Case Prioritization: Value Is Not Enough

Stephanie Coffey

5/15/2016

2016 AAPOR

Austin, TX

Case Prioritization at Census

- Motivated by current survey environment
  - Rising costs, falling response rates, burden concerns
  - Exploring different features and interventions
- Ongoing experiments since 2013
  - 2013 Decennial Census Test
  - Survey of Income Program Participation (SIPP): Q2 2016
  - National Health Interview Survey (NHIS): Q3 2016
- Requires Decisions
  - How to operationalize?
  - How to prioritize?

Case Prioritization

- Allows for Differential Effort
  - Field Surveys
    - Interviewers spend more effort on certain cases
  - Multimode Surveys
    - Differential recruitment through mode assignment
- By Time or Overall
- Need a Way to Assign Priority
  - Measure of value is commonly used
  - Value is often based on a model (not always!)
  - Assign value and sort

Determining Value of a Case

- Related to data collection goals
  - Minimize weight variation?
    - Prioritize by base weight, w_i.
  - Maximize response rate?
    - Prioritize by response propensity, $\hat{\rho}_i^R$.
  - Balance respondent population?
    - Prioritize by balancing propensity, $\hat{\rho}_i^B$.
  - Improve coverage of hard to reach cases
    - Prioritize predicted movers, spawns, new households
- Models are based on dynamic data – value is dynamic

Value Isn't Enough

- Value doesn't consider likelihood of response
- Prioritizing by value only
  - "Throw good money after bad"
  - Continue attempting the same cases
- Consider expected value instead:
  - $E[value_i] = \hat{\rho}_{it}^{R}(value_i) + (1 - \hat{\rho}_{it}^{R})(0)$
  - You only realize the value if the case is completed
- Requires estimate of response propensity
- Prior presentation illustrates bias early

Initial Illustration

- Need data
  - Three months of cases from monthly Census surveys
  - Planning database data (geographic frame data)
  - Contact history data (interim and final outcome data)
- Need two estimates of response propensity at time *t*
  - Estimate based on “true” parameters at end, ρ_{it}^{R}
  - Estimate based on collected data through *t-1*, $\hat{\rho}_{it}^{R}$
- Need estimates of value
  - Case 1: Case values are the randomly assigned $[1,5]$
  - Case 2: Case values correlated with ρ_{it}^{R} $(r = -0.40)$

Initial Illustration

- Three prioritization schemes for each value
  - $E\left[value_{it}|\rho_{it}^{R}\right]$ based on “true” prioritization
  - $E\left[value_{it}|\hat{\rho}_{it}^{R}\right]$ based on “t-1” prioritization
  - $value_{it}$ only: ignoring response propensity
- Evaluation questions
  - How much overlap is there between methods?
  - How much value do you lose amongst prioritized cases?

Results: Prioritization Agreement

Value Assignment	% Cases Prioritized		
	5%	15%	25%
Uncorrelated (Random)	94%	87%	75%
Correlated (Negatively)	92%	78%	68%

Results: Prioritization Agreement

Value Assignment	% Cases Prioritized		
	5%	15%	25%
Uncorrelated (Random)	94%	87%	75%
Correlated (Negatively)	92%	78%	68%

- Partial information does not select the same cases
  - As more cases are prioritized differences increase
  - Correlation between R and V causes larger differences

Results:
Percent Expected Value Obtained

Value Assignment	% Cases Prioritized		
	5%	15%	25%
Uncorrelated (Random)	76%	85%	86%
Correlated (Negatively)	74%	84%	86%

Results:
Percent Expected Value Obtained

Value Assignment	% Cases Prioritized		
	5%	15%	25%
Uncorrelated (Random)	76%	85%	86%
Correlated (Negatively)	74%	84%	86%

- Partial information results in lower realized expected value
- Prioritized cases ended up…
  - Being less valuable than predicted by data through t-1
  - Ended up being more difficult to convert than expected

Conclusions

- Response rates are falling and costs are increasing
  - Make decisions about resource allocation
  - Case prioritization is one example
- Knowledge and data can help make these decisions
  - Reliable estimates of survey parameters is important!
  - Unreliable information may lead to suboptimal choices
- Bayesian methods can help mitigate this problem
  - Leverage external information to improve parameter estimation

Thank you!!

Contact Info:

Stephanie Coffey

stephanie.coffey@census.gov