Problem Set 3

(Submit through the Hub by 12pm November 2nd)

1. Suppose you collect data from a survey on wages, education, experience, and gender. In addition, you ask for information about marijuana usage. The original question is: ``On how many separate occasions last month did you smoke marijuana?"

   (a) Write an equation that would allow you to estimate the effects of marijuana usage on wage, while controlling for other factors. You should be able to make statements such as, ``Smoking marijuana five more times per month is estimated to change wage by x%."

   (b) Write a model that would allow you to test whether drug usage has different effects on wages for men and women. How would you test that there are no differences in the effects of drug usage for men and women?

   (c) Suppose you think it is better to measure marijuana usage by putting people into one of four categories: nonuser, light user (1 to 5 times per month), moderate user (6 to 10 times per month), and heavy user (more than 10 times per month). Now, write a model that allows you to estimate the effects of marijuana usage on wage.

   (d) Using the model in part (c), explain in detail how to test the null hypothesis that marijuana usage has no effect on wage.

   (e) What are some potential problems with drawing causal inference using the survey data that you collected?

2. Use the data in nbasal.RData for this exercise.

   (a) Estimate a linear regression model relating points per game to experience in the league and position (guard, forward, or center). Include experience in quadratic form and use centers as the base group. Report the results (including SRF, the sample size, and R-squared).

   (b) Holding experience fixed, does a guard score more than a center? How much more? Is the difference statistically significant?

   (c) Now, add marital status to the equation. Holding position and experience fixed, are married players more productive (based on points per game)?

   (d) Add interactions of marital status with both experience variables. In this expanded model, is there strong evidence that marital status affects points per game?

   (e) Estimate the model from part (c) but use assists per game as the dependent variable. Are there any notable differences from part (c)? Discuss.

Important: Please also submit the relevant portions of your log file (delete errant commands and output).