Optimum Nonresponse Subsampling Rate for the American Community Survey

Anthony Tersine and Michael Starsinic
US Census Bureau

Outline

- American Community Survey Design
- Methodology
- Results
- Conclusions

American Community Survey Design

- Large Monthly Survey
  - 250,000 Unique Addresses per Month

    (3 Million Unique Per Year)
- Sample Spread Across the Entire Country
- Mail Survey With Telephone Follow-up (CATI)
- 1/3 of Nonrespondents Followed Up In Person (CAPI)
  - 2/3 of Nonmailable Addresses

American Community Survey Design

- Mail Component
  - Initial Letter
  - Questionnaire
  - Reminder Card
  - Second Questionnaire

  - Telephone Failed Edit Follow-Up Operation
    - Incomplete Forms
    - Large Households (6 or more)

Methodology

- Determine the cost function
  - Data collection costs for housing units in US
- Determine the variance function
  - Choose a reliability
  - Solve for the sample size (n)
  - Only as a function of the sampling parameters
- Replace n in cost function
- Minimize the resulting function

Definitions and Costs

- Mail
- Telephone
- Personal Visit

Mail Definitions

n	3,000,000	total annual sample
P_d	0.96	proportion of sample mailable
P_o	0.90	proportion of sample in occupied housing units
R_{mf}	1/3	fraction of mail returns needing TFEFU
R_m	0.50	proportion of mailables returned
R_{m2}	0.40	proportion of mail returns needing second mailing
R_{mo}	0.56	proportion of occupied deliverables returned

Mail Costs

C_{m0}	3.92	cost for each mailout case
C_{mr}	14.85	additional cost for each mail return case
C_{mb}	8.88	cost for mailback and processing returns
C_{m2}	2.33	cost for each second mailing
C_{mf}	15.10	cost for each TFEFU

The value of C_{mr} is calculated as follows:

$$C_{mr} = C_{mb} + R_{mf}\, C_{mf} + R_{m2}\, C_{m2}$$
$$= 8.88 + (1/3) * 15.10 + 0.4 * 2.33$$
$$= 14.85$$

Telephone Definitions

e_t	0.32	proportion of mail non-returns eligible for CATI (good phone numbers)
f_t	1.00	proportion of mail non-returns selected for CATI (current value)
R_t	0.60	proportion of CATI eligible cases interviewed
R_{to}	0.75	proportion of occupied CATI eligible cases interviewed

Telephone Costs

C_{ti}	50.94	cost for each telephone interview
C_{tni}	12.73	cost for each telephone noninterview

Personal Visit Definitions

f_{pd}	1/3	fraction of mailable noninterviews selected for CAPI
f_{pu}	2/3	fraction of non-mailables selected for CAPI
R_p	0.86	proportion of CAPI cases interviewed
R_{po}	0.82	proportion of occupied CAPI cases interviewed (assume all vacants interviewed) N_{pio} / n_p
N_{pio}	298,342	number of occupied interviews in CAPI
n_p	363,840	number of occupied units selected in CAPI

Personal Visit Costs

C_{pi}	145.58	cost for each personal visit interview
C_{pni}	72.79	cost for each personal visit noninterview

Sample Proportions

- proportion of occupied units represented by mail respondents

  $s_m = 0.533333$

- proportion of occupied units represented by CATI interviews

  $s_t = 0.102400$

Sample Proportions

- proportion of occupied units represented by CAPI universe

  $s_p = 0.364267$

- s_p can be split into two components representing mailable and unmailable addresses

Sample Proportions

- proportion of occupied units represented by unmailable CAPI cases

  $s_{pu} = 0.040000$

- proportion of occupied units represented by mailable CAPI cases

  $s_{pd} = 0.324267$

Sample Sizes

- number of sample cases representing occupied unit mail respondents

  $n_m = 1{,}440{,}000$

- number of sample cases representing occupied unit CATI interviews

  $n_t = 276{,}480$

Sample Sizes

- number of sample cases representing occupied unit CAPI universe

  $n_p = 363{,}840$

- n_p can be split into two components representing mailable and unmailable addresses

Sample Sizes

- number of sample cases representing CAPI universe of unmailable occupied units

  $n_{pu} = 72{,}000$

- number of sample cases representing CAPI universe of mailable occupied units

  $n_{pd} = 291{,}840$

Cost per Interview - Mail

$$C_m = C_{m0} / R_m + C_{mr} + [(1 - R_m) / R_m] C_{m2}$$

$$= 3.92 / 0.5 + 14.85 + [(1 - 0.5) / 0.5] * 2.33$$

$$= 25.02$$

Cost per Interview - CATI

$$C_t = C_{ti} + [(1 - R_t) / R_t]\, C_{tni}$$

$$= 50.94 + [(1 - 0.6) / 0.6] * 12.73$$

$$= 59.43$$

Cost per Interview - CAPI

$$C_p = C_{pi} + [(1 - R_p) / R_p]\, C_{pni}$$

$$= 145.58 + [(1 - 0.86) / 0.86] * 72.79$$

$$= 157.43$$

Optimization of Subsampling Rates

- Optimize the subsampling rates f_t, f_{pd}, and f_{pu}
- Minimize cost/variance function
- Use Cauchy-Schwartz inequality
- Two ways
  1. Calculated f_t
  2. Set $f_t = 1$

Results – Optimal Rates 1

- $f_t = 0.648863$
- $f_{pd} = 0.519043$
- $f_{pu} = 0.374116$

Results – Optimal Rates 2

- $f_t = 1.0$
- $f_{pd} = 0.372223$
- $f_{pu} = 0.413479$

Results – Variances 1

Variable	Current Rates	Actual Rates (Option 1)	Rounded Rates (Option 2)
f_t	1.000000	0.648863	0.666667
f_{pd}	0.333333	0.519043	0.500000
f_{pu}	0.666667	0.374116	0.400000
SE	0.020979	0.021577	.021565
CV	20.98%	21.58%	21.57%
90% CI	6.55%, 13.45%	6.45%, 13.55%	6.45%, 13.55%
Total Cost	115,800,000	105,950,000	106,100,000

Results – Variances 2

Variable	Actual Rates (Option 3)	Rounded Rates 1 (Option 4)	Rounded Rates 2 (Option 5)
f_t	1.000000	1.000000	1.000000
f_{pd}	0.372223	0.400000	0.333333
f_{pu}	0.413479	0.400000	0.400000
SE	0.020599	0.020258	0.021210
CV	20.60%	20.26%	21.21%
90% CI	6.61%, 13.39%	6.67%, 13.33%	6.51%, 13.49%
Total Cost	117,950,000	122,140,000	111,580,000

Future Research

- Look at the affect on small areas
- Sensitivity analysis
- Parameters after full implementation of ACS

Conclusions

- Efficiency could be improved
  - Start subsampling in CATI
- Decrease costs by $10 million (Option 2)
  - Almost 3 percent larger standard error
- Decrease costs by $4 million (Option 5)
  - 1 percent larger standard error

Anthony Tersine

Anthony.G.Tersine.Jr@census.gov