level of a classification factor, mixed-effects models flexibly represent the covariance structure induced by the grouping of the data.

In this chapter we present an overview of linear mixed-effects (LME) models, introducing their basic concepts through the analysis of several real-data examples, starting from simple models and gradually moving to more complex models. Although the S code to fit these models is shown, the purpose here is to present the motivation for using LME models to analyze grouped data and not to concentrate on the software for fitting and displaying the models. This chapter serves as an appetizer for the material covered in later chapters: the theoretical and computational methods for LME models described in Chapter 2 and the linear mixed-effects modeling facilities available in the `nlme` library, covered in detail in Chapter 4.

The examples described in this chapter also serve to illustrate the breadth of applications of linear mixed-effects models.

## 1.1 A Simple Example of Random Effects

The data shown in Figure 1.1 are from an experiment in nondestructive testing for longitudinal stress in railway rails cited in Devore (2000, Example 10.10, p. 427). Six rails were chosen at random and tested three times each by measuring the time it took for a certain type of ultrasonic wave to travel the length of the rail. The only experimental setting that changes between the observations is the rail. We say these observations are arranged in a *one-way classification* because they are classified according to a single characteristic—the rail on which the observation was made. These data are described in greater detail in Appendix A.26.

The quantities the engineers were interested in estimating from this experiment are the average travel time for a “typical” rail (the *expected travel time*), the variation in average travel times among rails (the *between-rail variability*), and the variation in the observed travel times for a single rail (the *within-rail variability*). We can see from Figure 1.1 that there is considerable variability in the mean travel time for the different rails. Overall the between-rail variability is much greater than the within-rail variability.

The data on the rails experiment are given in an object called `Rail` that is available with the `nlme` library. Giving the name `Rail` by itself to the S interpreter will result in the data being displayed.

```
> Rail
Grouped Data: travel ~ 1 | Rail
  Rail travel
  1   1    55
  2   1    53
  3   1    54
  . . .
  17  6    85
  18  6    83
```

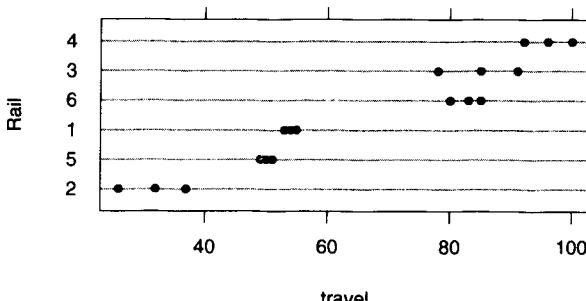


FIGURE 1.1. Travel time in nanoseconds for ultrasonic head-waves in a sample of six railroad rails. The times shown are the result of subtracting 36,100 nanoseconds from the original observation.

As would be expected, the structure of the data is quite simple—each row corresponds to one observation for which the rail and the travel time are recorded. The names of the variables in the data frame are `Rail` and `travel`. There is also a formula, `travel ~ 1 | Rail`, associated with the data. This formula is discussed in Chapter 3, where we describe structures available in the `nlme` library for representing grouped data.

Data from a one-way classification like the rails example can be analyzed either with a fixed-effects model or with a random-effects model. The distinction between the two models is according to whether we wish to make inferences about those particular levels of the classification factor that were used in the experiment or to make inferences about the population from which these levels were drawn. In the latter case the “levels” usually correspond to different subjects or different plots or different experimental units of some sort.

To illustrate the importance of accounting for the classification factor when modeling grouped data such as the rails example, we initially ignore the grouping structure of the data and assume the simple model

$$y_{ij} = \beta + \epsilon_{ij}, \quad i = 1, \dots, M, \quad j = 1, \dots, n_i, \quad (1.1)$$

where  $y_{ij}$  is the observed travel time for observation  $j$  on rail  $i$ ,  $\beta$  is the mean travel time across the population of rails being sampled, and the  $\epsilon_{ij}$  are independent  $\mathcal{N}(0, \sigma^2)$  error terms. The number of rails is  $M$  and the number of observations on rail  $i$  is  $n_i$ . In this case  $M = 6$  and  $n_1 = n_2 = \dots = n_6 = 3$ . The total number of observations is  $N = \sum_{i=1}^M n_i = 18$ .

The `lm` function is used to fit the single-mean model (1.1) in S. Its first argument is a formula describing the model and its second argument is a data frame containing the variables named in the model formula.

```
> fm1Rail.lm <- lm( travel ~ 1, data = Rail )
> fm1Rail.lm
Call:
lm(formula = travel ~ 1, data = Rail)

Coefficients:
(Intercept)
66.5

Degrees of freedom: 18 total; 17 residual
Residual standard error: 23.645
```

As is typical with S, we do not produce output directly from the fitting process. Instead we store the fitted model as an object called `fm1Rail.lm` then cause this object to be displayed. It contains the parameter estimates  $\hat{\beta} = 66.5$  and  $\hat{\sigma} = 23.645$ .

The boxplots of the residuals from the `fm1Rail.lm` fit by rail number, displayed in Figure 1.2, illustrate the fundamental problem with ignoring

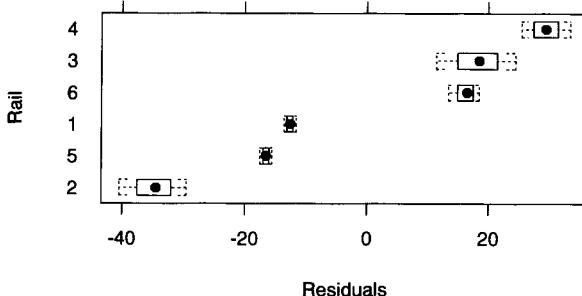


FIGURE 1.2. Boxplots of residuals by rail number for the `lm` fit of the single-mean model (1.1) to the data from the rail experiment.

the classification factor when modeling grouped data: the “group effects” are incorporated into the residuals (which, in this case, have identical signs for each rail), leading to an inflated estimate of the within-rail variability.

The “rail effects” indicated in Figure 1.2 may be incorporated into the model for the travel times by allowing the mean of each rail to be represented by a separate parameter. This *fixed-effects* model for the one-way classification is written

$$y_{ij} = \beta_i + \epsilon_{ij}, \quad i = 1, \dots, M, \quad j = 1, \dots, n_i, \quad (1.2)$$

where the  $\beta_i$  represents the mean travel time of rail  $i$  and, as in (1.1), the errors  $\epsilon_{ij}$  are assumed to be independently distributed as  $\mathcal{N}(0, \sigma^2)$ . We can again use `lm` to fit (1.2).

```
> fm2Rail.lm <- lm( travel ~ Rail - 1, data = Rail )
> fm2Rail.lm
Call:
lm(formula = travel ~ Rail - 1, data = Rail)
```

#### Coefficients:

Rail12	Rail15	Rail11	Rail16	Rail13	Rail14
31.667	50	54	82.667	84.667	96

Degrees of freedom: 18 total; 12 residual

Residual standard error: 4.0208

A `-1` is used in the model formula to prevent the default inclusion of an intercept term in the model. As expected, there is considerable variation in the estimated mean travel times per rail. The residual standard error obtained for the fixed-effects model (1.2),  $\hat{\sigma} = 4.0208$ , is about one-sixth of the corresponding estimate obtained for the single-mean model (1.1), indicating that the `fm2Rail.lm` model has successfully accounted for the rail effects. This is better illustrated by the boxplots of the `fm2Rail.lm` residuals

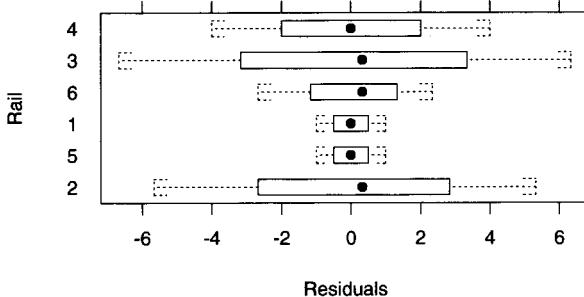


FIGURE 1.3. Boxplots of residuals by rail number for the `lm` fit of the fixed-effects model (1.2) to the data from the rail experiment.

by rail number, shown in Figure 1.3. The residuals are now centered around zero and have considerably smaller magnitudes than those in Figure 1.2.

Even though the fixed-effects model (1.2) accounts for the rail effects, it does not provide a useful representation of the rails data. Its basic problem is that it only models the specific sample of rails used in the experiment, while the main interest is in the population of rails from which the sample was drawn. In particular, `fm2Rail1.lm` does not provide an estimate of the between-rail variability, which is one of the central quantities of interest in the rails experiment. Another drawback of this fixed-effects model is that the number of parameters in the model increases linearly with the number of rails.

A *random-effects* model circumvents these problems by treating the rail effects as random variations around a population mean. The following re-parameterization of model (1.2) helps motivate the random-effects model for the rails data. We write

$$y_{ij} = \bar{\beta} + (\beta_i - \bar{\beta}) + \epsilon_{ij}, \quad (1.3)$$

where  $\bar{\beta} = \sum_{i=1}^6 \beta_i / 6$  represents the average travel time for the rails in the experiment. The random-effects model replaces  $\bar{\beta}$  by the mean travel time over the *population of rails* and replaces the deviations  $\beta_i - \bar{\beta}$  by random variables whose distribution is to be estimated.

A random-effects model for the one-way classification used in the rails experiment is written

$$y_{ij} = \beta + b_i + \epsilon_{ij}, \quad (1.4)$$

where  $\beta$  is the mean travel time across the population of rails being sampled,  $b_i$  is a random variable representing the deviation from the population mean of the mean travel time for the  $i$ th rail, and  $\epsilon_{ij}$  is a random variable representing the deviation in travel time for observation  $j$  on rail  $i$  from the mean travel time for rail  $i$ .

To complete the statistical model, we must specify the distribution of the random variables  $b_i, i = 1, \dots, M$  and  $\epsilon_{ij}, i = 1, \dots, M; j = 1, \dots, n_i$ . We begin by modeling both of these as independent, constant variance, normally distributed random variables with mean zero. The variances are denoted  $\sigma_b^2$  for the  $b_i$ , or “between-rail” variability, and  $\sigma^2$  for the  $\epsilon_{ij}$ , or “within-rail” variability. That is,

$$b_i \sim \mathcal{N}(0, \sigma_b^2), \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2). \quad (1.5)$$

This model may be modified if it does not seem appropriate. As described in Chapter 4, we encourage using graphical and numerical diagnostic tools to assess the validity of the model and to suggest ways in which it could be modified. To start, however, we will use this simple model.

This model with two sources of random variation,  $b_i$  and  $\epsilon_{ij}$ , is sometimes called a *hierarchical* model (Lindley and Smith, 1972; Bryk and Raudenbush, 1992) or a *multilevel* model (Goldstein, 1995). The  $b_i$  are called *random effects* because they are associated with the particular experimental units—rails in this case—that are selected at random from the population of interest. They are *effects* because they represent a deviation from an overall mean. That is, the “effect” of choosing rail  $i$  is to shift the mean travel time from  $\beta$  to  $\beta + b_i$ . Because observations made on the same rail share the same random effect  $b_i$ , they are correlated. The covariance between observations on the same rail is  $\sigma_b^2$  corresponding to a correlation of  $\sigma_b^2 / (\sigma_b^2 + \sigma^2)$ .

The parameters of the statistical model created by combining (1.4) and (1.5) are  $\beta$ ,  $\sigma_b^2$ , and  $\sigma^2$ . Note that the number of parameters will always be three, irrespective of the number of rails in the experiment. Although the random effects,  $b_i, i = 1, \dots, M$  may behave like parameters, formally they are just another level of random variation in the model so we do not “estimate” them as such. We will, however, form predictions  $\hat{b}_i$  of the values of these random variables, given the data we observed.

### 1.1.1 Fitting the Random-Effects Model With `lme`

The `lme` function from the `nlme` library for S can be used to fit linear mixed-effects models, using either *maximum likelihood* (ML) or *restricted maximum likelihood* (REML). These estimation methods for the parameters in LME models are described in detail in §2.2.

A typical call to `lme` is similar to a call to `lm`. As in `lm`, the first two arguments to `lme`, `fixed` and `data`, give the model for the expected response (the fixed-effects part of the model) and the object containing the data to which the model should be fit. The third argument, `random`, is a one-sided formula describing the random effects and the grouping structure for the model. Another important argument is `method`. Specifying `method = "ML"` produces maximum likelihood fits while `method = "REML"`, the default, produces restricted maximum likelihood fits.

Many variations in the specifications of linear mixed-effects models for `lme` are possible, as shown later in this and other chapters. Details of all the possible arguments and their forms are given in Appendix B.

We obtain the restricted maximum likelihood fit of the model given by (1.4) and (1.5) to the `Rail` data with

```
> fm1Rail.lme <- lme(travel ~ 1, data = Rail, random = ~ 1 | Rail)
```

The first argument indicates that the response is `travel` and that there is a single fixed effect, the intercept. The second argument indicates that the data will be found in the object named `Rail`. The third argument indicates that there is a single random effect for each group and that the grouping is given by the variable `Rail`. Note that there is a variable or column `Rail` within the data frame that is also named `Rail`. Because no estimation method is specified, the default, "REML", is used.

We can query the fitted `lme` object, `fm1Rail.lme`, using different accessor functions, also described in detail in Appendix B. One of the most useful of these is the `summary` function

```
> summary(fm1Rail.lme)
Linear mixed-effects model fit by REML
Data: Rail
  AIC    BIC   logLik
128.18 130.68 -61.089

Random effects:
Formula: ~ 1 | Rail
          (Intercept) Residual
StdDev:      24.805   4.0208

Fixed effects: travel ~ 1
                Value Std.Error DF t-value p-value
(Intercept) 66.5     10.171 12  6.5382 <.0001

Standardized Within-Group Residuals:
      Min        Q1        Med        Q3       Max
-1.6188 -0.28218  0.035693  0.21956  1.6144

Number of Observations: 18
Number of Groups: 6
```

We see that the REML estimates for the parameters have been calculated as

$$\hat{\beta} = 66.5, \quad \hat{\sigma}_b = 24.805, \quad \hat{\sigma} = 4.0208,$$

corresponding to a log-restricted-likelihood of  $-61.089$ . The estimated mean travel time  $\hat{\beta}$  is identical to the estimated intercept in the `fm1Rail.lm` fit.

and the estimated within-rail standard deviation  $\hat{\sigma}$  is identical to the residual standard error from `fm2Rail.lm`. This will not occur in general; it is a consequence of the `Rail` data being a *balanced* one-way classification that has the same number of observations on each rail. We also note that the estimated between-rail standard deviation  $\hat{\sigma}_b$  is similar to the residual standard error from the `fm1Rail.lm` fit.

The output of the `summary` function includes the values of the *Akaike Information Criterion (AIC)* (Sakamoto, Ishiguro and Kitagawa, 1986) and the *Bayesian Information Criterion (BIC)* (Schwarz, 1978), which is also sometimes called *Schwarz's Bayesian Criterion (SBC)*. These are model comparison criteria evaluated as

$$\begin{aligned} AIC &= -2 \log \text{Lik} + 2n_{\text{par}}, \\ BIC &= -2 \log \text{Lik} + n_{\text{par}} \log(N), \end{aligned}$$

where  $n_{\text{par}}$  denotes the number of parameters in the model and  $N$  the total number of observations used to fit the model. Under these definitions, “smaller is better.” That is, if we are using AIC to compare two or more models for the same data, we prefer the model with the lowest AIC. Similarly, when using BIC we prefer the model with the lowest BIC.

To examine the maximum likelihood estimates we would call `lme` with the same arguments as for `fm1Rail.lme` except for `method = "ML"`. A convenient way of fitting such alternative models is to use the `update` function where we only need to specify the arguments that are different from those in the earlier fit.

```
> fm1Rail.lmeML <- update( fm1Rail.lme, method = "ML" )
> summary( fm1Rail.lmeML )
Linear mixed-effects model fit by maximum likelihood
Data: Rail
      AIC      BIC   logLik
 134.56  137.23 -64.28

Random effects:
Formula: ~ 1 | Rail
          (Intercept) Residual
StdDev:     22.624   4.0208

Fixed effects: travel ~ 1
              Value Std.Error DF t-value p-value
(Intercept) 66.5     9.554 12  6.9604 <.0001

Standardized Within-Group Residuals:
      Min        Q1        Med        Q3       Max
-1.611 -0.28887  0.034542  0.21373  1.6222

Number of Observations: 18
Number of Groups: 6
```

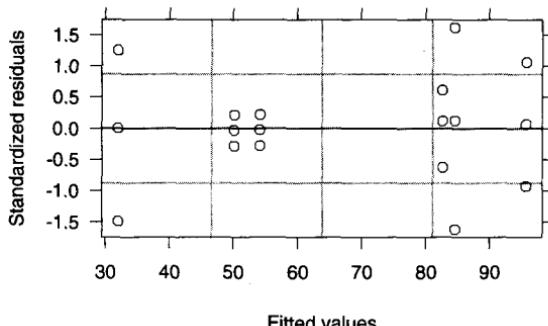


FIGURE 1.4. Standardized residuals versus the fitted values for the REML fit of a random-effects model to the data from the rail experiment.

Notice that the ML estimate of  $\sigma$  is 4.0208, the same as the REML estimate. Equality of the ML and REML estimates of  $\sigma$  occurs for this simple model, but will not occur in general. The ML estimate of  $\sigma_b$ , 22.624, is smaller than the REML estimate, 24.805. Finally the ML estimate of  $\beta$ , 66.5, is the same as the REML estimate. Again, exact equality of the ML and REML estimates of the fixed effects need not occur in more complex models, but it is commonplace for them to be nearly identical.

### 1.1.2 Assessing the Fitted Model

The fitted model can, and should, be examined using graphical and numerical summaries. One graphical summary that should be examined routinely is a plot of the residuals versus the fitted responses from the model. This plot is used to assess the assumption of constant variance of the  $\epsilon_{ij}$ . Because this plot is a common diagnostic, it is the default `plot` method for a fitted `lme` model. That is, it is produced by the simple call

```
> plot( fm1Rail.lme ) # produces Figure 1.4
```

The standardized residuals, shown on the vertical axis in Figure 1.4, are the raw residuals,  $e_{ij} = y_{ij} - \hat{\beta} - \hat{b}_i$ , divided by the estimated standard deviation,  $\hat{\sigma}$ , of the  $\epsilon_{ij}$ .

In this plot we are looking for a systematic increase (or, less commonly, a systematic decrease) in the variance of the  $\epsilon_{ij}$  as the level of the response increases. If this is present, the residuals on the right-hand side of the plot will have a greater vertical spread than those on the left, forming a horizontal “wedge-shaped” pattern. Such a pattern is not evident in Figure 1.4.

With more complicated models there are other diagnostic plots that we may want to examine, as discussed in Chapter 4.

We should also examine numerical summaries of the model. A basic summary is a set of confidence intervals on the parameters,  $\beta$ ,  $\sigma$  and  $\sigma_b$ , as produced by the `intervals` function.

```
> intervals( fm1Rail.lme )
Approximate 95% confidence intervals

Fixed effects:
      lower est. upper
(Intercept) 44.339 66.5 88.661

Random Effects:
  Level: Rail
      lower est. upper
sd((Intercept)) 13.274 24.805 46.354

Within-group standard error:
      lower est. upper
2.695 4.0208 5.9988
```

We can see that there is considerable imprecision in the estimates of all three of these parameters.

Another numerical summary, used to assess the significance of terms in the fixed-effects part of the model, is produced by the `anova` function

```
> anova( fm1Rail.lme )
      numDF denDF F-value p-value
(Intercept)     1     12   42.748 <.0001
```

In this case, the fixed-effects model is so simple that the analysis of variance is trivial. The hypothesis being tested here is  $\beta = 0$ . The *p*-value, which is that probability of observing data as unusual as these or even more so when  $\beta$  actually is 0, is so small as to rule out this possibility. Regardless of the *p*-value, the hypothesis  $\beta = 0$  is of no practical interest here because the data have been shifted by subtracting 36,100 nanoseconds from each measurement.

## 1.2 A Randomized Block Design

In the railway rails example of the last section, the observations were classified according to one characteristic only—the rail on which the observation was made. In other experiments we may have more than one classification factor for each observation. A *randomized block design* is a type of experiment in which there are two classification factors: an *experimental* factor for which we use fixed effects and a *blocking* factor for which we use random effects.

The data shown in Figure 1.5 and available as the object `ergoStool` in

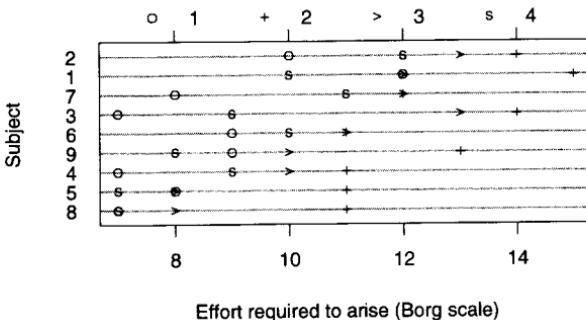


FIGURE 1.5. Effort required (Borg scale) to arise from a stool for nine different subjects each using four different types of stools. Different symbols, shown in the key at the top of the plot, are used for the different types of stools.

the `nlme` library are from an ergometrics experiment that has a randomized block design. The experimenters recorded the effort required by each of nine different subjects to arise from each of four types of stools. We want to compare these four particular types of stools so we use fixed effects for the `Type` factor. The nine different subjects represent a sample from the population about which we wish to make inferences so we use random effects to model the `Subject` factor.

From Figure 1.5 it appears that there are systematic differences between stool types on this measurement. For example, the `T2` stool type required the greatest effort from each subject while the `T1` stool type was consistently one of the low effort types. The subjects also exhibited variability in their scoring of the effort, but we would expect this. We say that `Subject` to be a *blocking factor* because it represents a known source of variability in the experiment. `Type` is said to be an *experimental factor* because the purpose of the experiment is to determine if there are systematic differences in the level of effort to arise from the different types of stools.

We can visually compare the magnitude of the effects of the `Type` and `Subject` factors using a “design plot”

```
> plot.design( ergoStool ) # produces Figure 1.6
```

This plot is produced by averaging the responses at each level of each factor and plotting these averages. We see that the variability associated with the `Type` factor is comparable to the variability associated with the `Subject` factor. We also see that the average effort according to stool type is in the order  $T1 \leq T4 \leq T3 \leq T2$ .

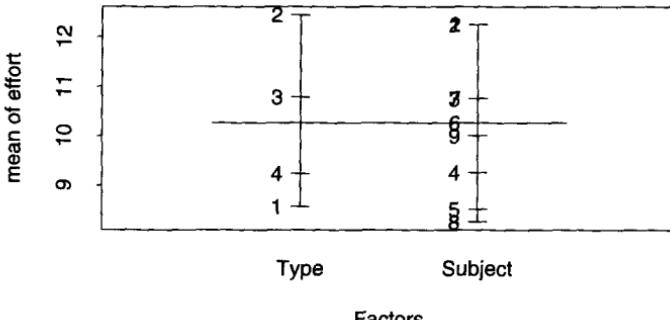


FIGURE 1.6. Design plot for the data in the stool ergometric experiment. The mean of the response (`effort`) is plotted for each level of each of the factors `Type` and `Subject`.

### 1.2.1 Choosing Contrasts for Fixed-Effects Terms

A model with fixed effects  $\beta_j$  for the `Type` factor and random effects  $b_i$  for the `Subject` factor could be written

$$\begin{aligned} y_{ij} &= \beta_j + b_i + \epsilon_{ij}, \quad i = 1, \dots, 9, \quad j = 1, \dots, 4, \\ b_i &\sim \mathcal{N}(0, \sigma_b^2), \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2), \end{aligned} \tag{1.6}$$

or, equivalently,

$$\begin{aligned} \mathbf{y}_i &= \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i b_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, 9, \\ b_i &\sim \mathcal{N}(0, \sigma_b^2), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}), \end{aligned}$$

where, for  $i = 1, \dots, 9$ ,

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ y_{i4} \end{bmatrix}, \quad \mathbf{X}_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{Z}_i = \mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{\epsilon}_i = \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \epsilon_{i3} \\ \epsilon_{i4} \end{bmatrix}.$$

This form of fixed-effects matrix  $\mathbf{X}_i$  is sometimes called the *cell means* form because the  $j$ th component of  $\boldsymbol{\beta}$  represents what would be the mean effort to arise from the  $j$ th type of stool if the whole population were tested.

These  $\beta_j$  have a simple interpretation, but are not convenient to use when assessing differences between stool types. To make it easier to assess these differences we use an alternative form of the  $\mathbf{X}_i$  matrices with one column representing some “overall mean” or reference level and three columns representing changes between the types of stools. The three columns representing the changes are called the *contrasts*. There are several different choices available for these contrasts (Venables and Ripley, 1999, §6.2). In S-PLUS, the default choice for unordered factors, such as the `Type` factor, is the *Helmert* contrasts

```
> contrasts( ergoStool$Type )
 [,1] [,2] [,3]
 1    -1   -1   -1
 2     1   -1   -1
 3     0    2   -1
 4     0    0    3
```

(In R the default contrasts for an unordered factor are the “treatment” contrasts, which are described below.)

The  $\mathbf{X}_i$  matrices for a given set of contrasts can be displayed with the `model.matrix` function. To save space we show the  $\mathbf{X}_1$  matrix only.

```
> ergoStool1 <- ergoStool[ ergoStool$Subject == "1", ]
> model.matrix( effort ~ Type, ergoStool1 ) # X matrix for Subject 1
  (Intercept) Type1 Type2 Type3
 1           1   -1   -1   -1
 2           1    1   -1   -1
 3           1    0    2   -1
 4           1    0    0    3
```

Using the Helmert contrasts shown above, the components of  $\beta$  represent:

- $\beta_1$ —Mean level of effort for the four stool types.
- $\beta_2$ —Half the difference between T2 and T1.
- $\beta_3$ —One third of the difference between T3 and the average of T1 and T2.
- $\beta_4$ —One fourth of the difference between T4 and the average of T1, T2, and T3.

Fitting the model in this form with `lme` produces

```
> fm1Stool <-
+   lme(effort ~ Type, data = ergoStool, random = ~ 1 | Subject)
> summary( fm1Stool )
...
Random effects:
 Formula: ~ 1 | Subject
          (Intercept) Residual
 StdDev:      1.3325   1.1003

Fixed effects: effort ~ Type
                Value Std.Error DF t-value p-value
(Intercept) 10.250    0.48052 24  21.331 <.0001
      Type1   1.944    0.25934 24   7.498 <.0001
      Type2   0.093    0.14973 24   0.618  0.5421
      Type3  -0.343    0.10588 24  -3.236  0.0035

Correlation:
```

```
(Intr) T1 T2
T1 0
T2 0 0
T3 0 0 0
```

## Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-1.802	-0.64317	0.057831	0.701	1.6314

Number of Observations: 36

Number of Groups: 9

By convention, the coefficient corresponding to the first column in the  $X_i$ , which is the column of 1's, is called the *intercept*. The name originated with models like the analysis of covariance model of §1.4 where a straight-line model for each group is written in terms of its slope and its intercept with the  $y$ -axis. In those cases, this parameter is the  $y$ -intercept. For the model considered here, the parameter labelled `(Intercept)` is the estimate of mean effort for all four types of stools across the population. The other three parameters, labelled `Type1`, `Type2`, and `Type3`, are described above. Their individual interpretations are not as important as the collective variability among the stool types they represent. The significance of this variability, and hence the overall significance of the `Type` term, is assessed with the `anova` function.

```
> anova( fm1Stool )
      numDF denDF F-value p-value
(Intercept)     1     24  455.01 <.0001
      Type       3     24   22.36 <.0001
```

On some occasions we may want to switch to other contrasts that provide more meaningful parameter estimates for the experiment. For example, if stool type `T1` was a “standard” stool and we wished to compare the other types to this standard type, we could use the contrasts called the *treatment* contrasts. These contrasts represent the change from the first level of the factor to each of the other levels.

One way to cause the treatment contrasts to be used is to reset the `contrasts` option. Its value should be a vector of two character strings. The first string is the name of the function to use for factors, such as `Type`, and the second is the function to use for ordered factors, which are described in §1.6.

```
> options( contrasts = c( factor = "contr.treatment",
+                           ordered = "contr.poly" ) )
> contrasts( ergoStool$Type )
 2 3 4
1 0 0 0
2 1 0 0
```

```

3 0 1 0
4 0 0 1
> fm2Stool <-
+  lme(effort ~ Type, data = ergoStool, random = ~ 1 | Subject)
> summary( fm2Stool )
Linear mixed-effects model fit by REML
Data: ergoStool
    AIC      BIC   logLik
 133.13  141.93 -60.565

Random effects:
Formula: ~ 1 | Subject
          (Intercept) Residual
StdDev:     1.3325   1.1003

Fixed effects: effort ~ Type
            Value Std.Error DF t-value p-value
(Intercept) 8.5556   0.57601 24 14.853 <.0001
Type2       3.8889   0.51868 24   7.498 <.0001
Type3       2.2222   0.51868 24   4.284  0.0003
Type4       0.6667   0.51868 24   1.285  0.2110

Correlation:
  (Intr) Type2 Type3
Type2 -0.45
Type3 -0.45   0.50
Type4 -0.45   0.50   0.50
...
> anova( fm2Stool )
            numDF denDF F-value p-value
(Intercept)      1     24  455.01 <.0001
      Type       3     24   22.36 <.0001

```

Although the individual parameter estimates for the `Type` factor are different between the two fits, the `anova` results are the same. The difference in the parameter estimates simply reflects the fact that different contrasts are being estimated. The similarity of the `anova` results indicates that the overall variability attributed to the `Type` factor does not change. In each case, the row labelled `Type` in the analysis of variance table represents a test of the hypothesis

$$H_0 : \beta_2 = \beta_3 = \beta_4 = 0,$$

which is equivalent to reducing model (1.6) to

$$\mathbf{y}_i = \mathbf{1}\boldsymbol{\beta} + \mathbf{Z}_i b_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, 9, \quad b_i \sim \mathcal{N}(0, \sigma_b^2), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}).$$

This reduced model is invariant under the change in contrasts.

There is a more subtle effect of changing from one form of  $\mathbf{X}_j$  matrix to another—the value of the REML criterion changes. As described in §2.2.5,

when the model parameters are kept at a fixed value, a change in the  $\mathbf{X}_j$  matrices results in change in the value of the restricted likelihood function. Even though we converge to the same variance component estimates  $\hat{\sigma} = 1.1003$  and  $\hat{\sigma}_b = 1.3325$ , the value of the estimation criterion itself changes. Because the AIC and BIC criteria are based on the REML criterion, they will also change.

As a consequence, when using REML estimation we can only use likelihood ratio tests or comparisons of AIC or BIC for models with the same fixed-effects structure and the same contrasts for any factors used in the fixed-effects structure.

We can fit the “cell means” parameterization of the model if we add the term `-1` to the formula for the fixed effects. This causes the column of 1’s to be removed from the model matrices  $\mathbf{X}_j$ ,

```
> model.matrix( effort ~ Type - 1, ergoStool1 )
   Type1 Type2 Type3 Type4
1      1     0     0     0
2      0     1     0     0
3      0     0     1     0
4      0     0     0     1
```

and the fitted model is now expressed in terms of the mean effort for each stool type over the population

```
> fm3Stool <-
+ lme(effort ~ Type - 1, data = ergoStool, random = ~ 1 | Subject)
> summary(fm3Stool)
Linear mixed-effects model fit by REML
Data: ergoStool
      AIC      BIC      logLik
133.13 141.93 -60.565

Random effects:
Formula: ~ 1 | Subject
          (Intercept) Residual
StdDev:      1.3325  1.1003

Fixed effects: effort ~ Type - 1
                Value Std.Error DF t-value p-value
Type1    8.556   0.57601 24  14.853  <.0001
Type2   12.444   0.57601 24  21.604  <.0001
Type3   10.778   0.57601 24  18.711  <.0001
Type4    9.222   0.57601 24  16.010  <.0001

Correlation:
           Type1 Type2 Type3
Type2  0.595
Type3  0.595  0.595
Type4  0.595  0.595  0.595
```

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-1.802	-0.64317	0.057831	0.701	1.6314

Number of Observations: 36

Number of Groups: 9

This change in the fixed-effects structure does change the `anova` results.

```
> anova( fm3Stool )
  numDF denDF F-value p-value
Type      4     24  130.52 <.0001
```

The hypothesis being tested by `anova` for this model is

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0,$$

which is equivalent to reducing model (1.6) to

$$\mathbf{y}_i = \mathbf{Z}_i b_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, 9, \quad b_i \sim \mathcal{N}(0, \sigma_b^2), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}).$$

That is, the hypothesis  $H_0$  completely eliminates the fixed-effects parameters from the model so the mean response across the population would be zero. This hypothesis is not meaningful in the context of this experiment.

To reiterate, some general principles to keep in mind regarding fixed-effects terms for factors are:

- The overall effect of the factor should be assessed with `anova`, not by examining the `t-value`'s or `p-value`'s associated with the fixed-effects parameters. The `anova` output does not depend on the choice of contrasts as long as the intercept term is retained in the model.
- Interpretation of the parameter estimates for a fixed-effects term depends on the contrasts being used.
- For REML estimation, likelihood-ratio tests or comparisons of AIC or BIC require the same fixed-effects structure and the same choice of contrasts in all models.
- The “cell means” parameters can be estimated by adding `-1` to a model formula but this will usually make the results of `anova` meaningless.

### 1.2.2 Examining the Model

As in the rail example, we should examine the fitted model both graphically and numerically. The `intervals` function provides an indication of the precision of the estimates of the variance components.

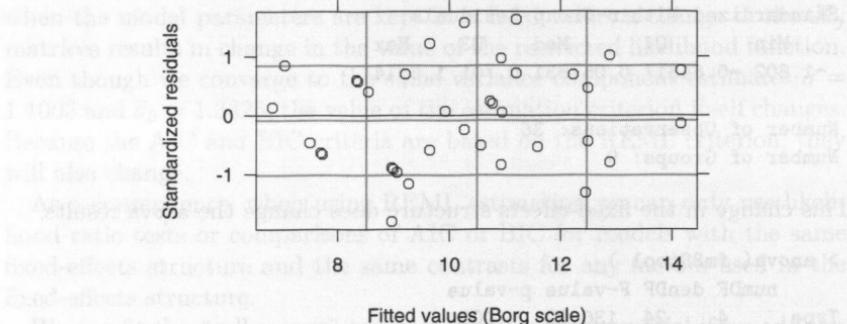


FIGURE 1.7. Standardized residuals versus the fitted values for the REML fit of a random-effects model to the data in the ergometric experiment on types of stools.

```
> intervals( fm1Stool )
Approximate 95% confidence intervals

Fixed effects:
      lower       est.       upper
(Intercept) 9.25825 10.250000 11.24175
  Type1    1.40919  1.944444  2.47970
  Type2   -0.21644  0.092593  0.40162
  Type3   -0.56111 -0.342593 -0.12408

Random Effects:
  Level: Subject
          lower     est.     upper
sd((Intercept)) 0.74923 1.3325 2.3697

Within-group standard error:
      lower     est.     upper
  0.82894 1.1003 1.4605
```

We see that  $\sigma$  is estimated relatively precisely, whereas  $\sigma_b$  can vary by a factor of about 3, which is a factor of 9 if we express the estimates as variances.

The plot of the standardized residuals versus the fitted values, shown in Figure 1.7, does not indicate a violation of the assumption of constant variance for the  $\epsilon_{ij}$  terms.

Figure 1.7 shows the overall behavior of the residuals relative to the fitted values. It may be more informative to examine this behavior according to the `Subject` factor or according to the `Type` factor, which we can do by providing an explicit formula to the `plot` method for the fitted `lme` model. The formula can use functions such as `resid` or `fitted` applied to the fitted model. As a shortcut, a `"."` appearing in the formula is interpreted as the

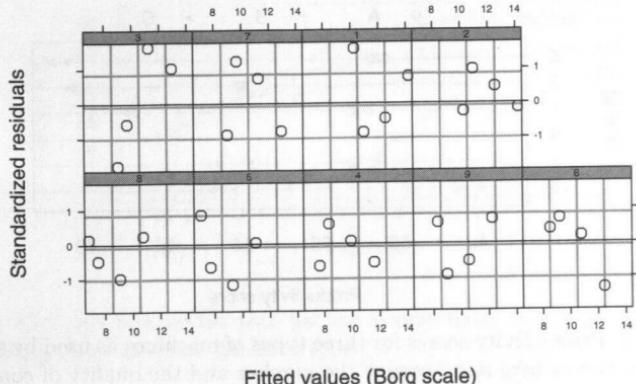


FIGURE 1.8. Standardized residuals versus the fitted values by `Subject` for the REML fit of a random-effects model to the data in the ergometric experiment on types of stools.

fitted model object itself. Thus, to plot the standardized, or “Pearson,” residuals versus the fitted values by `Subject`, we use

```
> plot( fm1Stool, # produces Figure 1.8
+       form = resid(., type = "p") ~ fitted(.) | Subject,
+       abline = 0 )
```

The argument `abline = 0` adds a horizontal reference line at  $y = 0$  to each panel. In its more general form, the value of the `abline` argument should be a numeric vector of length two giving the intercept and the slope of the line to be drawn on each panel. Diagnostic plots for assessing the adequacy of lme fits are discussed in detail in §4.3.

### 1.3 Mixed-Effects Models for Replicated, Blocked Designs

In the ergometric experiment on the types of stools, each subject tried each type of stool once. We say this design is *unreplicated* because only one observation is available at each combination of experimental conditions. In other experiments like this it may be feasible to take replicate measurements. For example, the `Machines` data, described in Milliken and Johnson (1992, Chapter 23) and shown in Figure 1.9, gives the productivity score for each of six randomly chosen workers tested on each of three different machine types. Each worker used each machine three times so we have three *replicates* at each set of conditions.

In Figure 1.9 we can see that there are strong indications of differences between machines and also some indications of differences between workers.

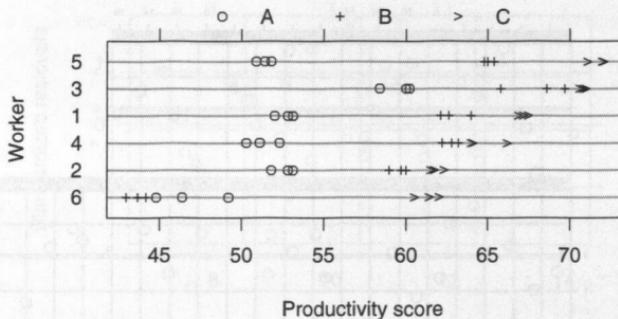


FIGURE 1.9. Productivity scores for three types of machines as used by six different workers. Scores take into account the number and the quality of components produced.

We note that there is very little variability in the productivity score for the same worker using the same machine.

As we did for the experiment on the types of stools, we will model the subject or **Worker** factor with random effects and the type or **Machine** factor with fixed effects. The replications in this experiment will allow us to assess the presence of *interactions* between worker and machine. That is, we can address the question of whether the effect of changing from one type of machine to another is different for different workers.

The comparative dotplot in Figure 1.9 allows us to see patterns across the workers and to see differences between machines within each worker. However, the possibility of interactions is not easy to assess in this plot. An alternative plot, called an *interaction plot*, shows the potential interactions more clearly. It is produced by averaging the scores for each worker on each machine, plotting these averages versus the machine type, and joining the points for each worker. The function `interaction.plot` in S creates such a plot. It is most easily called after `attach`'ing the data frame with the data so the variables in the data frame can be accessed by name.

```
> attach( Machines ) # make variables in Machines available by name
> interaction.plot( Machine, Worker, score, las = 1 ) # Figure 1.10
> detach() # undo the effect of 'attach( Machines )'
```

(The optional argument `las = 1` to `interaction.plot` alters the “label style” on the vertical axis to a more appealing form.)

If there were no interactions between machines and workers, the lines in the interaction plot would be approximately parallel. The lines in Figure 1.10 do not seem to be parallel, especially relative to the variability within the replicates that we can see in Figure 1.9. Worker 6 has an unusual pattern compared to the other workers.

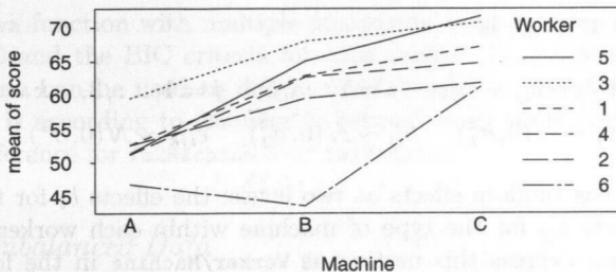


FIGURE 1.10. An interaction plot for the productivity scores for six different workers using three different machine types.

### 1.3.1 Fitting Random Interaction Terms

A model without interactions has the same form as the model for the ergometric experiment.

$$y_{ijk} = \beta_j + b_i + \epsilon_{ijk}, \quad i = 1, \dots, 6, \quad j = 1, \dots, 3, \quad k = 1, \dots, 3, \\ b_i \sim \mathcal{N}(0, \sigma_b^2), \quad \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2). \quad (1.7)$$

There is a fixed effect for each type of machine and a random effect for each worker. As before, the fixed effects for the machines will be re-coded as an intercept and a set of contrasts when we fit this model as

```
> fm1Machine <-  
+   lme(score ~ Machine, data = Machines, random = ~ 1 | Worker)  
> fm1Machine  
Linear mixed-effects model fit by REML  
Data: Machines  
Log-restricted-likelihood: -145.23  
Fixed: score ~ Machine  
(Intercept) Machine1 Machine2  
      59.65     3.9833    3.3111  
  
Random effects:  
Formula: ~ 1 | Worker  
          (Intercept) Residual  
StdDev:      5.1466    3.1616  
  
Number of Observations: 54  
Number of Groups: 6
```

Because the workers represent a random sample from the population of interest, any interaction terms modeling differences between workers in changing from one machine to another will also be expressed as random effects. The model incorporating the random interaction terms,  $b_{ij}, i =$

$1, \dots, 6$ ,  $j = 1, \dots, 3$ , is

$$y_{ijk} = \beta_j + b_i + b_{ij} + \epsilon_{ijk}, \quad i = 1, \dots, 6, \quad j = 1, \dots, 3, \quad k = 1, \dots, 3,$$

$$b_i \sim \mathcal{N}(0, \sigma_1^2), \quad b_{ij} \sim \mathcal{N}(0, \sigma_2^2), \quad \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2).$$

This model has random effects at two levels: the effects  $b_i$  for the worker and the effects  $b_{ij}$  for the type of machine within each worker. In a call to `lme` we can express this nesting as `Worker/Machine` in the formula for the random effects. This expression is read as “`Worker` and ‘`Machine` within `Worker`’”. We can update the previous model with a new specification for the random effects.

```
> fm2Machine <- update( fm1Machine, random = ~ 1 | Worker/Machine )
> fm2Machine
Linear mixed-effects model fit by REML
Data: Machines
Log-restricted-likelihood: -109.64
Fixed: score ~ Machine
(Intercept) Machine1 Machine2
      59.65     3.9833    3.3111

Random effects:
Formula: ~ 1 | Worker
          (Intercept)
StdDev:        4.7814

Formula: ~ 1 | Machine %in% Worker
          (Intercept) Residual
StdDev:        3.7294   0.96158

Number of Observations: 54
Number of Groups:
  Worker Machine %in% Worker
       6           18
```

This model has produced a value of the REML criterion of  $-109.64$ , which is considerably greater than that of `fm1Machine`,  $-145.23$ . The `anova` function, when given two or more arguments representing fitted models, produces likelihood ratio tests comparing the models.

```
> anova( fm1Machine, fm2Machine )
Model df      AIC      BIC  logLik  Test L.Ratio p-value
fm1Machine     1  5 300.46 310.12 -145.23
fm2Machine     2  6 231.27 242.86 -109.64 1 vs 2   71.191 <.0001
```

The likelihood ratio statistic comparing the more general model (`fm2Machine`) to the more specific model (`fm1Machine`) is huge and the *p*-value for the test is essentially zero, so we prefer `fm2Machine`.

The `anova` function with multiple arguments also reproduces the values of the AIC and the BIC criteria for each model. As described in §1.1.1, these criteria can be used to decide which model to prefer. Because the preference is according to “smaller is better,” both these criteria show a strong preference for `fm2Machine` over `fm1Machine`.

### 1.3.2 Unbalanced Data

The `Machines` data are balanced in that every `Worker` is tested on every `Machine` exactly three times. Milliken and Johnson (1992) analyze this example both as balanced data and as unbalanced data. To obtain the unbalanced data, they randomly deleted ten observations, as indicated in their Table 23.1. They observe that the software they used to estimate the random effects components (SAS PROC VARCOMP) did not produce sensible maximum likelihood estimates (although the current version of this software does). The `lme` function does produce sensible maximum likelihood estimates or restricted maximum likelihood estimates from the unbalanced data.

```
> ## delete selected rows from the Machines data
> MachinesUnbal <- Machines[ -c(2,3,6,8,9,12,19,20,27,33), ]
> ## check that the result is indeed unbalanced
> table(MachinesUnbal$Machine, MachinesUnbal$Worker)
  6 2 4 1 3 5
A 3 2 2 1 1 3
B 3 3 3 1 2 2
C 3 3 3 3 3 3
> fm1MachinesU <- lme(score ~ Machine, data = MachinesUnbal,
+   random = ~ 1 | Worker/Machine )
> fm1MachinesU
Linear mixed-effects model fit by REML
Data: MachinesUnbal
Log-restricted-likelihood: -92.728
Fixed: score ~ Machine
(Intercept) Machine1 Machine2
      59.648     3.9812    3.3123

Random effects:
Formula: ~ 1 | Worker
          (Intercept)
StdDev:        4.7387

Formula: ~ 1 | Machine %in% Worker
          (Intercept) Residual
StdDev:        3.7728    0.9332

Number of Observations: 44
```

```
Number of Groups: 6
Worker Machine %in% Worker: 18
> intervals( fm1MachinesU )
Approximate 95% confidence intervals
```

#### Fixed effects:

	lower	est.	upper
(Intercept)	55.2598	59.6476	64.0353
Machine1	1.5139	3.9812	6.4485
Machine2	1.8940	3.3123	4.7307

#### Random Effects:

Level: Worker	lower	est.	upper
sd((Intercept))	2.2162	4.7388	10.132

Level: Machine	lower	est.	upper
sd((Intercept))	2.4091	3.7728	5.9084

#### Within-group standard error:

	lower	est.	upper
	0.71202	0.9332	1.2231

The estimates of the standard deviations and the confidence intervals on these parameters look reasonable when compared to those from the full data set. The techniques used in `lme` for parameter estimation do not depend on the data being balanced.

However, for either balanced or unbalanced data we must have sufficient information in the data to be able to estimate the variance components and the fixed-effects parameters. We can fit a model with random interaction effects to the `Machines` data because there are replications. If we tried to fit a nested model to unreplicated data, such as the `ergoStool` data, it may appear that we are successful until we examine the intervals on the variance components.

```
> fm4Stool <- lme( effort ~ Type, ergoStool, ~ 1 | Subject/Type )
> intervals( fm4Stool )
Approximate 95% confidence intervals
```

#### Fixed effects:

	lower	est.	upper
(Intercept)	9.25825	10.250000	11.24175
Type1	1.40919	1.944444	2.47970
Type2	-0.21644	0.092593	0.40162
Type3	-0.56111	-0.342593	-0.12408

**Random Effects:**

Level: Subject

	lower	est.	upper
sd((Intercept))	0.74952	1.3325	2.3688

Level: Type

	lower	est.	upper
sd((Intercept))	0.05386	0.99958	18.551

Within-group standard error:

	lower	est.	upper
4.3603e-07	0.45988	485050	

Apparently the standard deviation  $\sigma$  could vary over twelve orders of magnitude!

If we write this model for these data, taking into account that each subject only tries each type of stool once, we would have

$$y_{ij} = \beta_i + b_j + b_{ij} + \epsilon_{ij}, \quad i = 1, \dots, 3, \quad j = 1, \dots, 6, \\ b_j \sim \mathcal{N}(0, \sigma_1^2), \quad b_{ij} \sim \mathcal{N}(0, \sigma_2^2), \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2).$$

We can see that the  $b_{ij}$  are totally confounded with the  $\epsilon_{ij}$  so we cannot estimate separate standard deviations for these two random terms. In fact, the estimates reported for  $\sigma$  and  $\sigma_2$  in this model give a combined variance that corresponds to  $\hat{\sigma}^2$  from fm1Stool.

```
> (fm1Stool$sigma)^2
[1] 1.2106
> (fm4Stool$sigma)^2 + 0.99958^2
[1] 1.2107
```

The lesson here is that it is always a good idea to check the confidence intervals on the variance components after fitting a model. Having abnormally wide intervals usually indicates problems with the model definition. In particular, a model with nested interaction terms can only be fit when there are replications available in the data.

### 1.3.3 More General Models for the Random Interaction Effects

In the model (1.3.1), the random interaction terms all have the same variance  $\sigma_2^2$ . Furthermore, these random interactions are assumed to be independent of one another, even within the same subject. A more general model could treat the random interactions for each subject as a vector and allow the variance-covariance matrix for that vector to be estimated from the set of all positive-definite matrices.

To express this model we return to the matrix/vector representation used in (1.2.1). We define  $\mathbf{y}_i$  to be the entire response vector for the  $i$ th subject,

$\beta$  to be the three-dimensional vector of fixed-effects parameters for the population,  $b_i$  to be the three-dimensional vector of random effects for the  $i$ th subject,  $X_i$  to be the  $9 \times 3$  fixed-effects design matrix for subject  $i$ , and  $Z_i$  to be the  $9 \times 3$  random-effects design matrix for subject  $i$ . The general form of the model is then

$$\begin{aligned} \mathbf{y}_i &= \mathbf{X}_i\beta + \mathbf{Z}_i b_i + \epsilon_i, \quad i = 1, \dots, 6, \\ b_i &\sim \mathcal{N}(\mathbf{0}, \Psi), \quad \epsilon_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}), \end{aligned}$$

where  $\Psi$  is a positive-definite, symmetric  $3 \times 3$  matrix.

To be more specific we must define the matrices  $X_i$  and  $Z_i$  or, equivalently, define the formulae that generate these matrices as model matrices. As discussed in §1.2.1, in the fixed-effects we generally use a formula that creates  $X_i$  with a single column of 1's and two columns of contrasts. We could do the same for the  $Z_i$  but, because the random effects are assumed to have an expected value of  $\mathbf{0}$  anyway, it is often more informative to use a formula such as `~ Machine - 1` that removes the intercept column.

Sample model matrices, evaluated on the Worker1's data only, are

```
> Machine1 <- Machines[ Machines$Worker == "1", ]
> model.matrix( score ~ Machine, Machine1 )      # fixed-effects X_i
   (Intercept) Machine1 Machine2
 1           1       -1      -1
 2           1       -1      -1
 3           1       -1      -1
 19          1        1      -1
 20          1        1     -1
 21          1        1     -1
 37          1        0      2
 38          1        0      2
 39          1        0      2
> model.matrix( ~ Machine - 1, Machine1 )      # random-effects Z_i
   MachineA MachineB MachineC
 1           1       0      0
 2           1       0      0
 3           1       0      0
 19          0       1      0
 20          0       1      0
 21          0       1      0
 37          0       0      1
 38          0       0      1
 39          0       0      1
```

The fitted model using this formulation is

```
> fm3Machine <- update( fm1Machine, random = ~Machine - 1 |Worker)
> summary( fm3Machine )
Linear mixed-effects model fit by REML
```

```

Data: Machines
      AIC    BIC   logLik
231.89 251.21 -105.95

Random effects:
Formula: ~ Machine - 1 | Worker
Structure: General positive-definite
          StdDev   Corr
MachineA 4.07928 MachnA MachnB
MachineB 8.62529 0.803
MachineC 4.38948 0.623 0.771
Residual 0.96158

Fixed effects: score ~ Machine
      Value Std.Error DF t-value p-value
(Intercept) 59.650    2.1447 46  27.813 <.0001
Machine1     3.983    1.2104 46   3.291  0.0019
Machine2     3.311    0.5491 46   6.030 <.0001
Correlation:
      (Intr) Machn1
Machine1  0.811
Machine2 -0.540 -0.453

Standardized Within-Group Residuals:
      Min        Q1        Med        Q3        Max
-2.3935 -0.51378  0.026908  0.47245  2.5334

Number of Observations: 54
Number of Groups: 6

```

This model can be compared to the previous two models using the multi-argument form of `anova`.

```

> anova( fm1Machine, fm2Machine, fm3Machine )
      Model df    AIC    BIC   logLik   Test L.Ratio p-value
fm1Machine     1 5 300.46 310.12 -145.23
fm2Machine     2 6 231.27 242.86 -109.64 1 vs 2  71.191 <.0001
fm3Machine     3 10 231.89 251.21 -105.95 2 vs 3   7.376  0.1173

```

Because the *p*-value for the test comparing models 2 and 3 is about 12%, we would conclude that the fit `fm3Machine` is not significantly better than `fm2Machine`, taking into account the fact that `fm3Machine` requires four additional parameters in the model.

The AIC criterion is nearly the same for models 2 and 3, indicating that there is no strong preference between these models. The BIC criterion does indicate a strong preference for model 2 relative to model 3. In general BIC puts a heavier penalty than does AIC on having more parameters in the model. Because there are a total of ten parameters in model 3 compared

to six parameters in model 2, the BIC criterion will tend to prefer model 2 unless model 3 provides a substantially better fit.

## 1.4 An Analysis of Covariance Model

Traditionally, the term *analysis of variance* has been applied to models for a continuous response as it relates to various classification factors for the observations. The model for the rails example described in §1.1

$$y_{ij} = \beta + b_i + \epsilon_{ij}, \quad i = 1, \dots, M, \quad j = 1, \dots, n_i,$$

$$b_i \sim \mathcal{N}(0, \sigma_b^2), \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

is an example of an analysis of variance model with random-effects terms. A *linear regression model*, such as

$$y_i = \beta_1 + \beta_2 x_i + \epsilon_i, \quad i = 1, \dots, N, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2),$$

relates a continuous response (the  $y_i$ ) to one or more continuous covariates (the  $x_i$ ).

The term *analysis of covariance* designates a type of model that relates a continuous response to both a classification factor and to a continuous covariate. If  $y_{ij}$  is the  $j$ th observation in the  $i$ th group of data and  $x_{ij}$  is the corresponding value of the covariate, an analysis of covariance model with a random effect for the intercept would be

$$\begin{aligned} y_{ij} &= \beta_1 + b_i + \beta_2 x_{ij} + \epsilon_{ij}, \quad i = 1, \dots, M, \quad j = 1, \dots, n_i, \\ b_i &\sim \mathcal{N}(0, \sigma_b^2), \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2). \end{aligned} \tag{1.8}$$

This model combines a random-effects analysis of variance model with a linear regression model.

### 1.4.1 Modeling Simple Linear Growth Curves

A common application of random-effects analysis of covariance models is in modeling *growth curve* data—the results on different subjects of repeated measurements of some characteristic over time. The terms *repeated measures* and *longitudinal data* are also applied to such data.

A classic example of such data, given in Potthoff and Roy (1964), is a set of measurements of the distance from the pituitary gland to the pterygomaxillary fissure taken every two years from 8 years of age until 14 years of age on a sample of 27 children—16 males and 11 females. The data, available as the S object `Orthodont` and shown in Figure 1.11, were collected by orthodontists from x-rays of the children’s skulls. The pituitary gland and the pterygomaxillary fissure are two easily located points on these x-rays.

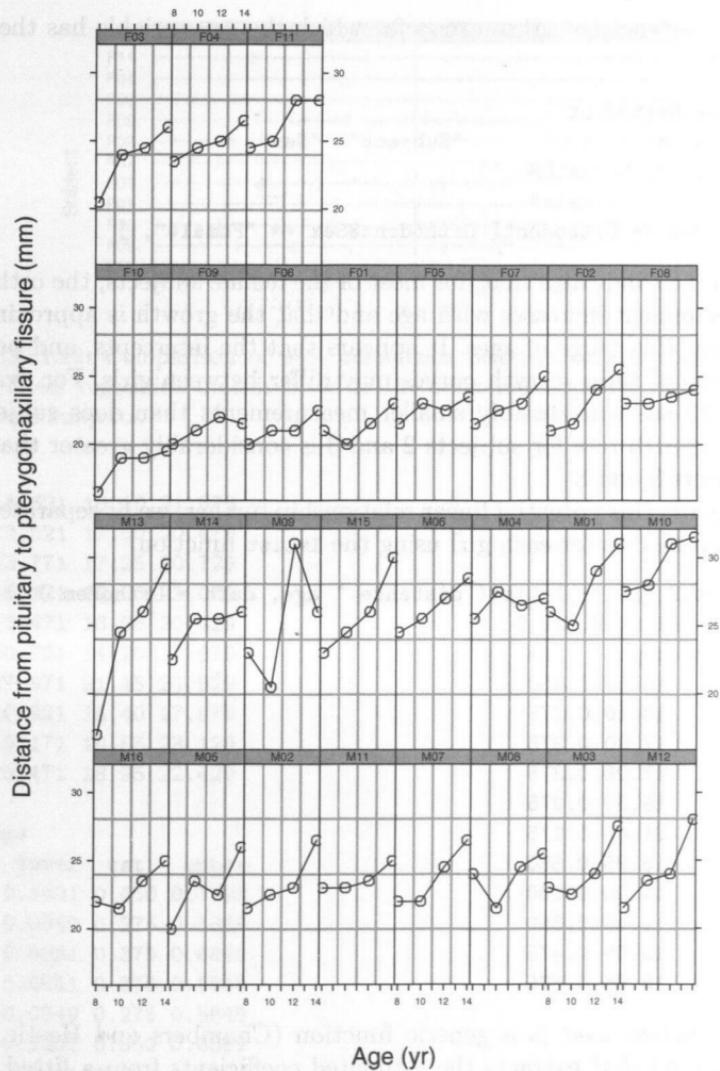


FIGURE 1.11. Distance from the pituitary to the pterygomaxillary fissure versus age for a sample of 16 boys (subjects M01 to M16) and 11 girls (subjects F01 to F11). The aspect ratio for the panels has been chosen to facilitate comparison of the slope of the lines.

From Figure 1.11 it appears that there are qualitative differences between boys and girls in their growth patterns for this measurement. In Chapter 4 we will model some of these differences, but for now it is easier to restrict our modeling to the data from the female subjects only. To extract the data for the females only we first check on the names of the variables in the `Orthodont` object, then check for the names of the levels of the variables

Sex, then extract only those rows for which the `Sex` variable has the value "Female".

```
> names( Orthodont )
[1] "distance" "age"      "Subject"  "Sex"
> levels( Orthodont$Sex )
[1] "Male"     "Female"
> OrthoFem <- Orthodont[ Orthodont$Sex == "Female", ]
```

Figure 1.11 indicates that, for most of the female subjects, the orthodontic measurement increases with age and that the growth is approximately linear over this range of ages. It appears that the intercepts, and possibly the slopes, of these growth curves may differ between girls. For example, subject 10 has considerably smaller measurements than does subject 11, and the growth rate for subjects 2 and 3 is considerably greater than that for subjects 5 and 8.

To explore this potential linear relationship further, we fit separate linear regression models for each girl using the `lmList` function.

```
> fm1OrthF.lis <- lmList( distance ~ age, data = OrthoFem )
> coef( fm1OrthF.lis )
   (Intercept)    age
F10       13.55 0.450
F09       18.10 0.275
F06       17.00 0.375
F01       17.25 0.375
F05       19.60 0.275
F08       21.45 0.175
F07       16.95 0.550
F02       14.20 0.800
F03       14.40 0.850
F04       19.65 0.475
F11       18.95 0.675
```

The function `coef` is a generic function (Chambers and Hastie, 1992, Appendix A) that extracts the estimated coefficients from a fitted model object. For an `lmList` object the coefficients are returned as a matrix with one row for each of the groups of observations.

We might wish to consider whether we need to allow different slopes for each girl. There are formal statistical tests to assess this and we will discuss them later. For now we can proceed informally and examine individual confidence intervals on the parameters. As we have seen, the `intervals` function is used to create confidence intervals on the parameters in an object representing a fitted model.

```
> intervals( fm1OrthF.lis )
, , (Intercept)
      lower  est.  upper
F10 10.071 13.55 17.029
```

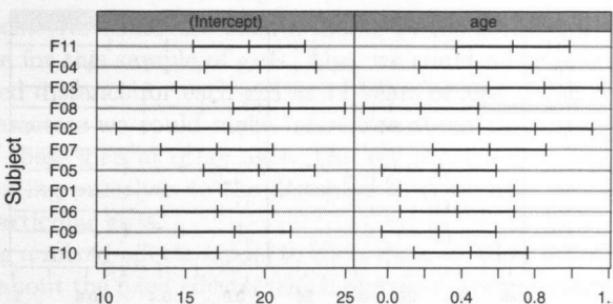


FIGURE 1.12. Comparison of 95% confidence intervals on the coefficients of simple linear regression models fitted to the orthodontic growth curve data for the female subjects.

```
F09 14.621 18.10 21.579
F06 13.521 17.00 20.479
F01 13.771 17.25 20.729
F05 16.121 19.60 23.079
F07 13.471 16.95 20.429
F02 10.721 14.20 17.679
F08 17.971 21.45 24.929
F03 10.921 14.40 17.879
F04 16.171 19.65 23.129
F11 15.471 18.95 22.429
```

```
, , age
      lower   est.   upper
F10  0.1401  0.450  0.7599
F09 -0.0349  0.275  0.5849
F06  0.0651  0.375  0.6849
F01  0.0651  0.375  0.6849
F05 -0.0349  0.275  0.5849
F07  0.2401  0.550  0.8599
F02  0.4901  0.800  1.1099
F08 -0.1349  0.175  0.4849
F03  0.5401  0.850  1.1599
F04  0.1651  0.475  0.7849
F11  0.3651  0.675  0.9849
```

As often happens, displaying the intervals as a table of numbers is not very informative. We find it much more effective to plot these intervals using

```
> plot( intervals( fm1OrthF.lis ) ) # produces Figure 1.12
```

Figure 1.12 is of interest as much for what it does not show as for what it does show. First, consider what the figure does show. We notice that the intervals for the intercepts are all the same width, as are the intervals

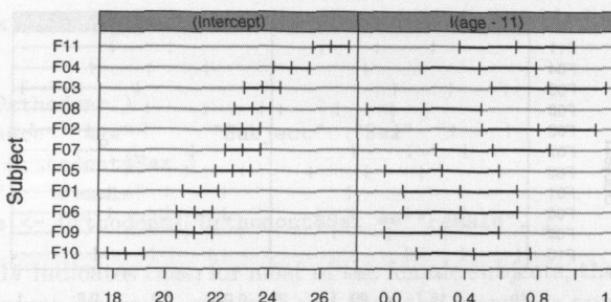


FIGURE 1.13. Comparison of 95% confidence intervals on the coefficients of simple linear regression models fitted to the centered orthodontic growth curve data for the female subjects.

for the slope with respect to age. This is a consequence of having *balanced* data; that is, all the subjects were observed the same number of times and at the same ages. We also notice that there is considerable overlap in the set of intervals for the slope with respect to age. It may be feasible to use a model with a common slope.

The surprising thing about Figure 1.12 is that it does not show the substantial differences in the intercepts that Figure 1.11 would lead us to expect. Furthermore, even though we have ordered the groups from the one with the smallest average distance (subject F10) to the one with the largest average distance (subject F11), this ordering is not reflected in the intercepts. Finally, we see that the pattern across subjects in the intervals for the intercepts is nearly a reflection of the pattern in the intervals for the slopes.

Those with experience analyzing regression models may already have guessed why this reflection of the pattern occurs. It occurs because all the data were collected between age 8 and age 14, but the intercept represents a distance at age 0. The extrapolation back to age 0 will result in a high negative correlation (about  $-0.98$ ) between the estimates of the slopes and their corresponding intercept estimate.

We will remove this correlation if we center the data. In this case, we would fit the distance as a linear function of  $\text{age} - 11$  so the two coefficients being estimated are the distance at 11 years of age and the slope or growth rate. If we fit this revised model and plot the confidence intervals

```
> fm2OrthF.lis <- update(fm1OrthF.lis, distance ~ I(age - 11))
> plot(intervals(fm2OrthF.lis)) # produces Figure 1.13
```

then these intervals (Figure 1.13) show the expected trend in the (Intercept) term, which now represents the fitted distance at 11 years.

To continue with the analysis of these data we could fit a regression model to the centered data with a common growth rate but separate intercepts for each girl. Before doing that we should consider what we could infer from

such a model. We could use such a model to make inferences about the growth rate for this sample of girls. Also, we could make inferences about the expected distance for each girl at 11 years of age. Using combinations of the parameters we could make inferences about the expected distance for each of these girls at other ages. The key point is that we are in some ways restricting ourselves to the distances that we have or could observe on these particular girls.

By fitting a mixed-effects model to these data we allow ourselves to make inferences about the fixed effects, which represent average characteristics of the population represented by these subjects, and the variability amongst subjects. A call to `lme` to fit linear growth curves with common slopes but randomly distributed shifts to the girls' orthodontic data is

```
> fm1OrthF <-  
+   lme( distance ~ age, data = OrthoFem, random = ~ 1 | Subject )  
> summary( fm1OrthF )  
Linear mixed-effects model fit by REML  
Data: OrthoFem  
      AIC      BIC    logLik  
 149.22 156.17 -70.609  
  
Random effects:  
 Formula: ~ 1 | Subject  
          (Intercept) Residual  
 StdDev:     2.0685  0.78003  
  
Fixed effects: distance ~ age  
              Value Std.Error DF t-value p-value  
(Intercept) 17.373  0.85874 32  20.230 <.0001  
       age     0.480  0.05259 32   9.119 <.0001  
Correlation:  
   (Intr)  
age -0.674  
  
Standardized Within-Group Residuals:  
      Min        Q1        Med        Q3        Max  
-2.2736 -0.70902  0.17282  0.41221  1.6325  
  
Number of Observations: 44  
Number of Groups: 11
```

We could also fit a model with the formula `distance ~ I(age - 11)` but, because of the requirement of a common slope, for model building purposes the properties of the centered model are essentially equivalent to the uncentered model. Using the uncentered model makes it easier to compare with other models described below.

The model being fit would be expressed in matrix notation as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, 11,$$

$$\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \Psi), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}),$$

with matrices

$$\mathbf{X}_1 = \dots = \mathbf{X}_{11} = \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{bmatrix}, \quad \mathbf{Z}_1 = \dots = \mathbf{Z}_{11} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

The two-dimensional fixed-effects vector  $\boldsymbol{\beta}$  consists of the mean intercept,  $\beta_1$ , for the population and the common slope or growth rate,  $\beta_2$ . The one-dimensional random-effects vectors,  $\mathbf{b}_i, i = 1, \dots, 11$ , describe a shift in the intercept for each subject. Because there is a common growth rate, these shifts are preserved for all values of age. The matrix  $\Psi = \sigma_b^2$  will be a  $1 \times 1$  matrix in this case. It represents the variance of the measurements in the population at a fixed value of age.

The REML estimates for the parameters are

$$\hat{\sigma}_b = 2.0685, \quad \hat{\sigma} = 0.78003, \quad \hat{\beta}_1 = 17.373, \quad \hat{\beta}_2 = 0.480.$$

To obtain the maximum likelihood estimates we use `method = "ML"`.

```
> fm1OrthFM <- update( fm1OrthF, method = "ML" )
> summary( fm1OrthFM )
Linear mixed-effects model fit by maximum likelihood
Data: OrthoFem
      AIC      BIC   logLik
 146.03  153.17 -69.015

Random effects:
Formula: ~ 1 | Subject
          (Intercept) Residual
StdDev:     1.9699  0.76812

Fixed effects: distance ~ age
                Value Std.Error DF t-value p-value
(Intercept) 17.373   0.85063 32  20.423 <.0001
           age  0.480   0.05301 32   9.047 <.0001

Correlation:
  (Intr)
age -0.685

Standardized Within-Group Residuals:
    Min      Q1      Med      Q3      Max 
-2.3056 -0.71924 0.17636 0.4258 1.6689
```

Number of Observations: 44

Number of Groups: 11

Notice that, to the accuracy printed here, the estimates of the fixed-effects parameters are the same for ML and REML. The ML estimates of the standard deviations,  $\hat{\sigma}_b = 1.9699$  and  $\hat{\sigma} = 0.76812$  are smaller than the corresponding REML estimates. This is to be expected—the REML criterion was created to compensate for the downward bias of the maximum likelihood estimates of variance components, so it should produce larger estimates.

We have made the assumption of a common slope or growth rate for all the subjects. To test this we can fit a model with random effects for both the intercept and the slope.

```
> fm2OrthF <- update(fm1OrthF, random = ~ age | Subject)
```

The predictions from this model are shown in Figure 1.14. We compare the two models with the `anova` function.

```
> anova(fm1OrthF, fm2OrthF)
      Model df     AIC     BIC logLik  Test L.Ratio p-value
fm1OrthF     1  4 149.22 156.17 -70.609
fm2OrthF     2  6 149.43 159.85 -68.714 1 vs 2   3.7896 0.1503
```

Because the *p*-value for the second model versus the first is about 15%, we conclude that the simpler model, `fm1OrthF`, is adequate.

### 1.4.2 Predictions of the Response and the Random Effects

The derivation of predicted values for the response and for the random effects in the linear mixed-effects model is described in §2.5. We can extract the *best linear unbiased predictions* (BLUPs) of the random effects from the fitted model with the `random.effects` function.

```
> random.effects(fm1OrthF)
      (Intercept)
F10    -4.005329
F09    -1.470449
F06    -1.470449
F01    -1.229032
F05    -0.021947
F07    0.340179
F02    0.340179
F08    0.702304
F03    1.064430
F04    2.150807
F11    3.599309
```

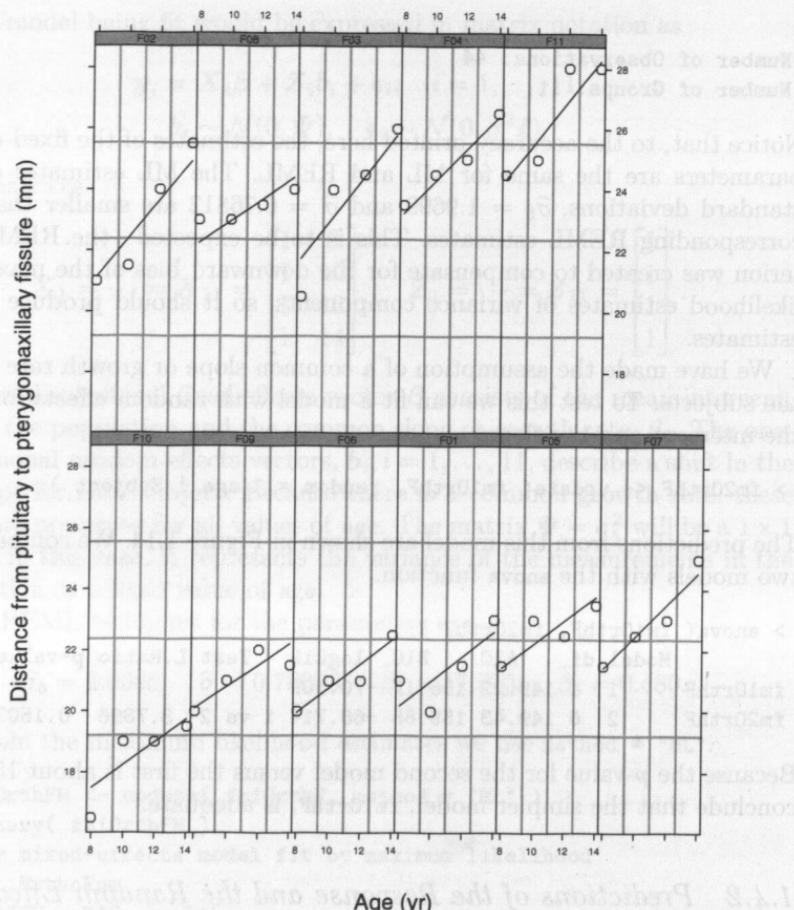


FIGURE 1.14. Original data and fitted linear relationships from a mixed-effects model for the girls' orthodontic data. This model incorporates random effects for both the slope and the intercept.

The shorter name `ranef` is a synonym for `random.effects`.

```
> ranef( fm10OrthFM )
   (Intercept)
F10    -3.995835
F09    -1.466964
F06    -1.466964
F01    -1.226119
F05    -0.021895
F07     0.339372
F02     0.339372
F08     0.700640
F03     1.061907
```

```
F04    2.145709
F11    3.590778
```

The `coefficients` function (or its shorter form `coef`) is used to extract the coefficients of the fitted lines for each subject. For the fitted model `fm10rthF` the intercept of the fitted line for subject  $i$  is  $\hat{\beta}_1 + \hat{b}_i$  and the slope is  $\hat{\beta}_2$ .

```
> coef(fm10rthF)
   (Intercept)      age
F10        13.367 0.47955
F09        15.902 0.47955
F06        15.902 0.47955
F01        16.144 0.47955
F05        17.351 0.47955
F07        17.713 0.47955
F02        17.713 0.47955
F08        18.075 0.47955
F03        18.437 0.47955
F04        19.524 0.47955
F11        20.972 0.47955
```

Looking back at the BLUPs of the random effects for the ML and REML fits, we can see that they are very similar. The same will be true of the coefficients for each subject and hence for the fitted lines themselves. To show this we can plot either the estimated BLUPs or the estimated coefficients. The `compareFits` function is helpful here because it allows us to put both sets of coefficients on the same panels.

```
> plot(compareFits(coef(fm10rthF), coef(fm10rthFM))) # Figure 1.15
```

In Figure 1.15 each line corresponds to one subject. In the left panel the estimated intercepts from the REML fit are shown as open circles while those from the ML fit are shown as +'s. The two estimates for each subject are essentially identical. In the right panel the estimates for the coefficient with respect to `age` are shown. Because there is no random effect associated with this coefficient, the estimates do not vary between subjects. Again, the ML estimates and the REML estimates are essentially identical.

We may also want to examine the predictions for each subject from the fitted model. The `augPred` function produces predictions of the response for each of the groups over the observed range of the covariate (i.e. the range 8–14 for `age`). These predictions are augmented with the original data to produce a plot of the predicted lines for each subject superposed on the original data as in Figure 1.16.

```
> plot(augPred(fm10rthF), aspect = "xy", grid = T) # Fig. 1.16
```

Further diagnostic plots, such as plots of residuals versus the fitted values by subject (not shown), did not indicate any serious deficiencies in this

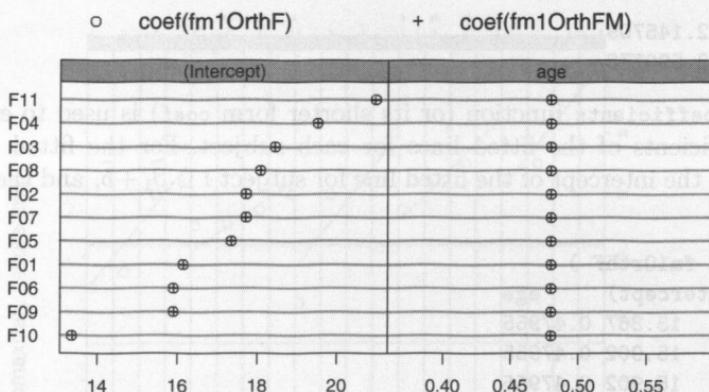


FIGURE 1.15. A comparison of the coefficients of the fitted lines for each female subject in the orthodontic example. The two sets of coefficients are from the restricted maximum likelihood fit ( $\text{fm1OrthF}$ ) and the maximum likelihood fit ( $\text{fm1OrthFM}$ ).

model. We will return to this data set in Chapter 4 where we will fit a combined model to the data for both the males and the females. This allows us to check for sex-related differences in the growth patterns.

## 1.5 Models for Nested Classification Factors

In the **Machines** example of §1.3 we introduced the concept of nested random effects to model an interaction between a fixed-effects factor and a random-effects factor. Nested random-effects terms are also used when we have nested classification factors.

Data from an experiment on the pixel intensity in computerized tomography (CT) scans, available as the object **Pixel**, are shown in Figure 1.17 and are described in Appendix A.24. The experimenters injected each of ten dogs with a dye contrast then recorded the mean pixel intensities from CT scans of the right and left lymph nodes in the axillary region on several occasions up to 21 days post injection.

Each observation is classified according to the **Dog** and the **Side** of the **Dog** on which it was made. The nature of the experiment is such that the left and right sides are expected to be different, but the difference is not expected to be systematic in terms of left and right. That is, for one dog the left side may have greater pixel intensities than the right, while for another dog the opposite may be true. Thus **Dog** and **Side** are considered to be nested classification factors. We will associate random-effects terms with the **Dog** factor, and with the **Side** factor nested within **Dog**.

Figure 1.17 indicates that the intensities generally increase then decrease over time, reaching a peak after about 10 days. There is, however, consid-

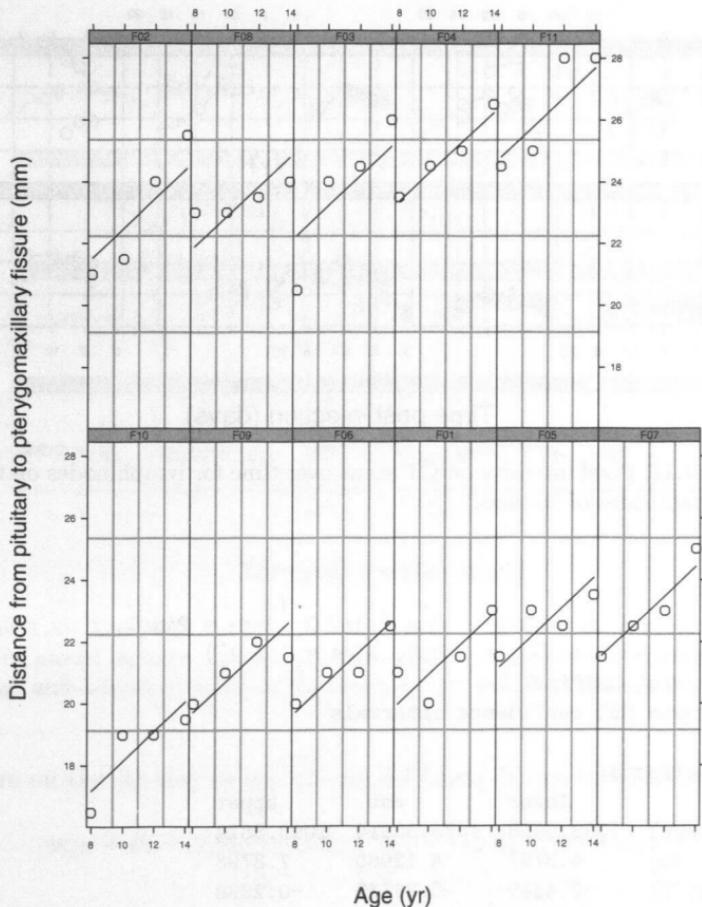


FIGURE 1.16. Original data and fitted growth curves for each female subject in the orthodontic example. The fitted curves are from a restricted maximum likelihood fit of the analysis of covariance model.

erable variability between dogs in this pattern. Within the same dog, the left and the right side generally follow the same pattern over time but often with a vertical shift between sides.

We will start with a quadratic model with respect to the day covariate so we can model the pattern of reaching a peak. We use random-effects for both the intercept and the linear term at the Dog level and a single random effect for the intercept at the Side within Dog level. This allows the overall pattern to vary between dogs in terms of the location of the peak, but not in terms of the curvature at the peak. The only difference between sides for the same dog will be a shift in the intercept.

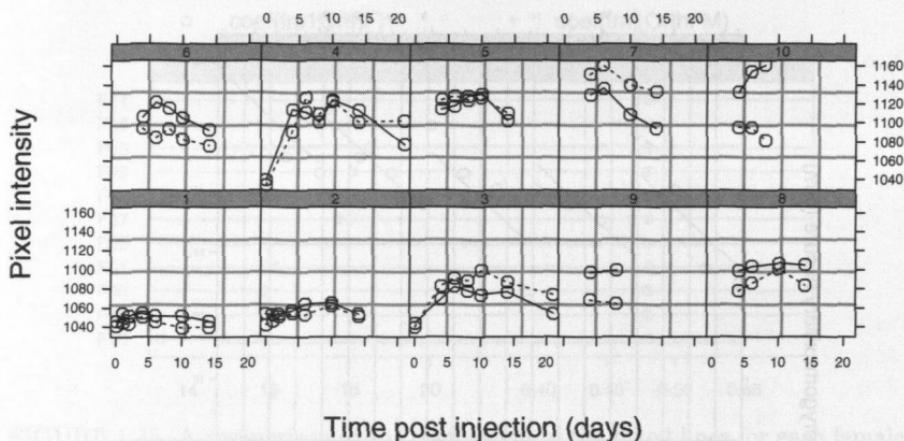


FIGURE 1.17. Pixel intensity on CT scans over time for lymph nodes on the left and the right sides of 10 dogs.

```
> fm1Pixel <- lme( pixel ~ day + day^2, data = Pixel,
+   random = list( Dog = ~ day, Side = ~ 1 ) )
> intervals( fm1Pixel )
```

Approximate 95% confidence intervals

Fixed effects:

	lower	est.	upper
(Intercept)	1053.0968	1073.33914	1093.5814
day	4.3797	6.12960	7.8795
I(day^2)	-0.4349	-0.36735	-0.2998

Random Effects:

Level: Dog

	lower	est.	upper
sd((Intercept))	15.92849	28.36994	50.52918
sd(day)	1.08085	1.84375	3.14514
cor((Intercept),day)	-0.89452	-0.55472	0.19138

Level: Side

	lower	est.	upper
sd((Intercept))	10.417	16.824	27.173

Within-group standard error:

lower	est.	upper
7.6345	8.9896	10.585

```
> plot( augPred( fm1Pixel ) ) # produces Figure 1.18
```

If we write the pixel intensity of the  $j$ th side's on the  $i$ th dog at the  $k$ th occasion as  $y_{ijk} i = 1, \dots, 10; j = 1, 2; k = 1, \dots, n_{ij}$ , and the time of the

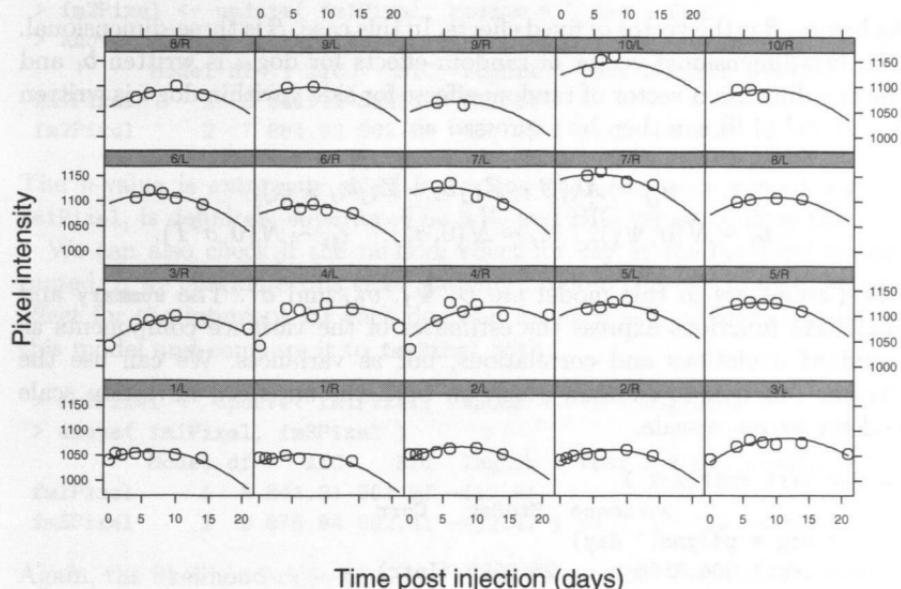


FIGURE 1.18. The fitted curves from a quadratic model fit to the pixel intensity data. The model includes random effects in the intercept and the linear term for each dog, and a random effect in the intercept for each side of each dog.

$k$ th scan on the  $i$ th dog as  $d_{ik}$ , the model being fit can be expressed as

$$y_{ijk} = \beta_1 + \beta_2 d_{ik} + \beta_3 d_{ik}^2 + b_{i,1} + b_{i,2} d_{ik} + b_{ij} + \epsilon_{ijk}, \quad (1.9)$$

$$i = 1, \dots, 10, \quad j = 1, 2, \quad k = 1, \dots, n_{ij}.$$

To describe the variance and covariance terms in the model, we consider the  $n_{ij}$ -vector  $\mathbf{y}_{ij}$  of intensities measured on side  $j = 1, 2$  within dog  $i = 1, \dots, 10$ . In this experiment  $n_{ij}$ , the number of observations on side  $j$  of dog  $i$ , does not depend on  $j$  but does depend on  $i$ . For example, dog 9 was scanned on only two occasions but dogs 3 and 4 were each scanned on seven occasions. The model can be represented in terms of design matrices  $\mathbf{X}_{ij}$  for the fixed effects,  $\mathbf{Z}_{ij}$  for the random effects for side  $j$  within dog  $i$ , and  $\mathbf{Z}_{i,j}$  for the random effects for dog  $i$  on the measurements for side  $j$  within dog  $i$ .

Because both sides are scanned at the same times these matrices depend on  $i$  but not on  $j$ . For example, because dog 8 was scanned on days 4, 6, 10, and 14,

$$\mathbf{X}_{81} = \mathbf{X}_{82} = \begin{bmatrix} 1 & 4 & 16 \\ 1 & 6 & 36 \\ 1 & 10 & 100 \\ 1 & 14 & 196 \end{bmatrix}, \quad \mathbf{Z}_{8,1} = \mathbf{Z}_{8,2} = \begin{bmatrix} 1 & 4 \\ 1 & 6 \\ 1 & 10 \\ 1 & 14 \end{bmatrix}, \quad \mathbf{Z}_{81} = \mathbf{Z}_{82} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

As before,  $\beta$  is the vector of fixed effects. In this case,  $\beta$  is three-dimensional. The two-dimensional vector of random effects for dog  $i$  is written  $b_i$  and the one-dimensional vector of random effects for side  $j$  within dog  $i$  is written  $b_{ij}$ . Model (1.9) can then be expressed as

$$\begin{aligned} y_{ij} &= X_{ij}\beta + Z_{i,j}b_i + Z_{ij}b_{ij} + \epsilon_{ij}, \\ b_i &\sim \mathcal{N}(\mathbf{0}, \Psi_1), \quad b_{ij} \sim \mathcal{N}(0, \sigma_2^2), \quad \epsilon_{ij} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I). \end{aligned}$$

The parameters in this model are  $\beta$ ,  $\Psi_1$ ,  $\sigma_2^2$ , and  $\sigma^2$ . The `summary` and `intervals` functions express the estimates of the variance components as standard deviations and correlations, not as variances. We can use the `VarCorr` function to examine them on both the standard deviation scale and the variance scale.

```
> VarCorr( fm1Pixel )
          Variance StdDev   Corr
Dog = pdSymm(~ day)
(Intercept) 804.8535    28.3699 (Intr)
           day     3.3994    1.8437 -0.555
Side = pdSymm(~ 1)
(Intercept) 283.0551    16.8242
Residual    80.8131     8.9896
```

### 1.5.1 Model Building for Multilevel Models

As when modeling data with a single level of random effects, we should evaluate whether the fixed-effects structure and the random-effects structure are adequate to describe the observed data. We should also check if we have incorporated unnecessary terms in the model.

The summary table for the fixed-effects terms

```
> summary( fm1Pixel )
...
Fixed effects: pixel ~ day + day^2
              Value Std.Error DF t-value p-value
(Intercept) 1073.3    10.172 80  105.52 <.0001
           day      6.1     0.879 80    6.97 <.0001
I(day^2)    -0.4     0.034 80   -10.82 <.0001
...
...
```

indicates that the quadratic term is highly significant. In a polynomial model like this we would generally retain the linear term and the intercept term if we retain the quadratic term. Thus, we will accept the fixed-effects model as it is and go on to examine the random-effects terms.

The first question to address is whether there is a need to have a random effect for each `Side` within each `Dog`. We can fit the previous model without the random effect for `Side` and compare the two fits with `anova`.

```
> fm2Pixel <- update(fm1Pixel, random = ~ day | Dog)
> anova(fm1Pixel, fm2Pixel)
    Model df     AIC     BIC   logLik   Test L.Ratio p-value
fm1Pixel     1  8 841.21 861.97 -412.61
fm2Pixel     2  7 884.52 902.69 -435.26 1 vs 2  45.309 <.0001
```

The *p*-value is extremely small indicating that the more general model, `fm1Pixel`, is definitely superior. The AIC and BIC values confirm this.

We can also check if the random effect for `day` at the `Dog` level is warranted. If we eliminate this term the only random effects will be a random effect for the intercept for each dog and for each side of each dog. We fit this model and compare it to `fm1Pixel` with

```
> fm3Pixel <- update(fm1Pixel, random = ~ 1 | Dog/Side)
> anova(fm1Pixel, fm3Pixel)
    Model df     AIC     BIC   logLik   Test L.Ratio p-value
fm1Pixel     1  8 841.21 861.97 -412.61
fm3Pixel     2  6 876.84 892.41 -432.42 1 vs 2  39.629 <.0001
```

Again, the likelihood-ratio test, and the AIC and BIC criteria, all strongly favor the more general model, `fm1Pixel`.

Earlier we stated that there does not appear to be a systematic difference between the left and the right sides of the dogs. For some dogs the left side produces higher pixel densities while for other dogs the right side does. We can check that this indeed is the case by adding a term for `Side` to the fixed effects.

```
> fm4Pixel <- update(fm1Pixel, pixel ~ day + day^2 + Side)
> summary(fm4Pixel)
...
Fixed effects: pixel ~ day + I(day^2) + Side
            Value Std.Error DF t-value p-value
(Intercept) 1073.3   10.171 80 105.53 <.0001
      day      6.1    0.879 80    6.97 <.0001
    I(day^2)   -0.4    0.034 80   -10.83 <.0001
      Side     -4.6    3.813  9    -1.21  0.2576
...
```

With a *p*-value of over 25% the fixed-effects term for `Side` would not be considered significant.

Finally, we would examine residual plots such as Figure 1.19 for deficiencies in the model. There are no alarming patterns in this figure.

## 1.6 A Split-Plot Experiment

Multiple nested levels of random effects are also used in the analysis of *split-plot* experiments such as that represented by the `Oats` data, shown in

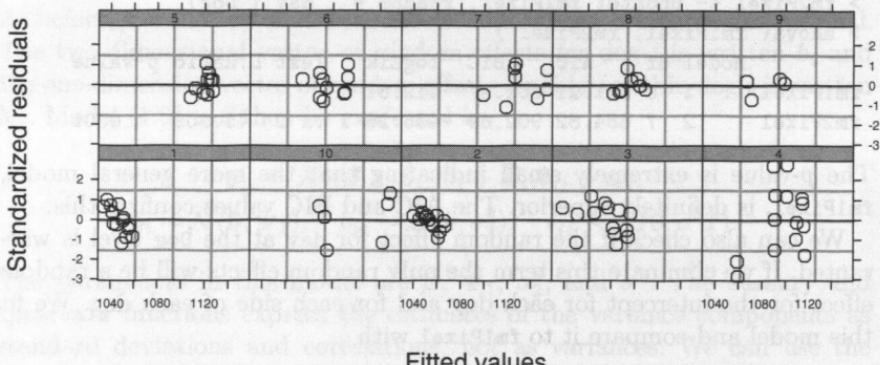


FIGURE 1.19. Standardized residuals versus fitted values by dog for a multilevel mixed-effects model fit to the pixel data..

Figure 1.20. As described in Appendix A.15, the treatment structure in this experiment was a  $3 \times 4$  full factorial, with three varieties of oats and four nitrogen concentrations. The term *full factorial* means that every variety was used with every nitrogen concentration.

The agricultural plots for this experiment were grouped into six blocks, each with three plots. Each plot was subdivided into four subplots. The varieties were randomly assigned to the plots within each block. The nitrogen concentrations were randomly assigned to the subplots within each plot.

Physically, there are three levels of grouping of the experimental units: block, plot, and subplot. Because the treatments are randomly assigned at each level of grouping, we may be tempted to associate random effects with each level. However, because there is only one yield recorded for each subplot we cannot do this as we would saturate the model with random effects. We use a random intercept at each of the block and the whole plot levels.

Generally, we begin modeling a split-plot experiment using fixed effects for each of the experimental factors and for their interaction. For this experiment the `nitro` factor is recorded as a numeric variable. If we wish to allow general patterns in the dependencies of `yield` on `nitro` we should “coerce” it to a factor using, say, `factor(nitro)`. In this particular example, there is a natural ordering of the levels of nitrogen applied so it makes sense to coerce `nitro` to an *ordered factor* using `ordered(nitro)`. As the name implies, an ordered factor is a factor for which there is a natural ordering of the levels. One consequence of using an ordered factor instead of a factor is that the default contrasts for an ordered factor are orthogonal polynomial contrasts. The first contrast estimates the linear trend, the second estimates the quadratic effect orthogonal to the linear term, and so on.

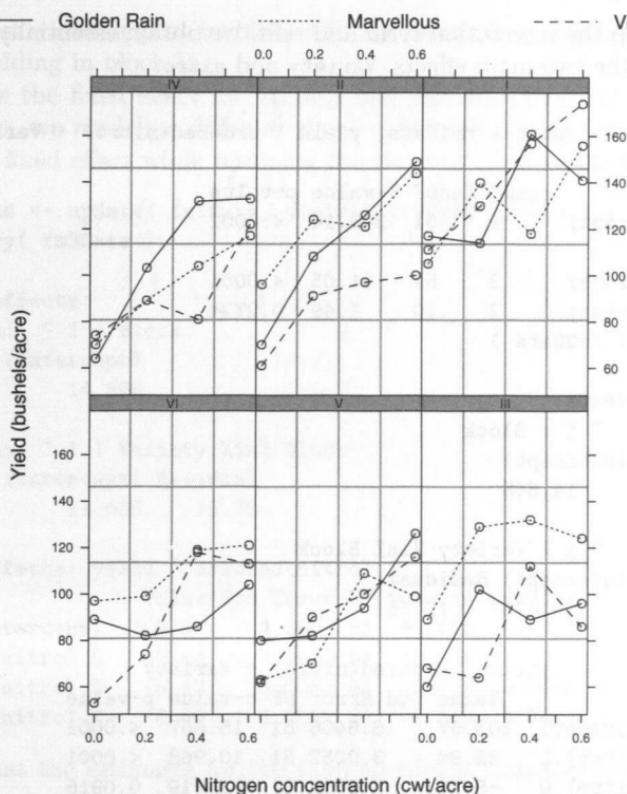


FIGURE 1.20. Yield in bushels/acre of three different varieties of oats at four different concentrations of nitrogen (hundred weight/acre). The experimental units were arranged into six blocks, each with three whole-plots subdivided into four subplots. One variety of oats was used in each whole-plot with all four concentrations of nitrogen, one concentration in each of the four subplots. The panels correspond to the blocks.

The model with fixed effects for both experimental factors and for their interaction and with random effects for both the Block factor and the Variety (whole-plot) factor is fit with

```
> fm10ats <- lme( yield ~ ordered(nitro) * Variety, data = Oats,
+   random = ~ 1 | Block/Variety )
> anova( fm10ats )

```

	numDF	denDF	F-value	p-value
(Intercept)	1	45	245.15	<.0001
ordered(nitro)	3	45	37.69	<.0001
Variety	2	10	1.49	0.2724
ordered(nitro):Variety	6	45	0.30	0.9322

The anova results indicate that nitro is a significant factor, but that neither Variety nor the interaction between Variety and nitro are significant.

If we drop the interaction term and refit, we obtain essentially the same results for the two main effects, **Variety** and **nitro**.

```
> fm20ats <- update( fm10ats, yield ~ ordered(nitro) + Variety )
> anova( fm20ats )
      numDF denDF F-value p-value
(Intercept)     1      51   245.14 <.0001
ordered(nitro)    3      51    41.05 <.0001
Variety         2      10     1.49  0.2724
> summary( fm20ats )
...
Random effects:
Formula: ~ 1 | Block
(Intercept)
StdDev: 14.645

Formula: ~ 1 | Variety %in% Block
(Intercept) Residual
StdDev: 10.473 12.75

Fixed effects: yield ~ ordered(nitro) + Variety
      Value Std.Error DF t-value p-value
(Intercept) 103.97  6.6406 51 15.657 <.0001
ordered(nitro).L 32.94  3.0052 51 10.963 <.0001
ordered(nitro).Q -5.17  3.0052 51 -1.719  0.0916
ordered(nitro).C -0.45  3.0052 51 -0.149  0.8823
Variety1      2.65  3.5395 10  0.748  0.4720
Variety2     -3.17  2.0435 10 -1.553  0.1515
...
```

In this model there is a random effect for **Variety %in% Block** as well as a fixed effect for **Variety**. These terms model different characteristics of the response. The random effects term, as a nested random effect, is allowing for different intercepts at the level of plots within blocks. The fact that each plot is planted with one variety means that we can use the **Variety** factor to indicate the plot as long as we have **Variety** nested within **Block**. As seen in Figure 1.20 the yields in one of the plots within a block may be greater than those on another plot in the same block for all levels of **nitro**. For example, in block III the plot that was planted with the Marvellous variety had greater yields than the other two plots at each level of **nitro**. The random effect at the level of **Variety %in% Block** allows shifts like this that may be related to the fertility of the soil in that plot, for example.

On the other hand, the fixed-effects term for **Variety** is used to model a systematic difference in the yields that would be due to the variety of oats planted in the plot. There do not appear to be such systematic differences. For example, even though the plot planted with the Marvellous variety is

the highest yielding plot in block III, the Marvellous plot is one of the lowest yielding in block V.

Because the fixed effect for `Variety` and the random effect for `Variety %in% Block` are modeling different types of behavior, it makes sense to remove the fixed effect while retaining the random effect

```
> fm30ats <- update( fm10ats, yield ~ ordered( nitro ) )
> summary( fm30ats )
...
Random effects:
Formula: ~ 1 | Block
          (Intercept)
StdDev:    14.506

Formula: ~ 1 | Variety %in% Block
          (Intercept) Residual
StdDev:    11.039    12.75

Fixed effects: yield ~ ordered(nitro)
                Value Std.Error DF t-value p-value
(Intercept) 103.97    6.6406 51 15.657 <.0001
ordered(nitro).L 32.94    3.0052 51 10.963 <.0001
ordered(nitro).Q -5.17    3.0052 51 -1.719 0.0916
ordered(nitro).C -0.45    3.0052 51 -0.149 0.8823
```

We see that the estimates for the random-effects variances and the fixed-effects for `nitro` have changed very little, if at all.

We can now examine the effect of nitrogen in more detail. We notice that the linear term, `ordered(nitro).L`, is highly significant, but the quadratic and cubic terms (.Q and .C extensions) are not. To remove the cubic and quadratic terms in the model, we simply revert to using `nitro` as a numeric variable.

```
> fm40ats <-
+   lme( yield ~ nitro, data = Oats, random = ~ 1 | Block/Variety )
> summary( fm40ats )
...
Random effects:
Formula: ~ 1 | Block
          (Intercept)
StdDev:    14.506

Formula: ~ 1 | Variety %in% Block
          (Intercept) Residual
StdDev:    11.005    12.867

Fixed effects: yield ~ nitro
                Value Std.Error DF t-value p-value
(Intercept) 81.872    6.9453 53 11.788 <.0001
nitro      73.667    6.7815 53 10.863 <.0001
```

```

Correlation:
  (Intrc
nitro -0.293

Standardized Within-Group Residuals:
    Min      Q1      Med      Q3     Max
-1.7438 -0.66475  0.017104  0.54299  1.803

Number of Observations: 72
Number of Groups:
  Block Variety %in% Block
    6           18

```

With `VarCorr` and `intervals` we can examine the variance components and their confidence intervals for this model

```

> VarCorr( fm40ats )
    Variance StdDev
  Block = pdSymm(~ 1)
(Intercept) 210.42     14.506
  Variety = pdSymm(~ 1)
(Intercept) 121.10     11.005
  Residual   165.56     12.867
> intervals( fm40ats )
Approximate 95% confidence intervals

```

```

Fixed effects:
    lower   est.   upper
(Intercept) 67.942 81.872 95.803
  nitro       60.065 73.667 87.269

```

Random Effects:

Level: Block

	lower	est.	upper
sd((Intercept))	6.6086	14.506	31.841

Level: Variety

	lower	est.	upper
sd((Intercept))	6.408	11.005	18.899

Within-group standard error:

	lower	est.	upper
10.637	12.867	15.565	

We can see that the random effects at the `Block` and `plot` levels account for a substantial amount of the variability in the response. Although the standard deviations of these random effects are not estimated very precisely, it does not appear reasonable that they could be zero. To check this we would fit models without these random effects and use likelihood ratio tests to compare them to `fm40ats`. We do not show that here.

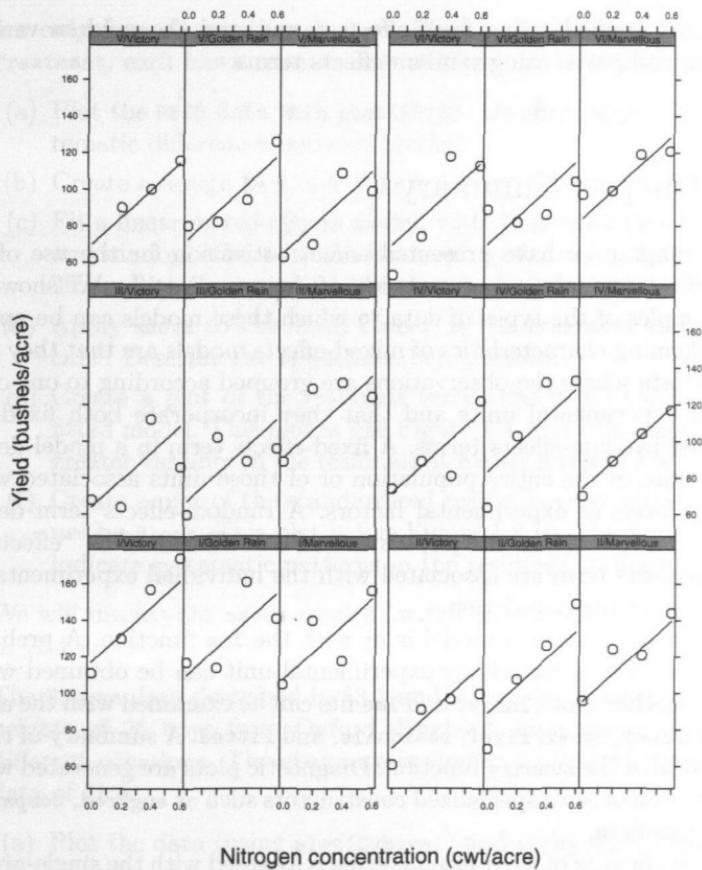


FIGURE 1.21. Observed and predicted yields in bushels/acre for three different varieties of oats at four different concentrations of nitrogen (hundred weight/acre) by block and variety. Although the model has a random effect for variety, the whole-plot factor, there is no fixed effect for variety.

The modeling of the dependence on nitrogen level by a simple linear term appears adequate. Plots of the original data and the fitted curves, obtained with

```
> plot( augPred( fm4Oats ), aspect = 2.5, layout = c(6, 3),
+       between = list( x = c(0, 0, 0.5) ) ) # produces Figure 1.21
```

do not show any systematic lack of fit. (The extra arguments in the `plot` call are used to enhance the appearance of the plot. They are described in §3.3.)

We could (and did) examine other common diagnostic plots to check for inadequacies in this model, but did not find any. We now have a simple, adequate model to explain the dependence of the response on both the levels

of nitrogen applied, using fixed-effects terms, and the random variability in blocks and plots, using random-effects terms.

## 1.7 Chapter Summary

In this chapter we have presented some motivation for the use of linear mixed-effects models and an overview of their application. We showed several examples of the types of data to which these models can be applied.

The defining characteristics of mixed-effects models are that they are applied to data where the observations are grouped according to one or more levels of experimental units and that they incorporate both fixed-effects terms and random-effects terms. A fixed-effects term in a model describes the behavior of the entire population or of those units associated with repeatable levels of experimental factors. A random-effects term describes the distribution within the population of a coefficient. The “effects” in a random-effects term are associated with the individual experimental units sampled from the population.

A linear mixed-effects model is fit with the `lme` function. A preliminary list of fitted linear models by experimental unit can be obtained with the `lmList` function. Both `lmList` and `lme` fits can be examined with the accessor functions `coef`, `ranef`, `fixef`, `residuals`, and `fitted`. A summary of the fit is obtained with the `summary` function. Diagnostic plots are generated with the `plot` function or with specialized constructors such as `augPred`, `compareFits`, and `comparePred`.

The significance of fixed-effects terms is assessed with the single-argument form of the `anova` function or directly from the the `summary` function. Different forms of random-effects terms can be compared by fitting different models and comparing them with the multiple-argument form of `anova`.

The purpose of this chapter is to present the motivation for using LME models with grouped data and to set the stage for later chapters in the book, dealing with the theory and computational methods for LME models (Chapter 2) and the linear mixed-effects modeling facilities in the `nlme` library (Chapter 4).

## Exercises

1. The `PBIB` data (Appendix A.22) are from an agricultural experiment that was laid out as a partially balanced incomplete block design. This is described in more detail in §2.4.2. The roles of the variables in these data are indicated by the names: `response`, `Treatment`, and `Block`. The structure is similar to that of the `ergoStool` data but, because there

are only four observations in each block and there are 15 levels of `Treatment`, each block receives only a subset of the treatments.

- (a) Plot the `PBIB` data with `plot(PBIB)`. Do there appear to be systematic differences between blocks?
- (b) Create a design plot, like Figure 1.6 (p. 14), for the `PBIB` data.
- (c) Fit a linear mixed-effects model, with fixed effects for `Treatment` and random effects for the intercept by `Block`, to the `PBIB` data. The call to `lme` would be like that used to fit `fm1Stool` in §1.2.
- (d) Apply `anova` to the fitted model. Is the `Treatment` term significant? Describe the hypothesis being tested.
- (e) Create a plot of the residuals versus the fitted values for this fitted model. This plot is like Figure 1.4. Does this plot indicate greater variance in the residuals at higher levels of the response?
- (f) Create a plot of the standardized residuals versus the fitted values by `Block`. This plot is like Figure 1.8 (p. 21). Does this plot indicate systematic patterns in the residuals by `Block`?

We will discuss the `anova` results for this fitted model in more detail in §2.4.2.

2. The `Oxboys` data described in §3.1 and Appendix A.19 consist of the heights of 26 boys from Oxford, England, each measured on nine different occasions. The structure is similar to that of the `OrthoFem` data of §1.4.
  - (a) Plot the data (using `plot(Oxboys)`) and verify that a simple linear regression model gives a suitable representation of the boys' growth patterns. Do there appear to be significant differences in the individual growth patterns?
  - (b) Fit a simple linear regression model to `height` versus `age` using the `lm` function, ignoring the `Subject` effects. Obtain the boxplots of the residuals by `Subject` with `bwplot(Subject ~ resid(object), Oxboys)`, where `object` should be replaced with the name of the fitted `lm` object. Explain the observed pattern.
  - (c) Use the `lmList` function to fit separate simple linear regression models for each `Subject`, using a call similar to the one used to produce `fm10rthF.lis` in §1.4. Compare the boxplots of the residuals by `Subject` for the `lmList` fit (obtained with `plot(object, Subject ~ resid(.))`, with `object` replaced with the name of the `lmList` object) to those obtained for the `lm` fit. Compare also the residual standard errors from the two fits and comment.
  - (d) Plot the individual confidence intervals on the parameters estimated in the `lmList` fit and verify that both the intercept and the slope vary significantly with `Subject`.

- (e) Use the `lme` function to fit an LME model to the data with random effects for both the intercept and the slope, using a call similar to the one used to obtain `fm1OrthF` in §1.4. Examine the boxplots of the residuals by `Subject`, comparing them to those obtained for the `lm` and `lmList` fits.
- (f) Produce the plot of the standardized residuals versus fitted values (`plot(object)`) and the normal plot of the standardized residuals (`qqnorm(object)`). (In both cases `object` should be replaced with the name of the `lme` object.) Can you identify any departures from the model's assumptions?
- (g) Plot the augmented predictions for the `lme` fit (obtained with `plot(augPred(object))`). Do the linear models for each subject appear adequate?
- (h) Another way of assessing the linear models for each subject is to plot the residuals versus `age` by `Subject` (use `plot(object, resid(.) ~ age | Subject)`, replacing `object` with the name of the `lme` object). Several subjects have a noticeable “scooping” pattern in their residuals, indicating the need for a model with curvature.
- (i) Use the `lmList` function to fit separate quadratic models for each subject. A quadratic model in `age`, as shown in `fm1Pixel` of §1.5, would be fit with `lmList(height ~ age + age^2, Oxbboys)`.
- (j) Examine a plot of the confidence intervals on coefficients from this second `lmList` fit. Are there indications that the coefficients differ between subjects? Are the quadratic coefficients significantly different from zero for some subjects?
- (k) Fit the full mixed-effects model corresponding to the last `lmList` fit. The model will have linear and quadratic terms for `age` in the fixed-effects and the random effects. A simple way to describe this model is `lme(object)` replacing `object` with the name of the `lmList` fit.
- (l) Check residual plots and numerical summaries for this `lme` model. Do there appear to be deficiencies in the fit? Do there appear to be terms in the model that could be eliminated?
3. The LME model used for the `Pixel` data in §1.5 uses random effects for the intercept and the slope at the `Dog` level and a single random effect for the intercept at the `Side` within `Dog` level. We did not discuss there how that random-effects model was chosen. The `lmList` function can be used with multilevel data to investigate which terms in an LME model require random effects.
- (a) Use `lmList` to fit a separate quadratic model in `day` for each `Dog`. Print the fitted object and examine the estimated coefficients.

Can you explain the error message printed in the `lmList` fit? Notice that `lmList` was able to recover from the error and proceed to normal completion.

- (b) Plot the individual confidence intervals for the coefficients in the `lmList` fit. Verify that only the intercept and the linear coefficient seem to vary significantly with `Dog`.
  - (c) Use the `level` argument to `lmList` to fit separate quadratic models in `day` for each `Side` within `Dog` (use `Dog/Side` as the grouping expression and set `level=2`). Print the fitted object using `summary` and explain the missing values (`NA`) for the standard errors of `Dog` 10.
  - (d) Plot the individual confidence intervals for the coefficients in the `lmList` fit by `Side` within `Dog` and verify that there is more variation among the intercepts and the linear coefficients than among the quadratic coefficients.
  - (e) Fit an LME model with random effects for the intercept and the linear term at both levels of grouping. Compare the resulting `lme` fit to the `fm1Pixel` object in §1.5 using `anova`. Which model should be the preferred?
4. The `Alfalfa` data described in Appendix A.1 is another example of a split-plot experiment. The structure is similar to that of the `Oats` data of §1.6: a  $3 \times 4$  full factorial on varieties of alfalfa and date of third cutting is used with 6 blocks each subdivided into 4 plots according to a split-plot arrangement. The whole-plot treatments are given by the varieties and the subplot treatments by the date of third cutting.
- (a) Plot the data (using `plot(Alfalfa)`). Do there appear to be cutting dates that are consistently worse/better than the others? What can you say about the block-to-block variation in the yields?
  - (b) Use `lme` to fit a two-level LME model with grouping structure `Block/Variety`, including a single random intercept for each level of grouping (i.e., `random = ~1 | Block/Variety`). Assume a full factorial structure with main effects and interactions for the fixed effects (i.e., `fixed = Yield ~ Date * Variety`). Use the *treatment* contrasts (`options(constraints = c("contr.treatment", "contr.poly"))`) to get more interpretable coefficients for the fixed effects.
  - (c) Examine the significance of the terms in the model using `anova`, verifying that there are no significant differences between varieties and no significant interactions between varieties and cutting dates.

- (d) Because the data are balanced, a similar ANOVA model can be fit using `aov` and the `Error` function (use `aov(Yield ~ Date * Variety + Error(Block/Variety), Alfalfa)`). Compare the results from the `aov` and `lme` fits, in particular the  $F$ -values and  $p$ -values for testing the terms in the fixed-effects model (these are obtained for the `aov` object using the `summary` function). In this case, because of the balanced structure of the data, the REML fit (obtained with `lme`) and the ANOVA fit (obtained with `aov`) are identical.
- (e) Refit the LME model using fixed effects for `Date` only (a simple way to do this is to use `update(object, Yield ~ Date)`, where `object` should be replaced with name of the previous `lme` object). Print the resulting object using `summary` and investigate the differences between the cutting dates (recall that, for the *treatment* contrasts, the coefficients represent differences with respect to the cutting date labelled `None`). Can you identify a trend in the effect of cutting date on yield?
- (f) Examine the plot of the residuals versus fitted values and the normal plot of the residuals. Can you identify any departures from the LME model's assumptions?