District of Columbia
STATE DATA CENTER
MONTHLY BRIEF
July 2012

By
Daniel Sheres

This report is based on data produced by the U.S. Census Bureau and may differ from data produced by other entities.

Understanding the American Community Survey Data

Introduction

If you used the U.S. Census Bureau's website to search for data recently, you most likely have noticed that much of the data are now presented in the American Community Survey. The American Community Survey (also known as the ACS), is a nationwide survey designed to provide data users with annual demographic, social, economic and housing characteristics that have historically been available only once a decade from the decennial census sample. Although the topics covered by the ACS are similar to those that were covered by the decennial sample, the method of collecting the data is very different and introduces a mark difference in results and interpretation of the data. The purpose of this report is to provide data users with some guidelines on the interpretation and use of the ACS data.

Continuous measurement has long been viewed as a possible alternative method for collecting detailed information on the characteristics of population and housing in the United States. However, it was not considered a practical alternative to the decennial census long form until the early 1990s when the demands for current, nationally consistent data from a wide variety of users led federal government policymakers to consider the feasibility of collecting social, economic, and housing data continuously throughout the decade. The American Community Survey (ACS) was implemented to meet these demands. Population and housing data from the full sample size ACS, which is approximately 2.6 percent of the population or 2.3 percent of housing units, began in 2005.

The ACS is based on a questionnaire that is sent each month to a sample of addresses in the United States. At the end of each calendar year, these data are pooled and estimates are produced. Since the size of a geographic area determines the sample size, only larger areas with population 65,000 or more people receive 1-year ACS estimates products. Smaller areas with population between 20,000 and 65,000 receive 3-year ACS estimates products; and for places with less than 20,000 people, five years of data need to be collected to provide reliable estimates, hence the 5-year ACS estimates products. These estimates which are based on data collected over one, three, and five years are referred to as "period" estimates, meaning that numbers represent an area's characteristics for a specific period of time. This differs from the decennial census, where the data are pegged to an April 1 reference point.

Understanding the ACS

There are conceptual differences between a sample survey that is taken during the decennial census and a series of monthly surveys that are taken throughout the decade. One of these differences relates to sample size. For the decennial census, the sample size used in the 2000 census for example was one in six housing units or 17 percent. The full implementation of the ACS, which began in 2005, samples approximately 2.9 million housing unit addresses annually, nationally. Thus, the sample size is much smaller at one in 45 or 2.2 percent of housing units. For the District of Columbia, the ACS sample size is approximately 6,000 housing units annually, or one in 49 or 2.2 percent of housing units. Smaller sample sizes impact small area estimates adversely by increasing the variability or uncertainty of the estimate. This uncertainty is referred to as "sampling error," and it means that the estimate derived from the sample survey will likely differ from the values that would have been obtained if the entire population was included in the survey, as well as from values that would have been obtained had a different set of sample units been selected. Understanding statistical sampling and the ramifications of working with sample data are key to using the ACS data successfully.

Sampling errors can be expressed quantitatively in different ways. To gain the most value from the data presented in the ACS, four of these expressions are discussed in this report: 1) sampling error, 2) margin of error, 3) confidence intervals, 4) and rare events.

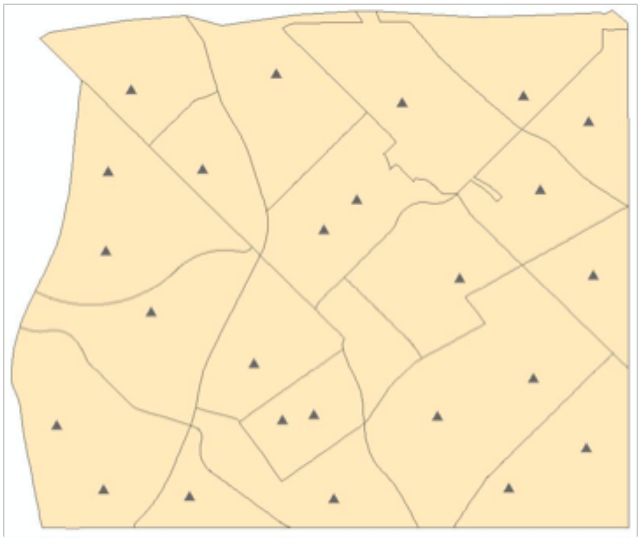
Sampling Error

Inherent in every survey is sampling error–error that results from the uncertainty of the estimate. Sampling error is present whenever the entire population is not surveyed and indicates that the population may not be accurately represented by the sample. A better understanding of sampling error can be gained from the following example:

Figure 1 represents a hypothetical study area. It is assumed that the area has an estimated population of 2,500 households. In order to perform a one percent random sample of the area, 25 households (sampling locations) are required; these are indicated in Figure 1 by the small triangles.

Random sampling can often lead to unevenly distributed sampling locations as depicted in Figure 1. Error is introduced when there are areas with more sampling locations than other areas, a very common sampling scenario. In Figure 1, there are six areas with more than one sampling location, 13 areas with one sampling location, and one area did not get sampled at all. Areas which have more than one sampling location can be described more accurately by statistics. The general rule is that more sampling locations means there will be less errors.

Figure 1. Study Area with Sampling Locations



Note that 6 areas have more than one sampling location, 13 areas have a single sampling location, and 1 area did not get sampled.

Margin of Error

For any survey where random sampling is employed, there will be some error in the estimates. This error is represented by the Margin of Error (MOE). MOE is the difference between an estimate and its upper or lower limit, known as confidence bounds. Confidence bounds can be created by adding the MOE to the estimate for the upper bound and subtracting the MOE from the estimate for the lower bound. Estimates with smaller MOE are more accurate, while estimates with larger MOE are less accurate. ACS estimates are always presented with a MOE. For example, Table 1 indicates that of the people aged 25 and over in the District, the estimated number of high school graduates is 84,569 and the MOE is plus or minus 4,333. This means that the actual number of high school graduates may be as high as 88,902 or may be as low as 80,236. In reality, the number might fall outside of this range of values, but with the estimate and the MOE together, there is some level of confidence in the possible range of the true value, or number of high school graduates in this case.

Table 1. Margin of Error Example 1

Educational Attainment – Population 25 years and over	Estimate	Margin of Error (MOE)
High school graduate (includes equivalency)	84,569	+/-4,333

Source: US Census Bureau American Community Survey 2010 1-Year Estimates

With the three different ACS products, there are varying levels of accuracy depending on the chosen statistic. For some statistics the 5-year data will be the more accurate than the 3-year or the 1-year estimate.

In Table 2, the number for high school graduates is presented for the three ACS products. Notice that the MOE for the 5-year (+/- 2,035) is smaller than the MOE for the 3-year (+/- 2,857) which is smaller than the MOE for the 1-year (+/- 4,333). This is as a result of the increased sample size in the 5-year survey as compared to the 1-year and 3-year surveys.

Table 2. Margin of Error Example 2

Educational Attainment – Population 25 years and over	ACS 1-Year 2010		ACS 3-Year 2008-2010		ACS 5-Year 2006-2010	
	Estimate	MOE	Estimate	MOE	Estimate	MOE
High school graduate (includes equivalency)	84,569	+/-4,333	80,363	+/-2,857	79,659	+/-2,035

Source: US Census Bureau American Community Survey 2010 1-Year, 3-Year, and 5-Year Estimates

Confidence Intervals

A confidence interval (CI) is the range that is expected to contain the average value of the characteristic that would result over all possible samples with a known probability. This probability is called the "level of confidence" or "confidence level." The Census Bureau statistical standard for published data is to use the 90-percent confidence level. Thus, the MOEs published with the ACS estimates correspond to a 90-percent confidence level. For example, given the MOE at the 90-percent confidence level used in constructing the estimated number of high school graduates in the ACS 1-Year 2010 in Table 3, the user could be 90 percent certain that the total number of high school graduates in the District was between 80,236 and 88,902.

The choice of confidence level is a matter of preference, balancing risk for the specific application. A 90-percent confidence level implies a ten percent chance of an incorrect inference, in contrast with a five percent and one percent chance if using the 95-percent and the 99-percent confidence levels, respectively. Therefore, if the impact of an incorrect conclusion is substantial, the user should consider increasing the confi-

Table 3. Margin of Error at the 90%, 95% and 99% Confidence Intervals

ACS 1-Year 2010	Estimate	MOE - 90%	Estimate	MOE - 95%
High school graduates (or equiv)	84,569	+/-4,333	+/- 5,163	+/- 6,785

Note: MOEs are larger for the 95% and 99% confidence intervals.

dence level. To convert to a 95 percent confidence interval, one has to multiply the ACS 90-percent MOE by 1.19; to convert to a 99 percent confidence interval, one has to multiply the ACS 90-percent MOE by 1.56. In Table 3, the results of that conversion are presented using the same data as in Table 2.

Rare Events

Random sampling naturally lends itself to higher accuracy with common events (large segments of the population) and lower accuracy with rare events (small segments of the population). For example, as a percentage of the population, there are fewer Native Americans in the District of Columbia than there are Whites or Blacks. When sampled, the relative rarity of Native Americans present introduces more errors in the data.

The logic is, if one were to pick a house in the District at random, it would be much more likely to pick a house with either a White or Black family than to pick a house with a Native American family. The reason for this is that Native Americans represent 0.3 percent (1,835 people) of the District's population in 2010, a relatively small component. Table 4 shows that the smaller or more rare the estimate, the higher the MOE tends to be. In some cases, the MOE can actually exceed the estimate; so one must take this into consideration when using this data.

Notice also that the smaller estimates have higher relative margins of errors (shown as MOE as percent of estimate). Where the estimates are large (e.g., White and Black or African American), the MOE is relatively small, 1.8 percent and 0.7 percent, respectively. However, when the estimates are smaller (e.g.; American Indian and Alaska Native), the MOE is larger relative to the estimate (39.8 percent). The rule of thumb is: the smaller the MOE is compared to the estimate, the more accurate the estimate. It is important to note that when the MOE exceeds the estimate, the data are not likely to be accurate and should either not be used or used with a note of caution.

Tips to Remember

1) Depending on the desired statistics, different data sets will give different levels of accuracy. For relatively rare events or smaller geography and/or populations, the 3-year ACS and 5-year ACS will generally provide more accurate estimates and smaller MOE.

2) Be wary of data where the MOE is more than 50 percent of the estimate. The estimate will most likely be unreliable. In these instances, one may choose to use alternative data sources.

3) Consider the geography. For state and national levels, ACS data are available in the 1-year, 3-year, and 5-year products. ACS data for the 3-year product are available for some counties and heavily populated areas. For the District of Columbia as a whole, ACS data are available in the 1-year, 3-year and 5-year products. However, for smaller geographies, like the ward and census tract levels, the ACS data are only available in the 5-year product.

4) Wherever possible, include MOE with published estimates.

5) Convert published ACS MOE to 95-percent or 99-percent confidence intervals if there is a need to minimize risk.

6) It is not advisable to compare overlapping data periods and different periods of data products. For example, compare 1-year estimates to other 1-year estimates, and 3-year estimates to other 3-year estimates. Do not compare 1-year estimates to 3-year estimates. Similarly, with overlapping periods, for example, do not compare 2005-2007 estimates with 2006-2008 estimates; years 2006 and 2007 overlap in both products. Rather, it is better to compare 2005-2007 with 2008-2010 3-year estimates, where there are no overlapping years.

Table 4. Margin of Error in Estimates of Rare Events

ACS 1-Year 2009	Race		
ACS Demographic and Housing Estimates	Estimate	Margin of Error (+/-)	MOE as Percent of Estimate
Total population	599,657	-	-
White	232,247	4,198	1.8%
Black or African American	319,119	2,130	0.7%
Asian	17,268	980	5.7%
American Indian and Alaska Native	1,835	730	39.8%
Native Hawaiian and Other Pacific Islander	539	368	68.3%

Source: US Census Bureau

For more information contact:
D.C. Office of Planning State Data Center
1100 4th Street SW, Suite E650, Washington, DC 20024
202.442.7600 ph 202.442.7638 fax
www.planning.dc.gov

District of Columbia
Office of Planning