

Mixed models in R using the lme4 package

Part 4: Longitudinal data, modeling interactions

Douglas Bates

University of Wisconsin - Madison
and R Development Core Team
<Douglas.Bates@R-project.org>

Merck, Rahway, NJ
Sept 23, 2010

1 Longitudinal data: sleepstudy

Outline

- 1 Longitudinal data: sleepstudy
- 2 A model with random effects for intercept and slope

Outline

- 1 Longitudinal data: sleepstudy
- 2 A model with random effects for intercept and slope
- 3 Conditional means

Outline

- 1 Longitudinal data: sleepstudy
- 2 A model with random effects for intercept and slope
- 3 Conditional means
- 4 Conclusions

Outline

- 1 Longitudinal data: sleepstudy
- 2 A model with random effects for intercept and slope
- 3 Conditional means
- 4 Conclusions
- 5 Other forms of interactions

Outline

- 1 Longitudinal data: sleepstudy
- 2 A model with random effects for intercept and slope
- 3 Conditional means
- 4 Conclusions
- 5 Other forms of interactions
- 6 Summary

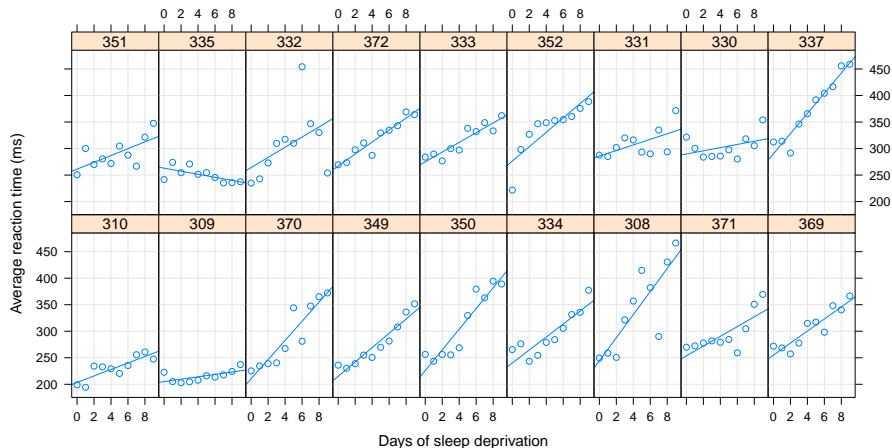
Simple longitudinal data

- *Repeated measures* data consist of measurements of a response (and, perhaps, some covariates) on several *experimental* (or observational) *units*.
- Frequently the experimental (observational) unit is Subject and we will refer to these units as “subjects”. However, the methods described here are not restricted to data on human subjects.
- *Longitudinal* data are repeated measures data in which the observations are taken over time.
- We wish to characterize the response over time within subjects and the variation in the time trends between subjects.
- Frequently we are not as interested in comparing the particular subjects in the study as much as we are interested in modeling the variability in the population from which the subjects were chosen.

Sleep deprivation data

- This laboratory experiment measured the effect of sleep deprivation on cognitive performance.
- There were 18 subjects, chosen from the population of interest (long-distance truck drivers), in the 10 day trial. These subjects were restricted to 3 hours sleep per night during the trial.
- On each day of the trial each subject's reaction time was measured. The reaction time shown here is the average of several measurements.
- These data are *balanced* in that each subject is measured the same number of times and on the same occasions.

Reaction time versus days by subject



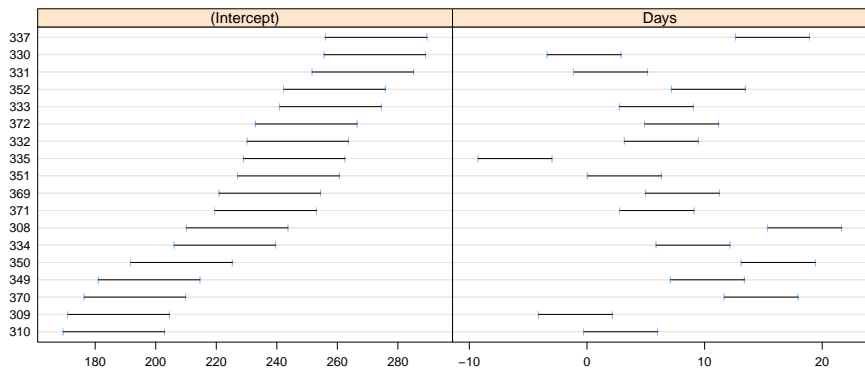
Comments on the sleep data plot

- The plot is a “trellis” or “lattice” plot where the data for each subject are presented in a separate panel. The axes are consistent across panels so we may compare patterns across subjects.
- A reference line fit by simple linear regression to the panel’s data has been added to each panel.
- The aspect ratio of the panels has been adjusted so that a typical reference line lies about 45° on the page. We have the greatest sensitivity in checking for differences in slopes when the lines are near $\pm 45^\circ$ on the page.
- The panels have been ordered not by subject number (which is essentially a random order) but according to increasing intercept for the simple linear regression. If the slopes and the intercepts are highly correlated we should see a pattern across the panels in the slopes.

Assessing the linear fits

- In most cases a simple linear regression provides an adequate fit to the within-subject data.
- Patterns for some subjects (e.g. 350, 352 and 371) deviate from linearity but the deviations are neither widespread nor consistent in form.
- There is considerable variation in the intercept (estimated reaction time without sleep deprivation) across subjects – 200 ms. up to 300 ms. – and in the slope (increase in reaction time per day of sleep deprivation) – 0 ms./day up to 20 ms./day.
- We can examine this variation further by plotting confidence intervals for these intercepts and slopes. Because we use a pooled variance estimate and have balanced data, the intervals have identical widths.
- We again order the subjects by increasing intercept so we can check for relationships between slopes and intercepts.

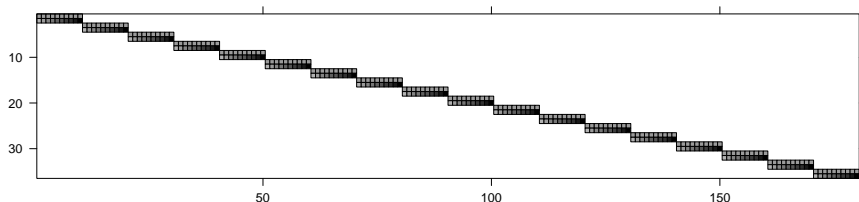
95% conf int on within-subject intercept and slope



These intervals reinforce our earlier impressions of considerable variability between subjects in both intercept and slope but little evidence of a relationship between intercept and slope.

A preliminary mixed-effects model

- We begin with a linear mixed model in which the fixed effects $[\beta_1, \beta_2]^T$ are the representative intercept and slope for the population and the random effects $\mathbf{b}_i = [b_{i1}, b_{i2}]^T, i = 1, \dots, 18$ are the deviations in intercept and slope associated with subject i .
- The random effects vector, \mathbf{b} , consists of the 18 intercept effects followed by the 18 slope effects.



Fitting the model

```
> (fm1 <- lmer(Reaction ~ Days + (Days|Subject), sleepstudy))
```

```
Linear mixed model fit by REML ['merMod']
```

```
Formula: Reaction ~ Days + (Days | Subject)
```

```
Data: sleepstudy
```

```
REML criterion at convergence: 1743.628
```

```
Random effects:
```

| Groups | Name | Variance | Std.Dev. | Corr |
|----------|-------------|----------|----------|-------|
| Subject | (Intercept) | 612.09 | 24.740 | |
| | Days | 35.07 | 5.922 | 0.066 |
| Residual | | 654.94 | 25.592 | |

```
Number of obs: 180, groups: Subject, 18
```

```
Fixed effects:
```

| | Estimate | Std. Error | t value |
|-------------|----------|------------|---------|
| (Intercept) | 251.405 | 6.825 | 36.84 |
| Days | 10.467 | 1.546 | 6.77 |

```
Correlation of Fixed Effects:
```

```
(Intr)
```

```
Days -0.138
```

Terms and matrices

- The term `Days` in the formula generates a model matrix \mathbf{X} with two columns, the intercept column and the numeric `Days` column. (The intercept is included unless suppressed.)
- The term `(Days|Subject)` generates a vector-valued random effect (intercept and slope) for each of the 18 levels of the `Subject` factor.

A model with uncorrelated random effects

- The data plots gave little indication of a systematic relationship between a subject's random effect for slope and his/her random effect for the intercept. Also, the estimated correlation is quite small.
- We should consider a model with uncorrelated random effects. To express this we use two random-effects terms with the same grouping factor and different left-hand sides. In the formula for an `lmer` model, distinct random effects terms are modeled as being independent. Thus we specify the model with two distinct random effects terms, each of which has `Subject` as the grouping factor. The model matrix for one term is intercept only (1) and for the other term is the column for `Days` only, which can be written `0+Days`. (The expression `Days` generates a column for `Days` and an intercept. To suppress the intercept we add `0+` to the expression; `-1` also works.)

A mixed-effects model with independent random effects

Linear mixed model fit by REML [`'merMod'`]

Formula: `Reaction ~ Days + (1 | Subject) + (0 + Days | Subject)`

Data: `sleepstudy`

REML criterion at convergence: 1743.669

Random effects:

| Groups | Name | Variance | Std.Dev. |
|---------|-------------|----------|----------|
| Subject | (Intercept) | 627.57 | 25.051 |
| Subject | Days | 35.86 | 5.988 |
| | Residual | 653.58 | 25.565 |

Number of obs: 180, groups: Subject, 18

Fixed effects:

| | Estimate | Std. Error | t value |
|-------------|----------|------------|---------|
| (Intercept) | 251.405 | 6.885 | 36.51 |
| Days | 10.467 | 1.560 | 6.71 |

Correlation of Fixed Effects:

(Intr)

Days -0.184

Comparing the models

- Model `fm1` contains model `fm2` in the sense that if the parameter values for model `fm1` were constrained so as to force the correlation, and hence the covariance, to be zero, and the model were re-fit, we would get model `fm2`.
- The value 0, to which the correlation is constrained, is not on the boundary of the allowable parameter values.
- In these circumstances a likelihood ratio test and a reference distribution of a χ^2 on 1 degree of freedom is suitable.

```
> anova(fm2, fm1)
```

```
Data: sleepstudy
```

```
Models:
```

```
fm2: Reaction ~ Days + (1 | Subject) + (0 + Days | Subject)
```

```
fm1: Reaction ~ Days + (Days | Subject)
```

| | Df | AIC | BIC | logLik | deviance | Chisq | Chi | Df | Pr(>Chisq) |
|-----|----|--------|--------|---------|----------|--------|-----|----|------------|
| fm2 | 5 | 1762.0 | 1778.0 | -876.00 | 1752.0 | | | | |
| fm1 | 6 | 1763.9 | 1783.1 | -875.97 | 1751.9 | 0.0639 | 1 | | 0.8004 |

Conclusions from the likelihood ratio test

- Because the large p-value indicates that we would not reject f_{m2} in favor of f_{m1} , we prefer the more parsimonious f_{m2} .
- This conclusion is consistent with the AIC (Akaike's Information Criterion) and the BIC (Bayesian Information Criterion) values for which "smaller is better".
- We can also use a Bayesian approach, where we regard the parameters as themselves being random variables, is assessing the values of such parameters. A currently popular Bayesian method is to use sequential sampling from the conditional distribution of subsets of the parameters, given the data and the values of the other parameters. The general technique is called *Markov chain Monte Carlo* sampling.
- We will expand on the use of likelihood-ratio tests in the next section.

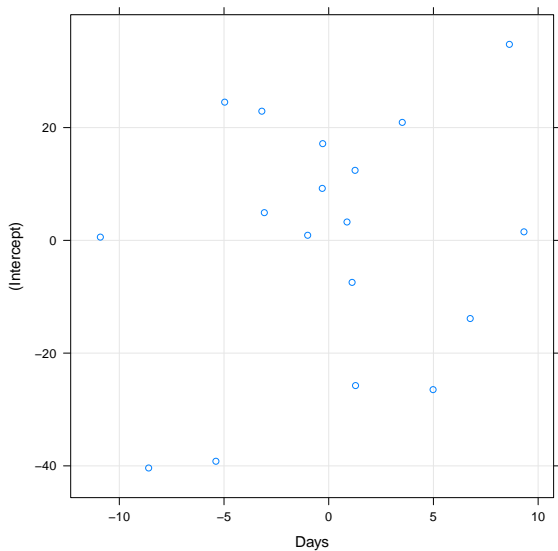
Conditional means of the random effects

```
> (rr2 <- ranef(fm2))
```

```
$Subject
```

| | (Intercept) | Days |
|-----|-------------|-------------|
| 308 | 1.5126970 | 9.3234891 |
| 309 | -40.3738989 | -8.5991690 |
| 310 | -39.1810430 | -5.3877904 |
| 330 | 24.5189049 | -4.9686457 |
| 331 | 22.9144338 | -3.1939348 |
| 332 | 9.2219740 | -0.3084936 |
| 333 | 17.1561217 | -0.2872073 |
| 334 | -7.4517336 | 1.1159900 |
| 335 | 0.5787245 | -10.9059661 |
| 337 | 34.7679297 | 8.6276159 |
| 349 | -25.7543244 | 1.2806878 |
| 350 | -13.8650351 | 6.7564004 |
| 351 | 4.9159801 | -3.0751329 |
| 352 | 20.9290434 | 3.5122096 |
| 369 | 3.2586475 | 0.8730507 |
| 370 | -26.4758270 | 4.9837864 |
| 371 | 0.9056475 | -1.0052929 |

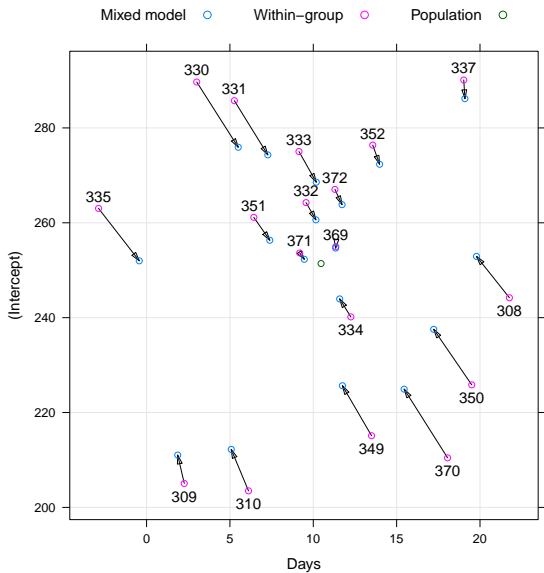
Scatterplot of the conditional means



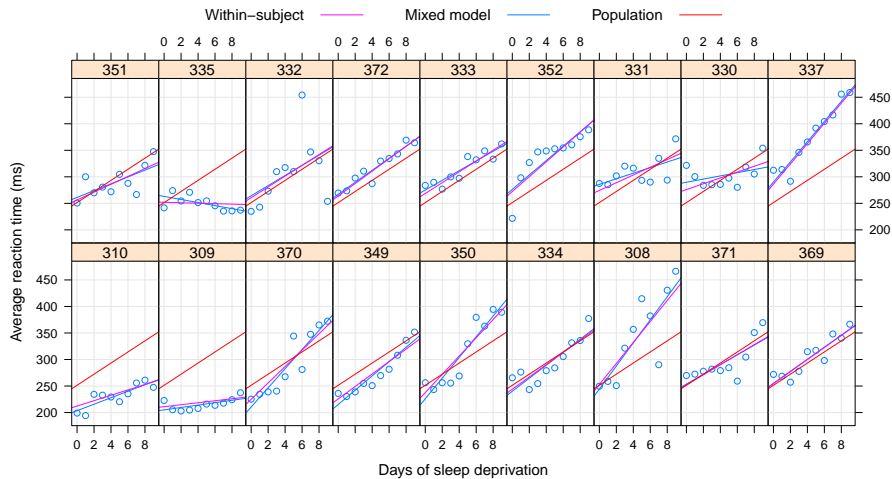
Comparing within-subject coefficients

- For this model we can combine the conditional means of the random effects and the estimates of the fixed effects to get conditional means of the within-subject coefficients.
- These conditional means will be “shrunk” towards the fixed-effects estimates relative to the estimated coefficients from each subject’s data. John Tukey called this “borrowing strength” between subjects.
- Plotting the shrinkage of the within-subject coefficients shows that some of the coefficients are considerably shrunk toward the fixed-effects estimates.
- However, comparing the within-group and mixed model fitted lines shows that large changes in coefficients occur in the noisy data. Precisely estimated within-group coefficients are not changed substantially.

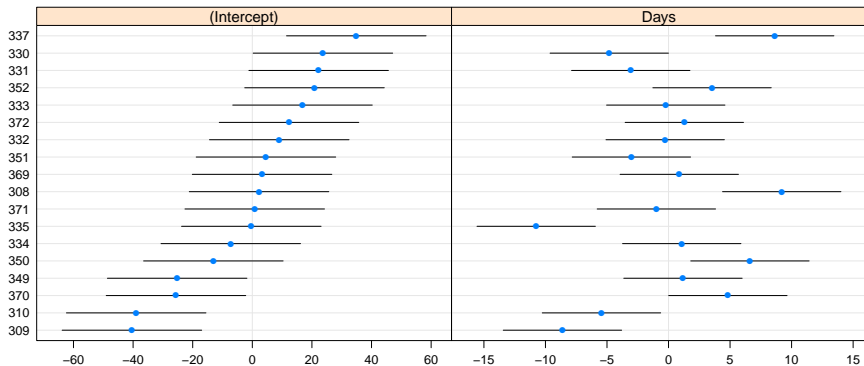
Estimated within-group coefficients and BLUPs



Observed and fitted



Plot of prediction intervals for the random effects



Each set of prediction intervals have constant width because of the balance in the experiment.

Conclusions from the example

- Carefully plotting the data is enormously helpful in formulating the model.
- It is relatively easy to fit and evaluate models to data like these, from a balanced designed experiment.
- We consider two models with random effects for the slope and the intercept of the response w.r.t. time by subject. The models differ in whether the (marginal) correlation of the vector of random effects per subject is allowed to be nonzero.
- The “estimates” (actually, the conditional means) of the random effects can be considered as penalized estimates of these parameters in that they are shrunk towards the origin.
- Most of the prediction intervals for the random effects overlap zero.

Random slopes and interactions

- In the `sleepstudy` model fits we allowed for random effects for Days by Subject.
- These random effects can be considered as an interaction between the fixed-effects covariate Days and the random-effects factor Subject.
- When we have both fixed-levels categorical covariates and random-levels categorical covariates we have many different ways in which interactions can be expressed.
- Often the wide range of options provides “enough rope to hang yourself” in the sense that it is very easy to create an overly-complex model.

The Multilocation data set

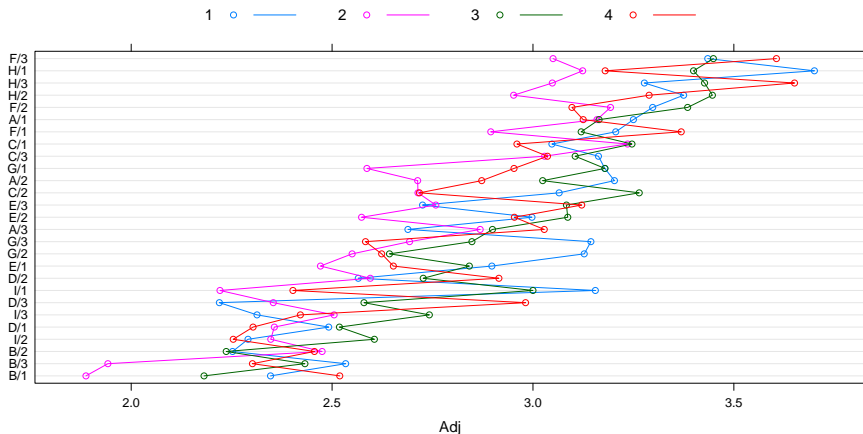
- Data from a multi-location trial of several treatments are described in section 2.8 of Littell, Milliken, Stroup and Wolfinger (1996) **SAS System for Mixed Models** and are available as Multilocation in package SASmixed.
- Littell et al. don't cite the source of the data. Apparently Adj is an adjusted response of some sort for 4 different treatments applied at each of 3 blocks in each of 9 locations. Because Block is implicitly nested in Location, the Grp interaction variable was created.

```
> str(Multilocation)
```

```
'data.frame': 108 obs. of 7 variables:
```

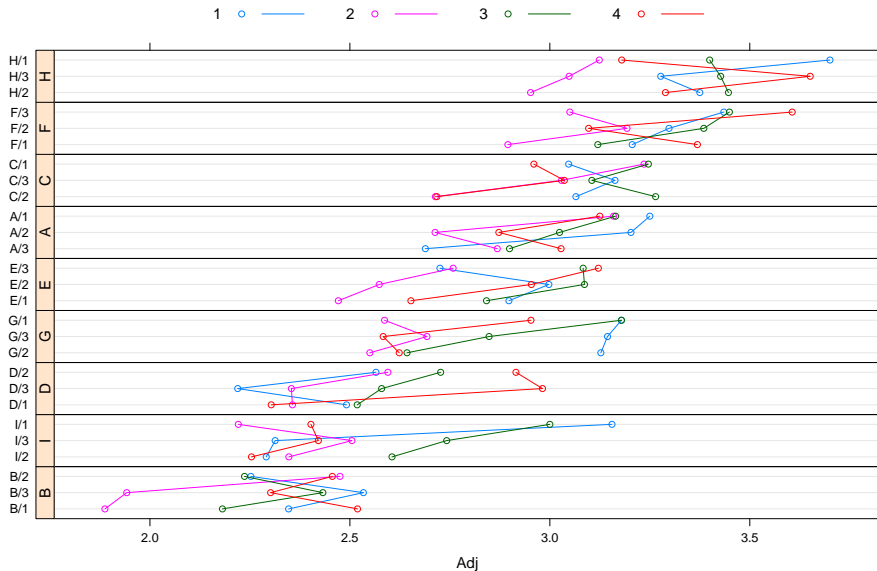
```
$ obs      : num  3 4 6 7 9 10 12 16 19 20 ...
$ Location: Factor w/ 9 levels "A","B","C","D",...: 1 1 1 1 1 1..
$ Block   : Factor w/ 3 levels "1","2","3": 1 1 1 1 2 2 2 2 3 ..
$ Trt     : Factor w/ 4 levels "1","2","3","4": 3 4 2 1 2 1 3 ..
$ Adj     : num  3.16 3.12 3.16 3.25 2.71 ...
$ Fe      : num  7.1 6.68 6.83 6.53 8.25 ...
$ Grp     : Factor w/ 27 levels "A/1","A/2","A/3",...: 1 1 1 1 ..
```

Response by Grp and Trt



- From this one plot (Littell et al. do not provide any plots but instead immediately jump into fitting several “cookie-cutter” models) we see differences between locations, not as much between blocks within location, and treatment 2 providing a lower adjusted response.

Response by Block and Trt within Location



Fixed-levels categorical covariates and “contrasts”

- In this experiment we are interested in comparing the effectiveness of these four levels of `Trt`.
- That is, the levels of `Trt` are fixed levels and we should incorporate them in the fixed-effects part of the model.
- Unlike the situation with random effects, we cannot separately estimate “effects” for each level of a categorical covariate in the fixed-effects and an overall intercept term.
- We could suppress the intercept term but even then we still encounter redundancies in effects for each level when we have more than one categorical covariate in the fixed-effects.
- Because of this we estimate coefficients for $k - 1$ “contrasts” associated with the k levels of a factor.
- The default contrasts (called `contr.treatment`) measure changes relative to a reference level which is the first level of the factor. Other contrasts can be used when particular comparisons are of interest.

A simple model for Trt controlling for Grp

```
> print(fm3 <- lmer(Adj ~ Trt + (1|Grp), Multilocation), corr=)
```

```
Linear mixed model fit by REML ['merMod']
```

```
Formula: Adj ~ Trt + (1 | Grp)
```

```
Data: Multilocation
```

```
REML criterion at convergence: 31.5057
```

```
Random effects:
```

| Groups | Name | Variance | Std.Dev. |
|----------|-------------|----------|----------|
| Grp | (Intercept) | 0.11092 | 0.3331 |
| Residual | | 0.03672 | 0.1916 |

```
Number of obs: 108, groups: Grp, 27
```

```
Fixed effects:
```

| | Estimate | Std. Error | t value |
|-------------|----------|------------|---------|
| (Intercept) | 2.92401 | 0.07395 | 39.54 |
| Trt2 | -0.24637 | 0.05215 | -4.72 |
| Trt3 | 0.02544 | 0.05215 | 0.49 |
| Trt4 | -0.05834 | 0.05215 | -1.12 |

Interpretation of the results

- We see that the variability between the Location/Block combinations (levels of Grp) is greater than the residual variability, indicating the importance of controlling for it.
- The contrast between levels 2 and 1 of Trt, labeled Trt2 is the greatest difference and apparently significant.
- If we wish to evaluate the “significance” of the levels of Trt as a group, however, we should fit the trivial model and perform a LRT.

```
> fm4 <- lmer(Adj ~ 1 + (1|Grp), Multilocation)
> anova(fm4, fm3)
```

Data: Multilocation

Models:

fm4: Adj ~ 1 + (1 | Grp)

fm3: Adj ~ Trt + (1 | Grp)

| | Df | AIC | BIC | logLik | deviance | Chisq | Chi | Df | Pr(>Chisq) |
|-----|----|--------|--------|----------|----------|-------|-----|----|------------|
| fm4 | 3 | 49.731 | 57.777 | -21.8654 | 43.731 | | | | |
| fm3 | 6 | 26.951 | 43.044 | -7.4756 | 14.951 | 28.78 | 3 | | 2.491e-06 |

Location as a fixed-effect

- We have seen that Location has a substantial effect on Adj. If we are interested in these specific 9 locations we could incorporate them as fixed-effects parameters.
- Instead of examining 8 coefficients separately we will consider their cumulative effect using the single-argument form of anova.

```
> anova(fm5 <- lmer(Adj ~ Location + Trt + (1|Grp), Multilocat
```

Analysis of Variance Table

| | Df | Sum Sq | Mean Sq | F value |
|----------|----|--------|---------|---------|
| Location | 8 | 7.3768 | 0.92210 | 25.115 |
| Trt | 3 | 1.2217 | 0.40725 | 11.092 |

An interaction between fixed-effects factors

- We could ask if there is an interaction between the levels of Trt and those of Location considered as fixed effects.

```
> anova(fm6 <- lmer(Adj ~ Location*Trt + (1|Grp), Multilocation
```

Analysis of Variance Table

| | Df | Sum Sq | Mean Sq | F value |
|--------------|----|--------|---------|---------|
| Location | 8 | 6.9475 | 0.86843 | 25.1147 |
| Trt | 3 | 1.2217 | 0.40725 | 11.7774 |
| Location:Trt | 24 | 0.9966 | 0.04152 | 1.2008 |

```
> anova(fm5, fm6)
```

Data: Multilocation

Models:

fm5: Adj ~ Location + Trt + (1 | Grp)

fm6: Adj ~ Location * Trt + (1 | Grp)

| | Df | AIC | BIC | logLik | deviance | Chisq | Chi Df | Pr(>Chisq) |
|-----|----|---------|--------|--------|----------|--------|--------|------------|
| fm5 | 14 | -24.504 | 13.046 | 26.252 | -52.504 | | | |
| fm6 | 38 | -11.146 | 90.775 | 43.573 | -87.146 | 34.642 | 24 | 0.07388 |

Considering levels of Location as random effects

```
> print(fm7 <- lmer(Adj ~ Trt + (1|Location) + (1|Grp), Multiloc)
```

Linear mixed model fit by REML ['merMod']

Formula: Adj ~ Trt + (1 | Location) + (1 | Grp)

Data: Multilocation

REML criterion at convergence: 1.8978

Random effects:

| Groups | Name | Variance | Std.Dev. |
|----------|-------------|----------|----------|
| Grp | (Intercept) | 0.005085 | 0.07131 |
| Location | (Intercept) | 0.114657 | 0.33861 |
| Residual | | 0.036715 | 0.19161 |

Number of obs: 108, groups: Grp, 27; Location, 9

Fixed effects:

| | Estimate | Std. Error | t value |
|-------------|----------|------------|---------|
| (Intercept) | 2.92401 | 0.11953 | 24.462 |
| Trt2 | -0.24637 | 0.05215 | -4.724 |
| Trt3 | 0.02544 | 0.05215 | 0.488 |
| Trt4 | -0.05834 | 0.05215 | -1.119 |

Is Grp needed in addition to Location?

- At this point we may want to check whether the random effect for Block within Location is needed in addition to the random effect for Location.

```
> fm8 <- lmer(Adj ~ Trt + (1|Location), Multilocation)
> anova(fm8, fm7)
```

Data: Multilocation

Models:

fm8: Adj ~ Trt + (1 | Location)

fm7: Adj ~ Trt + (1 | Location) + (1 | Grp)

| | Df | AIC | BIC | logLik | deviance | Chisq | Chi | Df | Pr(>Chisq) |
|-----|----|---------|--------|--------|----------|--------|-----|----|------------|
| fm8 | 6 | 0.25442 | 16.347 | 5.8728 | -11.746 | | | | |
| fm7 | 7 | 0.39496 | 19.170 | 6.8025 | -13.605 | 1.8595 | | 1 | 0.1727 |

- Apparently not, but we may want to revisit this issue after checking for interactions.

Ways of modeling random/fixed interactions

- There are two ways we can model the interaction between a fixed-levels factor (`Trt`) and a random-levels factor (`Location`, as we are currently viewing this factor).
- The first, and generally preferable, way is to incorporate a simple scalar random-effects term with the interaction as the grouping factor.
- The second, more complex, way is to use vector-valued random effects for the random-levels factor. We must be careful when using this approach because it often produces a degenerate model, but not always obviously degenerate.

Scalar random effects for interaction

```
> (fm9 <- lmer(Adj ~ Trt + (1|Trt:Location) + (1|Location), Mu
```

```
Linear mixed model fit by maximum likelihood ['merMod']
```

```
Formula: Adj ~ Trt + (1 | Trt:Location) + (1 | Location)
```

```
Data: Multilocation
```

```
      AIC      BIC  logLik deviance  
2.2544 21.0293  5.8728 -11.7456
```

```
Random effects:
```

| Groups | Name | Variance | Std.Dev. |
|--------------|-------------|-----------|-----------|
| Trt:Location | (Intercept) | 1.962e-20 | 1.401e-10 |
| Location | (Intercept) | 1.029e-01 | 3.207e-01 |
| Residual | | 3.930e-02 | 1.982e-01 |

```
Number of obs: 108, groups: Trt:Location, 36; Location, 9
```

```
Fixed effects:
```

| | Estimate | Std. Error | t value |
|-------------|----------|------------|---------|
| (Intercept) | 2.92401 | 0.11351 | 25.759 |
| Trt2 | -0.24637 | 0.05396 | -4.566 |
| Trt3 | 0.02544 | 0.05396 | 0.472 |
| Trt4 | -0.05834 | 0.05396 | -1.081 |

Both interaction and Block-level random effects

```
> (fm10 <- update(fm9, . ~ . + (1|Grp)))
```

```
Linear mixed model fit by maximum likelihood ['merMod']
```

```
Formula: Adj ~ Trt + (1 | Trt:Location) + (1 | Location) + (1 | Grp)
```

```
Data: Multilocation
```

| AIC | BIC | logLik | deviance |
|--------|---------|--------|----------|
| 2.3564 | 23.8134 | 6.8218 | -13.6436 |

```
Random effects:
```

| Groups | Name | Variance | Std.Dev. |
|--------------|-------------|-----------|----------|
| Trt:Location | (Intercept) | 0.0007769 | 0.02787 |
| Grp | (Intercept) | 0.0056193 | 0.07496 |
| Location | (Intercept) | 0.1011950 | 0.31811 |
| Residual | | 0.0345787 | 0.18595 |

```
Number of obs: 108, groups: Trt:Location, 36; Grp, 27; Location, 9
```

```
Fixed effects:
```

| | Estimate | Std. Error | t value |
|-------------|----------|------------|---------|
| (Intercept) | 2.92401 | 0.11322 | 25.826 |
| Trt2 | -0.24637 | 0.05229 | -4.712 |
| Trt3 | 0.02544 | 0.05229 | 0.487 |
| Trt4 | -0.05834 | 0.05229 | -1.116 |

Scalar interaction random effects are still not significant

```
> anova(fm10, fm8)
```

Data: Multilocation

Models:

fm8: Adj ~ Trt + (1 | Location)

fm10: Adj ~ Trt + (1 | Trt:Location) + (1 | Location) + (1 | Grp)

| | Df | AIC | BIC | logLik | deviance | Chisq | Chi | Df | Pr(>Chisq) |
|------|----|---------|--------|--------|----------|-------|-----|----|------------|
| fm8 | 6 | 0.25442 | 16.347 | 5.8728 | -11.746 | | | | |
| fm10 | 8 | 2.35640 | 23.813 | 6.8218 | -13.644 | 1.898 | | 2 | 0.3871 |

- We have switched to ML fits because we are comparing models using `anova`. In a comparative `anova` any REML fits are refit as ML before comparison so we start with the ML fits.
- In model `fm9` the estimated variance for the scalar interaction random effects was exactly zero in the ML fit. In `fm10` the estimate is positive but still not significant.

Vector-valued random effects

- An alternative formulation for an interaction between `Trt` and `Location` (viewed as a random-levels factor) is to use vector-valued random effects.
- We have used a similar construct in model `fm1` with vector-valued random effects (intercept and slope) for each level of `Subject`.
- One way to fit such a model is

```
> fm11 <- lmer(Adj ~ Trt + (Trt|Location) + (1|Grp), Multilevel)
```

but interpretation is easier when fit as

```
> fm11 <- lmer(Adj ~ Trt + (0+Trt|Location) + (1|Grp), Multilevel)
```

Examining correlation of random effects

- The random effects summary for fm11

```
AIC      BIC      logLik deviance
15.8244  58.7385   8.0878 -16.1756
```

Random effects:

```
Groups   Name          Variance Std.Dev.  Corr
Grp      (Intercept)  0.006352 0.0797
Location Trt1          0.119330 0.3454
          Trt2          0.093347 0.3055   0.984
          Trt3          0.104075 0.3226   0.994 0.996
          Trt4          0.099934 0.3161   0.921 0.967 0.941
Residual          0.031647 0.1779
```

Number of obs: 108, groups: Grp, 27; Location, 9

shows very high correlations between the random effects for the levels of Trt within each level of Location.

- Such a situation may pass by unnoticed if estimates of variances and covariances are all that is reported.
- In this case (and many other similar cases) the variance-covariance matrix of the vector-valued random effects is effectively singular.

Singular variance-covariance for random effects

- When we incorporate too many fixed-effects terms in a model we usually find out because the standard errors become very large.
- For random effects terms, especially those that are vector-valued, overparameterization is sometimes more difficult to detect.
- The REML and ML criteria for mixed-effects models seek to balance the complexity of the model versus the fidelity of the fitted values to the observed responses.
- The way “complexity” is measured in this case, a model with a singular variance-covariance matrix for the random effects is considered a good thing - it is optimally simple.
- When we have only scalar random-effects terms singularity means that one of the variance components must be exactly zero (and “near singularity” means very close to zero).

Detecting singular random effects

- The `Lambda` slot in a `merMod` object is the triangular factor of the variance-covariance matrix.
- We can directly assess its condition number using the `kappa` (condition number) or `rcond` (reciprocal condition number) functions. Large condition numbers are bad.
- We do need to be cautious when we have a large number of levels for the grouping factors because `Lambda` will be **very** large (but also very sparse). At present the `kappa` and `rcond` functions transform the sparse matrix to a dense matrix, which could take a very long time.

```
> kappa(fm11@re@Lambda)
```

```
[1] Inf
```

```
> rcond(fm11@re@Lambda)
```

```
[1] 0
```

Using verbose model fits

- An alternative, which is recommended whenever you have doubts about a model fit, is to use `verbose=TRUE` (the lines don't wrap and we miss the interesting part here).

```
npt = 17 , n = 11
rhobeg = 0.2 , rhoend = 2e-07
  0.020:  41:      -9.00509;0.533967  1.75302  0.993757  1.29209  1.11595  1.
  0.0020: 122:     -16.1171;0.446798  1.91490  1.66791  1.77935  1.62279  0.
  0.00020: 223:     -16.1600;0.444435  1.93613  1.69396  1.80554  1.63429  0.
  2.0e-05: 417:     -16.1756;0.448007  1.94192  1.69057  1.80287  1.63609  0.
  2.0e-06: 527:     -16.1756;0.447998  1.94182  1.69048  1.80284  1.63602  0.
  2.0e-07: 555:     -16.1756;0.447997  1.94182  1.69047  1.80284  1.63601  0.
```

At return

```
575:      -16.175574: 0.447997  1.94182  1.69047  1.80284  1.63601  0.303262  0
```

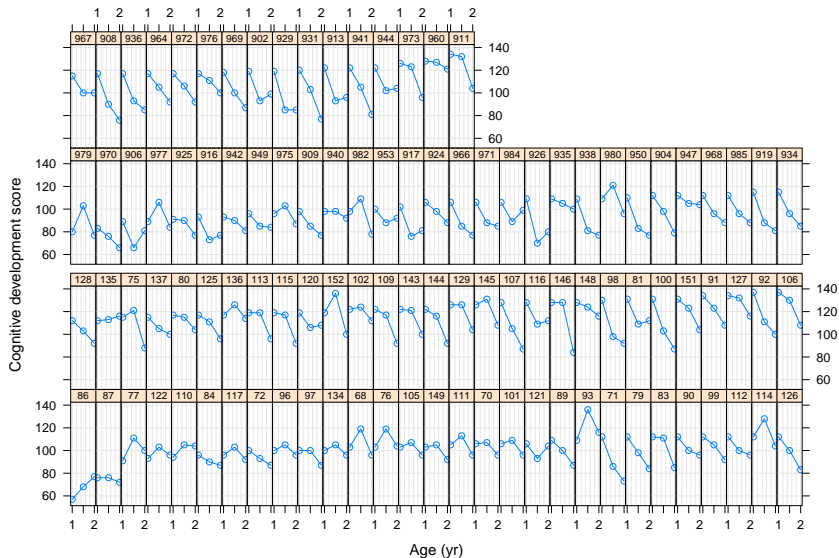
```
> fm11@re@theta
```

```
[1] 0.4479973  1.9418232  1.6904725  1.8028408  1.6360132
[6] 0.3032622  0.1790634  0.6128859  0.0797213 -0.3249881
[11] 0.0000000
```

What to watch for in the verbose output

- In this model the criterion is being optimized with respect to 11 parameters.
- These are the variance component parameters, θ . The fixed-effects coefficients, β , and the common scale parameter, σ , are at their conditionally optimal values.
- Generally it is more difficult to estimate a variance parameter (either a variance or a covariance) than it is to estimate a coefficient. Estimating 11 such parameters requires a considerable amount of information.
- Some of these parameters are required to be non-negative. When they become zero or close to zero (2.7×10^{-7} , in this case) the variance-covariance matrix is degenerate.
- The @re@lower slot contains the lower bounds. Parameter components for which @re@lower is $-\text{Inf}$ are unbounded. The ones to check are those for which @re@lower is 0.

Data on early childhood cognitive development



Fitting a model to the Early data

- The Early data in the `mLmRev` package are from a study on early childhood cognitive development as influenced by a treatment. These data are discussed in **Applied Longitudinal Data Analysis** (2003) by Singer and Willett.
- A model with random effects for slope and intercept is

```
> Early <- within(Early, tos <- age-0.5)
> fm12 <- lmer(cog ~ tos+trt:tos+(tos|id), Early, verbose=TRUE)
```

```
npt = 7 , n = 3
```

```
rhobeg = 0.2 , rhoend = 2e-07
```

```
0.020: 11:      2368.50; 1.09296 -0.173139 0.0953204
0.0020: 30:      2364.50; 1.48770 -0.374305 0.0138819
0.00020: 42:      2364.50; 1.48462 -0.372458 0.00762182
2.0e-05: 58:      2364.50; 1.48417 -0.372319 0.00114304
2.0e-06: 74:      2364.50; 1.48420 -0.372480 0.00000
2.0e-07: 80:      2364.50; 1.48420 -0.372481 0.00000
```

```
At return
```

```
87:      2364.5016: 1.48420 -0.372481 1.72050e-07
```

Fitted model for the Early data

Linear mixed model fit by REML ['merMod']

Formula: cog ~ tos + trt:tos + (tos | id)

Data: Early

REML criterion at convergence: 2364.502

Random effects:

| Groups | Name | Variance | Std.Dev. | Corr |
|----------|-------------|----------|----------|--------|
| id | (Intercept) | 166.40 | 12.900 | |
| | tos | 10.48 | 3.237 | -1.000 |
| Residual | | 75.54 | 8.691 | |

Number of obs: 309, groups: id, 103

Fixed effects:

| | Estimate | Std. Error | t value |
|-------------|----------|------------|---------|
| (Intercept) | 120.783 | 1.824 | 66.22 |
| tos | -22.470 | 1.494 | -15.04 |
| tos:trtY | 7.646 | 1.447 | 5.28 |

Here is it obvious that there is a problem. However, Singer and Willett did not detect this in model fits from SAS PROC MIXED or MLWin, both of which reported a covariance estimate.

Other practical issues

- In some disciplines there is an expectation that data will be analyzed starting with the most complex model and evaluating terms according to their p-values.
- This can be appropriate for carefully balanced, designed experiments. It is rarely a good approach on observational, imbalanced data.
- Bear in mind that this approach was formulated when graphical and computational capabilities were very limited.
- A more appropriate modern approach is to explore the data graphically and to fit models sequentially, comparing these fitted models with tests such as the LRT.

Fixed-effects or random-effects?

- Earlier we described the distinction between fixed and random effects as dependent on the repeatability of the levels.
- This is the basis for the distinction but the number of levels observed must also be considered.
- Fitting mixed-effects models requires data from several levels of the grouping factor. Even when a factor represents a random selection (say sample transects in an ecological study) it is not practical to estimate a variance component from only two or three observed levels.
- At the other extreme, a census of a large number of levels can be modeled with random effects even though the observed levels are not a sample.

Summary

- In models of longitudinal data on several subjects we often incorporate random effects for the intercept and/or the slope of the response with respect to time.
- By default we allow for a general variance-covariance matrix for the random effects for each subject.
- The model can be restricted to independent random effects when appropriate.
- For other interactions of fixed-effects factors and random-effects grouping factors, the general term can lead to estimation of many variance-covariance parameters. We may want to restrict to independent random effects for the subject and the subject/type interaction.