Douglas M. Bates

lme4: Mixed-effects modeling with R

February 24, 2010

Springer

Contents

List of Figures

Chapter 3
Models Incorporating Covariates

In the previous chapter we fit models having different configurations of simple, scalar random effects but always with the same fixed-effects specification of an intercept, or overall mean response, only.

It is common in practice to have several fixed-effects terms involving one or more covariates in the specification of a linear mixed model. Indeed, the purpose of fitting a linear mixed model is often to draw inferences about the effects of the covariates while appropriately accounting for different sources of variability in the responses.

In this chapter we demonstrate how fixed-effects terms are incorporated in a linear mixed model and how inferences about the effects of the covariates are drawn from a fitted linear mixed model.

3.1 Models for the `ergoStool` data

Problems 2.1 and 2.2 in Chap. 2 involve examining the structure of the `ergoStool` data from the `MEMSS` package

```
> str(ergoStool)
```

```
'data.frame':        36 obs. of  3 variables:
 $ effort : num  12 15 12 10 10 14 13 12 7 14 ...
 $ Type   : Factor w/ 4 levels "T1","T2","T3",..: 1 2 3 4 1 2 3 4 1 2 ...
 $ Subject: Factor w/ 9 levels "A","B","C","D",..: 1 1 1 1 2 2 2 2 3 3 ...
```

and plotting these data, as in Fig. 3.1. These data are from an ergometrics experiment where nine subjects evaluated the difficulty to arise from each of four types of stools. The measurements are on the scale of perceived exertion developed by the Swedish physician and researcher Gunnar Borg. Measurements on this scale are in the range 6-20 with lower values indicating less exertion.

From Fig. 3.1 we can see that all nine subjects rated type `T1` or type `T4` as requiring the least exertion and rated type `T2` as requiring the most exertion.

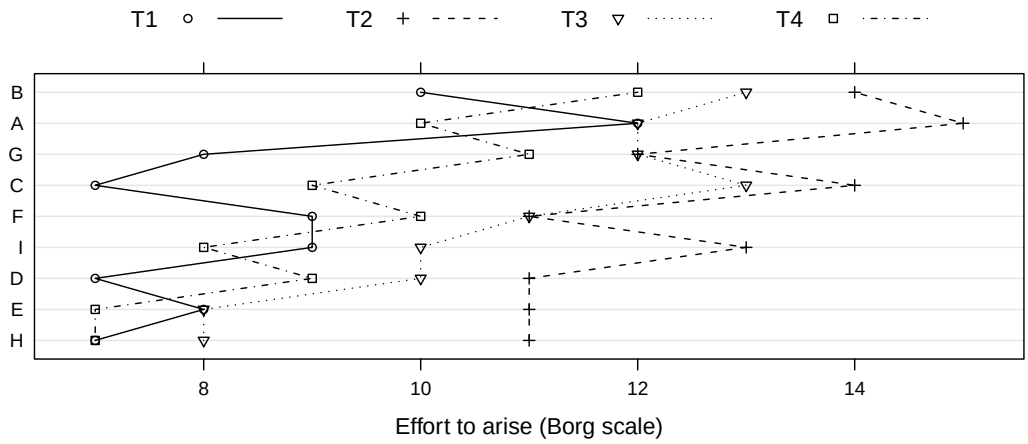

Fig. 3.1 Subjective evaluation of the effort required to arise (on the Borg scale) by 9 subjects, each of whom tried each of four types of stool.

Type `T3` was perceived as requiring comparatively little exertion by some subjects (H and E) and comparatively greater exertion by others (F, C and G).

Problem 2.3 involves fitting and evaluating a model in which the effects of both the `Subject` and the `Type` factors are incorporated as random effects. Such a model may not be appropriate for these data where we wish to make inferences about these particular four stool types. According to the distinction between fixed- and random-effects described in Sect. 1.1, if the levels of the `Type` factor are fixed and reproducible we generally incorporate the factor in the fixed-effects part of the model.

Before doing so, let's review the results of fitting a linear mixed model with random effects for both `Subject` and `Type`.

3.1.1 Random-effects for both `Subject` *and* `Type`

A model with random effects for both `Subject` and `Type` is fit in the same way that we fit such in Chap. 2,

```
> (fm06 <- lmer(effort ~ 1 + (1|Subject) + (1|Type), ergoStool, REML=0))

Linear mixed model fit by maximum likelihood
Formula: effort ~ 1 + (1 | Subject) + (1 | Type)
   Data: ergoStool
   AIC   BIC logLik deviance
 144.0 150.4 -68.01    136.0
Random effects:
 Groups   Name        Variance Std.Dev.
 Subject  (Intercept) 1.7040   1.3054
 Type     (Intercept) 2.2645   1.5048
 Residual             1.2128   1.1013
```

```
Number of obs: 36, groups: Subject, 9; Type, 4

Fixed effects:
            Estimate Std. Error t value
(Intercept)  10.2500     0.8883   11.54
```

from which we determine that the mean effort to arise, across stool types and across subjects, is 10.250 on this scale, with standard deviations of 1.305 for the random-effects for the `Subject` factor and 1.505 for the `Type` factor.

One question we would want to address is whether there are "significant" differences between stool types, taking into account the differences between subjects. We could approach this question by fitting a reduced model, without random effects for `Type`, and comparing this fit to model `fm06` using a likelihood-ratio test.

```
> fm06a <- lmer(effort ~ 1 + (1|Subject), ergoStool, REML=0)
> anova(fm06a, fm06)

Data: ergoStool
Models:
fm06a: effort ~ 1 + (1 | Subject)
fm06: effort ~ 1 + (1 | Subject) + (1 | Type)
      Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
fm06a  3 164.15 168.90 -79.075
fm06   4 144.02 150.36 -68.011 22.128      1  2.551e-06
```

The p-value in this test is very small, indicating that the more complex model, `fm06`, which allows for differences in the effort to arise for the different stool types, provides a significantly better fit to the observed data.

In Sect. 2.2.4 we indicated that, because the constraint on the reduced model, $\sigma_2 = 0$, is on the boundary of the parameter space, the p-value for this likelihood ratio test statistic calculated using a χ^2_1 reference distribution will be conservative. That is, the p-value one would obtain by, say, simulation from the null distribution, would be even smaller than the p-value, 0.0000026, reported by this test, which is already very small.

Thus the evidence against the null hypothesis ($H_0 : \sigma_2 = 0$) and in favor of the alternative, richer model ($H_a : \sigma_2 > 0$) is very strong.

Another way of addressing the question of whether it is reasonable for σ_2 to be zero is to profile `fm06` and examine profile zeta plots (Fig. 3.2) and the corresponding profile pairs plot (Fig. 3.3).

We can see from the profile zeta plot (Fig. 3.2) that both σ_1, the standard deviation of the `Subject` random effects, and, σ_2, the standard deviation of the `Type` random effects, are safely non-zero. We also see that σ_2 is very poorly determined. That is, a 95% profile-based confidence interval on this parameter, obtained as

```
> confint(pr06)[".sig02",]

    2.5 %    97.5 %
0.7925434 3.7958504
```

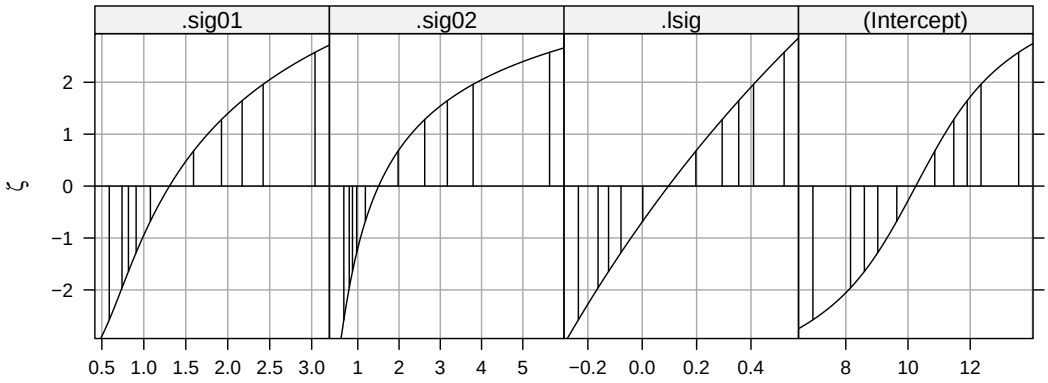

Fig. 3.2 Profile zeta plot for the parameters in model `fm06` fit to the `ergoStool` data

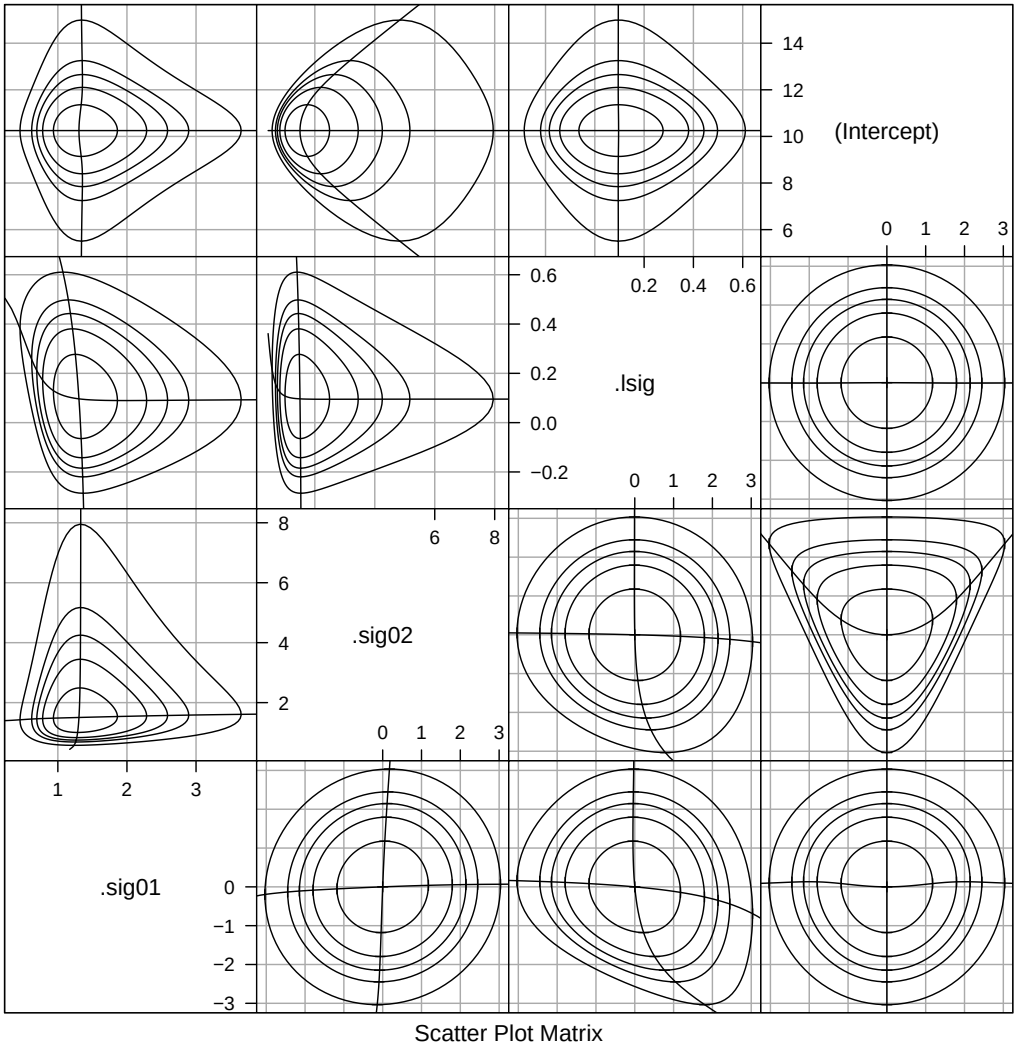

Fig. 3.3 Profile pairs plot for the parameters in model `fm06` fit to the `ergoStool` data

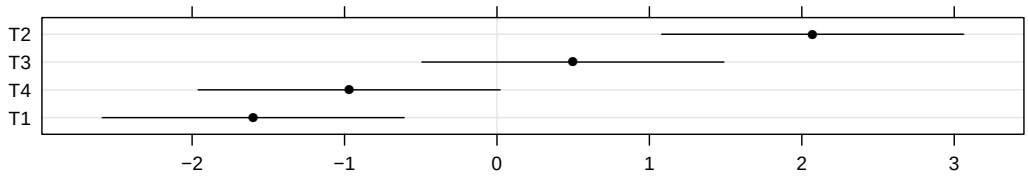

Fig. 3.4 95% prediction intervals on the random effects for `Type` from model `fm06` fit to the `ergoStool` data

is very wide. The upper end point of this 95% confidence interval, 3.796, is more than twice as large as the estimate, $\widehat{\sigma_2} = 1.505$.

A plot of the prediction intervals on the random effects for `Type` (Fig. 3.4) confirms the impression from Fig. 3.1 regarding the stool types. Type `T2` requires the greatest effort and type `T1` requires the least effort. There is considerable overlap of the prediction intervals for types `T1` and `T4` and somewhat less overlap between types `T4` and `T3` and between types `T3` and `T2`.

In an analysis like this we begin by asking if there are any significant differences between the stool types, which we answered for this model by testing the hypothesis $H_0 : \sigma_2 = 0$ versus $H_a : \sigma_2 > 0$. If we reject H_0 in favor of H_a — that is, if we conclude that the more complex model including random effects for `Type` provides a significantly better fit than the simpler model — then usually we want to follow up with the question, "Which stool types are significantly different from each other?". It is possible, though not easy, to formulate an answer to that question from a model fit such as `fm06` in which the stool types are modeled with random effects, but it is more straightforward to address that question when we model the stool types as fixed-effects parameters, which we do next.

3.1.2 Fixed Effects for `Type`*, Random for* `Subject`

To incorporate the `Type` factor in the fixed-effects parameters, instead of as a grouping factor for random effects, we remove the random-effects term, `(1|Type)`, and add `Type` to the fixed-effects specification.

```
> (fm07 <- lmer(effort ~ 1 + Type + (1|Subject), ergoStool, REML = 0))

Linear mixed model fit by maximum likelihood
Formula: effort ~ 1 + Type + (1 | Subject)
   Data: ergoStool
   AIC   BIC logLik deviance
 134.1 143.6 -61.07    122.1
Random effects:
 Groups   Name        Variance Std.Dev.
```

```
 Subject  (Intercept) 1.5782   1.2563
 Residual             1.0761   1.0374
Number of obs: 36, groups: Subject, 9

Fixed effects:
            Estimate Std. Error t value
(Intercept)   8.5556     0.5431  15.754
TypeT2        3.8889     0.4890   7.952
TypeT3        2.2222     0.4890   4.544
TypeT4        0.6667     0.4890   1.363

Correlation of Fixed Effects:
       (Intr) TypeT2 TypeT3
TypeT2 -0.450
TypeT3 -0.450  0.500
TypeT4 -0.450  0.500  0.500
```

It appears that the last three levels of the `Type` factor are now modeled as fixed-effects parameters, in addition to the `(Intercept)` parameter, whose estimate has decreased markedly from that in model `fm06`. Furthermore, the estimates of the fixed-effects parameters labeled `TypeT2`, `TypeT3` and `TypeT4`, while positive, are very much smaller than would be indicated by the average responses for these types.

It turns out, of course, that the fixed-effects parameters generated by a factor covariate do not correspond to the overall mean and the effect for each level of the covariate. Although a model for an experiment such as this is sometimes written in a form like

$$y_{ij} = \mu + \alpha_i + b_j + \varepsilon_{ij}, \quad i = 1,\ldots,4,\ j = 1,\ldots 9 \tag{3.1}$$

where i indexes the stool type and j indexes the subject, the parameters $\{\mu, \alpha_1, \alpha_2, \alpha_3, \alpha_4\}$, representing the overall mean and the effects of each of the stool types, are redundant. Given a set of estimates for these parameters we would not change the predictions from the model if, for example, we added one to μ and subtracted one from all the α's. In statistical terminology we say that this set of parameters is not *estimable* unless we impose some other conditions on them. The estimability condition $\sum_{i=1}^{4} \alpha_i = 0$ is often used in introductory texts.

The approach taken in R is not based on redundant parameters that are subject to estimability conditions. While this approach may initially seem reasonable, in complex models it quickly becomes unnecessarily complex to need to use constrained optimization for parameter estimation. Instead we incorporate the constraints into the parameters that we estimate. That is, we reduce the redundant set of parameters to an estimable set of *contrasts* between the levels of the factors.

3.1.2.1 The default contrasts generated for a factor

Although the particular set of contrasts used for a categorical factor can be controlled by the user, either as a global option for a session (see `?options`) or by the optional `contrasts` argument available in most model-fitting functions, most users do not modify the contrasts, preferring to leave them at the default setting, which is the "treatment" contrasts (`contr.treatment`) for an unordered factor and orthogonal polynomial contrasts (`contr.poly`) for an ordered factor. You can check the current global setting with

```
> getOption("contrasts")
```

```
        unordered           ordered
"contr.treatment"      "contr.poly"
```

Because these were the contrasts in effect when model `fm07` was fit, the particular contrasts used for the `Type` factor, which has four levels, correspond to

```
> contr.treatment(4)
```

```
  2 3 4
1 0 0 0
2 1 0 0
3 0 1 0
4 0 0 1
```

In this display the rows correspond to the levels of the `Type` factor and the columns correspond to the parameters labeled `TypeT2`, `TypeT3` and `TypeT4`.

The values of `Type` in the data frame, whose first few rows are

```
> head(ergoStool)
```

```
  effort Type Subject
1     12   T1       A
2     15   T2       A
3     12   T3       A
4     10   T4       A
5     10   T1       B
6     14   T2       B
```

combined with the contrasts produce the model matrix $\mathbf{X}$, whose first few rows are

```
> head(with(env(fm07), X))
```

```
6 x 4 Matrix of class "dgeMatrix"
     (Intercept) TypeT2 TypeT3 TypeT4
[1,]           1      0      0      0
[2,]           1      1      0      0
[3,]           1      0      1      0
[4,]           1      0      0      1
[5,]           1      0      0      0
[6,]           1      1      0      0
```

We see that the rows of $\mathbf{X}$ for observations on stool type `T1` have zeros in the last three columns; the rows for observations on stool type `T2` have a 1 in the second column and zeros in the last two columns, and so on. As before, the `(Intercept)` column is a column of 1's.

When we evaluate $\mathbf{X}\beta$ in the linear predictor expression, $\mathbf{X}\beta+\mathbf{Zb}$, we take the p elements of the fixed-effects parameter vector, β, whose estimate is

```
> fixef(fm07)
```

```
(Intercept)       TypeT2       TypeT3       TypeT4
  8.5555556    3.8888889    2.2222222    0.6666667
```

and the p elements of a row of the matrix $\mathbf{X}$, multiply corresponding components and sum the resulting products. For example, the fixed-effects predictor for the first observation (stool type `T1`) will be

$$8.5556\times 1+3.8889\times 0+2.2222\times 0+0.6667\times 0=8.5556$$

and the fixed-effects predictor for the second observation (stool type `T2`) will be

$$8.5556\times 1+3.8889\times 1+2.2222\times 0+0.6667\times 0=12.4444$$

We see that the parameter labeled `(Intercept)` is actually the fixed-effects prediction for the first level of `Type` (i.e. level `T1`) and the second parameter, labeled `TypeT2`, is the difference between the fixed-effects prediction for the second level (`T2`) and the first level (`T1`) of the `Type` factor.

Similarly, the fixed-effects predictions for the `T3` and `T4` levels of `Type` are $8.5556+2.2222=10.7778$ and $8.5556+0.6667=9.2222$, respectively, as can be verified from

```
> with(env(fm07), head(as.vector(X %*% fixef)))
```

```
[1]  8.555556 12.444444 10.777778  9.222222  8.555556 12.444444
```

The fact that the parameter labeled `TypeT2` is the difference between the fixed-effects prediction for levels `T2` and `T1` of the `Type` factor is why we refer to the parameters as being generated by *contrasts*. They are formed by contrasting the fixed-effects predictions for some combination of the levels of the factor. In this case the contrast is between levels `T2` and `T1`.

In general, the parameters generated by the "treatment" contrasts (the default for unordered factors) represent differences between the first level of the factor, which is incorporated into the `(Intercept)` parameter, and the subsequent levels. We say that the first level of the factor is the *reference* level and the others are characterized by their shift relative to this reference level.

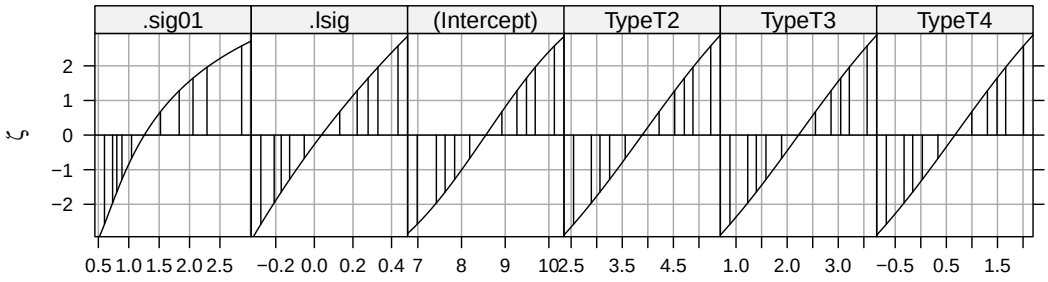

Fig. 3.5 Profile zeta plot for the parameters in model `fm06` fit to the `ergoStool` data

3.1.2.2 Profiling the contrasts

Because some of the contrasts that are of interest to us correspond to fixed-effects parameter values, we can profile the fitted model (Fig. 3.5) and check, say, the confidence intervals on these parameters.

```
> confint(pr07, c("TypeT2", "TypeT3", "TypeT4"))

            2.5 %   97.5 %
TypeT2  2.8953043 4.882473
TypeT3  1.2286377 3.215807
TypeT4 -0.3269179 1.660251
```

According to these intervals, and from what we see from Fig. 3.5, types `T2` and `T3` are significantly different from type T1 (the intervals do not contain zero) but type `T4` is not (the confidence interval on this contrast contains zero).

However, this process must be modified in two ways to provide a suitable answer. The most important modification is to take into account the fact that we are performing *multiple comparisons* simultaneously. We describe what this means and how to accomodate for it in the next subsection. The other problem is that this process only allows us to evaluate contrasts of the reference level, `T1`, with the other levels and the reference level is essentially arbitrary. For completeness we should evaluate all six possible contrasts of pairs of levels.

We can do this by refitting the model with a difference reference level for the `Type` factor and profiling the modified model fit. The `relevel` function allows us to change the reference level of a factor.

```
> pr07a <- profile(lmer(effort ~ 1 + Type + (1|Subject),
+                       within(ergoStool, Type <- relevel(Type, "T2")),
+                       REML = 0))
> pr07b <- profile(lmer(effort ~ 1 + Type + (1|Subject),
```

```
+                       within(ergoStool, Type <- relevel(Type, "T3")),
+                       REML = 0))
```

The other constrasts of interest are

```
> confint(pr07a, c("TypeT3", "TypeT4"))

            2.5 %     97.5 %
TypeT3 -2.660251 -0.6730821
TypeT4 -4.215807 -2.2286377

> confint(pr07b, "TypeT4")

           2.5 %    97.5 %
TypeT4 -2.54914 -0.561971
```

from which would conclude that type `T2` requires significantly greater effort than any of the other types at the 5% level (because none of the 95% confidence intervals on contrasts with `T2` contain zero) and that types `T3` and `T4` are significantly different at the 5% level.

However, we must take into account that we are performing multiple, simulataneous comparisons of levels.

3.1.2.3 Multiple comparisons

In the technical definition of a confidence interval we regard the end points as being random (because they are calculated from the random variable which is the observed data) and the value of the parameter or the contrast of interest as being fixed (because it is determined from the fixed, but unknown, values of the parameters). Thus we speak of the probability that an interval covers the true parameter value rather than the probability that the parameter falls in the interval. The distinction may seem, and probably is, somewhat pedantic. We introduce it here simply to clarify the term "coverage probability" used throughout this section.

We have evaluated six possible pairwise comparisons of the four levels of the `Type` factor. A 95% confidence interval on a particular contrast has, in theory, a 5% probability of failing to cover the true difference. That is, if the difference between two levels was in fact zero, there would still be a 5% probability that a 95% confidence interval on that contrast would not include zero. When we consider the coverage of the six intervals contrasting all possible pairs of stool types we usually have in mind that there should be a 95% probability of all six intervals covering the true, but unknown, differences in effort for the stool types. That is, we think of the coverage probability as applying to the simultaneous coverage of the family of intervals, not to the coverage of one specific interval.

But the intervals calculated in the previous section were based on 95% coverage for each specific interval. In the worst case scenario the family-wise coverage could be as low as $1 - 0.05 * 6 = 0.70$ or 70%. For factors with more

than four levels there are even more possible pairwise comparisons (for k levels there are $k(k-1)/2$ possible pairs) and this worst-case coverage probability is even further from the nominal level of 95%.

Several methods have been developed to compensate for multiple comparisons in the analysis of linear models with fixed effects only. One of the simplest, although somewhat coarse, compensations is the Bonferroni correction where the individual intervals are chosen to have a greater coverage probability in such a way that the "worst-case" probability is the desired level. With six comparisons to get a family-wise coverage probability of 95% the individual intervals are chosen to have coverage of

```
> (covrge <- 1 - 0.05/6)

[1] 0.9916667
```

or a little more than 99%. We can specify this coverage level for the individual intervals to ensure a family-wise coverage of at least 95%.

```
> rbind(confint(pr07, c("TypeT2","TypeT3","TypeT4"), covrge),
+       confint(pr07a, c("TypeT3","TypeT4"), covrge),
+       confint(pr07b, "TypeT4", covrge))

          0.417 %   99.583 %
TypeT2  2.5109497  5.2668280
TypeT3  0.8442830  3.6001613
TypeT4 -0.7112726  2.0446058
TypeT3 -3.0446059 -0.2887275
TypeT4 -4.6001614 -1.8442831
TypeT4 -2.9334948 -0.1776164
```

We again reach the conclusion that the only pair of stool types for which zero is within the confidence interval on the difference in effects is the (T1,T4) pair but, for these intervals, the family-wise coverage of all six intervals is at least 95%.

There are other, perhaps more effective, techniques for adjusting intervals to take into account multiple comparisons. The purpose of this section is to show that the profile-based confidence intervals can be extended to at least the Bonferroni correction.

The easiest way to apply other multiple comparison adjustment methods is to model both the Type and the Subject factors with fixed effects, which we do next.

3.1.3 Fixed Effects for Type *and* Subject

Even though the subjects in this study are chosen as representatives of a population, many statisticians would regard Subject as a *blocking factor* in the experiment and fit a model with fixed-effects for both Type and Subject. A blocking factor is a known source of variability in the response. We are

not interested in the effects of the levels of the blocking factor—we only wish to accomodate for this source of variability when comparing levels of the experimental factor, which is the `Type` factor in this example.

We will discuss the advantages and disadvantages of the fixed- versus random-effects choice for the `Subject` factor at the end of this section. For the moment we proceed to fit the fixed-effects model, for which we could use the `lm` function or the `aov` function. These two functions produce exactly the same model fit but the `aov` function returns an object of class `"aov"` which extends the class `"lm"`, providing more options for examining the fitted model.

```
> summary(fm08 <- aov(effort ~ Subject + Type, ergoStool))

            Df Sum Sq Mean Sq F value    Pr(>F)
Subject      8 66.500  8.3125  6.8662 0.0001061
Type         3 81.194 27.0648 22.3556 3.935e-07
Residuals   24 29.056  1.2106
```

As seen above, the `summary` method for objects of class `"aov"` provides an analysis of variance table. The order in which the terms are listed in the model formula can affect the results in this table, if the data are unbalanced, and we should be cautious to list the terms in the model in the appropriate order, even for a balanced data set like the `ergoStool`. The rule is that blocking factors should precede experimental factors because the contributions of the terms are assessed sequentially. Thus we read the rows in this table as measuring the variability due to the `Subject` factor and due to the `Type` factor after taking into account the `Subject`. We want to assess the experimental factor after having removed the variability due to the blocking factor.

If desired we can assess individual coefficients by applying the summary method for `"lm"` objects, called `summary.lm` to this fitted model. For example, the coefficients table is available as

```
> coef(summary.lm(fm08))

                Estimate Std. Error       t value     Pr(>|t|)
(Intercept)  1.055556e+01  0.6352554  1.661624e+01 1.146367e-14
SubjectB     6.005604e-15  0.7780258  7.719029e-15 1.000000e+00
SubjectC    -1.500000e+00  0.7780258 -1.927957e+00 6.577394e-02
SubjectD    -3.000000e+00  0.7780258 -3.855913e+00 7.577323e-04
SubjectE    -3.750000e+00  0.7780258 -4.819892e+00 6.565314e-05
SubjectF    -2.000000e+00  0.7780258 -2.570609e+00 1.678081e-02
SubjectG    -1.500000e+00  0.7780258 -1.927957e+00 6.577394e-02
SubjectH    -4.000000e+00  0.7780258 -5.141218e+00 2.907651e-05
SubjectI    -2.250000e+00  0.7780258 -2.891935e+00 8.010522e-03
TypeT2       3.888889e+00  0.5186838  7.497610e+00 9.753420e-08
TypeT3       2.222222e+00  0.5186838  4.284348e+00 2.562927e-04
TypeT4       6.666667e-01  0.5186838  1.285304e+00 2.109512e-01
```

but often the individual coefficients are of less interest than the net effect of the variability due to the levels of the factor, as shown in the analysis of variance table. For example, in the summary of the coefficients shown above the `(Intercept)` coefficient is the predicted response for the reference subject

(subject A) on the reference stool type (type T1). Other coefficients generated by the Subject term are the differences from the reference subject to other subjects. It is not clear why we would want to compare all the other subjects to subject A.

One of the multiple comparison methods that we can apply to fm08 is Tukey's Honest Significant Difference (HSD) method

```
> TukeyHSD(fm08, which = "Type")

  Tukey multiple comparisons of means
    95% family-wise confidence level
Fit: aov(formula = effort ~ Subject + Type, data = ergoStool)

$Type
            diff        lwr        upr     p adj
T2-T1  3.8888889  2.4580431  5.3197347 0.0000006
T3-T1  2.2222222  0.7913764  3.6530680 0.0013709
T4-T1  0.6666667 -0.7641791  2.0975125 0.5807508
T3-T2 -1.6666667 -3.0975125 -0.2358209 0.0182037
T4-T2 -3.2222222 -4.6530680 -1.7913764 0.0000115
T4-T3 -1.5555556 -2.9864013 -0.1247098 0.0295822
```

from which we reach essentially the same conclusions as before, although perhaps less arduously.

3.1.4 Fixed Effects Versus Random Effects for Subject

These three analyses provoke the question of whether to use fixed-effects parameters or random effects for the Subject factor. That is, should Subject be treated as a blocking factor or as a sample of levels from a population. If the sole purpose of the experiment is to rate the stool types relative to each other then it may be more convenient to consider Subject as a blocking factor in the experiment and incorporate it as the first term in a fixed-effects model for this completely balanced data set from a designed experiment. Strictly speaking, the inferences about the stool types that we would draw apply to these particular nine subjects only, not to a general population, but in practice it is not too much of a stretch to think of them as applying to a population *unless* we want to predict the score that a general member of the population would give to a particular stool type. We can formulate the prediction but assessing the variability in the prediction is difficult when using fixed-effects models.

Random effects are preferred, perhaps even necessary, if we wish to make inferences that apply to the population of potential users. Also, when we have unbalanced data or large samples, the flexibility of mixed models becomes important in stabilizing the estimation of parameters. The estimation of parameters in fixed-effects models by least squares is somewhat rigid. In

theory it requires that the columns of the model matrix, $\mathbf{X}$, are linearly independent. In practice, it is very difficult to determine if the columns of a large model matrix are indeed linear independent or, equivalently, if the rank of $\mathbf{X}$ is exactly p. The best we can do is determine if the condition number, κ, of $\mathbf{X}$ (the ratio of the largest to smallest singular values — see `?kappa`) is sufficiently small to trust the numerical linear algebra results. One way of thinking of κ is as a "magnification factor" for numerical perturbations. In the worst-case scenario, perturbations of relative size ε in the elements of $\mathbf{y}$ results in perturbations of relative size $\kappa\varepsilon$ in the coefficients $\widehat{\beta}$.

When the number of columns, p, of $\mathbf{X}$ is small the condition number tends to be small.

```
> kappa(fm08, exact = TRUE)
```

```
[1] 10.76924
```

However, when p is large the condition number of $\mathbf{X}$ tends to become very large and evaluation of fixed-effects parameter estimates and their standard errors is an ill-conditioned problem.

Calculations involving the random effects model matrix, $\mathbf{Z}$, are not as sensitive to ill-conditioning. The numerical accuracy is determined by the condition number of the sparse Cholesky factor, $\mathbf{L}_\theta$, defined in (**??**) which is less than the condition number of $\Lambda_\theta^\top\mathbf{Z}^\top\mathbf{Z}\Lambda_\theta$, even when Λ_θ is singular.

```
> kappa.tri(as(as(env(fm06)$L, "sparseMatrix"), "matrix"), exact = TRUE)
```

```
[1] 4.839924
```

The evaluation of $\mathbf{L}_\theta$ is an example of *regularization* methods for solving ill-posed problems or to prevent overfitting.

3.2 Covariates Affecting Mathematics Score Gain

West et al. [2007] provides comparisons of several different software systems for fitting linear mixed models by fitting sample models to different data sets using each of these software systems. The `lmer` function from the `lme4` package is not included in these comparisons because it was still being developed when that book was written.

In this section we will use `lmer` to fit models described in Chap. 4 of West et al. [2007] to data on the gain in mathematics scores for students in a selection of classrooms in several schools.

```
> str(classroom)
```

```
'data.frame':        1190 obs. of  11 variables:
 $ sex     : Factor w/ 2 levels "M","F": 2 1 2 1 1 2 1 1 2 1 ...
 $ minority: Factor w/ 2 levels "N","Y": 2 2 2 2 2 2 2 2 2 2 ...
 $ mathkind: int  448 460 511 449 425 450 452 443 422 480 ...
```

```
 $ mathgain: int  32 109 56 83 53 65 51 66 88 -7 ...
 $ ses     : num  0.46 -0.27 -0.03 -0.38 -0.03 0.76 -0.03 0.2 0.64 0.13 ...
 $ yearstea: num  1 1 1 2 2 2 2 2 2 2 ...
 $ mathknow: num  NA NA NA -0.11 -0.11 -0.11 -0.11 -0.11 -0.11 -0.11 ...
 $ housepov: num  0.082 0.082 0.082 0.082 0.082 0.082 0.082 0.082 0.082 0.082 ..
 $ mathprep: num  2 2 2 3.25 3.25 3.25 3.25 3.25 3.25 3.25 ...
 $ classid : Factor w/ 312 levels "1","2","3","4",..: 160 160 160 217 217 217 ..
 $ schoolid: Factor w/ 107 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...
```

The response modeled in Chap. 4 of West et al. [2007] is `mathgain`, the difference between a student's mathematics score at the end of grade one and at the end of kindergarten. To allow comparisons with the results in West et al. [2007] we will use that response in this section.

There is one observation per student. In the terminology of the multilevel modeling literature the levels of variability correspond to student (level 1), classroom (level 2) and school (level 3) with the assumption that classroom is nested within school. The concept of "levels" can only be applied to models and data sets in which the grouping factors of the random effects form a nested sequence. In crossed or partially crossed configurations there are no clearly defined levels.

At this point we should check if there is implicit nesting. That is, are the levels of the `classid` factor nested within `schoolid` factor. We could simply create the interaction factor to avoid the possibility of implicit nesting but it saves a bit of trouble if we check before doing so

```
> with(classroom, lme4a:::isNested(classid, schoolid))
```

```
[1] TRUE
```

A model with simple, scalar random effects and without any fixed-effects terms (other than the implicit intercept) is called the "unconditional model" in the multilevel modeling literature. We fit it as

```
> (fm09 <- lmer(mathgain ~ (1|classid) + (1|schoolid), classroom))
```

```
Linear mixed model fit by REML
Formula: mathgain ~ (1 | classid) + (1 | schoolid)
   Data: classroom
 REML
11769
Random effects:
 Groups   Name        Variance Std.Dev.
 classid  (Intercept)   99.232  9.9615
 schoolid (Intercept)   77.483  8.8024
 Residual             1028.235 32.0661
Number of obs: 1190, groups: classid, 312; schoolid, 107

Fixed effects:
            Estimate Std. Error t value
(Intercept)   57.427      1.443   39.79
```

The results from this model fit using the REML criterion can be compared to Table 4.6 (page 156) of West et al. [2007].

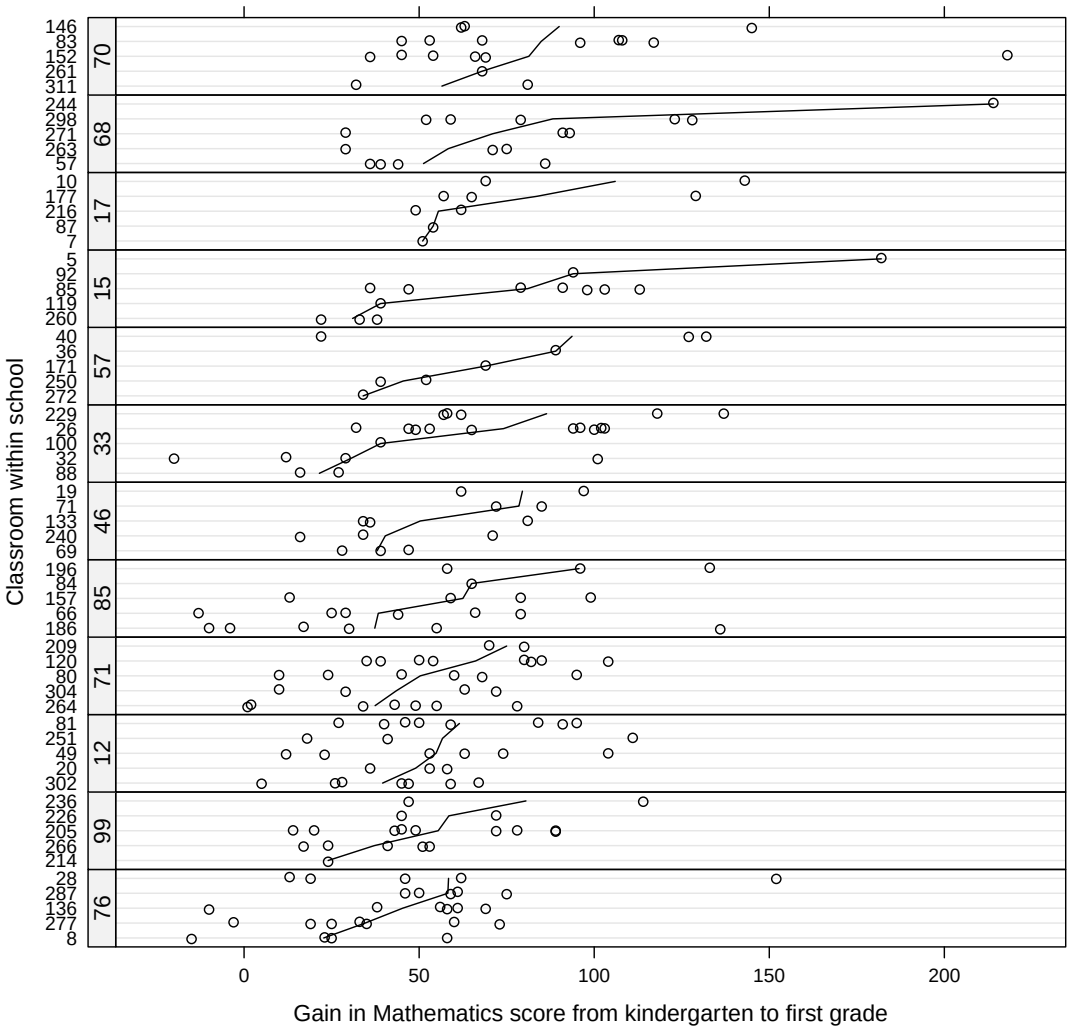

Fig. 3.6 Comparative dotplots of gain in the mathematics scores in classrooms within schools.

It seems that the `housepov` value is a property of the school. We can check this by considering the number of unique combinations of `housepov` and `schoolid` and comparing that to the number of levels of `schoolid`. For safety we check the number of levels of `factor(schoolid)` in case there are unused levels in `schoolid`.

```
> with(classroom, length(levels(factor(schoolid))))

[1] 107

> nrow(unique(subset(classroom, select = c(schoolid, housepov))))

[1] 107
```

In some formulations of multilevel models or hierarchical linear models it is important to associate covariates with different levels in the hierarchy.

```
> (fm7 <- lmer(mathgain ~ 1 + I(mathkind-450) + sex + minority + ses
+              + housepov + (1|classid) + (1|schoolid), classroom))

Linear mixed model fit by REML
Formula: mathgain ~ 1 + I(mathkind - 450) + sex + minority + ses + housepov +      (1 | class
   Data: classroom
 REML
11378
Random effects:
 Groups   Name        Variance Std.Dev.
 classid  (Intercept)  81.555   9.0308
 schoolid (Intercept)  77.761   8.8182
 Residual             734.420  27.1002
Number of obs: 1190, groups: classid, 312; schoolid, 107

Fixed effects:
                   Estimate Std. Error t value
(Intercept)        73.17077    2.80273  26.107
I(mathkind - 450)  -0.47086    0.02228 -21.133
sexF               -1.23459    1.65743  -0.745
minorityY          -7.75587    2.38499  -3.252
ses                 5.23971    1.24497   4.209
housepov          -11.43920    9.93736  -1.151

Correlation of Fixed Effects:
            (Intr) I(-450 sexF   mnrtyY ses
I(mthk-450) -0.233
sexF        -0.279 -0.032
minorityY   -0.492  0.153 -0.015
ses         -0.105 -0.165  0.019  0.144
housepov    -0.555  0.035 -0.009 -0.184  0.078
```

A profile plot of the parameters in model `fm7` is shown in Fig. 3.7

3.3 Rat Brain example

```
> ftable(xtabs(activate ~ animal + treatment + region, ratbrain))

                  region    BST     LS    VDB
animal  treatment
R100797 Basal            458.16 245.04 237.42
        Carbachol        664.72 587.10 726.96
R100997 Basal            479.81 261.19 195.51
        Carbachol        515.29 437.56 604.29
R110597 Basal            462.79 278.33 262.05
        Carbachol        589.25 493.93 621.07
R111097 Basal            366.19 199.31 187.11
        Carbachol        371.71 302.02 449.70
R111397 Basal            375.58 204.85 179.38
        Carbachol        492.58 355.74 459.58
```

Description of the Rat Brain data should go here.

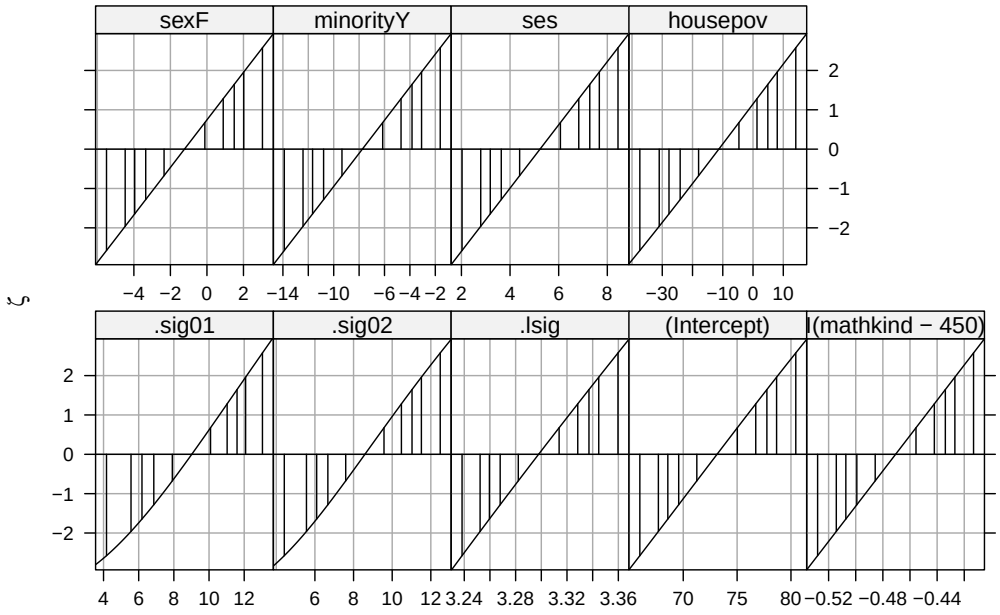

Fig. 3.7 Profile plot of the parameters in model `fm4`.

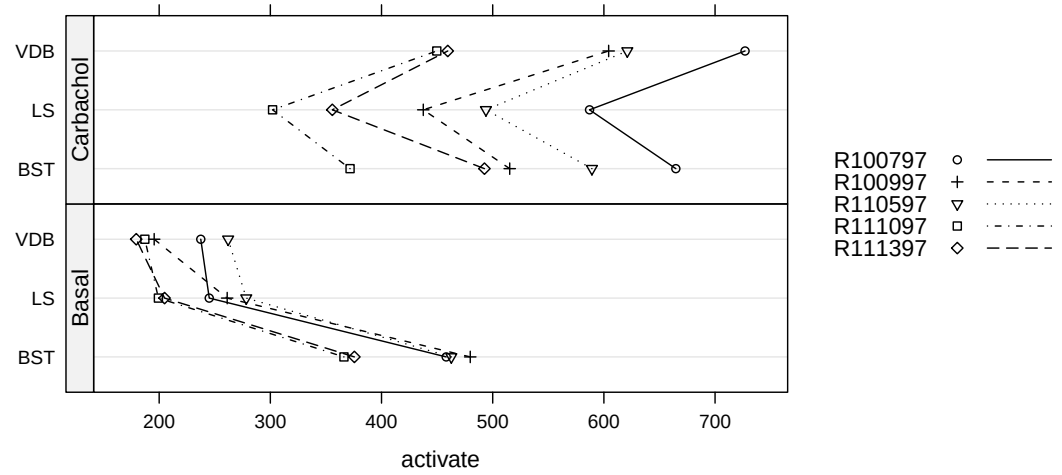

Fig. 3.8 Activation of brain regions in rats

References

Douglas M. Bates and Donald G. Watts. *Nonlinear Regression Analysis and Its Applications*. Wiley, Hoboken, NJ, 1988. ISBN 0-471-81643-4.

Gregory Belenky, Nancy J. Wesensten, David R. Thorne, Maria L. Thomas, Helen C. Sing, Daniel P. Redmond, Michael B. Russo, and Thomas J. Balkin. Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. *Journal of Sleep Research*, 12:1–12, 2003.

G.E.P. Box and G.C. Tiao. *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, MA, 1973.

Bill Cleveland. *Visualizing Data*. Hobart Press, Summit, NJ, 1993.

Owen L. Davies and Peter L. Goldsmith, editors. *Statistical Methods in Research and Production*. Hafner, 4th edition, 1972.

Tim Davis. An approximate minimal degree ordering algorithm. *SIAM J. Matrix Analysis and Applications*, 17(4):886–905, 1996.

Tim Davis. *Direct Methods for Sparse Linear Systems*. SIAM, Philadelphia, PA, 2006.

Friedrich Leisch. Sweave: Dynamic generation of statistical reports using literate data analysis. In Wolfgang Härdle and Bernd Rönz, editors, *Compstat 2002 — Proceedings in Computational Statistics*, pages 575–580. Physica Verlag, Heidelberg, 2002. URL `http://www.stat.uni-muenchen.de/~leisch/Sweave`. ISBN 3-7908-1517-9.

José C. Pinheiro and Douglas M. Bates. *Mixed-effects Models in S and S-PLUS*. Springer, 2000.

J. Rasbash, W. Browne, H. Goldstein, M. Yang, and I. Plewis. *A User's Guide to MLwiN*. Multilevel Models Project, Institute of Education, University of London, London, 2000.

Stephen W. Raudenbush and Anthony S. Bryk. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage, 2nd edition, 2002. ISBN 0-7619-1904-X.

Y. Sakamoto, M. Ishiguro, and G. Kitagawa. *Akaike Information Criterion Statistics*. Reidel, Dordrecht, Holland, 1986.

Deepayan Sarkar. *Lattice: Multivariate Data Visualization with R*. Springer, 2008.

G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.

Brady T. West, Kathleen B. Welch, and Andrzej T. Gałecki. *Linear Mixed Models: A Practical Guide Using Statistical Software*. Chapman and Hall/CRC, Boca Raton, FL, 2007. ISBN 1-58488-480-0.