Tech Talk: Aggregating Data from the American Community Survey – from the January 2011 CSDC Network News

With the release of the first 5-year American Community Survey (ACS), many users will want to aggregate small levels of geography into larger areas. In fact, the Census Bureau encourages users to aggregate census tract or smaller areas to increase the data reliability rather than to look at the small areas individually.

When you aggregate areas, you essentially create a new estimate and should recalculate the margin of error (MOE) for that estimate. There are many resources available to assist you:

Compass Handbooks:
The ACS Compass Handbooks from the Census Bureau discuss the formulas to recalculate the MOE for aggregated counts, derived proportions (percentages), and derived ratios (such as, persons per household or mean income) in Appendix 3. They are online at http://www.census.gov/programs-surveys/acs/guidance.html

The Search for Significance:
A discussion of measuring and using a test of significance is online at www.incontext.indiana.edu/2008/nov-dec/4.asp

Statistical Calculators:
If you are only aggregating up to one or two areas, look into "statistical calculators" to do the calculations for you. These calculators are Excel files that already include the formulas – you just enter the estimate values. One of the better ones is online at:

> From the New York State Data Center – www.trbcensus.com/notes/StatisticalCalculationsMenu.xls

However, none of these resources discuss how to develop estimates of medians, such as median household income. Medians are tricky - you cannot just "add them up" when you aggregate geographic areas. Medians are not arithmetically derived, so although you can create an average of the medians, it is not a true median. Median will need to be recalculated for each aggregated area using data in detailed tables by ranges. We have prepared an article, which is attached with this Newsletter, about how to estimate a median from data in ranges and how to calculate the MOE. We have also developed an Excel "Median Calculator" which is also attached.

A key thing to remember - When analyzing data using an MOE calculated from a formula, you should be aware that formulas are actually approximations that overstate the MOE compared to the more precise methods based on actual survey returns used by the Census Bureau. Therefore, the calculated MOEs will be higher, or more conservative, than those found in the published tabulations for like-sized areas. This knowledge may affect the levels of error you are willing to accept when making a decision on whether to use a particular figure.