MLM Nested Group Project Part 2ab (Spring 2020)

1. Refit the model in Part 1 that has all fixed effects as well as random intercepts (in schools and classrooms). Recall that `math1st = mathkind + mathgain` is the outcome. The model is `math1st ~ housepov+yearstea+mathprep+mathknow+ses+sex+minority+(1|schoolid/classid,REML=T)`
   a. Construct the residual that removes only the 'fixed effects' then subtract it from the outcome; call this residual: resFE
      i. R hint 1: `predict` has an option to generate the prediction based on the fixed effects only.
      ii. R hint 2: If you decide to add a column to your dataframe with resFE, note that predict only generates predictions for cases uses in the model *after listwise deletion*.
2. Show that this residual is not independent within schools in some manner.
3. a. Construct the residual that utilizes the BLUPs for the random effects using the R command `residuals`.
      i. Call the new residual resFE_RE.
4. a. Show that these new residuals, resFE_RE are MUCH LESS (if not completely un-) correlated within school, using the same method as before (boxplot?)
5. a. Generate the two sets of BLUPs (for random effects zeta0 and eta0)
   b. Examine these for normality (include evidence), and comment.
6. Returning to the classroom data.
   a. Fit a slightly more complicated model with the same fixed effects, but now add a random slope for minority, correlated with the random intercept, at the school level (keep the classroom level random intercept).

   b. Construct the residual (individual, level 1) and the BLUPs for the remaining random effects. Call the new residual resFE_RE as before.
   c. Examine all error estimates (individual level residuals, BLUPs (school and classroom level) for normality (and comment).
   d. Plot zeta0 vs. zeta1 to see whether the estimated correlation is consistent with the observed. Briefly comment. e.Track down those odd points in the scatterplot. What schools are they? Do they have anything in common? (you should comment)
7. Make a *person-period* file with math score (Kindergarten and First grade). That is, `math0 <- mathkind ; math1 <- mathkind+mathgain` (you have to make this work in the dataframe). Using reshape in R, you have to be careful to specify the name of the math variable (math0 and math1) as *varying*.
8. We ignore classrooms in this analysis, but keep it in the notation
   a. Fit a model with math as outcome, and fixed effect for time trend (year), and random intercepts for schools.
   b. Write down the model.
   c. Add random intercepts for child
   d. Write down the model.
9. Report original and new variance estimates for $\sigma^2_{\zeta_0}$ (between schools) and σ^2_ϵ (within schools):
   a. Compute a pseudo R^2 relating the between school variation and ignoring between students in the same school. In other words, what fraction of the between-school variance in the first model is 'explained' by the addition of a student random effect?
   b. Does the total variation stay about the same (adding between children within schools variance as well, to the second model results)?
10. Add a random slope (ζ_1) for time trend (year) within schools (uncorrelated with random intercept (ζ_0))

   a. Generate the BLUPs for the random effects and examine whether the independence between ζ_0 and ζ_1 is REFLECTED in a scatterplot of these two sets of effects.
   b. Compute V_S(year=0) and V_S(year=1). Since there are only two years, this is a form of heteroscedasticity in the random effects.
      i. In which year is there more between school variation, net of all else (year=0 or year=1)?

11. If you ran the model separately BY YEAR, and removed the year trend from the model, would you get the same estimates for the variance between schools? TRY IT

12. Rerun the last nested longitudinal model, allowing correlation between intercept and slope.
    a. Is the correlation signif.?
    b. Compute V_S(year=0) and V_S(year=1) for this new model (your formula should include covariance terms).
       i. Is this result (and thus model) more consistent with the separate grade analysis? You are implicitly testing model fit here.