Handling Missing Data in Environmental Surveys

Ruben A. Smith

Virginia M. Lesser

Department of Statistics
Oregon State University

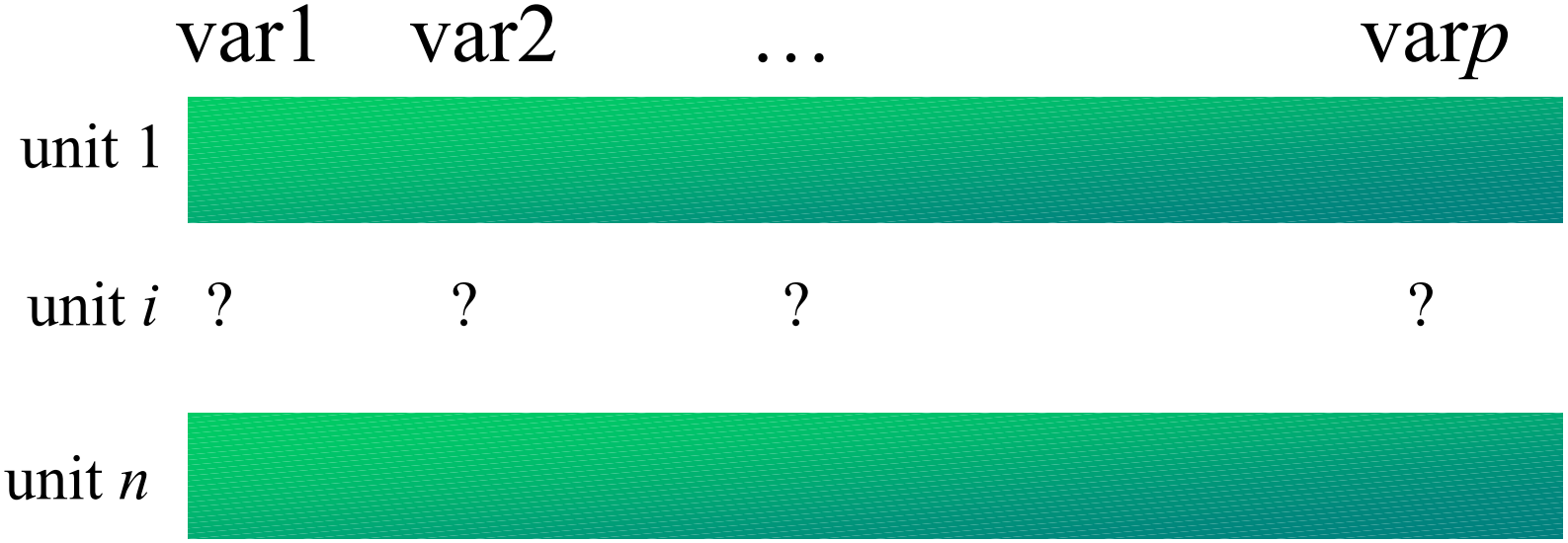
This presentation was supported under STAR Research Assistance Agreement No. CR-829096 awarded by the U.S. Environmental Protection Agency to Oregon State University. It has not been formally reviewed by EPA. The views expressed in this document are solely those of authors and EPA does not endorse any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

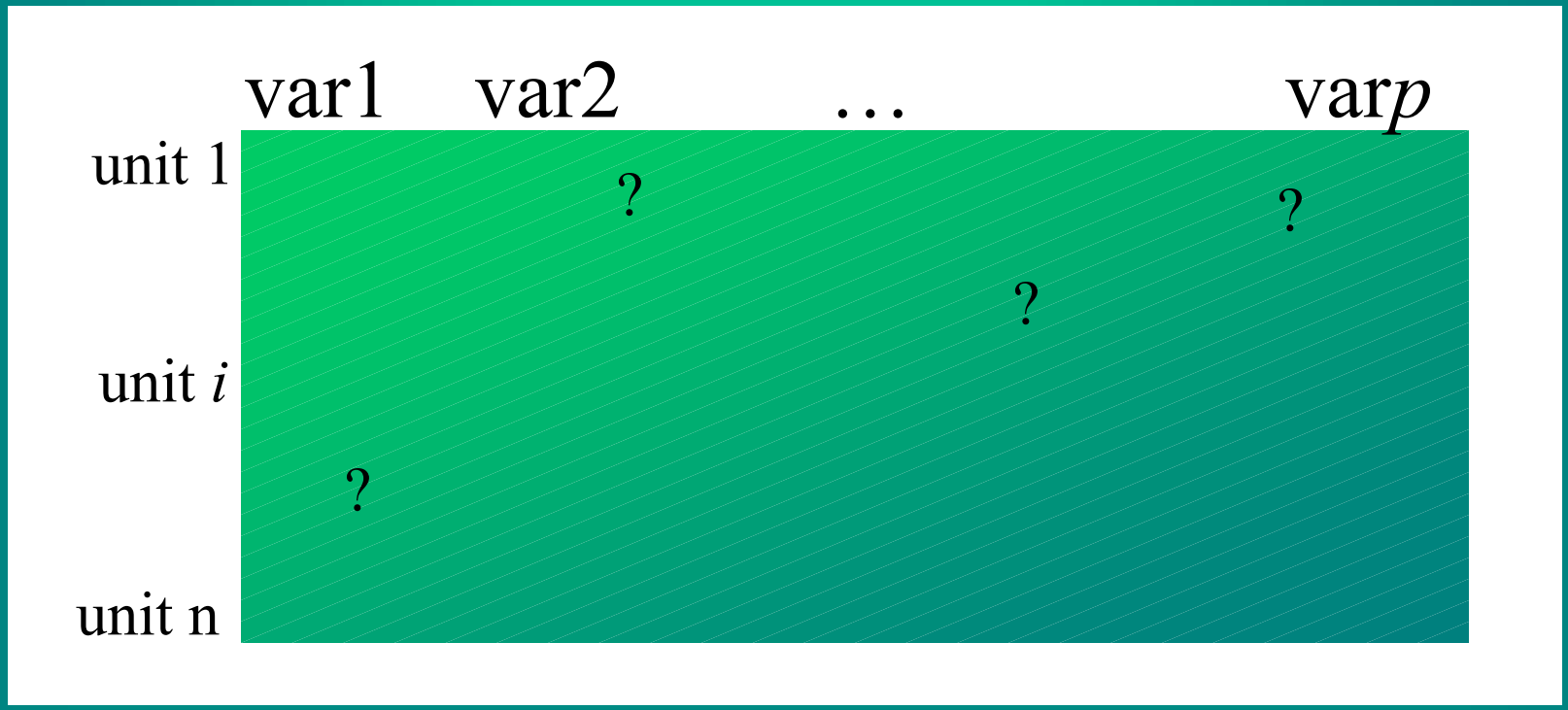
Types of Missing Data

- Missing data ≡ Nonresponse
- Unit nonresponse



Types of Missing Data

- Item nonresponse



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, λ
  - Difference between respondents-nonrespondents

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

- Adjustment procedures to reduce nonresponse bias.

Adjustment for Unit Nonresponse

- Weighting Methods

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

π_i Inclusion probability

ϕ_i Probability of response

Weighting Methods for Unit Nonresponse

Weighting class adjustment:

- ϕ_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:

- ϕ_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”

  - 1995-1996 EMAP North Dakota prairie wetlands survey.

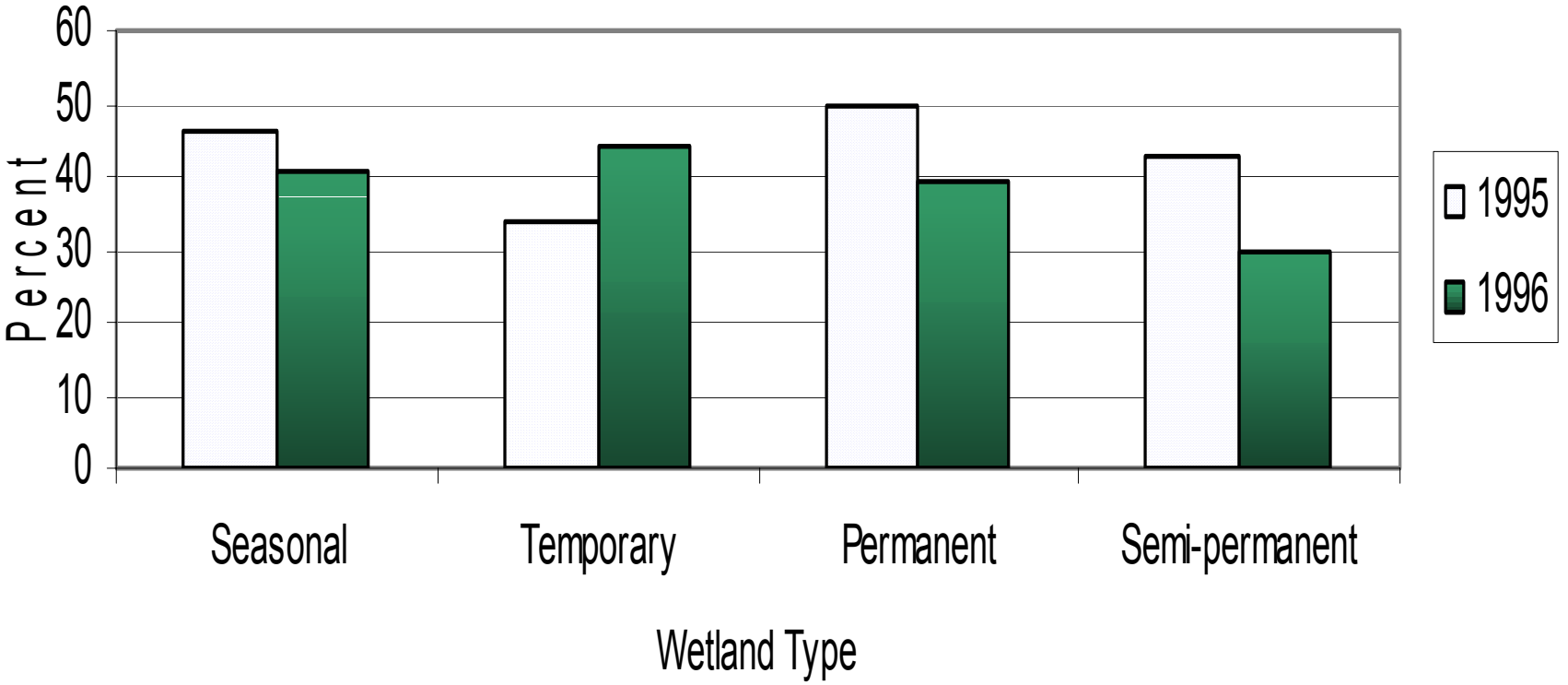
Illustration - Background

- 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.

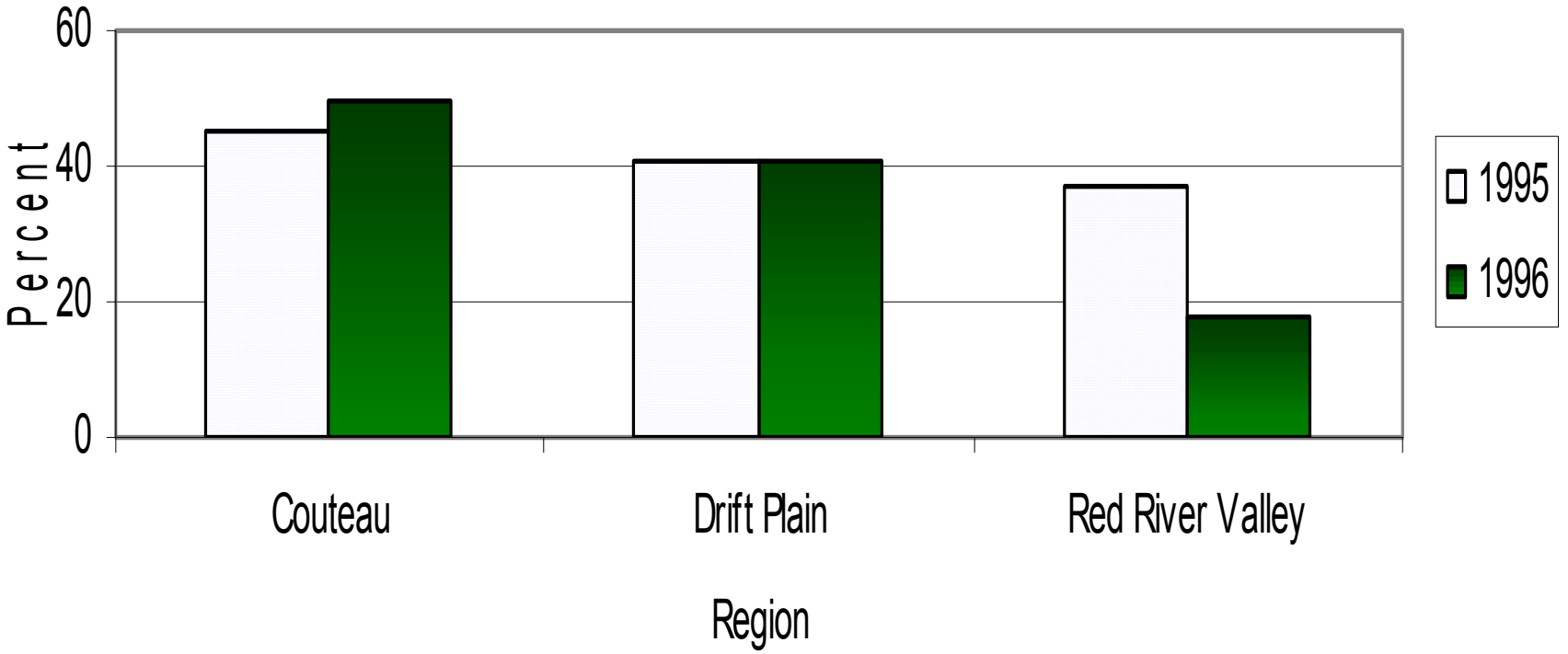
Response Disposition 1995/1996 EMAP North Dakota Prairie Wetlands Studies

Result	1995	1996
Private Landowners		
Agreed to access	43%	40%
Refused access	36%	37%
Undeliverable	2%	2%
Not returned/no contact	16%	14%
Public Land	3%	7%
Total	100%	100%

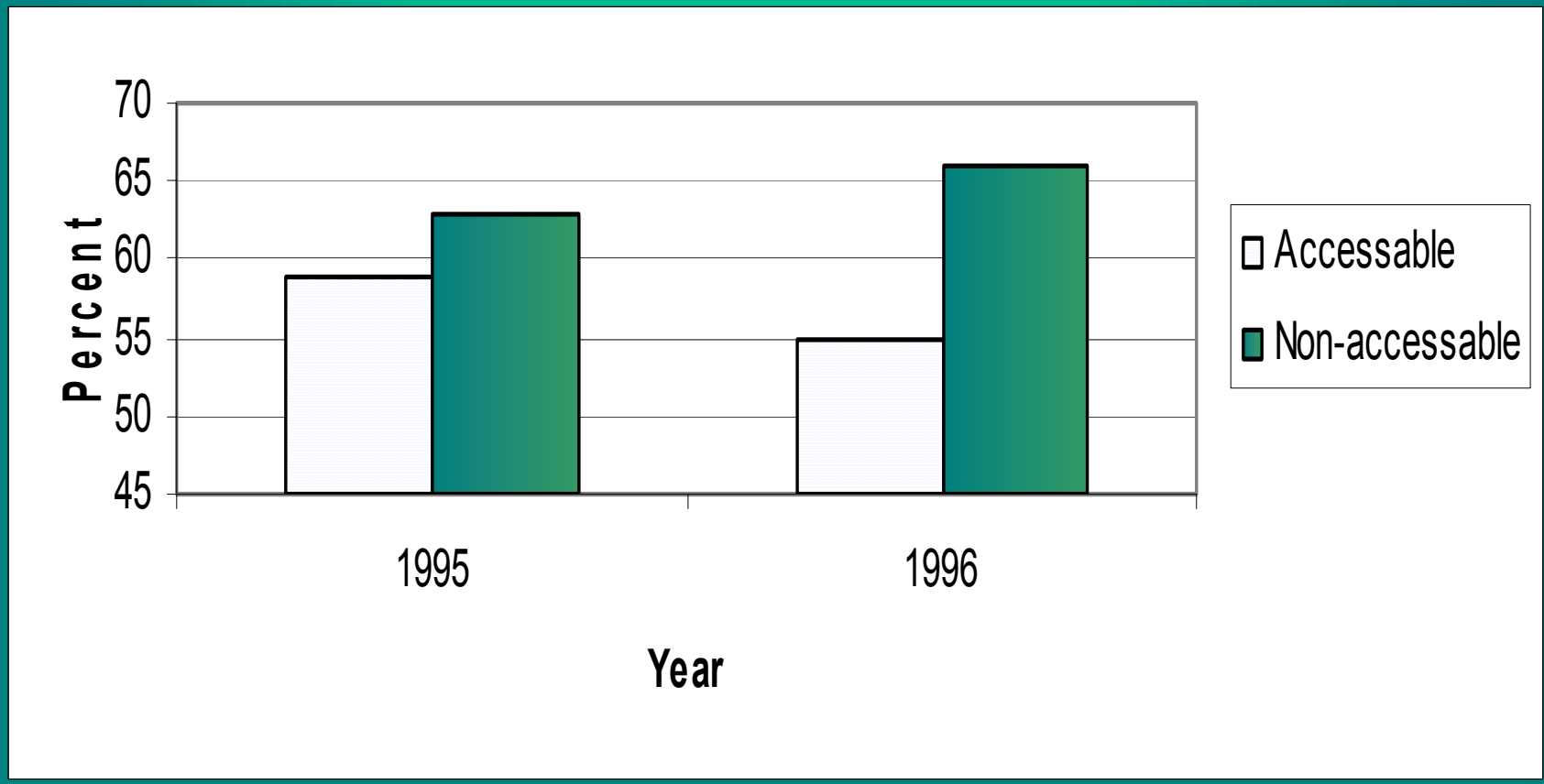
Access Rate for 1995/1996 North Dakota Prairie Wetlands Survey by Wetland Type



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



Percent of Cropland Bordering Wetlands for 1995/1996 North Dakota Prairie Wetlands Survey



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.