Understanding and Using the U.S. Census Bureau's American Community Survey

CHAMPAIGN COUNTY rpc REGIONAL PLANNING COMMISSION

The American Community Survey (ACS) is a nationwide continuous survey that is designed to provide communities with reliable and timely demographic, housing, social and economic data. However, sample size becomes a critical issue when interpreting the data. In some cases, unreliable data is reported. In order to understand and use the data appropriately, the Census Bureau provides the Margin of Error (MOE) figure which allows the user to determine the sampling error and relative reliability, calculate an estimate with a different confidence interval, properly aggregate data, and determine statistical significance in the change of an estimate. The following information is available through the American Community Survey.

Demographic Characteristics
Age
Sex
Hispanic Origin
Race
Relationship to Householder
(e.g., spouse)

Economic Characteristics
Income
Food Stamps Benefit
Labor Force Status
Industry, Occupation, and Class of Worker
Place of Work and Journey to Work
Work Status Last Year
Vehicles Available
Health Insurance Coverage*

Social Characteristics
Marital Status and Marital History*
Fertility
Grandparents as Caregivers
Ancestry
Place of Birth, Citizenship, and Year of Entry
Language Spoken at Home
Educational Attainment and School Enrollment
Residence One Year Ago
Veteran Status, Period of Military Service, and VA Service-Connected Disability Rating*
Disability

Housing Characteristics
Year Structure Built
Units in Structure
Year Moved Into Unit
Rooms
Bedrooms
Kitchen Facilities
Plumbing Facilities
House Heating Fuel
Telephone Service Available
Farm Residence

Financial Characteristics
Tenure (Owner/Renter)
Housing Value
Rent
Selected Monthly Owner Costs

<u>Sample sizes and reporting periods for geographies in Champaign County (as of 2010)</u>

	Number of areas related to Champaign County, Illinois	1 year (65,000)	3 year (20,000)	5-year (all areas)
States	1	1	1	1
Congressional Districts	1	1	1	1
Public Use Microdata Areas	?	?	?	?
Metropolitan Statistical Area	1	1	1	1
County	1	1	1	1
Urban Area	1	1	1	1
School Districts	14	1	2	14
Places	23	1	2	23
Townships	30	1	2	30
Zip Code Areas	?	NA	NA	All
Census Tracts	?	NA	NA	All
Census Block Groups	?	NA	NA	All

Margin of Error - Describes the precision of the estimate at a given level of confidence. The Census Bureau reports statistics at a 90 percent confidence level, however there may be instances when a confidence level of 95 or 99 percent is more desirable.

The margin of error (MOE) and a known confidence level is used to interpret the precision of the estimate. The Census Bureau publishes MOE's at a 90% confidence level. This implies a 10 percent chance of incorrect inference for all estimates. Using a 99% confidence interval will implies only a 1 percent chance of incorrect inference. Conversion of published ACS Margin of Error (MOE) can be accomplished using conversion factors and the existing MOE.

Adjustment Factors Associated With Margins of Error for Commonly Used Confidence Levels
90 Percent: 1.645 95 Percent: 1.960 99 Percent: 2.576
Census Bureau standard for published MOE is 90 percent.

To calculate a confidence level of 95% use the following equation:

$$MOE_{95} = \frac{1.960}{1.645} MOE_{ACS}$$

To calculate a confidence level of 99% use the following equation:

$$MOE_{99} = \frac{2.576}{1.645} MOE_{ACS}$$

Standard Error - Measures the variability of an estimate due to sampling. The standard error is mainly used to determine other statistics including the coefficient of variation and statistical significance.

Estimates derived from a sample (such as estimates from the ACS or the decennial census long form) will generally not equal the population value, since not all members of the population were measured in the survey. The standard error (SE) provides a measure of the extent that an estimate derived from the sample survey can be expected to deviate from this population value. Smaller SE values mean that all possible samples would result in similar estimates. Larger SE values mean that different samples may have vastly different estimates.

To calculate the standard error divide the Margin of Error by its adjustment factor. For published ACS data (at the 90 percent confidence level) this is 1.645.

$$SE = \frac{MOE_{ACS}}{1.645}$$

Confidence Interval - The range of values that is expected to contain the average value for the characteristic.

This is useful when graphing estimates to display sample variabilities. The confidence interval is calculated using the MOE. The MOE is added and subtracted from the sample estimate to obtain two numbers representing the confidence interval. For confidence levels other than 90 percent, a different margin of error must be calculated before the confidence interval can be calculated.

Coefficient of Variation or Relative Reliability - Measures the relative precision and provides a more effective measure for determining the usability of an estimate. The lower the coefficient of variation (CV), the higher the reliability of the estimate.

The CV provides a measure of the relative amount of sampling error associated with the sample estimate. This relative measure is useful for comparing usability of a range of estimates. For example, large populations may have larger margins of error than small populations but be more reliable because the variation is smaller. The CV allows the data user to asses if a set of estimates have comparable reliability. X is the estimate value in the equation below.

$$CV = \frac{SE}{\hat{X}} \times 100$$

Margin of Error for Aggregated Count Data -

The ACS allows the use of unique estimates called derived estimates. These are generated by aggregating reported estimates across geographic areas or population sub groups. Margin of error is not provided for aggregated estimates and therefore needs to be calculated. This is calculated by square root of the sum of squared margin of errors. The letter 'c' in the equation below represents each estimate that will be included in the aggregation.

$$MOE_{agg} = \pm\sqrt{\sum_c MOE_c^2}$$

Other Calculations

Similar to the aggregated count data, margin of error must be calculated when using sample data for proportions, ratios, products, and percent change. The following equations should be used to calculate the margin of error when calculating estimates.

Derived Proportion

Required information

- $Estimate_{(num)}$, $Estimate_{(den)}$
- MOE_{num}, MOE_{den}

1) Obtain the MOE for each estimate.
2) Divide the estimates to calculate the derived proportion.
3) Square the MOE_{num}, the MOE_{den}, and derived proportion.
4) Multiply the squared MOE_{den} and the squared derived proportion.
5) Subtract the result of (4) from the squared MOE_{num}.
6) Calculate the absolute value of the square root for the result of step (5).
7) Divide the result of (6) by the estimate that was used as the denominator for the proportion.

Caution: There are rare instances where this formula will fail (the value under the square root will be negative). If that happens, use the formula for derived ratios in the next section which will provide a conservative estimate of the MOE.

$$MOE_p = \frac{\pm\sqrt{MOE_{num}^2 - (\hat{p}^2 * MOE_{den}^2)}}{\hat{X}_{den}}$$

where MOE_{num} is the MOE of the numerator.

MOE_{den} is the MOE of the denominator.

$\hat{p} = \frac{\hat{X}_{num}}{\hat{X}_{den}}$ is the derived proportion.

$\hat{X}_{num}$ is the estimate used as the numerator of the derived proportion.

$\hat{X}_{den}$ is the estimate used as the denominator of the derived proportion.

Derived Ratios - Estimate 1 / Estimate 2

Required information

- $Estimate_{(num)}$, $Estimate_{(den)}$
- $MOE_{(num)}$, $MOE_{(den)}$

1) Obtain the MOE for each estimate.
2) Square the ratio of the estimates (R), the MOE_{num} and the MOE_{den}.
3) Multiply the squared MOE_{den} by the squared ratio of the estimates.
4) Add the result of (3) to the squared MOE_{num}.
5) Take the square root of the result of (4).
6) Divide the result of (5) by the estimate that was used as the denominator for the ratio.

$$MOE_R = \frac{\pm\sqrt{MOE_{num}^2 + (\hat{R}^2 * MOE_{den}^2)}}{\hat{X}_{den}}$$

where MOE_{num} is the MOE of the numerator.

MOE_{den} is the MOE of the denominator.

$\hat{R} = \frac{\hat{X}_{num}}{\hat{X}_{den}}$ is the derived ratio.

$\hat{X}_{num}$ is the estimate used as the numerator of the derived ratio.

$\hat{X}_{den}$ is the estimate used as the denominator of the derived ratio.

Product of Estimates - Estimate 1 x Estimate 2

Required information

- Estimate(a) , Estimate(b)
- MOE(a) , MOE(b)

1) Obtain the MOE for each estimate.
2) Square the estimates and MOE's.
3) Multiply the first estimate² by the second estimate's MOE².
4) Multiply the second estimate² by the first estimate's MOE².
5) Add the results from (3) and (4).
6) Take the square root of (5).

$$MOE_{A\times B} = \pm\sqrt{A^2 \times MOE_B^2 + B^2 \times MOE_A^2}$$

where A and B are the first and second estimates, respectively.

MOE_A is the MOE of the first estimate.

MOE_B is the MOE of the second estimate.

Percent Change - Estimate 2 / Estimate 1

Required information

Sample Estimate(X_1) , Sample Estimate(X_2)
MOE_1 , MOE_2

1) Determine the percent change of the two estimates by dividing $estimate_2$ by $estimate_1$.
2) Square MOEs for each estimate and the percent change value (R).
3) Multiply the squared MOE_{den} by the percent change squared, add it to the squared MOE_{num}.
4) Take the square root of (3), divide by $estimate_{den}$.

$$\hat{R} = \frac{\hat{X}_2}{\hat{X}_1}$$

$$MOE_R = \frac{\pm\sqrt{MOE_{num}^2 + (\hat{R}^2 * MOE_{den}^2)}}{\hat{X}_{den}}$$

Determining Statistical Significance

Required information

- *Estimate(X_1) , Estimate (X_2)*
- *Error (SE_1) , Error (SE_2)*
- *Critical Value for Confidence Level (Z)*

1) Calculate the SE for each estimate (positive MOE divided by their adjustment factor value).
2) Square the SE's.
3) Sum the squared SE's.
4) Calculate the square root of the sum of the squared SE's.
5) Calculate the difference between the two estimates.
6) Divide (5) by (4).
7) Compare the absolute value of (6) with the critical value for the desired level of confidence (1.645 for 90 percent, 1.960 for 95 percent, 2.576 for 99 percent).
8) If this value is greater than the critical value, the difference between the two estimates can be considered statistically significant at the selected confidence level.

$$\left| \frac{\hat{X}_1 - \hat{X}_2}{\sqrt{SE_1^2 + SE_2^2}} \right| > Z_{CL}$$

Practical Application of this Guide

Income

An example showing the use of these statistical measures to determine the usefulness of ACS income estimates is provided below.

1. Estimates and MOE's are obtained directly from the Census Bureau's American Fact Finder website.

2. The CV provides a statistic used to determine the reliability of an estimate. Values in the example are below 20, so they are considered to have good reliability. SE is used to calculate CV

3. The z-value from the statistical significance calculation is compared to the critical value (1.645). Values greater than the critical value are significantly different. In this example Median family income increased between 2008 and 2010 while median household income did not.

Measure of Income	2008		2010	
	Estimate	MOE	Estimate	MOE
Median Household Income	43,985	+/-3,749	45,254	+/-2,735
Median Family Income	62,477	+/-4,125	69,580	+/-3,963

Calculate Reliability	2007		2010	
	SE	CV	SE	CV
Median Household Income	2,279	5.18	1,663	3.67
Median Family Income	2,508	4.01	2,409	3.46

Statistical Significance			
	Num	Den	=z
Median Household Income	1,269	2,821	0.45
Median Family Income	7,103	3,478	2.04

Cautions

Comparison within the same time period
If one estimate includes another estimate as a subset (Population of State and Population of a County in that State), the statistical test may incorrectly find a lack of statistical significance. If the two estimates are strongly correlated, it is acceptable to ignore the partial dependence. However, if a more exact test of significance is necessary, one must account for correlation as well.

Comparison across time periods
Users should not compare single-year estimates with multiyear estimates or comparing multiyear estimates of different lengths. Another issue is that group quarters population are only included in estimates after 2005. This is of primary importance for areas with significant group quarters populations.

Comparison of overlapping periods
Ideally, comparisons are made on nonoverlapping years. For example, comparison of 2005-2007 and 2006-2008 data include 2006 and 2007 data in both estimates. The contribution of the overlapping years is subtracted when the estimate of differences is calculated. This calculation can be used but with caution. It should not be interpreted as a reflection of change between the 2 last years. The following equation can be used to account for the simple overlap. C is the fraction of overlapping years. For 2005-2009 and 2007-2011 C = 3/5 = 0.6.

$$SE(\hat{X}_1 - \hat{X}_2) \cong \sqrt{(1-C)}\sqrt{SE_1{}^2 + SE_2{}^2}$$

Comparison with 2000 Census Data

A conservative approach to testing for statistical significance when comparing ACS and Census 2000 estimates that avoids deriving the SE for the Census 2000 estimate would be to assume the SE for the Census 2000 estimate is the same as that determined for the ACS estimate. The result of this approach would be that a finding of statistical significance can be assumed to be accurate (as the SE for the Census 2000 estimate would be expected to be less than that for the ACS estimate), but a finding of no statistical significance could be incorrect. In this case the user should calculate the census long-form standard error and follow the steps to conduct the statistical test.

Comparison with 2010 Census Data

The critical factor that must be considered when comparing ACS estimates encompassing 2010 with the 2010 Census is the potential impact of housing and population controls used for the ACS. As the housing and population controls used for 2010 ACS data will be based on the Population Estimates Program where the estimates are benchmarked on the Census 2000 counts, they will not agree with the 2010 Census population counts for that year. The 2010 population estimates may differ from the 2010 Census counts for two major reasons—the true change from 2000 to 2010 is not accurately captured by the estimates and the completeness of coverage in the 2010 Census is different than coverage of Census 2000. The impact of this difference will likely affect most areas and states, and be most notable for smaller geographic areas where the potential for large differences between the population controls and the 2010 Census population counts is greater.