Your Guide

Using American Community Survey Data by Steve Beleu

Abstract:
This guide gives users the basic information that is needed to get the most out of the American Community Survey (ACS) data available on the Census Bureau website. It details information as "period of time" data versus "point of time" data, compares ACS data with Decennial Census data, compares different groups of ACS data with each other, discusses four types of sampling errors, four types of non-sampling errors, quality measures, data swapping, and advises librarians on how to work with their customers so that they get the data needed, understand the data, and work with the data accurately. The last page of this guide provides a summary of suggestions for working with the ACS. You may freely duplicate this guide for use at your reference desk or for distribution to your to your library customers.

Oklahoma
Department of
Libraries

Your Guide to Using
American Community Survey Multiyear Estimates

Keywords: Census, Decennial Census, American Community Survey, sampling errors, non-sampling errors, quality measures, data swapping

The American Community Survey (ACS) multi-year estimates are a data product of the Census Bureau that is different from the Decennial Census data file. The differences arise from the fact that ACS uses data gathered within a "period of time" estimate rather than a "point of time" count, which is the way that Decennial Census data are gathered every ten years.

A period of time estimate, also known as a period estimate, gathers data that describes the average characteristics of a Census geographical area over a period of time rather than on a specific date. By the end of a year, the U.S. Census Bureau will have gathered data from every survey address. For ACS 1-year period estimates, that period of time is the entire calendar year of 12 months; for ACS 3-year multi-year estimates, that period of time is the entire calendar year for three years (36 months); for the upcoming ACS 5-year multi-year estimates, that period of time is the entire calendar year for five years (60 months). Obviously, a 1-year period has a smaller sample size than a 3-year period, and a 3-year period has a smaller sample size than a 5-year period. However, a 1-year estimate will be more current than a 3-year estimate, and a 3-year estimate more current than a 5-year estimate. The basic difference is that the Decennial Census is an actual count of the population; but as a survey, the ACS is a report on the average characteristics of the population over a specific period of time.

Data are reported for Census geographies only if they have a large enough population that meets the following "population threshold" sizes:

> 50 to 20,000 population – has 5-year estimates
> 20,000 to 65,000 population – has both 3-year and 5-year estimates
> 65,000 + population – has 1-year, 3-year, and 5-year estimates

One note about the "50 to 20,000 population" threshold mentioned above: The U.S. Census has a minimum threshold value for any geography below which they will not report data in order to preserve the privacy of individuals. The thresholds for Census

datasets have been either 50 people or 100 people, and will continue to vary by specific data product, although ACS 5-year estimates will have no threshold because the data has such high Margins of Error that they serve to protect the privacy of data users—which is good for privacy protection, but bad for data users.

When do you use these multi-year estimates? When data for a 1-year estimate for any geography are not available; when the Margin of Error (*see below for explanation*) for a 1-year estimate is unacceptably high; and to get data for small population groups that you could not get otherwise.

The Basics

The U.S. Census Bureau gathers data for the ACS by mailing ACS forms, which are about the same size as Decennial Census forms, to residences and group quarters. People who live in urban and suburban areas should receive a form about once every nine years; people who live in rural areas should receive a form about once every five years. Note that these lengths of time are not fixed in stone – if the Census Bureau needs more data, this could change. When the Census Bureau receives the completed ACS forms back, they compile data for 12-month periods as 1-year estimates, 36-month periods as 3-year estimates, and 60-month periods as 5-year estimates. None of the data is averaged, which means that 1-year data are not averaged for 12 months, 3-year data are not averaged for 36 months, and 5-year data are not averaged for 60 months. The only ACS data adjusted is the financial data, which is adjusted forward to reflect what existed on January 1st of the final year of either the 3-year or the 5-year group of data. This is done to reflect the effect of inflation.

The forerunner of the American Community Survey was the "*2000 Supplemental Survey*" that was taken in January and December of 2000. There was also a "*2001 Supplemental Survey*." Both surveys provided data only for geographies of 250,000 or more. During the same period of time, the first actual "*American Community Survey 2000*" was also taken. This first ACS provided data only for geographies with populations of 65,000 or more. From this first ACS through the ACS of 2006, data exists only for geographies of 65,000 or more.

The ACS of 2007 continues to provide data for geographies of 65,000 or more, but the first multi-year estimates, 2005 – 2007 American Community Survey Multiyear Estimates, provides data for all geographies with populations of 20,000 and above. After the 2010 Census, the first five-year multi-year estimates of 2005 – 2009 American Community Survey Multiyear Estimates will provide data for all geographies with populations of 50 or more.

Due to lack of Congressional funding, data for such "Group Quarters" as military base barracks and other base housing, college and university residential housing, residential hospitals, nursing homes, prisons and jails, and other group quarters were not collected until the 2006 American Community Survey. There are geographies in which this may not make a difference, but in most communities it can make a difference.

> **Example:** in 2005, Norman, Oklahoma was listed as the third largest city in the state because no data were collected about the soldiers at the military base of Fort Sill in Lawton, Oklahoma. When these data were collected in the 2006 ACS, Lawton resumed its former ranking as the third largest city in Oklahoma for that year.

ACS data contains values that have been suppressed to protect the privacy of individuals in 1- and 3-year data. The unsuppressed data are identified as "Base" tables and are noted with a "B" coding; the suppressed data are identified as "Collapsed" tables and are noted with a "C" coding.   "B" data are the most detailed for all geographies and all topics. To create "C" data, two or more lines of "B" data are merged, in a process that the Census Bureau calls "collapsing," into a single line of "C" data. The process of "merging" data consists of taking several tables of "Base" table data that, if published, might be used to identify individuals. These are then combined into one table.

> **Example:** Data from several Base tables for people aged 10-14, 15-17, and 18-19 are combined into one table that gives values for people from age 10-19. Census tells us that because Collapsed Tables are made from combining several Base Tables, they may not be as geographically specific as Base Tables, but may consist of data that are more reliable. It is also possible that data that would have existed at the "Base" level will not exist at the "Collapsed" level for specific geographies.

What to do and what not to do

Because of the nature of the data sets, you will not be able to give a customer data for a city of, for example, 51,000 in December except for the actual ACS multi-year estimate group of years. You must identify the data correctly by telling them that the data are for "ACS 2005-2007" but not "ACS 2005" or "ACS 2006" or "ACS 2007."

You and your customers must correctly label the data citations to reflect this multi-year nature. ACS is a paradigm shift in gathering, reporting, and using data. You can not derive data for any one year of a multi-year estimate, since the data are not averaged and are additionally adjusted in various ways.

What do you do about getting data for a geography if the Census geography has changed over time due to a city annexing a town or de-annexing a part of town, or if the population increases drastically in an area or if it falls drastically, etc? It is best to contact your local State Data Center. They will do the detective work to tell you how to adjust geographies to get comparative geographies, i.e., geographies that are similar or identical with each other from one Census to another.

Examples: a particular Census Tract in the 2000 Census may be the same Census Tract in the 2010 Census, but if an area has suffered from population decline that Census Tract may have been merged into other Census Tracts. Similarly, if an area has grown one Census Tract may have been become two or three Census Tracts. The following link takes you to a list of State Data Centers (click on the map) - http://www.census.gov/sdc/

Do not "Cherry-pick" data

As a result of American Community Surveys differing slightly from each other: you will need to coach your customers not to "cherry-pick" data from non-matching ACS multi-year estimate groups. For example, someone may mix 1-year with 3-year and 5-year estimates. We will need to tell them that this is bad data practice and advise them what the "Best Practices" are.

The Census data community has discussed the fact that a five-year multi-year estimate is going to be more reliable than a three-year estimate, and a three-year estimate is going to be more reliable than a 1-year estimate for the simple reason that more data have been gathered in a multi-year estimate.

Data users will learn how to choose multi-year estimates when the margin of error for data they are working with is larger than they feel they should use in a 1-year estimate; after 2010 they may choose a five-year estimate if the margin of error is larger than that in a three-year estimate. They may also choose to use a multi-year estimate if they are working in a smaller geography that has no 1-year estimate available for it.

Currency vs. Reliability

There is also a "currency versus reliability" trade-off here. The most recent data will always be 1-year data that are from a large sample size rather than multi-year data that are from a smaller sample size. The reason why 1-year data are more recent is that the Census Bureau will publish it earlier than it will publish multi-year estimates.

Adjustments for Inflation

Data that includes such financial information as income, rent, home value, and the cost of energy are adjusted for inflation to reflect the last data year of any multi-year estimate. Example: data for income in 2005-2007 ACS data are adjusted to reflect inflation in 2007. These adjustments are made in accordance with Consumer Price Index methodology.

Geographic Boundaries

All multi-year estimates are based on the geographic boundaries that existed on January 1 of the last data year of any multi-year estimate. 2010 Census boundaries will be used in the ACS data released in 2011.

How to Compare ACS Multiyear Estimates with Each Other

This is simple: you can only compare 1-year estimates with other 1-year estimates; 3-year estimates with other 3-year estimates; and 5-year estimates with other 5-year estimates. We know that our customers may wish to "cherry-pick" data. Nevertheless, our professional ethics forbid us from doing the same, and we should also help customers make accurate "Best Practice" comparisons.

How to Compare ACS Multiyear Estimates Across Time Periods

Since the components of the data estimates in any 3-year or 5-year period will change from year to year depending on survey participation for any geography (state, county, Census tract, etc.), comparisons of time periods that do not overlap each other are more reliable than time periods that overlap each other. Example: it is best to compare 2005-2007 ACS data with other 2005-2007 ACS data, or 2006-2008 ACS data with other 2006-2008 ACS data.

This means that it is also best to compare ACS data used to get a run of historical data in sequential ACS data periods. Example: use 2005-2007 ACS data, then use 2008-2010 ACS data to get a full six year's worth of data rather than using 2005-2007 ACS data with 2006-2009 ACS data and 2008-2010 ACS data.

How to Compare ACS Data with Census 2000 and Census 2010 Data

You are not comparing apples with oranges in comparing these datasets, but you are comparing different types of apples. Always include notes in your reports that you are comparing ACS data with Census 2000 or Censes 2010 data so that your reader will know that your data report only points out general trends, not specific realities.

Four Types of "Sampling Errors" You Must Know About to Work with ACS Data

A sampling error means this and nothing more: there is an uncertainty associated with ACS data because it is gathered as a "sample" survey of a population rather than the full "count" of a population, such as the 2000 Census or 2010 Census. All data surveys exhibit these sampling errors. Of the following measures, Margin of Error is what your customers will probably work with most.

Margin of Error (MOE) – This measures the precision of a data estimate reflected in the level of confidence that is reported for that data table (such as "90%, 96%, 98%" etc.). Example: a 90% MOE for a data table means that the ACS data and the actual characteristics of the population value differ by no more than 10%. Note: this is not the same as saying that your customer can be 90% sure that the data they are working with is reliable! This is not what Margin of Error is, but is a commonly made mistake. MOE can help customers assess the reliability of a data estimate.

Where To Find MOE – *this is reported in ACS tables with the data*

Confidence Interval (CI) – Confidence Interval determines a range of data that is said to be accurate for the population and geography being measured. This is reported as a range of accuracy percentage. This CI will differ between different ACS surveys, such as ACS 2005-2007 having a different CI than ACS 2006-2008, and ACS 2006-2007 having a different CI than either of the earlier ACS surveys. Example: the Census Bureau is 94% certain that the Confidence Interval for a Census Tract in ACS 2005-2007 is between 54.5% and 59.9%. This is a measurement of the accuracy of different ACS surveys as compared with each other, and the way your customers might possibly use this is that they could choose to use an ACS survey that has a higher Confidence Interval than an ACS survey that has a lower Confidence Interval for a specific geography.

Standard Error (SE) – This measures the variability of an estimate, and thus the accuracy, of an estimate due to sampling. This reflects the basic accuracy of a survey, and is used along with Margin of Error to compute the Confidence Interval.

Coefficient of Variation (CV) – This measures the "sampling errors" that are associated with a sample estimate. In general, the larger the sample size of any survey, the smaller the sampling error; and the smaller the sample size, the larger the sampling error.

Again, we need to advise our customers about these sampling errors regardless of whether they choose to take them into account or ignore them. We also need to advise our customers about how to use these in a "Best Practice" mode. Example: if a researcher does not like the data for a particular geography that is reported as "9.9," and the Margin of Error noted for that data estimate is "7.2", they could either add or subtract up to "7.2" to that "9.9" data value to get a range of data between 2.7 and 17.1. This is not a demonstration of "Best Practice" in using ACS data, and our professional duty is to advise against it.

General rule: the smaller the population group is for a geography, the larger the size of the Standard Error and Confidence Intervals. For more information about the accuracy of ACS data see:

http://www.census.gov/acs/www/UseData/Accuracy/Accuracy1.htm

Four Types of Non-Sampling Errors

- Non-response errors,
- Response errors,
- Processing errors, and
- Coverage errors (read: mistakes-in-coverage errors).

Quality Measures

http://www.census.gov/acs/www/UseData/sse/

Note: you will also find a "Quality Measures" link on the top right screen of the ACS section of the American FactFinder homepage.

The Census website features a page titled How to Use the Data: Quality Measures that contains information about the quality of *American Community Survey* data for your state. You can also choose "Nation" and "All States" to see data on one screen for the nation, every state, the District of Columbia, and Puerto Rico. Quality measure data for older years of ACS are available on the ACS website at:

http://www.census.gov/acs/www/UseData/Accuracy/Accuracy1.htm

- **Size of Sample** – how large was the sample in your state as measured by the number of housing unit addresses that received ACS forms? Measured by the number of "initial addresses/group quarters sample selected" and the "final interviews" received.

- **Coverage Rate** – what was the coverage rate in your state? Coverage rate includes "under-coverage" when people in housing units do not have a chance of being selected to receive and return an ACS form, for any reason (e.g., because of postal problems), and "over-coverage" when people in housing units have more than one chance of being selected to receive and return an ACS form (e.g., because of people who own a vacation home). Measured by the percentage of housing units in a state that were covered and also by the percentage of the male and the female population that were covered.

- **Response Rate** – how many people in your state returned their ACS forms to the Census Bureau? Census actually measures the rate of nonresponse. Measured by percentage, it includes geographic area details for these categories of nonresponse: refusal, unable to locate, no one home, temporarily absent, language problem, insufficient data, and "other."

• **Completeness of Data** –how complete were the data that were used to create an estimate for a state? This also measures nonresponse, but for the missing data that occurs when someone who has returned an ACS form fails to answer a question, or could not be contacted by a Census enumerator. Measure by percentage for different topics.

Quality Measures Reported in ACA Data

When you do a search in the "Detailed tables" tools for 2006 ACS data, 2007 ACS data, or 2005-2007 ACS the first two tables that will be included for any search are "B00001 - Unweighted sample count of the population" and "B00002 - Unweighted sample housing units"

Data Swapping

"Data swapping" is a tool that the Census uses to protect the privacy of individuals when they report data for such small geographies as Census blocks.

> **Example:** if one or two individuals in a first specific Census block are Asian-American, and one or two individuals in a neighboring second specific Census block are African-Americans, the Census Bureau will "swap" the data between the first and second blocks to protect their privacy.

This is only done at a very specific level so that totals for the larger area of the Census tract, or perhaps for the Census Block Group, add up accurately. We need to remember and advise our customers that data for one or two people at a small geography, may have been swapped to conceal information about the people who live there.

Miscellaneous

• The Census Bureau will begin using the geographies of the 2010 Census in 2011 ACS data.

• The term "statistically significant" when used in the context of Census does not mean the same thing as "statistically significant" when applied to society. In the context of Census data, "statistically significant" only means that the data are significant in the context of numbers, not the world we live in. However, this mistake is frequently made by the media to create news stories.

Websites about ACS

ACS Homepage - http://www.census.gov/acs/www/

Guides on how to use the ACS :
http://www.census.gov/acs/www/UseData/Compass/handbook_def.html

Comparing ACS data to other data such as the Decennial Census:
http://www.census.gov/acs/www/UseData/compACS.htm
(Be sure to look at the "table crosswalks" which gives table numbers that correspond to each other between the Decennial Census and the ACS.)

Glossary of terms for different American Community Surveys:
http://www.census.gov/acs/www/UseData/Def.htm

Papers, Presentation, and Webinars:
http://www.census.gov/acs/www/AdvMeth/Papers/Papers1.htm

Using Multiyear Estimates:
http://www.census.gov/acs/www/UseData/myeoverview.html

Subscribe to the ACS Alerts to know what is happening in the ACS:
http://www.census.gov/acs/www/Special/Alerts.htm

To access ACS data from American FactFinder:
http://factfinder.census.gov/home/saff/main.html?_lang=en

You should access ACS data from American FactFinder, not the ACS homepage. The "Access Data" link on the ACS homepage links you to American FactFinder. Note that the "Access Data" link also provides three additional means of accessing the data:

1. by FTP protocol,
2. by PUMS (Public Use Microdata Samples), and
3. by custom tabulations done by the Census Bureau that you will pay for.

The Do's and Don'ts of Working with ACS Data – Summary

- Best practice: only compare 1-year estimates with other 1-year estimates, 3-year estimates with other 3-year estimates, and 5-year estimates with other 5-year estimates. If you should make such cross-comparisons, tell this to your readers in a footnote.

- The threshold for reporting data for any geography is between 50 and 100 people, so data may not be available for, say, a town of 35 people.

- Remember that American Community Survey before the 2005-2007 ACS did not contain data for any group quarters, such as military housing, university housing, nursing homes, etc.

- "Base" data tables are the most detailed data for any geography, but may not be available for that geography; you will probably still have to create a "Collapsed" data table to use for that geography, but not in all cases.

- Label your ACS data correctly. If you get data from an ACS multi-year survey, such as the 2005 – 2007 ACS, you can not label the data as either '2005" "2006" or "2007." You must label it as "ACS 2005 – 2007" data.

- Contact your State Data Center for help with geographies that may have changed from a past survey to a recent survey.

- Do not "cherry-pick" your data by mixing the data from one survey with the next.

- Best practice: do not compare an ACS survey with another ACS survey that covers only part of the time period of the first survey; do not overlap surveys, such as comparing data from the 2005-2007 ACS with the data from the 2006-2008 ACS.

- Pay attention to "Sampling Errors," especially to Margin of Error (MOE), and contact your State Data Center for help with calculating and working with other sampling errors.

When in doubt about how to use ACS surveys,
contact your State Data Center for help: http://www.census.gov/sdc/

Acknowledgements

Steve Beleu has been with the Oklahoma Department of Libraries since 1979. He became Regional Depository Librarian in1983. He has been a Fellow of the National Center for Education Statistics since 2002, and is also Director of the Oklahoma State Data Center Coordinating Agency. The Oklahoma Department of Libraries was named Federal Depository Library of the Year, 2009. He specializes in developing and presenting workshops about online federal information.