Computation of Survey Weights: Bridging Theory and Practice

Matthew DeBell
American National Election Studies
Stanford University
This Talk

Theory of weighting
Practice
Problems
Improvements made
Improvements to make
This Talk

Theory of weighting
  – What weighting does
  – Theoretical implementation
  – Limitations (what weights don’t do)

Practice
Problems
Improvements made
Improvements to make
What Weights Do

• How many people are represented by each respondent
What Weights Do

• How many people are represented by each respondent

• Fix random error (sampling error)
What Weights Do

- How many people are represented by each respondent
- Fix random error (sampling error)
- Adjust for selection probability
What Weights Do

• How many people are represented by each respondent
• Fix random error (sampling error)
• Adjust for selection probability
• Adjust for nonresponse (Cochran 1968)
What Weights Do

• How many people are represented by each respondent
• Fix random error (sampling error)
• Adjust for selection probability
• Adjust for nonresponse (Cochran 1968)
• Fix non-random error, \textit{within limits}
Why Weights Are Required

• Unequal probability of household selection
  – Area and oversampling
Why Weights Are Required

• Unequal probability of household selection
  – Area and oversampling

• Unequal probability of respondent selection within households of different sizes
Why Weights Are Required

• Unequal probability of household selection
  – Area and oversampling
• Unequal probability of respondent selection within households of different sizes
• Unequal response rates
Why Weights Are Required

• Unequal probability of household selection
  – Area and oversampling
• Unequal probability of respondent selection within households of different sizes
• Unequal response rates
• Known errors
  – Differences from population benchmarks
Theoretical Implementation

• Weight in steps
Theoretical Implementation

• Weight in steps
  – Household probability
    • RDD: Inverse of number of phone lines
    • ABS: Probability of address inclusion
Theoretical Implementation

• Weight in steps
  – Household probability
    • RDD: Inverse of number of phone lines
    • ABS: Probability of address inclusion
  – Person probability (weight by household size)
Theoretical Implementation

• Weight in steps
  – Household probability
    • RDD: Inverse of number of phone lines
    • ABS: Probability of address inclusion
  – Person probability (weight by household size)
  – Nonresponse in observable categories
Theoretical Implementation

• Weight in steps
  – Household probability
    • RDD: Inverse of number of phone lines
    • ABS: Probability of address inclusion
  – Person probability (weight by household size)
  – Nonresponse in observable categories
  – Post-stratify on key factors
Theoretical Implementation

• Weight in steps
  – Household probability
    • RDD: Inverse of number of phone lines
    • ABS: Probability of address inclusion
  – Person probability (weight by household size)
  – Nonresponse in observable categories
  – Post-stratify on key factors

(Alternatives such as propensity scores)
Limitations

Weights do not...
Limitations

Weights **do not**

- fix non-coverage
Limitations

Weights **do not**

- fix non-coverage
- fix extreme nonresponse bias
Limitations

Weights **do not**

- fix non-coverage
- fix extreme nonresponse bias
- fix nonresponse bias for factors uncorrelated with weighting factors
Limitations

Weights **do not**

- fix non-coverage
- fix extreme nonresponse bias
- fix nonresponse bias for factors uncorrelated with weighting factors
- knowably fix errors on factors with unknown population benchmarks (e.g. party ID)
Limitations

Weights **do not**

- fix non-coverage
- fix extreme nonresponse bias
- fix nonresponse bias for factors uncorrelated with weighting factors
- knowably fix errors on factors with unknown population benchmarks (e.g. party ID)
- come for free: weights increase variance
This Talk

Theory of weighting

➢ Practice

Problems

Improvements made

Improvements to make
This Talk

Theory of weighting

- Practice
  - Few implementation specifics in the literature
  - Inconsistent methods
  - Historical example: ANES
  - Other studies
  - User awareness

Problems

Improvements made

Improvements to make
Few Specifics in General Literature

• Groves et al. (2008) *Survey Methodology*
Few Specifics in General Literature

• Groves et al. (2008) *Survey Methodology*
  – has two paragraphs on poststratification
Few Specifics in General Literature

• Groves et al. (2008) *Survey Methodology*  
  – has two paragraphs on poststratification

• Other “comprehensive” and “step by step” guides have little or no mention of weights  
Few Specifics in General Literature

• Groves et al. (2008) Survey Methodology
  – has two paragraphs on poststratification

• Other “comprehensive” and “step by step” guides have little or no mention of weights

• Statistical literature covers statistical theory, not standardizing practice or a detailed guide
Inconsistent Methods in Practice

• Practices are quite varied (Voss, Gelman, & King 1995)
American National Election Studies

• Pre- and Post-election surveys since 1948
• Face-to-face interviews (mostly)
• Cluster samples (about 40-70 PSUs)
• 1,000 to 2,500 cases per survey
ANES Example

• Earlier studies: no weights
ANES Example

• Earlier studies: no weights
• Some studies: household size
ANES Example

- Earlier studies: no weights
- Some studies: household size
- Others: rudimentary cell-based poststratification
ANES Example

- Earlier studies: no weights
- Some studies: household size
- Others: rudimentary cell-based poststratification
- Since 2008: poststratified with raking
Other studies

• Inconsistency
• Lack of documentation (often, not always)
User (Un)Awareness

• ANES literature
This Talk

Theory of weighting
Practice

➤ Problems

Improvements made
Improvements to make
The Problems

• Average survey researcher does not know how to weight well
• Survey statisticians are doing it ad hoc
• Results:
  – not (always) transparent
  – not (often) replicable
  – not (often) comparable
  – not (necessarily) optimal
The Need

We should want to promote four things:
The Need

We should want to promote four things

• More accessible guidance
The Need

We should want to promote four things
• More accessible guidance
Methods that are
• Transparent
The Need

We should want to promote four things
• More accessible guidance
Methods that are
• Transparent
• Replicable
The Need

We should want to promote four things

• More accessible guidance

Methods that are

• Transparent
• Replicable
• Comparable

Achievable through a standard of practice
Caveats

• A “standard” does not mean closing off alternatives
• A “standard” is only a starting point or frame of reference, not the last word
• There is more than one way to compute legitimate weights
This Talk

Theory of weighting
Practice
Problems

➤ Improvements made
Improvements to make
This Talk

Theory of weighting
Practice
Problems

➤ Improvements made
  – ANES committee
  – ANES procedure
  – “anesrake” tool

Improvements to make
ANES Committee

- Doug Rivers, *Stanford University* (chair)
- Martin Frankel, *Baruch College, CUNY*
- Colm O’Muircheartaigh, *Univ. of Chicago*
- Charles Franklin, *University of Wisconsin*
- Andrew Gelman, *Columbia University*
ANES Procedure: Aim

• Describe a general approach to weighting that everyone can use.

• Take the guess work and art out of weighting.
ANES Procedure: selection probabilities

- Standard steps:
  - Unequal probability of household selection
    - Area and oversampling
  - Unequal probability of respondent selection within households of different sizes
  - Unequal response rates
ANES Procedure: benchmark comparison & raking

• Raking to population benchmarks
• Step 1: conduct benchmark comparison
  – Comparably measured variables such as
    • Age
    • Sex
    • Race/ethnicity
    • Educational attainment
    • Turnout and vote choice
Interpret Benchmark Comparison
(hypothetical data)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Survey</th>
<th>Benchmark</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>57</td>
<td>52</td>
<td>5</td>
</tr>
<tr>
<td>Black</td>
<td>9</td>
<td>13</td>
<td>-4</td>
</tr>
<tr>
<td>College graduate</td>
<td>32</td>
<td>26</td>
<td>6</td>
</tr>
<tr>
<td>Married</td>
<td>54</td>
<td>53</td>
<td>1</td>
</tr>
</tbody>
</table>
Rake, Assess, Repeat

- Rake to correct errors exceeding threshold (5pts)
- Assess errors on all relevant factors
- Assess design effects
- Repeat raking as necessary
Other Details

• Cap weights at 5 by default
• Don’t bother limiting small values
• Raking variables should have 6 or fewer categories, with >5% of cases in each
• Include cases with missing values
Advantages of ANES Procedure

- Explicitly spelled out in detail
- Standardized and systematic, especially for comparing benchmarks and identifying raking factors
“ANESRAKE”

- R package developed by Josh Pasek
- Online tool developed by Guarav Sood

http://www.stanford.edu/group/iriss/cgi-bin/anesrake/raking.php

- Practical: automated, fast, free
This Talk

Theory of weighting
Practice
Problems
Improvements made

➢ Improvements to make
This Talk

Theory of weighting
Practice
Problems
Improvements made

➢ Improvements to make
  – Transparency & replication
  – Address disclosure risk
  – Comparing methods
Transparency & Replication

• Prioritizing transparent & replicable methods
• Hallmarks of science
• Prerequisites for
  – Scientific progress (cumulative & disseminated knowledge)
  – Convincing argument
• Factor for funding?
• Raise some concerns about disclosure risk
Disclosure Risk: Concept

• Disclosure risk is the possibility that confidential respondent information could become known to a data intruder.
Disclosure Risk: Obligations

Based on

• Promises to respondents
• Ethical professional obligations
• IRB requirements
• Funding agency mandates
• University requirements
Disclosure Risk: Weighting Issues

• Weights may reveal respondent characteristics such as exact size of Census tract or other sampling area where R lives
Disclosure Risk: Mitigation

• Restricted access to weight components
  – But this limits transparency
Disclosure Risk: Mitigation

• Restricted access to weight components
  – But this limits transparency

• Jittering or rounding weight data
  – But this limits accuracy
Disclosure Risk: Mitigation

• Restricted access to weight components
  – But this limits transparency
• Jittering or rounding weight data
  – But this limits accuracy
• This is an area for further research
Lastly: Comparing Methods

- Small literature on comparing alternate weighting methods (e.g. Kalton & Flores-Cervantes 2003)
- Hope for comparisons of the recommended transparent & replicable methods to alternative transparent & replicable methods
Recap

• Weights are essential & theory is well founded

• In practice implementation is uneven

• We’ve made free tools to produce weights by replicable, comparable, transparent methods

• Researchers
  – need more guidance
  – should be held to a high standard

• Disclosure risk needs to be mitigated

• Methods need to be compared
Thank you

Computation of Survey Weights: Bridging Theory and Practice

Matthew DeBell
American National Election Studies
Stanford University