Survey Sampling
Statistics 4234/5234 — Fall 2018

Assignment 3

Reading:

By Tuesday, September 25, read Appendix A and Chapters 1–3 of *Sampling: Design and Analysis, second edition*; by Sharon L. Lohr.

For Tuesday, October 2, read Chapter 4 (pages 117–155) of Lohr.

Homework 3:

The following problems are due in class on Tuesday, October 9. Homework can also be submitted to the course mailbox in Room 904 SSW by 5:00pm on Wednesday, October 10.

1. Suppose that a city has 90,000 dwelling units of which 35,000 are houses, 45,000 are apartments, and 10,000 are condominiums.

   (a) You believe that the mean electricity usage is about twice as much for houses as for apartments or condominiums, and that the standard deviation is proportional to the mean so that $S_1 = 2S_2 = 2S_3$. How would you allocate a stratified sample of 900 observations if you wanted to estimate the mean electricity consumption for all households in the city?

   (b) Now suppose that you take a stratified random sample with proportional allocation and want to estimate the overall proportion of houses in which energy conservation is practiced. Suppose 45% of house dwellers, 25% of apartment dwellers, and 12% of condominium residents practice energy conservation. What gain would the stratified sample with proportional allocation offer over an SRS?

   Calculate $V_{\text{SRS}}(\hat{p}_{\text{SRS}})/V_{\text{prop}}(\hat{p}_{\text{str}})$, and explain what this quantity represents.

2. Compute and interpret an approximate 95% confidence interval for the total number of refereed publications by the 807 university faculty, based on the following simple random sample of 50 faculty members.

Refereed publications	0	1	2	3	4	5	6	7	8	9	10
Faculty members	28	4	3	4	4	2	1	0	2	1	1

3. Install the R package `SDaA` and obtain the data file `agsrs`. Use these data, a simple random sample of $n = 300$ counties from the population of $N = 3078$ counties in the United States, to estimate the total number of acres devoted to farms in the U.S. in 1987. Give a 95% confidence interval.

4. The data file `agstrat`, also available in the `SDaA` package, contains a stratified random sample of counties, using the four census regions of the United States — Northeast, North Central, South, and West — as strata.

h	Stratum	Total counties, N_h	Sampled counties, n_h
1	Northeast	220	21
2	North Central	1054	103
3	South	1382	135
4	West	422	41
	Total	3078	300

Estimate the total number of acres of land devoted to farms in the U.S. in 1987, along with a 95% confidence interval. How does your interval compare to the one computed from a simple random sample?

5. Proportional allocation was used for the stratified sample in the previous problem. You should have found, however, that variability was much higher in the West than in the other regions. Using the estimated variances, and assuming that the sampling costs are the same in each stratum, find an optimal allocation for a stratified sample of size 300.

6. Select a stratified sample of size 300 from the data file `agpop`, using your allocation in the previous exercise (enter `set.seed(5234)` just before you take the sample, to ensure your result is reproducible). Estimate the total number of acres devoted to farming in the United States in 1987, and give the standard error of your estimate. How does this standard error compare with that found in Exercise 4? Give a 95% confidence interval.

7. Researchers used a stratified sample to estimate the number of otter dens along the 1400-km coastline of Shetland, UK. The coastline was divided into 237 5-km sections (excluding five sections that were predominantly buildings), and each section was assigned to the stratum whose terrain type predominated. Then sections were chosen randomly from the sections in each stratum. In each section chosen, the investigators counted the total number of dens in a 110-m-wide strip along the coast. The data are in the data file `otters`, which is available in the R package `SDaA`. The population sizes for the strata are

h	Stratum	Total sections, N_h	Sections counted, n_h
1	Cliffs over 10m	89	19
2	Agriculture	61	20
3	Not 1 or 2, peat	40	22
4	Not 1 or 2, non-peat	47	21

Estimate the total number of otter dens in Shetland, along with a standard error for your estimate. Give a 95% confidence interval.