Tech Talk: Using the Margin of Error – from the July 2011 CSDC Network News

The data provided in the American Community Survey (ACS) are based on a sample of households in a specific geographic area. Data based on a sample are estimates of the true value you would have gotten if you had asked the question of every person in the geographic area. The more people you ask within an area, the closer the estimate will be to the true value for the total population. Probability theory is used to work out a predication of how confident users can be that the statistics from our sample are representative of the entire population. That prediction is expressed as the margin of error.

For the ACS, the Census Bureau provides the margin of error (MOE) at the 90% confidence interval (CI) with every estimate. This MOE is a calculated range that the Census Bureau is 90% sure that the true value is within the range of the estimate plus and minus the MOE. For example, an estimate of 5,400 has a MOE of 200, so we can be 90% sure that the actual value is between 5,200 and 5,600.

So What Does the MOE Tell Me?

We often speak of the "reliability" of an estimate. Or we say the sample size is too small to produce reliable estimates. But how do we measure the reliability of an estimate?

Just looking at the MOE by itself is not enough. An MOE of 1,800 may be good or bad. You must compare the size of the MOE to the size of the estimate. The easiest way is to compare the MOE to the estimate by expressing the MOE as a percent of the estimate. For purposes of this discussion, let's call that percentage a "relative MOE". So if the MOE of 1,800 is provided for an estimate of 3,600, the MOE is 50 percent of the estimate. But if it is for an estimate of 180,000, the MOE is 1 percent of the estimate.

> Side Note: Strictly speaking in statistical theory, you would calculate the coefficient of variance (CV) to measure reliability. The coefficient of variation expresses the standard error as a percentage of the sample mean. However, this article is targeted to the common data user who has only the estimate and the MOE and does not have a degree in statistics. So we are going to use the relative MOE as our test of reliability.

A problem arises when the estimate for a variable is zero. In this case, the MOE can be misleading because the range for the confidence interval will contain both positive and negative values. The Census Bureau provides a uniform value that applies in all cases within a State when the variable of interest is zero. For California, this value is 132.

But what is the range of the relative MOE for an estimate to be reliable or usable? Unfortunately the answer is the ever vague - "it depends".

A low percentage indicates a more reliable estimate. But there are no steadfast rules as to what constitutes a reliable estimate. The Census ACS Compass products suggest that users should be cautious about using an estimate if the coefficient of variance (CV) is greater than 15 percent which translates to a relative MOE of about 25 percent. However, individual data users must determine what a reasonable MOE threshold is specific to their needs, risk-outcomes, and use. Some users may be forced to use an estimate with a relative MOE of 50 percent when better data are not available and the information is required. Ultimately is up to the user to assess the quality of each ACS estimate and decide whether a particular estimate is suitable to his/her needs.

When the California SDC publishes our own research reports from the ACS, if the estimate has a relative MOE of 50 percent or higher, we will not publish it. Estimates with a relative MOE between 33 to 49 percent are italicized and we generally warn users that they should use caution with the estimate.

The question of reasonableness comes into play based on the question of what is at stake - lives, millions of dollars, or just informational knowledge. The end user must evaluate what the consequences would be if the variable numbers were different. Most NIH, CDC efficacy reports usually cap things at 10 percent CV (MOE of 16 percent) for instance, but that rate reduces severely the data from the ACS that is acceptable, but, of course, they are talking about disease and health.

A very good case study on using and interpreting the ACS data for small areas can be found in the *ACS Compass Handbook: What Users of Data for Rural Areas Need to Know* https://www.census.gov/content/dam/Census/library/publications/2009/acs/ACSRuralAreaHandbook.pdf

Using the MOE in Programs and Grants

Many times the ACS data is used to qualify for specific programs or on grant applications, in these cases, the applicant should contact their grant program officer about how to use the MOE and how it should be provided and applied. There may be a way to use the MOE to the applicant's advantage. For example in program A, the median household income of the applicant city must be below the median household income of the state. Here's the data:

	Med HHinc	MOE	90% CI
City	$61,254	+/-1,823	$59,431 – $63,077
State	$60,392	+/-154	$60,238 – $60,546

In this example, the estimate for median household income for the city is higher than the state's. However, if we use the margin of error and the confidence interval, one could argue that the state's income could be as high as $60,546 (the high end of the confidence interval range) and the city's could be as low as $59.431 (the low end of the CI range). Therefore, it could be argued statistically that the city qualifies for the program.

Comparing the MOE of the 2000 Census with ACS

While many data users were initially dismayed by the large MOEs present in the first 5-year ACS tract level data released last year, they forget that the 2000 census long form is also based on a sample and is subject to error. This has historically been overlooked by data users, not because it was not an issue, but rather because the Census Bureau made it difficult to determine the MOE for census data.

It is true that error measurements for the decennial census are lower than the ACS but, for many small geographies and characteristics, the MOE of census data are still far too high and too much faith may have been placed in decennial census tract data (much less Block Group level data). Ironically, the increased awareness of sampling variability due to the ACS might actually be a positive step towards understanding the inherent limitations of sampled data.