Working With the American Community Survey: A Reference Guide for Congressional Staff

Prepared for a Brookings Briefing on the Census

Know Your Constituency:

Congressional District and State Profiles From the 2005 American Community Survey

September 28 and 29, 2006

Cynthia Taeuber

CM Taeuber & Associates

Table of Contents

Section One: Overview of the American Community Survey 1
- Introduction
- When Data Will Become Available
- Geographic Concepts
- Useful References

Section Two: Finding Thematic Maps from American FactFinder 8

Section Three: Understanding Margins of Error 9

Section Four: Comparing Census 2000 and ACS 2005 14
- General Comparability
- Item-Specific Comparability

Section One: Overview of the American Community Survey

Adapted from: "American Community Survey Data in Community Planning," by Cynthia M. Taeuber, forthcoming.

See: http://www.census.gov/acs/www/SBasics/What/What1.htm

Introduction

The American Community Survey (ACS) was developed by the Census Bureau to replace the decennial census "long form," that is, the demographic, social, economic, and housing characteristics of the population and housing stock that is required by hundreds of federal laws and court cases.

- The ACS is the decennial long form spread out over 10 years – that is, the data collection occurs throughout the decade rather than just once in ten years. Traditionally, the states and communities have had to use an aging snapshot for 10-to-12 years until the next census was taken.

- The ACS replaces the old snapshot with a video. Eventually, the American Community Survey will provide estimates, updated *every year*, of the distribution of *characteristics* of the population and housing in geographic areas.

- The ACS will provide what communities have lacked to inform policy and evaluate results, that is, current estimates of population and housing *characteristics*, and measurements of the level and direction of *change*.

- The questions and the data products are essentially the same as those you saw for Census 2000.

How many people? What are they like?

Be sure you got the point that, starting in 2010, the decennial census counts people and housing units only. The American Community Survey does not count -- it is a sample that gives you estimates about the characteristics of the population and the housing stock – what the survey estimates they are like.

In the years between the censuses, the Census Bureau's has a partnership program with the states to estimate how the number of people and housing units change in counties and states (see: http://www.census.gov/popest/estimates.php). Between censuses, the ACS may improve current estimates of the *number* of people in small areas (such as census tracts and school districts) by furnishing current demographic distributions for use in the population estimate models.

2010 Census – counts people and housing units

Population estimates – estimates number of people

American Community Survey – estimates demographic, social, economic characteristics of people and the housing stock

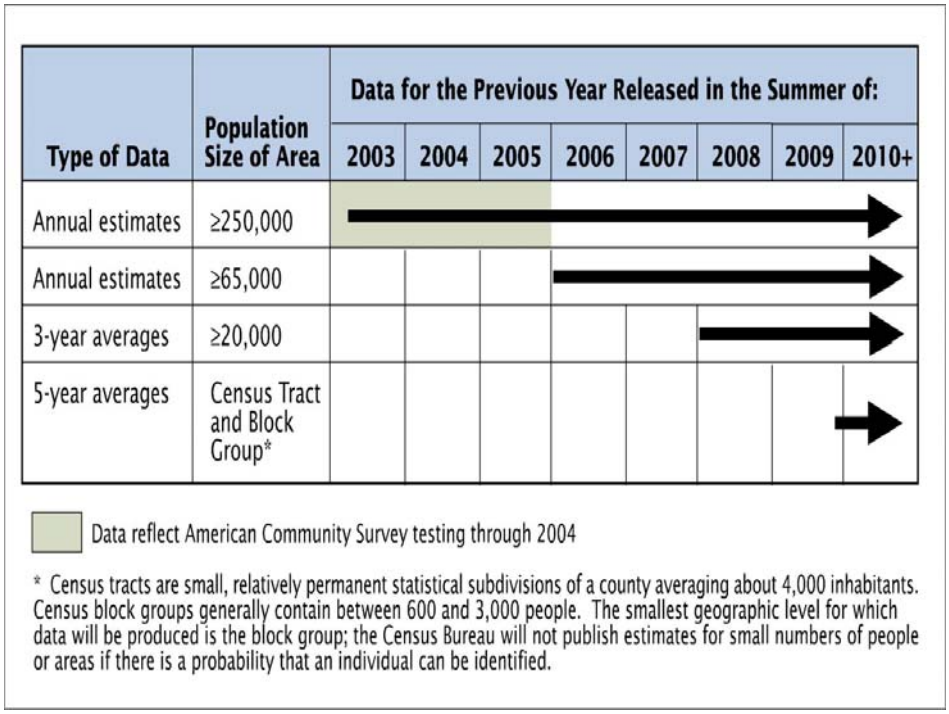
The ACS provides profiles of *small population groups* in states and regions, such as poor children under age 5, or teenage mothers and whether they are in school, working or unemployed. With the ACS, you will be able to track trends and the direction of change for population groups.

When Data Will Become Available

Areas will receive annually updated estimates based on population size. The chart below shows when ACS data will be released for different size areas.

- Beginning with the release in 2006, and every year thereafter, the ACS will have single-year population and housing profiles for areas and population groups of 65,000 or more people. Data covering each calendar year will be released in the second half of the next calendar year. For example, 2005 ACS data are being released in August-November 2006.

- For smaller areas and population groups, it will take 3 to 5 years to accumulate data to provide accurate estimates. Information for areas of 20,000 or more people will be available starting in 2008 (that is, the 3-year average for the data collected over the period 2005-2007). The information will be updated every year thereafter. Information for areas of less than 20,000 people first will become available in 2010 (the 5-year average for data collected over the years 2005-2009).
- After 2010, single-, 3-year, and 5-year averages will be updated every year. Then the United States will finally have a community information resource that shows change over time.



In 2010 –

Areas with 65,000 or more people will have three estimates:

- 1-year averages for 2009
- 3-year averages for 2007 – 2009
- 5-year averages for 2005 – 2009

Areas with 20,000 to 64,999 people will have two estimates:

- 3-year averages for 2007 – 2009
- 5-year averages for 2005 – 2009

Areas with less than 20,000 people will have one estimate:

- 5-year averages for 2005 – 2009

In the 2011 releases, the averages will be updated. For example, the 5-year averages will be for data collected from 2006 – 2010. Notice that four of the five years are the same as those as the 2010 release and so the estimates will be highly correlated.

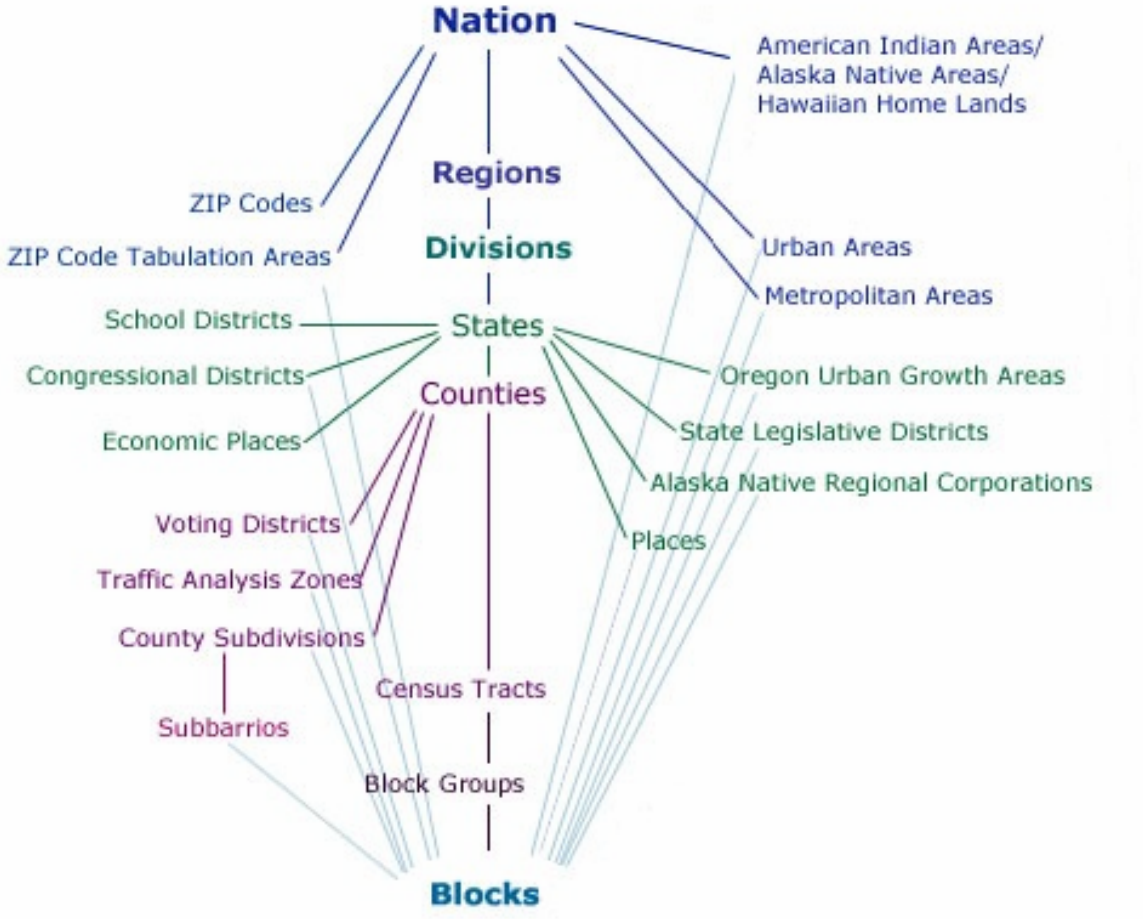
Areas with more than one average will have analytic flexibility and more information to understand trends than they ever had before. Before the ACS, communities had to rely on decennial census information that was 12 years old before it was replaced, and the intercensal population estimates of changes in the number of people.

The old information from the decennial census had an obvious problem – it was old. Even when the nation is said to be "doing well" economically, for example, there are parts of the country that are not doing well. The extent and distribution of trouble spots was not easy to determine from the statistical system of the past. Some likened the situation to putting one's head in a freezer and feet in a fire and then concluding, "On average, I feel just fine." The problem is how to determine where the fire is and where the freezer is – that is, which areas need what, how much, and when. Communities have been relying on historical data, acting reactively rather than proactively. They need a modern information system to anticipate or prevent situations rather than try to mop up after the fact. With an information system that was not current, of high quality, and comparable across areas, it has often been difficult to adjust policies in time to meet needs, establish priorities, evaluate results, and plan for the future.

Updated information every year opens new possibilities for using data in ways that are different from how the historical data from the long form have been used. Researchers are constantly developing ways to use the updated trends provided by the ACS in needs assessments, in econometric and predictive models, and to improve estimates of characteristics such as disability and poverty. Geographic Information Systems (GIS) can use the population and housing information to help community officials visualize differences among areas and to actively manage programs for results.

Geographic Concepts

The Census Bureau's diagram below shows geographic relationships:
(http://factfinder.census.gov/home/en/epss/census_geography.html)



You can find the official, detailed definitions of geographic terms and concepts at:
http://www.census.gov/geo/www/tiger/glossry2.pdf.

This is a good place to go to learn, for example, the difference between "urban" and "metropolitan" and their counterparts, rural and nonmetropolitan.

Useful References

American Community Survey website (homepage): http://www.census.gov/acs/www/

American Community Survey User's Guide for each year – under the "Using the Data" tab on the ACS website. For 2005:
http://www.census.gov/acs/www/Products/users_guide/index.htm

American Community Survey Questionnaires:
http://www.census.gov/acs/www/SBasics/SQuest/SQuest1.htm

American Community Survey Glossary:
http://www.census.gov/acs/www/UseData/Def.htm

Subject Definitions: http://www.census.gov/acs/www/UseData/Def.htm

Group quarters: *Background* -
http://www.census.gov/acs/www/SBasics/GQ/index.htm

and *GQ definitions* - http://www.census.gov/acs/www/UseData/GQ/def.htm

Accuracy of the Data (American Community Survey)
http://www.census.gov/acs/www/UseData/Accuracy/Accuracy1.htm

Data Quality (American Community Survey):
http://www.census.gov/acs/www/UseData/sse/index.htm

Nonsampling Errors in the American Community Survey:
http://www.census.gov/acs/www/Downloads/ACS/accuracy2003.pdf

Response rates in the American Community Survey:
http://www.census.gov/acs/www/UseData/sse/index.htm

Operation Plan for the American Community Survey:
http://www.census.gov/acs/www/SBasics/op_plan.htm

Puerto Rico ACS: *Basics* - http://www.census.gov/acs/www/SBasics/FlyerPR.htm
Group Quarters- http://www.census.gov/acs/www/SBasics/GQ/index.htm

Public Use Microdata, American Community Survey:
http://factfinder.census.gov/home/en/acs_pums.html

Publications about the ACS:

http://www.census.gov/acs/www/AdvMeth/Papers/Papers1.htm

American FactFinder:
http://factfinder.census.gov/home/saff/main.html?_lang=en

American FactFinder Help: http://factfinder.census.gov/home/en/epss/toc.html

2000 Census Gateway: http://www.census.gov/main/www/cen2000.html

2000 Questionnaire: http://www.census.gov/dmd/www/2000quest.html

2000 Census Glossary:
http://factfinder.census.gov/home/en/epss/glossary_a.html

Accuracy of the Data, Census 2000, Chapter 8,
www.census.gov/prod/cen2000/doc/sf3.pdf. To compute confidence intervals, you will need to use Table C for the design factors but the documentation on that site neglected to provide the link which is, for Arizona, www.census.gov/prod/cen2000/doc/tablec-az.pdf. For other states, change the two-character state abbreviation "az" in the URL.

Brief, Overview of Race and Hispanic Origin
http://www.census.gov/prod/2001pubs/c2kbr01-1.pdf

Overview of Race and Hispanic Origin
http://www.census.gov/population/www/cen2000/briefs.html

Population Estimates: http://www.census.gov/popest/estimates.php

State Data Centers (many develop convenient tables from the decennial census and American Community Survey for their states that can be downloaded into spreadsheets):
http://www.census.gov/sdc/www/

Census Information Centers (show data for specific population groups):
http://www.census.gov/contacts/www/c-cics.html

Geography, census – the framework on which the data rest:
http://www.census.gov/geo/www/index.html

Geographic Terms and Concepts: http://www.census.gov/geo/www/tiger/glossry2.pdf

Geographic Types in the American FactFinder:
http://factfinder.census.gov/home/en/epss/census_geography.html

Census 2000 Geographic Products and Information:
http://www.census.gov/geo/www/census2k.html

Geographic Reference Resources: http://www.census.gov/geo/www/reference.html

Reference Maps - see the "User's Guide" for each year under the "Using the Data" tab. For 2005:
http://www.census.gov/acs/www/Products/users_guide/acs_2005_reference_maps.htm

FIPS Codes for Places, Counties, Congressional Districts, and More Use to attach county and place names, for example (places, such as Portland, OR, can cross county boundaries): http://www.census.gov/geo/www/index.html

U.S. Gazetteer Files - search for a place or ZIP code

for 2000: http://www.census.gov/geo/www/gazetteer/places2k.html

for 1990: http://www.census.gov/cgi-bin/gazetteer\

County and City Data Book: 2000 -- http://www.census.gov/prod/www/ccdb.html

Historical statistics: http://www2.census.gov/prod2/statcomp/

Historical population counts from decennial censuses:
http://www.census.gov/population/www/censusdata/hiscendata.html

Other Census Bureau Demographic Surveys:
http://www.census.gov/main/www/sur_demo.html

Census Bureau Catalog: http://www.census.gov/prod/www/abs/catalogs.html

Other federal data sets: www.fedstats.gov

Statistics, Third Edition by David A. Freedman, Robert Pisani, and Roger Purves, WW Norton, NY, 1998. The 4th edition will be available in 2007. This is my favorite statistics book for understanding the logic behind statistical techniques and how to do them. Freedman, et.al. write in plain English and include many examples. There is extensive discussion of sampling and non-sampling errors and confidence intervals. It is a low-tech approach to learning statistics with words and pictures and no formulas.

Section Two: How to Find Thematic Maps from American FactFinder

http://factfinder.census.gov

Main >

Data Sets >

American Community Survey 2005 >

Thematic Maps >

Geography type (State) >

Highlight state name - Next >

Select Themes – Show Result >

Display map by - 109th Congress

Section Three: Understanding Margins of Error

The American Community Survey profile report below (Figure 1) shows educational attainment for people 25 years and older, estimates of the number of people who have attained increasingly higher levels of education. In this Congressional District, for example, an estimated 42,201 people got past the 8th grade but did not receive a high school diploma or its equivalent.

Notice the second column marked "Margin of Error." That's there because these are estimates from a sample survey and not a full count from a census. The margin of error is the plus or minus – in this example, the estimate is 42,201 people plus or minus 3,716 people. You need to know about the margins of error of estimates so you don't make a dumb mistake by saying two areas are different when they aren't, or that something has changed when it hasn't.

I want to offer a second strong suggestion—show tables and graphs as percentage distributions rather than just as numbers. These are estimates, not exact counts and it's easier to make comparisons with percentages. You can download the data into a spreadsheet so you can let software do the arithmetic. Just click in the upper right corner of the page showing the data.

To get those percentages, just divide the category's estimate by the total. For example, the percentage of people 25 years and older in this District who dropped out of school between the 9th and 12th grade is estimated to be 10.1 percent. Notice I also added that it was plus or minus 0.9 percentage points—that means you can be highly confident that the true percentage of high school graduates in this District is 9.2 to 11.0 percent of the population 25 years and older. You don't necessarily need to show the range, but YOU need to know it. Knowing the range of the sampling error helps you evaluate the usefulness of the estimate for your purposes – for small areas or population groups, the ranges can be large – sometimes too large. You also need to know how the range of the sampling error to compare this District with other Districts. In Figure 2 below, you can see my approach to computing this range. The Brookings web site provides access to Figure 2 as an Excel spreadsheet, with embedded formulas for you to use.

Here is what we learn by doing this: About 4 percent of adults in this District are functionally illiterate with less than a 9th grade education. Another 10 percent dropped out of high school sometime after 8th grade – they might be a target group for GED classes and then a community college. About 1 in 3 people have only a high school degree – another target group for a community college. About 1 in 4 people in this Congressional District have at least a bachelor's degree – 15 percent with a bachelor's degree and another 9 percent with a graduate or professional degree.

Figure 1: American Community Survey Profile Report

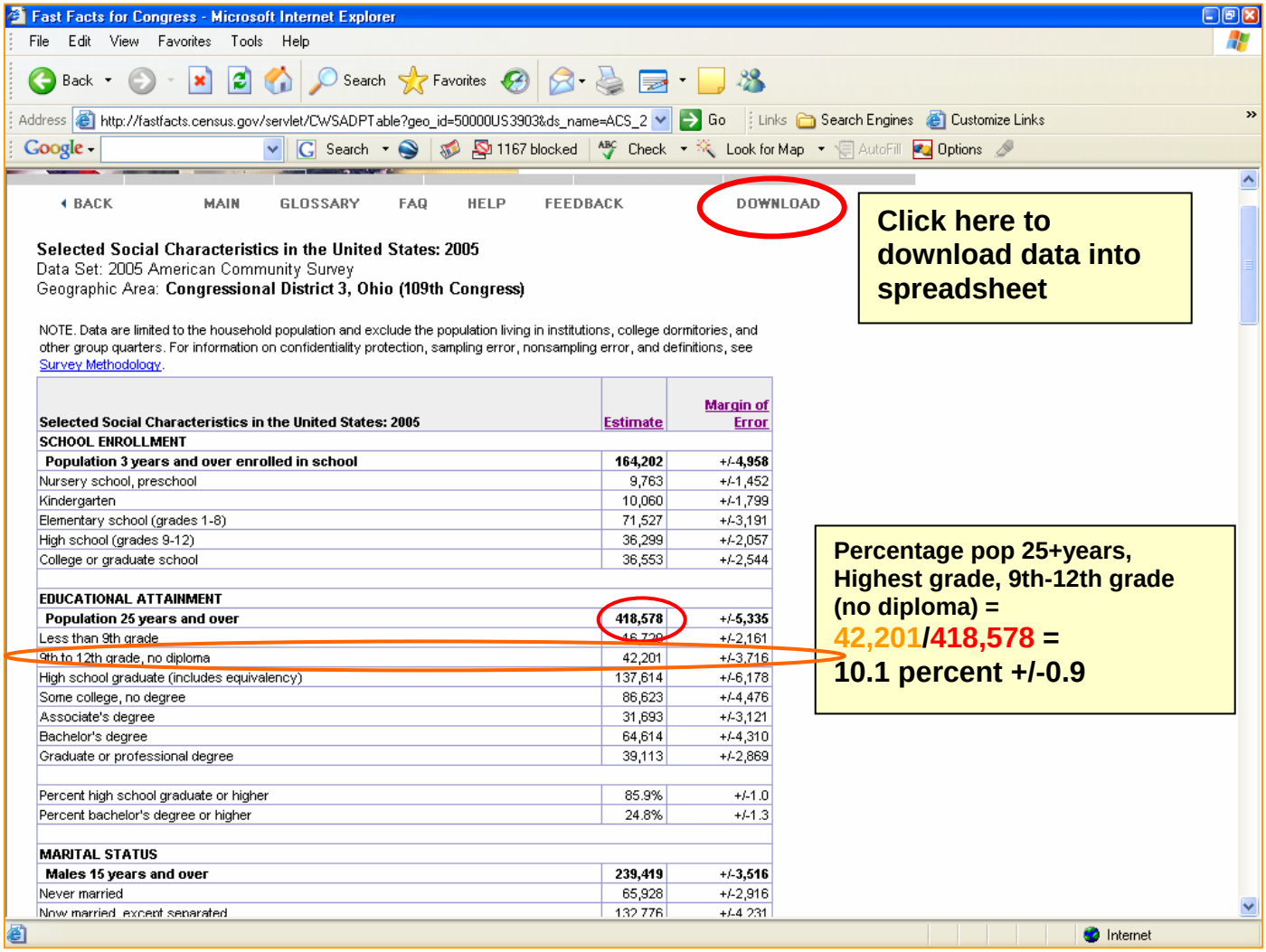
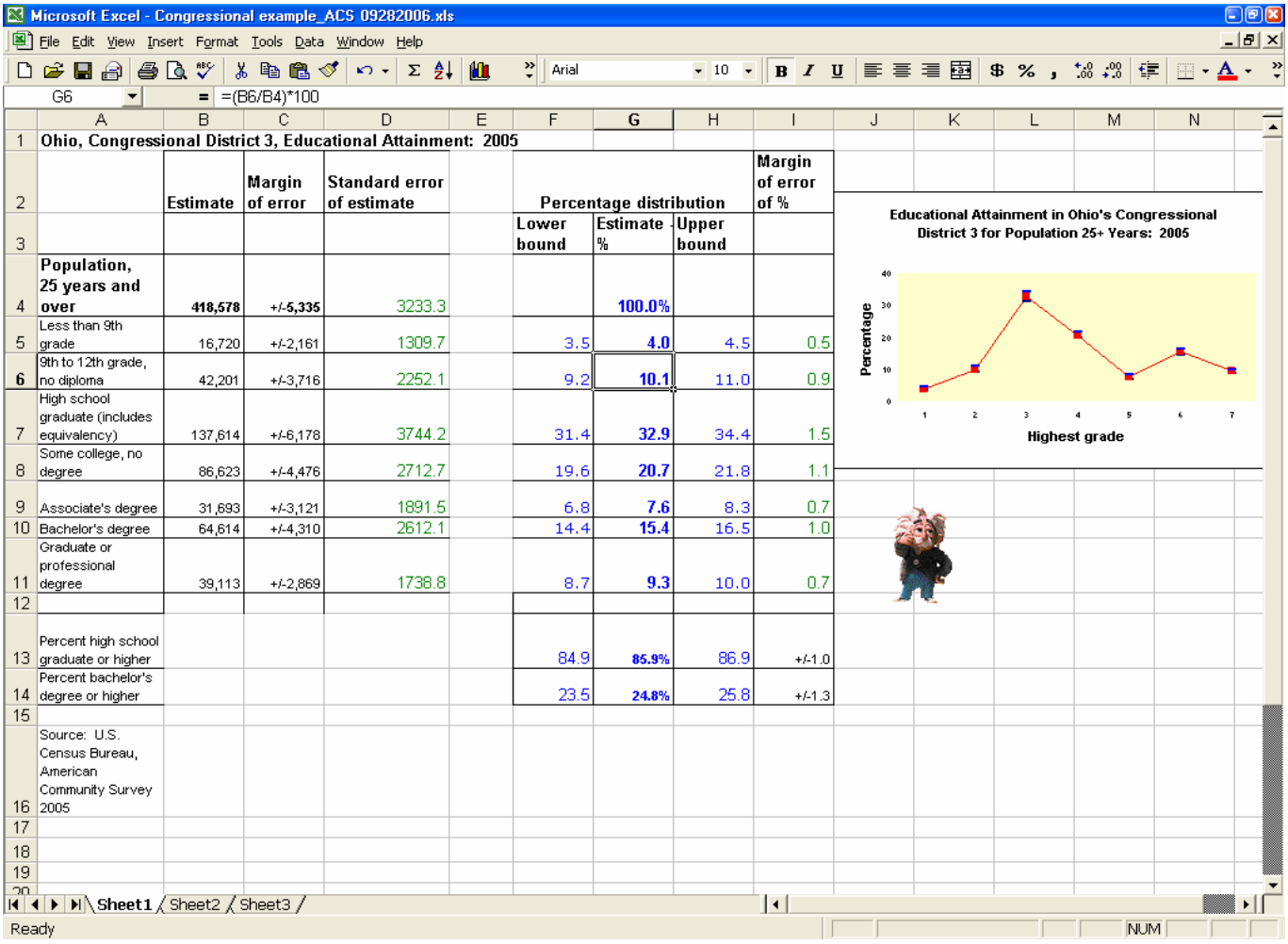


Figure 2: Computing Margin of Error



The best way to understand the data is to look at the percentage distributions rather than just the numbers. To do that, I downloaded the data from the FastFacts page into a spreadsheet. Figure 2 shows what we have for the distribution of educational attainment in Ohio's Congressional District 3. Again, the Excel spreadsheet on the Brookings website provides the formulas for computing percentage distributions and margins of error.

Notice that the working graph shows you the estimates and the margins of error so you know what the range is for an estimate given the sampling error—that's the lower and upper bounds of what statisticians call the "90-percent confidence interval." Because these are data from a sample rather than a full count, we say we are highly confident—there are 9 chances in 10—that the value you would have gotten from a complete count lies within this range. It's nothing heavy—just a little arithmetic that the computer can do.

When we want to understand what is going on in an area, we compare its current estimate with those for other areas, or for itself over time. The heart of safely making comparisons—of saying that areas are significantly different from each other or that an area significantly changed over time—is to do some easy arithmetic. This is that hardest math you have to do—add and subtract the margin of error from the estimate so you can get the range within which the estimate falls. A good rule of thumb: If one estimate range overlaps another, don't say they are significantly different.

Figure 2 shows that in 2005 in Ohio's Congressional District 3, an estimated 10.1 percent of people 25 and older had completed at least 9th grade, but left school before they received a diploma. The margin of error is given as 0.9 and so we compute the range of the estimate:

Find the range of the estimate

LOWER bound of confidence interval =
Estimate - Margin of error
10.1 – 0.9 = 9.2

UPPER bound of confidence interval =
Estimate + Margin of error
10.1 + 0.9 = 11.0

In our example, 10.1 is the estimate and the margin of error is 0.9, so the range of the estimate is 9.2 to 11.0.

Because these are estimates and not exact counts, when you are writing up the results, use the word "about" and round off the numbers. In this example, you could say "In 2005, about 1 in 10 people in this Congressional District dropped out of high school sometime after the 8th grade and never received a high school diploma or its equivalent."

Figure 3 is your cheat sheet that summarizes the information you need to compute sampling error so you can make comparisons between different areas and over time using data sets from the American Community Survey. In the ACS products through 2004, the Census Bureau shows the upper and lower bounds of the range – you don't have to compute anything. In the 2005 products, they show the margin of error only. That's useful if you are looking at just one number, but if you want to make comparisons, you need to compute the range of the estimates, the confidence intervals.

Figure 3: Summary of Formulas Related to Sampling Error

Upper bound of confidence interval = Estimate + Margin of error (i.e., sampling error)

Lower bound of confidence interval = Estimate - Margin of error

Margin of error = Upper bound of confidence interval – estimate

Margin of error = Estimate - Lower bound of confidence interval

Margin of error = Standard error * 1.65

Section Four: Comparing Census 2000 and ACS 2005

Prepared by New York State Dept of Economic Development State Data Center

General Comparability

Issue	Census 2000	ACS 2005
Universe	• Most tables include ALL residents (there may be age, race, etc. restrictions) • Some tables are limited to the household population • If geography is correct, PUMS or Advanced Query System can be used to generate tables limited to the household population but there are other issues.	• Tables are limited to the HOUSEHOLD POPULATION ONLY (there may be additional restrictions such as age or race) • *This will have a greater impact on areas with large group quarters facilities.*
Residence Rules	• Usual place of residence – Self identification of where you live most of the year. College students are supposed to be counted at their college address. • Does not capture seasonality or second home location.	• Current Residence – Counted at the sampled address if lived there most of the time in the last TWO MONTHS. College students might be counted at parent's house during the later summer months. • *This will have a greater impact on areas with large seasonal populations such as college and resort communities.* • Seasonal destination areas may have population counted that would have lived elsewhere April 1, Census.

Collection Procedures Non-Response Follow-up	• Primarily personal visit by short-term, moderately trained employees, being paid a relatively low wage for the area. • Emphasis on counting number of people at address.	• Conducted by long-term, highly-trained employees, being paid reasonably well for the area. • Telephone follow-up in second month if possible. Personal visit in third month usually to 1/3 of addresses not responding. • Emphasis is on collecting characteristics of individuals in the household.
Proxy Respondents	• Allowed - People living outside the address (landlords, neighbors, etc.) are allowed to provide information.	• NOT Allowed - ONLY people living at the address can answer the questions.
Reliability	• Statistical reliability is generally not reported, but can be calculated • Generally, fairly small compared to the estimate. (Confidence intervals on data from PUMS will be larger.)	• Reported as Margin of Error (MOE). Must be calculated for user derived data • Can be VERY large compared to the estimate. • MUST be used for comparing areas or change over time.

Item-Specific Comparability

Total Population – Not Comparable – Unless otherwise noted, the 2000 Census reports the total resident population. The ACS reports only the household population. The ACS does NOT yet include group quarters.

Age – Concept is comparable but the Census reports age as of April 1, 2000. The ACS reports age as of the survey month.

Gender - Comparable

Race – Comparable

Hispanic or Latino Origin – Comparable

Household Relationship – Comparable though some categories are different. The Census distinguishes between Natural-born, Adopted, and Step sons and daughters while the ACS has only one category "Son or Daughter". The Census also distinguishes between Parent-in-low and Son or Daughter-in-law while the ACS includes only the category "In-Law".

Average Household/Family Size – Comparable though the different residency rules will likely impact household and family size.

School Enrollment – Concept is comparable but the reference period for Census is April 1, 2000 while the ACS reference period is the survey month. May also be affected by the ACS not including group quarters population.

Educational Attainment – Concept is comparable but the reference period for Census is April 1, 2000 while the ACS reference period is the survey month.

Marital Status – Concept is comparable but the reference period for Census is April 1, 2000 while the ACS reference period is the survey month. May also be affected by the ACS not including group quarters population.

Fertility – Not included in the 2000 Census

Grandparents – Comparable

Veterans Status – Comparable

Disability - Not comparable because the question was redesigned and because of the lack of non-institutional group quarters population in the 2005 ACS.

Residence 1 Year Ago – Not comparable because the Census question related to residence 5 years ago.

Place of Birth/Nativity – Comparable

Citizenship – Comparable

Year of Entry – Comparable

Region of Birth - Comparable

Income – Not Comparable – While the concepts are similar, there are several issues that make the data not comparable: differences in the time periods for which data are collected in the ACS versus the Census; adjustments for inflation in the ACS data; accuracy of the respondents' answer; and the rates of imputation when the Census Bureau cannot get answers to these questions.

> *Reference Period* – The 2005 ACS asks respondents for their income over the 12 months prior to completing the questionnaire. The 2000 Census asks respondents about their income in calendar year 1999.
>
> *Inflation Adjustment* – Since the income data on the ACS is collected over an entire year, it refers to incomes received over a 23 month period (12 months prior to January 2005 through 12 months prior to December 2005). The Census Bureau adjusts incomes to represent the same time period using the Bureau of Labor Statistics' Consumer Price Index for all urban consumers – research series (CPI-U-RS).
>
> *Accuracy of the Respondent's Answer* – Respondent accuracy can depend on the relative stability of the respondent's income and their ability to recall changes, especially if there are major fluctuations in their income. The 2000 Census asks about income in 1999 at a time when most respondents have the information needed to complete their income taxes. It may be more difficult for an ACS respondent to recall income over the previous 12 months.
>
> *Imputation of Non-Response* – Historically, rates of imputation for non-response in the ACS have been much lower than in the 2000 Census because of the use of highly training interviewers in the ACS. In the 2000 Census about 33 percent of all New York residents 15 and older had income imputed versus about 25 percent in the ACS.

Household and family incomes – Not comparable - Concepts are comparable but in addition to the issues above, they are also affected by differences in household composition due to the different residence rules used in the 2000 Census and the ACS

Per Capita and Aggregate Incomes – Not comparable - Concepts are comparable but the 2005 ACS excludes the incomes of people living in non-institutional group quarters (college dormitories, military barracks, etc.) and uses the household population as the base while the 2000 Census includes these incomes and uses the total population as the base.

Source of Income – Not comparable - Concepts are comparable in that definitions are consistent between the Census and the ACS, but the data are impacted by all of the general income qualifications.

Poverty Status – Not comparable - Since poverty status is based on income, it is subject to all of the problems described under income. Additionally, poverty status in the 2000 Census refers to poverty status during 1999. In the ACS poverty status refers to the 12 months prior to completing the questionnaire. The difference in residency rules can also affect the number of people in a family, changing the poverty threshold for that family, but might not have much of an effect on the family's income.

Employment Status – Not Comparable - The concept is comparable in that both the 2000 Census and the ACS ask for employment status last week. However, in the Census the reference week is the last week in March whereas the ACS reflects an annual average collected throughout the year.

Means of Travel to Work – Generally comparable - In areas with large seasonal workforces or commuting methods affected by weather, such as walking, the data may not be comparable due to different reference periods. Also, "public transportation" included taxicabs in the 2000 Census but the 2005 ACS excluded this category.

Industry, Occupation, and Class of Worker - Generally comparable - In areas with large seasonal workforces the data may not be comparable due to different reference periods.

Weeks Worked – Comparable

Hours Worked – Comparable

Place of Work – Comparable

Time Leaving Home – Comparable

Travel Time to Work – Comparable