Statistics 200B
SPRING 1997

Lab VI: Babies and Smoke (Part II)

In the first set of investigations, you may have found some relationships between smoking and babies' weights. But smoking is not the only variable that has impact on babies' weights. In this lab we will do some regression analysis to see the predicability of birth weight by smoking and other variables.

1. Make several scatterplots of pairs of the variables. Do you see any variables that may help to explain the variation in birthweight?

2. Use your findings to fit linear models with birth weight as the response variable. What is the single best predictor for birth weight? Check your model assumptions with residuals from your fit. Do you see any pattern in the residuals? Do the linear models seem appropriate for the data?

3. Try a linear model that includes the indicator variable for whether the mother smokes. Provide an interpretation for the coefficient for the indicator variable. Diagram the design matrix. How do your findings compare to the Surgeon General's statement about smoking being "a more significant determinant of birthweight than the mother's prepregnancy weight, height, ..."?

4. For the model that includes gestational age and an indicator for whether the mother smokes, we see that holding gestational age constant, smoking is associated with some reduction in birth weight. Can we conclude then that smoking causes reduced birth weight? Do you think that smoking causes reduced birth weight?

Summarize your conclusion or findings from the above analysis. For all models you fit, please clearly state your model formulations, assumptions and implications. Attach diagnostic plots for each of them. Which one (ones) of them do you think "the best"? For what reason?