**Part 1: Sampling noise in a fixed population**

1. Develop some data generating process for data X’s and for outcome Y.
2. Write a do-file that creates a fixed population of 10,000 individual observations and generate random X’s for them (use set seed to make sure it will always create the same data set). Create the Ys from the Xs with a true relationship and an error source. Save this data set in your Box folder.
3. Write a do-file defining a program that: (a) loads this data; (b) randomly samples a subset whose sample size is an argument to the program; (c) performs a regression of Y on X; and (e) returns the N, beta, SEM, p-value, and confidence intervals into r().
4. Using the simulate command, run your program 500 times each at sample sizes N = 10, 100, 1,000, and 10,000. Load the resulting data set of 2,000 regression results into Stata.
5. Create at least one figure and at least one table showing the variation in your beta estimates depending on the sample size, and characterize the size of the SEM and confidence intervals as N gets larger.
6. Fully describe your results in your README file, including figures and tables as appropriate.

**Part 2: Sampling noise in an infinite superpopulation.**

1. Write a do-file defining a program that: (a) randomly creates a data set whose sample size is an argument to the program following your DGP from Part 1 including a true relationship and an error source; (b) performs a regression of Y on one X; and (c) returns the N, beta, SEM, p-value, and confidence intervals into r().
2. Using the simulate command, run your program 500 times each at sample sizes corresponding to the first twenty powers of two (ie, 4, 8, 16 ...); as well as at N = 10, 100, 1,000, 10,000, 100,000, and 1,000,000. Load the resulting data set of 13,000 regression results into Stata.
3. Create at least one figure and at least one table showing the variation in your beta estimates depending on the sample size, and characterize the size of the SEM and confidence intervals as N gets larger.
4. Fully describe your results in your README file, including figures and tables as appropriate.
5. In particular, take care to discuss the reasons why you are able to draw a larger sample size than in Part 1, and why the sizes of the SEM and confidence intervals might be different at the powers of ten than in Part 1. Can you visualize Part 1 and Part 2 together meaningfully, and create a comparison table?