POL-GA 1251
Quantitative Political Analysis II
Homework 1

Due at the beginning of class Thursday next week.

1. Suppose a population P of N units indexed by $i = 1, ..., N$. Now suppose that I do something ridiculous: I randomly sample, with equal probability for all units, only *one* unit from the population. For this unit, i, I flip a fair (50-50) coin. If it comes up heads, I assign treatment and if tails, I assign control. Indicate treatment status by $D_i = 1$ if treated and $D_i = 0$ if control. I record treatment the status, D_i. Potential outcomes for $D_i = 0, 1$ are given by (Y_{1i}, Y_{0i}), respectively. The average treatment effect for the population (ATE) is given by

$$\rho = \mathrm{E}\left[Y_{1i} - Y_{0i}\right].$$

For the one unit in my sample, I observe the usual

$$Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}.$$

I then compute

$$\hat{\rho} = \begin{cases} 2Y_i & \text{if } D_i = 1 \\ -2Y_i & \text{if } D_i = 0 \end{cases} = 2D_i Y_i - 2(1 - D_i) Y_i.$$

Given this research design, I want to know the following.

1. True or false: $\hat{\rho}$ is unbiased for ρ. Demonstrate. (5 points)
2. True or false: $\hat{\rho}$ is consistent for ρ as $N \to \infty$. Explain (you can explain in words). (5 points)

2. Suppose a population P in which units are characterized by an "instrumental variable," $Z_i \in \{0,1\}$ and an "endogenous treatment function" that depends on the instrumental variable, $D_i(Z_i) : \{0,1\} \to \{0,1\}$.

For some units in P, $D_i(1) = D_i(0) = 1$ (always-takers), for others, $D_i(1) = 1$ and $D_i(0) = 0$ (compliers), and for a third subset of P, $D_i(1) = D_i(0) = 0$ (never takers). (We assume that there are no "defiers.")

Potential outcomes, (Y_{1i}, Y_{0i}), are defined with respect to $D_i(Z_i)$ (not Z_i). Therefore, what we observe is $Y_i = D_i(Z_i)Y_{1i} + [1 - D_i(Z_i)]Y_{0i}$. In other words, Z_i has no effect on Y_i except through its effect on $D_i(Z_i)$. This part of what it means to say that Z_i is an "instrument" for D_i; it is called the "exclusion restriction" assumption for instrumental variables.

The instrument, Z_i, is randomly assigned such that $Z_i \perp\!\!\!\perp (Y_{1i}, Y_{0i})$ and $Z_i \perp\!\!\!\perp (D_i(0), D_i(1))$. $D_i(Z_i)$ is obviously endogenous to Z_i, with values depending on whether i is a complier, always-taker, or never-taker.

The average treatment effect for the population from which i is drawn is $\rho = \mathrm{E}[Y_{1i} - Y_{0i}]$, as usual. Consider a strategy for estimating ρ that uses Z_i as a control variable. Specifically, we consider a "stratified estimator" that controls for Z_i by (i) partitioning the sample by values of Z_i, then (ii) taking the difference in treated and control means within each of these strata, and then (iii) combining these stratum-specific estimates with a weighted average, where we weight each stratum contribution by the share of the P in each stratum.[1]

True or false: this estimation strategy using Z_i identifies ρ. Explain why in intuitive terms and then provide a proof for your answer. (10 points)

[1] Here's how to express this formally: To estimate $E[A]$ by conditioning on a variable, B, via stratification simply means that one computes conditional expectations holding B fixed to various values, and then marginalizes over the distribution of B. So, for B discrete, we would estimate $E[A|B = b]$ holding B fixed to all possible values of b, and then marginalize as,

$$\sum_b E[A|B = b]Pr[B = b].$$

3. The November 2017 issue of the APSR contains articles by (i) Kostelka, (ii) Graham et al., (iii) Lerman et al., (iv) Healy et al., and (v) Charnysh and Finkel that make either explicit or implicit empirical claims about causal effects. Create a table that in which you briefly (a sentence or two) answer each of the following questions for each of these studies:

1. What is/are the causal effect/s of interest?
2. What identification strategy/ies (implicit or explicit) do the authors use?
3. What would be an ideal hypothetical experiment (don't worry if it isn't feasible or realistic) that would allow you to get at this causal relationship more convincingly?

(10 points)

4. This exercise has you use simulation to obtain an heuristic understanding of the central limit theorem. Open the dataset pop.dta and make histograms of the variables X_i and Y_i. Are they skewed? Symmetric?

Now, carry out a simulation study on convergence to normality for the mean of X_i and mean of Y_i computed from 1,000 replicates of simple random samples (without replacement) of sizes $N = 10, 50, 250, 500$.

Then, simulate a set of randomized experiments that randomly assign a treatment variable, D_i, such that $M = 10, 50, 250, 500$ units are assigned to treatment ($D_i = 1$) and equivalent numbers to control ($D_i = 0$) (in which case $N = 2M$). For each value of M, run 1,000 replicates. In each experiment, we record a value W_i equal to X_i for treated units and Y_i for control units,

$$W_i = D_i X_i + (1 - D_i) Y_i$$

and compute the difference in means across treated and control, $\hat{\rho}$,

$$\hat{\rho} = \frac{\sum_{i:D_i=1} X_i}{M} - \frac{\sum_{i:D_i=0} Y_i}{M}$$

Study the convergence of $\hat{\rho}$ to normality.

For the two simulations, examine convergence to normality by creating graphs that lay a normal distribution with the appropriate mean and variance over histograms for the different values of N (for the first two simulations) and M (for the last simulation). Briefly comment on how quickly the means and differences-in-means converge to normality as the sample size increases. (10 points)