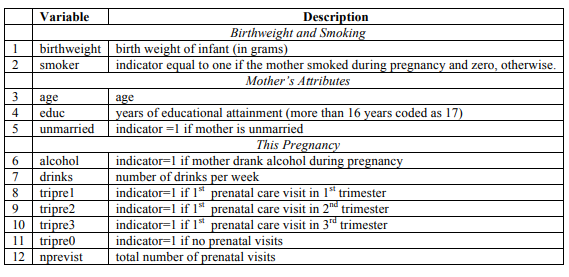
# PB4A7 Seminar 5

***About the Dataset:***

The datafile Birthweight\_Smoking is from the 1989 linked National Natality-Mortality Detail

files, which contains a census of infant births and deaths. The data in bw\_smoking.data are for

births in Pennsylvania in 1989. These data were provided by Professor Douglas Almond, Kenneth Chay, and David Lee and are a subset of the data used in their paper “The Costs of Low Birth Weight,” Quarterly Journal of Economics, August 2005, 120(3): 1031-1083. The file contains 3,000 observations on the variables described below:



Exercise 4 questions:

An important problem facing health officials is estimating the effect of cigarette smoking on infant health. One measure of infant health is birthweight: a very low birthweight can put an infant at risk of contracting various illnesses.

1. Write out a model via a linear equation that represents a regression model that will estimate the effect of smoking on infant birthweight. What is the likely sign for ?
2. Estimate the model. What is the estimated effect of smoking on birthweight?
3. Are there factors other than cigarette smoking that affect birth weight and are likely to be correlated with smoking. List 2 such factors and explain why may this affect the causal interpretation in our first model.
4. Regress birthweight on smoker, alcohol and nprevist. Is the estimated effect of smoking substantially different?
5. Are there any changes to the R-square of the regression model?
6. How should you interpret the coefficient on Nprevist? Does it measure the causal effect of prenatal visits on birthweight?
7. Add in tripre0-3 in the regression model. What happens?