POL-GA 1251

Quantitative Political Analysis II

Homework 1

Due at the beginning of class Monday next week.

I. 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

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

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

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

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)

II. Take the most recent issue of the APSR (November 2018) and find two empirical articles. Create a table 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. How would you describe the "ideal intervention" or experiment (don't worry if it isn't realistic or ethical) that this paper's analysis is trying to evaluate? Is this intervention well defined in terms of "manipulability"? Discuss.

(5 points)

III. This exercise has you use monte carlo simulation to obtain an heuristic understanding of the central limit theorem.

1. Open the dataset pop.dta and make histograms of the variables X_i and Y_i. Are they skewed? Symmetric? (2 points)

2. Carry out a simulation study on convergence to normality for the mean of X_i and mean of Y_i. Do this by taking a 1,000 replicates of simple random samples (without replacement) of sizes $N = 10, 50, 250, 500$, and computing the respective means for each replicate and sample size. 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. Briefly comment on how quickly the means and differences-in-means converge to normality as the sample size increases. (8 points)

3. Carry out a simulation study on convergence of the difference in means for a randomized experiment. 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, $\hat{\rho}$,

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

   Study the convergence of $\hat{\rho}$ to normality by creating graphs that lay a normal distribution with the appropriate mean and variance over histograms for the different values of M. Briefly comment on how quickly the means and differences-in-means converge to normality as the sample size increases. (10 points)