POL SCI 231b (Spring 2017): Problem Set 1

Prof. Thad Dunning/GSI Natalia Garbiras-Díaz

Dept. of Political Science

University of California, Berkeley

Due Friday, January 27 in section

Please email only **one** copy of your solution set per group, with the names of all group members on it. Please use the following email address: nataliagarbirasdiaz+231b@berkeley.edu. If you need to turn in a hard copy, please do so in section and only turn in one per group.

Remember to work out the problems on your own, before you meet with your group to agree on solutions.

1. (From Freedman, Pisani, and Purves—hereafter FPP). A gambler will play roulette 50 times, betting a dollar on four joining numbers each time (like 23, 24, 26, 27 in FPP 4th edition, figure 3, p. 282). If one of these four numbers comes up, she gets the dollar back, together with winnings of $8. If any other number comes up, she loses the dollar. So this bet pays 8 to 1, and there are 4 changes in 38 of winning. Her net gain in 50 plays is like the sum of ____ draws from the box ____. Fill in the blanks; explain.

2. (From FPP). You are thinking about playing a lottery. The rules: you buy a ticket, choose, 3 different numbers from 1 to 100, and write them on the ticket. The lottery has a box with 100 balls numbered from 1 through 100. Three balls are drawn at random without replacement. If the numbers on the balls are the same as the numbers on your ticket, you win. (Order doesn't matter). If you decide to play, what is your chance of winning? Briefly explain your answer.

3. (From FPP). (a) Four draws are going to be made at random with replacement from a box with five tickets in it. The tickets are labelled **1**, **2**, **2**, **3**, **3**, respectively. Find the chance that **2** is drawn at least once. (b) Repeat (a), if the draws are made at random without replacement.

4. There are currently 12 students registered for our class. (a) Suppose I take a survey of the 12 students to find out who is left-handed. Is the number of left-handers a random variable? (b) Suppose I find that there are 8 right-handers and 4 left-handers. Is the difference statistically significant? Explain your answers. (Note: we are covering hypothesis testing on Thursday 1/26, but you may already have the knowledge you need to answer this question.).

5. Consider a randomized experiment in which N units are assigned to a treatment or control group. Denote the m units assigned to treatment by $i = 1, ..., m$, with $Y^T = \frac{\sum_i^m Y_i}{m}$, and the $N - m$ units assigned to control by $i = m + 1, ..., N$, with $Y^C = \frac{\sum_{m+1}^N Y_i}{N-m}$. Here, Y_i is an observed outcome. Let $\overline{Y(1)} \equiv \sum_i^N \frac{Y_i(1)}{N}$ and $\overline{Y(0)} \equiv \sum_i^N \frac{Y_i(0)}{N}$. (Note that $\equiv$ means "identically equal to.") Now, answer the following questions and briefly explain your answers:

   (a) What is the quantity $\overline{Y(1)} - \overline{Y(0)}$? Is this an estimator and if so, what does it estimate?

   (b) What is the quantity $\frac{1}{N} \sum_i^N (Y_i(1) - \overline{Y(1)})^2$? Is this an estimator and if so, what does it estimate?

   (c) What is the quantity $Y^T - Y^C$? Is this an estimator and if so, what does it estimate?

   (d) For each of the estimators you identified in parts (a)-(c), show that the estimator is unbiased for its estimand (i.e., the parameter it estimates).

6. Suppose we draw $i = 1, ..., n$ tickets at random with replacement from a box. The mean of the tickets in the box is μ, and their variance is σ^2. Denote the value of a given draw of a ticket by Y_i and the average of all n tickets by $\overline{Y}$. For each of the following statements, say which aspects of this sampling process must be invoked for the statement to true. Explain your answers, using English as well as algebra (where appropriate).

   (a) $E(Y_1) = \mu$

   (b) $E(Y_2) = \mu$

   (c) $E(\overline{Y}) = \mu$

   (d) $\text{Var}(Y_i) = \sigma^2$

   (e) $\text{Var}(\overline{Y}) = \sigma^2/n$

   (f) $Y_i \sim$ i.i.d. (Read $\sim$ as "is distributed")

   (g) Y_i is a random variable

7. Now suppose we draw n tickets at random without replacement. Which of the statements in parts (a)-(g) of the previous question are true and which are false? Explain your answers.

8. (R exercise) Consider the dataset called "potential outcomes" that is loaded at the Problem Set 1 folder on the class bCourses page. Write and turn in R code that does the following (please comment your code to indicate what each line or chunk of code is doing), and answer the conceptual questions as comments in your code:

1. Find the difference in average potential outcomes under treatment and control, for the study group. What is another term for this quantity? Is this an estimator or a parameter, and why?
2. Simulate an experiment in which 1/2 of the units are assigned to the treatment group and 1/2 are assigned to control. Calculate the difference of means between the treatment and control group.
3. Replicate step (2) 10,000 times, saving the difference of means for each replicate.
4. Plot a histogram for the 10,000 differences of means, and add a vertical line showing the average causal effect you calculated in (1).
4. Is the difference of means in 2. a parameter or an estimator? If the latter, what does it estimate, and does the evidence suggest that it is biased or unbiased? Explain your answer.
5. What is the standard error of the difference of means? Use your code and plot to answer this question, and explain your answer.
6. Repeat the simulation in the previous exercise but simulating experiments that vary the proportion of units in the treatment and control group.

- 1/5 of the units are assigned to treatment, 4/5 to control.
- 1/3 of the units are assigned to treatment, 2/3 to control.