36-402/608 Homework #11 due 10:30AM 4/8

1. Dyads (60 points)

   You must use SAS for this problem! Use the DDFM=SATTERTH option.

   This problem is a study of income in married couples in Massachusetts. Use this code to load the data in dyads.dat.

```
DATA dyads;
    INFILE "dyads.dat" FIRSTOBS=2;
    INPUT agentNum dyadNum Agender Acollege Aage Pgender Page Pcollege income;
    Aage30 = Aage-30;
    Page30 = Page-30;
RUN;
```

   Use only the adjusted age variables to obtain a more interpretable intercept.

   The explanatory variables are gender (1=female, -1=male), college (indicator variable of college graduate), and age (in years). "A" indicates "agent" and "P" indicates partner. First do some EDA to get familiar with the study variables. Then select an appropriate dyadic random effect model. Justify the need for a random per-dyad intercept. Don't try any random slopes (or serial correlation). Start with the largest possible fixed effects model without interactions. Remove unneeded fixed effects.

   Turn in a brief description of what you learned about the study from the EDA, the BIC values you used in your model selection, the output and SAS code for the final model including a residual plot and your interpretation of it, and a brief interpretation of the estimated parameters. Also state any one interaction that would be worth studying, and what a small p-value would indicate for that interaction.

2. Schools (40 points)

   You must use SAS for this problem! Use the DDFM=SATTERTH option.

   File fundses.dat has data from a study of the relationship between a test score and school socio-economic status and funding level. Schools were randomly selected, and one randomly selected 7th grade classroom was tested from each school. The variables are school id number, school funding level, school average SES, and individual test score in that order with one line per student. The funding is in thousands of dollars per student. Adjust the funding variable so that the intercept will correspond to \$10,000 per student. The per school SES variable is a z-score (mean 0, sd=1).

Run the mixed model with a random per-school intercept and fixed effects for the other explanatory variables without interaction and using REML. (You don't need to do EDA. Don't do model selection. You don't need to make residual plots.) Turn in the SAS code, the SAS output, and your interpretations of the parameters.