

Homework Assignment #8

This is an exercise, not to be handed in. However, give it a shot, and then run the code in the solutions.

These problems are taken from the book by Pinheiro and Bates: *Mixed Effects Models in S and Splus*, p. 53ff. The data are available in the R-package *nlme*. This package can be downloaded from the CRAN webpage (<http://cran.r-project.org/>), or directly installed by using the R command `install.packages("nlme")`. After you installed the package, load the package using `library(nlme)`. The commands `data(Oxboys)` and `data(Alfalfa)` will load the data frames you need for the homework. Please contact me (ingo@jhu.edu) if you have trouble with any of the above. Please also note that the Pinheiro/Bates book uses Splus to obtain the results. There are some minor differences to R. In the *Oxboys* problem, parts (i) and (k), R requires $I(\text{age}^2)$ to specify a quadratic term, instead of age^2 in Splus. R also uses the treatment contrasts by default, so there is no need to specify those in the second part of the problem using the *Alfalfa* data. I altered some details in the phrasing of the problems to take those differences into account.

When you plot some of those grouped data and want to get rid of the grey background, use

```
> library(lattice)
> trellis.device(col=F)
```

to get a white background (which is a lot easier on the eye).

1. The *Oxboys* data consist of the heights of 26 boys from Oxford, England, each measured on nine different occasions. The structure is similar to that of *OrthoFem* in the Introduction to mixed-effects model handout.
 - (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*. 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 a linear mixed effects model to the data with random effects for both the intercept and the slope. 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)`), where *object* should be replaced with the name of the `lme` object in both cases. 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 should 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* would be fit with `lmList(height~age+I(age^2),Oxboys)`.
 - (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 numeric 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?
2. The *Alfalfa* data is an example of a so-called split-plot experiment. The structure is similar to that of the *Oats* data in the handouts: a 3×4 full factorial design 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? Also plot the interactions.
 - (b) Use *lme* to fit a two-level linear mixed effects model with grouping structure *Block/Variety*, including a single random intercept for each level of grouping (which means `random=~1|Block/Variety`). Assume a full factorial structure with main effects and interactions for the fixed effects (which means that you have `fixed=Yield~Date*Variety`).
 - (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 can be fit using *aov* and the *Error* function (consider the command `aov(Yield~Date*Variety+Error(Block/Variety),Alfalfa)`). Compare the results from the *aov* and *lme* fits, in particular the F and p-values for testing the terms in the fixed effects model. 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 linear mixed effects model using fixed effects for *Date* only (a simple way to do this is to use the command `update(object,Yield~Date)`, where *object* should be replaced with the name of the previous *lme* object). Print the resulting object using *summary*, and investigate the differences between cutting dates. 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 linear mixed effects model's assumptions?