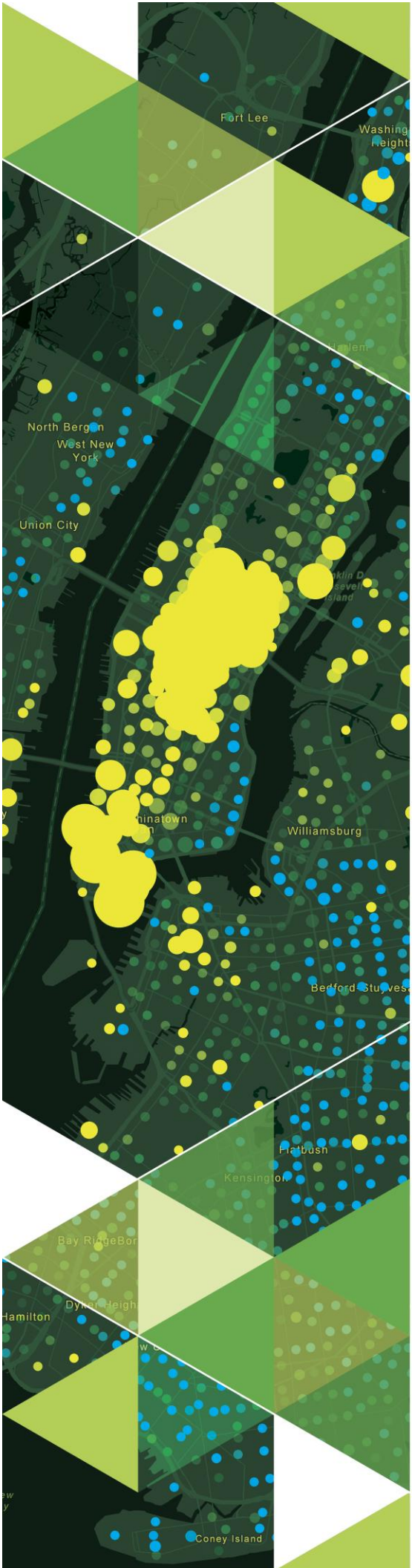

AN ESRI
WHITE PAPER

JUNE 2020

Methodology Statement: 2014–2018 American Community Survey

380 New York Street
Redlands, California 92373-8100 USA
909 793 2853
info@esri.com
esri.com

esri | THE SCIENCE OF WHERE®

Table of Contents

Methodology Statement: 2014–2018 American Community Survey

Introduction

The 2014–2018 data from the American Community Survey (ACS) is now available from Esri. Esri provides reports, data enrichment, and thematic mapping for ACS estimates in standard geographies, current ZIP codes, and user-defined polygons. Reports include two summary profiles, Population and Housing. Esri's reports/maps are designed to simplify the data and enhance its usability with reliability thresholds. Online help is also provided to explain the data.

The ACS is the de facto replacement for sample data from the decennial census. The 2010 Census eliminated the long form. Those who want data on income and poverty status, school enrollment, journey to work, household type and relationships, languages spoken, migration, citizenship, disability, health insurance, ancestry, military service, or housing characteristics must turn to the American Community Survey.[1]

The US Census Bureau was testing this replacement before Census 2000; however, the full rollout of the ACS did not happen until 2005. The first release of ACS data for *all* counties, plus tracts and block groups (BG), was in December 2010. Earlier releases of ACS data (2006 through 2009) were only available for larger geographic areas.

What's New

For the first time, Esri is making ACS data available for Puerto Rico. All of the ACS data that is available for the US is also now available for Puerto Rico. In addition, Esri has made additional data available for Puerto Rico that is of interest to users. Reports for Puerto Rico include population and housing reports similar to the US as well as an additional report containing Puerto Rico information called Key Population and Household Facts.

ACS versus Census 2000: What's the Difference?

The first thing that you notice on an ACS report or map is the additional number shown for the margin of error (MOE). The margin of error represents the confidence interval for an ACS estimate. There were no margins of error reported for Census 2000 sample data. The MOE epitomizes the main difference between Census 2000 and ACS data—the precision of the estimates.

The subjects included in the ACS are similar to the Census 2000 sample; however, the method of collecting the data is very different, which introduces conspicuous differences in the results. The Census 2000 sample represented approximately 1 in 6 housing units at one point in time, April 1, 2000. The ACS represents

[1] General information about the American Community Survey is summarized here. However, this is the Census Bureau's data. More information is available from the ACS handbooks at https://www.census.gov/programs-surveys/acs/guidance/handbooks.html.

approximately 1 in 40 housing units annually and continuous measurement of demographic characteristics through monthly surveys. Releases for all areas down to block groups represent averages over 60 months, or five years (2014–2018, inclusive).

There are important distinctions between sample data provided in conjunction with a census and sample data collected throughout the year, every year. The differences are summarized here, with references to additional documentation for the curious data user. There are three key differences between Census 2000 sample data and ACS estimates:

- Data collection: Ongoing monthly surveys (ACS) vs. single survey (Census 2000)
- Time frame: Period estimates (ACS) vs. point estimates (Census 2000)
- Sample size: 1 in 40 housing units (ACS) vs. 1 in 6 housing units (Census 2000)

These differences in methodology can affect comparisons of the data over and above the demographic change that occurred between 2000 and the latter half of the decade.

Data Collection/ Methodology

The continuous data collection of the ACS necessitates changes in variable definitions, sample weighting, and sizes:

- Residency rules are different. The ACS defines a resident by a two-month rule. The census rule is "usual place of residence" or wherever a person spends most of the year. ACS data may include seasonal populations in addition to year-round residents.
- Date-specific variables, like employment, represent monthly averages, including seasonal variations.
- Since income is also collected over the course of the previous 12 months, it must be adjusted by the Consumer Price Index to represent a calendar year.
- Migration is now measured from one year ago, not five years ago.
- Survey samples must be weighted by *estimates* for states, counties, or places, not census *counts* for states, counties, tracts, and block groups. (Estimates are subject to error.)
- Sample sizes are smaller than previous decennial census samples, since the data is collected from continuous surveys of the population, not once every 10 years.

Time Frame

Small monthly samples must be pooled to provide suitable estimates for the smallest areas. Areas with populations fewer than 20,000, including tracts and block groups, require 60 months of surveys. Even one-year ACS data (for areas with populations greater than 65,000) requires a 12-month sample. ACS estimates

are all period estimates representing an interval of time, not a single date like April 1, 2010.

Interpreting the change between April 1, 2000, and a five-year average for 2014–2018 may be difficult.

- Annual rates of change cannot be calculated, precluding comparison to any other periods in time.
- Overlapping multiyear periods are likely to challenge data users who try to calculate change between periods.

Sample Size

The much smaller sample sizes of the ACS (1 in 40 compared to 1 in 6 in 2000) affect data reporting and produce much larger sampling errors.

- Smaller sample sizes require less detail in the data reported. For example, age by income in 2000 was reported for seven different age groups (10-year intervals, such as 25–34 years). ACS age by income is reported for four age groups (15–24, 25–44, 45–64, and 65+ years).
- Some values for medians, per capita income, and the aggregates used to determine averages are missing from the ACS database, especially at the block group level.
- Sampling errors must be reported as margins of error, because the variability of the estimates is increased with smaller sample sizes. In some cases, the sampling error can exceed the estimate.

ACS data looks like Census 2000 sample data, but the resemblance is superficial. Continuous measurement and significantly smaller sample sizes yield less precise measures of common variables than Census 2000 sample data. All survey-based estimates are subject to sampling error and uncertainty. Any sample will differ from the total population because it represents just a fraction of the total. Census 2000 sample data represented a larger share of the population, and sampling errors were not reported. However, the Census Bureau deems it necessary to report measures of sampling error with all ACS estimates.

Margin of Error

The margin of error enables data users to measure the range of uncertainty around each estimate. This range can be calculated with 90 percent confidence by taking the estimate +/- the MOE. For example, if the ACS reports an estimate of 100 +/- 20, then there is a 90 percent chance that the value for the total population falls between 80 and 120. The larger the MOE, the lower the precision of the estimate and the less confidence one should have that the estimate is close to the true population value.

The MOE measures the variability of an estimate due to sampling error. Simply, sampling error occurs when only part of the population is surveyed to estimate the total population. There will always be differences between the sample and the total. Statistically, sampling error measures the differences between multiple samples of the same population and differences within a sample of the population. Sampling

error is directly related to sample size. The larger the sample size, the smaller the sampling error. Different areas are sampled at different rates to make the sample representative of the total population. Due to these complex sampling techniques, estimates in some areas have more sampling error than estimates in other areas. All MOEs are approximations of the true sampling error in an area and should not be considered exact. In addition, MOEs do not account for nonsampling error in the data and therefore should be thought of as a lower bound of the total error in a survey estimate.

The ACS reports MOEs with estimates for most standard census geographies. ACS estimates of total population and collapsed age, sex, and Hispanic origin estimates are controlled to annual estimates from the census' Population Estimates Program (PEP) for counties or groups of less populous counties. For the ACS in Puerto Rico, estimates are controlled by age and sex only.[2] Since these estimates are directly controlled to independent estimates, there is no sampling error, and MOEs are zero. However, controlling a period estimate to the average of five-point estimates imparts additional errors in the data that are not measured by MOEs.

In some areas, missing values are prevalent for medians and the aggregate estimates used to calculate averages.[3] When estimates are zero, the Census Bureau models the MOE calculation by comparing ACS estimates to the most recent census counts and deriving average weights for states and the country.[4] At the state, county, tract, or block group level, state-specific MOEs for zero estimates will be the same regardless of the base of the table.

Geography

Most ACS geography corresponds to boundaries as of January 1, 2018. ACS geography is generally consistent with 2010 geography and the areas available with Esri's 2020 Updated Demographics; however, there are differences. The inventory of county subdivisions has changed since 2010, which is included in ACS but not in Esri's updates. The ACS place inventory includes changes since the release of TIGER 2018, while the place inventory in Esri's updates corresponds to TIGER 2019 places.

ACS data for congressional districts represents the boundaries from the 116th Congress. ACS data for Core Based Statistical Areas (CBSAs) reflects definitions from the Office of Management and Budget from August 2017. Esri's updates for CBSAs reflect definitions from September 2018. Due to the large number of differences between the August 2017 and September 2018 definitions, Esri has made ACS data available for the latest CBSA boundaries. CBSA estimates are calculated by summing the estimates of the county components, and MOEs are

[2] A Comparison of the American Community Survey and the Puerto Rico Community Survey, https://www.census.gov/content/dam/Census/programs-surveys/acs/Library/OutreachMaterials/ACSFlyers/A%20Comparison%20of%20the%20ACS%20and%20the%20PRCS.pdf

[3] Please note, with this latest 2014–2018 ACS data release, missing data affects a broader number of variables within an isolated geographic area due to a survey error. ACS suppressed data for select characteristics like labor force and employment, migration, and poverty in Rio Arriba County, New Mexico, due to a data collection error. For more information on this specific issue, please refer to https://www.census.gov/programs-surveys/acs/technical-documentation/errata/125.html.

[4] US Census Bureau, "Variance Estimation," *Design and Methodology American Community Survey* (Washington, DC: US Government Printing Office, 2010), 12-4–12-5.

calculated using the approximation formula.[5] Due to this, estimates and MOEs may vary from what can be found on data.census.gov. For example, estimates and MOEs from data.census.gov reflect a 16-county CBSA for Virginia Beach-Norfolk-Newport News, VA-NC. Esri has made ACS data available at the latest delineation for this CBSA, which includes 19 counties.

Additionally, Esri has made ACS data available for designated market areas (DMAs), ZIP codes, and user-defined polygons. ACS data for ZIP codes is not provided by the Census Bureau, but Esri has created ZIP code data by aggregating the block group level ACS data using a block-to-block group apportionment methodology. ZIP code boundaries are current as of Q4 2019, and the source is HERE. For the US only, Esri produces ACS data for DMAs, representing the 2019–2020 definitions from Nielsen; this data is not provided by the Census Bureau either.

Esri and ACS

Clearly, ACS data differs from the familiar census sample data. To help data users understand the inconsistencies, Esri is providing reports, thematic mapping, and online help. All products include the display of MOEs for the estimates. The reports include two summaries (Population and Housing) for the US and three for Puerto Rico.

Esri's reports/maps are designed to simplify the data and enhance its usability including the following:

- Enhanced geographic coverage: user-defined polygons and ZIP codes
- Reliability thresholds to simplify interpretation of MOEs in summary profiles and mapping

Esri offers the ability to query ACS data for the most popular geographies—user-defined polygons and ZIP codes. Since these areas are not available from the Census Bureau, there are no tabulated MOEs. Estimating data for these custom areas requires aggregation of ACS estimates *and* recalculation of MOEs. Esri has developed algorithms to calculate MOEs using guidelines from the Census Bureau. These algorithms account for full and partial areas within the custom area.

There are several considerations to note when viewing MOEs for custom areas. As the number of estimates involved in the sum of a derived estimate increases, the approximate MOE becomes increasingly different from the MOE that would be derived directly from ACS microdata. The direction of this difference (positive or negative) is based on the correlation and covariance of the estimates. In addition, MOEs are not scalable. MOEs at smaller geographic levels do not add up to MOEs at larger levels. Therefore, analyses should always make use of the largest standard geographic unit possible. For example, if your area of interest includes 90 percent of

[5] Calculating Measures of Error for Derived Estimates, https://www.census.gov/content/dam/Census/library/publications/2018/acs/acs_general_handbook_2018_ch08.pdf

a county, the MOE for the total county will be more accurate than the MOE derived from county parts.

Medians and Averages

A median represents the middle of a distribution. Many variables are reported as distributions with median values such as contract rent, year householder moved in, or year structure built. The Census Bureau estimates medians from standard distributions that are *not* released to the public.[6] Therefore, the bureau's estimated medians will differ from medians that are calculated from the reported tables. For standard geographic areas, Esri displays the medians that are reported by the Census Bureau with its calculations of MOEs. Note that there are missing medians in the Census Bureau's tables, primarily for smaller areas like tracts and block groups. It is possible to find a distribution reported for a given variable, even if the median is missing. If the median is not reported by the Census Bureau for a standard geographic area, then Esri® reports display *N/A*, or not available.

Medians are shown for nonstandard areas like ZIP codes and polygons, which are not available from the Census Bureau. For these areas, Esri calculates the medians from the reported distributions. However, MOEs are not available.

Averages are commonly calculated from the aggregate value of a variable, such as the sum of all contract rent paid or the total number of vehicles reported, divided by the total number of cases (e.g., renter-occupied housing units or households). Aggregates may also be tabulated as missing by the Census Bureau, even if a distribution is reported for the area. If an aggregate value is missing, then an average cannot be determined and will be displayed as N/A whether for standard or nonstandard areas.

Summary Profiles/Mapping: Reliability of ACS Data

The summary reports display MOEs for the estimates plus an additional column that Esri has included to help data users interpret the MOEs relative to the estimates. Decisions about the quality of an estimate based on the MOE alone can be difficult. A reliability symbol is displayed on the reports to give the user some perspective on the MOE. The symbol is based on an estimate's coefficient of variation (CV) and is meant to be used as a quick reference to gauge the usability of an ACS estimate.

The CV is a measure of relative error in the estimate. It measures the amount of sampling error in the estimate relative to the size of the estimate itself. A large amount of sampling error in a small estimate will generally discount the usefulness of the estimate; however, a small amount of sampling error in a large estimate shows that the estimate is reliable.

The reliability is based on thresholds that Esri has established based on the usability of the estimates. Users should be aware that these are generalized thresholds:

[6] For more information on the standard distributions, see the Census Bureau's documentation at https://www2.census.gov/programs-surveys/acs/tech_docs/subject_definitions/2018_ACSSubjectDefinitions.pdf, Appendix A.

-  High Reliability: Small CVs (less than or equal to 12 percent) are flagged green to indicate that the sampling error is small relative to the estimate, and the estimate is reasonably reliable.[7]

-  Medium Reliability: Estimates with CVs between 12 and 40 are flagged yellow—use with caution.

- Low Reliability: Large CVs (over 40 percent) are flagged red to indicate that the sampling error is large relative to the estimate. The estimate is considered very unreliable.

- Some estimates do not indicate reliability. In these cases, either the estimate or MOE is missing, or the estimate is zero.

The amount of acceptable error in an estimate is subjective to the analysis at hand. Data users can compute a CV directly from the MOE; the CV is calculated as the ratio of the standard error to the estimate itself. To get the standard error, divide the MOE by 1.645 (for a 90 percent confidence interval). To calculate a CV, use the following equation:

$$CV = \frac{\left(\frac{MOE}{1.645}\right)}{ESTIMATE} \times 100$$

The CV is commonly expressed as a percentage. For example, if you have an estimate of 80 +/- 20, the CV for the estimate is 15.2 percent. This estimate should be used with caution, since the sampling error represents more than 15 percent of the estimate.

Summary

The American Community Survey is a product of its design. Data users (including vendors) cannot fix the differences that ensue from continuous measurement of the population in lieu of a decennial sample survey. Data users will have to balance the benefits of timely data with the drawbacks of estimate quality. To do this effectively, data users will have to make use of new tools to evaluate the quality of ACS data, such as MOEs, CVs, and tests for significant differences between samples.

In addition to statistical tools, the data user can employ larger areas of analysis or collapse some of the distributions if the reliability of the estimates is a problem. When comparing areas, the Census Bureau recommends focusing on percentages of distributions rather than estimate values.

Changes to the sample size, time frame, data collection, and survey methodology make ACS data something completely different from the sample data previously collected from the decennial census. When the Census Bureau reports sampling error with the survey estimates, it's time to pay attention to the differences.

[7] National Research Council, *Using the American Community Survey: Benefits and Challenges* (Washington, DC: The National Academies Press, 2007).

Glossary

ACS estimates incorporate new definitions that emphasize the importance of the statistical tools that are unique to survey estimates—and key to effective use of the data.

Coefficient of variation (CV): The CV measures the amount of sampling error relative to the size of the estimate, expressed as a percentage. A large amount of sampling error in a small estimate will generally discount the usefulness of the estimate; however, a small amount of sampling error in a large estimate shows that the estimate is reliable.

Confidence interval: The confidence interval is another way to measure the uncertainty of an estimate. The upper bound is the estimate plus the margin of error; the lower bound is the estimate minus the margin of error. (If the lower bound is negative, then zero is assumed for the lower bound.) Confidence intervals for ACS estimates represent a 90 percent certainty that the interval around the estimate includes the true population value.

Margin of error (MOE): The MOE is a measure of the variability of the estimate due to sampling error. MOEs enable the data user to measure the range of uncertainty for each estimate with 90 percent confidence. The range of uncertainty is called the confidence interval, and it is calculated by taking the estimate +/- the MOE. For example, if the ACS reports an estimate of 100 with an MOE of +/- 20, then you can be 90 percent certain the value for the estimate falls between 80 and 120.

Nonsampling error: All other survey errors that are not sampling errors are collectively classified as nonsampling error. This type of error includes errors from interviewers, respondents, coverage, nonresponse, imputation, and processing. Nonsampling error also includes unchecked methodological errors from controlling ACS estimates to independent population estimates.

Period estimates: These are estimates based on data collected over a period of time. ACS five-year data is collected monthly over 60 months and is sometimes referred to as a "rolling survey."

Point estimates: Point estimates are based on data collected at a single point in time. The decennial census refers to April 1 and captures a snapshot of the population at that time.

Reliability: These symbols represent threshold values that Esri has established from the coefficients of variation to designate the usability of the estimates:

 High Reliability: Small CVs (less than or equal to 12 percent) are flagged green to indicate that the sampling error is small relative to the estimate and the estimate is reasonably reliable.[8]

[8] National Research Council, *Using the American Community Survey: Benefits and Challenges* (Washington, DC: The National Academies Press, 2007).

Medium Reliability: Estimates with CVs between 12 and 40 are flagged yellow—use with caution.

Low Reliability: Large CVs (over 40 percent) are flagged red to indicate that the sampling error is large relative to the estimate. The estimate is considered very unreliable.

Residence rules: These rules are used to establish a primary residence to reduce duplication. The ACS defines a resident by a two-month rule. The census rule is "usual place of residence" or wherever a person spends most of the year. ACS data may include seasonal populations in addition to year-round residents.

Sampling error: Errors that occur from making inferences about the whole population from only a sample of the population are collectively referred to as sampling error. Sampling error measures the variability within each sample as well as the variability between all possible samples. All survey data has sampling error.

Statistical significance: Tests for statistical significance are used to determine if the difference between two survey estimates is real or likely due to sampling error alone. Statistical significance is shown at the 90 percent confidence level. Therefore, if estimate differences are statistically significant, there is less than a 10 percent chance that the difference is due to sampling error.

For more information about Esri's ACS data, call 1-800-447-9778.

Esri's Data Development Team

Led by chief demographer **Kyle R. Cassal**, Esri's data development team has a 35-year history of excellence in market intelligence. The team's economists, statisticians, demographers, geographers, and analysts produce independent small-area demographic and socioeconomic estimates and forecasts for the United States. The team develops exclusive demographic models and methodologies to create market-proven datasets, many of which are now industry benchmarks such as Tapestry™ Segmentation, Consumer Spending, Market Potential, and annual Updated Demographics. Esri demographics powers the ArcGIS® platform through dynamic web maps, data enrichment, reports, and infographics.

esri | THE SCIENCE OF WHERE®

Esri, the global market leader in geographic information system (GIS) software, offers the most powerful mapping and spatial analytics technology available.

Since 1969, Esri has helped customers unlock the full potential of data to improve operational and business results. Today, Esri software is deployed in more than 350,000 organizations including the world's largest cities, most national governments, 75 percent of Fortune 500 companies, and more than 7,000 colleges and universities. Esri engineers the most advanced solutions for digital transformation, the Internet of Things (IoT), and location analytics to inform the most authoritative maps in the world.

Visit us at esri.com.

Contact Esri

380 New York Street
Redlands, California 92373-8100 USA

1 800 447 9778
T 909 793 2853
F 909 793 5953
info@esri.com
esri.com

Offices worldwide
esri.com/locations

For more information, visit **esri.com/data/esri_data**.