Handling Missing Data
in
Environmental Surveys

Ruben A. Smith
Virginia M. Lesser

Department of Statistics
Oregon State University
This presentation was supported under STAR Research Assistance Agreement No. CR-829096 awarded by the U.S. Environmental Protection Agency to Oregon State University. It has not been formally reviewed by EPA. The views expressed in this document are solely those of authors and EPA does not endorse any products or commercial services mentioned in this publication.
Outline

• Missing data problem in environmental surveys.

• Methods for handling missing data.

• Illustration of missing data in environmental surveys.

• Future Activities.
Types of Missing Data

- Missing data ≡ Nonresponse
- Unit nonresponse

<table>
<thead>
<tr>
<th></th>
<th>var1</th>
<th>var2</th>
<th>…</th>
<th>varp</th>
</tr>
</thead>
<tbody>
<tr>
<td>unit 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>unit i</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>unit n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Types of Missing Data

- **Item nonresponse**

<table>
<thead>
<tr>
<th></th>
<th>var1</th>
<th>var2</th>
<th>…</th>
<th>varp</th>
</tr>
</thead>
<tbody>
<tr>
<td>unit 1</td>
<td></td>
<td></td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>unit i</td>
<td></td>
<td></td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>unit n</td>
<td></td>
<td></td>
<td></td>
<td>?</td>
</tr>
</tbody>
</table>
Missing Data

- In environmental surveys missing data may occur for different reasons.
- Unit nonresponse:
  - Inaccessibility of the site
  - Access denial
- Item or unit nonresponse
  - Data is lost or damaged
  - Failure of the measuring instruments
  - Addition of new sites to existing monitoring program
Mechanisms for Nonresponse

- Missing Completely at Random (MCAR)
  - Expect respondents and nonrespondents to be similar.

- Missing at Random (MAR)
  - Given covariates a model can be used to account for the nonresponse.

- Not Missing at Random (NMAR)
  - Complicated. The probability of being a nonrespondent depends on the unobserved response. Given covariates a model cannot completely adjust for the nonresponse.
Nonresponse Bias

• Combination:
  • Nonresponse rate, $\lambda$
  • Difference between respondents-nonrespondents

Bias $\approx \lambda(\bar{Y}_{\text{resp}} - \bar{Y}_{\text{nonresp}})$

• Adjustment procedures to reduce nonresponse bias.
Adjustment for
Unit Nonresponse

• **Weighting Methods**

The weight for a respondent: \( \frac{1}{\pi_i \phi_i} \)

\( \pi_i \) Inclusion probability

\( \phi_i \) Probability of response
Weighting Methods for Unit Nonresponse

Weighting class adjustment:

- $\phi_i$ is estimated by dividing the sample into classes using auxiliary variables, which are known for all units in the selected sample.

- Assumption: within each class the units have the same response probability.
Weighting Methods for Unit Nonresponse

Poststratification:

- \( \phi_i \) is estimated by dividing the population into classes (poststratum) using auxiliary variables (known only for respondents), and population counts.

- Assumption: within each class, the population elements have the same response probability.
Other Adjustment Procedures for Unit Nonresponse

- Model-based methods
  - A model is proposed for the complete data, accounting for the nonresponse mechanism.
  - Utility of model-based inferences depends on how closely the assumed model reflects the reality.
Adjustment for Item Nonresponse

• Delete observation
• Imputation methods:
  • Cell mean imputation
  • Hot/cold-deck imputation
  • Regression imputation
  • Multiple imputation
  • Neural networks.
Illustration

• Background:
Lesser, M. Virginia, (2001), “Applying Survey Research methods to account for denied access to research sites on private property”

Illustration - Background

• A stratified random sample was selected for each year.
• Objective: Provide estimates of prairie wetland condition and status by wetland class.
• On-site visits were necessary to obtain information on condition indicators.
### Response Disposition 1995/1996 EMAP North Dakota Prairie Wetlands Studies

<table>
<thead>
<tr>
<th>Result</th>
<th>1995</th>
<th>1996</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Private Landowners</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreed to access</td>
<td>43%</td>
<td>40%</td>
</tr>
<tr>
<td>Refused access</td>
<td>36%</td>
<td>37%</td>
</tr>
<tr>
<td>Undeliverable</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td>Not returned/no contact</td>
<td>16%</td>
<td>14%</td>
</tr>
<tr>
<td><strong>Public Land</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3%</td>
<td>7%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Access Rate for 1995/1996 North Dakota Prairie Wetlands Survey by Wetland Type

Seasonal
Temporary
Permanent
Semi-permanent

Wetland Type

Percent
0 10 20 30 40 50 60

1995 1996
Access Rate for 1995/1996 North Dakota Prairie Wetlands Survey by Region

Couteau Drift Plain Red River Valley

Percent

0 20 40 60

Couteau Drift Plain Red River Valley

Region

1995 1996
Future Activities

- Evaluate classical and model based adjustment procedures to account for missing data.

- Consider some spatial interpolation techniques within the context of single and multiple imputation.

- Data from Oregon Department of Fish & Wildlife (ODFW).
Future Activities

• Develop an user-friendly manual to deal with missing data in environmental surveys.