University of California, Davis  
Department of Statistics  

Spring 2011  
(Due Thursday, June 9)  
(with your in class final exam)  

Statistics 108  
Project  

The purpose of this project is to use the applied linear model techniques that we have learned in this course to analyze a real-life data and prepare a professional quality report. You may select any data set of your interest, which includes two or more explanatory variables. The following data set is a typical data that you may use for your report. Select as many variables as you wish to perform a regression analysis of the data. You can download this data from the course website.

A real estate expert is interested in analyzing the selling price and characteristics of suburban residential properties. Her interest lies in a new, large residential property development on the outskirts of a major city for which she has data on 30 properties that were sold recently. She developed a long list of possible explanatory variables including lot size, property taxes, assessed property value, various house characteristics (such as floor area, number of rooms, style of construction) and traffic access. After carefully sifting through a long list, she decided to include seven variables (given in homes.xls) in the pool:

- Property taxes (annual taxes, in dollars)
- House size (floor area, in square feet)
- Lot size (in acres)
- Attractiveness index
- Style (E, S or M)
- Selling price

For the selected data explain the problem of interest, give details of your response and explanatory variables and provide the source of the data. Select the ‘best’ model fitting the data using a model selection criterion. Test some hypotheses about the final model or compute some simultaneous confidence intervals for the regression coefficients. Comment on the fit of your final model based on the adjusted R². Find confidence interval(s) for the mean response or prediction interval(s) for certain level(s) of the explanatory variables and discuss the findings. Perform a complete residual analysis and test for multicollinearity. Write a four-page report (no appendix, no raw computer output) describing the results of your analysis. Your report may include the following sections:

- **Introduction:** Statement of the problem.
- **Material and Methods:** Description of the data and methods that you have used for the analysis.
- **Results:** Write about the results of your analysis. You can include tables, charts and graphs in this section and refer to those in explaining your results.
- **Conclusion and Discussion:** Highlight the main points and discuss.