The Lake County Challenge

Robert E. Fay
U.S. Census Bureau

General guidance on analyzing ACS

NRC (2007) (Citro and Kalton, eds.)

- Report examined approaches to analysis

- 10 guidelines in the Executive Summary

1st guideline: *Always examine margins of error before drawing conclusions from a set of estimates.*

General guidance

Beaghen and Weidman (2007)

- General overlap but some shift of emphasis

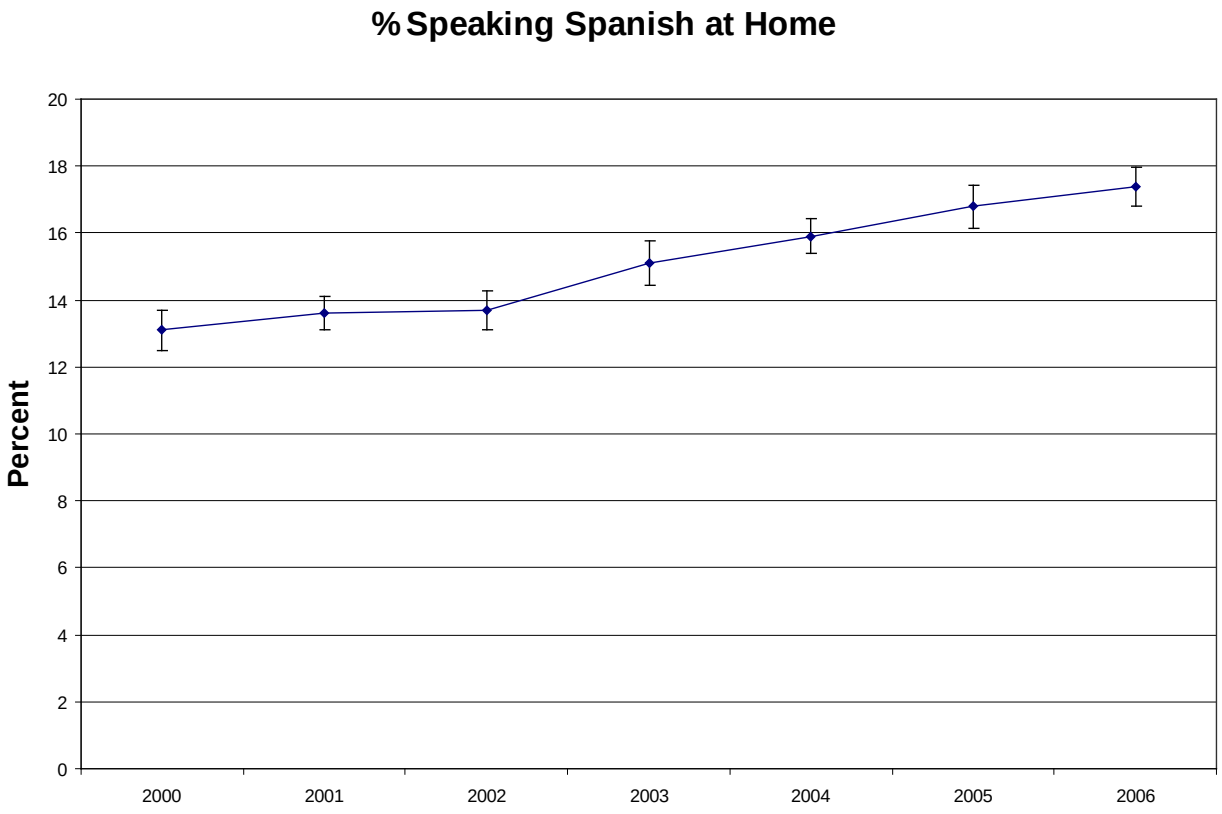
Example from Lake County, Illinois:

- % speak Spanish at home

- Universe: age 5+, including group quarters

- Asked in 1990, 2000 censuses, ACS

Lake County


% Speaking Spanish at Home
Percent
20
18
16
14
12
10
8
6
4
2
0
2000
2001
2002
2003
2004
2005
2006

The Lake County Challenge

Challenge in handout:

- Lake County: 644,599 in Census 2000
- Can we identify subcounty trends?
- Best professional effort standard

Draft paper

Analyses for

- 18 Townships
- 5 PUMAs

Doesn’t take up

- Places – too complicated geographically
- Tracts – too complicated statistically

Ambition: second paper on tracts

Benjamini and Hockberg (1995) JRSSB

"Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing"

- apply false discovery rate calculations to tract-level analysis

Townships

Newport
Lake Villa
Warren
Avon
Waukegan
Lake
Wauconda
Fremont
Libertyville
Shields
Cuba
Ela
Vernon
Moraine

Townships

Start with 1990 – 2000 trends

- Increases in all 18 townships (a few n.s.)
- ~ 1/2 in Waukegan
- ~ 2/3 in Waukegan, Avon, and Zion

  (26% of 2000 county population)

Townships

Propose (Table 1 from 1990/2000 data):

- Group 4: Waukegan ~21% growth
- Group 3: Avon+Zion ~9% growth
- Group 2: 7 townships with 3-6% growth
- Group 1: 8 townships < 3% growth

Townships

1-year ACS data can’t be used

Table 2: 1999-2001 vs. 2003-2005

Non-overlapping 3-year period estimates

Gaps: 11 out of 18 townships

Townships

Table 3: 1999-2003 vs. 2001-2005

Overlapping 5-year period estimates

No gaps, can construct groups 1-4

- Significant increases in all 4 groups
- Group 2 now accounting for larger share of growth

Townships

Table 4: Comparing annualized change

When trend nearly linear, annualized change for 11 townships *quite similar*

- 3-year 1999-2001 to 2003-2005 (4 years)
- 5-year 1999-2003 to 2001-2005 (2 years)

Both for estimates and standard errors

PUMAs

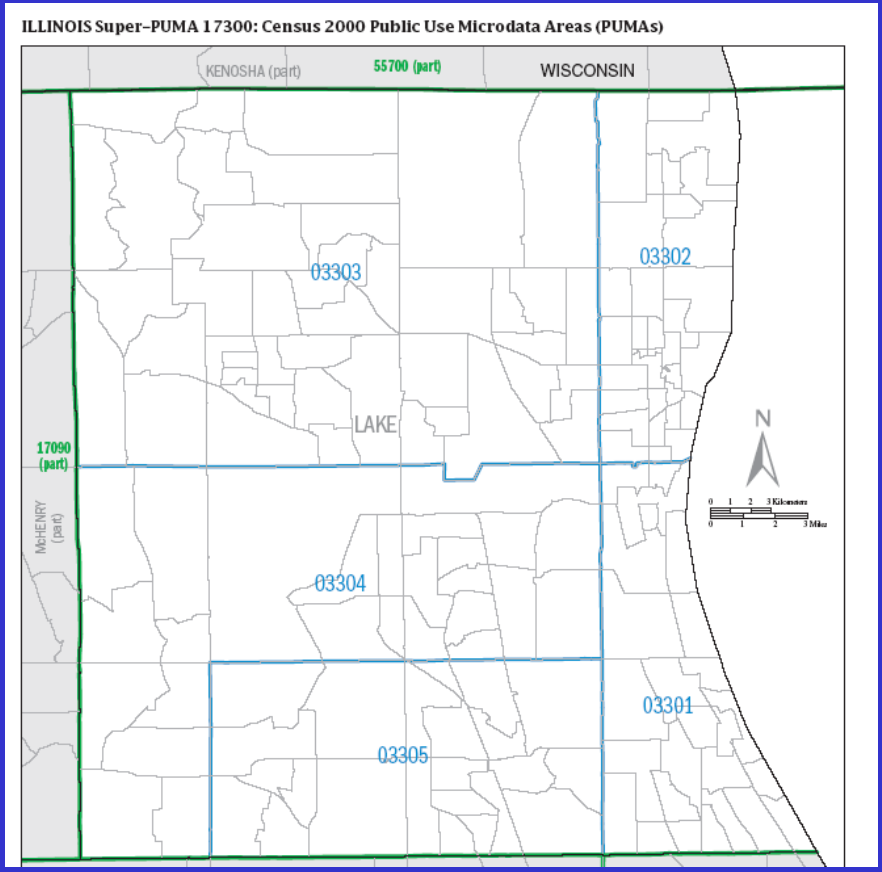
Public Use Microdata Areas

~ 100,000 population

ACS publishes annually

NRC (2007) recommended as possible level of analysis

PUMAs


ILLINOIS Super–PUMA 17300: Census 2000 Public Use Microdata Areas (PUMAs)
KENOSHA (part)
55700 (part)
WISCONSIN
03303
03302
LAKE
17090 (part)
McHENRY (part)
N
03304
03301
03305

PUMAs

Start with 1990 – 2000 trends

- Must approximate from townships
- AFF doesn’t provide tables
- Table 6 (in draft):

Table 6 PUMAs 1990-2000

PUMA	Pop 5+ 2000	1990 %	2000 %	Increase in %	s.e.	Increase in #	s.e.
03301	101,886	4.4%	7.3%	2.9%	0.4	3,023	375
03302	120,967	15.4%	32.6%	17.2%	0.7	23,413	702
03303	174,132	4.3%	8.9%	4.6%	0.4	10,556	418
03304	97,328	4.3%	9.0%	4.7%	0.5	5,407	366
03305	97,206	2.8%	5.1%	2.3%	0.4	2,891	295

PUMAs

1-year ACS data erratic, hard to analyze

3-year, 1999-2001 vs. 2003-2005,

- non-overlapping, in Table 7

5-year, 1999-2003 vs. 2001-2005,

- overlapping, in Table 8

Table 9 Annualized Change

PUMA	1-yr rate pers/yr	s.e.	t-test	1-yr rate pers/yr	s.e.	t-test
03301	418	225	1.9	334	268	1.2
03302	2,275	400	5.7	2,360	435	5.4
03303	2,140	358	6.0	2,156	348	6.2
03304	912	298	3.1	969	289	3.4
03305	75	210	0.4	108	209	0.5

Discussion: Guidelines

- Case study calls into question recommendation to avoid analysis of overlapping periods.
- In other respects, does case study fall within NRC guidance?

Discussion: Geographic level

- NRC report suggests difficult to use ACS to track change, except for large areas
- The case study appears to agree

- grouped townships into larger areas

- PUMA-level analysis possible

Discussion: Helping users

- Possible consideration: PUMA results from 1990 and 2000 censuses
- Standard errors for simple aggregates
- Possible displays of differences, trends?

Discussion: Tools

- The analysis was time consuming
- Primarily in Excel, but new study would require almost starting over
- Possible role for the R statistical software? Another approach?