

---

# Survey Research: Errors & Limits

---


URBPL 5/6010: Urban Research  
University of Utah  
Pam Perlich

# Research Design

Identify purpose, audience, budget, & objectives



Define analytical construct / model to:

- Address your research questions
  - Structure your approach
- 

Develop data or identify existing data that is required for the project

**Build error checking into every step**

---

# Constraints

- Budget
  - Professional Standards
  - Ethics
  - Institutional context / barriers
  - Legal standards
  - Confidentiality
  - Technical limitations
  - Transparency
-

---

# Types of Data

- Cross-sectional
  - Point in time
  - The following discussion mainly applied to this type of data
  - Survey or census
- Panel data
  - Follows individuals across time
  - Errors are different – attrition, for example

---

# Data Collection Errors

- All data collected from censuses and surveys will have error
- “Population” is a moving target
- Researchers attempt to
  - Measure or estimate error
  - Minimize error
- Errors cannot be avoided
  - Attempt to identify and minimize

---

# Types of Survey Errors

- **Nonsampling** / response errors
  - ❑ Coverage errors
  - ❑ Content / reporting errors
  - ❑ Nonreporting
- **Sampling** errors
  - ❑ Sampling or probabilistic error
  - ❑ Other errors due to sampling

---

# Nonsampling Errors:

## Gross Coverage Errors

- Erroneous exclusion – individuals were omitted who should have been included → undercount
- Erroneous inclusion – individuals were included who should not have been
  - Duplications – Multiple responses from one individual
  - Curbstoning – Fictitious responses from nonexistent people
  - Geocoding errors – Incorrect address
  - Out-of-scope cases – e.g. including tourists in survey targeting permanent residents

---

# Causes of Coverage Errors

- Errors of the professional staff
  - Poor questionnaire, failure to follow-up on nonrespondents, address list error
- Actions / inactions of respondents
  - Avoidance, carelessness, misrepresentation
- Problems of the physical environment
  - Irregular housing/living arrangements, dangerous or isolated area
- Misc. others – machine errors, accidents
  
- Siegel: This list of “The causes of omission . . .” (are) “not claimed to be mutually exclusive or comprehensive.”

---

# Reduction of Coverage Error

- Substitution – if multiple unsuccessful attempts have been made to get the interview
  - If age, sex, race, relationship to householder are known, substitute data is often used
  - This becomes an estimation problem
- Coverage improvement programs
  - Target to augment the original count
  - 1990 Census: Shelter and Night Street Enumeration, etc.

---

# Undercount Especially Affects

- Minorities
- Poor
- Homeless
- Undocumented
- Isolated

---

# Undercount Controversy

- Population counts
    - Congressional apportionment and redistricting
    - Funding formulas
  - Mayors of large cities – 1970 and onward
  - Major lawsuits to adjust counts
  - 1990 census produce two sets of data
    - Official counts
    - Adjusted counts
  - Harvey Choldin (1994) *Looking for the Last Percent: The Controversy Over Census Undercounts.*
-

# Measurement of Coverage Rates

## Decennial Census Population Net Undercount Rates From Demographic Analysis: 1940 to 2000

|                                | Year |      |      |      |      |      |      |
|--------------------------------|------|------|------|------|------|------|------|
|                                | 1940 | 1950 | 1960 | 1970 | 1980 | 1990 | 2000 |
| Total .....                    | 5.4  | 4.1  | 3.1  | 2.7  | 1.2  | 1.8  | 0.1  |
| Black .....                    | 8.4  | 7.5  | 6.6  | 6.5  | 4.5  | 5.7  | 2.8  |
| Difference: Black - Total. ... | 3.0  | 3.4  | 3.5  | 3.2  | 3.3  | 3.9  | 2.7  |

Jon R. Clark and Darlene A. Moul. (2004) *Census 2000 Testing, Experimentation, and Evaluation Program Topic Report No. 10, TR-10, Coverage Improvement in Census 2000 Enumeration*, U.S. Bureau of the Census.

“This is the first time we have measured an overall population overcount in the history of the U.S. Census.”

**Percent of Net Undercount for Major Groups in Census 2000 Using A.C.E. Revision II and Demographic Analysis**

|  | A.C.E. w/o correlation bias adjustment | A.C.E. with correlation bias adjustment | Demographic analysis |
|--|--|---|----------------------|
| Total .....                            | -1.09                                  | -0.48                                   | 0.12                 |
| <b>Race: Hispanic Origin Domain</b>    |  |   |                      |
| Non Black .....                        | -1.18                                  | -0.80                                   | -0.29                |
| Black .....                            | -0.49                                  | 1.72                                    | 2.78                 |
| Hispanic .....                         | -                                      | 0.71                                    | -                    |
| Non Hispanic - Asian .....             | -                                      | -0.75                                   | -                    |
| American Indians on Reservations ..... | -                                      | -0.88                                   | -                    |
| <b>Tenure Domain</b>                   |  |   |                      |
| Owner .....                            | -                                      | -1.25                                   | -                    |
| Renter .....                           | -                                      | 1.14                                    | -                    |
| <b>Gender</b>                          |  |   |                      |
| Male .....                             | -0.87                                  | 0.37                                    | 0.86                 |
| Female .....                           | -1.30                                  | -1.30                                   | -0.60                |

A negative net undercount denotes a net overcount.  
 A dash (-) indicates the data were not available.

**Jon R. Clark and Darlene A. Moul. (2004) *Census 2000 Testing, Experimentation, and Evaluation Program Topic Report No. 10, TR-10, Coverage Improvement in Census 2000 Enumeration*, U.S. Bureau of the Census.**

---

# Content / Reporting Errors

- Respondent error
  - Misrepresentation
  - Omission
  - Unacceptable answers
- Imputation Errors: Hot deck procedures
  - Missing characteristic is replaced with information for the last similar person surveyed.
- Survey taker error
- Data processing errors
- **LACK OF A PERFECT STANDARD**

---

# Imputation of 100% Population Items: Census 2000

| Resident Population         | United States | Share  |
|-----------------------------|---------------|--------|
| Total:                      | 277,980,752   | 100.0% |
| No items allocated          | 250,880,895   | 90.3%  |
| One or more items allocated | 27,099,857    | 9.7%   |

9.7 percent of population records had one or more characteristic estimated with imputation

---

# Errors Due to Sampling

- Sampling error is the probabilistic error associated with estimates based on samples of the population
  - Confidence intervals
  - Control of sampling errors
    - Large enough samples
    - Over-sampling
  - Report the confidence intervals
- Biased sample selection (non-random sample)

---

# Sampling Error

- Point estimate is  $100,000 \pm 2.25\%$  at 95% CI
  - 95% CI = 2 standard errors
  - There is a 95% probability that the “true” value lies between 97,750 and 102,250
- Larger sample size → reduces interval

Table 10.  
**Percentage of People in Poverty by State Using 2- and 3-Year Averages: 2002 to 2004**

(People as of March of the following year)

| State                           | 3-year average<br>2002–2004 |   | 2-year average |   |             |   | Change in percentage<br>points (2003–2004<br>average less<br>2002–2003 average) <sup>1</sup> |   |
|---------------------------------|-----------------------------|---|----------------|---|-------------|---|--|---|
|                                 |                             |   | 2002–2003      |   | 2003–2004   |   |  |   |
|                                 | Percentage                  | 90-percent<br>confidence<br>interval <sup>2</sup> (±) | Percentage     | 90-percent<br>confidence<br>interval <sup>2</sup> (±) | Percentage  | 90-percent<br>confidence<br>interval <sup>2</sup> (±) | Percentage   | 90-percent<br>confidence<br>interval <sup>2</sup> (±) |
| <b>United States</b> .....      | <b>12.4</b>                 | <b>0.2</b>  | <b>12.3</b>    | <b>0.2</b>  | <b>12.6</b> | <b>0.2</b>  | <b>*0.3</b>  | <b>0.2</b>  |
| Alabama .....                   | 15.5                        | 1.5   | 14.7           | 1.6   | 16.0        | 1.7   | 1.2  | 1.4   |
| Alaska .....                    | 9.2                         | 1.2   | 9.2            | 1.3   | 9.4         | 1.4   | 0.2  | 1.2   |
| Arizona .....                   | 13.8                        | 1.4   | 13.5           | 1.7   | 13.9        | 1.6   | 0.4  | 1.4   |
| Arkansas .....                  | 17.6                        | 1.6   | 18.8           | 1.9   | 16.4        | 1.8   | *-2.3  | 1.6   |
| California .....                | 13.2                        | 0.6   | 13.1           | 0.7   | 13.2        | 0.7   | 0.1  | 0.6   |
| Colorado .....                  | 9.8                         | 1.1   | 9.7            | 1.2   | 9.9         | 1.4   | 0.1  | 1.2   |
| Connecticut .....               | 8.8                         | 1.1   | 8.2            | 1.1   | 9.1         | 1.3   | 0.9  | 1.1   |
| Delaware .....                  | 8.5                         | 1.2   | 8.2            | 1.3   | 8.2         | 1.4   | -  | 1.2   |
| District of Columbia .....      | 16.8                        | 1.7   | 16.9           | 1.9   | 16.7        | 2.1   | -0.2   | 1.7   |
| Florida .....                   | 12.3                        | 0.7   | 12.6           | 0.9   | 12.2        | 0.9   | -0.5   | 0.7   |
| <b>Section of table omitted</b> |                             |   |                |   |             |   |  |   |
| South Carolina .....            | 14.0                        | 1.4   | 13.5           | 1.5   | 13.8        | 1.7   | 0.3  | 1.4   |
| South Dakota .....              | 12.5                        | 1.3   | 12.1           | 1.4   | 13.0        | 1.5   | 1.0  | 1.3   |
| Tennessee .....                 | 14.9                        | 1.4   | 14.4           | 1.7   | 15.0        | 1.6   | 0.5  | 1.4   |
| Texas .....                     | 16.4                        | 0.8   | 16.3           | 1.0   | 16.7        | 1.0   | 0.4  | 0.8   |
| Utah .....                      | 9.6                         | 1.2   | 9.5            | 1.3   | 9.5         | 1.4   | -  | 1.2   |
| Vermont .....                   | 8.8                         | 1.1   | 9.2            | 1.2   | 8.2         | 1.3   | -1.0   | 1.2   |
| Virginia .....                  | 9.8                         | 1.1   | 10.0           | 1.3   | 9.7         | 1.2   | -0.3   | 1.0   |
| Washington .....                | 11.7                        | 1.2   | 11.8           | 1.4   | 12.0        | 1.4   | 0.2  | 1.2   |
| West Virginia .....             | 16.1                        | 1.4   | 17.1           | 1.6   | 15.8        | 1.6   | -1.3   | 1.4   |
| Wisconsin .....                 | 10.2                        | 1.1   | 9.2            | 1.1   | 11.0        | 1.3   | *1.9   | 1.1   |
| Wyoming .....                   | 9.6                         | 1.2   | 9.4            | 1.3   | 9.9         | 1.5   | 0.5  | 1.3   |

Source: U.S. Census Bureau, Current Population Survey, 2003 to 2005 Annual Social and Economic Supplements.

### Three-Year Average Poverty Rate by State: 2002 to 2004

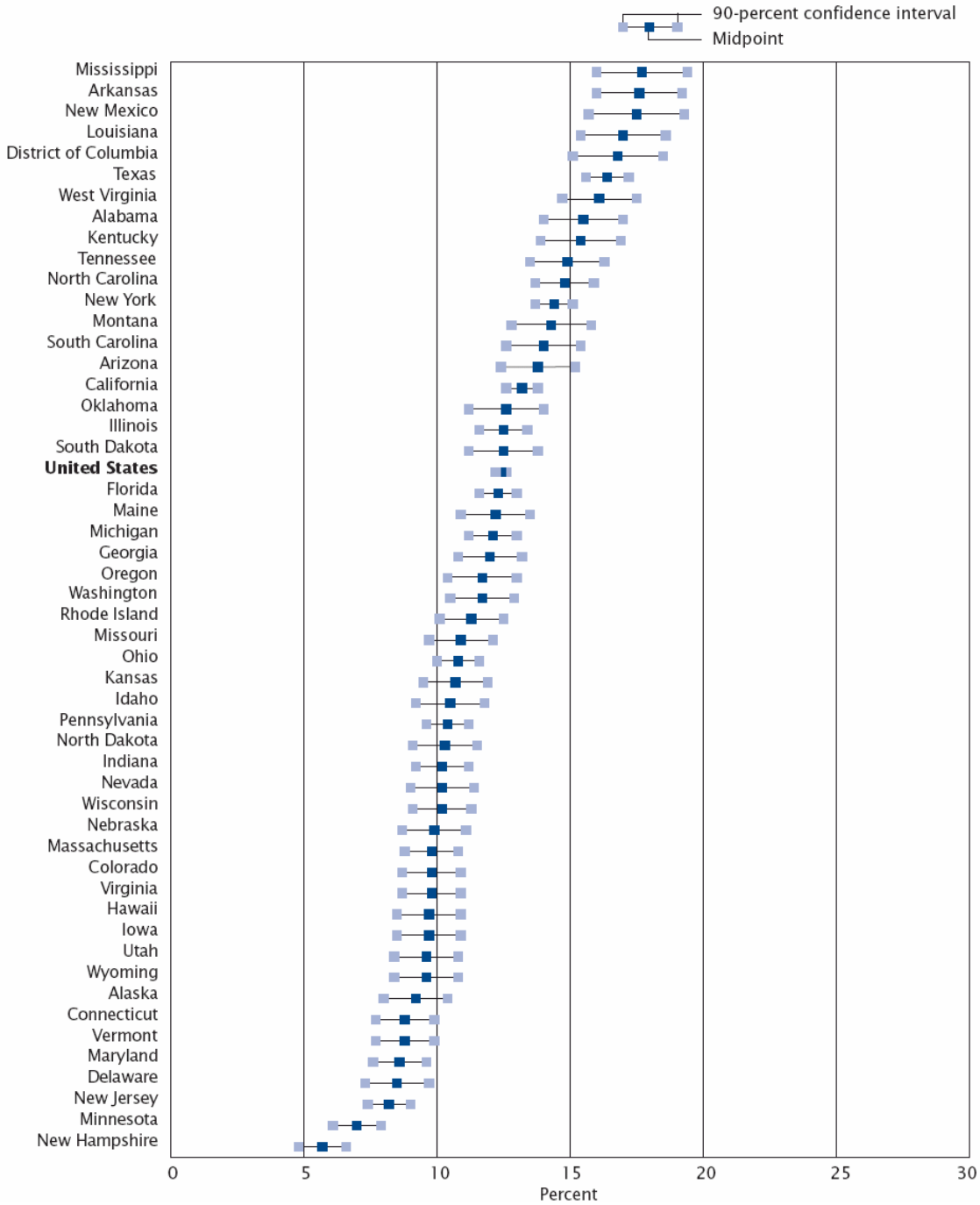
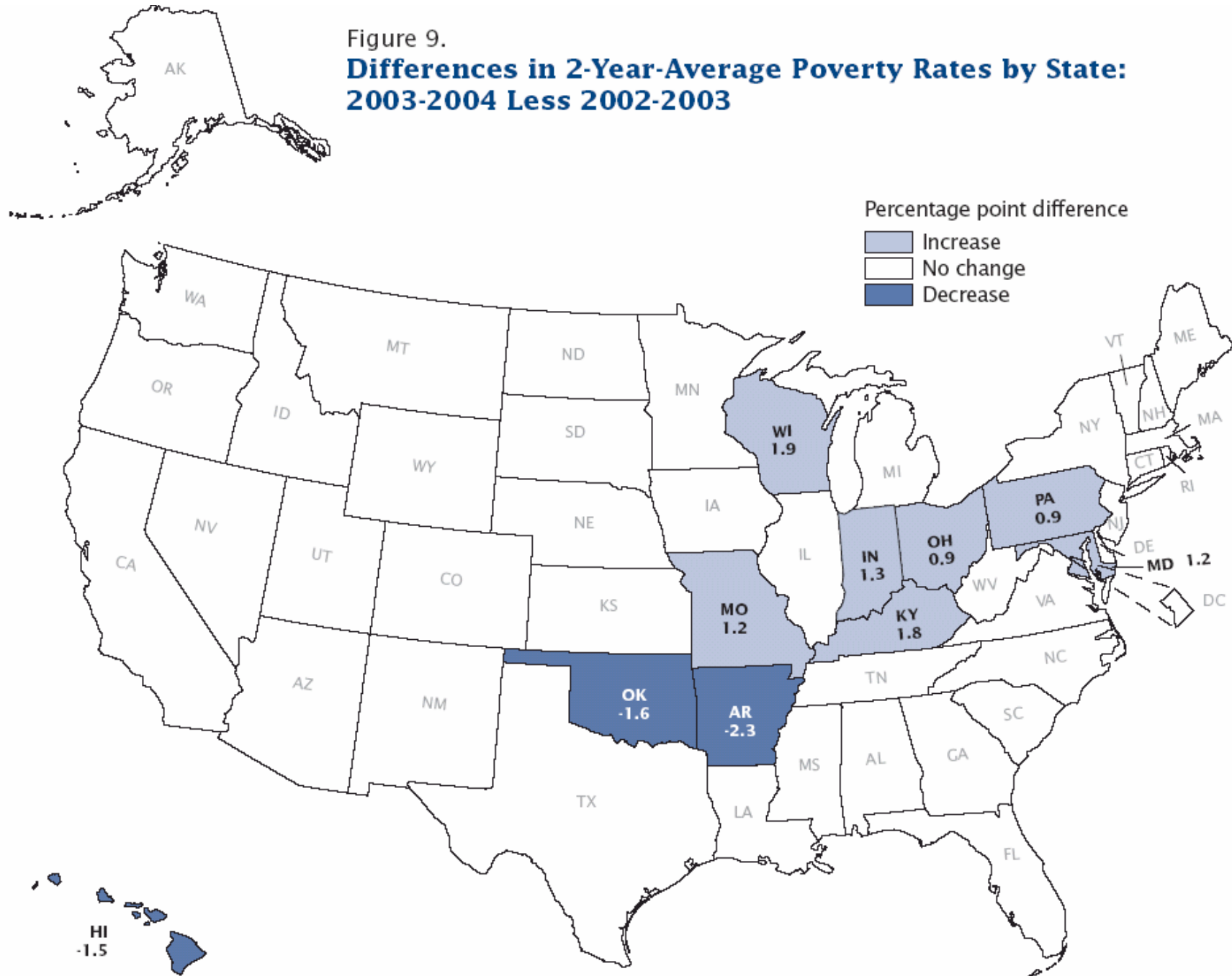


Figure 9.  
**Differences in 2-Year-Average Poverty Rates by State:  
2003-2004 Less 2002-2003**



Source: U.S. Census Bureau, Current Population Survey, 2003 to 2005 Annual Social and Economic Supplements.

---

# Confidentiality Issues

- Privacy is protected by law
- Agencies must report micro data in a way that the identity of individuals cannot be disclosed
- Maximize data availability
- PUMS – micro data
- Disclosure limitation procedures
- Patriot Act ??????

---

# Conclusions

- Errors are unavoidable – models, surveys, data
  - Develop procedures to minimize these
  - Document your methods and work so that others can replicate it
  - Be honest about the limitations of your data, surveys, and models
  - Limited resources (including time) constrain your research strategy => Give adequate priority (resources) to error reduction in your research agenda / program.
-