

Constructing and Implementing Sampling Procedures in Census Experiments

Kelly Mathews, U.S. Census Bureau

American Association for Public Opinion Research Annual Conference

Virtual | June 11-12, 2020

Any views expressed are those of the author and not those of the U.S. Census Bureau.

Experiments

- Extending the census environment to the mailing materials (*Mail*)
 - Measure of interest: self-response rate
 - Shares national control panel with *Self-Response* experiment
 - Includes a geographic cluster sample
- Optimization of self-response in the 2020 Census (*Self-Response*)
 - Measure of interest: self-response rate
 - Shares national control panel with *Mail*
 - No geographic cluster sample
- Three additional experiments canceled after the program's sampling plan was well established

Determining Minimum Sample Sizes

$$n \geq \frac{\left((Z_{\alpha^*/2} + Z_{\beta})^2 (p_1(1 - p_1) + p_2(1 - p_2)) \times deff \right)}{\delta^2}$$

n = minimum sample size

δ = minimum detectible difference

α^* = alpha level adjusted for multiple comparisons (Bonferroni)

$Z_{\alpha^*/2}$ = critical value for set alpha level assuming a two-sided test

Z_{β} = critical value for set beta level

p_1 = proportion for group 1

p_2 = proportion for group 2

$deff$ = design effect due to unequal weighting

Determining Minimum Sample Sizes

Experiment	Geographic clustering?	Stratification?	Alpha (α)	Design Effect	Detectable Difference	Minimum Sample Size of Housing Units
Shared National Control Panel (<i>Mail and Self-Response</i>)	No	Yes, by contact strategy	0.014	1.75	0.03	41,500
Geographic Control Panel (<i>Mail</i>)	Yes, by mail route	No	0.050	2.5	0.03	11,000
	Yes, by tract	Yes, by contact strategy	0.050	2.5	0.03	21,500
Evaluation of Self-Response (<i>Self-Response</i>)	No	Yes, by contact strategy	0.014	1.75	0.03	57,100
Extending the Census Environment (<i>Mail</i>)	Yes, by mail route	No	0.050	2.5	0.03	11,000
	Yes, by tract	Yes, by contact strategy	0.050	2.5	0.03	21,500
	No	Yes, by contact strategy	0.014	1.75	0.03	41,500

Sampling Challenges

- Experiments with different designs needed to be sampled from the same frame
 - Stratification differed by experiment: contact strategy, demographic characteristics, vacancy rates
 - Different experiment panels required different sampling units: housing unit, census tract, mail carrier route
- Frame available in pieces
 - Eligible address list released in three waves over four months
 - Final frame size not known until third wave
 - Each wave consisted of 51 state-level files

Sampling Challenges - continued

- Sample requirements were due before the program was finalized because of time needed for programming and testing
 - Unable to change sort order to control for tract and housing unit characteristics prior to sampling for the *Self-Response* experiment
 - Unable to randomly assign treatment after selection of mail carrier route
 - Unable to reduce the number of subframes created for sampling as experiments were canceled

Dealing with Three Waves

- The sampling procedures were written for each of the three waves of housing unit frame data
- Borrowing a similar method from the sampling for the American Community Survey, the nation is divided into seven subframes: one for each treatment and an extra frame
- Each tract in the first wave was systematically randomly assigned to a subframe for sampling
 - Each individual state file was sorted by sub-state geography and the type and language of 2020 Census mailing materials
 - Tracts were then sequentially assigned to one of the seven subframes

Still Dealing with Three Waves

- Tracts in the second and third wave also assigned to a subframe for sampling based on if they were 'new' or 'previously assigned'
 - 'New' tracts, meaning they were never assigned a subframe, were assigned using the same method as tracts in the first wave
 - 'Previously assigned' tracts, meaning the tract was already assigned a subframe in a previous wave, was assigned the same subframe
- Any tract selected for an experiment that was in subsequent waves also had to be assigned to the selected experiment
 - If a tract is selected for an experiment in the first wave and that tract is also in the second wave, then the second wave tract is automatically assigned to the same experiment it was in the first wave

Dealing with State-Level Files

- In order to preserve the sampling order for the next state file, the last treatment panel number assigned along with total number of sampling units eligible for sampling in the current state are retained
 - This allows for a defined start to be calculated for the next state file to continue the sampling process
- We were able to choose the order in which the states would be sampled

Implementing Sampling Procedures

- Those who received the 2020 Census self-response mailings are in scope, about 95 percent¹ of the country
- Started with five experiments with six different sets of treatments for sampling which account for six of the seven sampling subframes
 - Every Door Direct Mailer (Direct Mailer) for *Mail* experiment – sampling mail carrier routes
 - Geographic mailing materials for *Mail* experiment – sampling tracts
 - *Self-Response* and *Mail* experiments – sampling housing units
 - Tailored Contact (canceled) – sampling tracts
 - Vacant Crowdsourcing (canceled) – sampling tracts
 - Citizenship (canceled) – sampling housing units

Sampling for Subframe 1: Direct Mailer

- Supported the direct mailer panel in the *Mail* experiment
- Sorted and sampled at the mail carrier route level, which is below the ZIP code level and is not nested within census tract
- Selected mail carrier routes were systematically assigned to one of two panels: control and treatment
 - Selected mail carrier routes could exist in the other subframes
 - Additionally, a mail carrier cannot deliver a direct mailer to only a subset of housing units on their route
 - Therefore, all housing units on a selected mail carrier route, regardless of subframe, are assigned to a panel

Sampling for Subframe 2: Geographic Cluster

- Supported the sticker wearable incentive panel in the *Mail* experiment
- Removed all housing units already selected for the direct mailer
- Only tracts that met a density criteria were eligible for sampling in order to control for the wide variation in tracts
 - Based on tract land area and number of housing units in the tract
- Sample stratified by 2020 Census contact strategy; sorted by geography along with a 2020 Census mailing language indicator

Sampling for Subframe 3: National Sample

- Supported all but the direct mailer panel in both the *Mail* and *Self-Response* experiments
- Removed all housing units already selected for the direct mailer
- Sample stratified by 2020 Census contact strategy; sorted by geography, a 2020 Census mailing language indicator, and an indicator of selection for the American Community Survey or the Post-Enumeration Survey

Issues during Production Sampling

- Not enough housing units available for selection in a state
- Adjusting sampling for Hard-to-Count direct mailer
- Second wave file had unexpected size

Experiments within the 2020 Census Experiments and Evaluation Program Panel Outline

- Overview of Previous Experiments During the Decennial Census
- Overview and Experimental Design of the 2020 Census Program for Evaluations and Experiments
- Constructing and Implementing Sampling Procedures in Census Experiments
- Operational Considerations for Mailing Materials in Census Experiments
- Improving Internet response rates for the 2021 Canadian Census of Population

Thank you!

kelly.m.mathews@census.gov