Practical Tools Sampling Project

Stacey Frank & Chendi Zhao

2022-03-26

Contents

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

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 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 number of interviews per age group is listed in Table 1.

Table 1: Desired Age Domain Sample Sizes

Age Group	n (Respondents)
18-44 years	100
45-64 years	100
65+ years	100
Total	300

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 sample design uses census tracts as PSUs, block groups as SSUs, and persons as elements. A three-stage cluster sample was drawn, with a PPS sample of 15 tracts, a PPS sample of 1 block group within each tract, and a simple random sample of persons within each block group.

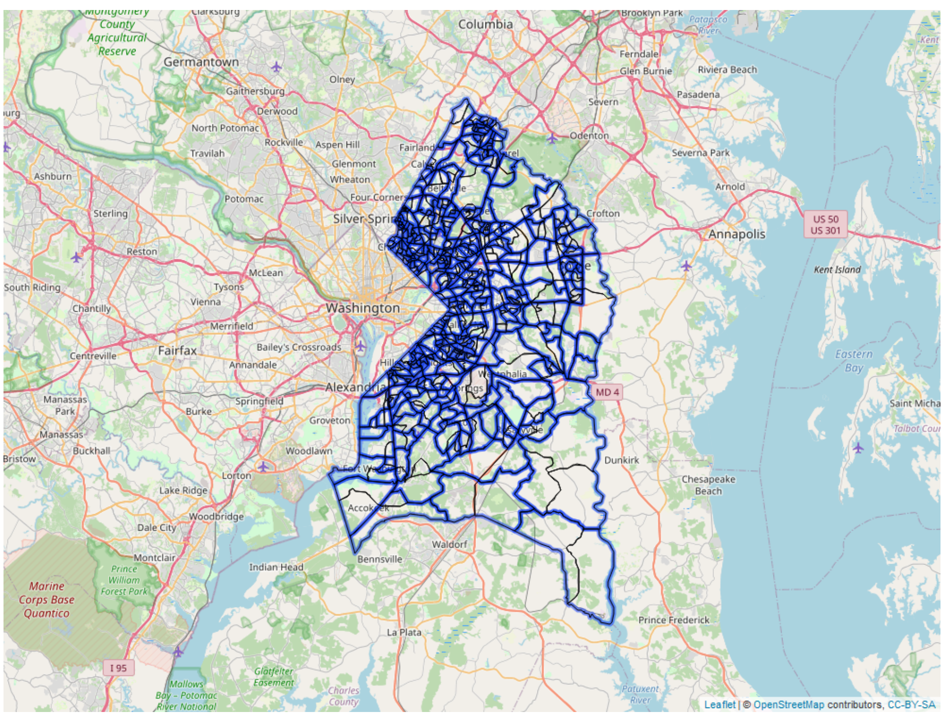

Figure 1: Map of Prince George's County, MD Tracts and Block Groups

Steps: a. Compute composite MOS and selection probability for each PSU and SSU b. Quality checks c. Combining uneligible/feasible SSU d. Repeat step a and construct the sample frame

Objectives: 1.self-weighting 2.equal workload in each PSU

Tracts as PSU; Block groups as SSU; Persons as elements. Domains(d):Age Groups

##Population Information

Textbook 10.5

The purposes of the composite MOS are to get: 1. Self-weighting samples from each of several domains 2. Equal workload in each PSU, i.e., same total sample size in each PSU (across all domains) 3. PSU selection probabilities that give "credit" for containing domains that are relatively rare in the population.

##Calculate composite MOS for Tract and BG (Textbook 10.5) ### See Excel "MOS Tract" & "MOS BG"

Note: The notations in the formulas are not 100% correct, should refer to Textbook P285 ("More Than Two Stages of Sampling")

The composite MOS for PSU(Tract) is defined to be (10.1) $S_i = \sum_D^{d=1} f_d N_{ij}(d)$
The total composite MOS for each PSU(Tract)can be written as $S = \sum_{i \epsilon U} \sum_D^{d=1} f_d N_i(d)$
$=\sum_D^{d=1} f_d \sum_{i \epsilon U} N_i(d)$
The selection probability within each PSU(Tract) $\pi_i = q * (f_d/S_i)$ The desired sample of each PSU(Tract)can be calculated by $\pi_i * N_i$

The composite MOS for SSU (BG) is defined to be (10.1) $S_i = \sum_D^{d=1} f_d N_{ij}(d)$
The total composite MOS for each SSU (BG) can be written as $S_{i+} = \sum_{i \epsilon U} \sum_D^{d=1} f_d N_{i+}(d)$
$=\sum_D^{d=1} f_d \sum_{i \epsilon U} N_{i+}(d)$
The selection probability within each SSU (BG) $\pi_i \pi_{j|i} = q * (S_{i+} * S_{ij})/(S_{++} * S_{i+})$ The desired sample of each SSU (BG) can be calculated by $\pi_i \pi_{j|i} * N_{ij}$

##Check Feasibility of each Tract and BG (Textbook P286) $f(d) \leq S(i)/\bar{n}$

There is one tract and several BGs do not meet the criteria above–Need to check the map and decide whether to drop or combine the BGs (the unfeasible tract actually contains unfeasible BGs, so it might be okay to just deal with the issue in BG-level)

##Then, recalculate the selection probability and desired sample size in each PSU and SSU; Construct weights

##Quality Control Checks (Textbook 11.1) $q_{ij}^*(d) \leq Q_{ij}(d)$ $\bar{\bar{q}} \leq Q_{ij}$ for each SSU $\bar{n}\bar{\bar{q}} \leq Q_i$ for each PSU π_i,$\pi_{j|i}$,$\pi_{k|ij}$ less or equal to 1

Note: There are 13 tracts that only have one BG—- we are not using Sampford's method.

However, notice that this sample design of selecting tracts first, followed by a single BG per tract, is not the same as selecting BG's directly. If we selected BG's directly using Sampford's method, all pairs of BG's would have non-zero joint selection probabilities. Since we select tracts and then 1 BG per tract, the joint selection probability of any two BG's in a given tract is zero.

Sample Selection

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

Appendix