Project 2

Most student achievement models take into account three type of variables: Student ability or aptitude (usually a measure of previous achievement), student characteristics (e.g. home background, attitudes. etc.), and various combinations of teacher and classroom variables. In large scale surveys of educational achievement, a considerable number of teacher classroom, and school variables are typically available. The researcher then has the difficult task of selecting variables to represent dimensions of instructional process affecting student outcome.

The data that you will be working with in this project were collected as part of a study of mathematics (the Second International Mathematics Study) conducted under the auspices of the International Association for the Evaluation of Educational Achievement. The observations in the dataset are 8th graders from randomly sampled classrooms in the US who completed mathematics achievement tests at the beginning and at the end of the 1981-82 academic year. Students also answered questioners regarding their attitudes toward mathematics. Their math teachers completed questionnaires describing their teaching practice and experience, the type of resources they use and how they use them, their perception of students??? characteristics and ability as well as their perception of what makes teaching effective.

You can download the dataset (called `sims`) here.

```
download.file("http://stat.duke.edu/courses/Spring13/sta101.001/projects/sims.Rdata", destfile = "sims.Rdata")
load("sims.Rdata")
```

If you are interesting in learning more about the data, you can read about the Second International Mathematics Study here. Below is a description of the variables:

Identifier

1. `NO`: student ID

Student math ability

2. `POSAVG`: average posttest score on arithmetic, algebra, and geometry tests [response variable]
3. `PREAVG`: average pretest score on arithmetic, algebra, and geometry tests

Student background and attitudes

4. `YSEX`: gender of student (male and female)
5. `YFEDUC`: education level of student's father
6. `YMEDUC`: education level of student's mother
7. `YHOMLAN`: English used at home
8. `YMOREED`: expected number of years for continued education
9. `YGOWO`: "I can get along well in everyday life without math"
10. `YUSEDAY`: "Math is useful in everyday life"
11. `ETHNIC`: ethnic background

Teacher and classroom characteristics

12. `TEXPTSCH`: teaching experience of teacher (years)
13. `TEXPMTH`: math teaching experience of teacher (years)
14. `TEDMATH`: math credits (math training) of teacher
15. `COBJLOG`: "How much emphasis do you put on understanding the logical structure of proof?"
16. `COBJPRE`: "How much emphasis do you put on understanding the nature of proof?"
17. `COBJSCI`: "How much emphasis do you put on developing an awareness of the importance of math in basic and applied science?"
18. `CTMATH`: "How difficult a subject is math to teach?"

Opportunity to learn outside of school Opportunity to learn outside of school includes having a tutor, parental help, going to an after school program for homework help, etc.

19. `GEOMEOTL`: geometry learning opportunity
20. `ARITHOTL`: arithmetic learning opportunity
21. `ALGEBOTL`: algebra learning opportunity
22. `CLPOSAVG`: average class posttest score on arithmetic, algebra, and geometry tests
23. `CLPREAVG`: average class pretest score on arithmetic, algebra, and geometry tests

You are welcomed to use the data set as is, or manipulate the variables (collapsing levels, etc.), or create new variables as you see fit.

Poster & presentation

Your results will be displayed on a poster during your scheduled lab time on Monday, April 22. Each group will have 4 minutes scheduled during which all group members should be at the poster to give a formal presentation. All group members must contribute to this presentation. Note that this is too short a time to go through in detail all of your analysis and findings. Your poster should be showcasing those. The presentation should only highlight what you think are most interesting findings, or something that you believe may be unique about your approach.

For the rest of the poster session, you will be wandering around, assessing the other group???s posters and asking questions. The lab will be broken up so each person will spend 1/(# of members in team) of the remaining time by your own poster to answer questions. This is part of your grade, so every group member should feel comfortable answering questions regarding all aspects of the project.

Paper

Maximum of 12 pages (including figures, etc.). Due by Wednesday, April 24, at 5pm. Hard copy and submission on Sakai. Be conservative in what you include in your paper, 12 pages is not very long.

```
download.file("http://stat.duke.edu/courses/Spring13/sta101.001/projects/prj2.Rmd", destfile = "prj2.Rmd")
```

Components of project

You can focus on whatever aspects of the data you find interesting, but as a minimum you must include each of the following components:

1. **Introduction:** Explain your data, including implications for the scope of inference.
2. **Univariate analysis:** Analysis of variables of interest, one at a time
3. **Bivariate analysis:** Exploring and assessing relationships between two variables at a time (response vs. explanatory or explanatory vs. explanatory) using appropriate methods
4. **Multiple regression:**
   (a) Decide on a "best" model for predicting the response variable. You do not need to explain every step you took to arrive at this model, but should give some indication of why you chose the model you did.
   (b) For your best model, are the conditions met? If not, what are the implications?
   (c) Using your best model, obtain a predicted value for your response variable for a "typical" student. (You should define what you consider "typical" and why.)
5. **Conclusion**

After providing the description of your dataset and research question in the introduction you must apply what you have learned about descriptive statistics, graphical methods, correlation and regression, and hypothesis testing and confidence intervals to your dataset. The goal is not to do an exhaustive data analysis i.e., do not calculate every statistic and procedure you have learned for every variable, but rather let me know that you are proficient at using R at a basic level,

and that you are proficient at interpreting and presenting the results. Focus on methods that help you begin to answer your research question. Also pay attention to your presentation. Neatness, coherency, and clarity will count.

Your write up must also include a roughly one page conclusion and discussion. This will require a summary of what you have learned about your research question along with statistical arguments supporting your conclusions. Also critique your own methods and provide suggestions for improving your analysis. Issues pertaining to the reliability and validity of your data, and appropriateness of the statistical analysis should be discussed here. A paragraph on what you would do differently if you were able to start over with the project or what you would do next if you were going to continue work on the project should also be included.

Grading

Grading of the project by the professor and TAs will take into account the following:

- Content - What is the quality of research and/or policy question and relevancy of data to those questions?
- Correctness - Are statistical procedures carried out and explained correctly?
- Writing and Presentation - What is the quality of the statistical presentation, writing and explanations?
- Creativity and Critical Thought - Is the project carefully thought out? Are the limitations carefully considered? Does it appear that time and effort went into the planning and implementation of the project?

Your grade for the project will be based on the following components:

- 40% Poster (poster, presentation, answers to questions)
- 40% Paper
- 10% Classmates??? grades (based on poster session)
- 10% Team member evaluation

Team Member Evaluation: All of the other components are common to the entire group. This is your chance to be rewarded (or punished) for contributing (or not contributing) to the group. You will rate each of your group member???s contribution to the project on a scale of 0 - 5. You will receive an email with a link to a Qualtrics poll to submit your evaluations, these must be submitted by Wednesday, April 24, at 5pm, as well. These ratings are anonymous to your teammates – please be honest. The average of these scores will be your points for this 10%. For grades less than 3, please provide some explanation. If any individual gets an average peer grade less than 1, this person will receive half the grade of the rest of the group.

A general breakdown of scoring is as follows:

> 90%-100% - Outstanding effort. Student understands how to apply all statistical concepts, can put the results into a cogent argument, can identify weaknesses in the argument, and can clearly communicate the results to others.
>
> 80%-89% - Good effort. Student understands most of the concepts, puts together an adequate argument, identifies some weaknesses of their argument, and communicates most results clearly to others.
>
> 70%-79% - Passing effort. Student has misunderstanding of concepts in several areas, has some trouble putting results together in a cogent argument, and communication of results is sometimes unclear.
>
> 60%-69% - Struggling effort. Student is making some effort, but has misunderstanding of many concepts and is unable to put together a cogent argument. Communication of results is unclear.
>
> Below 60% - Student is not making a sufficient effort.

Honor Code:

You may not discuss this project in any way with anyone outside your group, besides the professor and TAs. Failure to abide by this policy will result in a 0 for all groups involved.