Assn4stat402

Ronald Barry

9/18/2021

Problem 1:

I want to survey a group of students on how much time they are spending on homework and studying. I have a frame of students that I can contact, along with their age, gender, major and year in college. Do you think it would be worth stratifying on one of these variables? Why or why not? If you were to use a variable to make strata, which would you pick and why?

Problem 2

In the Alaska Department of Fish and Game paper (included with this HW) they use either stratification or poststratification. Which one did they use? What were the strata? Why did they do this?

Problem 3

We really wish to estimate the total expenditures for fuel in a city with $N = 80000$ households. We use random phone dialing to find a SRS of $n = 400$ households. However, we can divide the city into three strata, which we think will include houses with low fuel expense (first stratum), medium fuel expense (second stratum) and high fuel expense (third stratum) based on the typical temperatures and ages of houses in the three regions. We know the size of each stratum ($N1 = 20000, N2 = 30000, N3 = 30000$). However, we can't take a SRS of each stratum since phone numbers are only roughly related to address (and we are using random phone dialing).

However, we can sort the $n = 400$ sampled households into three samples, one from each stratum. We get the following:

- Stratum One: $\bar{x}_1 = \$2500.00, s_1 = \$500.00, n_1 = 120.$
- Stratum Two: $\bar{x}_2 = \$4000.00, s_2 = \$750.00, n_2 = 150.$
- Stratum Three: $\bar{x}_3 = \$5500.00, s_3 = \$750.00, n_3 = 130.$

a. What type of sampling is this?

b. Why is this better than just ignoring the strata and considering this to be an old, boring SRS of size $n = 400$ which we did in the beginning of the course?

c. Find a 95 percent confidence interval for the true mean fuel expenditure in the city. What IS a 95 percent confidence interval (yep, I'll keep asking this until everyone gets it right).

d. Suppose that we comlete the study and decide, before beginning the analysis, to just analyze it without sorting into strata (just using the SRS approach from early in the course). Would this be valid? Why or why not?

Problem 4

We want to estimate the concentration of available nitrogen in the soils of a region. Cold, wet soils generally have more available nitrogen, but also soils with nitrogen fixing (alder) plants or areas showing high productivity might have higher soil nitrogen. We think we can very easily classify plots of ground into either low or high nitorgen plots just by looking at them, but the actual soil sampling and analysis is expensive.

To lower cost, we'll do the following:

1. divide the region into $N = 20000$ reasonably-sized plots. (2) take a SRS of size $m = 500$ plots which we will visit and rapidly classify into either high N stratum or low N stratum (actually this would probably be done as a systematic sample, not an SRS, which we'll see later).
2. We find that $m1 = 300$ of these plots are classified as low nitrogen and $m2 = 200$ plots as high nitrogen.
3. Now we take an SRS of size $n1 = 30$ from the low nitrogen (we hope) plots and, independently, $n2 = 50$ from the high nitrogen plots. We get the following:

Classified as low nitrogen: $\bar{x}_1 = 30ppm, s_1 = 10ppm$ Classified as high nitrogen: $\bar{x}_2 = 40ppm, s_2 = 15ppm.$

a. Does it appear that the stratification will help us much? If we decide it doesn't, can we just pretend we took a SRS of size 500 and ignore the stratification?

b. Find a 95 percent confidence interval for the average nitrogen concentration.