National Woodland Owner Survey: Statistical Design and Methods

John Coulston
USDA Forest Service, Southern Research Station
Timber Products Monitoring National Leader
Laundry List of Topics

• Probability proportional to size (PPS); the more land that is held by an owner, the higher the probability of an owner receiving a questionnaire
• Questionnaire design. Topics covered, testing of questions and changes over time.
• Determination of sampling intensities
• Implementation methods. How responses are solicited from respondents, response rates, and non-response assessments.
• Illustrate how standard errors differ between state-level, sub-state regions, and county-level estimates from data collected

National

• What does PPS sampling look like?
• How do you determine sample size for a state?
  • Given the sample size what are the implications for sub-state estimates?
• What questions do you ask and why?
• How do you implement?
• How do you deal with non-response?
What does PPS look like?

• Sampling with probability proportional to size (PPS) gives “larger” elements a greater chance of selection than “smaller” ones.

• Specifically, the probability of selecting an element is directly proportional to its size.

• If one element is twice as large as another, it will have double the chance of being sampled.

• In short the probability of inclusion is \( \pi_i \propto \frac{\text{Size}_i}{\sum_{i=1}^{I} \text{Size}_i} \)

• As such higher weight is given to the “larger” elements \( w_i = \frac{1}{\pi_i} \)

• \( w_i \) and \( \pi_i \) are also a function of the number of samples taken (\( n \))

• But… actual inclusion probability depends on how one draws the sample.
What does PPS look like?

- 15 owners in an 256 unit area
- Owners vary in size from 1 unit to 128 units
- If you throw a dart randomly who would you hit?
- What about more darts?
What does PPS look like?.

Hypothetical Population of Owners: 30 sample locations

<table>
<thead>
<tr>
<th>owner</th>
<th>area</th>
<th>( \pi_i )</th>
<th>forest area&lt;sub&gt;i&lt;/sub&gt;</th>
<th>forest Mgt Plan&lt;sub&gt;i&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>128</td>
<td>1</td>
<td>57.0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>64</td>
<td>1</td>
<td>41.5</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>1</td>
<td>10.4</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>0.25</td>
<td>3.0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>0.25</td>
<td>2.6</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>0.0625</td>
<td>0.8</td>
<td>0</td>
</tr>
</tbody>
</table>

Estimation

\[
\hat{Y} = \sum_{i=1}^{I} \frac{y_i}{\pi_i}
\]

Forest area =
\[\sum \text{forestArea}_i / \pi_i = 143 \quad (\text{SE}=2.05)\]

Other estimates made in a similar fashion.
PPS: Leveraging the FIA sample

- The land use for each FIA sample location is known.
- For those FIA locations that are ‘forest use’ the owner class is determined.
  - E.g. family forest = forest land use and owned by individual and family including trusts, estates, and family partnerships
- Stratification process makes separate groups of family forest, small and large corporate.
- Reduces variance
## Stratification Process

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Land use</th>
<th>Ownership Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family forest land</td>
<td>Forest land</td>
<td>Individual and family, including trusts, estates, and family partnerships</td>
</tr>
<tr>
<td>Corporate forest land</td>
<td>Forest land</td>
<td>Corporate, including Native Corporations in Alaska and private universities</td>
</tr>
<tr>
<td>Other private forest land</td>
<td>Forest land</td>
<td>Non governmental conservation / natural resources organization and unincorporated partnerships / associations / clubs</td>
</tr>
<tr>
<td>Tribal forest land</td>
<td>Forest land</td>
<td>Native American (Indian)</td>
</tr>
<tr>
<td>Federal forest land</td>
<td>Forest land</td>
<td>Forest service and other federal agencies</td>
</tr>
<tr>
<td>State forest land</td>
<td>Forest land</td>
<td>State</td>
</tr>
<tr>
<td>Local and other non-federal forest land</td>
<td>Forest land</td>
<td>Local (county, municipal, etc.) and other non-federal public</td>
</tr>
<tr>
<td>Nonforest land and non-Census water</td>
<td>Nonforest land and non-Census water</td>
<td>--</td>
</tr>
</tbody>
</table>
What does PPS look like?.

- Implications of PPS sample arising from a ~ simple random sample (equal probability)
  - Inclusion probabilities must be calculated in a specific manor
  - Inclusion probabilities are estimated
  - Weights are estimated

- Inclusion probability of each ownership*

\[ \pi_i = \frac{a_i n_h}{A_h n_{hi}} \]

- Weight of each ownership

\[ w_i = \frac{1}{\pi_i} \]

- Estimated weights requires bootstrap approaches for SE estimation

*Formula for 100% response. \( \pi_i \) is adjusted for non-response in the survey.
How do you determine sample size for a state?

• Statistical perspective
  • Sample size depends on variability of the population
  • And... acceptable level of precision (e.g. standard error of the estimate)

• Reality
  • Interaction of $ and precision

• Can approximate decision based on data available via GitHub and NWOS R package
Given the sample size what are the implications for sub-state estimates?

- Given the stratification process:
  - Sub-state estimates are domain estimates
  - Similar to other domains (e.g. owner with > 10 ac of forest)
- Question similar to the previous one (how many samples) but need to test differently.
What questions do you ask and why?

Types of questions

• Questions are designed to estimate the parameters of interest
• The parameters can be grouped into the following
  • How many private forest owners there are?
  • Why they own land?
  • What they have done with their forests in the past?
  • What they plan to do with their forests in the future?

Testing questions

• Initial content from listening sessions with users groups
  • Historically used focus groups to test the instruments tested specific words in these focus groups (e.g. forest vs. woodland)
• Now use cognitive interviews
  • “think out loud” interviews with landowners to see how they respond and where they may be having challenges
  • Based on these interviews (and item non-response analyses), one of the biggest recent changes has been to minimize the use of skip patterns
• Pre-testing
  • Pilot tests involving a few hundred private landowners where we use our full implementation methods
How do you implement?

• Primary contact mode is mail (because that is the only information available for all landowners)

• We use a 4-wave contact
  • 1. Intro postcard
  • 2. First questionnaire package with cover letter, questionnaire, and business reply envelope
  • 3. Thank you/reminder postcard
  • 4. Second questionnaire package with modified cover letter, questionnaire, and business reply envelope

• Telephone follow-up mode
  • For a subset of mail nonrespondents telephone interviews are performed by NASS
  • Prior to 2019 most of the questions on the questionnaire were asked in interviews
  • From 2019 onwards it will be a select set of questions (about 5 or 6 of the 40 or so questions)
How do you deal with non-response (after implementation)?

- Two types of non-response: unit and item
- Item: generally imputed
- Unit:
  - Inclusion probabilities and a function $n$
  - Weights are a function of $n$
  - Weights are adjusted given the ratio of actual responses to intended samples
Some remarks

• Documentation on methods was reasonably available and straightforward.
  • R package being available facilitated custom analysis

• Efficiency of PPS approach depends on correlation between owner area and owner behavior.
  • Assume ‘intensifications’ can be done at State partner request and contribution

• Partners are a big part of FIA
  • Includes States and NGOs (for example)
  • Could be opportunities to lesson data collection loads through upward reporting by land owner groups. Analogous with TPO program.