POLSCI 787 **Due Date:**
Fall 2017 **Problem Set 1** **Fri Sep 22, 2017, 11:59pm**

- You may work in (small) groups while solving this assignment.
- Submit individual solutions via Canvas in one PDF file collecting everything (e.g., derivations, figures, tables, computer code).
- Use `RMarkdown` to make all calculations and create a report with all the solutions and answers.
- Please provide written explanations whenever those are requested; reporting only code and output automatically leads to ✓⁻.

Question 1

Imagine a treatment effect model that, instead of being additive and constant, is multiplicative and constant. The model is $Y_i(1) = \tau Y_i(0)$ for all $i = 1, 2, \cdots N$, with observed outcome Y_i. Assume that there is an N-dimensional vector $\mathbf{Z}$ of zeros and ones that assigns a binary treatment to the N units. The treatment assignment $\mathbf{Z}$ is determined randomly according to a known random mechanism, so that its distribution is known. Discuss how to test hypotheses of the type $H_0 : Y_i(1) = \tau_0 Y_i(0)$ based solely on the randomization distribution of $\mathbf{Z}$. How would you construct adjusted outcomes in this case? How would you use these adjusted outcomes to test these hypotheses? How would you construct confidence intervals for τ?

Question 2

Consider the following variation of the Lady Tasting Tea example covered in Chapter 2 of Rosenbaum (2002) (pages 21-23 and 29-30). The Lady tastes six cups, 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 both that there are exactly three milk-first cups and three tea-first cups, and that the cups will be presented to her randomly.

(a) Do not allow the Lady to make any mistakes, i.e. reject the null hypothesis if she makes one mistake or more. What is 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?

(b) Now allow the Lady to make one mistake, i.e. to classify a milk-first cup as a tea-first cup. What is now the significance level for a test that rejects the null hypothesis of no ability to discriminate?

(c) Now allow the Lady to make two mistakes, i.e. to classify two milk-first cups as a tea-first cups. What is now the significance level for a test that rejects the null hypothesis of no ability to discriminate?

(d) How useful is this experimental design?

Question 3

For Questions 3, 4, and 6, you will use the database `AR_data_distribution.dta`, which contains data from Arkansas state senators during the legislative session that met in 2003. After being elected, these 35 senators were randomly assigned to serve a term of 2 years or 4 years. We refer to the 2-year term condition as "treatment", and to the 4-year term condition as "control". The randomization mechanism fixed the number of treated to be 18 and the number of controls to be 17, and gave each possible treatment assignment an equal probability of occurring. Each row is a different senate district which is represented by a different senator, and the columns are different variables that contain information about the senators and their districts. Some variables are covariates, this is, they are determined before the treatment assignment, and some variables are outcomes, this is, they are determined after the treatment assignment. (For details about

the experiment, you can look at Titiunik (2016)).

The description of the variables in the dataset is in the file `codebook_AR_data_CLEAN.txt`.

This exercise will look at some descriptive statistics of the dataset.

(a) Read the data, look at its dimensions, look at its variables. How many variables do you have? How many observations? [Hint: use the R function `read.dta` in the library `foreign` to load the data.]

(b) Report the minimum, maximum, mean and standard deviation of at least ten covariates in the dataset, separately for treatments and controls (this is, separated by both values of the `dshort_term` variable). Are treatment and control similar in terms of these characteristics? Should they be similar? Why?/Why not? Just look at the descriptive statistics to answer these question, do not perform hypothesis tests for now. [Hint: use the R functions `min,max,mean` and `sd`. And to make your life easy, you may choose to use the function `apply`.]

(c) For each covariate that you analyzed above, present the test statistic and p-value from a parametric two-sided t-test of the hypothesis that the mean in the control group is equal to the mean in the treatment group. Do you see any statistically significant difference? What does this tell you about the validity of the experiment? You are not allowed to use a canned t-test routine to calculate your answer, but you can check your results with the R function `t.test`.

Question 4

In this question, we will make **Fisherian** inferences using the Arkansas data, that

is, inferences that see the potential outcomes as fixed and are based solely on the exact distribution of the treatment assignment (these inferences do not rely on any large-sample approximations).

(a) As mentioned above, the randomization mechanism fixed the number of treated to be 18 and the number of controls to be 17, and gave each possible treatment assignment an equal probability of occurring. What was this probability? How many possible treatment assignments are there? Is it feasible to enumerate them all?

(b) Now let's test the sharp null hypothesis for the abstention rate outcome (variable `abs_rate`). You may use part (1) of the R code `RandomizationInference-Fisher-vs-Neyman.R` as a guide. Report (an approximation to) the p-value for the sharp null hypothesis of no treatment effect using the difference in means as a test-statistic. Can you reject the null hypothesis that the abstention rate under 2-year terms is equal to the abstention rate under 4-year terms for all senators? What is the p-value?

(c) Now test the sharp null hypothesis and report the approximate exact p-value for the total number of bills introduced during the session (variable `bills_intro`). Can you reject the null hypothesis that the number of bills introduced under 2-year terms is equal to the number of bills introduced under 4-year terms for all senators? What is the p-value?

(d) Repeat the calculations in subquestions (b) and (c) but using the Wilcoxon rank sum statistic instead of the difference in means. See Rosenbaum (2002), Chapter 2, page 32 for a definition and an example of how to calculate it. Do your conclusions change? What does this evidence tell you about the sharp null? Are you more or less confident with the conclusions above, given the evidence from the rank-sum statistic? [Hint: the R function `rank` may be helpful.]

(e) Write an `R` function to perform the calculations in (b) and (d) for any possible variable. Use your function to repeat your calculations. Do you get the same results?

(f) Use the function you wrote in the previous question to perform **Fisherian** covariate balance tests. That is, redo the tests in question 3(c) using Fisherian inference. Are you testing the same hypothesis in both cases? Explain.

Question 5

In the famous study, LaLonde (1986) examined the impact of the National Supported Work (NSW) Demonstration on the post-training income. The data was from a randomized control experiment, hence treatment effect could be easily estimated. The program was a federally and privately funded program implemented in the mid-1970s to provide work experience for a period of 6-18 months to individuals who had faced economic and social problems prior to enrollment in the program. Candidates eligible for the NSW were randomized into the program between March 1975 and July 1977. The file `LaLonde 1986.csv` contains the data used in Dehejia and Wahba (1999), which is a subset of the initial LaLonde (1986), where earnings in 1974 is available (number of observations: 445).

Consider the data from `Lalonde_1986.csv`. For $i = 1, 2, ..., n$, with $n = 445$, let $Y_i = earn78 \in \mathbb{R}_+$ and $Z_i \in \{0, 1\}$. Employ a causal inference framework where potential outcomes are non-random, and the only source of randomness is the treatment status of each unit i (i.e. the probability law for the random variable Z_i). Assume throughout that $\mathbf{Z} = (Z_1, Z_2, ..., Z_n)'$ follows a fixed margins (or complete randomization) distribution, with N_1 treated units and $n - N_1$ control units.

(a) **Neyman's Approach**

Consider the difference-in-means statistic:

$$\begin{aligned} T_{DM} &= \bar{Y}_1 - \bar{Y}_0 \\ \bar{Y}_t &= \frac{1}{N_t}\sum_{i=1}^{n} D_i(t) Y_i \\ N_t &= \sum_{i=1}^{n} D_i(t) \\ D_i(t) &= \mathbb{1}[Z_i = t] \\ t &= 0, 1 \end{aligned}$$

i.) Show that:

$$\mathbb{E}[T_{DM}] = \frac{1}{n}\sum_{i=1}^{n} Y_i(1) - \frac{1}{n}\sum_{i=1}^{n} Y_i(0) =: \tau_{ATE}$$

Compute an unbiased estimator of the average treatment effect using the `Lalonde_1986.csv` data.

ii.) It can be shown that, under appropriate regularity conditions,

$$\begin{aligned} \frac{T_{DM} - \tau_{ATE}}{\sqrt{\mathbb{V}[T_{DM}]}} &\to_d \mathcal{N}(0,1) \\ \mathbb{V}[T_{DM}] &\leq \frac{\bar{S}_0^2}{N_0} + \frac{\bar{S}_1^2}{N_1} \\ \bar{S}_t^{\,2} &= \frac{1}{N_t - 1}\sum_{i=1}^{n} D_i(t)(Y_i - \bar{Y}_t) \\ t &= 0, 1 \end{aligned}$$

Construct an asymptotically conservative 95% confidence interval for the average treatment effect using the `Lalonde_1986.csv` data.

(b) **Fisher's approach**

Consider the difference-in-means test statistic (denoted by T_{DM}), the Kolmogorov-Smirnov test statistic (denoted by T_{KS}), and the 75th quantile test statistic

(denoted by T_{75q}). In order to define T_{KS} and T_{75q}, we define an estimator for the cumulative densitive function (CDF) for each group:

$$\begin{aligned}\hat{F}_0(y) &= \frac{1}{N_0}\sum_{i:Z_i=0}\mathbb{1}[Y_i \leq y] \\ \hat{F}_1(y) &= \frac{1}{N_1}\sum_{i:Z_i=1}\mathbb{1}[Y_i \leq y]\end{aligned}$$

Then, T_{KS} is defined as:

$$T_{KS} = \sup_y \left|\hat{F}_1(y) - \hat{F}_0(y)\right| = \max\left\{\left|\hat{F}_1(Y_i) - \hat{F}_0(Y_i)\right|\right\}_{i=1}^{n}$$

The 75th quantile test statistic is defined as:

$$T_{75q} = \left|\hat{F}_1^{-1}(0.75) - \hat{F}_0^{-1}(0.75)\right|$$

Answer the following questions. For all programming involved, write programs without using canned functions that calculate the statistics involved (difference-in-means, Kolmogorov-Smirnov, or quantiles.)

i.) Give a short intuitive explanation of the difference between what T_{DM}, T_{KS}, and T_{75q} are measuring.

ii.) Report (an approximation to) the p-value for the sharp null hypothesis of no treatment effect using the three statistics described above and the `Lalonde_1986.csv` data.

iii.) Report (an approximation to) the p-value for the sharp null hypothesis of a constant treatment effect $c \in \mathbb{R}\backslash\{0\}$ (that you choose) using the three statistics described above and the `Lalonde_1986.csv` data.

iv.) Using a constant treatment effect model (and SUTVA) construct a finite-sample valid 95% confidence interval for the treatment effect using the three statistics described above and the `Lalonde_1986.csv` data.

Question 6

In this question, we will make **Neyman** inferences using the Arkansas data, based on some of the answers you found in the previous question. Use a Neyman approach to answer all questions below.

(a) Calculate the mean abstention rate (variable `abs_rate`) in the treatment group, and the mean abstention rate in the control group. What is the difference in means? Is this an unbiased estimator of the average treatment effect in this sample?

(b) Assume we are interested in testing whether this average treatment effect is zero. What is the general expression of the variance of this estimator? Why is it not possible to estimate this variance? Propose a conservative estimator of the unfeasible unknown variance. Use this estimator to estimate an upper bound on the variance of the difference in average abstention rates between the two groups of senators.

(c) Test the hypothesis that the average abstention rate is equal in both groups.

(d) Construct a 95% confidence interval for the finite-sample average treatment effect on abstention rates.

References

Dehejia, Rajeev H., and Sadek Wahba. 1999. "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs." *Journal of American Statistical Association* 94 (448): 1053-1062.

LaLonde, Robert J. 1986. "Evaluating the econometric evaluations of training programs with experimental data." *The American Economic Review*: 604–620.

Rosenbaum, Paul R. 2002. *Observational studies*. Springer.

Titiunik, Rocío. 2016. "Drawing Your Senator From a Jar: Term Length and Legislative Behavior." *Political Science Research and Methods* 4 (2): 293–316.