PS C236A / Stat C239A
Midterm Exam
Due: December 4, 2012

Instructions

This exam is due at the beginning of class (2:10 pm) on Tuesday, December 4. 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 **must** submit your midterm answers in paper form to class. This material should include all .R output, figures, and tables needed to answer the computing portion of the exam. We will not read computer code to find your answer, however, you **must** submit a fully executable version of all .R code to <`jahenderson[at]berkeley.edu`>. 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 - Rice).

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 three papers studying the effects of media bias on political attitudes and choices. The first paper is "The Fox News Effect", by S. DellaVigna and E. 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 J. Ladd and G. Lenz (LL), and can be found here `http://sekhon.berkeley.edu/causalinf/papers/LaddLenzBritish.pdf`. And the third paper is "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions", by A. Gerber, D. Karlan, and D. Bergan (GKB), and can be found here `http://sekhon.berkeley.edu/causalinf/papers/GerberNewspapers.pdf`.

Please write a page or two addressing the following questions:

a. Compare the identification strategies of the three papers. Which strategy do you find the most convincing? The least? Why?

b. Given the different types of interventions being studied (e.g., biased media exposure v. change in media bias, television v. newspaper media, etc), in what sense are the findings across these three studies 'comparable'? Do these studies give us useful information to test the same theoretical claim or different theoretical claims?

c. Imagine at a future point in time, Fox News expanded to every major cable and media market in the US. Would we expect to see a similar media effect as measured by DVK as a result of this national expansion? Why or why not? What would be an analogous type of issue in the studies conducted by LL and GKB? Do any of the three studies seem more robust to this issue than the others?

d. Which paper do you find the most 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 Ladd and Lenz's "Exploiting a Rare Shift" data. The unit of observation is the individual respondent in a UK election survey, and the treatment under study is whether an individual is a reader of a newspaper that switched its party endorsement from Tory to Labour in the run-up to the 1997 election. The main outcome is change in Labour party vote support between 1992 and 1997. To control for confounding, the authors condition on a number of covariates (listed in Table 3 and Table 1A of their paper) that may predict both readership and party voting behavior.

The Ladd and Lenz data is available here: `http://sekhon.berkeley.edu/causalinf/data/midterm.dta`. The variables are described in the following file: `http://sekhon.berkeley.edu/causalinf/data/midterm_codebook.xsls`

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 of being a typical reader of a newspaper that switched party endorsement (from Tory to Labour) on the *change* in Labour party vote support between 1992 and 1997. In doing so, select a set of relevant covariates to condition on. In matching, first use a custom loss function and then use GenMatch's default loss function. Provide some justification for your custom loss function. Choose the matched dataset with the best balance on the relevant covariates. Are the media effects on voting you estimate significantly different from zero? What are the mean differences in change in party vote suport you recover after matching? What are the three *worst* balanced covariates in this best-matched dataset? What are the standardized mean differences across matched treated and control on these three covariates?

b. Using the best-matched dataset from part (a), stratify the matched-pairs to include only those treated individuals who also are also habitual readers of a newspaper that switched its party endorsement. Check balance on this 'stratified' dataset using `MatchBalance`. Does balance change considerably in this dataset, compared with that recovered in (a)? Now, use GenMatch to match on the same covariates used in (a), utilizing habitual readers as the treatment indicator. Does balance in this matched dataset improve compared to the 'stratified' matched data? What media effects do you recover in these two matched datasets? Are these different from that found in (a)?

c. Choose one matched dataset from (a) or (b) that you think is the most convincing in recovering conditional exchangeability (for either habitual or typical readers), and conduct two robustness tests of the conditional exchangeability assumption. The first robustness check should be a Rosenbaum sensitivity test using the `rbounds` package in R. The second robustness check either should be a post-matching parametric bias adjustment on the matched data (e.g., a probit regression including covariates and treatment to model the outcome on the matched data), or a placebo test of the effect of treatment on a prior party vote outcome before and after matching. What is the Γ magnitude of confounding due to an unobserved covariate in the Rosenbaum sensitivity test at which the estimated treatment effect is indistinguishable from zero? How does this Γ compare to the imbalance recovered in the best-balanced dataset in (a)? Are these robustness tests convincing that conditional exchangeability holds?

d. Repeat the analysis in part (a), this time using the *level* of Labour party vote support in 1997 (rather than change in vote support). Is this estimate consistent with the one recovered in (a)? Is this causal estimate more or less persuasive than the difference-in-difference estimate you recovered in (a)? Overall, what do we learn about the effects of media from this analysis?

e. [BONUS QUESTION] Fully replicate the regression analysis in Table 1A (excluding the 1992 instrument column), on both the *level* and *change* in party vote support. That is, do the bivariate analysis, the exact matching on the same covariates used by Ladd and Lenz, and the GenMatch analysis on the same coveriates, and also perform linear adjustment on each matched data set. Can you replicate the table exactly? If not, which parts can you replicate exactly? How confident are you that this analysis is recovering an unbiased estimate of the persuasive effect of media on vote choice behavior? Does replicating the analysis change your assessment?