Homework #3

POSC 3410 - Quantitative Methods in Political Science

Steven V. Miller

This homework assignment will refer to the `1-simulate-population.R` script I wrote and made available on the Github repository for the course website.[1] This script and the underlying data are patterned from a blog post I wrote in March 2020, and I encourage students to read it since it mimicks the Pollock (2016) chapter on this topic quite nicely.[2] Briefly, this is a simulated population of people who are evaluating some hypothetical (and polarizing) political figure through a "thermometer rating." Higher values in a thermomete rating indicate more "warmth", approximating related but different concepts of "approval" and "support." The data are simulated such that $\mu = 40.01578$ and $\sigma = 40.24403$ for data drawn from a beta distribution. The ensuing data are then bound between 0 and 100 and rounded to integers.

Answer the following questions in two to three *complete* sentences.

1. Get a histogram of these thermometer data. Describe what you see. Your answer should describe the distribution of the data as if you were actually talking a population of subjects (i.e. people) in a hypothetical country/state/territorial unit assessing their feelings toward a hypothetical political figure.
2. Get the mean and median of this thermometer rating (as two estimates of central tendency). Get the standard deviation. What do you see? Recall you are describing a population of cases.
3. Get 5 random samples of size 10. Get the mean for each of those samples. What do you see in these sample means, knowing in advance what the true population parameters are? How would you describe the value of using a handful of samples of this size to estimate the population?
4. Get 100,000 random samples of size 10. What is that mean of sample means? How close is it to the actual population mean?
5. Consider question 4 and explain why this is the case.

Extra Credit

6. Let's get 10 samples of varying sizes: 10, 25, 100, 400, 1,000, 2,000, 4,000, and 10,000. Calculate the means of those and plot them along with associated 95% confidence intervals. If you had only one opportunity to sample this population and had to weigh the conceivable cost versus the benefit of the sample size, what size would you choose and why?

---

[1]https://github.com/svmiller/posc3410/blob/master/homeworks/hw3/1-simulate-population.R
[2]http://svmiller.com/blog/2020/03/normal-distribution-central-limit-theorem-inference/