**Sampling Project—Task #3**

If you have changed your mind about what you want to sample, please briefly redo Tasks #1 and #2 and turn them in as well with this task.

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.

**What is your research question:**

What proportion of Chinese character books in the East Asian Library were published before 1949?

1. **What is your sampling plan?**

My elements are the books in each book shelf. I plan to take a Cluster Random Sample of book shelves from N = 953 shelves. The clusters will be diverse since each shelf will be from a different subject. There are 12 major subjects of books in the East Asian Library. The call numbers are organized according to these subjects (B, C, DS, G, H, J, L, N, NK, PL, PN, and Z). The subjects with the most shelves are B, DS, and PL, so I will sample 4 shelves from each of those subjects. For the rest of the subjects I will only sample 1 shelf. So, I will start sampling from the B shelf until I get to the Z shelf.

1. **How will you handle problems with elements in your sampling frame?**

If a book does not have its publication year, I will only include the book in my estimation of the total number of books.

1. **Which type of sampling design are you going to use, specifically?**

I am planning to use Cluster Sampling.

1. **Why did you select this approach? What are the pros and cons of it?**

I selected this approach because it is a good method for conducting a sampling project in a library setting. It is very difficult for me to list all of the books in the East Asian Library because there are thousands of books, the majority of which are Chinese. Another reason I chose this approach is because books are organized into book shelves, so this makes it easier for me to create clusters for sampling. The main advantage of this approach is that it saves time compared to other sampling methods. The main disadvantage of cluster sampling is that it provides less precision than SRS or stratified sampling.

1. **What level of precision do you want to achieve in producing your estimate of the population characteristic that you specified in Task 2?**

Estimate of variance:

0.0046

1. **Given your desired level of precision, how big a sample should you select?** Here you should do a calculation. You may need to do a small pilot study to get an estimate of the population variance.

Since I am calculating a proportion under Cluster Sampling, I need to use the following formula to determine my sample size:

Sample size:

Pilot study:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Shelf # | 33 | 115 | 132 | 165 | 189 | 230 | 240 | 551 | 921 |
| Cluster shelf | C | G | H | J | L | N | NK | PN | Z |
| # of books published before 1949 | 1 | 0 | 0 | 1 | 1 | 4 | 3 | 0 | 0 |
| Total | 187 | 226 | 206 | 133 | 254 | 215 | 261 | 251 | 261 |

|  |  |  |  |
| --- | --- | --- | --- |
| Shelf # | 5 | 40 | 450 |
| Cluster shelves | B | DS | PL |
| # of books published before 1949 | 4 | 2 | 4 |
| Total | 178 | 104 | 208 |
| Shelf # | 13 | 59 | 680 |
| # of books published before 1949 | 1 | 2 | 7 |
| Total | 130 | 154 | 247 |
| Shelf # | 26 | 62 | 789 |
| # of books published before 1949 | 6 | 11 | 2 |
| Total | 161 | 178 | 133 |
| Shelf # | 29 | 92 | 821 |
| # of books published before 1949 | 1 | 2 | 6 |
| Total | 198 | 198 | 214 |