BMI 826 Causal Inference
Homework 2
Due Monday, Feb 24, in class,

1. **NHEFS example from H&R: consider complete data only**

   (a) Use logistic regression to estimate a propensity score for quitting smoking. You do not have to use the model suggested in the book. Examine if there are some overlap between the propensity scores of the two groups (smoking and non-smoking) based on the model. Include the summary for your logistic regression and the overlap plot for your logistic regression.

   (b) Stratification: classify individuals in 5 strata of approximately equal size, what is the estimated effect of smoking cessation on weight gain? How about 10 strata?

   (c) With careful model selection, sometimes it is possible to create a logistic regression that perfectly predicts group membership (for example, people with A = 1 have high propensity scores and people with A = 0 have low propensity scores). This would seem like a very effective logistic regression model but is not useful for causal inference. Explain why a result of this type is problematic for a propensity score analysis.

2. **Project.** You can work alone or in a group of two. You can: (1) focus on a topic that we dont cover; (3) review the literature in an area where causal arguments may play a role; (3) carry out an analysis of data from an observational study using methods in the class; or (4) conduct a research project. For this assignment your responsibility is to identify your partner (if any), list one (or more) possible topics, and write a paragraph on what you are thinking about for the topic(s). (I say one or more topics in case you want feedback on more than one idea).