STATISTICAL ANALYSIS OF LONGITUDINAL DATA I

EXPLORATORY ANALYSIS

LAB 5

Spring 2012

This lab focuses on exploratory analysis using (1) a step-up method based on hypothesis testing, and (2) a reasonable subset examination using fit indices based on the BIC. We make three assumptions in this lab.

- The shape of the growth curve is unknown.

- The number of static predictors is unknown.

- The number of random effects is unknown.

You will use the `sleep5` data frame, so the static predictors are `GPA` and `Female`. To simplify things, you will consider the static predictors as single effects (not their interaction).

Read in the Data: The Rdata file for this lab is `sleepLab05.Rdata` (note this is not a raw data file). It has the same variables as the long format file you made in Lab 01. Download the file from the course website and make a note of where this file resides on your hard drive. An Rdata file is loaded into R using the `load()` function. The data frame has the name `sleep5`, which you can see using the `ls()` function after loading the Rdata file. The response variable is `Reaction`, which is the average reaction time for a number of cognitive tests. The time metric is `Days` $(0, 1, \ldots, 9)$, and the static predictors are `gpa` and one version of sex, `Female` $(0 = \text{male}, 1 = \text{female})$. The id variable is `SubNum`.

The script for the lab is `Lab-05.R` available on the course website. You should invoke R and open `Lab-05.R` as a script file. You should also invoke your word processor or document processor as you will need to copy and paste output from R. Be sure to regularly save both your script file and the file in your word/document processor. In addition, recall that comments are preceded by # and you should use comments to create a file title and denote different sections in your script file. Include all the output and graphs in your word processor file.

To load the Rdata file, type the following code in your script file and run it. You should also load the `lme4` and `lattice` packages for the data analysis to follow.

```
#######################################################
## Tailor the first line to your system
#######################################################

> load(file = "/.../sleepLab05.Rdata")
> ls()
```

1

```
#######################################################
## Load packages
#######################################################

> library(lme4)
> library(lattice)

#######################################################
## xyplot
#######################################################

> xyplot(Reaction ~ Days | SubNum, type = c("g", "l", "p"),
    data = sleep5)
```

`/.../` indicates you replace … with the location of the file. For example„ on my computer the entire syntax would be `/Users/andrewz/Documents/EPsy8282/sleeplab.Rdata`. For Windows users, replace the pathname with `C:\\...\\`. Be sure to use double back-slashes.

An indication you successfully loaded the file is the absence of any error messages. That is, success is indicated by a new prompt line in R (i.e., >). If you did not successfully read in the file, then check your syntax and try again.

The `ls()` function shows the objects in the Rdata file, which are data frames. If you followed the directions in Lab 1, the long format data frame is named `sleep.long3`. If you renamed this data frame, then be sure to use the new name in the syntax below. It is a good idea to run `head(sleep.long3)` to see the top of the data set.

Assignment Guidelines: In each section you are directed to produce output, which you should copy from R and paste into your word/document processor. Please label the sections as indicated below and use the question numbering as indicated. All questions are worth 1 point. There is a total of 12 points possible. The majority of this lab focuses on your ability to produce the proper syntax for each approach. Thus, four of the points will be awarded for correct syntax and output. There is an external R script for this lab. Please attach a printed copy of your syntax file to the document in which you respond to the questions.

Step-Up Analysis

For the step-up analysis, you will focus on the use of the nested chi-squared test which is also known as the likelihood ratio test (LRT). Here are the steps of the analysis:

- Fit the intercept-only model (one fixed effect, one random effect).

- Add polynomial time terms (fixed effects) using models with only random intercepts.

- Add single static predictors one-at-a-time in this order: `Female`, `GPA` (typically you would need a justification for this order, here it is simply given).

- For the selected fixed effects model from the above steps add random effects up to the highest order time polynomial.

- For the selected model from the previous step, add static predictor by time polynomial effects.

Due to the nature of the analysis, you will have to supply the syntax for the steps following the first two. For each step, use the `anova()` function. Be sure to use $\alpha = 0.05$ when you test fixed effects and $\alpha = 0.10$ when you test random effects (variance components). The following syntax is provided for the initial steps. See the comments for further directions and hints.

```
#######################################################
#######################################################
##
## Nested Chi-squared Selection
##
#######################################################
#######################################################


####################################
## Step 1: Fit intercept-only model.
####################################

> lmer.0 <- lmer(Reaction ~ 1 + (1 | SubNum), data = sleep5)


####################################
## Step 2: Add time polynomials.
####################################

> lmer.1 <- lmer(Reaction ~ 1 + Days + (1 | SubNum), data =
    sleep5)
> anova(lmer.0, lmer.1)

> lmer.2 <- lmer(Reaction ~ 1 + Days + I(Days^2) + (1 |
    SubNum), data = sleep5)
> anova(lmer.1, lmer.2)

> lmer.3 <- lmer(Reaction ~ 1 + Days + I(Days^2) + I(Days^3)
    + (1 | SubNum), data = sleep5)
> anova(lmer.2, lmer.3)

> lmer.4 <- lmer(Reaction ~ 1 + Days + I(Days^2) + I(Days^3)
    + I(Days^4) + (1 | SubNum), data = sleep5)
> anova(lmer.3, lmer.4)
```

```
#######################################################
## Step 3: After selecting the model from last step,
##         add single static predictors one at a time
##         in this order: female, then gpa.
#######################################################

> # (Add your syntax here.)


#######################################################
## Step 4: After selecting the model from the last step,
##         add random effects up to the highest order
##         time polynomial.
#######################################################

> # (Add your syntax here.)


#######################################################
## Step 5: After selecting the model from the last step,
##         add static predictor by time effects.
#######################################################

> # (Add your syntax here.)
```

1. Using the labels from the data frame (`Reaction`, etc.), write the HLM of the model you selected in Step 2. Be sure to use proper notation consistent with your textbook and lecture notes.

2. In Step 2, the highest order polynomial considered was a 4th order polynomial. Should even higher order models be considered? Explain why or why not.

3. In the step-up analysis, are there any concerns in using `anova()`? Explain.

4. Using the labels from the data frame (`Reaction`, etc.), write the HLM of the final model in Step 5. Be sure to use proper notation consistent with your textbook and lecture notes.

Subset Model Selection using BIC

For purposes of consistency, you will be supplied with the models in the subset to be considered in the model selection. A two-step process is involved in the subset model selection:

- Using models with only random intercepts, evaluate the fit of the fixed effects models in the reasonable subset. (The reasonable subset is defined for you, but you would have to determine this in your own analysis.)

- Once the optimal model is identified from the last step, you will consider the same fixed effects as the optimal model but with additional random effects up to the highest order time polynomial.

Syntax for the first step is provide for you. You must provide the syntax for the second step. Be sure to use the REML solution for the second step (see the lecture notes).

```
#########################################################
#########################################################
##
## BIC Selection
##
#########################################################
#########################################################


####################################
## Step 1: Fixed Effects Models
####################################

## No Static Predictors:
> lmer.1 <- lmer(Reaction ~ 1 + Days + (1 | SubNum), data =
    sleep5)
> lmer.2 <- lmer(Reaction ~ 1 + Days + I(Days^2) + (1 |
    SubNum), data = sleep5)

## Static Predictors, Intercept:
> lmer.3 <- lmer(Reaction ~ 1 + Days + Female + (1 | SubNum),
    data = sleep5)
> lmer.4 <- lmer(Reaction ~ 1 + Days + GPA    + (1 | SubNum),
    data = sleep5)
> lmer.5 <- lmer(Reaction ~ 1 + Days + Female + GPA + (1 |
    SubNum), data = sleep5)

## Static Predictors, Time Interactions:
> lmer.6 <- lmer(Reaction ~ 1 + Days + Female + Days*Female +
    (1 | SubNum), data = sleep5)
> lmer.7 <- lmer(Reaction ~ 1 + Days + GPA + Days*GPA + (1 |
    SubNum), data = sleep5)
> lmer.8 <- lmer(Reaction ~ 1 + Days + GPA + Female +
    Days*GPA + Days*Female + (1 | SubNum), data = sleep5)
> lmer.9 <- lmer(Reaction ~ 1 + Days + I(Days^2) + Female +
    Days*Female + I(Days^2)*Female + (1 | SubNum), data =
    sleep5)
> lmer.10 <- lmer(Reaction ~ 1 + Days + I(Days^2) + GPA +
    Days*GPA + I(Days^2)*GPA + (1 | SubNum), data = sleep5)
> lmer.11 <- lmer(Reaction ~ 1 + Days + I(Days^2) + GPA +
    Female + Days*Female + Days*GPA + I(Days^2)*GPA +
    I(Days^2)*Female + (1 | SubNum), data = sleep5)
```

```
####################
## Fit Statistics:
####################
> fits <- c(
  BIC(logLik(lmer.1, REML = FALSE)),
  BIC(logLik(lmer.2, REML = FALSE)),
  BIC(logLik(lmer.3, REML = FALSE)),
  BIC(logLik(lmer.4, REML = FALSE)),
  BIC(logLik(lmer.5, REML = FALSE)),
  BIC(logLik(lmer.6, REML = FALSE)),
  BIC(logLik(lmer.7, REML = FALSE)),
  BIC(logLik(lmer.8, REML = FALSE)),
  BIC(logLik(lmer.9, REML = FALSE)),
  BIC(logLik(lmer.10, REML = FALSE)),
  BIC(logLik(lmer.11, REML = FALSE))
  )
F <- data.frame(1:11)
colnames(F) <- "Model"
F$BIC <- fits


################
# Compute delta:
################

> F$delta <- fits - min(fits)


################
# Sort by delta:
################

> F <- F[order(F$delta), ]


####################
# Plot delta by model
####################

> plot(x = 1:11, y = F$delta, type = "b", xaxt = "n", xlab =
    "Model", ylab = expression(Delta))
> axis(1, at = 1:11, labels = F$Model)
> abline(h = 2, lty = 2)
```

```
#######################################################
## Add random effects using fixed effects model from last
    step:
#######################################################

> # (Add your syntax here.)
```

5. In the subset analysis, the highest order polynomial is the quadratic polynomial. Why might you not include higher order polynomials (e.g., cubic)?

6. Explain in detail what this syntax does: `BIC(logLik(lmer.1, REML=FALSE))`.

7. Explain in detail what the plot in Step 1 indicates.

8. If the final model is different than the final model in the step-up analysis, then write its HLM using the labels from the data frame. If the models are the same, then simply state this is so.