Causal Inference
Problem Set 1

Due Wednesday April 4th

Please 1) Write up your answers to all questions (including computational questions and any related graphics), 2) include your `R` code *with comments*, in your write-up, and 3) email your write up with an easy-to-recognize file names (e.g., `ps1_grimmer.pdf`). Please ensure that all of these are completed *before* class begins. For any problems that require calculations, please show your work. You are encouraged to work in groups, but you should write up the problem set alone.

Problem 1

Review of unbiasedness and consistency.

(a) Consider a vector y of length N, where $y_i \in \mathbb{R}$ and are fixed numbers (*not* random variables) $\forall\ i$. Now consider a situation in which y_is are being randomly sampled from y (without replacement), and define a random vector z, also of length N, where z_i takes a value of 1 if y_i is sampled and 0 otherwise. For this and all subsequent parts in Problem 1, assume a sample of size n.

Explain what the sample space of the random vector z is, and specify how many elements (i.e. possible outcomes) the sample space contains.

(b) Denote the mean of the y_is as μ_y.
What is the expected value of $\sum_{i=1}^{N} \frac{z_i y_i}{n}$? Is it unbiased for μ_y?

Does your answer to change if we drop the assumption of random sampling? Why or why not?

And if $c \neq 0$ is a constant, is $\frac{c}{n} + \sum_{i=1}^{N} \frac{z_i y_i}{n}$ an unbiased estimator of μ_y? What is its expectation?

(c) Derive the probability limit of $\frac{c}{n} + \sum_{i=1}^{N} \frac{z_i y_i}{n}$. Be very explicit in your use of relevant theorems. Is this estimator consistent for μ_y?

(d) Show that $s^2 = \frac{1}{n-1} \sum z_i (y_i - \bar{y})^2$ is a biased estimator of σ_y^2, the variance of y.

(e) Is $\hat{\sigma}^2 = \frac{1}{n} \sum z_i (y_i - \bar{y})^2$ an unbiased estimator of σ_y^2?

(f) **[Extra credit]** Derive the probability limit of $\hat{\sigma}^2 = \frac{1}{n} \sum z_i (y_i - \bar{y})^2$. Is $\hat{\sigma}^2$ a consistent estimator of σ_y^2?

Problem 2

Review of distribution of the sample mean.

(a) Using `R`, draw 2000 observations from a non-normal distribution with finite mean and variance: specifically, use an exponential distribution with mean one. Call this vector X, and for the rest of the problem consider it fixed (that is, do not redraw it).

(b) Provide a histogram of X and overlay a kernel density estimate.

(c) Draw 1000 random samples, each of size $n = 10$, from the original X (do not redraw X; keep the original 2000 observations). In each of these 1000 samples, estimate the mean and standard error of the mean (*not* the standard deviation of the sample!). Use these mean and standard error estimates to estimate a 95% confidence interval for the mean in each of the 1000 samples (that is, 1000 confidence intervals). Use a normal approximation for the distribution of your mean estimator. Finally, calculate how often these confidence intervals contain the true mean—that is, the mean of the original X vector. In other words: what proportion of the 1000 confidence intervals contain the mean of the original X vector?

(d) Repeat part (c) using samples of size of $n = 25, 50, 100, 200$. For each sample size, what proportion of the 1000 confidence intervals contain the true mean of X? What is going on here?

(e) Repeat parts (c) and (d), using the t distribution (rather than the normal approximation) to select the critical values for the confidence intervals. Are there any meaningful changes? Why or why not?

(f) Use a histogram or kernel density estimator to plot the distribution of the sample means themselves from above in the cases when $n = 10$ and when $n = 200$. Do they look different? Why?

(g) What important theorem from statistics explains the shape of the distribution of sample means when $n = 200$? Briefly describe what this theorem guarantees and why it is so important.

Problem 3

Consider the model:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

where $\mathbf{y}$ is an Nx1 vector of realizations of an outcome variable, $\mathbf{X}$ is an Nx2 matrix where the first column is a constant (a 1xN vector of 1's) and the second column is a single variable X (a 1xN vector of realizations of X, each of which could be denoted x_i), $\mathbf{b}$ is a 2x1 vector of unknown parameters and $\mathbf{e}$ is an Nx1 vector of errors. The variable X takes on the value of 1 if a given observation receives an experimental treatment and 0 otherwise. There are m treated observations and $N - m$ untreated observations.

Prove that the second element of the vector $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is equivalent to the difference between the mean value of $\mathbf{y}$ for treated observations and the mean value of $\mathbf{y}$ for untreated observations. Recall that for any nonsingular 2x2 matrix:

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$\mathbf{A}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Problem 4

To reinforce the intuition behind the potential outcomes framework, consider the fictional data set "POdata.csv." In these fictional data, we observe an outcome for each unit both under treatment and under control (which, again, is usually impossible in the real world).

(a) Define individual level treatment effects and explain the fundamental problem of causal inference.

(b) Define the Average Treatment Effect (ATE) and calculate the ATE in these data.

(c) Plot the distribution of the individual treatment effects. Does the treatment seem to have an effect? How well is it captured by the ATE?