

With the release of the first 5-year American Community Survey (ACS), many users will want to aggregate small levels of geography into larger areas. In fact, the Census Bureau encourages users to aggregate census tract or smaller areas rather than to look at the areas individually.

When you aggregate areas, you essentially create a new estimate and must calculate the margin of error (MOE) for that new estimate. You can find the formulas to recalculate the MOE for aggregated counts, derived proportions (percentages), and derived ratios (such as, persons per household) in Appendix 3 of the ACS Compass Handbooks at <http://www.census.gov/programs-surveys/acs/guidance/handbooks.html>. But these publications do not include formulas used to re-calculate medians.

Medians are **NOT** measures that you can just “add up” when you aggregate geographic areas. Medians are not arithmetically derived, so, although you can create an average of the medians, it is not a true median. The median will need to be recalculated for each aggregated area.

Since we are using the Summary File and do not have access to the actual survey returns, the only way to recalculate a median is to use range tables. The smaller the width of the range; the better the calculated median will be - so use the most detailed table you can find. The following example recalculates median household income for an aggregation of areas using Table B19001 – Household Income.

### How to Calculate a Median from Range Data:

- (1) Aggregate the number of households in each range in Table B19001 for the selected areas. Calculate a cumulative total and a cumulative percent. Your result should look like this:

Range start	Range end	Households	Cumulative Total	Cumulative Percent	
<b>TOTAL</b>		2,068			
(2,500)	9,999	186	186	9.0	Bottom Range
10,000	14,999	78	264	12.8	
15,000	19,999	98	362	17.5	
20,000	24,999	287	649	31.4	
25,000	29,999	142	791	38.3	<i>p_lower</i>
30,000	34,999	90	881	42.6	
35,000	39,999	107	988	47.8	
40,000	44,999	104	1,092	52.8	Mid-Range
45,000	49,999	178	1,270	61.4	
50,000	59,999	106	1,376	66.6	<i>p_upper</i>
60,000	74,999	177	1,553	75.1	
75,000	99,999	262	1,815	87.8	
100,000	124,999	77	1,892	91.5	
125,000	149,999	100	1,992	96.4	
150,000	199,999	58	2,050	99.2	
200,000	250,001	18	2,068	100.0	Top Range

- (2) Determine the mid-point:

$$\text{Total Households} \div 2 = 2068 \div 2 = 1034$$

The 1,034<sup>th</sup> household is the mid-point.

- (3) Determine the range holding the mid-point:

Using the running total, determine that the 1,034<sup>th</sup> household is one of the households in the \$40,000 to \$44,999 range – highlighted above. This is the mid-range. [Note: If the median is in the bottom or the top range, see the section “Top and Bottom Ranges” below for additional information.]

- (4) How many of the households (HH) in the mid-range are needed to reach the mid-point?  
(Mid-point) - (total HH in the previous smaller range, \$35,000 to \$39,999) =  
1034 - 988 = 46.  
So we need 46 more households in the mid-range to equal the mid-point.
- (5) What is the proportion of the number of households in the \$40,000 to \$44,999 range that would be needed to get to the mid-point?  
(Result from Step 4) ÷ (number of HH in mid-range) = 46 ÷ 104 = 0.4423
- (6) Apply this proportion to width of the mid-range dollar figure:
- The width of \$40,000 to \$44,999 range is \$5,000.
  - The proportion \* the width of the range = 0.4423 \* \$5000 = \$2,212
- (7) Calculate the new median:  
Beginning of the mid-range + the dollar result from step 6 = \$40,000 + \$2,212 = \$42,212

The calculated median for the aggregated geography is \$42,212. This figure isn't exact – it assumes an even distribution of incomes within the range and it assumes that there are no duplicate values for the households in the range.

### How to Calculate a Margin of Error for a Re-calculated Median:

Using our example above, let's calculate the margin of error (MOE) for \$42,212.

- (A) Approximate the standard error (SE) of a 50 percent proportion.

$$SE(50\text{ percent}) = DF * \sqrt{\frac{99}{B} * 50^2}$$

Where:

*B* = the total households (base total)

*DF* = the Design Factor from the PUMS Accuracy Statement, which is 1.5 for Income.

Example:

$$SE(50\text{ percent}) = 1.5 * \text{Square Root of } ((99/2068) * 50^2) = 16.4$$

- (B) Subtract and add the standard error determined in step (A) to 50 percent.

$$p_{\text{lower}} = 50 - SE(50\text{ percent}) = 33.6$$

$$p_{\text{upper}} = 50 + SE(50\text{ percent}) = 66.4$$

- (C) Determine the range(s) in the distribution that contain  $p_{\text{lower}}$  and  $p_{\text{upper}}$ . If  $p_{\text{lower}}$  and  $p_{\text{upper}}$  fall in different ranges, follow step (D). If  $p_{\text{lower}}$  and  $p_{\text{upper}}$  fall in the same range, go to step (E).

[Note: If the median is in the lowest or the highest range, see the section "Top and Bottom Ranges" below for additional information.]

- (D) If  $p_{\text{lower}}$  and  $p_{\text{upper}}$  fall into different ranges, do the following:
- Define A1 as the smallest value in that range.
  - Define A2 as the smallest value in the next (higher) range.
  - Define C1 as the cumulative percent of units strictly less than A1.

- Define C2 as the cumulative percent of units strictly less than A2.

Use the following formulas to approximate the lower and upper bounds for a confidence interval about the median:

$$LowerBound = \left[ \frac{p\_lower - C1}{C2 - C1} \right] * (A2 - A1) + A1$$

$$UpperBound = \left[ \frac{p\_upper - C1}{C2 - C1} \right] * (A2 - A1) + A1$$

- For the category containing p\_lower, define A1, A2, C1, and C2. Use these values and the formula above to obtain the LowerBound.

Example:

$$A1 = 25,000 \qquad C1 = 31.4$$

$$A2 = 30,000 \qquad C2 = 38.3$$

$$LowerBound = [ (33.6 - 31.4) / (38.3 - 31.4) ] * (30,000 - 25,000) + 25,000 = 26,594$$

- For the category containing p\_upper, define new values for A1, A2, C1, and C2 and use these values and the formula above to obtain the UpperBound.

Example:

$$A1 = 50,000 \qquad C1 = 61.4$$

$$A2 = 60,000 \qquad C2 = 66.6$$

$$UpperBound = [ (66.4 - 61.4) / (66.6 - 61.4) ] * (60,000 - 50,000) + 50,000 = 59,615$$

(E) If p\_lower and p\_upper fall into the same range, it is only necessary to identify one set of A1, A2, C1 and C2 (see Step D) for the single range. The same values are used in calculations for both the LowerBound and the UpperBound.

(F) Use the LowerBound and UpperBound estimated in steps (D) or (E) to approximate the standard error (SE) of the median.

$$SE(median) = 0.5 * (UpperBound - LowerBound)$$

Example:

$$SE(median) = 0.5 * (59,615 - 26,594) = 16,511$$

(G) Calculate the margin of error at the 90% confidence interval:

$$Margin\ of\ error = +/- (1.645 * SE(median))$$

Example:

$$Margin\ of\ error = +/- (1.645 * 16,511) = +/- \$27,161$$

A "Median Calculator" is available in Excel to calculate the new median and the margin of error for that estimate. If the Excel file was not sent with this article, contact the California State Data Center at 916-323-4086 to request the file.

## Comparing MOE from Formulas versus MOE from Actual Census Returns:

When analyzing data using an MOE that have been calculated from a formula, the user should be aware that the formulas are actually approximations that overstate the MOE compared to the more precise methods based on the actual survey returns that the Census Bureau uses. Therefore, the calculated MOEs will be higher, or more conservative, than those found in published tabulations for similarly-sized areas. This knowledge may affect the levels of error you are willing to accept when making a decision on whether or not to use a particular figure.

To demonstrate this, the data we used in our example above are actual values for the Palermo Census Designated Place (CDP), a small CDP in Butte County. The table below provides the published median and MOE from the Census Bureau using survey returns and the calculated data using the formulas for a small area like Palermo CDP and a large area like Los Angeles County.

	Published by Census Bureau	Calculated from formula	Percent Difference
<b>Palermo CDP</b>			
Total Households	2,067		
Median Household Income	\$41,477	\$42,212	1.8%
Margin of Error	+/- \$4,972	+/- \$27,161	446%
<b>Los Angeles County</b>			
Total Households	3,178,266		
Median Household Income	\$54,828	\$55,073	0.5%
Margin of Error	+/- \$244	+/- \$879	260%

As you can see, although the method for calculating the median provides a comparable figure, the calculated MOE is much higher than the published MOE.

## Top and Bottom Ranges:

If p\_lower or p\_upper are in the lowest or highest range, you will need to find a bottom or top value to estimate the width of the range. This is the value where the “Less than \$10,000” range would start or the “\$250,000 or more” range would top out. The Census Bureau has made recommendations for these figures and they are called “jam values”. They can be found in Appendix C of the Technical documentation for the specific ACS file. This documentation can be found on the Census Bureau’s ACS technical documentation website at <http://www.census.gov/programs-surveys/acs/technical-documentation.html>. For our example, median household income is in Table B19013 and the jam values are:

Table ID	Upper Limit	Upper Limit Meaning	Lower Limit	Lower Limit Meaning
B19013	250,001	250,000+	-2,499	-2,500

Prepared on January 26, 2011, and revised on April 1, 2016, by:

California State Data Center  
 Demographic Research Unit  
 Department of Finance  
 915 L Street, 8<sup>th</sup> Floor  
 Sacramento, CA 95814  
 Phone: 916-323-4086