**PSYC 5670: Multilevel modeling**

**Homework #3**

**Due 11/14/2019**

Dataset: external\_pp

Directions: Include all R code and relevant output in your submission

1) Import the data into R.

2) Prepare a table of descriptive statistics for the externalizing variable by time. Hint: use dplyr’s group\_by() and summarise() functions to calculate descriptive by time point, then print the results nicely with kable() or stargazer(summary=F).

3) Write down, in both two-level (decomposed) and reduced-form, the model equations for the unconditional growth model using TIME as the TIME index and EXTERNAL as the response variable. Explain the meaning of each term in the model.

4) Use ggplot to visualize the data, mapping TIME to x, EXTERNAL to y, ID to facets, and FEMALE to color. What do you notice about the plots? What results might you expect to obtain when you fit statistical models to this data?

5) Any pattern that might exist in the data is difficult to see because the male and female subjects are not ordered. We will produce two separate plots and display them side by side. First make a plot of only the male subjects and assign it to object p1. (You can use dplyr’s filter commend to select only males inside your ggplot call like this:

p1 <- ggplot(data=filter(df, FEMALE==0), aes(…)

Do the same for females and assign that plot to p2.

Then install the cowplot package. Use cowplot::plot\_grid(p1, p2) to display the plots side by side.

Hints: You will want to remove color from the aesthetic in each plot because the subsetted datasets will not contain any variation on FEMALE. You can color-code the fitted regression lines manually by adding the argument color=”red” to the geom\_smooth() call for the female plot and color=”blue” to the corresponding call for male plot. These arguments should not go inside aes() because you aren’t mapping a variable to color anymore, you are manually specifying the color for the fitted lines.

6) Fit the baseline multilevel growth model that you wrote down in #3 to the data and report and interpret the results.

7) Fit a model estimating the main effect of FEMALE on externalizing.

8) Fit a model testing the FEMALE by TIME interaction

9) Estimate the intercept and time slope for subject #3. (Note: you will need to combine that subject’s empirical Bayes residual with the marginal estimates of the intercept and slope (from the model fixed effects coefficients) to obtain these values).

10) Make a scatterplot of the EB intercepts on the x-axis versus the EB time slopes on the y-axis. Explain what this means and how this plot is related to the random effects estimates from the model output.

11) Make histograms of the EB intercepts and time slopes. Given that the multilevel model assumes that the random effects are normally distributed, is there any reason here to be concerned about an assumption violation?