

8. CALCULATING MEASURES OF ERROR FOR DERIVED ESTIMATES

The U.S. Census Bureau publishes a wide range of American Community Survey (ACS) estimates through the data products provided in data.census.gov and the ACS Summary File. However, for some applications, data users may still need to construct custom ACS estimates and their associated margins of error (MOEs). For example, users may want to aggregate data across geographic areas or population subgroups, or may need to derive proportions, ratios, and percent change measures that are not provided in published ACS data products.

TIP: Data users can also aggregate data across geographic areas—or across population subgroups—to produce more reliable estimates in cases where the underlying component areas have estimates with relatively large standard errors (SEs).

When such derived estimates are generated, the user must calculate the associated MOEs. As described in the section on “Understanding Error and Determining Statistical Significance,” MOEs are needed to calculate SEs and coefficients of variation (CVs) and to determine statistical significance. The steps outlined in the sections below illustrate how to calculate the MOEs for derived counts, proportions, percentages, ratios, and percent change, which can then be translated into SEs and CVs.

Calculating Measures of Error for Aggregated Count Data

Aggregating Data Across Geographic Areas

For some applications, data users may want to know the number of people with certain characteristics within a particular geographic region that is not included as a standard geographic area in ACS products. For example, a user may be interested in the number of never-married females within a tri-county area. The example below shows how to calculate the MOE, SE, and CV for such a derived estimate.

To calculate the MOE for aggregated count data:

1. Obtain the MOE of each component estimate.
2. Square the MOE of each component estimate.
3. Sum the squared MOEs.
4. Take the square root of the sum of the squared MOEs.

The result is the MOE for the aggregated count. Algebraically, the MOE for the aggregated count is calculated as:

$$\text{MOE}(\hat{X}_1 + \hat{X}_2 + \dots + \hat{X}_n) = \pm \sqrt{[\text{MOE}(\hat{X}_1)]^2 + [\text{MOE}(\hat{X}_2)]^2 + \dots + [\text{MOE}(\hat{X}_n)]^2} \tag{1}$$

The example below shows how to calculate the MOE and SE for the estimated number of never-married females living in the three Virginia counties/independent cities that border Washington, DC (Fairfax and Arlington counties, Alexandria City) from the 2015 ACS 1-year estimates as shown in Table 8.1.

Table 8.1. Data for Example 1 From Three Virginia Counties/Independent Cities: 2015		
Characteristic	Estimate	MOE
Never-married females living in Fairfax County (Component 1)	135,173	±3,860
Never-married females living in Arlington County (Component 2)	43,104	±2,642
Never-married females living in Alexandria City (Component 3)	24,842	±1,957
Source: U.S. Census Bureau, data.census.gov, Table B12001: “Sex by Marital Status for the Population 15 Years and Over.”		

The aggregate estimate is:

$$\begin{aligned}\hat{X} &= \hat{X}_{\text{Fairfax}} + \hat{X}_{\text{Arlington}} + \hat{X}_{\text{Alexandria}} = \\ &135,173 + 43,104 + 24,842 = 203,119\end{aligned}$$

Obtain MOEs of the component estimates:

$$\begin{aligned}\text{MOE (Fairfax)} &= \pm 3,860 \\ \text{MOE (Arlington)} &= \pm 2,642 \\ \text{MOE (Alexandria)} &= \pm 1,957\end{aligned}$$

Using formula (1), calculate the MOE for the aggregate estimate:

$$\begin{aligned}\text{MOE}(\hat{X}_1 + \hat{X}_2 + \hat{X}_3) &= \pm \sqrt{(3,860)^2 + (2,642)^2 + (1,957)^2} = \\ &\pm \sqrt{25,709,613} = \pm 5,070\end{aligned}$$

Thus, the derived estimate of the number of never-married females living in the three Virginia counties/independent cities that border Washington, DC, is 203,119, and the MOE for the estimate is $\pm 5,070$.

The SE of this derived estimate can be calculated from the SEs of the component estimates as follows:

1. Calculate the SE of each component estimate from its MOE using:

$$\text{SE} = \frac{\text{MOE}}{1.645} \quad (2)$$

$$\begin{aligned}\text{SE (Fairfax)} &= 3,860 / 1.645 = 2,347 \\ \text{SE (Arlington)} &= 2,642 / 1.645 = 1,606 \\ \text{SE (Alexandria)} &= 1,957 / 1.645 = 1,190\end{aligned}$$

2. Calculate the SE of the aggregate estimate:

$$\text{SE}(\hat{X}_1 + \hat{X}_2 + \dots + \hat{X}_n) = \sqrt{[\text{SE}(\hat{X}_1)]^2 + [\text{SE}(\hat{X}_2)]^2 + \dots + [\text{SE}(\hat{X}_n)]^2} \quad (3)$$

With the three component estimates in this example, this becomes:

$$\begin{aligned}\text{SE}(\hat{X}_1 + \hat{X}_2 + \hat{X}_3) &= \sqrt{(2,347)^2 + (1,606)^2 + (1,190)^2} = \\ &\sqrt{9,500,878} = 3,082\end{aligned}$$

To assess the reliability of this derived estimate, users may find it helpful to calculate the CV as follows:

$$\begin{aligned}\text{CV} &= \frac{\text{SE}}{\text{Estimate}} * 100 \\ \text{CV} &= \frac{3,082}{203,119} * 100 = 0.015 * 100 = 1.5\%\end{aligned} \quad (4)$$

This CV indicates that the sampling error of this estimate is very small relative to the estimate itself, so the number of never-married females residing in the Virginia tri-county area bordering Washington, DC, can be considered a very reliable estimate.

However, users should note that this method for calculating the MOE and SE for aggregated count data is an approximation, and caution is warranted because this method does not consider the correlation or covariance between the component estimates. This method may result in an overestimate or underestimate of the derived estimate's SE depending on whether the component estimates are highly correlated in either a positive or negative direction. As a result, the approximated SE may not match the result from a direct calculation of the SE that does include a measure of covariance, such as the following:

$$SE(\hat{X}_1 \pm \hat{X}_2) = \sqrt{[SE(\hat{X}_1)]^2 + [SE(\hat{X}_2)]^2 \pm 2cov(\hat{X}_1, \hat{X}_2)} \quad (5)$$

Data users should also be aware that as the number of estimates involved in a sum or difference increases, the results of the approximation formula become increasingly different from the SE derived directly from the ACS microdata. Users are encouraged to work with the fewest number of estimates possible. If there are estimates involved in a sum that are controlled in the weighting, then the approximate SE can be increasingly different. Several examples are provided to demonstrate issues associated with approximating the SEs when summing large numbers of estimates together in the latest ACS Worked Examples for Approximating Margins of Error document.⁶⁸

Aggregating Data Across Population Subgroups

For some applications, data users may wish to combine data across population subgroups, especially in ACS tables where some groups have low counts and large MOEs. Before aggregating categories in a Detailed Table to create a derived estimate, users should check to make sure there is not a collapsed version of the same table already available in data.census.gov or the ACS Summary File. The MOEs in the published, collapsed tables will be more accurate than those users can approximate using the methods described in this section.

The example below illustrates the results from aggregating household income categories from a Detailed Table in data.census.gov for Loudoun County, Virginia, from the 2015 ACS 1-year estimates. Income categories are organized into three subgroups for the purpose of this example.

⁶⁸ U.S. Census Bureau, American Community Survey (ACS), Worked Examples for Approximating Margins of Error, <www.census.gov/programs-surveys/acs/technical-documentation/code-lists.html>.

Table 8.2. Data for Example 2 From Loudoun County, Virginia: 2015

Household income category	Subgroup #	Estimate	MOE	SE	CV
Less than \$10,000	-	2,163	±812	494	22.8
\$10,000 to \$14,999	-	1,178	±504	306	26.0
\$15,000 to \$19,999	1	1,502	±743	452	30.1
\$20,000 to \$24,999	1	1,995	±722	439	22.0
\$25,000 to \$29,999	2	1,756	±685	416	23.7
\$30,000 to \$34,999	2	1,781	±631	384	21.5
\$35,000 to \$39,999	3	2,708	±1,007	612	22.6
\$40,000 to \$44,999	3	1,981	±647	393	19.9
\$45,000 to \$49,999	3	2,581	±996	605	23.5
\$50,000 to \$59,999	-	6,590	±1,109	674	10.2
\$60,000 to \$74,999	-	6,861	±1,288	783	11.4
\$75,000 to \$99,999	-	14,391	±1,810	1,100	7.6
\$100,000 to \$124,999	-	14,790	±1,944	1,182	8.0
\$125,000 to \$149,999	-	12,735	±1,341	815	6.4
\$150,000 to \$199,999	-	18,167	±1,890	1,149	6.3
\$200,000 or more	-	29,380	±2,053	1,248	4.2

Source: U.S. Census Bureau, data.census.gov, Table B19001: "Household Income in the Past 12 Months (In 2015 Inflation-Adjusted Dollars)."

After calculating the SEs and CVs shown in Table 8.2, a data user might want to combine some of the household income categories in the lower end of the income distribution to reduce the CVs and create more reliable estimates. For example, the income categories in Subgroup 1 (\$15,000 to \$19,999 and \$20,000 to \$24,999) could be combined to form the first new subgroup, the two in Subgroup 2 (\$25,000 to \$29,999 and \$30,000 to \$34,999) to form a second new subgroup, and the three in Subgroup 3 (\$35,000 to \$39,999, \$40,000 to \$44,999, and \$45,000 to \$49,999) to form a third new subgroup. Following the same steps used to combine data across three geographic areas in Example 1, derived estimates and MOEs for the three new subgroups could be calculated as follows:

1. Aggregate counts to form new subgroup estimates:

- Subgroup 1 (\$15,000 to \$24,999) = 1,502 + 1,995 = 3,497
- Subgroup 2 (\$25,000 to \$34,999) = 1,756 + 1,781 = 3,537
- Subgroup 3 (\$35,000 to \$49,999) = 2,708 + 1,981 + 2,581 = 7,270

2. Using formula (1), calculate MOEs for each new subgroup estimate:

$$\text{MOE (Subgroup 1)} = \pm\sqrt{(743)^2 + (722)^2} = \pm\sqrt{1,073,333} = \pm1,036$$

$$\text{MOE (Subgroup 2)} = \pm\sqrt{(685)^2 + (631)^2} = \pm\sqrt{867,386} = \pm 931$$

$$\begin{aligned}\text{MOE (Subgroup 3)} &= \pm\sqrt{(1,007)^2 + (647)^2 + (996)^2} = \\ &\pm\sqrt{2,424,674} = \pm 1,557\end{aligned}$$

3. Calculate the SE of each new subgroup estimate using formula (2):

- SE (Subgroup 1) = 1,036 / 1.645 = 630
- SE (Subgroup 2) = 931 / 1.645 = 566
- SE (Subgroup 3) = 1,557 / 1.645 = 947

4. Using formula (4), calculate the CV for each new subgroup estimate:

- CV (Subgroup 1) = (630 / 3,497) * 100 = (0.180) * 100 = 18.0 percent
- CV (Subgroup 2) = (566 / 3,537) * 100 = (0.160) * 100 = 16.0 percent
- CV (Subgroup 3) = (947 / 7,270) * 100 = (0.130) * 100 = 13.0 percent

Aggregating across income categories increases the reliability of the new subgroup estimates relative to the original component estimates. For example, the CV for the \$15,000 to \$19,999 category is 30 percent. When this category is combined with the \$20,000 to \$24,999 category, the CV for the derived estimate for the new Subgroup 1 drops to 18 percent—lower than the original CV for this category by itself. Similarly, the CVs for the component estimates of Subgroup 3 range from 20 percent to 24 percent, while the CV for the derived estimate for Subgroup 3 (\$35,000 to \$49,999) is only 13 percent.

Data.census.gov includes a collapsed income distribution within Demographic Profile DP03 that makes it possible to assess the difference in results from the approximation method illustrated above versus a direct estimation method that accounts for covariance between the component estimates of the subgroups. The relevant portion of this data.census.gov table, from the 2015 ACS 1-year estimates, is shown in Table 8.3 below.

Table 8.3. Data From Demographic Profile DP03 for Loudoun County, Virginia: 2015

Household income category	Subgroup #	Estimate	MOE	SE	CV
\$15,000 to \$24,999	1	3,497	±1,037	630	18.0
\$25,000 to \$34,999	2	3,537	±973	591	16.7
\$35,000 to \$49,999	3	7,270	±1,554	945	13.0

Source: U.S. Census Bureau, data.census.gov, Table DP03: "Selected Economic Characteristics."

While the MOEs, SEs, and CVs for new Subgroup 1 and Subgroup 3 are almost identical to those derived from the approximation method, this is not the case for new Subgroup 2. The user-derived MOE is ±931 compared with the published MOE of ±973, and the derived CV is only 16.0 percent compared with a CV of 16.7 percent based on the published MOE for Subgroup 2. The approximation method slightly underestimates the MOE and SE for Subgroup 2 because it does not account for some covariance between the two component estimates. Although the differences between the user-derived and published MOE's are small in this example, they can vary substantially in other cases, particularly for linear combinations of multiple estimates. Examples illustrating this problem are provided in the ACS Worked Examples for Approximating Margins of Error document (see section on "Issues with Approximating the Standard Error of Linear Combinations of Multiple Estimates").⁶⁹

⁶⁹ U.S. Census Bureau, American Community Survey (ACS), Worked Examples for Approximating Margins of Error, <www.census.gov/programs-surveys/acs/technical-documentation/code-lists.html>.

Calculating Measures of Error for User-Derived Proportions and Percentages

When data users create derived estimates for aggregated count data, they are often interested in calculating additional measures such as proportions and percentages. For example, the user who calculated the *number* of never-married females in the tri-county area in northern Virginia might also want to know what *proportion* of all females in this region have never been married. With proportions, the numerator is a subset of the denominator. In this case, the number of never-married females is a subset of all females. This derived proportion would be calculated as the number of never-married females aged 15 and older divided by the total number of females aged 15 and older. This proportion could also be converted to a percentage by multiplying by 100. The 2015 ACS 1-year estimates of the total number of females aged 15 and older in Fairfax County, Arlington County, and Alexandria City and their respective MOEs are shown in Table 8.4.

Table 8.4. Data for Example 3 From Three Virginia Counties/Independent Cities: 2015		
Characteristic	Estimate	MOE
Total females aged 15 and older living in Fairfax County (Component 1)	466,037	±391
Total females aged 15 and older living in Arlington County (Component 2)	97,360	±572
Total females aged 15 and older living in Alexandria City (Component 3)	67,101	±459
Source: U.S. Census Bureau, data.census.gov, Table B12001: "Sex by Marital Status for the Population 15 Years and Over."		

Using the data for never-married females from Example 1 in this section with the data in Table 8.4 yields the following:

Proportion = (Never-married females / Total females)

(203,119 / (466,037 + 97,360 + 67,101)) = (203,119 / 630,498) = 0.322

To calculate the MOE of this proportion, we need the MOEs of the numerator (203,119) and the denominator (630,498). We already calculated the MOE of the number of never-married females in Example 1 in this section as ±5,070. Using formula (1), the MOE of the denominator is calculated as:

$MOE(\hat{X}_1 + \hat{X}_2 + \hat{X}_3) = \pm\sqrt{(391)^2 + (572)^2 + (459)^2} = \pm\sqrt{690,746} = \pm 831$

If we define the proportion as $\hat{P} = \hat{X}/\hat{Y}$, then the MOE of this proportion is approximated as:

$MOE(\hat{P}) = \frac{1}{\hat{Y}} \sqrt{[MOE(\hat{X})]^2 - (\hat{P}^2 * [MOE(\hat{Y})]^2)}$ (6)

However, the Census Bureau provides ACS estimates as percentages rather than proportions in data.census.gov. For this example, the proportion of never-married females (0.322) can be converted to the percentage of never-married females (32.2%) by multiplying by 100. If we define this percentage as $\hat{Q}$, then the

$MOE(\hat{Q}) = 100\% \times MOE(\hat{P})$

Substituting the estimates for the numerator and denominator, and their respective MOEs in formula (6), and multiplying by 100 yields:

$MOE(\hat{Q}) = 100 * \left[\frac{1}{630,498} \sqrt{(5,070)^2 - [(0.322)^2 * (831)^2]} \right] = 0.8\%$

Users should note that if the value under the square root is negative, then substitute a “plus” for the “minus” sign under the square root in formula (6). This modified formula is the same as the formula for the MOE of a ratio, which will be discussed in the next section.

Using formula (2), the SE of the percentage estimate of never-married females is $0.80 / 1.645 = 0.488$. CVs can also be calculated for derived percentages using formula (4) and are interpreted in the same way as for derived count estimates. In this example, the CV for the percentage of never-married females is $(0.488 / 32.2) * 100 = 1.5$ percent, indicating that this is a very reliable estimate with a standard error that is much smaller than the estimated percentage.

Calculating Measures of Error for Derived Ratios

Ratios are used to compare two estimates where the numerator is not a subset of the denominator. For example, the data user in Example 1 may want to compare the number of never-married males in the tri-county area to the number of never-married females. To do this, the user must first obtain the component estimates and MOEs for males, and calculate the derived estimate and MOE of the number of never-married males in the tri-county area. These data, from the 2015 ACS 1-year estimates, are shown in Table 8.5 below.

Table 8.5. Data for Example 4 From Three Virginia Counties/Independent Cities: 2015		
Characteristic	Estimate	MOE
Never-married males living in Fairfax County (Component 1).....	156,720	±4,222
Never-married males living in Arlington County (Component 2).....	44,613	±2,819
Never-married males living in Alexandria City (Component 3).....	25,507	±2,259

Source: U.S. Census Bureau, data.census.gov, Table B12001: “Sex by Marital Status for the Population 15 Years and Over.”

Using the derived estimate and MOE calculated in Example 1 in this section, and the data from Table 8.5, the ratio would be calculated as:

$$\text{Ratio} = (\text{Never-married males} / \text{Never-married females})$$

$$\text{Ratio} = ((156,720 + 44,613 + 25,507) / 203,119) = (226,840 / 203,119) = 1.117$$

This means that there were approximately 112 never-married men for every 100 never-married women living in the three Virginia counties in 2015.

Using formula (1), the MOE of the derived estimate of never-married males is calculated as:

$$\text{MOE}(\hat{X}_1 + \hat{X}_2 + \hat{X}_3) = \pm\sqrt{(4,222)^2 + (2,819)^2 + (2,259)^2} = \pm 5,557$$

A ratio has a slightly different estimator for the MOE than a proportion. If $\hat{R} = \hat{X}/\hat{Y}$, where $\hat{X}$ = the number of never-married males and $\hat{Y}$ = the number of never-married females, then the MOE of this ratio is approximated as:

$$\text{MOE}(\hat{R}) = \frac{1}{\hat{Y}} \sqrt{[\text{MOE}(\hat{X})]^2 + (\hat{R}^2 * [\text{MOE}(\hat{Y})]^2)} \quad (7)$$

Substituting the estimates for the numerator and denominator, and their respective MOEs in this formula yields:

$$\text{MOE}(\hat{R}) = \frac{1}{203,119} \sqrt{(5,557)^2 + [(1.117)^2 * (5,070)^2]} = 0.039$$

Using formula (2), the SE of the ratio is $0.039 / 1.645 = 0.024$. Using formula (4), the CV for the ratio of never-married males to never-married females is $(0.024 / 1.117) * 100 = 2.1$ percent, indicating that this is a very reliable estimate with an SE that is much smaller than the value of the estimated ratio.

Calculating Measures of Error for Derived Estimates of Percent Change

One of the most important benefits of annual releases of ACS estimates is the ability it provides for data users to analyze change over time. A frequent application is to calculate the percent change from one time period to another. For example, users may want to calculate the percent change from a 2009–2013 ACS 5-year estimate to a 2014–2018 ACS 5-year estimate. Normally, the current estimate is compared with the older estimate.

If the current estimate = $\hat{X}$ and the earlier estimate = $\hat{Y}$, then the MOE for percent change is calculated as follows:

$$\text{MOE}\left(\frac{\hat{X} - \hat{Y}}{\hat{Y}} * 100\right) = 100 * \text{MOE}\left(\frac{\hat{X}}{\hat{Y}} - 1\right) = 100 * \text{MOE}\left(\frac{\hat{X}}{\hat{Y}}\right) \quad (8)$$

Formula (8) reduces to a ratio, so the ratio formula (7) described in the section above should be used to calculate the MOE. As a caveat, users should be aware that this formula does not take into account the correlation when calculating a change between two overlapping time periods. To calculate standard errors for overlapping ACS multiyear estimates see the section on “Understanding Error and Determining Statistical Significance.”

Calculating Measures of Error for the Product of Two Estimates

In some instances, data users may need to derive an estimate by multiplying a published count by a published percentage. For example, a data user might be interested in the number of 1-unit detached owner-occupied housing units in a geographic area. In 2015, the number of owner-occupied housing units in the United States was 74,506,512 with an MOE of 228,238, and the percent of 1-unit detached owner-occupied housing units was 82.4 (0.824) with an MOE of 0.1 percent (0.001).⁷⁰ So, the number of 1-unit detached owner-occupied housing units was $74,506,512 * 0.824 = 61,393,366$. The formula to calculate the MOE of a product is:

$$\text{MOE}(\hat{X} * \hat{Y}) = \sqrt{(\hat{X}^2 * [\text{MOE}(\hat{Y})]^2) + (\hat{Y}^2 * [\text{MOE}(\hat{X})]^2)} \quad (9)$$

Substituting the estimates and their respective MOEs in formula (9) yields:

$$\begin{aligned} & \text{MOE}(\hat{X} * \hat{Y}) \\ & \sqrt{[(74,506,512)^2 * (0.001)^2] + [(0.824)^2 * (228,238)^2]} = 202,289 \end{aligned}$$

To obtain the lower and upper bounds of the 90 percent confidence interval around 61,393,366, add and subtract the MOE from 61,393,366. Thus, the 90 percent confidence interval for this estimate is $[61,393,366 - 202,289]$ to $[61,393,366 + 202,289]$ or 61,191,077 to 61,595,655. Using formula (2), the SE of the product is $202,289 / 1.645 = 122,972$. Using formula (4), the coefficient of variation for this derived estimate is $(122,972 / 61,393,366) * 100 = 0.2$ percent, indicating that the derived estimate is very reliable.

⁷⁰ U.S. Census Bureau, ACS 1-year estimates in data.census.gov, Table S2504: “Physical Housing Characteristics for Occupied Housing Units.”

Calculating Measures of Error Using Variance Replicate Tables

Advanced users may be interested in the *Variance Replicate Tables*, first released for the 2010–2014 ACS 5-year data in July 2016.⁷¹ These augmented ACS Detailed Tables include sets of 80 replicate estimates, which allow users to calculate measures of error for derived estimates using the same methods that are used to produce the published MOEs in data.census.gov. These methods incorporate the covariance between estimates that the approximation formulas in this document leave out.

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 will be released on an annual basis, shortly after the release of the standard 5-year data products.

Variance Replicate Documentation, including lists of tables and summary levels, is available on the Census Bureau's Web site.⁷²

⁷¹ U.S. Census Bureau, American Community Survey (ACS), Variance Replicate Tables, <www.census.gov/programs-surveys/acs/data/variance-tables.html>.

⁷² U.S. Census Bureau, American Community Survey (ACS), Variance Replicate Tables Documentation, <www.census.gov/programs-surveys/acs/technical-documentation/variance-tables.html>.