

Practical Tools Sampling Project

Stacey Frank & Chendi Zhao

2022-03-26

Contents

Introduction	1
Sample Design	2
Target Population and Goal of Sample Design	2
Method of Selection	2
Sample Selection	3
Composite Measure of Size and Selection Probability	3
Quality Control Checks	3
Selected units and characteristics of each	4
Selection Probabilities of units at each stage of selection	5
Description of how persons should be selected from area listings	6
Anticipated Precision	6
Variance Estimation	6
Maps	6
Appendix	6

Introduction

This report will outline the process for sample design and selection for a sample of census tracts, block groups, and persons from Prince George's County, Maryland. This sample was designed to allow for estimates of the proportion of persons in different age groups who have civic awareness. Civic awareness will be measured in a survey by asking respondents questions about the name of their district representative in the U.S. House of Representatives, the name of their local delegate to the Maryland house of Delegates, and other indicators.

A three-stage cluster sample was drawn, with probability proportional to size (PPS) selection of 15 primary sampling units (PSUs), PPS selection of 1 secondary sampling unit (SSU) within each PSU, and a simple random sample (SRS) selection of elements within each SSU.

First, this report will explain the overall sample design and the method of assigning measure of size to PSUs and SSUs. Next, we will describe the method of sample selection and the units that were selected. Lastly, we will discuss the precision of estimates that can be anticipated from this sample, and the process for correctly measuring the variance of estimates in the achieved sample.

Sample Design

Target Population and Goal of Sample Design

The target population for this study is the adult (18+) non-institutionalized population of Prince George’s County, Maryland. The sample frame is the United States 2010 decennial census. The population for this study’s sampling frame includes approximately 657,421 persons.

The primary goal of this sample design is to allow the estimation of the proportion of the Prince George’s County, Maryland population that has certain markers of civic awareness. The client desires to conduct this analysis within three age groups: people aged 18-44, people aged 45-64, and people aged 65 or over.

The desired total sample size is 300 persons which was split equally among the three age groups. To achieve the desired sample size - which reflects completed questionnaires - one needs to account for non-response. The response rates for the three age groups are anticipated to be 0.60, 0.70 and 0.85, respectively. After adjusting to account for the non-response, the new desired total sample size is 428, with 167, 153, and 118 persons in each age group. Thus, the new overall sampling rate f becomes 0.00065, calculated by $428/657421$. The sampling rate for each age group, f_d can also be obtained using the same formula. The population, desired number of completed interviews, desired sample size, and sampling fraction per age group is listed in Table 1.

Table 1: Desired Age Domain Sample Sizes

Age Group	Population	n	Expected Response Rate	Target Sample Size	Sampling Rate
18-44 years	350725	100	0.6	167	0.00048
45-64 years	225183	100	0.7	153	0.00064
65+ years	81513	100	0.8	118	0.00145
Total	657421	300		428	0.00065

Method of Selection

Given that the goal of this study is to measure civic awareness within these three age domains, a composite measure of size was used in sampling that accounted for the prevalence of persons within these age groups within each cluster. Using this method of selection should ensure that a targeted number of respondents per age group will be achieved in the final sample. Secondary goals of this sample are to achieve these domain sample sizes while also achieving a self-weighting sample within the three age groups and also creating an equal interviewer workload within each PSU. The equal workload for each tract can be calculated by $\bar{q} = 428/(15 * 1) \approx 28.5333$.

As specified by the client, this sample design uses census tracts as PSUs, block groups as SSUs, and persons as elements. In total, there are 218 tracts and 523 block groups in the sampling frame. We will use the composite measure of size (MOS) method to meet the sampling goals. This method can also provide PSU selection probabilities that give “credit” for containing domains that are relatively rare in the population. To be specific, a three-stage cluster sample was drawn, with systematic sampling with probabilities proportional to size in PSUs and SSUs, and a simple random sample of persons within each block group.

Description of sample frame (units, data available, data source) Descriptive statistics for frame

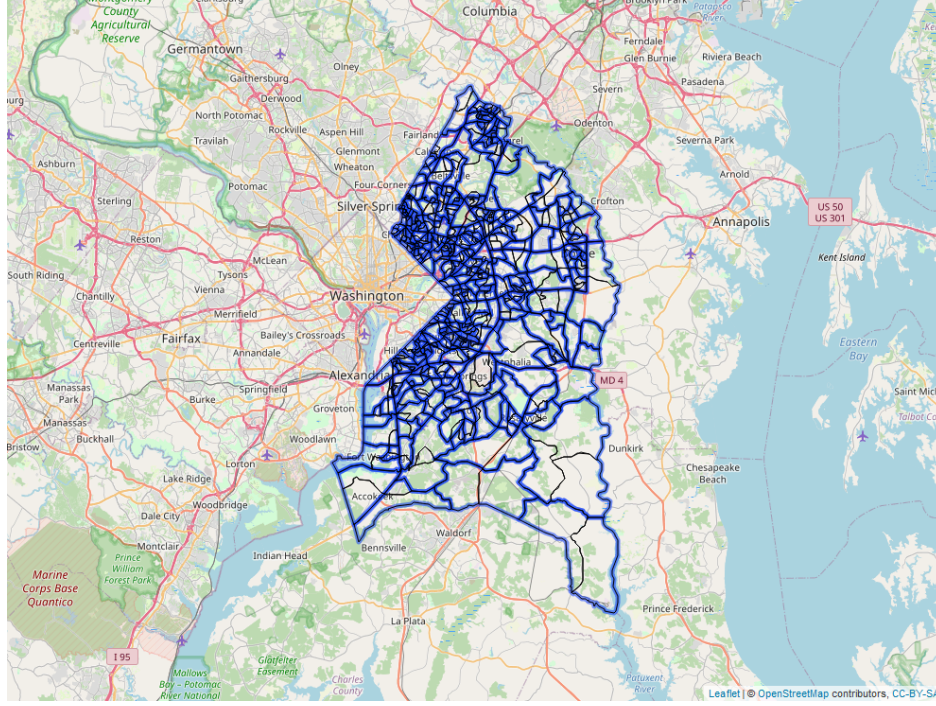


Figure 1: Map of Prince George's County, MD Tracts and Block Groups

Chendi: I put the descriptives in Table 1 above...

Discuss whether a design that uses the same response rate for all persons could be used to achieve the sample size targets for each age group. (Pros and Cons)

Sample Selection

Composite Measure of Size and Selection Probability

The composite MOS for each PSU_{ij} , $S_i = \sum_{j \in U_i} S_{ij} = \sum_d f_d Q_i(d)$, where S_{ij} is the composite MOS for SSU_j in PSU_i and Q_i is defined as number of elements in PSU i that are in domain d. Summing the S_i will give us the total composite MOS, which should be equal to the total desired sample size, 428.

Given that both PSUs and SSUs are sampled with probabilities proportional to the composite MOS, the selection probability of SSU_{ij} is defined as $\pi_i \pi_{k|ij} = mn S_{ij} / S$, where m is the number of sample PSUs and n is the number of sample SSUs in each PSU. Then, we are able to calculate the desired number to be selected from domain d in each SSU with $q_{ij}^*(d) = \bar{q} f_d / S_{ij}$. It is worth to mentioning that \bar{q} is constant in each sampling stage.

Quality Control Checks

After obtaining the information above, we did quality control checks to ensure that the desired sample size is possible for each SSU. The four criterion include:

- (1). $q_{ij}^*(d) \leq Q_{ij}(d)$ for every SSU and domain, $q_{ij}^*(d)$ where is the expected number of sample persons in SSU_{ij} from domain d.
- (2). $\bar{q} \leq Q_{ij}$ for each SSU
- (3). $n \bar{q} \leq Q_i$ for each PSU
- (4). $\pi_i, \pi_{j|i}, \pi_{k|ij}$ less or equal to 1

In the current sample frame, seven block groups listed in Table 2 were detected to be undersized. Based on the map of Prince George’s County, these unqualified areas include an air base, golf course, park land, and a university campus. Therefore, we combined them with the nearest block group within the tract to ensure each cluster met the minimum criteria for selection.

The first three block groups in Table 2 are the only SSUs within that tract. After combining them, the new block group still had a desired sample size larger than the actual population in domain 3. We decided to keep the new group in the frame, since there were no other block groups within the tract that it could be combined with. If this new block group is sampled, we would sample more persons in domain 3 in the next sampled block group to achieve the expected sample size. Block Group 240338024082 and 240338035192 were combined with 240338024082 and 240338035191, respectively. Block Group 240338072002 and 240338072003 are a university campus so there are mainly young adults living there. If we combine them together, there will still be insufficient sample for domain 2 and 3. Also, the population will be very disproportionately contributed across the domains. Therefore, we combined 240338072002 with 240338072001 and 240338072003 with 240338072004.

Table 2: Unqualified Block Groups

NO.	Block Group	Total Units	Domain 1	Domain 2	Domain 3
1	240338011041	0	0	0	0
2	240338011042	0	0	0	0
3	240338011043	2973	1734	183	8
4	240338024082	8	5	2	0
5	240338035192	55	24	2	1
6	240338072002	5219	5200	0	0
7	240338072003	6585	6551	9	9

Selected units and characteristics of each

The sampled block groups are listed in Table 3 with the information of household and population in each domain. We noticed that the workloads are not integers, which means that when the samples of persons within a sample block groups are selected, the sampling need to be done using fixed rates not fixed sample sizes.

Table 3: Sample Result

NO.	Selected Block Group	Total Units	Total Household	Domain 1	Domain 2	Domain 3	Workload
1	Block Group 1, Census Tract 8001.03, Prince George’s County, Maryland	670	311	246	236	59	28.53
2	Block Group 2, Census Tract 8002.06, Prince George’s County, Maryland	2441	1056	1148	509	86	28.53
3	Block Group 2, Census Tract 8002.12, Prince George’s County, Maryland	848	509	101	376	363	28.53
4	Block Group 2, Census Tract 8004.03, Prince George’s County, Maryland	1021	417	301	301	218	28.53
5	Block Group 2, Census Tract 8004.13, Prince George’s County, Maryland	1934	720	703	593	112	28.53
6	Block Group 3, Census Tract 8005.11, Prince George’s County, Maryland	1202	531	484	313	96	28.53
7	Block Group 1, Census Tract 8005.18, Prince George’s County, Maryland	1140	342	308	444	66	28.53

NO.	Selected Block Group	Total Units	Total Household	Domain 1	Domain 2	Domain 3	Workload
8	Block Group 3, Census Tract 8006.07, Prince George's County, Maryland	2012	659	686	656	113	28.53
9	Block Group 2, Census Tract 8009, Prince George's County, Maryland	668	330	170	239	135	28.53
10	Block Group 1, Census Tract 8012.10, Prince George's County, Maryland	2104	799	679	755	243	28.53
11	Block Group 2, Census Tract 8012.17, Prince George's County, Maryland	2371	683	684	765	481	28.53
12	Block Group 2, Census Tract 8013.11, Prince George's County, Maryland	1979	679	690	630	145	28.53
13	Block Group 3, Census Tract 8014.07, Prince George's County, Maryland	1002	376	294	322	171	28.53
14	Block Group 2, Census Tract 8017.01, Prince George's County, Maryland	1664	625	511	553	246	28.53
15	Block Group 1, Census Tract 8018.01, Prince George's County, Maryland	1966	912	882	486	176	28.53

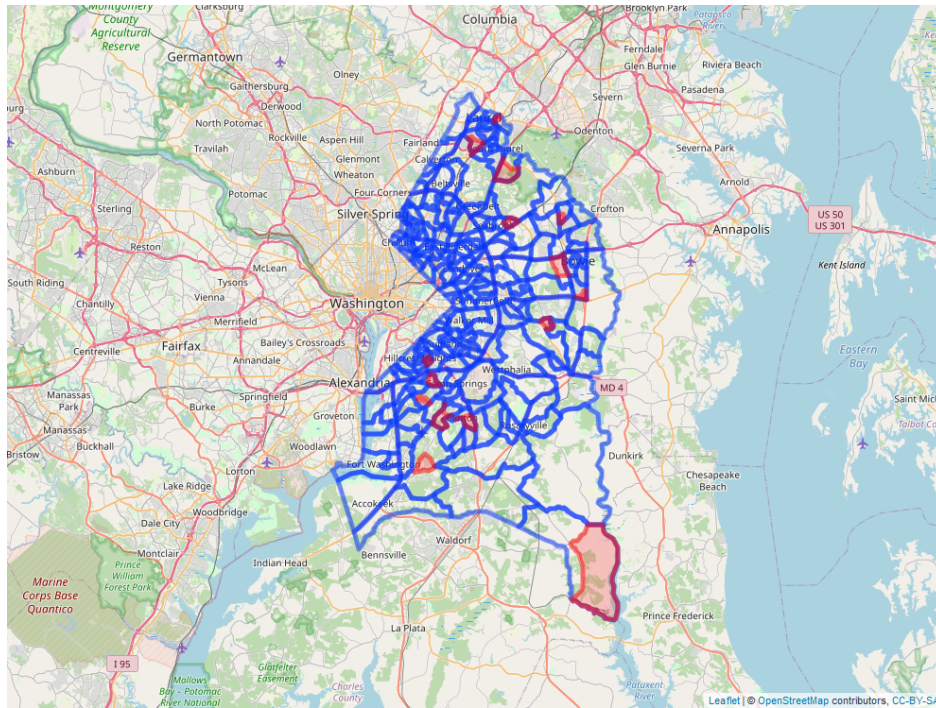


Figure 2: Map of Sampled Block Groups in Prince George's County, MD

Selection Probabilities of units at each stage of selection

Table 4: Selection Probability and Weights

	Min	1st Quantile	Median	Mean	3st Quantile	Max
π_i	0.0235412	0.0529464	0.0658597	0.0688073	0.0818276	0.2678272
π_{ij}	0.0123509	0.0203309	0.0306116	0.0285907	0.0338900	0.0528435
$1/\pi_i$	3.7337504	12.2210103	15.1841080	16.4287669	18.8870896	42.4786709
$1/\pi_{ij}$	80.9655144	49.1863021	32.6673973	34.9764297	29.5072488	18.9237925

.....

Description of how persons should be selected from area listings

Anticipated Precision

The fact that only 1 BG is selected per tract might raise the question of whether variances can be estimated with this design. We can still estimate design-variances because the number of first-stage units is 15, the number of sample tracts. See Textbook 9.2.1

Variance Estimation

Maps

Appendix

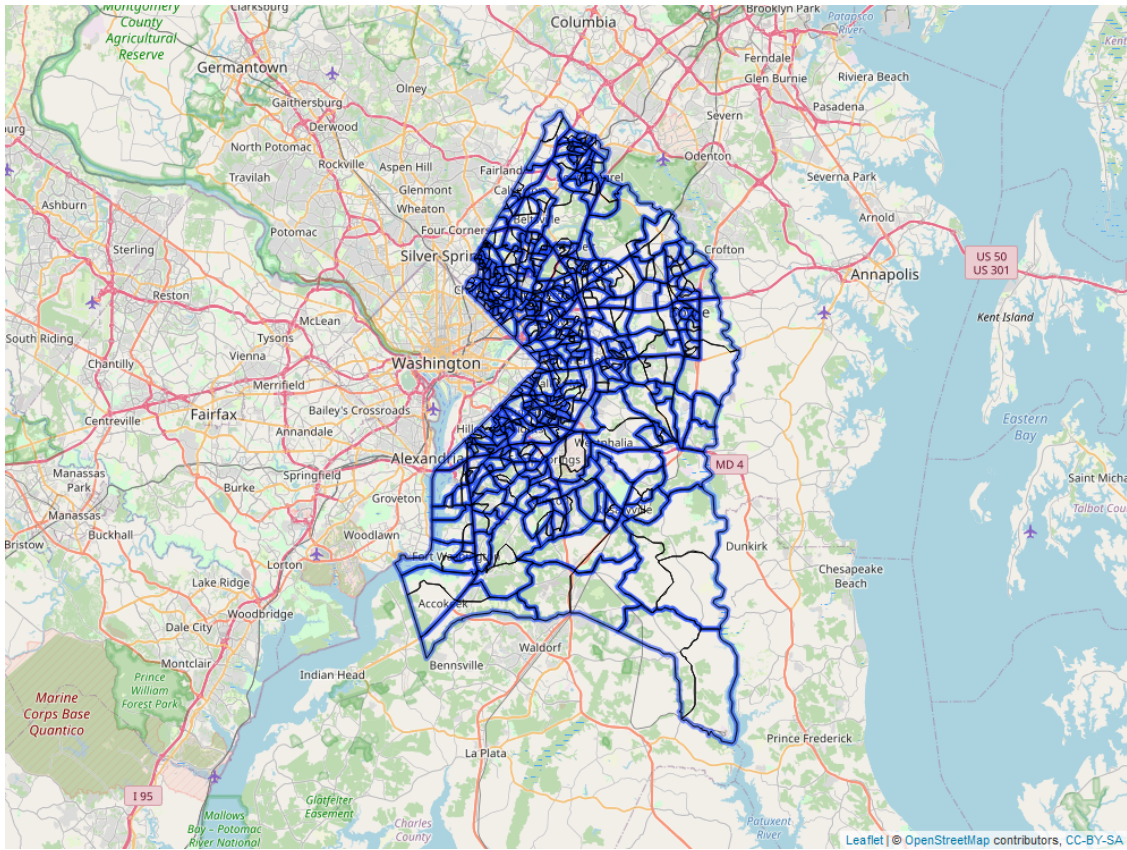


Figure 3: Map of Prince George's County, MD Tracts and Block Groups

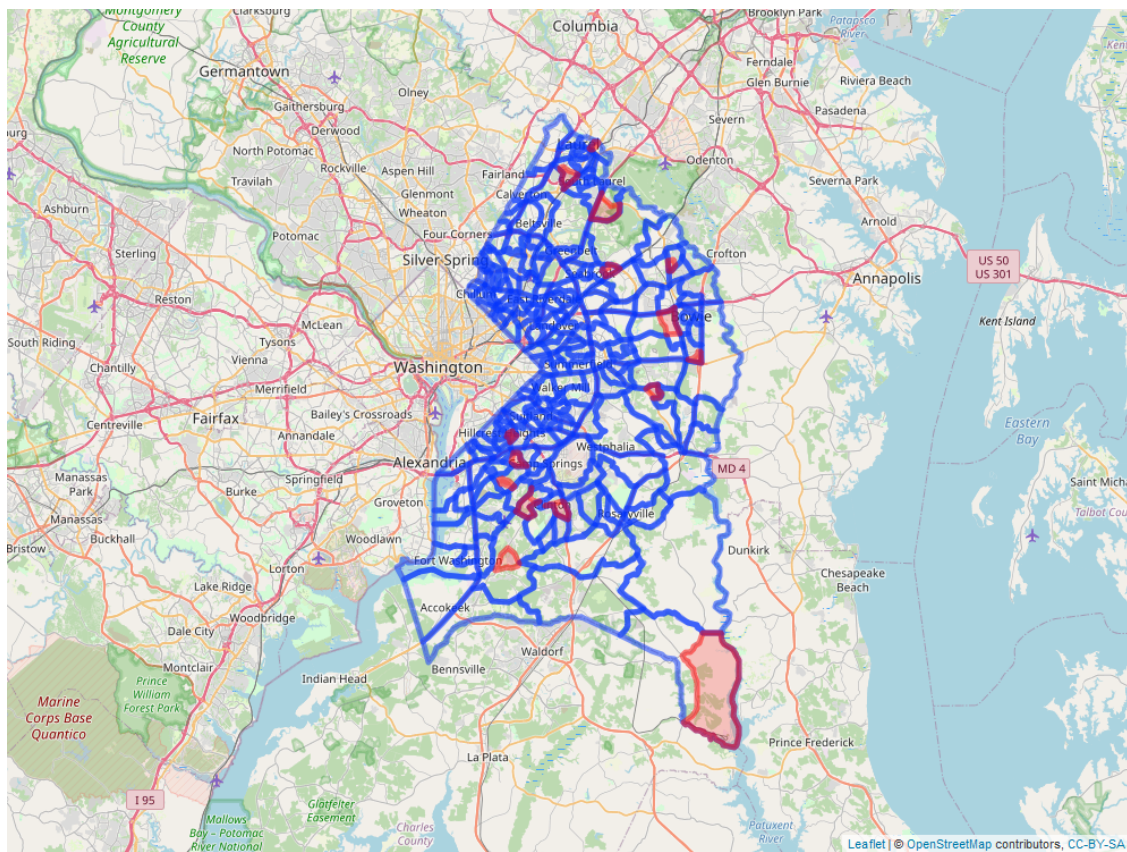


Figure 4: Map of Sampled Block Groups in Prince George's County, MD