POL SCI 231b (Spring 2017):

Problem Set 8

Spring 2017

University of California, Berkeley

Prof. Thad Dunning/GSI Natalia Garbiras-Díaz

Due Friday, April 21st, at 10:00 AM (before section)

Each group should turn in its problem set solution by email to Natalia Garbiras-Díaz (nataliagarbirasdiaz+231b@berkeley.edu) by Friday at 10 AM. Please work out the problems on your own, before you meet with your group to agree on solutions.

1. **The bootstrap** An analyst assumes the following regression model:

$$Y = X\beta + \epsilon, \tag{1}$$

where Y is an $n \times 1$ vector of observable random variables. Here, X is a fixed $n \times p$ matrix with a vector of 1's as the first column, and ϵ is mean-zero vector of i.i.d. random variables with $\text{var}(\epsilon_i) = \sigma^2$. The OLS estimator for this model is $\hat{\beta} = (X'X)^{-1}X'Y$. The residuals from the OLS fit are $e = Y - X\hat{\beta}$.

Suppose the analyst uses the procedure described by Freedman (2009, Chapter 8) to bootstrap the regression model. In particular, for the kth bootstrap replicate, she samples at random with replacement from the vector e to produce an $n \times 1$ vector of bootstrap errors, $\epsilon_{(k)} = \{\epsilon_{(k)1}, ..., \epsilon_{(k)n}\}'$. For each bootstrap replicate, she then constructs $Y_{(k)} = X\hat{\beta} + \epsilon_{(k)}$ and fits the OLS estimator, $\hat{\beta}_{(k)} = (X'X)^{-1}X'Y_{(k)}$. There are 1,000 bootstrap replicates. Finally, let

$$
\begin{aligned}
\hat{\epsilon}_{(k)} &= Y_{(k)} - X\hat{\beta}_{(k)}, \\
s_k^2 &= \frac{\hat{\epsilon}'_{(k)}\hat{\epsilon}_{(k)}}{n-p}, \\
\hat{\beta}_{\text{ave}} &= \frac{1}{1000}\sum_{k=1}^{1000}\hat{\beta}_{(k)}, \text{ and} \\
V &= \frac{1}{1000}\sum_{k=1}^{1000}[\hat{\beta}_{(k)} - \hat{\beta}_{\text{ave}}][\hat{\beta}_{(k)} - \hat{\beta}_{\text{ave}}]'.
\end{aligned}
$$

Say whether the following statements are true or false, and most importantly, explain your answers (a correct answer with an incorrect explanation does not get full credit!):

(a) $E(\epsilon_{(k)}) = 0_{n\times 1}$.

(b) $E(\hat{\beta}_{(k)}) = \hat{\beta}$.

(c) $E(s_k^2) = \sigma^2$.

(d) $E(s_k^2) = \frac{1}{n}e'e$.

(e) $E(s_k^2) = \frac{1}{n-p}e'e$.

(f) $E(V) = \sigma^2(X'X)^{-1}$

(g) The square roots of the diagonal elements of V are the bootstrap standard errors.

(h) The sample SD of the $\hat{\beta}_{(k)}$'s is a good approximation to the SE of $\hat{\beta}$.

(i) $\hat{\epsilon}_{(k)} \perp X$ for all k.

(j) The bootstrap can provide evidence that the original data were produced according to equation (1).

(k) The bootstrap can provide evidence that $E(\hat{\beta}) = \beta$ if the original data were produced according to equation (1), with i.i.d. errors, $E(\epsilon_i) = 0$ and $\text{var}(\epsilon_i) = \sigma^2$.

2. Miguel and Fisman (2006: 1020) are interested in how cultural norms influence corruption. They write, “Until 2002, diplomatic immunity protected [United Nations] diplomats from parking enforcement actions, so diplomats' actions were constrained by cultural norms alone. We find a strong effect of corruption norms: diplomats from high-corruption countries (on the basis of existing survey-based indices) accumulated significantly more unpaid parking violations." They refer to this study as a “unique natural experiment."

   (a) What is the treatment variable in this study? It is plausibly assigned as-if at random? Is this plausibly a natural experiment?

(b) Consider three threats to the substantive or theoretical relevance of the intervention here, as discussed in Dunning (2012: Chapter 10): external validity, idiosyncrasy, and bundling. Which of these do you consider to be the most important here and why?

(c) What if diplomats from richer countries tend to have paid parking spaces? What violation(s) of the natural-experimental setup would this imply?

3. **Regression Discontinuity Design**.

   (a) Researcher A, analyzing a regression discontinuity design, uses all of the data inside of a given window to fit the following regression using OLS:

   $$E(Y_i|R_i, T_i) = \alpha_1 + \alpha_2 T_i + \alpha_3 R_i + \alpha_4 T_i * R_i, \tag{2}$$

   where T_i is the treatment indicator and R_i is the value of the running variable for i Researcher B chooses the same window but splits the data into treatment and control group and runs two separate OLS regressions:

   $$E(Y_i|R_i, T_i) = \beta_1 + \beta_2 R_i \quad \text{if } T_i = 1 \tag{3}$$

   and

   $$E(Y_i|R_i, T_i) = \beta_3 + \beta_4 R_i \quad \text{if } T_i = 0 \tag{4}$$

   Write the coefficients in equation (2) in terms of the β_j coefficients in equations (3) and (4).

   (b) **RD estimation function.** Modify Hidalgo's RD "estimate" function such that it calculates difference in means estimates using your own t-test function.

   (c) **Balance tests/F-test.**

   - Using your new function and Hidalgo's data, choose a bandwidth and produce a table replicating the individual balance tests in the paper.
   - **The F-test.** You will use the F-test to evaluate whether the entire set of pre-treatment covariates can predict treatment assignment (you should continue to work with the subset of data included in the window you chose for the previous ponti). The null hypothesis here is that the coefficients from a regression of treatment assignment on all of the pre-treatment covariates are all zero.

     i. Fit the big model (including all the pre-treatment covariates) and the small model by OLS and compute the sums of squares that are needed for the test: $\|e\|^2$, $\|X\hat{\beta}\|^2$, and $\|X\hat{\beta}^{(s)}\|^2$ using the matrix commands in R.

     ii. Use your results to calculate the F-statistic.

     iii. Write your own $F - test$ function and calculate a p-value using randomization inference.You will need to specify a sharp null, then reshuffle treatment and control labels (above and below the threshold), calculating an F-statistic each time to get the full distribution of F-statistics under the sharp null. Finally, compare the realized F-statistic to the distribution to calculate the randomization p-value.

iv. Is $\|Y\|^2 = \|X\hat{\beta}^{(s)}\|^2 + (\|X\hat{\beta}\|^2 - \|X\hat{\beta}^{(s)}\|^2) + \|e\|^2$? Coincidence or math fact?

- **Optional:** Use the *DCdensity* function in the *rdd* package to test for sorting.[1]

(d) Download `rddata.Rda` file from bcourses. The file contains electoral outcomes for 2000 and 2004 for 6000 municipalities. You will estimate party incumbency advantage.

i. Use the 2000 electoral outcomes to set up an RDD where you keep the winner and the runner up parties in each municipality. Define the running variable as the margin of victory of the winner.

ii. Create an RDD plot with margin in the x axis and vote share in 2004 in the y axis. You can recycle Hidalgo's code to do this.

iii. Some of the outcome data is missing. Assess whether it is missing at random or systematically. Discuss a few strategies to deal with the missing data.

iv. Use your RD function to estimate a LATE. Is there a party incumbency advantage?

v. Does this effect vary by party? Calculate heterogeneous treatment effects. Interpret your results.

4. There is a study group of 10 subjects in a randomized controlled experiment, in which 7 of the subjects are assigned at random to treatment and 3 are assigned to the control group. In answering the questions below, you may use *R* as a calculator where appropriate (except where noted below), but discuss your work. Observed data on the response variable look as follows:

Assigned to Treatment	Assigned to Control
3	–
2	–
5	–
6	–
3	–
4	–
5	–
–	2
–	4
–	3

(a) Construct a box model for this experiment, drawing on our discussion of the Neyman urn model. What is in the box?

[1] You can see what the code needed for the test by checking the github repository

(b) Define the average causal effect in terms of the model you constructed in (a). Then estimate the average causal effect, using the data in the table.

(c) Estimate the standard error of your estimate in (b). To do this, use the “conservative variance formula" discussed in readings and lectures.

(d) Suppose you want to know whether the estimated effect in (b) is statistically significant. First, explain carefully what is meant by “statistically significant." Use the words “null hypothesis" in your answer. Explain what is the null hypothesis, in terms of your box model in (a).

(e) Should you use a t-test to assess statistical significance in this study? Why or why not? Explain carefully.

(f) Conduct the t-test. What is the p-value?

(g) Suppose you instead want to use randomization inference to assess statistical significance. What null hypothesis do you need to assume? How does this differ from the null hypothesis you defined in (d)? How else does randomization inference differ from the t-test in (e)-(f)? Is randomization inference more appropriate than a t-test, and why or why not?

(h) Now, use randomization inference in R to assess the statistical significance of the estimated effect in (b). (Here, use code from section or problem sets if possible, rather than an R package such as `ri`. Set the seed to 12345 before you run your code). Is your p-value the same or different as in part (f)?

(i) Now, suppose an investigator assumes the OLS model:

$$Y_i = \alpha + \beta T_i + \epsilon_i, \tag{5}$$

where T_i is a 0-1 variable, with 1 indicating that a subject was assigned to treatment. Make a list of the “usual OLS assumptions."

(j) What are the differences between the Neyman urn model in part (a) and the OLS model in part (e)? What assumptions are shared by the models?

(k) Do you think the usual OLS assumptions are satisfied in this problem? Why or why not? Which assumptions are the most plausible? What assumptions are less plausible? Explain your answers carefully.

(l) Under the OLS model, what is $E(Y_i|T_i = 0)$? How about $E(Y_i|T_i = 1)$?

(m) Denote the design matrix as X. What is a typical row of this matrix? What size is X? Denote the response variable as Y. What size is Y?

(n) Calculate $X'X$, $(X'X)^{-1}$, $X'Y$, and $(X'X)^{-1}X'Y$. Use $(X'X)^{-1}X'Y$ to estimate α and β. Here, you should work out the matrices by hand (e.g., don't use R commands such as `solve`).

(o) Express $(\hat{Y}|T_i = 1) - (\hat{Y}|T_i = 0)$ in terms of your estimates $\hat{\alpha}$ and/or $\hat{\beta}$. How does this difference compare to your answer in (b)? Comment briefly.

(p) Now use the usual OLS formula to attach estimated standard errors to $\hat{\alpha}$ and $\hat{\beta}$. (Here, you can use R matrix commands such as `solve` if needed—but calculate terms explicitly and explain your work).

(q) Attach a standard error to the difference $(\hat{Y}|T_i = 1) - (\hat{Y}|T_i = 0)$ you found in (o). How does this compare to your estimated standard error in (c)? Explain why they are the same or different.

5. The *statistical power* of a test is the probability that it will reject the null hypothesis, given that the null hypothesis is false. In this question, you are asked to calculate statistical power in an experiment.

   Thus, consider Tables 1 and 4 in Dunning and Harrison (2010). The first row of Table 4 compares respondents' evaluations of co-ethnic politicians who either are or are not their joking cousins. Dunning and Harrison report an estimated treatment effect of $\hat{\tau} = 0.49$ on their 1-7 scale, with an estimated standard error of 0.22. In the questions below, take 0.22 as your best guess of the true standard error of $\hat{\tau}$. Suppose that the Ns in the first row of Table 1 are large enough that a central limit theorem approximately applies. (You can use *R* freely if needed to answer any of the following questions).

   (a) Fill in the following question marks with the correct answers: under the null hypothesis, the distribution of $\hat{\tau}$ is $\mathbf{?}(\mathbf{?}, \mathbf{?})$.

   (b) Suppose that the true treatment effect is indeed $\tau = 0.49$. Fill in the question marks with the correct answers: under this alternative, the distribution of $\hat{\tau}$ is $\mathbf{?}(\mathbf{?}, \mathbf{?})$.

   (c) Now suppose that we repeated Dunning and Harrison's experiment infinitely many times under identical conditions (just suppose). For each experiment, we use a z-score to conduct a test of the null hypothesis. For what fraction of these experiments would we reject the null hypothesis that $\tau = 0$, if in truth $\tau = 0.49$?

   (d) Now suppose instead that the true treatment effect is 1.0. What is the power of the test? Compare your answer to (c). Do you have more power against smaller effects or bigger effects?

   (e) Now, suppose Dunning and Harrison are filling out a grant application before they do their study. The funders ask them to calculate the power of their test. They don't know the true effect size but they guess (based on the extensive previous experimental literature on cousinage) that the standard error of the estimated effect is 0.22. Then, they consider two alternative proposals: one in which they presume a true effect of 0.49, and another in which they presume an effect of 1.0. Which proposal do you think the donor would be more likely to fund, and why?

6. We will now turn to an example in R in which we use simulations to evaluate the statistical power of an experiment, as a function of sample size. To do this, you will fix the N and the effect size and simulate 500 hundred experiments and calculate the proportion for which we can reject the null hypothesis. You will then vary the N to see how this proportion varies. Before running your, set the seed to 54321.

   (a) You first need to create your assumed "box" for the simulation. You will start by fixing N = 100. Now, use R to draw the control potential outcomes from a normal distribution with mean = 0 and sd = 1. To construct the potential outcomes under treatment, add .25 to each control potential outcome.

(b) Now take the box as fixed and:

  i. generate a treatment vector by randomly assigning half of the units to the treatment group and the other half to control.

  ii. get the "observed" outcomes under that vector.

  iii. use your t-test function to test the hypothesis that the ATE is different from zero.

  iv. store an indicator of whether you reject the null hypothesis (take $\alpha = 0.05$).

(c) Repeat the procedure in (b) 500 times. Indicate the power of your test.

(d) Repeat (a)-(c) for N= 300, N= 500, N= 700, and N= 900.

(e) Plot statistical power as a function of sample size, using results from (b)-(d).

7. Professor Smedley is interested in the effects of owning property on policy attitudes. Smedley theorizes that owning property will make people more supportive of market-based economic policies. After conducting extensive fieldwork in Guatemala, he believes he has found the perfect natural experiment. Following the passage of a new law in parliament, squatters in certain houses in Guatemala City were given formal titles to their land. The new law stated that all squatters could go to the municipal office and request property rights. According to Smedley's research, the officials would then grant property rights in a haphazard (as if random) manner. Smedley conducts a survey of some squatters. He only includes in his final dataset those squatters who applied for property rights. Smedley collects the following variables (which he has made publicly available in 'Smedley.RData").

```
x1 - Indicator for being awarded property rights at time t0
x2 - Indicator for if the applicant just ahead in line of the squatter
 was granted property rights at time t0.
x3 - Fraction of household income from agriculture at time t0
x4 - Indicator for minority status  at time t0
x5 - Indicator for TV ownership at time t0
x6 - Indicator for living in southern neighborhood at time t0
x7 - Indicator for voted for Mayor at time t0
y1 - A measure of support for liberal economic policies at time t1.
y2 - Neighbor's support for liberal economic policies at time t1.
```

Smedley assumes the Neyman model. To estimate the average causal effect, he regresses $y1$ on $x1$ and finds that being awarded property rights increases support for liberal economic policies by about 2.7 units. Note that `t1` is 2 months after `t0`.

Your task is to evaluate the strength of Smedley's design using balance and placebo tests. Do the assumptions of Smedley's analysis hold?