PS C236A / Stat C239A
Problem Set 2
Due: Sept. 28, 2012

Instructions

This assignment is due **4 pm Friday, Sept. 28.** You may submit your analytical work 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 of the assignment, you <u>must</u> submit a fully executable version of all .R code, along with any data used in the code (excepting that provided through the course webpage) to the email above. All files for each assignment sent electronically should be included in one omnibus email, with the subject line containing the course and homework number, and your last name (e.g., PS239A/STAT236A: HW2 - Obama).

You are encouraged to work together in groups to complete the assignments. However, you must hand in your own individual answers. Photocopies and other reproductions of someone else's answers are not acceptable. Please also list the names of everyone with whom you have collaborated on this assignment.

Problem 1: The Lady Tasting Tea Consider the following variation of the Lady Tasting Tea example that we discussed in class. The Lady tastes eight cups of tea, four of which have milk added first and four of which have tea added first. The cups are organized into matched pairs and for each pair, a fair coin is flipped to determine which gets milk first. The Lady knows the design, meaning that she knows there is one milk-first cup and one tea-first cup in each matched pair.

a. In the case where the Lady makes one mistake (classifies one milk-first cup as a tea-first cup), what is the p-value for a test under the null hypothesis that the Lady has no ability to discriminate the order in which milk is added to tea?

b. Pretend that you mistakenly thought that assignment of milk-first or tea-first was completely randomized, i.e. that there was no randomization within matched pairs, but rather across all cups. If the Lady makes one mistake, what p-value would you calculate for a test under the null hypothesis that the Lady has no ability to discriminate the order in which milk is added to tea? Is this p-value different from the one calculated in part (a)? Why or why not?

Now, instead of using fixed margins, let's imagine that we conduct the Lady Tasting Tea experiment under binomial randomization *without* a fair coin. There are *six* cups, $C_c = \{C_1, C_2, ..., C_6\}$, (indexed by c) with the following vector of probabilities of having milk first, $p_c = \{p_1, p_2, ..., p_6\}$, and $1 - p_c = \{1 - p_1, 1 - p_2, ..., 1 - p_6\}$ probabilities of having tea added first. The Lady does not know the values of p_c, but does know that the cups are assigned randomly under binomial randomization

c. First fix $p_c = 2/3$, for all c cups. Which null hypothesis would we prefer: The Lady has no ability to identify milk-first cups or The Lady has no ability to identify tea-first cups. Why?

d. Now fix $p_c = \{0.45, 0.5, 0.55, 0.8, 0.85, 0.9\}$. If the Lady makes one mistake, now what is the p-value for a test under the null hypothesis that the Lady has no ability to discriminate the order of milk first? What if a researcher, years later, came across this data and assumed $p_c = 0.5$, $\forall c$? How much would this bias the inferences the researcher draws from the experiment? What is the expected number of cups with milk first under the true assignment mechanism (rounded to the nearest cup)? How probable is it to realize this (rounded) expected number of milk-first cups with $p_c = .5$? Is this rare?

e. Continuing with this example, *fix the margins* so that the number of cups with milk first is held at *three*, but the probability of each cup is determined by the above non-fair coin in (d) (i.e., the orderings are a function of these individual probabilities). Assume the experimental draw is $C_c = \{0, 1, 1, 0, 1, 0\}$, and the Lady selects two milk-first cups correctly, with an associated p-value, q_1, under the null of no ability. If we repeat the experiment and randomly assign cup order as $C_c = \{0, 0, 1, 0, 1, 1\}$ and again the Lady gets two correct with p-value q_2, is $q_1 = q_2$? Why or why not?

Problem 2: Catholic School – I In an observational study of the effects of attending a Catholic school, the central dependent variable of interest is a binary variable, Y_i, which indicates whether or not student i graduated from high school. The treatment variable, T_i, indicates Catholic school attendance. In a very large sample of students, half attended Catholic school and half did not. You observe that the treated students have a graduation rate of .7 and the control students have a graduation rate of .5. You wish to estimate the average treatment effect of attending Catholic school. Assume that your sample is large enough to make sampling variability negligible.

a. Without making any assumptions about the relationship between the students' potential outcomes and treatment assignment, what is the largest possible value of the ATE? What is the smallest possible value of the ATE? What is the difference between these two values? Will this difference between the maximum and minimum possible ATE always be the same, irregardless of the specific observed values of the outcome variable?

b. Again making no assumptions about treatment assignment, assume that Catholic school does not prevent any student from graduating. What is the largest possible value of the ATE? What is the smallest possible value of the ATE?

Problem 3: Catholic School – II To address this question more precisely, researchers randomly sample n students, collecting the variable X for students who attend Catholic (treateds) and non-Catholic schools (controls). The researchers then exactly match 6 controls to 6 treated students on X, producing the data in the Table below. Assume *unconfoundedness* conditional on $\{X, U\}$, so that units are exchangeable across Catholic and non-Catholic attendance given X and U. Also assume that the conditionality in T_i follows a logit distribution, $\pi_i/(1 - \pi_i) = \exp(X_i\beta + \gamma U_i)$, where π_i is student i's probability of attending Catholic school, $U_i \in \{0, 1\}$ is a binary variable, and γ and β are an additive parameters.

Table 1: Catholic School Graduation Data

Strata (S)	Y	T	X	U
1	1	1	.89	1
2	1	1	-.25	1
3	1	1	.67	1
4	1	1	-.11	1
5	0	1	.13	1
6	1	1	.73	1
1	0	0	.89	1
2	0	0	-.25	0
3	0	0	.67	0
4	0	0	-.11	0
5	1	0	.13	0
6	0	0	.73	1

a. Ignore the strata for a moment. What is Fisher's test statistic for this data? Assume $\beta = 0$ and $\gamma = 0$. What is the permutation p-value for this statistic, under the sharp null of no difference in graduation outcomes for Catholic and non-Catholic schools? Now assume $\beta = 1.45$ and $\gamma = 0$. What is the p-value under Fisher's sharp null?

b. Turning to the stratification analysis, what is the McNemar test statistic for the matched-pair data? Again assume $\beta = 1.45$ and $\gamma = 0$. What is the permutation p-value for this statistic, under the sharp null of no difference in graduation outcomes in the matched pairs?

c. Assume that matching only eliminated the bias in X, and thus $\gamma > 0$. Assuming $\beta = 1.45$, at what minimum (positive) value of γ would we *fail to reject* the sharp null of no effect given the matched-pair design at the $p \geq .05$ level?

Problem 4 Consider an observational study, where $Z_i = 1$ if unit i is in the treatment group and $Z_i = 0$ if unit i is in the control group. Let X be a vector of observed pretreatment covariates. Write $X_{Z=1}$ for the observed covariates of the units in the treatment group. Similarly, let $X_{Z=0}$ be the observed covariates in the control group. Let r_1 be outcome under treatment and r_0 be the outcome under control. Assume the following:

$$r_0 \perp\!\!\!\perp Z | X_{Z=1}$$

$$P(Z = 1 | X_{Z=1}) < 1$$

Suppose you know the propensity score $e(X) = P(Z = 1)$ for all units i. With these assumptions, can conditioning on the propensity score estimate the ATT without bias? Prove it mathematically and describe your logic in words. What additional assumption would we need in order to estimate the ATE without bias?

Problem 5: In this problem, you will analyze a famous experiment conducted by Leonard Wantchekon in Benin in 2001. Wantchekon wanted to examine the effectiveness of different types of campaign messages on voting behavior in a presidential election. For details, see:

`http://www.princeton.edu/~lwantche/Clientelism_and_Voting_Behavior_Wantchekon.pdf`

Wantchekon convinced the campaigns of the major presidential candidates to randomize the messages they employed in 24 villages. The three treatment conditions were as follows:

1. *Public Policy:* Wantchekon describes this treatment condition as: "It was decided that any public policy platform would raise issues pertaining to national unity and peace, eradicating corruption, alleviating poverty, developing agriculture and industry, protecting the rights of women and children, developing rural credit, providing access to the judicial system, protecting the environment, and/or fostering educational reforms."

2. *Clientelist*: Wantchekon describes this treatment as: "A clientelist message, by contrast, would take the form of a specific promise to the village, for example, for government patronage jobs or local public goods, such as establishing a new local university or providing financial support for local fishermen or cotton producers."

The data has been modified for the assignment, but the basic structure of the experiment was *block* randomization. For the purposes of the assignment, villages were divided into groups of 2 based on geography and treatment status was randomized within the 8 groups of 2. The outcome variable is the vote share of the candidate participating in the experiment. The only covariate is the number of registered voters. In the dataset, `block` indicates block group, `reg.voters` is the registered voters covariate,`vote.pop`is the outcome variable, `treat` is a variable indicating treatment status.

In this problem, we are interested in the difference between the clientelist and public policy conditions.

a. Estimate the effect the clientelist message compared to the public policy message, using the ITT estimator and the regression estimator. For the regression estimate, include block level dummy variables in your regression equation.

a. Now test the sharp null of no treatment effect using randomization inference. Use two test statistics: Wilcoxon's signed rank test (Rosenbaum 2002, pg. 32) and the difference in means. What are the two sided p-values under these two tests?

a. Under the assumption of a constant, additive, treatment effect, use randomization inference to find a 95% confidence interval of the treatment effect. Use the signed rank as your test statistic. See pages 44-46 in Rosenbaum (2002).

d. What can you conclude about the effectiveness of clientelistic appeals in Benin?

e. Bonus: Perform randomization inference with covariance adjustment. How does this effect your results? For a very good article on covariance adjustment with randomization inference, see:

> Rosenbaum, Paul. 2002. "Covariance Adjustment in Randomized Experiments and Observational Studies." *Statistical Science* 17(3): 286-327.