PS C236A/ Stat C239A
Problem Set 3
Due: Sept. 25, 2009

Problem 1: Imagine a randomized experiment where 6 out of 30 students in a classroom are randomly exposed to a new type of teaching method. The 24 other students are exposed to the conventional teaching method. Under the Neyman-Rubin model and assuming SUTVA, there are 60 potential outcomes. Drop the SUTVA assumption (sometimes called the no-interference assumption). How many potential outcomes are there? For a succinct definition of SUTVA, see page 41-42 of Rosenbaum's *Observational Studies*.

There are 593,775 possible treatment assignments. Without SUTVA, each child has a different potential outcome for every possible treatment assignment. As a result, there are a total of $30 \times 59,3775 = 17,813,250$ potential outcomes.

Problem 2: The Lady Tasting Tea Consider the following variation of the Lady Tasting Tea example that we discussed in class. The Lady tastes six cups of tea, three of which have milk added first and three of which have tea added first. The cups are presented to the Lady in random order. The Lady knows the design, meaning she knows there are exactly three milk-first cups and three tea-first cups which will presented to her randomly.

a. In the case where the Lady makes no mistakes, 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?

   The following answers will be derived using Rosenbaum's test-statistic notation for randomization inference. Since we allow the Lady to taste 6 cups, $N = 6$, and there is only one stratum, so $S = 1$. A treatment assignment is a 6-tuple containing three 1s and three 0s. Let Z_i be an indicator variable equal to one if unit i receives treatment. Let $\mathbf{Z}$ be an N-dimensional column vector whose elements are the Z_i for all units. The set of treatment assignments Ω contains all possible arrangements of the three 1s and three 0s, so $|\Omega| = \binom{6}{3} = 20$. The fact that treatment assignment was random means that $Pr(\mathbf{Z} = \mathbf{z}) = \frac{1}{20} \; \forall \, \mathbf{z} \in \Omega$.

   Let the Lady's outcome for cup i be defined as r_i, where $r_i = 1$ if the lady classifies cup i as milk first and $r_i = 0$ if the lady classifies cup i as tea first. This means that $\mathbf{r} = (r_1, \ldots, r_6)^T$. Since the Lady must classify exactly 3 cups as milk first and 3 cups as tea first, we have $1^T\mathbf{r} = 3$. The test-statistic is the number of cups correctly identified:

$$\begin{aligned} t(\mathbf{Z}, \mathbf{r}) &= \mathbf{Z}^T\mathbf{r} + (1 - \mathbf{Z})^T(1 - \mathbf{r}) \\ &= \mathbf{Z}^T\mathbf{r} + \mathbf{1}^T\mathbf{1} - \mathbf{1}^T\mathbf{r} - \mathbf{Z}^T\mathbf{1} + \mathbf{Z}^T\mathbf{r} \\ &= \mathbf{Z}^T\mathbf{r} + 6 - 3 - 3 + \mathbf{Z}^T\mathbf{r} \\ &= 2\mathbf{Z}^T\mathbf{r} \end{aligned}$$

   Under perfect agreement between the Lady's choices and the way milk and tea are added to the cups, the observed value of the test-statistic is

$$T = 2\mathbf{Z}^T\mathbf{r} = 2 \cdot 3 = 6$$

There is only one assignment $\mathbf{z} \in \Omega$ of the cups that leads to perfect agreement with the Lady's choice. Therefore, the significance level is

$$Pr(t(\mathbf{Z}, \mathbf{r}) \geq 6) = \frac{|\{\mathbf{z} \in \Omega : t(\mathbf{z}, r) \geq 6\}|}{20} = \frac{1}{20} = 0.05$$

b. 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?

Now we allow the lady to make one mistake (i.e. to correctly identify only 4 cups). We saw that there was only one assignment $\mathbf{z} \in \Omega$ that led to perfect agreement. How many $\mathbf{z} \in \Omega$ assignments lead to four agreements? To get exactly four agreements, we must change a 0 by 1 in $\mathbf{r}$. Since there are three 1s and three 0s there are $\binom{3}{1} \times \binom{3}{1} = \frac{3!}{(3-1)!1!} \times \frac{3!}{(3-1)!1!} = 9$ ways in which we can have 4 correctly identified cups. Another way of saying this is that there are 9 assignments with exactly $t(\mathbf{Z}, \mathbf{r}) = 4$ agreements. Since there is one assignment with $t(\mathbf{Z}, \mathbf{r}) = 6 > 4$ agreements, so there are 10 assignments leading to four or more agreements. Thus, the significance level in this case is

$$Pr(t(\mathbf{Z}, \mathbf{r}) \geq 4) = \frac{|\{\mathbf{z} \in \Omega : t(\mathbf{z}, r) \geq 4\}|}{20} = \frac{10}{20} = 0.5$$

Note that this is no longer a small probability. From observing that she classifies four cups correctly, we are not justified in inferring that she does have the ability to discriminate between the cups.

Now, instead of using fixed margins, let's imagine that we conduct the Lady Tasting Tea experiment under binomial randomization. There are 6 cups, and each has a probability $p = 1/2$ of having milk first and a $1 - p$ probability of having tea added first. The Lady does not know the value of p, but does know that the cups are assigned randomly under binomial randomization.

c. In the case where the Lady makes no mistakes, 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?

In an experiment that consists of N trials, the probability that the number of successes, Y, is equal to a particular value, y, is:

$$Pr(Y = y) = \binom{N}{y} p^y (1-p)^{N-y} \qquad y = 0, 1, 2, \ldots, N$$

Therefore, the chance that the lady correctly identifies 6 out of 6 cups under this experimental design is

$$Pr(Y = 6) = \binom{6}{6} \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^0 = \frac{1}{64} \approx 0.0156$$

Note that this is much lower than the 0.05 chance we observed in the previous experimental design.

d. 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?

Now we allow one mistake, which no longer implies that the lady incorrectly classifies two cups (under binomial randomization, one mistake is only one mistake because we don't have the fixed margins). The probability of 5 successes is

$$Pr(Y = 5) = \binom{6}{5} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^1 = \frac{6}{64} \approx 0.0937$$

The significance level for a test that rejects the null hypothesis that the lady has no ability to discriminate the order in which milk is added to tea is

$$Pr(t(\mathbf{Z}, \mathbf{r}) \geq 5) = \frac{|\{\mathbf{z} \in \Omega : t(\mathbf{z}, r) \geq 5\}|}{64} = \frac{1+6}{64} = \frac{7}{64} \approx 0.1093$$

e. Make the best argument you can for the fixed margin design. Also make the best argument you can for the binomial randomization design.

Perhaps we believe that the Lady needs to have a true comparison to determine which cups are milk first and which cups are tea first. Then we may prefer the fixed margin design because it guarantees that there will be some milk first cups and some tea first cups. However, it is not always a very sensitive design, as we've seen in this homework and in class. The binomial design has the advantage that it is a more sensitive design, however there is always a chance that we will have all milk first cups or all tea first cups.

f. *Bonus:* Now suppose that each cup as a probability $p = 1/3$ of having milk first. 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?

The change in this problem is that the probability of observing a milk-first cup and a tea-first cup is no longer equal. The important part of this question is to realize that because of this asymmetry, we can manipulate our null hypothesis to increase the sensitivity of our test. Imagine that the null hypothesis is that the lady has no ability to identify milk-first cups. In this case, the probability of success is $p = \frac{1}{3}$. The probability that the lady correctly identifies the six cups is

$$Pr(Y = 6) = \binom{6}{6}\left(\frac{1}{3}\right)^6\left(\frac{2}{3}\right)^0 = \frac{1}{729} \approx 0.0013$$

If we allow the lady one mistake

$$Pr(Y = 6) = \binom{6}{5}\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right)^1 = \frac{12}{729} \approx 0.0164$$

Therefore the significance level of the test is:

$$Pr(t(\mathbf{Z}, \mathbf{r}) \geq 5) = \frac{|\{\mathbf{z} \in \Omega : t(\mathbf{z}, r) \geq 5\}|}{729} = \frac{13}{729} \approx 0.0177$$

Now, imagine that the null hypothesis is that the lady has no ability to identify tea-first cups. in this case, the probability of success is $p = \frac{2}{3}$. The probability that the lady correctly identifies the six cups is:

$$Pr(Y = 6) = \binom{6}{6}\left(\frac{2}{3}\right)^6\left(\frac{1}{3}\right)^0 = \frac{64}{729} \approx 0.087$$

And if we allow one mistake:

$$Pr(Y = 6) = \binom{6}{5}\left(\frac{2}{3}\right)^5\left(\frac{1}{3}\right)^1 = \frac{192}{729} \approx 0.2633$$

Therefore, the significance level of the test is:

$$Pr(t(\mathbf{Z}, \mathbf{r}) \geq 5) = \frac{|\{\mathbf{z} \in \Omega : t(\mathbf{z}, r) \geq 5\}|}{729} = \frac{256}{729} \approx 0.3503$$

So we see that depending on what null hypothesis we choose, the significance level of the test varies quite a bit. Since it is less likely to observe milk-first cups, using milk-first cups in the null hypothesis will provide us with a more sensitive test. Now, it may not make sense in this case to have asymmetrical nulls, in a sense because if the lady can identify milk-first cups, then she can identify tea-first cups This is true, but we could modify the events such that they are not complementary nulls anymore. In any case, the point of the problem remains: what we choose as our null hypothesis determines the significance level of our test to reject the null. So, we should be careful when choosing our nulls!

Problem 3: According to David Freedman, under what circumstances might you get gains (reduction in asymptotic variance) from adjusting experimental data using a multiple regression (i.e. a regression with some set of covariates, Z, in addition to treatment assignment)? Under what circumstances might adjusting experimental data increase asymptotic variance? Explain briefly.

Freedman derives the gains from adjustment as:

$$\frac{\Delta}{np(1-p)}, \qquad \text{where}\Delta = (\overline{\alpha Z})\Big[(\overline{\alpha Z}) + 2(1-2p)(\overline{\beta Z})\Big]$$

From this, we can see that if $p = \frac{1}{2}$, then by adjusting the data, the adjustment is either positive or neutral. Also, if our "strict null" hypothesis holds, then we will have positive gains from adjustment unless $(\overline{\alpha Z}) = 0$. Note, the strict null means that $T_i = C_i + b \,\forall\, i$. This means that there is a constant additive treatment effect. The variance in $\hat{b}_{ITT}$ because there is still variance in the C_i from one subject to another, and the deviation of C_i from the population average $\bar{C}$ is captured by α_i. If $(\overline{\alpha Z}) = 0$, then the remaining variation in C_i is orthogonal to the covariate, and so there are no gains to adjusting under the "strict null". Note too that the "strict null" of constant additive treatment effect is the same assumption that we make with linear regression.

When $p \neq \frac{1}{2}$, then adjustment may hurt. Assuming that $(\overline{\beta Z})$ is large and positive, then if $(\overline{\alpha Z}) > 0$ and $p > \frac{1}{2}$, then there will be negative gains. Similarly, if $(\overline{\alpha Z}) < 0$ and $p < \frac{1}{2}$, there will be negative gains. In this world, in which the "strict null" doesn't hold and we don't necessarily have a balanced design, then the gains from adjustment are not clear. Regression adjustment may help or hurt, but it all depends on how the treatment effect and variation in C_i and T_i vary with the covariate as well as the relative size of the treatment group and control group. Often this is the scenario that we are working in.

Problem 4: 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.nyu.edu/gsas/dept/politics/faculty/wantchekon/research/WP_0331.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."
3. *Both*: These villages received both types of messages.

The structure of the experiment was *block* randomization. Villages were divided into groups of 3 based on geography and treatment status was randomized within the 8 groups of 3. 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 mainly 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. Note that the block structure of this experiment will affect how you calculate these quantities. See page 47 of Rosenbaum for how to compute the ITT estimator with blocking. For the regression estimate, include block level dummy variables in your regression equation.

   See solution code.

b. Now test the sharp null of no treatment effect using randomization inference. Use two test statistics: the rank sum test and the difference in means (see Rosenbaum (2002), Chapter 2). If you don't want to enumerate every possible treatment assignment, just sample a large number of draws from the randomization distribution. What are the two sided p-values under these two tests?

   See solution code.

c. Now perform randomization inference ignoring the block structure of the experiment using difference in means as your test statistic. In other words, pretend that treatment status was allocated randomly without regard to the block variable, sometimes known as "complete randomization". Are these results different from the result obtained in part b? Is this method of randomization inference valid?

   Because we ignore the block structure of the data, the test is less powerful and will produce conservative p-values. As a result, we will not over-reject the null hypothesis when it is true, making the test valid in a strict sense. On the other hand, a better strategy would be to use the correct method of following the actual randomization procedure since it will be more powerful and not over-reject the null.

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

   We can reject the null that the clientelist treatment had no effect. The point estimate suggests that, in aggregate, the clientelist is more effective at garnering votes than the public policy-oriented strategy.

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.

See code.