**Sampling Project—Task #4**

Please answer the following questions briefly but fully. You can use this form by simply typing your answers in under each question. This is to be submitted in class at the time of the final.

1. Why did you choose to do this project and what was the research question that you were trying to answer?

My research question is what proportion of the books in the UCLA Career Center library are published in California.Usually, California does not have that many publishers so my guess is that the proportion should be rather low. I choose to do this project as UCLA Career Center library is the smallest library on campus so sampling this library would not very time-consuming.

1. What was your target population, your sampling frame, and your sampling units? How did you go about enumerating them?

My target population is the books in the UCLA Career Center library. I enumerating them by calculating the proportion of books in on 20 shelves. The sampling unit is the bookshelf and sampling frame is the library catalog in the Career Center library.

1. What was your sampling plan, and how did you go about sampling? Did any issues come up that you had to deal with? Describe the strengths and weaknesses of your sampling plan. How could it have been improved?

My sampling frame is cluster sampling. The computation to calculate the needed cluster size is a little bit tricky. By using cluster sampling, my way of collecting data is rather efficient compared to other methods like SRS. Also because the cluster (which is the shelves) is pretty homogeneous so cluster sampling. However, cluster sampling provides less precision than SRS or stratified sampling. I would improve it by using two-stage cluster sampling as it is more efficient because it might involve less enumeration.

1. Please describe the method of analysis that you decided to use to analyze your observations.

My method of analysis is to use mean and confidence interval to understand the UCLA Career Center library are published in California. Also, I would make a graph and table to further analyze my observation.

1. What is your estimate of the population parameter of interest. Please show your calculations. If you have used Stata or R or SAS to develop your estimates, please include a copy of the output.

I used R STUDIO.

Output:

*1 - level Cluster Sampling design*

*With (20) clusters.*

*svydesign(id = ~per, data = lib, fpc = ~fpc)*

*Probabilities:*

*Min. 1st Qu. Median Mean 3rd Qu. Max.*

*0.06642 0.06642 0.06642 0.06642 0.06642 0.06642*

So it tells me 6.642% of books in UCLA Career Center library are published in California

1. Generate an estimate of the sampling variance and a confidence interval for your estimate. If you have used Stata or R or SAS to develop your estimates, please include a copy of the output.

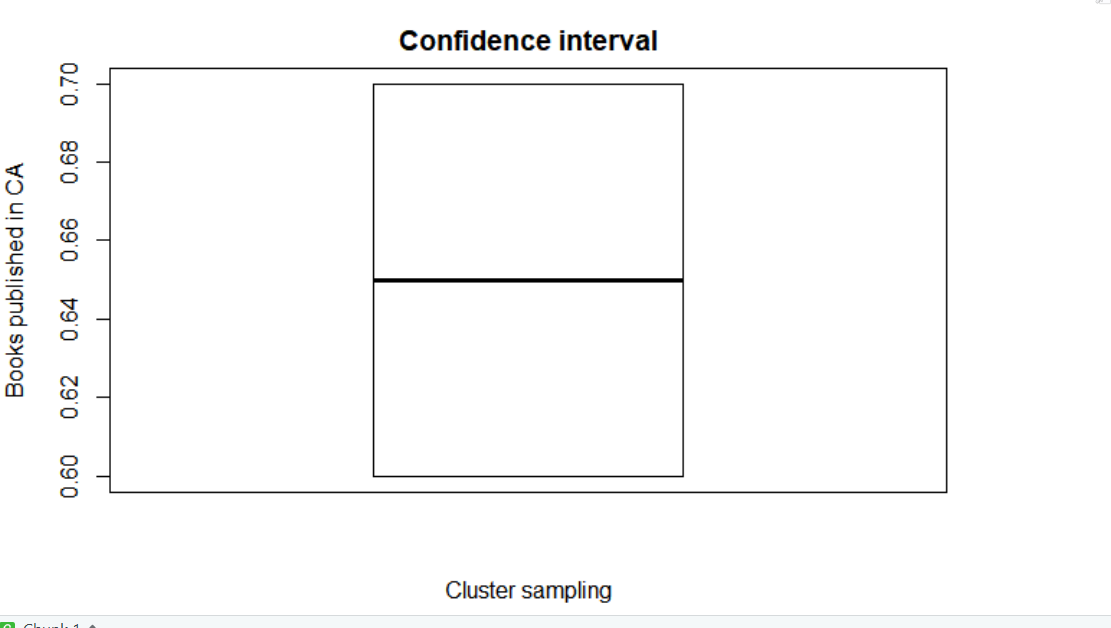
Output:

*2.5 % 97.5 %*

*per 0.060187 0.072601*

So confidence interval is between 0.060187 to 0.072601.

1. Please clearly and concisely present your results using both a graph and a table.



1. Discuss your findings briefly. Please use your sample estimates to make an inference about the population parameter. Given your experiences in collecting the data, do you have any concerns about nonsampling error and other sources of bias? Discuss the strengths and limitations of your estimate.

I found out that only a small proportion of books is published in California. I did encounter some problems with the non-response error where i could not find the location of the publication of a book (I treated these books as published outside of California). I would say my estimate is pretty precise though using a stratified sampling or SRS could improve the precision. Also the non-response error further limit the precision of my estimate

1. Please attach a copy of your sampling frame enumeration and your raw data to this report.