1. Consider now an extended version of our Amazon Echo example from lectures 6 and 7. We again want to know the impact of purchasing an Echo on the likelihood a household subscribes to Amazon Prime.  Consider now the data from the first spreadsheet in the Problem Set 3 data file ([Prob Set 3.xlsx](https://iu.instructure.com/courses/1944230/files/112402668/download?wrap=1)[download](https://iu.instructure.com/courses/1944230/files/112402668/download?download_frd=1)).  Here, we have data on household income, education, size, and age.
   1. Use a linear probability model to generate propensity scores for these data, using all demographic information.
      1. What is the mean of the propensity scores you generated?
      2. What is the mean propensity score for each treatment status (Echo owner vs. non-owner)?
   2. Using a probit model for matching and focusing on ETT, what is the estimated impact of owning an Echo on subscriptions?
   3. Is there any evidence that there are still meaningful differences between Echo owners and non-owners left unaddressed by the propensity score matching?
   4. When using matching to get causal effects, if there is evidence that there are still meaningful differences between Echo owners and non-owners left unaddressed by the propensity score matching, what key assumption does this violate?
   5. Suppose our matching method was to take all of the untreated households (i.e., Echo non-owners), and match them with similar households who were owners. We then measure the difference in their subscription rates for Amazon Prime.  Given the matching method we’ve discussed in class measures the ETT, what, in words, would this alternative approach measure?

1. Consider now the Yelp example discussed briefly in lecture 8. The data from the second spreadsheet in the Problem Set 3 ([Prob Set 3.xlsx](https://iu.instructure.com/courses/1944230/files/112402668/download?wrap=1)[download](https://iu.instructure.com/courses/1944230/files/112402668/download?download_frd=1)) data file contain information on Sales, average Yelp rating, and Yelp stars for a number of comparable restaurants (say, mid-level-priced, American food).
   1. In the data, what is the:
      1. Outcome?
      2. Forcing variable?

* Treatment variable?

1. Execute regression discontinuity to determine whether a change from 1 star to 2 stars impacts sales. What do you conclude?
2. If you widen the bandwidth of your regression discontinuity estimator, what are the consequences?
3. Execute regression discontinuity for changes from 2 to 3 stars, 3 to 4 stars, and 4 to 5 stars. What are your findings?
4. Perform a validity check of your data, and explain what you find.