

What is the government doing to encourage people to fill out the Census?

Sampling in the Census

Stephen E. Fienberg

Statistics 36-303

March 31, 2010

History of Sampling in the Census

- 1940 – 1 in 10 long form introduced
- 1947 – Price reported on undercount based on comparison with draft records for young men (larger for Blacks than for whites)
- 1950 – first post-enumeration survey to check on accuracy of census counts
- 1955 – demographic analysis introduced
- 1960 – repeat of post-enumeration survey

History of Sampling in the Census-II

- **1960s – Civil Rights Act**
- 1970 – use of sampling to correct a pair of problems with raw counts
- 1970 – debate over undercount intensifies
- **1970s – use of census data for funds allocation**
- **1970s – growth of Hispanic immigration and the illegal alien problem**
- 1980 – plans for use of a pair of CPS surveys in place of separate PES to check on accuracy

History of Sampling in the Census-III

- 1982 – decision in NYC lawsuit regarding adjustment
- 1987 – adjustment debate begins anew
- 1990, 1990s and 2000
 - Focus of *Who Counts?*
- Separate issue in the 1990s on sampling for non-response follow-up

Methods for Census Adjustment

- Demographic Analysis
- Dual Systems Estimation
- Administrative Records
- ????

Demographic Analysis

- For socio-demographic sub-populations (e.g., sex, race)
- # people in 2010
= # people in 2000 + births – deaths + immigration
- emigration
- Refinements by working with age cohorts

Census Adjustment Controversy and 1990 Post Enumeration Survey

Stephen E. Fienberg

Statistics 36-303

March 31, 2010

What Do Following Populations Have in Common?

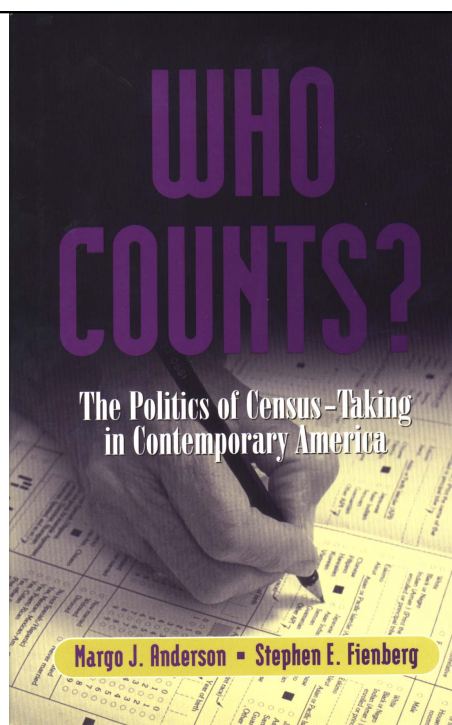
- fish*
- penguins
- homeless
- prostitutes in Glasgow
- Italians with diabetes*
- killings in Kosovo
- people in the U.S.*
- people with HIV virus
- adolescent injuries in Pittsburgh, PA
- WWW

9

Margo J. Anderson and
Stephen E. Fienberg (1999).

Who Counts? The Politics of Census-Taking in Contemporary America.

Russell Sage Foundation.



Example 1: Prevalence of Diabetes

Bruno et al. (1994) used 4 sources for ascertainment of diabetes in Casale Monferrat, Northern Italy

s1: diabetes clinic and/or family physicians

s2: patients discharged with diagnosis from hospitals

s3: insulin or oral hypoglycaemic prescriptions

s4: requests for reimbursement for insulin and reagent strips

11

Diabetes Example (cont.)

	S1	Yes	Yes	No	No
	S2	Yes	No	Yes	No
S3	s4				
Yes	Yes	58	46	14	8
Yes	No	157	650	20	182
No	Yes	18	12	7	10
No	No	104	709	74	-

$n = 2069$

12

Example 2: Fish in a Lake

- 200 fish caught 1st time.
- 150 fish caught 2nd time.
- Of 150 fish in 2nd sample, 125 were among 200 counted in 1st sample.

13

Example 2: Fish in a Lake

- 200 fish caught 1st time.
- 150 fish caught 2nd time.
- Of 150 fish in 2nd sample, 125 were among 200 counted in 1st sample.
- Total number of fish caught:
 $= 200 + (150 - 125) = 225$
- But how many fish have gone undetected?

14

Ex. 2: Fish in a Lake (cont.)

- Proportion of fish in 2nd sample also in 1st:
 $= 125/150 = 5/6$
- Generalize from sample to population:
 $(5/6) \hat{N} = 200$
 $\hat{N} = (6/5) 200 = 240$
- This is method of capture-recapture due to Petersen, Lincoln, Schnabel, etc.

15

Capture Recapture Model

		Sample 2		
		In	Out	Total
Sample 1	In	x_{11}	x_{12}	x_{1+}
	Out	x_{21}	$x_{22}??$	x_{2+}
Total		x_{+1}	x_{+2}	$x_{++}=N??$

$$\hat{N} = x_{1+} x_{+1} / x_{11}$$

16

Formal Details (cont.)

Alternatively, we can set

$$x_{11} x_{22} / x_{12} x_{21} = 1$$

Estimate unobserved x_{22} by

$$\hat{x}_{22} = x_{12} x_{21} / x_{11}$$

so that

$$\hat{N} = x_{11} + x_{12} + x_{21} + x_{12} x_{21} / x_{11} = x_{1+} x_{+1} / x_{11}$$

17

Example 1: Diabetes Looking at Pairs of Lists

Pair	\hat{N}
s1, s2	2,351
s1, s3	2,185
s1, s4	2,262
s2, s3	2,057
s2, s4	803
s3, s4	1,555

- Estimated s.e.'s are on the order of 100.
- Only 3 of 6 estimates exceed $n = 2069$.

18

Diabetes Example: What is Going Wrong?

- Independence of lists in the pairs!

19

Two-Sample Assumptions

- Random samples
- Closed population
- Perfect matching (no tag loss)
- Independence
- Homogeneity

- How do we check on assumptions?
- The problem of the “wiley trout.”

20

Applying the Ideas in the Census

- Dual systems estimation
 - “Sample 1”= census
 - “sample 2”=post-enumeration survey
- Two kinds of errors:
 - Erroneous enumerations
 - Omissions
- If we could first “fix up” the EE problem we could use capture recapture to estimate omissions

21

Post-Enumeration Surveys

- Take a sample of blocks nationwide
- Redo the census for the everyone in the blocks
- Correct for the passage of time
- Match households and people within blocks
- Correct for EEs
- Do DSE
- Easier said than done!

22

Red Fish, Blue Fish

23

Census Adjustment Controversy

- 1980 PEP program used by New York city in law suit for adjustment
 - Used DSE and “smoothing” to carry estimates down to state and lower levels
 - Proposal was controversial
 - NYC lost!
- 1980s planning for 1990
 - The NAS-NRC panel
 - Planned for adjustment using a PES consisting of a sample of census blocks nationwide

24

The Test of Adjustment Related Operations 1986

- TARP in Mississippi and LA
 - NAS-NRC recommendation
 - Internal Bureau evaluation
- October 30, 1987 cancellation of PES for 1990
 - Congressional hearings
- New law suit brought by NYC, LA, etc. versus Department of Commerce
 - July 17, 1989 stipulation agreement
 - 300,000 HH PES was reinstated but at ½ size
 - See pp. 98-99 of *Who Counts?* for step by step procedure

25

1990 Census Adjustment Controversy