New American Community Survey Data Products for 2016

Jonathan Buttle
Demographic Research Unit
Department of Finance
October 13, 2016

In July 2016, the Census Bureau released three new data products derived from the American Community Survey (ACS). They include:

➢ **2014 ACS (1-year estimates) Supplement Estimates Tables –**
58 detailed subject tables for geographies with populations of 20,000 or more;

➢ **2010-2014 ACS (5-year estimates) Variance Replicate Estimates Tables –**
107 American Community Survey tables across 11 geographic levels augmented to include 80 variance replicate estimates that can be used to calculate margins of error for user-generated estimates; and

➢ **Statistical Testing Spreadsheet Tool –**
An Excel workbook designed to assist with the statistical comparison of American Community Survey estimates across geographies or years.

2

2014 ACS (1-year estimates) Supplemental Estimates Tables

➤ Consist of 58 detailed subject tables –
  • Tables are drawn from nearly all of the ACS primary categories;

➤ The Supplemental Estimates Tables are best used when –
  • Currency is more important than precision;

  • Analyzing smaller populations; and

  • Examining smaller geographies because the standard 1-year estimates are not available.

3

No tables included from Grandparents or Grandchildren. See the handout for a complete list of all tables.

Precision refers to increased statistical reliability for data from less populated areas or small population subgroups.

2014 ACS (1-year estimates) Supplemental Estimates Tables

➤ The supplemental tables are versions of tables already available for larger geographies.

Table K200201 - Race	Sacramento County	
2014 ACS Supplemental Estimates	Estimate	Margin of Error
Total:	1,482,026	*****
White alone	871,895	+/-8,170
Black or African American alone	150,307	+/-3,636
American Indian and Alaska Native alone	10,329	+/-3,187
Asian alone	228,084	+/-5,154
Native Hawaiian and Other Pacific Islander alone	16,250	+/-1,198
Some other race alone	112,118	+/-8,865
Two or more races	93,043	+/-6,881

Table B02001 - Race	Sacramento County	
2014 ACS (1-year estimates)	Estimate	Margin of Error
Total:	1,482,026	*****
White alone	871,895	+/-8,170
Black or African American alone	150,307	+/-3,636
American Indian and Alaska Native alone	10,329	+/-3,187
Asian alone	228,084	+/-5,154
Native Hawaiian and Other Pacific Islander alone	16,250	+/-1,198
Some other race alone	112,118	+/-8,865
Two or more races:	93,043	+/-6,881
Two races including Some other race	17,449	+/-2,915
Two races excluding Some other race, and three or more races	75,594	+/-6,503

2014 ACS (1-year estimates) Supplemental Estimates Tables

➢ Contain the same geographic summary levels as those in the ACS 1-year dataset (Summary Level in parenthesis) –

- United States (010)
- Region (020)
- Division (030)
- States, the District of Columbia, and Puerto Rico (040)
- County or equivalent (050)
- County Subdivision (060)
- Place (incorporated places and CDPs - 160)
- American Indian Area/Alaska Native Area/Hawaiian Home Land (250)

- Metropolitan Statistical/Micropolitan Statistical Area (310)
- Principal City of Metropolitan or Micropolitan Statistical Areas (312)
- Metropolitan Division (314)
- Combined Statistical Area (330)
- Urban Area (400)
- Congressional Districts - 114th Congress (500)
- Public Use Microdata Area (795)
- School District (Elementary, Secondary, Unified – 950, 960, 970)

5

2014 ACS (1-year estimates) Supplemental Estimates Tables

➢ Estimates are available for geographies with populations of 20,000 or more –
   • The publication threshold is determined by:
      › The estimate taken from the Census Bureau's population estimates program – For state, county, sub-county, and place; or
      › The ACS estimate - For all others summary levels (CDPs, etc.).

   • Estimates are derived from ACS surveys for that geography.

6

In addition, the estimate will be published if the area's total population is within 5% of the population threshold and the area was published in the previous year.

2014 ACS (1-year estimates) Supplemental Estimates Tables

➢ For California, geographic entities include:

- California;

- 50 out of 58 California counties (compared with 40 in the 2014 ACS 1-year dataset);

- 345 out of 1,516 California places (compared with 136 in the 2014 ACS 1-year dataset).
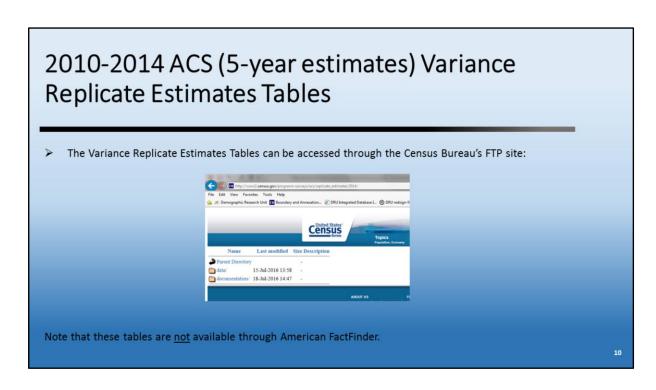
2014 ACS (1-year estimates) Supplemental Estimates Tables

➢ Estimates can be accessed through either:
  • American FactFinder; or
  • The Census Bureau's Application Programming Interface (API);

➢ The planned release date for the 2015 ACS Supplemental Estimates Tables is October 20, 2016.

8

The Census Bureau's API allows developers to design web and mobile apps that provides users with quick and easy access to data found in various Census Bureau public data sets. These data sets include the ACS 1-year, 3-year, and 5-year estimates and margins of error (various years), Decennial Census (1990-SF1 and SF3, 2000-SF1 and SF3, and 2010-SF1) and Current Population Survey poverty estimates (1959-2014).

More information is available at
https://www.census.gov/data/developers/data-sets/ACS-Supplemental-Data.html.

2010-2014 ACS (5-year estimates) Variance Replicate Estimates Tables

- ➢ The Variance Replicate Estimates Tables are augmented versions of 107 existing ACS tables across 11 geographic levels;

- ➢ Includes original published estimates and margins of error (MOE) for each characteristic;

- ➢ Also contains the estimate standard error and 80 variance replicate estimates for each characteristic.

2010-2014 ACS (5-year estimates) Variance Replicate Estimates Tables

➢ The Variance Replicate Estimates Tables can be accessed through the Census Bureau's FTP site:

Note that these tables are <u>not</u> available through American FactFinder.

Web address is: http://www2.census.gov/programs-surveys/acs/replicate_estimates/2014/

2010-2014 ACS (5-year estimates) Variance Replicate Estimates Tables

➢ The Variance Replicate Estimates Tables are organized by:

1. Geography; and
2. Table Number (i.e., subject);

➢ The Tables are downloaded as zipped csv files. When unzipped they are in this format (opened in Excel):

TBLID	GEOID	NAME	ORDER	TITLE	estimate	moe	CME	SE	Var_Rep1	Var_Rep2	Var_Rep3	Var_Rep80
B03001				HISPANIC OR LATINO ORIGIN BY SPECIFIC ORIGIN								
B03001				Universe: Total population								
B03001	04000US01	Alabama		1 Total:	4817678		*****	0	4817678	4817678	4817678	4817678
B03001	04000US01	Alabama		2 Not Hispanic or Latino	4625840	491 +/-491		298	4626038	4625693	4625857	4625732
B03001	04000US01	Alabama		3 Hispanic or Latino:	191838	491 +/-491		298	191640	191985	191821	191946
B03001	04000US01	Alabama		4 Mexican	128902	2815 +/-2,815		1711	129960	127905	129741	130266
B03001	04000US01	Alabama		5 Puerto Rican	14814	1516 +/-1,516		922	15488	14244	14462	14847
B03001	04000US01	Alabama		6 Cuban	4667	705 +/-705		429	4386	5067	4450	4405
B03001	04000US01	Alabama		7 Dominican (Dominican Republic)	920	290 +/-290		176	1064	956	957	886

2010-2014 ACS (5-year estimates) Variance Replicate Estimates Tables

➢ How are these tables useful?
  - Users can calculate MOEs for derived estimates across categories or geographies or for ratios, proportions, and percentages using a methodology similar to the one utilized during ACS data production (Successive Differences Replication-SDR);

  - Previously, computing MOEs for user-derived estimates from ACS summary tables require using approximation formulas.
    › The approximation formula entails squaring the MOE for each estimate being combined, summing the squares, and taking the square root of the sum.

12

2010-2014 ACS (5-year estimates) Variance Replicate Estimates Tables

➢ SDR Methodology –

- The Census Bureau employs successive differences replication to the derive the MOEs that appear in the summary tables;

- The calculation entails computing the official ACS estimates along with 80 pseudo-estimates or variance replicates;

- For the majority of ACS estimates, these values are used to form the variance of the estimate;

- This is the methodology used to calculate MOEs in the ACS Public Use Microdata Sample (PUMS);

- The variance estimate can be calculated without consideration of the form of the statistic (it can be a count, mean, ratio, etc.) or the complexity of the sampling and weighting procedure.

13

For details and a technical description of the variance estimation process employed by the Census Bureau, see: http://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/acs_design_methodology_ch12_2014.pdf

2010-2014 ACS (5-year estimates) Variance Replicate Estimates Tables

➢ Approximation formulas have issues –

- Approximation formulas do not account for the covariance between the estimates used in the calculation;

- Not accounting for the covariance between estimates can either understate or overstate the error if the estimates interact in either a positive or negative way.

Covariance is a measure of the interaction between two or more estimates. Specifically, it is a measure of how two or more variables or estimates change together. The Census Bureau does not provide the covariance for ACS estimates and if the covariance is very large it can result in a calculated MOE that differs greatly from the actual MOE.

2010-2014 ACS (5-year estimates) Variance Replicate Estimates Tables

➢ Derived MOEs using the approximation formula versus replicate weights –

- The estimate of the number of persons of Mexican, Salvadoran, or Guatemalan origin along with its MOE (from Table B03001) for California:

  › The estimate: 13,182,882

  › The MOE:
    ❖ Using the approximation formula: 23,830;
    ❖ Using replicate weights: 17,219.

2010-2014 ACS (5-year estimates) Variance Replicate Estimates Tables

➢ Calculating MOEs using replicate weights

1. Calculate the variance for the estimate.

The variance is calculated using the following formula:

$$Variance = \frac{4}{80} \sum_{i=1}^{80} (Estimate - Var\_Rep_i)^2$$

In English: the variance calculation is simply the sum of the squared differences of the estimate and each of the replicate weights multiplied by 4/80.

16

2010-2014 ACS (5-year estimates) Variance Replicate Estimates Tables

➢ Calculating MOEs using replicate weights (cont.)

2. Take the square root of the variance (also known as the standard error) and multiply by 1.645 to derive a 90 percent confidence level (the confidence level used by the Census Bureau);

• For different confidence levels, use different multipliers; i.e. –
  › 1.96 for a 95 percent confidence level, or
  › 2.58 for a 99 percent confidence level.

➢ Using replicate weights to calculate MOEs can be done in Excel, or with any of the popular statistical software packages (i.e., R, SAS, Stata, etc.);

17

2010-2014 ACS (5-year estimates) Variance Replicate Estimates Tables

➢ Examples of using Replicate Estimates Tables to calculate estimates and MOEs for derived measures:

- The number of non-relative persons living in family households with income below the poverty level in the past 12 months for Sacramento County; and

- The poverty rate of non-relative persons living in family households with incomes below the poverty level in the past 12 months.

18

2010-2014 ACS (5-year estimates) Variance Replicate Estimates Tables

➢ Table B17021 contains estimates for the poverty status of persons in the past 12 months by living arrangement . . .

TBLID	GEOID	NAME	ORDER	TITLE	estimate	moe	CME	SE	Var_Rep1	Var_Rep2	...	Var_Rep80
B17021	05000US06067	Sacramento County, California		1 Total:	1,427,006	1,484	1,484	902	1,427,183	1,426,508	...	1,426,451
B17021	05000US06067	Sacramento County, California		2 Income in the past 12 months below poverty level:	258,031	6,196	6,196	3,767	257,927	259,071	...	258,376
B17021	05000US06067	Sacramento County, California		3 In family households:	197,174	5,749	5,749	3,495	196,559	197,441	...	196,657
B17021	05000US06067	Sacramento County, California		4 In married couple families:	83,253	3,950	3,950	2,401	82,363	83,375	...	83,006
B17021	05000US06067	Sacramento County, California		5 All relatives	80,203	3,942	3,942	2,397	79,331	80,405	...	80,042
B17021	05000US06067	Sacramento County, California		6 Non-relatives	3,050	399	399	242	3,032	2,970	...	2,964
B17021	05000US06067	Sacramento County, California		7 In other families:	113,921	3,918	3,918	2,382	114,196	114,066	...	113,651
B17021	05000US06067	Sacramento County, California		8 Male householder, no wife present:	23,502	1,994	1,994	1,212	23,165	22,834	...	22,729
B17021	05000US06067	Sacramento County, California		9 All relatives	17,275	1,754	1,754	1,066	16,847	16,621	...	16,569
B17021	05000US06067	Sacramento County, California		10 Non-relatives	6,227	697	697	423	6,318	6,213	...	6,160
B17021	05000US06067	Sacramento County, California		11 Female householder, no husband present:	90,419	3,725	3,725	2,265	91,031	91,232	...	90,922
B17021	05000US06067	Sacramento County, California		12 All relatives	81,730	3,598	3,598	2,187	82,148	82,748	...	82,319
B17021	05000US06067	Sacramento County, California		13 Non-relatives	8,689	663	663	403	8,883	8,484	...	8,603

2010-2014 ACS (5-year estimates) Variance Replicate Estimates Tables

➢ And contains the data we will need to construct our estimates and MOEs . . .

TBLID	GEOID	NAME	ORDER	TITLE	estimate	moe	CME	SE	Var_Rep1	Var_Rep2	...	Var_Rep80
B17021	05000US06067	Sacramento County, California	1	Total:	1,427,006	1,484	1,484	902	1,427,183	1,426,508	...	1,426,451
B17021	05000US06067	Sacramento County, California	2	Income in the past 12 months below poverty level:	258,031	6,196	6,196	3,767	257,927	259,071	...	258,376
B17021	05000US06067	Sacramento County, California	3	In family households:	197,174	5,749	5,749	3,495	196,559	197,441	...	196,657
B17021	05000US06067	Sacramento County, California	4	In married couple families:	83,253	3,950	3,950	2,401	82,363	83,375	...	83,006
B17021	05000US06067	Sacramento County, California	5	All relatives	80,203	3,942	3,942	2,397	79,331	80,405	...	80,042
B17021	05000US06067	Sacramento County, California	6	Non-relatives	3,050	399	399	242	3,032	2,970	...	2,964
B17021	05000US06067	Sacramento County, California	7	In other families:	113,921	3,918	3,918	2,382	114,196	114,066	...	113,651
B17021	05000US06067	Sacramento County, California	8	Male householder, no wife present:	23,502	1,994	1,994	1,212	23,165	22,834	...	22,729
B17021	05000US06067	Sacramento County, California	9	All relatives	17,275	1,754	1,754	1,066	16,847	16,621	...	16,569
B17021	05000US06067	Sacramento County, California	10	Non-relatives	6,227	697	697	423	6,318	6,213	...	6,160
B17021	05000US06067	Sacramento County, California	11	Female householder, no husband present:	90,419	3,725	3,725	2,265	91,031	91,232	...	90,922
B17021	05000US06067	Sacramento County, California	12	All relatives	81,730	3,598	3,598	2,187	82,148	82,748	...	82,319
B17021	05000US06067	Sacramento County, California	13	Non-relatives	8,689	663	663	403	8,883	8,484	...	8,603

20

2010-2014 ACS (5-year estimates) Variance Replicate Estimates Tables

➢ Summing the estimates and Var_Rep1 to Var_Rep80 for non-related persons and calculating the percentages of the total persons with incomes below poverty level gives:

TBLID	GEOID	NAME	ORDER	TITLE	estimate	moe	CME	SE	Var_Rep1	Var_Rep2	...	Var_Rep80
B17021	05000US06067	Sacramento County, California		2 Income in the past 12 months below poverty level:	258,031	6,196	6,196	3,767	257,927	259,071	...	258,376
B17021	05000US06067	Sacramento County, California		Non-relatives	17,966	1,041			18,233	17,667	...	17,727
B17021	05000US06067	Sacramento County, California		Percent, Non-relatives	7.0%				7.1%	6.8%	...	6.9%

You'll notice the MOE next to the estimate for non-relatives. I calculated this MOE using the approximation formula as a comparison to the MOE we will calculate using SDR methodology. We'll come back to that in a moment.

2010-2014 ACS (5-year estimates) Variance Replicate Estimates Tables

➢ Applying the SDR variance formula to the estimates and variance replicate weights gives:

	ACS Estimate	Replicate Estimate	Difference	Difference Squared	Percent Estimate	Replicate Estimate	Difference (x100)	Difference Squared
Var_Rep1	17,966	18,233	-267	71,289	7.0%	7.1%	-0.1063	0.0113
Var_Rep2	17,966	17,667	299	89,401	7.0%	6.8%	0.1434	0.0206
Var_Rep3	17,966	18,336	-370	136,900	7.0%	7.2%	-0.2282	0.0521
Var_Rep4	17,966	18,515	-549	301,401	7.0%	7.2%	-0.2090	0.0437
...	...	...	...	...	...	...	...	...
Var_Rep80	17,966	17,727	239	57,121	7.0%	6.9%	0.1018	0.0104
		Sum of Differences Squared		8,850,871		Sum of Differences Squared		1.3776
		Variance		442,544		Variance		0.0689
		Standard Error		665		Standard Error		0.2625
		MOE (90% CL)		1,094		MOE (90% CL)		0.4317

This is an illustration of the how the calculation works using the SDR variance formula. The next series of slides will illustrate how to use Excel to calculate the MOE in a more user-friendly way.

2010-2014 ACS (5-year estimates) Variance Replicate Estimates Tables

- ➢ Approximation formula vs. SDR Method
  - As noted above, each method will yield different MOEs;
  - The first example illustrated that the SDR Method MOE can be smaller than the one derived from using the approximation formula;
  - This example illustrated the opposite – that the SDR Method MOE can be larger than the MOE derived using the approximation formula –
    - › MOE for non-relatives in family households:
      - › SDR: 1,094
      - › Approximation Formula: 1,041
  - The absence of the covariance between estimates can have an unpredictable effect on the magnitude of the resulting MOE using the approximation formula.
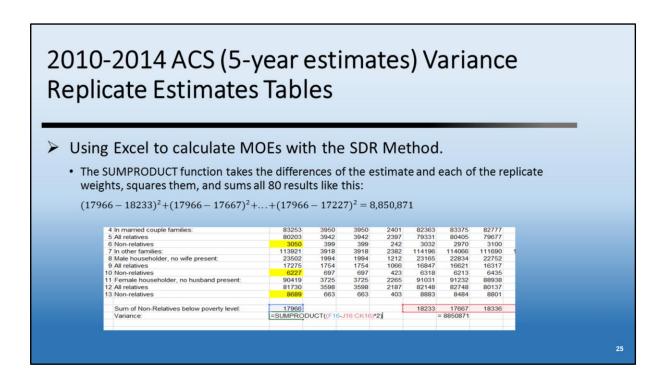
2010-2014 ACS (5-year estimates) Variance Replicate Estimates Tables

➢ Using Excel to calculate MOEs with the SDR Method.

- As noted above, the Tables are downloaded as csv files, which can be opened directly into Excel.

TBLID	GEOID	NAME	ORDER	TITLE	estimate	moe	CME	SE	Var_Rep1	Var_Rep2	Var_Rep3	Var_Rep80
B17021				POVERTY STATUS OF INDIVIDUALS IN THE PAST 12 MONTHS BY LIVING ARRANGEMENT								
B17021				Universe: Population for whom poverty status is determined								
B17021	05000US01001	Autauga County, Alabama	1 Total:		54712	139	139	84	54683	54712	54688	54735
B17021	05000US01001	Autauga County, Alabama	2 Income in the past 12 months below poverty level:		7006	935	935	569	7145	7018	7317	7043
B17021	05000US01001	Autauga County, Alabama	3 In family households:		5312	930	930	566	5342	5319	5518	5429
B17021	05000US01001	Autauga County, Alabama	4 In married couple families:		2030	792	792	481	2027	2100	2064	2172
B17021	05000US01001	Autauga County, Alabama	5 All relatives		1993	798	798	485	1969	2055	2001	2146
B17021	05000US01001	Autauga County, Alabama	6 Non-relatives		37	42	42	26	58	45	63	26
B17021	05000US01001	Autauga County, Alabama	7 In other families:		3282	719	719	437	3315	3219	3454	3257
B17021	05000US01001	Autauga County, Alabama	8 Male householder, no wife present:		175	130	130	79	153	129	214	187
B17021	05000US01001	Autauga County, Alabama	9 All relatives		118	110	110	67	90	65	173	138
B17021	05000US01001	Autauga County, Alabama	10 Non-relatives		57	47	47	28	63	64	41	49

The csv files are laid out in such a way that makes combining estimates easier (down columns) but can be challenging to make use of the SDR variance formula.

Excel to the rescue – the SUMPRODUCT function! The SUMPRODUCT function allows the user to take the sum of the squares of the differences of the estimate and each of the replicate weights in one step.

In the example above, the formula is =SUMPRODUCT((F16-J16:CK16)^2)

2010-2014 ACS (5-year estimates) Variance Replicate Estimates Tables

➢ Using Excel to calculate MOEs with the SDR Method.

- Multiplying the result of SUMPRODUCT by $\frac{4}{80}$ yields the estimate variance:
  $$\left(\frac{4}{80}\right) \times 8{,}850{,}871 \sim 442{,}544$$

- Taking the square root of the variance and multiplying by 1.645 gives the MOE:
  $$1.645 \times \sqrt{442{,}543.6} \sim 1{,}094$$

81730	3598	3598	2187	82148	82748	80137	80
8689	663	663	403	8883	8484	88	
17966				18233	17667	183	
=(4/80)*SUMPRODUCT((F16-J16:CK16)^2)				= 442543.6			

81730	3598	3598	2187	82148	82748	80137	80
8689	663	663	403	8883	8484	8801	
17966				18233	17667	18336	18
=SQRT((4/80)*SUMPRODUCT((F16-J16:CK16)^2))*1.645				= 1094.319			

2010-2014 ACS (5-year estimates) Variance Replicate Estimates Tables

➢ When NOT to use the SDR Method to calculate MOEs

- There are a number of special situations where this methodology is not suitable. They include:
    › Where a count estimate is zero, or a percent estimate is zero or 100 percent;
    › Where one or more of the estimates is controlled (i.e., the MOE in the file for that estimate is assigned a value of '*****');
    › Where the sum of non-controlled estimates equals a controlled value; and
    › A percent or ratio estimate with one or more undefined variance replicate estimates;

- Please refer to the Replicate Tables documentation for more details.

➢ 2011-2015 Replicate Estimates Tables Release: January 19, 2017

27

The documentation for the Variance Replicate Estimates Tables can be found at http://www2.census.gov/programs-surveys/acs/replicate_estimates/2014/documentation/5-year/

Statistical Testing Spreadsheet Tool

➢ An Excel workbook that assists users with comparing ACS estimates on a statistical basis;

➢ Consists of two spreadsheets – one for testing two estimates (up to 3,200 pairs) and one for testing multiple estimates (up to 150 estimates). The workbook also includes instructions and a worked example;

➢ Handles special characters, such as '+/-' in front of the MOE;

➢ Default is 90 percent confidence level but can test at the 95 or 99 percent confidence level.

28

Statistical Testing Spreadsheet Tool

➤ Examples –

• Comparing pairs of estimates across ACS years for the same geography:

Sacramento City	Percentage Persons in Poverty	MOE
2015 ACS	20.3	+/-1.7
2014 ACS	22.6	+/-1.8
2013 ACS	23.4	+/-2.1
2012 ACS	23.4	+/-1.7

Table DP-3, 2012-2015 ACS (1-year estimates)

• Comparing multiple estimates across geographies for the same ACS year:

	Percentage Persons in Poverty	MOE
El Dorado County, California	10.3%	+/-0.9
Nevada County, California	12.4%	+/-1.4
Placer County, California	8.9%	+/-0.6
Sacramento County, California	18.1%	+/-0.4
Sutter County, California	16.8%	+/-1.5
Yolo County, California	20.0%	+/-0.9
Yuba County, California	22.2%	+/-2.1

Table DP-3, 2010-2014 ACS (5-year estimates)

Statistical Testing Spreadsheet Tool

➤ Results (at 90% conf. level) –

The Tool shows that there is no statistical difference between the 2014 and the 2015 ACS percent persons in poverty and no statistical difference between the 2012 and the 2013 ACS percent persons in poverty;

Here the Tool illustrates that the percentages of persons below poverty for Sacramento and Sutter counties are not statistically different and that the percentages for Yolo and Yuba counties are not statistically different. The remaining estimates are statistically different from each other.

Questions? Contact Us!

Demographic Research Unit
California Department of Finance
http://www.dof.ca.gov/Forecasting/Demographics/

ficalpop@dof.ca.gov
(916) 323-4086

Thank You!