

## MINI-PROJECT

**Goal:** To give you the opportunity to dig deeply into concepts, procedures and interpretations that we have covered in class or in homework; or to consider something that we will not have time to cover.

**Your Project:** Prepare a written report that clearly communicates the framework for your work, what you have done, what you have found and a discussion (background, methods, results, discussion). Projects can have either an applied or a methodologic emphasis, but in either case must go beyond what you have prepared for this class, other classes, or for other purposes. Applied-emphasis projects consist of reading some background literature on an application, conducting and reporting data analyses using linear models. Methodologic-emphasis projects consist of reading some background literature and conducting a simulation, a mathematical analysis or a combination thereof.

**Report Specifications:** Your report must be in pdf format, fully double-spaced (main body, references, footnotes, appendices), using an 11 point, sanserif font (e.g., Arial) with one inch margins all around. The text content of your report should be between 6 and 8, double-spaced pages. Taken together, the number of tables and figures should not exceed 7. All tables and figures should have informative captions. Relevant computer output must be annotated and consolidated into the report. You may need to take advantage of the scanning option on the copying machine near my office.

**Requirements:** Email me a short proposal by **Thursday, February 28, 2008**. I will review it to ensure it is sufficiently educational and do-able in the short time available. I'll provide feedback and may ask that we meet to discuss it. Projects are to be your **individual work**. However, you may consult with Bruce and other students, with me other faculty (Please don't overly burden other faculty!).

**Datasets:** Any of the datasets on the course website can be used. Use of other datasets is allowed, subject to all of the following conditions:

- You must have permission to use the dataset.
- The dataset must immediately ready for computer analysis.
- You must be familiar with the application and the dataset.
- All IRB/HIPPA requirements must be met.

**Some Possibilities:** These are possibilities; don't be limited by them.

- Analyze a reasonably complicated dataset and identify your two best models. Clearly report and discuss your methods and results.
- Compare results for a variety of working covariance structures (independence, and others that you pre-specify). Compare estimates, reported and robust SEs **and** diagnostics (high leverage and high influence points).
- Read up and report on PRESS,  $C_p$  and  $R_{adj}^2$ . Analyze a dataset and compare approaches.
- Design a simulation study to assess the over-optimism with regard to prediction of a future dataset of models selected by all subset regression with the model selected by PRESS, or  $C_p$  or  $R_{adj}^2$ .
- Design a simulation study of the performance of LSEs when the residuals are not Gaussian or when the model is incorrectly specified. Also, develop the mathematics to explain some aspects of your results.

**Deadline:** Your report must be submitted electronically to my email address (tlouis@jhsph.edu) as a pdf attachment **no later than midnight, Monday, March 17, 2008**. **Hard-copy submissions will not be accepted**. I will acknowledge receipt of your report. If you do not receive an acknowledgement within 24 hours of submitting, resubmit and leave a voicemail at 410-614-7838.