PS C236A / Stat C239A
Practice Midterm

Instructions

This is an ungraded practice exam. The following instructions outline the expectations for the upcoming midterm.

The questions below will be graded as follows: True/False (I) 10%, analytical section (II, III, IV) 40%, and the empirical section (V) 50%. You may submit the analytical portion of the midterm either electronically or in paper form. Electronic versions must be sent as a .pdf to <jahenderson[at]berkeley.edu>. Paper copies should be placed in my mailbox in 210 Barrows. For the computing portion, you **must** submit a fully executable version of all .R code, along with any data used in the code to the email above. If you do not send an electronic version of your .R code, that portion of the midterm **will not** be graded. All files sent electronically should be included in one omnibus email, with the subject line containing the course number and your last name (e.g., PS239A/STAT236A: Midterm - Norquist).

Note: This exam is open book. However, during the exam, you are not allowed to communicate or cooperate with anyone in any way about the exam. Any questions should be asked directly to the Professor or the GSI. To repeat: you may not use study groups, online help forms, the writing center, or any other form of external help. If in doubt, ask.

I. True or False

Answer *True* or *False*. Explain your answer in a sentence or two.

1. A treatment was randomly assigned to a population. A researcher is investigating the impact of this treatment on an outcome Y, and she proceeds to estimate the average treatment effect (ATE) by:

   $$\widehat{ATE} = \frac{1}{N_1}\sum_{i=1}^{N_1} Y_{i1} - \frac{1}{N_0}\sum_{i=1}^{N_0} Y_{i0}$$

   Here, Y_{i0} is the observed outcome for unit i in the control group, Y_{i1} is the observed outcome for unit i in the treatment group, N_0 is the number of units in the control group and N_1 is the number of units in the treatment group. Even though she knows that $\widehat{ATE}$ is a consistent estimator of the ATE, she decides to verify this by running OLS regression. She estimates the model $Y = \alpha + \beta T + \epsilon$, (where T is the treatment dummy) and she finds $\hat{\beta} = \widehat{ATE}$. This is evidence that the randomization worked well, since OLS recovered the experimental benchmark.

2. A researcher is analyzing the effect of a treatment in a randomized experiment and uses a two-sample t-test (with unequal variances) to reject the null hypothesis of a 0 average treatment effect. The researcher could have tested the same null hypothesis with a randomization (permutation) test and his inference would not depend on any large sample approximations.

3. A group of researchers begin a drug trial to study the effectiveness of a particular psychostimulant in treating Attention Deficit Hyperactivity Disorder (ADHD). At time t, people in the study are randomly assigned either to treatment and receive the drug, or to control and receive a sugar pill placebo. Since compliance is a common problem in drug trials, the researchers included an additional intervention at $t+1$ aimed to increase compliance rates. In this second part, half the subjects were randomly assigned to an 'encouragement' condition, where they were counseled on the importance of taking their assigned pill dosage – the other half of the subjects received no such encouragement. Since both the drug and encouragement interventions are randomly assigned, it is

generally valid (without additional assumptions) to estimate the Intention-to-Treat (ITT) effect of the drug on changes in behavior at $t+2$, by pooling all people in the drug arm to measure the average outcome for the treated, and pooling people in the placebo to measure the average outcome for the controls, with the estimate of the ITT being the difference between these two averages.

II. Sample Selection

A political scientist wants to estimate the personal incumbency advantage. Let $Y_i(1)$ be the vote share in the next election of candidate i if he or she wins in the present election, and $Y_i(0)$ is the vote share of candidate i in the next election if he or she loses in the present election. Define incumbency advantage as:

$$E[\delta] = E[Y_i(1) - Y_i(0)]$$

Let D_i be an indicator (treatment) variable for whether or not candidate i wins election. If all candidates re-run in the subsequent election, then the political scientist would observe $Y_i(1)|D_i = 1$ and $Y_i(0)|D_i = 0$.

Unfortunately, not all candidates re-run in subsequent elections. Let R_i be an indicator variable for whether or not candidate i runs for office in the subsequent election. Thus, the political scientist can observe $Y_i(1)|(D_i = 1, R_i = 1)$ and $Y_i(0)|(D_i = 0, R_i = 1)$, but not $Y_i(1)|(D_i = 1, R_i = 0)$ and $Y_i(0)|(D_i = 0, R_i = 0)$.

Denote the proportion of treated units in the population as π and the proportion of candidates running for office in the next election as λ. Assume a very large sample so that sampling error is negligible.

a. The political scientist naively estimates the incumbency advantage using the following estimator:

$$\hat{\delta}_{naive} = E[Y_i|(D_i = 1, R_i = 1)] - E[Y_i|(D_i = 0, R_i = 1)]$$

   This is simply the average difference in vote shares between the winners and losers *who run again* in the following election. Without making any assumptions, if the estimand is the average treatment effect ($E[\delta]$) what is the bias in this estimator? Be sure to account for the bias resulting from "fundamental missingness" (unobservability of the counterfactual conditions) as well as the bias resulting from candidates not always rerunning.

   Hint: Decompose $E[Y_i(1)]$ as the weighted average of four causal types, based on their potential outcomes under treatment and control and whether or not they re-run. Do the same for $E[Y_i(0)]$.

b. Assume $(Y_i(1), Y_i(0)) \perp D_i$. With this assumption, what is the bias in the naive estimator? Under what conditions would the bias be 0?

c. Assume that incumbency advantage is bigger among candidates who re-run than those who do not re-run. Under this assumption, as well as the independence assumption made in part (b), what is the largest possible value of $E[\delta]$? What is the smallest?

III. Regression Discontinuity

Imagine a study where a treatment group of unemployed workers in San Francisco are given the opportunity to participate in a worker training program. Six months after the treatment is administered, the workers' employment status and yearly income is measured. You are asked to evaluate the effect of this program.

Please be explicit about your assumptions, how you would make your inferences, and the workers to which your inferences would apply. In your answers below, use mathematical notation where appropriate.

a. Suppose that the randomization to treatment occurred as follows: a randomly generated number X is drawn from a uniform distribution with the range [0, 4]. Units with $X \geq 2$ are given the treatment while units with $X < 2$ are denied treatment. All workers assigned to treatment are forced to attend the training program. Under this setup, what inferences could you make about the effect of the training program on the workers' employment status and income? What assumptions are required?

b. Now imagine that for ethical reasons, workers are compensated for having received a "bad draw" by being awarded monetary compensation inversely proportional to the random number X. So workers with a $X \approx 0$ receive a large sum of money and those with $X \approx 4$ receive very little. The workers are enrolled in the worker trainer program if $X \geq 2$. Under this setup what inferences could you make about the workers training program? How would you make these inferences?

c. Suppose the same set-up as in part (b) (including the compensation), except workers with $X \geq 2$ flip a coin – those who flip heads are enrolled in the worker trainer program, and those who flip tails are not. The value of X is observed, but it is not possible to know whether or not a worker actually participated in the program. What assumptions need to be made in order to bound the estimand in (b)?

IV. Media Bias

For this problem, you will compare the research design from two papers studying the effects of media bias on political attitudes and choices. The first paper is "The Fox News Effect", by Stefano DellaVigna and Ethan Kaplan (DVK), and can be found here `http://sekhon.berkeley.edu/causalinf/papers/DellaVignaFoxNews.pdf`. The second paper is "Exploiting a Rare Shift in Communication Flows to Document News Media Persuasion", by Jonathan Ladd and Gabe Lenz (LL), and can be found here `http://sekhon.berkeley.edu/causalinf/papers/LaddLenzBritish.pdf`.

Please write a page or two addressing the following questions:

a. Compare the identification strategies of the two papers. Does LL share similar weaknesses as DVK? Similar strengths? Do you find LL more or less convincing than DVK?

b. One potential issue in DVK is that the effect is measured in the aggregate at the township level. Do you think addressing the selection problem in DVK would be improved by analyzing individual- rather than township-level data? Does the individual-level data analysis in LL influence your judgement about the strengths or weaknesses of LL relative to DVK?

c. Which paper do you find more interesting, weighting both the scope and significance of the effect being estimated, as well as the *external* and *internal* validity of the respective estimates? Generally speaking, which study is more informative about the substantive impact of media bias on public opinion or vote choice?

V. Data and Matching

For this problem, you will perform several matching exercises using the "Fox News Effect" data. The unit of observation are towns in the US, and the treatment under study is the availability of Fox News during the 2000 election season. The outcome (`reppresfv2p00m96`) is the change in the Republican presidential vote share between 1996 and 2000. The dataset for this assignment only includes those towns with pre-treatment outcome data, i.e. the change in the Republican presidential vote share between 1988 and 1992 (`reppresfv2p92m88`). The treatment indicator (`foxnews2000`) has been defined as equal to one if the town's cable system carried the Fox News network before the 2000 election. The dataset includes a set of demographic covariates from the 2000 and 1990 census.

The Fox news data is available here: `http://sekhon.berkeley.edu/causalinf/data/hw6data.RData`. The variables are described in the following file: `http://sekhon.berkeley.edu/causalinf/data/hw6_codebook.txt`

For parts (a) - (e) below, be sure to explicitly set seeds to ensure that GenMatch recovers reproducible results, i.e. `set.seed` in general, and in GenMatch `unif.seed`, `int.seed`.

a. Estimate the causal effect for the treated of a town carrying Fox news on the change in Republican presidential vote share between 1996 and 2000. In doing so, select a set of covariates to condition on, being sure to include higher order terms and interactions you think are appropriate. Also include a propensity score when conditioning, and "orthogonalize" your other covariates using this propensity score. Report your balance statistics before and after matching using `MatchBalance`. Are these effects significant? What is the most interesting summary statistic when comparing change in Republican vote returns? How informative are mean differences? What are the mean differences?

b. Create a loss function in GenMatch that ensures that the function will not return a matched data set with worse balance on any variable in your BalanceMatrix than the balance obtained by matching on *just* your propensity score in part (a) – as judged by eQQ-plots and difference of means. Do this so that this property holds by design – i.e., it holds regardless of the dataset used. (*Hint*: To do this, you will have to both write a custom loss function and provide GenMatch with starting values for the covariate weights so that it begins with the matched dataset returned from using only the propensity score above.) Match again on your orthogongalized covariates from above using this loss function. Present balance before and after matching using `MatchBalance`.

c. Now match using the method from part (b) only using demographic covariates. Estimate the "treatment effect" of the introduction of Fox news prior to the 2000 election on the pre-treatment outcome of change in Republican presidential vote share between 1988 and 1992. This is known as a "placebo test". Can you recover a 0 ATT estimate using only demographic covariates as the conditioning set?

d. Overall, how do your results differ from those in DellaVigna and Kaplan (2007)? Are your results and their results comparable?

e. [BONUS QUESTION] Freed of the constraints in the previous parts, find the best matching method (and possibly post-matching adjustment model) to answer the substantive question at hand. How confident are you that this is an unbiased estimate of the Fox news effect? What do we learn about the effects of media bias from this analysis?