Accounting for Missing Data in the Census Coverage Measurement Survey

Federal Committee on Statistical Methodology Conference

November, 2013

Vincent Thomas Mule, Jr. and Scott Konicki

United States Census Bureau

Census Coverage Measurement (CCM) Survey

- Survey-based evaluation of the 2010 Census
- Net coverage of household population and housing units
  - Dual system estimation
- Components of Census Coverage
  - Correct and erroneous enumerations
  - Omissions

Accounting for Missing Data in CCM Estimation

- Focus on household population
- Highlight procedures for dual system estimation

Dual System Estimation

(Shaded in green: Can be measured)		Enumerated in PES ? (P Sample)		
		YES	NO	Total
Enumerated in Census ? (E sample)	YES	N_{11}	N_{12}	N_{1+}
	NO	N_{21}	N_{22}	N_{2+}
	Total	N_{+1}	N_{+2}	N

Dual System Estimation (cont.)

		Enumerated in PES ? (P Sample)		
		YES	NO	Total
Enumerated in Census ? (E sample)	YES	#2 Matches		#3 Correct Enumerations
	NO			
	Total	#1 P sample		Total Population

1. Population (P) Sample

Independent population interview
August to October 2010

Sources of Missing Data:

- Noninterviews
- Demographic characteristics
- Is the rostered person a resident on Interview Day? Inclusion status

Noninterviews

3.1 percent noninterviews

Noninterview adjustment procedure

- Cell-based approach
- Check: Noninterview < 2 x Interview
- Spreading to larger aggregates instead of collapsing

Missing Characteristics of Population Sample

Percentages of missing:

- 2.8% race
- 2.6% Hispanic origin
- 6.1% age
- 1.3% sex
- 2.5% relationship

Characteristic Imputation for Missing Characteristics

Census Edit and Imputation System

- Formed Census-day and Interview-day households
- Imputation of demographic characteristics and tenure

Unresolved Population Sample Inclusion Status

Is the person a resident of housing unit on Interview Day?

- 2.9% unresolved

Predictions for Unresolved Population Sample Inclusion Status

- 2010: Logistic regression modeling
- 2000: Weighted cell means
- 1990: Logistic regression modeling

Models for Unresolved Population Sample Inclusion Status

	Unresolved Group	Model
1	Reported Name and 2 Characteristics Census Day, Seasonal or Other address during Interview	All resolved cases Followup reasons, demographics, type of address and others Address flags used in model
2	Reported Name and 2 Characteristics No address collected	All resolved cases Followup reasons, demographics, type of address and others Address flags not used in model
3	Did not report name and 2 Characteristics	All resolved cases Where rostered in interview? Residence status Housing unit matching

2. Population Sample Match Status

Does the Population person match to a correct enumeration in the census?

Unresolved Match status in 2010

Mover Status	Number of Cases	Unresolved Match Status
Nonmover	317,000	0.0%*
Outmover	1,200	0.0%*
Inmover	26,400	6.7%

Predicted probabilities for Unresolved Match Status

$$\Pr(Match) = Pr(Inmover) \times \Pr(Match|Inmover) + (1 - \Pr(Inmover)) \times \Pr(Match|Not\ Inmover)$$

Probability	Approach
Pr(Inmover)	Weighted average
Pr(Match|Inmover)	Logistic regression of resolved inmovers
Pr(Match|Not Inmover)	Logist regression of resolved nonmovers and outmovers

3. Census Correct Enumeration Status

- Which census enumerations were correct enumerations?

4.8 percent unresolved enumeration status

Models for Unresolved Enumeration Status

	Unresolved Group	Model
1	Census Day, Seasonal or Other address during Interview or Duplicate found in census	All resolved cases Followup reason, demographics, proxy interview and others Address flags and duplication flag used in model
2	No address collected or duplicate found	All resolved cases Followup reason,demographics, proxy interview and others Address flags and duplication flag not used in model

Conclusion

Methods shown today to compensate for:

- Unit-level missing data
- Person-level missing data

Major changes since 2000

- Logistic regression models
- Same characteristic system as Census

Thank you

Thanks to Rachel Bray, Andrew Keller, Mark Seiss, Robert Sands and Colt Viehdorfer

Vincent.t.mule.jr@census.gov