Homework 02 Solutions

2022-02-05

library(arm) ## includes lme4,MASS, Matrix
library(ggplot2)

Problem 1

In lecture 06 (crossed random effects, cross-level interactions), six different models were fitted to the London Schools data, but only the lmer() model specification was given for each model. For each of these six models, write out the two-level MLM formulation of the model. You can use mathematical variable names $(y_i, x_{3i},$ etc.) or English variable names (LRT_i, etc.), as long as it is clear which variables in the data frame are in each of your six MLM models. Be sure to state all distribution assumptions, and be sure that the indexing (i, j, j[i] and so forth) is clear.

In all that follows we will use the definitions::

$$i = 1...n$$
 (students)
 $j = 1...J$ (schools)
 $j[i] = the school j that student i is in$

See my response on Piazza (https://piazza.com/class/kykc45ull1j7dh?cid=19) for general guidance on how to do these...

Part 1(a)

Random Intercept Only: Y ~ 1 + (1 | school)

$$Y_i = \alpha_{0[i]} + \epsilon_i , \quad \epsilon_i \stackrel{iia}{\sim} N(0, \sigma^2)$$

$$\alpha_{0j} = \beta_0 + \eta_{0j} , \quad \eta_{0j} \stackrel{iid}{\sim} N(0, \tau_0^2)$$

Part 1(b)

Random Intercept, Fixed Slope: Y ~ LRT + (1 | school)

$$Y_i = \alpha_{0[i]} + \beta_1 \cdot \text{LRT}_i + \epsilon_i , \quad \epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

$$\alpha_{0j} = \beta_0 + \eta_{0j} , \quad \eta_{0j} \stackrel{iid}{\sim} N(0, \tau_0^2)$$

Part 1(c)

Random Intercept, Random Slope: Y ~ LRT + (LRT | school)

$$Y_{i} = \alpha_{0[i]} + \alpha_{1j[i]} \cdot \operatorname{LRT}_{i} + \epsilon_{i} , \quad \epsilon_{i} \stackrel{iia}{\sim} N(0, \sigma^{2})$$

$$\alpha_{0j} = \beta_{0} + \eta_{0j} , \quad \eta_{0j} \stackrel{iid}{\sim} N(0, \tau_{0}^{2})$$

$$\alpha_{1j} = \beta_{1} + \eta_{1j} , \quad \eta_{1j} \stackrel{iid}{\sim} N(0, \tau_{1}^{2})$$

$$\operatorname{Corr}(\eta_{0j}, \eta_{1j}) = \rho$$

Part 1(d)

Random Intercept, Random Slope, CWC: Y ~ LRT.cwc + (1 + LRT.cwc | school)

Let \overline{LRT}_j be the sample average of LRT in school j and let $LRT.cwc_i = LRT_i - LRT.cwc_{j[i]}$. Then the MLM is

$$Y_{i} = \alpha_{0[i]} + \alpha_{1j[i]} \cdot \text{LRT.cwc}_{i} + \epsilon_{i} , \quad \epsilon_{i} \stackrel{iid}{\sim} N(0, \sigma^{2})$$

$$\alpha_{0j} = \beta_{0} + \eta_{0j} , \quad \eta_{0j} \stackrel{iid}{\sim} N(0, \tau_{0}^{2})$$

$$\alpha_{1j} = \beta_{1} + \eta_{1j} , \quad \eta_{1j} \stackrel{iid}{\sim} N(0, \tau_{1}^{2})$$

$$\text{Corr}(\eta_{0j}, \eta_{1j}) = \rho$$

Part 1(e)

Random Intercept, individual and group predictor: $Y \sim LRT + sch.avg + (1 + LRT | school)$ Using the same definitions as in the previous part, let $sch.avj_j = \overline{LRT}_j$. Then the MLM is

$$Y_{i} = \alpha_{0[i]} + \alpha_{1j[i]} \cdot \operatorname{LRT}_{i} + \beta_{2} \cdot \operatorname{sch.avg}_{j[i]} + \epsilon_{i} , \quad \epsilon_{i} \stackrel{iid}{\sim} N(0, \sigma^{2})$$

$$\alpha_{0j} = \beta_{0} + \eta_{0j} , \quad \eta_{0j} \stackrel{iid}{\sim} N(0, \tau_{0}^{2})$$

$$\alpha_{1j} = \beta_{1} + \eta_{1j} , \quad \eta_{1j} \stackrel{iid}{\sim} N(0, \tau_{1}^{2})$$

$$\operatorname{Corr}(\eta_{0j}, \eta_{1j}) = \rho$$

Part 1(f)

Random Intercept, Cross-Level Interaction: Y ~ LRT * sch.avg + (1 + LRT | school) With all the same definitions as above, the MLM is

$$\begin{split} Y_i &= \alpha_{0[i]} + \alpha_{1j[i]} \cdot \operatorname{LRT}_i + \beta_2 \cdot \operatorname{sch.avg}_{j[i]} + \beta_3 \cdot \operatorname{LRT}_i \cdot \operatorname{sch.avg}_{j[i]} + \epsilon_i \quad , \quad \epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2) \\ \alpha_{0j} &= \beta_0 + \eta_{0j} \quad , \quad \eta_{0j} \stackrel{iid}{\sim} N(0, \tau_0^2) \\ \alpha_{1j} &= \beta_1 + \eta_{1j} \quad , \quad \eta_{1j} \stackrel{iid}{\sim} N(0, \tau_1^2) \\ &\qquad \operatorname{Corr}(\eta_{0j}, \eta_{1j}) = \rho \end{split}$$

Problem 2

Return to the model on slide 10 of lecture 05 (random slopes, correlation and no centering). Use the AIC.ml, BIC.ml, and DIC.lm functions defined at the end of lecture 06 (crossed random effects, cross-level interactions) as needed for the following quesitons.

Part 2(a)

Write the multilevel model that has a cross-level interaction between log.uranium and floor. Interpret this model: what does it say about how log.uranium and floor influence log.radon in each house, and how is it different from the model without the cross-level interaction?

Here is the MLM with the cross-level interaction:

$$y_{i} = \alpha_{0[i]} + \alpha_{1j[i]} \cdot x_{i} + \beta_{2} \cdot u_{j[i]} + \beta_{3} \cdot x_{i} \cdot u_{j[i]} + \epsilon_{i} , \quad \epsilon_{i} \stackrel{iid}{\sim} N(0, \sigma^{2})$$

$$\alpha_{0j} = \beta_{0} + \eta_{0j} , \quad \eta_{0j} \stackrel{iid}{\sim} N(0, \tau_{0}^{2})$$

$$\alpha_{1j} = \beta_{1} + \eta_{1j} , \quad \eta_{1j} \stackrel{iid}{\sim} N(0, \tau_{1}^{2})$$

$$\operatorname{Corr}(\eta_{0j}, \eta_{1j}) = \rho$$

where $y_i = \log(radon)$ reading for the *i*th house, $x_i = 0$ if the radon level was measured in the basement; 1 if it was measured on the first floor, and j[i] is the county j that house i is in.

And here is the MLM without the cross-level interaction:

$$y_{i} = \alpha_{0[i]} + \alpha_{1j[i]} \cdot x_{i} + \beta_{2} \cdot u_{j[i]} + \epsilon_{i} , \quad \epsilon_{i} \stackrel{iid}{\sim} N(0, \sigma^{2})$$

$$\alpha_{0j} = \beta_{0} + \eta_{0j} , \quad \eta_{0j} \stackrel{iid}{\sim} N(0, \tau_{0}^{2})$$

$$\alpha_{1j} = \beta_{1} + \eta_{1j} , \quad \eta_{1j} \stackrel{iid}{\sim} N(0, \tau_{1}^{2})$$

$$\operatorname{Corr}(\eta_{0j}, \eta_{1j}) = \rho$$

In the model with the cross-level interaction, $u_{j[i]}$ affects the slope on x_i , which we can see by just collecting terms that involve a factor of x_i :

slope on $x_i = \beta_1 + \beta_2 \cdot u_{j[i]} + \eta_{1j[i]}$

whereas in the model without the cross-level interaction, the slope on x_i doesn't depend on $u_{i|i|}$ at all:

slope on
$$x_i = \beta_1 + \eta_{1j[i]}$$

Part 2(b)

Which model fits the data better, with or without the cross-level interaction between floor and log.uranium?

```
AIC.ml <- function(M) {AIC(update(M,REML=F))}
BIC.ml <- function(M) {BIC(update(M,REML=F))}</pre>
DIC.ml <- function(M) {extractDIC(update(M,REML=F))}</pre>
mn.radon <- read.table("mn-radon.txt",header=T)</pre>
lmer.0 <- lmer(log.radon ~ floor</pre>
                                                  + (1 + floor | county), data=mn.radon)
lmer.1 <- lmer(log.radon ~ floor + log.uranium + (1 + floor | county), data=mn.radon)</pre>
lmer.2 <- lmer(log.radon ~ floor * log.uranium + (1 + floor | county), data=mn.radon)</pre>
res <- rbind(AIC=sapply(list(lmer.0, lmer.1,lmer.2), AIC.ml),</pre>
             BIC=sapply(list(lmer.0, lmer.1,lmer.2), BIC.ml),
             DIC=sapply(list(lmer.0, lmer.1,lmer.2), DIC.ml))
colnames(res) <- c("(w/o log.uranium)", "(w/o interaction)", "(with interaction)")</pre>
t(round(res,2))
                                    BIC
                                             DIC
##
                            AIC
## (w/o log.uranium)
                       2173.11 2202.05 2161.11
```

As you can see, all three information criteria like log.uranium in the model, and they all prefer the model with a main effect, better than the model with the interaction.

2131.60 2165.37 2117.60

(with interaction) 2135.22 2173.81 2119.22

Part 2(c)

(w/o interaction)

Construct a new Level 2 variable nj that is equal to the number of houses sampled in each county. Like log.uranium, this variable will have the same value for each house in the county. Fit two models like the one on slide 10 of lecture 05, but with nj replacing log.uranium: one model without a cross-level interaction between nj and floor, and one model with the cross-level interaction. Of the four models (with nj or with log.uranium, and with or without the cross-level interaction) which one(s) fit the data best? Accompany your results with a short paragraph interpreting them.

```
nj <- with(mn.radon,unlist(lapply(split(county,county),</pre>
                                    function(x) rep(length(x),length(x))))
lmer.3 <- lmer(log.radon ~ floor + nj + (1 + floor | county), data=mn.radon)</pre>
lmer.4 <- lmer(log.radon ~ floor * nj + (1 + floor | county), data=mn.radon)</pre>
res <- rbind(AIC=sapply(list(lmer.1,lmer.2,lmer.3,lmer.4), AIC.ml),</pre>
             BIC=sapply(list(lmer.1,lmer.2,lmer.3,lmer.4), BIC.ml),
             DIC=sapply(list(lmer.1,lmer.2,lmer.3,lmer.4), DIC.ml))
colnames(res) <- c("lmer.1", "lmer.2", "lmer.3","lmer.4")</pre>
t(round(res,2))
##
              AIC
                       BIC
                               DIC
## lmer.1 2131.60 2165.37 2117.60
## lmer.2 2135.22 2173.81 2119.22
## lmer.3 2169.47 2203.23 2155.47
## lmer.4 2171.26 2209.85 2155.26
```

The two models with nj do quite a bit worse, on all three fit measures, than the models with log.uranium. And, as before, the model with only a main effect for log.uranium an no interaction is preferred by all of AIC, BIC, and DIC.

Problem 3

Although it is usual to keep the correlation between random effects in a multilevel model, there are times when we'd like to formally establish whether the data support having the correlation in the model or not. One way to do that is with a likelihood ratio test. Recall that the likelihood ratio test for testing model M_0 vs model M_1 , where M_0 is obtained from M_1 by linear constraint(s), computes the log-likelihood ratio test statistic

```
-2[\log(likelihood_{M_0}) - \log(likelihood_{M_1})]
```

where the likelihoods are evaluated at the maximum likelihood values of the parameters (e.g. as fitted by lmer()), and compares the value with a chi-squared distribution with degrees of freedom equal to the number of linear constraints. A statistically significant result rejects M_0 in favor of M_1 and conversely a nonsignificant result suggests that there is not enough evidence in the data to reject M_0 .

For the model on slide 10 of lecture 05 (random slopes, correlation and no centering), use a likelihood ratio test to determine whether there is sufficient evidence in the data to keep the correlation ρ between η_{0j} and η_{1j} in the model. (Note: For reasons that we will talk about next week, you need to set REML=FALSE in lmer() when you fit M_0 and M_1 for the likelihood ratio test. You will also find the logLik() function in R useful for this exercise.) Be sure to report the number of degrees of freedom of the test, the value of the test statistic, and the criterion you are using to determine statistical significance.

```
lmer.1 <- lmer(log.radon ~ floor + log.uranium + (1 + floor | county), data=mn.radon)
lmer.1.c0 <- lmer(log.radon ~ floor + log.uranium + (1 + floor || county), data=mn.radon)
## Since there is only one linear constraint needed (rho = 0) to get from lmer.1
## to lmer.1.c0, we will be referring the likelihood ratio test statistic to a
## chi-squared distribution on with df=1.
LL <- function(M) { logLik(update(M,REML=F)) }
LR.stat <- -2*(LL(lmer.1.c0) - LL(lmer.1))</pre>
```

Clearly, for any reasonable cutoff like 0.01 or 0.05, the p-value here is huge, so there is not enough evidence in the data to reject the null hypothesis that the model without correlation lmer.1.c0 fits adequately, vs the alternative lmer.1.

Problem 4

The file allvars.csv contains CD4 percentages (CD4PCT) for a set of 254 young children with HIV who were measured several times over a period of two years (A CD4 count measures the number of CD4 cells in your blood. It's used to check the immune system function in people with HIV (https://medlineplus.gov/lab-tests/cd4-lymphocyte-count/). The dataset also includes the ages of the children at each measurement. This is an example of *growth curve data* in which the Level 1 observations are CD4PCT's at each visit, and the Level 2 "group" is a child. How the CD4PCT's change over time for each child is the primary question of interest. Because of skewing, you should replace CD4PCT with sqrt(CD4PCT) for all of the following questions.

Part 4(a)

For the first 12 children only (newpid \leq 12), make a facet plot of CD4PCT vs VISIT number. Use different plotting point colors for children who have treatmnt=1 and treatmnt=0¹. Do the same for CD4PCT vs. visage² (the age of the child on each visit). Which is a better measure of time?

First we read in the data and do a little exploring and munging...

```
cd4 <- read.csv("allvar.csv",header=T)</pre>
```

```
## It's worth doing a little exploration of the data here,
## especially to look at missing values.
```

```
apply(cd4,2,function(x) mean(is.na(x)))
```

```
##
         VISIT
                                  VDATE.
                                             CD4PCT
                    newpid
                                                                      visage
                                                             arv
## 0.00000000 0.00000000 0.00000000 0.142743222 0.103668262 0.109250399
##
                    CD4CNT
      treatmnt
                                baseage
## 0.00000000 0.146730463 0.007177033
## we only need the variables newpid, VISIT, visage, CD4PCT and treatmnt
## so let's get rid of some (but not all) missing data problems
## by deleting the other variables
cd4 <- with(cd4,data.frame(newpid=newpid,</pre>
                            VISIT=VISIT,
                            visage=visage,
                            CD4PCT=CD4PCT,
                            treatmnt=treatmnt))
```

¹Presumably, "treatmnt" stands for some sort of treatment to improve childrens' HIV status.

²If you are using ggplot(), you may wish to set scales="free_x" in the facet_wrap() function.

```
## and since we still have some missing data, we will just
## delete the rows that continue to have NA's (not a great
## practice in general, but good enough for this exercise...)
cd4 <- cd4[!apply(cd4,1,function(x) any(is.na(x))),]
## set up the sqrt of the response variable...
cd4$sqrt.CD4PCT <- sqrt(cd4$CD4PCT)
## and for coloring ggplot elements it will better if treatmnt is a factor...
cd4$treatmnt <- as.factor(cd4$treatmnt)
## select the first 12 kids...
```

```
first.12 <- (cd4$newpid <= 12)</pre>
```

I will show the faceted graph for both VISIT and visage as the time variable...

```
Here it is using VISIT...
```

```
g <- ggplot(cd4[first.12,],aes(x=VISIT,y=sqrt.CD4PCT)) +
facet_wrap( ~ newpid) +
geom_point(aes(color=treatmnt))</pre>
```



```
... and here it is using visage as the time variable...
```



The graphs look very similar, but visage, which is the actual age of the child, would seem to have more physical/medical meaning than VISIT, which is just the sequential visit number of each visit. If you have a reason for preferring VISIT, go for it.

I will show the solutions below both ways (it will turn out not to make much difference!).

Part 4(b)

Build a multilevel model with random slopes and intercepts for all the children, using a measure of time (VISIT or visage, whichever you answered in part (a)) as a Level 1 predictor, and treatmnt as a Level 2 predictor. Report the estimated fixed effects and variances from the model, and add the fitted regression lines to your plot from part (a).

```
Here's the model, using VISIT as the measure of time:
display(lmer.1 <- lmer(sqrt.CD4PCT ~ 1 + VISIT + treatmnt + (1+VISIT|newpid), data=cd4))
## lmer(formula = sqrt.CD4PCT ~ 1 + VISIT + treatmnt + (1 + VISIT |
## newpid), data = cd4)
## coef.est coef.se
## (Intercept) 4.69 0.13
```

```
## VISIT
               -0.03
                          0.01
                0.18
                          0.18
## treatmnt2
##
## Error terms:
##
   Groups
             Name
                          Std.Dev. Corr
   newpid
             (Intercept) 1.40
##
##
             VISIT
                          0.05
                                   -0.10
##
   Residual
                          0.72
## ---
## number of obs: 1075, groups: newpid, 251
## AIC = 3140.7, DIC = 3100.6
## deviance = 3113.6
```

The overall intercept is $\hat{\beta}_0 = 4.69$, the overall slope on VISIT is $\beta_1 = -0.03$ and the overall treatment effect is $\beta_2 = 0.18$. The residual variances are $\hat{\sigma}^2 = (0.51)^2 = 0.26$, $\hat{\tau}_0^2 = (1.40)^2 = 1.96$ and $\hat{\tau}_1^2 = (0.05)^2 = 0.0025$, and $\hat{\rho}$ is a very modest -0.10. The only things that seem surprising are

- $\hat{\tau}_0 = 1.40$ seems large, considering that the most sqrt.CD4PCT values are between zero and 5;
- The treatmt effect, $\hat{\beta}_2 = 0.18$, is not significantly different from zero, even though the scatter plots of the first 12 children would suggest otherwise!

Next we overlay regression lines on the scatterplots from part (a). The regression lines track the data well (and even have different intercepts, even though the the treatmnt effect is nonsignificant in the model summary! [what part of the model could be helping here, do you suppose??])

(It was a little trickier than I expected, to overlay the regression lines and have them colored differntly for each of the two treatmnt levels, so have a look at the code below also.)

Here are the overlaid regression lines, using **VISIT** as the time variable:



Here's the model summary using visage as the measure of time:

```
display(lmer.2 <- lmer(sqrt.CD4PCT ~ 1 + visage + treatmnt + (1+visage|newpid),data=cd4))</pre>
## lmer(formula = sqrt.CD4PCT ~ 1 + visage + treatmnt + (1 + visage |
##
       newpid), data = cd4)
##
               coef.est coef.se
##
   (Intercept)
                5.34
                          0.18
                -0.23
                          0.04
##
  visage
## treatmnt2
                 0.23
                          0.18
##
## Error terms:
                          Std.Dev. Corr
##
    Groups
             Name
##
    newpid
             (Intercept)
                          1.45
                          0.29
##
                                    -0.53
             visage
##
    Residual
                          0.75
##
   ___
## number of obs: 1075, groups: newpid, 251
## AIC = 3152, DIC = 3119.5
## deviance = 3128.8
```

The story is much the same using visage as a measure time, as using VISIT as a measure of time. Again $\hat{\tau}_0$ is rather large, and the estimated treatmnt effect $\hat{\beta}_2$ is nonsignificant.

Here's the facets plot for the first 12 children, using visage for time. Again the regression lines are tracking differences in the children rather well, even though the treatmnt effect is nonsignificant.



Part 4(c)

Try to expand the model by adding a (cross-level) interaction between treatmnt and your measure of time. Comment on changes in the estimated fixed effects and variances for this model, vs. the model in part (b). Add the fitted regression lines from this model to your previous facets plot. Does treatmnt account for an interesting amount of the variation in slopes (or intercepts) among children in the study?

OK, here we go with the crosslevel interaction model using VISIT as the time variable:

```
display(lmer.3 <- lmer(sqrt.CD4PCT ~ 1 + VISIT * treatmnt + (1+VISIT|newpid), data=cd4))
## lmer(formula = sqrt.CD4PCT ~ 1 + VISIT * treatmnt + (1 + VISIT |
       newpid), data = cd4)
##
##
                   coef.est coef.se
## (Intercept)
                    4.71
                             0.13
## VISIT
                   -0.03
                             0.01
                             0.19
## treatmnt2
                    0.14
## VISIT:treatmnt2 0.01
                             0.01
##
## Error terms:
## Groups
             Name
                         Std.Dev. Corr
  newpid
             (Intercept) 1.40
##
##
             VISIT
                         0.05
                                  -0.10
                         0.72
##
  Residual
## ---
## number of obs: 1075, groups: newpid, 251
## AIC = 3149.4, DIC = 3093
## deviance = 3113.2
```

Same story as before: neither the coefficient on treatmnt nor the coefficient on the interaction is significantly different from zero, but $\hat{\tau}_0$ is somewhat large for the scale of the data.

Here are the regression lines plotted over the scatter plots using VISIT for time; again you can see that they are tracking differences between children that we might have thought would be detected by the treatment effect.

```
coef.3 <- coef(lmer.3)$newpid[,1:4] ## coef.3[,1] = beta0 + eta0 (random intcpt)</pre>
                                      ## coef.3[,2] = beta1 + eta1 (random slope on time)
                                      ## coef.3[,3] = beta2
                                                                      (coef on treatmnt)
                                      ## coef.3[,4] = beta3
                                                                      (coef on interaction)
tx <- with(cd4,sapply(split(treatmnt,newpid), function(x) x[1])) ## one per child</pre>
params <- data.frame(newpid=sort(unique(cd4$newpid)),</pre>
                      alpha0=coef.3[,1],
                      alpha1=coef.3[,2],
                     beta2=coef.3[,3],
                     beta3=coef.3[,4],
                      tx
                      )
g + geom_abline(data=params[params$newpid<=12,],aes(intercept=alpha0 + ifelse(tx==2,beta2,0),
                                                      slope=alpha1 + ifelse(tx==2,beta3,0),
```

color=tx))



Here's the cross level interaction model, using visage as time

```
display(lmer.4 <- lmer(sqrt.CD4PCT ~ 1 + visage * treatmnt + (1+visage|newpid),data=cd4))</pre>
## lmer(formula = sqrt.CD4PCT ~ 1 + visage * treatmnt + (1 + visage |
##
       newpid), data = cd4)
##
                     coef.est coef.se
## (Intercept)
                      5.34
                                0.21
## visage
                     -0.23
                                0.05
## treatmnt2
                       0.24
                                0.32
   visage:treatmnt2 0.00
                                0.08
##
##
## Error terms:
                           Std.Dev. Corr
##
    Groups
              Name
##
              (Intercept) 1.46
    newpid
                           0.30
                                     -0.53
##
              visage
                           0.75
    Residual
##
##
   ___
## number of obs: 1075, groups: newpid, 251
## AIC = 3157.3, DIC = 3116.3
## deviance = 3128.8
Again, no significant effects for visage, and a rather large \hat{\tau}_0.
```

Here's the facet plot...

```
coef.4 <- coef(lmer.4)$newpid[,1:4]</pre>
                                       ## coef.4[,1] = beta0 + eta0
                                                                       (random intcpt)
                                       ## coef.4[,2] = beta1 + eta1
                                                                       (random slope on time)
                                       ## coef.4[,3] = beta2
                                                                       (coef on treatmnt)
                                       ## coef.4[,4] = beta3
                                                                       (coef on interaction)
tx <- with(cd4,sapply(split(treatmnt,newpid), function(x) x[1])) ## one per child</pre>
params <- data.frame(newpid=sort(unique(cd4$newpid)),</pre>
                      alpha0=coef.4[,1],
                      alpha1=coef.4[,2],
                      beta2=coef.4[,3],
                      beta3=coef.4[,4],
                      tx
                      )
```





It does not appear that treatmnt is having a significant effects on the intercepts or the slopes. We could explore a little further with information criteria:

lmer.0 <- lmer(sqrt.CD4PCT ~ 1 + VISIT + (1+VISIT|newpid),data=cd4)</pre>

Warning in checkConv(attr(opt, "derivs"), opt\$par, ctrl = control\$checkConv, :
Model failed to converge with max|grad| = 0.00205385 (tol = 0.002, component 1)

```
res <- rbind(AIC=sapply(list(lmer.0,lmer.1,lmer.2,lmer.3,lmer.4), AIC.ml),
BIC=sapply(list(lmer.0,lmer.1,lmer.2,lmer.3,lmer.4), BIC.ml),
DIC=sapply(list(lmer.0,lmer.1,lmer.2,lmer.3,lmer.4), DIC.ml))
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00584617 (tol = 0.002, component 1)
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00584617 (tol = 0.002, component 1)
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00584617 (tol = 0.002, component 1)
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00584617 (tol = 0.002, component 1)
colnames(res) <- c("lmer.0","lmer.1", "lmer.2", "lmer.3","lmer.4")
t(round(res,2))
```

 ##
 AIC
 BIC
 DIC

 ##
 lmer.0
 3126.56
 3156.44
 3114.56

 ##
 lmer.1
 3127.63
 3162.49
 3113.63

 ##
 lmer.2
 3142.76
 3177.62
 3128.76

 ##
 lmer.3
 3129.21
 3169.06
 3113.21

 ##
 lmer.4
 3144.75
 3184.60
 3128.75

If we restrict our attention to lmer.1 through lmer.4, the winner seems to be lmer.1, with model formula sqrt.CD4PCT ~ 1 + VISIT + treatmnt + (1+VISIT | newpid). Comparing that with lmer.0 which just omits treatmnt, the simpler lmer.0 model seems to be favored; all of this is consistent with our earlier findings that treatmt doesn't seem to have a significant effect on either the slope on time, or the intercept, of the MLM's we have built.

NOTE (not to be graded):

What I suspect is going on is that the random part of the intercept, η_{0j} , is accounting for variation between treated and untreated children, that would otherwise show up in the treatmnt effect coefficient. We could check this out by removing the random part of the intercept from one of the models and comparing.

Below I have fitted the model with fixed intercept and random slope (note the 0+visage term in the lmer call; the 0 removes the random intercept η_{0j}). As you can see, the coefficient on treatmnt is now (barely) significant, and the model still tracks the data well.

Comparing all these models is also a good exercise in seeing the difference between getting a model that does well with prediction, vs getting a model that might be the "true" model. All of the models we've looked at seem to do well with in-sample prediction (out-of-sample would also be interesting to look at), but they obviously can't all be the "correct/true" model describing the effect of treatment and time on CD4PCT levels.

display(lmer.5 <- lmer(sqrt.CD4PCT ~ 1 + visage + treatmnt + (0+visage|newpid),data=cd4))

```
## lmer(formula = sqrt.CD4PCT ~ 1 + visage + treatmnt + (0 + visage |
##
       newpid), data = cd4)
##
               coef.est coef.se
## (Intercept)
                5.16
                          0.13
## visage
               -0.20
                          0.04
## treatmnt2
                0.33
                          0.14
##
## Error terms:
## Groups
             Name
                    Std.Dev.
```

```
newpid
             visage 0.42
##
## Residual
                     0.82
## ---
## number of obs: 1075, groups: newpid, 251
## AIC = 3297.1, DIC = 3267.3
## deviance = 3277.2
coef.5 <- coef(lmer.5)$newpid[,1:3] ## coef.5[,1] = beta0</pre>
                                                                      (fixed intcpt)
                                      ## coef.5[,2] = beta1 + eta1
                                                                      (random slope on time)
                                      ## coef.5[,3] = beta2
                                                                      (coef on treatmnt)
tx <- with(cd4,sapply(split(treatmnt,newpid), function(x) x[1])) ## one per child</pre>
params <- data.frame(newpid=sort(unique(cd4$newpid)),</pre>
                      beta0=coef.5[,1],
                      alpha1=coef.5[,2],
                      beta2=coef.5[,3],
                      tx
                      )
```



color=tx))