

### 3. SAMPLING ERROR IN THE ACS

Because the American Community Survey (ACS) is based on a sample, rather than all housing units and people, ACS estimates have a degree of uncertainty associated with them, known as sampling error. In general, the larger the sample, the smaller the level of sampling error. To help users understand the impact of sampling error on data reliability, the U.S. Census Bureau provides a “margin of error” (MOE) for each published ACS estimate. The MOE, combined with the ACS estimate, give users a range of values within which the actual, “real-world” value is likely to fall.

*TIP: Sometimes ACS data users ignore the issue of sampling variability, which can be problematic when analyzing differences across small area estimates. Rather than considering that the estimates are derived from a complex sample survey, the estimates have sometimes been treated as values for the population. Data users should be careful in drawing conclusions about small differences between two ACS estimates because they may not be statistically different.*

By presenting the MOE alongside the estimates, users can more easily determine whether differences they observe over time and space are statistically significant or within the bounds of random variation. The Census Bureau uses a 90 percent confidence level to determine the MOE in the published tabulations. Depending on the application, a user may wish to

increase the confidence level to 95 percent or 99 percent to conduct a more rigorous test of significant differences.

#### Tests of Statistical Significance for Aggregate ACS Estimates

The Census Bureau has produced a Statistical Testing Tool to make it easier for ACS data users to conduct tests of statistical significance when comparing ACS estimates (see Figure 3.1).<sup>26</sup>

The Statistical Testing Tool consists of an Excel spreadsheet that will automatically calculate statistical significance when data users are comparing two or more ACS estimates. Data users simply need to insert the ACS estimate(s) and associated MOE(s) into the correct columns and cells in the spreadsheet. The results are calculated automatically. The result “Yes” indicates that estimates are statistically different and the result “No” indicates the estimates are not statistically different.<sup>27</sup>

<sup>26</sup> U.S. Census Bureau, American Community Survey (ACS), Statistical Testing Tool, <[www.census.gov/programs-surveys/acs/guidance/statistical-testing-tool.html](http://www.census.gov/programs-surveys/acs/guidance/statistical-testing-tool.html)>.

<sup>27</sup> This tool only conducts statistical testing on the estimates keyed in by the data user for comparison within the spreadsheet and it does not adjust the MOE when making multiple comparisons, nor incorporate a Bonferroni correction or any other method in the results of the statistical testing.

Figure 3.1. Statistical Testing Tool

Label	First Estimate	First Margin of Error (MOE)	Second Estimate	Second Margin of Error (MOE)	Statistically Different?
1 Place your Geography of statistical variable keyword	48759	+/-2,395	28021	*****	Yes
2 in the gray column here. This text is here as an	6.2	+/-0.1	6	+/-0.3	No
3 example. You may delete the text before using the	22147046	+/-234,312	17997953	+/-199,112	Yes
4 spreadsheet.	-	**	27047	+/-230	N/A

Source: U.S. Census Bureau, American Community Survey (ACS), Statistical Testing Tool, <[www.census.gov/programs-surveys/acs/guidance/statistical-testing-tool.html](http://www.census.gov/programs-surveys/acs/guidance/statistical-testing-tool.html)>.

## Calculating Margins of Error for Custom (User-Derived) Estimates

In some cases, researchers will need to construct custom ACS estimates by combining data across multiple geographic areas or population subgroups, or it may be necessary to derive a new percentage, proportion, or ratio from published ACS data. For example, one way to address the issue of unreliable estimates for individual census tracts or block groups is to aggregate geographic areas, yielding larger samples and estimates that are more reliable. In such cases, additional calculations are needed to produce MOEs and to conduct tests of statistical significance for the derived estimates. The section on “Calculating Measures of Error for Derived Estimates” in the Census Bureau’s handbook on *Understanding and Using American Community Survey Data: What All Data Users Need to Know* provides detailed instructions on how to make these calculations.<sup>28</sup> Each ACS data release is also accompanied by “Accuracy of the Data” documentation that includes formulas for calculating MOEs (see Figure 3.2).<sup>29</sup>

<sup>28</sup> U.S. Census Bureau, *Understanding and Using American Community Survey Data: What All Data Users Need to Know*, <[www.census.gov/programs-surveys/acs/guidance/handbooks/general.html](http://www.census.gov/programs-surveys/acs/guidance/handbooks/general.html)>.

<sup>29</sup> U.S. Census Bureau, American Community Survey (ACS), Code Lists, Definitions, and Accuracy, <[www.census.gov/programs-surveys/acs/technical-documentation/code-lists.html](http://www.census.gov/programs-surveys/acs/technical-documentation/code-lists.html)>.

Users should note that some of the general formulas for calculating MOEs for derived estimates produce approximations rather than exact MOEs. Advanced users may be interested in the Variance Replicate Tables, first released for the 2010–2014 ACS 5-year estimates in July 2016.<sup>30</sup> These augmented ACS Detailed Tables include sets of 80 replicate estimates, which allow users to calculate MOEs for derived estimates using the same methods that are used to produce the published MOEs in the pre-made tables from the Census Bureau. These methods incorporate the covariance between estimates that the approximation formulas in the “Accuracy of the Data” document do not include.

The Variance Replicate Tables are available for a subset of the 5-year Detailed Tables for 11 geographic summary levels including the nation, states, counties, census tracts, and block groups. These tables are released on an annual basis, shortly after the release of the standard 5-year data products.

Variance Replicate Tables documentation, including lists of tables and summary levels, is available on the Census Bureau’s Web site.<sup>31</sup>

<sup>30</sup> U.S. Census Bureau, American Community Survey (ACS), Variance Replicate Tables, <[www.census.gov/programs-surveys/acs/data/variance-tables.html](http://www.census.gov/programs-surveys/acs/data/variance-tables.html)>.

<sup>31</sup> U.S. Census Bureau, American Community Survey (ACS), Variance Replicate Tables Documentation, <[www.census.gov/programs-surveys/acs/technical-documentation/variance-tables.html](http://www.census.gov/programs-surveys/acs/technical-documentation/variance-tables.html)>.

Figure 3.2. Code Lists, Definitions, and Accuracy

The screenshot shows the 'Code Lists, Definitions, and Accuracy' page on the U.S. Census Bureau website. The page features a navigation menu on the left with categories like 'About the Survey', 'Data', and 'Technical Documentation'. The main content area is titled 'Code Lists, Definitions, and Accuracy' and includes a year selector (2016-2012) and several sections: 'Code Lists', 'Subject Definitions', 'Group Quarters Definitions', 'Comparison Guidance', 'Instructions for Applying Statistical Testing', 'Statistical Testing Tool', and 'Accuracy of the Data'. The 'Accuracy of the Data' section is highlighted with a red box and lists links for '2012-2016 ACS Multiyear Accuracy (US)', '2012-2016 PRCS Multiyear Accuracy (Puerto Rico)', '2016 ACS 1-year Accuracy of the Data (US)', and '2016 PRCS 1-year Accuracy of the Data (Puerto Rico)'. The source information at the bottom of the screenshot is: 'Source: U.S. Census Bureau, American Community Survey (ACS), Code Lists, Definitions, and Accuracy, <[www.census.gov/programs-surveys/acs/technical-documentation/code-lists.html](http://www.census.gov/programs-surveys/acs/technical-documentation/code-lists.html)>.'

---

## Calculating Standard Errors for ACS PUMS Estimates

Researchers using the microdata files need to calculate their own estimates of standard error due to sampling, using either a generalized variance function (generalized standard errors) or by using the replicate weights (direct standard errors). The Census Bureau presents both approaches in their Accuracy of the Public Use Microdata Sample documentation.<sup>32</sup>

The Census Bureau notes that, “Direct standard errors will often be more accurate than generalized standard errors, although they may be more inconvenient for some users to calculate. The advantage of using replicate weights is that a single formula is

---

<sup>32</sup> U.S. Census Bureau, American Community Survey (ACS), PUMS Technical Documentation, <[www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html](http://www.census.gov/programs-surveys/acs/technical-documentation/pums/documentation.html)>.

used to calculate the standard error of many types of estimates.”<sup>33</sup>

With generalized standard errors, “design factors” are applied to “reflect the effects of the actual sample design and estimation procedures used for the ACS.”<sup>34</sup> There is a tendency for the formula for generalized standard errors to estimate higher standard errors than the replicate weights method. This means that by using the generalized formula instead of the replicate weights, one may be less likely to find statistically significant differences where they actually do exist.

---

<sup>33</sup> U.S. Census Bureau, Public Use Microdata Sample (PUMS), Accuracy of the Data, p. 16, 2016.

<sup>34</sup> U.S. Census Bureau, Public Use Microdata Sample (PUMS), Accuracy of the Data, p. 18, 2016.