STATISTICS 210 – Fall 2009

Final Exam: Take-home Data Analysis Question

- This is an exam. Please work independently. Do not ask your classmates questions, ask me instead. You can send email (sternh@uci.edu) or phone my office (949-824-1568) with questions.
- This data analysis is considered part of the final exam; it is worth 15% of the final exam grade.

EFFECTS OF SMOKING ON LUNG FUNCTION IN CHILDREN

Background. The effects of cigarette smoke exposure include impaired lung function and increased risk of heart disease and cancer. This is true for both active smoking by an individual or for passive smoke from a housemate or relative. The data used in this exam come from a series of studies that helped to establish the effect of passive smoke on children by following a group of children over time (a longitudinal study). We do not consider the longitudinal data; instead we focus on data from 654 children ages 4 to 19 collected at a single time point. Our goal is to determine the effect of cigarette smoking on lung function in children. We expect smoking will reduce lung function. A test measuring lung function is performed on each child. One difficulty in assessing the effect of smoking in children is that lung function develops as children grow. Since older children are more likely to smoke than younger children the age of a child could be a confounding factor. Experts also indicate that it might be physical measures (like height) that are more indicative of child development than chronological age.

The data are available on the course website in the file lung.txt / lung.xls / lung.csv. A list of the variables along with short descriptions is provided in the table below. There are no missing values which is quite remarkable for a real data set! (This likely means somebody cleaned the data before I got it.)

VARIABLE LIST		
Column	Name	Description
1	ID	numerical identification code for child
2	Age	age of child (years)
3	Lung	measure of lung function (liters)
4	Height	height of child (inches)
5	Sex	gender of child (Male/Female)
6	Smoker	current smoking status of child (Non/Current)

Your Assignment. Analyze these data to identify the effect of smoking on lung function. There is interest in knowing whether the effect of smoking is purely additive, or whether it depends on the level of other factors (e.g., age, height, sex). You should produce a short 2-5 page report (2-5 pages of text) describing your findings. The report should include a brief introduction to the problem, a discussion of your data analysis, and a summary addressing your main findings and any limitations of the study. The data analysis section should briefly describe the steps that you took in developing your model (i.e., preliminary data examination, important regression findings, relevant diagnostics, etc.) and important findings. You should include any tables/figures that you use to reach your conclusions (e.g., regression results, residual plots or diagnostics) but you do not need to include output for models that turn out to be irrelevant to your analysis. You should discuss any unusual observations and their impact on the results as well. Remember the report should integrate tables/output you want me to see. Additional output can be included as an appendix. **Remember:** I know that you have other exams and am not expecting you to spend the entire week on this data analysis.

SAS Hints: Remember that you put a $ after a variable name to identify that it is a character variable (like "Sex" and "Smoker" in the current data set). You will want to create indicator variables for these variables. You can do this by adding lines to the DATA step as follows:

```
smoke = 0;
if smoker="Current" then smoke = 1;
```