Understanding the Margin of Errors and the Coefficient of Variance in the American Community Survey

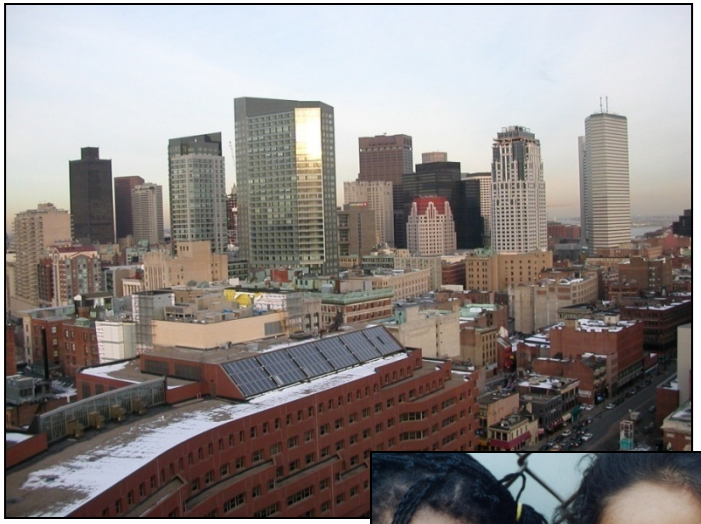
U.S. Census Bureau
Workshop at SACOG

Michael Burns
Deputy Regional Director

United States Census Bureau
Economic Statistics

American Community Survey

Four Main Types of Characteristics of the Population



Social

Economic

Housing

Demographic

United States Census Bureau

Economic Statistics

Expected improvements

- Five Year Coefficients of Variation (CVs) for typical tracts, by size where red > yellow > green

Tract Size Category	Average Tract Size	CVs before reallocation and sample expansion	CVs after reallocation, before sample expansion (2.9M)	CVs after reallocation and sample expansion (3.54M)
0 – 400	291	66%	41%	35%
401 – 1,000	766	41%	30%	25%
1,001 – 2,000	1,485	29%	29%	25%
2,000 – 4,000	2,636	26%	29%	25%
4,000 – 6,000	4,684	19%	29%	25%
6,000 +	8,337	15%	28%	25%

SAMPLING ERROR AND DEALING WITH MARGINS OF ERROR

United States Census Bureau
Economic Statistics

Probability Theory and Statistics

- All statistics are based on probability theory
- So if you do not like mathematical statistics, there are two French guys to blame:
- Pierre de Fermat and Blaise Pascal

Sample Design

- When designing a national survey, the Census Bureau has an advantage over all other research companies, even the large ones like NORC and RTI.
- We do the Census, so we not only have nationwide coverage of all population groups with their associated socio-economic characteristics, but also can draw a sample of housing units for a survey that is totally inclusive of all housing in the U.S.

How Can a Sample Represent the Whole Country?

Sample Design

- When designing a survey all you need to think about is chicken soup.
- How do you make chicken soup?
- Do you put 5 chickens in the soup or one chicken; a bunch of carrots or one carrot; and 2-3 stalks of celery or one stalk of celery?

Sample Design

Chicken Soup

- Water
- Chicken
- Celery
- Carrots
- Onion
- Garlic
- Salt
- Pepper
- Noodles
- Wine

Sample Design

- White
- African American
- Asian
- American Indian/Alaska Natives
- Hispanic
- Urban
- Rural
- Owner
- Renter
- Group Quarters

Proper Proportions

- Schichtung der Probenhilfen, die Veränderlichkeit in der Probenauswahl zu kontrollieren, nehmend dadurch die mathematische Veränderlichkeit im geschätzten Fehler ab (Fehlerspielraum (MOE)).
- Stratification of the sample helps to control the variability in the sample selection, thereby decreasing the mathematical variability in the estimated error (Margin of Error (MOE)).
- Doesn't the above sound like a bunch of gibberish?
- Let's get back to Chicken Soup!

Sample Design

Chicken Soup

- Water
- Chicken
- Celery
- Carrots
- Onion
- Garlic
- Salt
- Pepper
- Noodles
- Wine

Sample Design

- White
- African American
- Asian
- American Indian/Alaska Natives
- Hispanic
- Urban
- Rural
- Owner
- Renter
- Group Quarters

Stratification of the Sample

- Think of stratification as a fancy word that means groupings.
- The groupings are many since the grouping are cross tabulated when drawing the sample for all of our surveys, except for ACS.
  - White x rural x low income x homeowner
  - White x urban x medium income x renter
  - Afr Am x rural x high income x renter
  - Hispanic x urban x medium income x homeowner

ACS' Sample Stratification

- ACS has sixteen Strata
  - The strata are not cross tab on demographic characteristics, but on geographic size.
  - The strata are sorted by the size of addresses in each county by stratum and geographic order including tract, block, street name, and house number.

- The stratum assignment for a block is based on information about the set of geographic entities—referred to as sampling entities—which contain the block, or on information about the size of the census tract in which the block is located. Sampling entities are defined as:
  - Counties.
  - Places with active and functioning governments.
  - School districts.
  - American Indian Areas/Alaska Native Areas/Hawaiian Home Lands (AIANHH).
  - American Indian Tribal Subdivisions with active and functioning governments.
  - Minor civil divisions (MCDs) with active and functioning governments in 12 states

2012 Sampling – Summary Statistics (U.S.)

Sampling Stratum	Sampling Rate Definition	M12 Valid Addresses	S12 Valid Addresses	M12 Sampling Rate	S12 Sampling Rate	Final 2012 Sample
Totals	N/A	134,043,838	460,064	N/A	N/A	3,539,552
1	15%	1,211,251	3,310	15.00%	15.00%	181,355
2	10%	2,041,999	5,973	10.00%	10.00%	204,643
3	7%	3,982,496	12,068	7.00%	7.00%	279,459
4	2.8×BR	3,291,024	9,298	4.40%	2.74%	144,920
5	3.5×BR	152,940	974	5.50%	3.43%	8,429
6	0.92×3.5×BR	82,146	263	5.06%	3.16%	4,159
7	2.8×BR	5,058,766	10,661	4.40%	2.74%	222,649
8	0.92×2.8×BR	4,625,451	8,235	4.04%	2.52%	187,236
9	1.7×BR	21,774,868	40,398	2.67%	1.67%	581,816
10	0.92×1.7×BR	38,907,391	63,816	2.46%	1.53%	956,380
11	BR	14,229,122	223,043	1.57%	0.98%	225,643
12	0.92×BR	36,066,250	73,102	1.44%	0.90%	521,653
13	0.6×BR	489,081	1,695	0.94%	0.59%	4,613
14	0.92×0.6×BR	1,593,339	5,524	0.87%	0.54%	13,838
15	0.35×BR	83,946	120	0.55%	0.34%	463
16	0.92×0.35×BR	453,768	1,584	0.51%	0.32%	2,296

What are the Correct Proportions?

- The Census Bureau does the stratification based on:
  - Urban /Rural Designations
  - Sampling entities
- Stratifying the sample decreases the sample variability and thus decreases the Margin of Error.

One More Concept before We Discuss the Margin of Error: Standard Error

- The Standard Error measures the variability in the sample mean.

- We have to do a little more math to gain insight into how the Margin of Error works. We need to calculate the Standard Error, the formula is:

$$Standard\ Error = \frac{\sigma}{\sqrt{N}}$$

- The size of your sample effects the standard error and thus the Margin of Error (MOE). The larger your sample is, the smaller will be the Standard Error and therefore, the Margin of Error.

So what happens to the Standard Error when the # of addresses gets smaller in a sample?

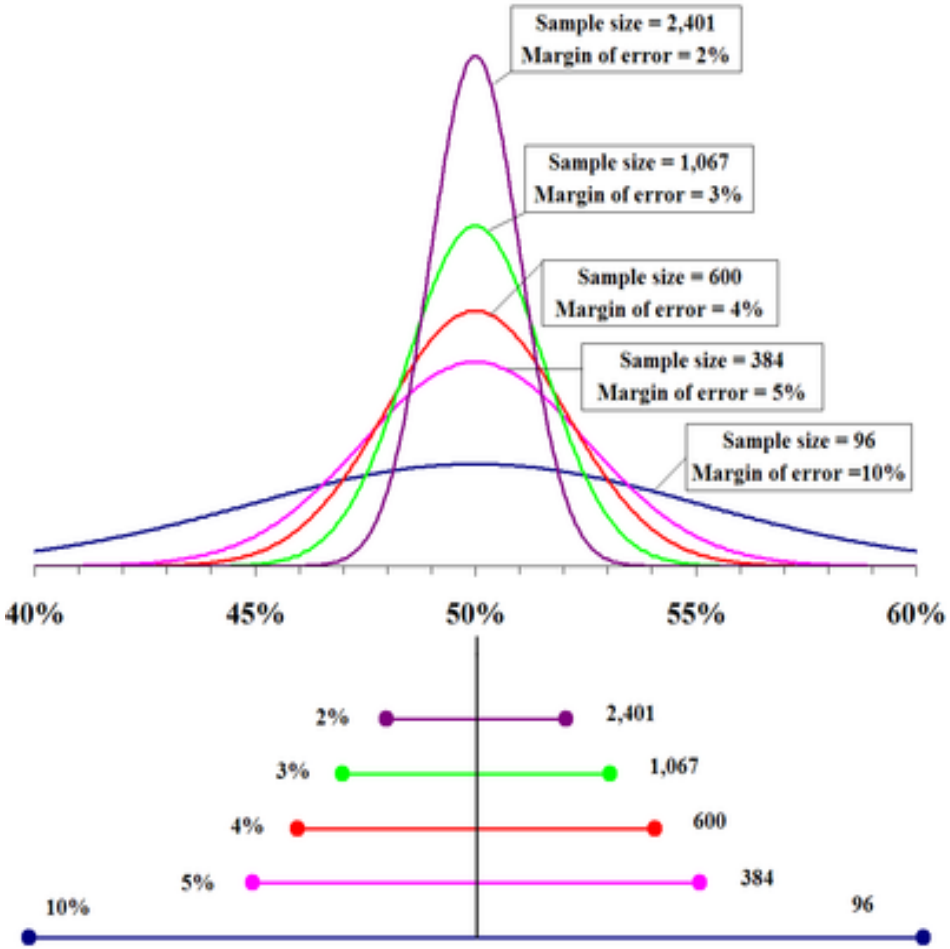
- Let's take an example: We are looking at household income in a U.S. State. The median household income is $56,384 and the standard deviation is $15,000. Let also say that the number of household in the State sample is 2,800,000 Hus. The standard error would be 8.9.
  - So let see what happens if we go down to the county level with 500,000 HUs. The standard error is 21.1
  - And if we go down to a city with 100,000 HUs? The Standard Error is 47.7
  - And if we go to a tract with 8000 HUs? The Standard Error is 168.1

Challenges of ACS
Sampling Error

- The *uncertainty* associated with an estimate that is based on data gathered from a sample of the population rather than the full population
- Margin of error (MOE) measures the precision of an estimate at a given level of confidence
- MOEs at the 90% confidence level for all published ACS estimates

Making Sense of The Margin of Error

- So the number of housing units in the sample has a direct effect on the Standard Error and the Margin of Error when choosing the confidence level of 90% on ACS.



Finally We can talk about the Margin of Error

- What is the Margin of Error:
  - Provides you with best estimation
    - A confidence level is used for the purpose of estimating a population parameter by using statistics (a single number that describes the population).
    - For example, the monthly unemployment rate for the country.
    - The Margin of Error is the amount of plus or minus that is attached to your sample results when you move from discussing the sample itself (the bowl of soup) to discussing the whole population (the large pot of soup) that the sample represents.

The Margin of Error

- The Margin of Error is not the chance a mistake was made.
- The Margin of Error measures the variation in the random samples due to chance.
- Because you did not interview all the housing units in the U.S., like you do in a census, you expect that your sample results will be “off” by a certain expect amount, just by chance.
- You acknowledge that your results could change with subsequent samples and that they are only accurate to within a certain range which is your Margin of Error (MOE).

Relating Margin of Error to Confidence Level

- ACS is at the 90% Confidence Level, which means?
  - I can draw 100 different ladles of soup (samples) from my big pot of soup (Total U.S. Population) and 90 ladles of soup will be within the parameter being studied→ Unemployment Rate
    - Unemployment rate is 8.4% ± 0.2
      - The range to account for the chance error which can be determined mathematically is 8.2% -- 8.6%.
      - That means I can take 90 ladles of soup from the big pot of soup and the unemployment rate will all fall with 8.2% to 8.6%
      - Only 10 ladles of soup(samples) would produce numbers outside of the 8.2% to 8.6% for the unemployment rate.

Margin of Error (MOE)

- **Adjusting your Confidence Level**
  - It is possible to construct margins of error with higher levels of confidence, such as 95 % or 99%.
  - This is done by adjusting the published margin of error.
- **Formula**
  - MOE = +/-1.645 x SE (90% level)
- **Values for other confidence levels**
  - 95% = 1.960
  - 99% = 2.576

Three Factors Effect the Size of the Margin of Error

- Three Factors:
  - The Confidence Level
  - The Sample Size
  - The Amount of Variability in the Population
- The ultimate goal when making an estimate using a confidence interval is to have a small margin of error.
  - The narrower the interval, the more precise the results are.

So why does ACS have such large MOEs at lower levels of Geography?

- Let's go back to chicken Soup and let's look at sample size:
  - State Level ACS Data

  - County Level ACS Data

  - City Level ACS Data

  - Tract Level ACS Data

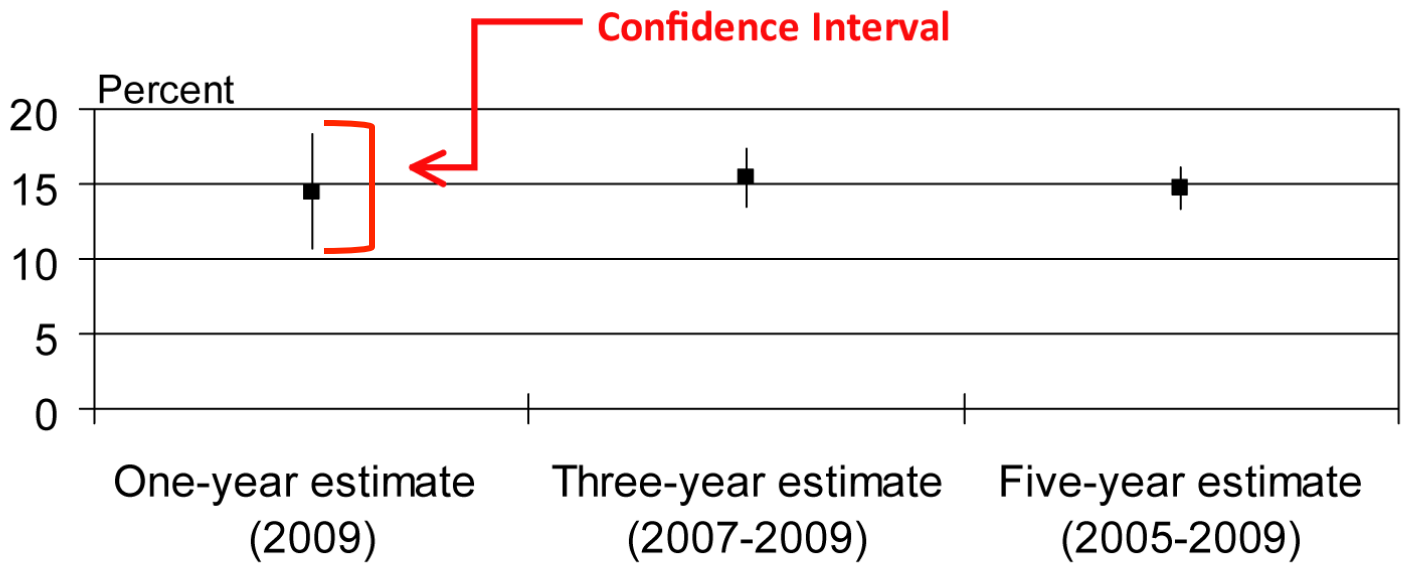
Interpreting the Data

United States Census Bureau
Economic Statistics

What is Reliability?

- Sampling Error is the uncertainty associated with an estimate that is based on data gathered from a sample of the population rather than the full population.
- Measures of sampling error give users an idea of how reliable, or precise, estimates are and speak to their fitness-for-use.
- Reliability is maximizing the inherent repeatability or consistency in an experiment.
  - Think of reliability in this vein. If your doctor checks your weight once and you get right back on the scale, you do not expect to see a difference or just a miniscule difference.
  - The closer the percent difference is to zero, the more reliable the measure.
  - But if you do see a large difference, then there is a reliability issue.

Reliability



Note: Fictional data

Measures of Sampling Error

- Standard Error (SE) – foundational measure of the variability of an estimate due to sampling
- Margin of Error (MOE) – precision of an estimate at a given level of confidence
- Confidence Interval (CI) - a range (based on a fixed level of confidence) that is expected to contain the population value of the characteristic
- Coefficient of Variation (CV) - The relative amount of sampling error associated with a sample estimate

Calculating Measures of Sampling Error

- At a 90 percent confidence level
  - Margin of Error MOE = SE x 1.645
  - Standard Error SE = MOE / 1.645
- Confidence Interval CI = Estimate +/- MOE
- Coefficient of Variance CV = SE / Estimate * 100%

Challenges of ACS

Margins of Error and Data Filtering

- We do not perform any data quality filtering for the 5-year ACS estimates.
- Check margins of error to ensure estimates have sufficient reliability for their intended use.
- You can improve the reliability of estimates by aggregating geographies or subpopulations.

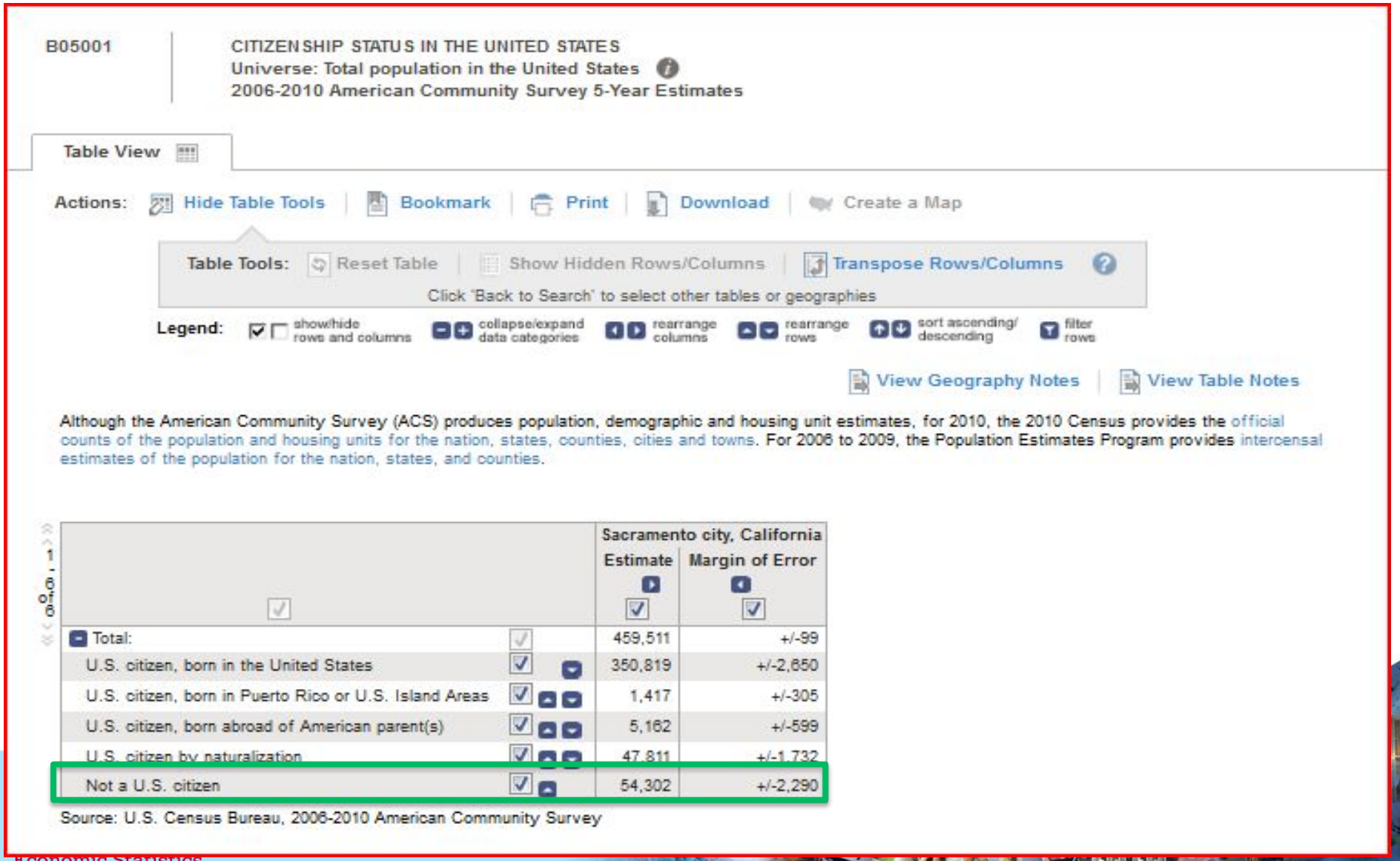
$$MOE_{agg} = \pm \sqrt{\sum_{c} MOE_{c}^{2}}$$

Example 1 – Assessing Utility

- Officials in Sacramento, CA are considering an outreach program to the non citizen population of the city. Officials need to know how many non-citizens are living in Sacramento, CA, but are concerned about how reliable the figure is. If there is high reliability, the city wants to institute an outreach program to teach new arrivals English at a reduced tuition.

- What do the 2006-2010 ACS 5-year estimates show?

Citizenship Status for Sacramento,CA



B05001

CITIZENSHIP STATUS IN THE UNITED STATES
Universe: Total population in the United States
2006-2010 American Community Survey 5-Year Estimates

Table View

Actions: Hide Table Tools | Bookmark | Print | Download | Create a Map

Table Tools: Reset Table | Show Hidden Rows/Columns | Transpose Rows/Columns

Click 'Back to Search' to select other tables or geographies

Legend: show/hide rows and columns; collapse/expand data categories; rearrange columns; rearrange rows; sort ascending/descending; filter rows

View Geography Notes | View Table Notes

Although the American Community Survey (ACS) produces population, demographic and housing unit estimates, for 2010, the 2010 Census provides the official counts of the population and housing units for the nation, states, counties, cities and towns. For 2006 to 2009, the Population Estimates Program provides intercensal estimates of the population for the nation, states, and counties.

	Sacramento city, California	
	Estimate	Margin of Error
Total:	459,511	+/-99
U.S. citizen, born in the United States	350,819	+/-2,650
U.S. citizen, born in Puerto Rico or U.S. Island Areas	1,417	+/-305
U.S. citizen, born abroad of American parent(s)	5,162	+/-599
U.S. citizen by naturalization	47,811	+/-1,732
Not a U.S. citizen	54,302	+/-2,290

Source: U.S. Census Bureau, 2006-2010 American Community Survey

Is the Reliability of the Data Good?

- City of Sacramento
  - Not a Citizen 54,302 ± 2290 (90% Confidence Level)
  - Which means (52,012 ← 54,302 → 56,592)

- Find the Standard Error (Standard Error SE = MOE / 1.645)
  - SE = 2290/1.645 ⟶ 1,392

- Coefficient of Variance CV = SE / Estimate * 100%
  - 1,392/54,302 x 100 = 2.5%

Expected improvements

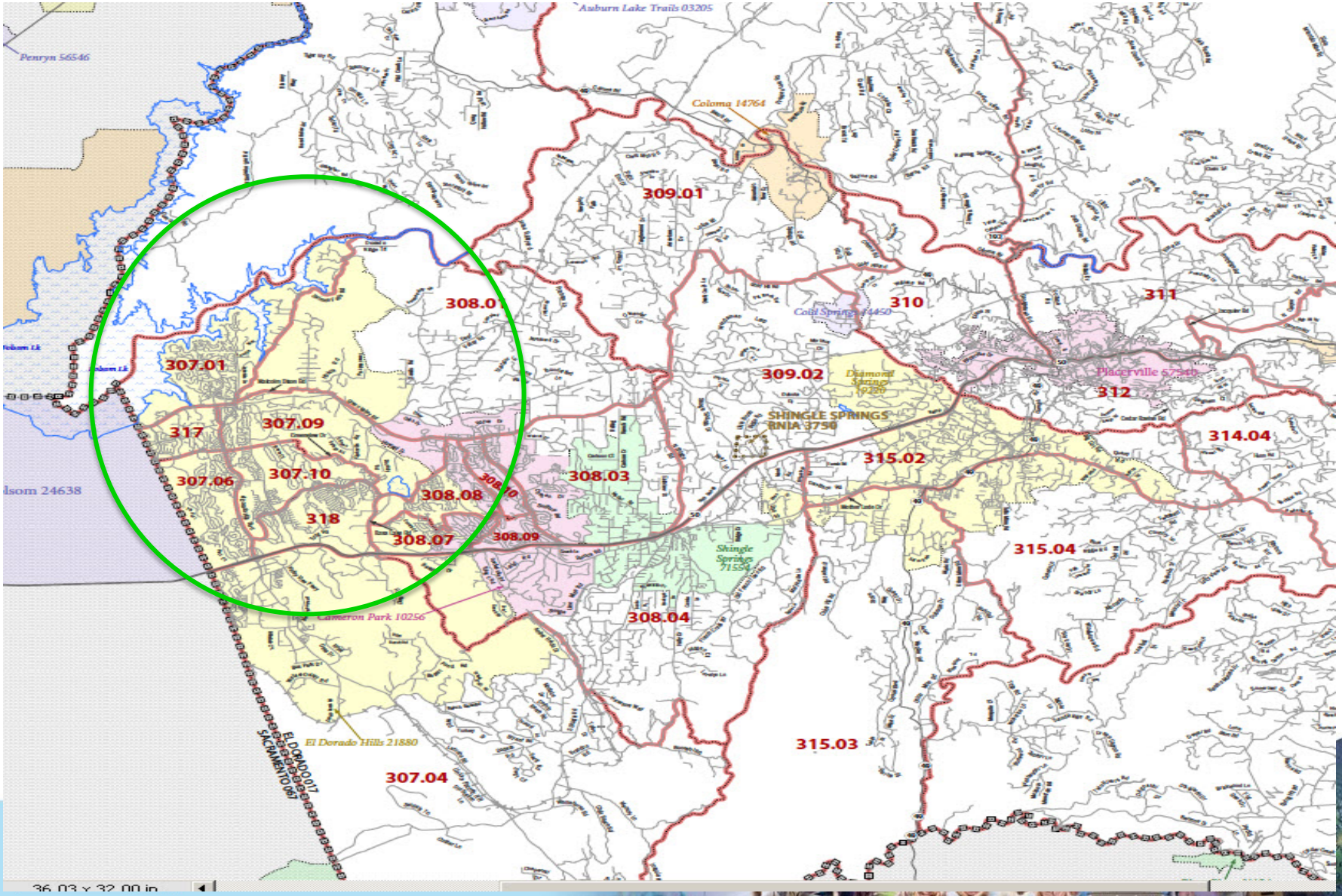
- Five Year Coefficients of Variation (CVs) for typical tracts, by size where red > yellow > green

Tract Size Category	Average Tract Size	CVs before reallocation and sample expansion	CVs after reallocation, before sample expansion (2.9M)	CVs after reallocation and sample expansion (3.54M)
0 – 400	291	66%	41%	35%
401 – 1,000	766	41%	30%	25%
1,001 – 2,000	1,485	29%	29%	25%
2,000 – 4,000	2,636	26%	29%	25%
4,000 – 6,000	4,684	19%	29%	25%
6,000 +	8,337	15%	28%	25%

Example 2
Consider combining geographic areas

- In the next example, we want a more reliable Coefficient of Variance for the receipt of Supplemental Security Income (SSI), Cash Public Assistance Income, or Food Stamps/SNAP in the past 12 months by Household Type for Children under 18 years in Households
- We are interested in Tracts 307.01, 307.06, 307.09, 307.10, 308.07, 308.08, 317 and 318.
- El Dorado County is applying for a grant in order to provide additional services for the county's children who live in households receiving some form of assistance.
- The grant writer first wants to see if they can use the data at the tract level or do they need to collapse cells to obtain a datum with improved reliability.

Example 2


Auburn Lake Trails 03205
Penryn 56546
Coloma 14764
309.01
308.0
310
311
307.01
309.02
Diamond Springs 17220
Placerville 67340
312
317
307.09
SHINGLE SPRINGS RNIA 3750
314.04
307.06
307.10
308.03
315.02
308.08
318
308.07
308.09
Shingle Springs 71554
315.04
Cameron Park 10256
308.04
El Dorado Hills 21880
315.03
307.04
EL DORADO 017
SACRAMENTO 067
36.03 x 32.00 in

Example 2

B9010 -- RECEIPT OF SUPPLEMENTAL SECURITY INCOME (SSI), CASH PUBLIC ASSISTANCE INCOME, OR FOOD STAMPS/SNAP IN THE PAST 12 MONTHS BY HOUSEHOLD TYPE FOR CHILDREN UNDER 18 YEARS IN HOUSEHOLDS				
Living in a HH w/ SSI, SNAP, etc	ESTIMATE	MOE	SE	CV
Tract 307.01	24	±35	21.27	88.6%
Tract 307.06	50	±69	41.94	83.9%
Tract 307.09	29	±37	22.49	77.6%
Tract 307.10	30	±49	29.78	99.3%
Tract 308.07	55	±55	33.43	60.8%
Tract 308.08	183	±119	72.34	39.5%
Tract 317	66	±107	65.04	98.6%
Tract 318	61	±71	37.08	60.1%

Example 2 - Calculations

B9010 -- RECEIPT OF SUPPLEMENTAL SECURITY INCOME (SSI), CASH PUBLIC ASSISTANCE INCOME, OR FOOD STAMPS/SNAP IN THE PAST 12 MONTHS BY HOUSEHOLD TYPE FOR CHILDREN UNDER 18 YEARS IN HOUSEHOLDS

Living in a HH w/ SSI, SNAP, etc	Estimate	MOE	MOE^2	Square root of sum
Tract 307.01	24	±35	1,225	
Tract 307.06	50	±69	4,761	
Tract 307.09	29	±37	1,369	
Tract 307.10	30	±49	2,401	
Tract 308.07	55	±55	3,025	
Tract 308.08	183	±119	14,161	
Tract 317	66	±107	11,449	
Tract 318	61	±71	5,041	
Combined	498	±208	43,432	208

Source: 2006-2010 ACS 5-Year Estimates

Example 2- Results

B9010 -- RECEIPT OF SUPPLEMENTAL SECURITY INCOME (SSI), CASH PUBLIC ASSISTANCE INCOME, OR FOOD STAMPS/SNAP IN THE PAST 12 MONTHS BY HOUSEHOLD TYPE FOR CHILDREN UNDER 18 YEARS IN HOUSEHOLDS

Living in a HH w/ SSI, SNAP, etc	HH ESTIMATE	MOE	SE	CV
Tract 307.01	24	±35	21.27	88.6%
Tract 307.06	50	±69	41.94	83.9%
Tract 307.09	29	±37	22.49	77.6%
Tract 307.10	30	±49	29.78	99.3%
Tract 308.07	55	±55	33.43	60.8%
Tract 308.08	183	±119	72.34	39.5%
Tract 317	66	±107	65.04	98.6%
Tract 318	61	±71	37.08	60.1%
Combined	498	±208	126	25.3%

Standard Error (SE) = MOE / 90% Confidence Interval. So 208 / 1.645 = 126 (SE)
Coefficient of Variance (CV) = Standard Error (SE) / HH Estimate. So 126 / 498 = 25.3%

Example 2 Summary

- Combining data for 8 neighboring tracts improved the reliability of the detailed data; collapsing this detail improved the estimate even more.

- Users need to consider the most important dimensions – geography or characteristic detail when considering collapsing.

ACS Calculator

Oklahoma Department of Commerce

OKLAHOMA
Department of Commerce

American Community Survey Calculators

Last updated: December 6, 2010

For further information, please contact:
Oklahoma Department of Commerce
Steve Barker
E-mail: steven_barker@okcommerce.gov
Phone: 1-405-815-5182 or 1-800-879-6552
www.okcommerce.gov

Quick Links

- ACS Made Easy
- Change Confidence Interval to 95% or 99% - Values
- Change Confidence Interval to 95% or 99% - Percentages
- Add ACS estimates together - Values
- Add ACS estimates together - Percentages
- Calculate a ratio
- Change a value to a percentage
- Change a percentage to a value
- Calculate percent change or percent difference
- Full text of Census advice used to develop these calculators
- Additional Information

Contact Info | ACS made easy | Confidence Interval - Value | Confidence Interval - Percent | Adding Data - Vali

http://www.okcommerce.gov/Data-And-Research/Demographic-And-Population-Data

Summary
Extrapolation to Large Data Sets

- Four Methods of Improving Reliability

1. Find a pre-existing table at a higher degree of aggregation

2. Collapse data cells to a higher degree of aggregation

3. Add geographies together (Example 2)

4. Collapse data cells and add geographies together

Questions?