Lab 4

Psychology 319 (GCM)

Instructions. Work through the lab, saving the output as you go. If you work in Microsoft Word, you can easily copy any graph to Word via the clipboard. Numerical output may also be copied easily by highlighting, moving it to the clipboard, then copying into Word. However, you should format R output in TrueType Courier New font so that it is *monospaced.* Output from this lab is to be handed in by Monday, February 22. Your output file should be named `LAST_FIRST_LAB4.DOC`, where `LAST` is your last name, and `FIRST` is your first name. Any additional files should have the same naming scheme, except the file extension should be correct. You may add any description text you wish after `LAB4` in the file name.

Preamble. Today's lab involves the use of R's capabilities to produce "empirical Bayes" estimates of individual's trajectories after fitting a model.

1 Introduction

A key aspect of multilevel modeling, as emphasized in standard textbooks, is its ability to generate improved estimates that take into account the reliability of information. For example, suppose you have a set of growth curve data on a group of individuals.

In some cases, you may be interested in computing individual trajectories. You can estimate individual trajectories using OLS, and the slopes and intercepts estimated this way will be unbiased. But will they be optimal?

In one sense, they won't. Another kind of estimation will yield better estimates. These estimates, known as "Empirical Bayes" (EB) estimates, have a specific optimality property. As Raudenbush and Bryk (2002, p. 67) put it,

> To be precise, this estimator is optimal in the specific sense that no other point estimator of β_{0j} has smaller expected mean square error, where the expectation is taken over the conditional distribution of β_{0j}, σ^2, and τ_{00}.

HLM will generate EB estimates automatically and put them in its level-2 residuals file. On the other hand, R is not so kind or convenient, and a little work is required.

Singer and Willett (p. 132–137) discuss how to calculate EB estimates in the context of the longitudinal study on alcohol use. In particular, they generate a hand-calculated example of EB estimates of the slope and regression line for Subject 23 under Model F. In what follows, we'll review the Singer-Willett calculations and demonstrate how to reproduce them for all 82 participants simultaneously.

2 Calculating Model-Based Estimates of Slope and Intercept

In model F, `PEER` is centered (we use a derived variable `CPEER`) while `COA` is not. The intercepts therefore represent a child of non-alcoholic parents whose peers at age 14 are average consumers (`PEER` $= 1.018$ and `COA` $= 0$).

The level-2 models are

$$\pi_{0i} = \gamma_{00} + \gamma_{01} COA_i + \gamma_{02} CPEER_i + \zeta_{0i} \quad (1)$$

$$\pi_{1i} = \gamma_{10} + \gamma_{12} CPEER_i + \zeta_{1i} \quad (2)$$

Make sure that you have the file *alcohol1_pp.txt* available in your working directory.

While HLM works in terms of the level-1 and level-2 breakdown, R works directly with the composite model specification. As we saw before, the composite model F given in Singer and Willett Table 4.2 can be fit in R routinely as follows:

```
> library(lme4)
> alcohol1 <- read.table("alcohol1_pp.txt", header=T, sep=",")
> attach(alcohol1)
> time <- age_14
> model.f <- lmer(alcuse ~ coa + cpeer + time + cpeer:time +
+  (time | id),REML=FALSE)
```

R does produce a list of subject-specific coefficients using its `coef` command. However, these are the coefficients for the composite model. To construct subject-specific estimates for slopes and intercepts, we need to use fixed effect coefficients and random effect estimates, as described by Singer and Willett in their equation 4.21, page 135, which unfortunately has typographical errors. As the errata on the Singer-Willett website notes, the formulas *should* read

$$\begin{aligned} \tilde{\pi}_{0i} &= \hat{\pi}_{0i} + \hat{\zeta}_{0i} \quad (3) \\ \tilde{\pi}_{1i} &= \hat{\pi}_{1i} + \hat{\zeta}_{1i} \quad (4) \end{aligned}$$

Note that $\tilde{\pi}_{0i}$ and $\tilde{\pi}_{1i}$ are the EB estimates for the intercept and slope, respectively, for the ith individual.

They are produced by constructing each person's $\hat{\pi}$s using Equations 1 and 2 with fixed effect estimates of the γs, then adding each person's estimated residual (random effect).

Singer and Willett perform this calculation for individual 23 on page 134–135. Here's how to do these calculations simultaneously for all the individuals in the study.

First, let's take a look at the fixed effects.

```
> fixef(model.f)

(Intercept)         coa       cpeer        time  cpeer:time
  0.3938747   0.5711965   0.6951827   0.2705847  -0.1513771
```

These are the numbers that Singer and Willett use in their Equations 4.20. We simply need to apply them to the values of COA and $CPEER$ for each individual.

There is a minor technical problem we have to overcome, however. That is, our data frame is in long format, so each individual has more than one line. One way of overcoming the problem is to reshape the `alcohol1` data frame, like this:

```
> persons <- reshape(alcohol1,direction="wide",timevar="age_14")
```

Now each person has one line. It is straightforward (if somewhat intricate) to construct the individual slopes and intercepts as follows. (Study the syntax carefully, and investigate how it works. In particular, note the method I employed to access the results of the `ranef` call.)

```
> intercepts <- fixef(model.f)[1] + fixef(model.f)[2]*persons$coa.1 +
+ fixef(model.f)[3] * persons$cpeer.1 + ranef(model.f)$id[,1]
> intercepts

 [1]  1.466501640  0.354769464  1.177198241  1.083144762  0.298896328
 [6]  2.222071008  1.594276367 -0.014144590  0.366621830  1.247484803
[11]  0.162541781  0.843896505  0.097601682  2.419908571  1.793085869
```

```
[16]  1.090901632  0.486112532  1.629196899 -0.014144590  2.658701201
[21]  1.351539598  2.172258380  0.588405995  0.173071615  2.043366406
[26]  0.794290955  1.664390172  0.298896328  0.354769464  1.218328913
[31]  1.970267115  0.858253072  0.565129077  0.228944750  1.033833684
[36]  1.046083872  1.710609215  0.614191581  0.048511542  0.620884051
[41]  0.839930528  0.048511542  0.048511542  0.167853721  1.660120231
[46]  0.127528087  0.235727746 -0.264529376 -0.264529376  0.178177338
[51] -0.264529376  0.235727746 -0.264529376 -0.208656240  0.277673420
[56]  0.774142568 -0.264529376  0.768695183  0.196337959  0.001703373
[61]  1.602944386  0.048511542 -0.264529376  0.547880081  0.444410087
[66]  1.444843525  0.367070815 -0.077313172  0.048511542  1.102008170
[71]  0.048511542  0.048511542 -0.264529376  0.726877523  0.361552459
[76] -0.264529376  0.471308322  0.048511542  0.864715525  0.651062127
[81]  0.277673420  0.822898005
```

```
> slopes <- fixef(model.f)[4] + fixef(model.f)[5]*persons$cpeer.1 +
+ ranef(model.f)$id[,2]
> slopes
```

```
 [1]  0.288971847  0.224867026  0.779000528  0.351961444 -0.021404448
 [6]  0.412723047  0.015094651  0.146727534  1.057308517 -0.056812246
[11]  0.925506323  0.204179404  0.639270482  0.303098136  0.275055559
[16]  0.207182256  0.036067140 -0.106377843  0.146727534 -0.310899997
[21]  0.205233679 -0.191853855  0.499425521  0.204199122  0.049949700
[26]  0.261266428 -0.156966626 -0.021404448  0.224867026  0.319122579
[31]  0.478644612  0.246808825  0.384347593  0.450470596 -0.218342322
[36]  0.327639242  0.580777677  0.599522373  0.077078265 -0.230338812
[41]  0.260813791  0.077078265  0.077078265  1.065301766  0.482612078
[46]  0.425358718  0.134549853  0.245210247  0.245210247  0.007435717
[51]  0.245210247  0.134549853  0.245210247  0.491481720 -0.046002899
[56]  0.641796438  0.245210247  0.630266536  0.728651322  0.650962287
[61]  0.227652433  0.077078265  0.245210247  0.742701206  0.413187757
[66]  0.759611298 -0.054250033  0.302681834  0.077078265 -0.251187908
[71]  0.077078265  0.077078265  0.245210247  0.604220574 -0.091053716
[76]  0.245210247  0.854916268  0.077078265  0.422540799  0.429776253
[81] -0.046002899  0.160921234
```

Problem 1. Using the same approach, obtain EB estimates of the individuals' slopes and intercepts for Model C fit to the alcohol data.