

Developing Guidelines Based on CVs for when One-Year Estimates Can Be Used Instead of Three-Year Estimates in the American Community Survey

Michael Ikeda¹

Statistical Research Division, U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC, 20233

Abstract

A method is outlined for developing guidelines for acceptable use of 1-year estimates rather than 3-year estimates for areas with more than 65,000 total population in the American Community Survey (ACS). The method is based on the coefficients of variation of the estimates. It is applied to estimates from the 2006 ACS and the results are presented.

Key Words: American Community Survey, One-Year Estimates, Three-Year Estimates, Guidelines

1. Introduction and Methodology

For areas with more than 65,000 total population both 1-year and 3-year estimates are now available for the American Community Survey (ACS). We want to develop simple guidelines, based on the coefficient of variation (CV), for when 1-year estimates can be used instead of 3-year estimates for these areas. We also want the general method to be applicable to developing guidelines for the use of 3-year estimates rather than 5-year estimates, and perhaps eventually for the use of 5-year estimates. In the case of 5-year estimates there are no alternative estimates that can be used, but it may still be useful to have information on when these estimates are likely to meet a desired CV.

For estimates that are basically estimated counts of people, households, or housing units, we seem to be able to obtain reasonable guidelines by setting cutoffs based on the *percentage that the estimate is of the total*. For 1-year estimates, we can use the empirical bivariate distribution of "estimated percent of total" and "estimated CVs" to set cutoffs so that:

- (a) most 1-year estimates with percentages above the cutoff will meet a desired CV, and
- (b) none or very few of these estimates will have CVs that exceed the desired CV by too much.

For a given target CV, if the percentage of total is greater than the cutoff, then the 1-year estimate can be used for any CV larger than the target; otherwise, the 3-year estimate can be used. We divide the areas into several population categories and set separate guidelines for each category.

To assign cutoffs we will divide the "percent of total" distribution into ranges. For each target CV we find the "percent of total" range where the percentage of CVs greater than the

¹This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical, methodological, and technical issues are those of the author and not necessarily those of the U.S. Census Bureau.

target exceeds 10%. We then set the cutoff at the upper limit for that "percent of total" range.

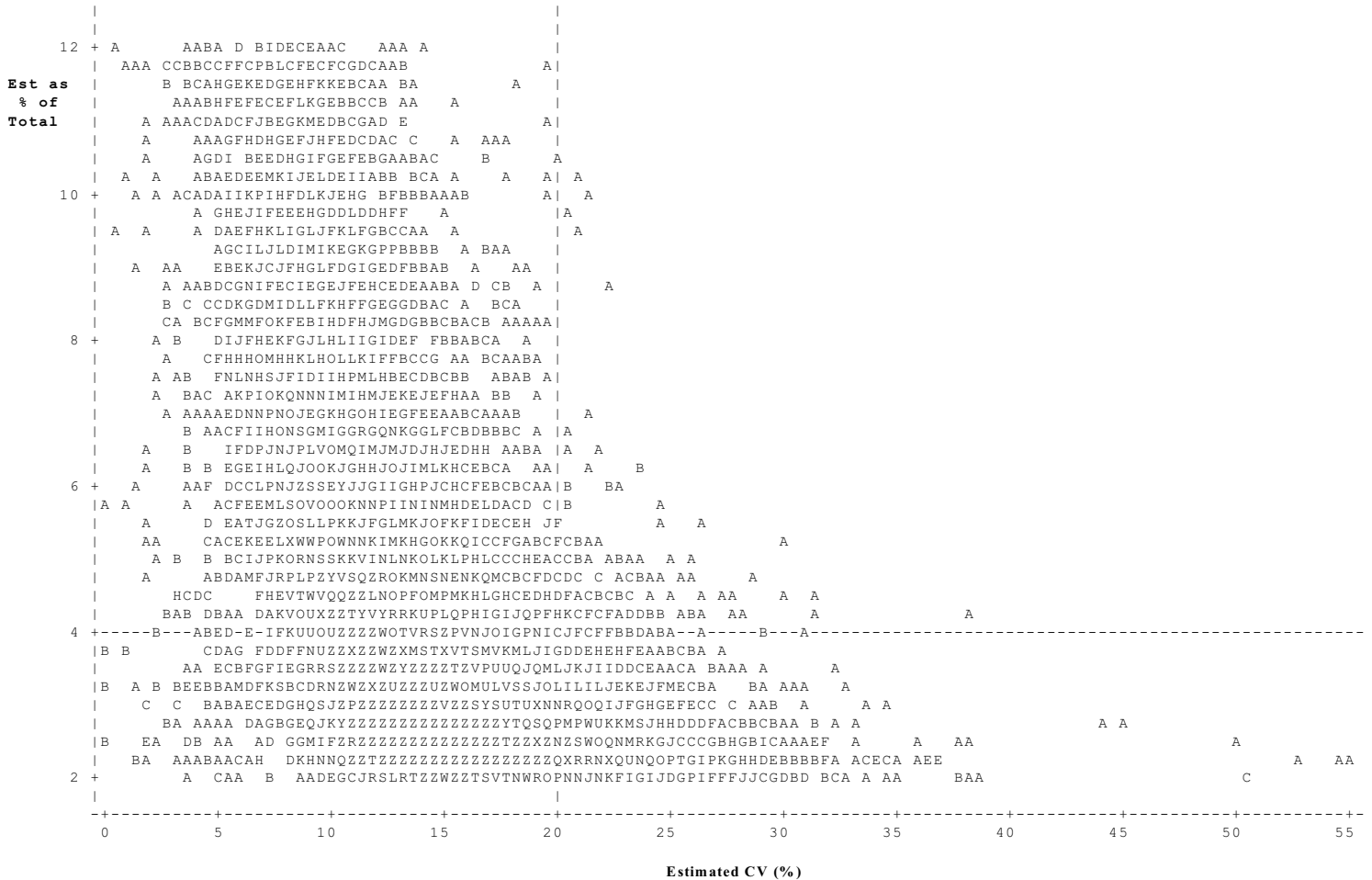
For estimates that are not counts, we will subdivide by type of estimate (aggregate total, median, ratio, gini coefficient) and use the empirical CV distribution for each estimate type. We again expect to divide the areas into separate population categories and set separate guidelines for each category. However, there are some noncount estimates with high CVs even for very large areas, so any general guidelines would need to be used with caution.

Figure 1 illustrates the concept behind the methodology for count estimates. For the clarity of the graph, the data shown in Figure 1 are **not** the data used in the rest of the paper, but instead are 2003-2005 data from the Multi-Year Estimates Study. This study was carried out before the ACS collected data nationwide and only covered a portion of the country and a subset of the estimates (U.S. Census Bureau, 2008). Figure 1 shows a "percent of total" range between 2% and 12% for a size classification that corresponds to a population from 45,000 to 64,999 for person estimates and approximately corresponds to that range for household and housing unit estimates. For illustrative purposes, Figure 1 shows the results of the calculation for the 3-year estimates from the Multi-Year Estimate Study. For this size category, and a target CV of 20%, the cutoff calculated for the 3-year estimates is 4% using the method outlined above. The horizontal and vertical lines show the cutoff and the target CV, respectively. The area on the upper right shows the estimates larger than the cutoff that have a CV larger than the target. The idea behind the method outlined above is that you want the number of estimates in the upper right quadrant to be a fairly low proportion of the number above the cutoff. Increasing the target moves the vertical line to the right and thus allows you to decrease the cutoff. Decreasing the target does the reverse.

Ikeda (2008) contains an earlier version of this paper.

Figure 1: Illustrative Graph, Three-Year Count Estimates from the American Community Survey Multi-Year Estimate Study Areas Mostly 45,000-64,999 population

Legend: A = 1 obs, B = 2 obs, etc. (Z = 26+ obs)



2. Results

Table 1 gives cutoffs for single-year count estimates obtained by applying the methodology outlined above to base table estimates from the 2006 American Community Survey (ACS). The base tables (called the detailed tables in American Factfinder) provide all of the detailed data on basic characteristics for each geographic area. Estimates at the U.S., region, and division summary levels were excluded from the calculation, as were geographic components (specified breakdowns such as urban, in metropolitan statistical area, etc.) within state. An overview of the 2006 ACS data is given in U.S. Census Bureau (2007) The file used for the research is an internal Census Bureau file that includes both estimates from published tables and estimates from tables filtered out due to variance concerns. A table is filtered out if the median CV for the table is greater than 0.61 (Baumgardner, 2006). The calculations in this paper did not take the results of variance filtering into account. This would probably tend to err on the side of safety when it comes to published estimates, as the estimates that are filtered out should tend to have high CVs. Tables 1a-1f expand the results

in Table 1 to give explicit "Yes/No" recommendations for the use of single-year estimates for each population range. Based on feedback from the Decennial Statistical Studies Division of the U.S. Census Bureau, Tables 1a-1f would be the preferable format for illustrating the guidelines.

Cutoffs were also calculated for the combined category 65,000-300,000. The cutoffs for the combined category were generally between the cutoffs for the two subcategories. Cutoffs were also calculated separately for person estimates and household /housing unit estimates for each population category. The cutoffs for person estimates tend to be slightly lower and the cutoffs for household/housing unit estimates tend to be slightly higher than the overall cutoffs, but not so much that separate tables would be required.

Table 1: Cutoffs by Percent of Total by Desired CV and Population Range for Single-Year Count Estimates

Population Range	Required Coefficient of Variation (CV)			
	5%	10%	20%	35%
65,000 to 150,000	75%	25%	7%	2.5%
150,001 to 300,000	40%	14%	3.5%	1.2%
300,001 to 1,000,000	16%	5%	1.2%	0.4%
1,000,001 to 3,000,000	7%	1.7%	0.4%	0.2%
3,000,001 to 15,000,000	2%	0.6%	0.2%	0.05%
15,000,001 to 40,000,000	0.6%	0.1%	0.04%	0.01%

Estimates for U.S., Region, and Division were not included in calculations.

Table 1a: Table of Guidelines on Use of One-Year Estimates for Geographies with 65,000 to 150,000 Population

Estimate as % of Total	Required Coefficient of Variation (CV)			
	5%	10%	20%	35%
less than 2.5%	No	No	No	No
2.5%-6.9%	No	No	No	Yes
7%-24.9%	No	No	Yes	Yes
25%-74.9%	No	Yes	Yes	Yes
75% and greater	Yes	Yes	Yes	Yes

Estimates for U.S., Region, and Division were not included in calculations.

Table 1b: Table of Guidelines on Use of One-Year Estimates for Geographies with 150,001 to 300,000 Population

Estimate as % of Total	Required Coefficient of Variation (CV)			
	5%	10%	20%	35%
less than 1.2%	No	No	No	No
1.2%-3.49%	No	No	No	Yes
3.5%-13.9%	No	No	Yes	Yes
14%-39.9%	No	Yes	Yes	Yes
40% and greater	Yes	Yes	Yes	Yes

Estimates for U.S., Region, and Division were not included in calculations.

Table 1c: Table of Guidelines on Use of One-Year Estimates for Geographies with 300,001 to 1,000,000 Population

Estimate as % of Total	Required Coefficient of Variation (CV)			
	5%	10%	20%	35%
less than 0.4%	No	No	No	No
0.4%-1.19%	No	No	No	Yes
1.2%-4.9%	No	No	Yes	Yes
5%-15.9%	No	Yes	Yes	Yes
16% and greater	Yes	Yes	Yes	Yes

Estimates for U.S., Region, and Division were not included in calculations.

Table 1d: Table of Guidelines on Use of One-Year Estimates for Geographies with 1,000,001 to 3,000,000 Population

Estimate as % of Total	Required Coefficient of Variation (CV)			
	5%	10%	20%	35%
less than 0.2%	No	No	No	No
0.2%-0.39%	No	No	No	Yes
0.4%-1.69%	No	No	Yes	Yes
1.7%-6.9%	No	Yes	Yes	Yes
7% and greater	Yes	Yes	Yes	Yes

Estimates for U.S., Region, and Division were not included in calculations.

Table 1e: Table of Guidelines on Use of One-Year Estimates for Geographies with 3,000,001 to 15,000,000 Population

Estimate as % of Total	Required Coefficient of Variation (CV)			
	5%	10%	20%	35%
less than 0.05%	No	No	No	No
0.05%-0.19%	No	No	No	Yes
0.2%-0.59%	No	No	Yes	Yes
0.6%-1.9%	No	Yes	Yes	Yes
2% and greater	Yes	Yes	Yes	Yes

Estimates for U.S., Region, and Division were not included in calculations.

Table 1f: Table of Guidelines on Use of One-Year Estimates for Geographies with 15,000,001 to 40,000,000 Population

Estimate as % of Total	Required Coefficient of Variation (CV)			
	5%	10%	20%	35%
less than 0.01%	No	No	No	No
0.01%-0.039%	No	No	No	Yes
0.04%-0.09%	No	No	Yes	Yes
0.1%-0.59%	No	Yes	Yes	Yes
0.6% and greater	Yes	Yes	Yes	Yes

Estimates for U.S., Region, and Division were not included in calculations.

To illustrate how these tables can be used, suppose a researcher is interested in characteristics for Lake County, Illinois. Lake County currently has an estimated population in the range 300,001 to 1,000,000 (644,356 in the 2000 Census) so Table 1c would be used. Suppose the researcher is satisfied if the CV is about 20%. Looking at Table 1c, any population estimate expected to be greater than 1.2% of the total population will usually meet the CV cutoff. Thus, for example, if the researcher is interested in an estimate of the foreign born population (14.8% of the total according to the 2000 Census) the one-year ACS estimate can be used, while for estimates that make up a small percentage of the county population (e.g., persons of Lithuanian ancestry) the one-year estimates would not be recommended.

Similar reasoning would be used for estimates of households, except that we would use the percentage of the total number of households. Thus, again using Lake County as an example, if the researcher is satisfied if the CV is about 20% then the one-year ACS estimate can usually be used for household estimates that are expected to be more than 1.2% of the total number of households such as families with income in the last 12 months below poverty level (just under 3% of the total number of households according to the 2006 ACS).

For the noncount estimates we use a somewhat analogous method. We again break the areas into population size categories. However, because "percent of total" is not really relevant for noncount variables, we set a single CV cutoff for each type of estimate within each size category. We consider four CV cutoff points: 5%, 10%, 20%, 35%. For each type of estimate within each population size category we find the smallest cutoff point for which the percentage of CVs greater than the cutoff point is less than 10%. That cutoff point is our CV cutoff for the given type of estimate and population size category. Table 2 lists the CV cutoffs. The one-year estimates are usually acceptable for any target CV above the cutoff. As mentioned above, however, the cutoffs should be used with caution since there are noncount estimates with large CVs even for areas with large populations.

Separate calculations were also done by breaking the 65,000 to 300,000 category into two subcategories (65,000 to 150,000 and 150,001 to 300,000). Dividing into two subcategories has little effect on the results.

Table 2: Single-Year Noncount Estimates--CV cutoff by Type of Estimate and Population Range

Population Range	Type of Estimate			
	Aggregate	Median	Ratio	Gini
65,000 to 300,000	---	---	35%	10%
300,001 to 1,000,000	35%	35%	20%	5%
1,000,001 to 3,000,000	20%	20%	20%	5%
3,000,001 to 15,000,000	20%	20%	10%	5%
15,000,001 to 40,000,000	5 %	10%	5%	5%

CV cutoff not listed if 50% or greater. Estimates for U.S., Region, and Division were not included in calculations.

Again using Lake County as an example, for medians the one-year estimates are usually acceptable if the desired CV is greater than 35%. Thus, if the researcher is interested in median household income and has a desired CV of 20% the one-year estimate probably should not be used.

The above results provide some rough guidance on when the one-year estimates are likely to be reliable enough to meet a given target CV. We can generally say that when the one-year estimates are expected to be sufficiently reliable, then the three-year estimates can also be expected to be sufficiently reliable. However, we cannot necessarily assume the reliability of the three-year estimates in those situations where the one-year estimates are not expected to be sufficiently reliable.

3. Future Research

We plan to do a similar analysis using base table data from the 2005-2007 ACS. We would perform the same types of calculations and calculate cutoffs for the three-year estimates in the same way we did for the one-year estimates. One complication is that ACS first included

Group Quarters data in 2006. Therefore, while there are three full years of data from the housing unit population in the 2005-2007 ACS, there are only two years of data from the Group Quarters population. This will tend to somewhat inflate the CV for estimates that include persons from Group Quarters. This should not usually have much of an effect on the calculations, since the Group Quarters population will usually be fairly small compared to the housing unit population.

While the one-year estimates are only available for areas above 65,000 in population, the three-year estimates are available for areas above 20,000 in population. We plan to start by calculating cutoffs for the areas above 65,000 in population, using the same population categories used for the one-year estimates. This will provide a comparison to the results using the one-year estimates. We may then proceed to calculate cutoffs for areas between 20,000 and 64,999 in population, probably dividing this range into two population categories.

Once we have completed the analysis for the three-year estimates, the obvious next step will be to conduct a similar analysis based on five-year base table estimates. However, the five-year estimates are not available at this time. Once they are available we could repeat the analysis on the five-year estimates. This analysis of the five-year estimates would probably be divided into three stages: the first stage would cover areas above 65,000 in population; the second stage would cover areas between 20,000 and 64,999; and the third stage would cover areas below 20,000 in population.

Acknowledgements

The author would like to thank Michael Beaghen and Lynn Weidman for helpful comments on earlier drafts of this paper.

References

- Internal Census Bureau Memorandum from S. Baumgardner to D. Hillmer (2006), "Final Report on "Research on Changes to the Current Base Tables Filtering Rules", " U.S. Census Bureau, September 26, 2006.
- Ikeda, M. (2008), "Developing Guidelines Based on CVs for when One-Year Estimates Can Be Used instead of Three-Year Estimates in the American Community Survey (ACS) for Areas with Populations of 65,000 or More," Study Series SSS2008/02, Statistical Research Division, U.S. Census Bureau, October 2008, www.census.gov/srd/papers/pdf/sss2008-02.pdf.
- U.S. Census Bureau (2007), "Quick Guide to the 2006 American Community Survey Products in American Factfinder," U.S. Census Bureau, September 2007, factfinder.census.gov/home/saff/aff_acs2006_quickguide.pdf.
- U.S. Census Bureau (2008), "About the Data (Methodology): Multiyear Estimates Study," American Community Survey Office, U.S. Census Bureau, last modified September 30, 2008, www.census.gov/acs/www/AdvMeth/Multi_Year_Estimates/overview.html.