Samping Proj #3 stat402

11/19/2021

Proposal

Stratified Two-Stage Cluster Sampling

First we can stratify the population of bookcases by sparsity. In each strata we will perform a Two-stage cluster sample. In the first stage we take an SRS of the bookcases, In the second stage we take an SRS(most likely a systematic sample) of books, count their pages and compute the mean number of pages per bookcase. Note that this means that we will need to count the total number of books in each bookcase that we sample.\
Recall the stratified sample estimator for the total of a population,

$$\hat{\tau}_{total} = \sum_{i=1}^{K} \hat{\tau}_i$$

$$Var(\hat{\tau}_{total}) = \sum_{i=1}^{K} Var(\hat{\tau}_i)$$

Here there would be N strata split by bookcase sparsity. N_i is bookcase totals in the ith strata. n_i is number of sampled bookcases in the ith strata. $\hat{\tau}_i$ is the total pages per strata computed by the two-stage cluster sample estimator. $Var(\hat{\tau}_i)$ is the variance from that estimator.

In the two-stage cluster sample we let N be the total number of bookcases, n is the number of sampled bookcases, M_i is the total number of books in the ith bookcase, m_i is the sample of books in the ith bookcase, $\bar{y}_i$ is the mean pages per book, and $\hat{s}_i^2$ is the variance of $\bar{y}_i$.

$$\hat{\tau}_i = \frac{N}{n} \sum M_i \bar{y}_i$$

Here is the variance estimator. Note we are substituting $n / \sum M_i$ since we don't know M the total number of books.

$$V(\hat{\tau}_i) = M^2 \left(\frac{N-n}{N} \left(\frac{N}{M} \right)^2 \frac{MSE}{n} + \frac{1}{nN} \left(\frac{N}{M} \right)^2 \sum M_i^2 \frac{M_i - m_i}{M_i} \frac{\hat{s}_i^2}{m_i} \right),$$

Simplifying we get,

$$V(\hat{\tau}_i) = (N-n)N \frac{MSE}{n} + \frac{N}{n} \sum M_i^2 \frac{M_i - m_i}{M_i} \frac{\hat{s}_i^2}{m_i}.$$

Where the MSE is equal to,

$$MSE = \frac{\sum (M_i \bar{y}_i - \frac{\hat{\tau}_i}{N})^2}{n-1}.$$

Strata Determination

I suppose we can walk around to each bookshelf and quickly rank it's sparcity. We could also determine strata by aisle sparcity.

Bookcase sampling

We can identify each bookcase via coordinates with the orgin at the northwestern most point of the Q-Z Block. Later we can rank each bookcase to form our strata. Then we can simply subset the dataframe by sparcity rank and sample the bookcases.

```r
## Bookcase totals by aisle from west to east
BookcasesByAisle <- c(21, 21, 26, 26, 26, 26,
                      26, 26, 26, 26, 21, 21,
                      26, 26, 22, 22, 15, 15,
                      15, 15, 13, 13, 15, 15,
                      15, 15, 12, 12)

## Generate Bookcase Coordinates
BookcaseCoordinatex <- vector()
BookcaseCoordinatey <- vector()
for (i in 1:length(BookcasesByAisle)){
  for (j in 1:BookcasesByAisle[i]){
    BookcaseCoordinatex <- append(BookcaseCoordinatex, i)
    BookcaseCoordinatey <- append(BookcaseCoordinatey, j)
  }
}


df <- data.frame(BookcaseCoordinatex, BookcaseCoordinatey)
head(df)
```

```
##   BookcaseCoordinatex BookcaseCoordinatey
## 1                   1                   1
## 2                   1                   2
## 3                   1                   3
## 4                   1                   4
## 5                   1                   5
## 6                   1                   6
```