

Comparison of the American Community Survey's Three-Year Averages and the Census Sample for a Sample of Counties.¹

Gregg Diffendal, Rita Petroni, Andre Williams
Gregg Diffendal, US Census Bureau, Washington DC 20233

Keywords Census Long Form, Graphical display, Census Profile Data

1. Introduction

This paper should assist data users to better understand the American Community Survey's (ACS) estimates and to demonstrate comparability with the Census 2000 sample estimates. We examine the Census 2000 sample and the ACS 3-year average estimates for a group of 36 counties. The ACS data were collected between 1999 and 2001 while the Census 2000 data were collected over a 6-month period in 2000.

The ACS, as carried out so far, is a sample of housing units and uses a questionnaire similar to the census long form. It is planned that the ACS will replace the census long form for 2010, freeing up census resources to work on counting the population without the need to collect the census sample data. Also the ACS provides annual updated estimates rather than the Census estimates available every ten years. The ACS sample for most counties in this 3-year study is from a systematic sample of about 15 percent of all housing units over the 3-year period. The samples for San Francisco, CA; Bronx, NY; Lake, IL; Broward, FL; and Franklin, OH were 9 percent of the housing units, and the Harris and Fort Bend, TX samples were 3 percent. The sample mimics the census long form sample with a higher sampling rate for less populous places and a lower sampling rate for large tracts not in small places.

The census long form was used to enumerate a sample of housing units and people living in group quarters. The data are available every ten years. Because of the large quantity of estimates for Census 2000, the census long form estimates were released over a 2-year span beginning in the summer of 2002. Estimates are available for all states, counties, places, tracts, and block groups.

There are a series of national level reports that address a similar set of comparisons at the national level and which will be helpful in understanding many of the differences noted here.

The paper begins with a discussion of the methodology we used to calculate the ACS and Census

2000 estimates. This includes percents, calculation of the standard errors, percent differences and determination of whether the differences may be due to sampling error alone. Then we discuss the analysis of the data and changes to the data to make them comparable, such as dropping the group quarters population from the census. We follow this with graphical displays of the differences between the ACS 3-year average and the Census 2000 estimates for a small number of tables. We finish with a discussion of the results.

2. Methodology

Hundreds of tables were produced for the Census 2000 and the ACS. To keep this work manageable, we limited ourselves to a small set of data tables covering a few major topics in the ACS and in Census 2000.

2.1 Changes to the Data

To make the ACS and census data comparable, a number of changes were made. The changes include the removal of the group quarter population from the Census 2000 data; rounding the Census 2000 numbers to prevent disclosure; using the Census 2000 sample data for the demographic tables even though the Census 2000 hundred percent counts were available; removal or collapsing of some items because the data were not directly comparable; and adjusting the ACS dollar amounts for inflation.

The focus of this paper is to examine the differences and similarities between the ACS and Census 2000. We are not interested in the count data itself since the total population and total number of housing units are used to control the population estimates for the ACS and the hundred percent Census 2000 counts are used to control the Census 2000 sample estimates. Instead we focus on the percent differences.

¹ This report is released to inform interested parties of research and to encourage discussion. The views expressed on operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

2.2 Weighting

The ACS weighting was done separately for each of the three years of sample. The population controls used the data from the population estimates program for 1999 and 2001. For 2000 we used the Census 2000 (without group quarters population) as the population control. For 1999 we used the Census 2000 counts projected back to 1999. These numbers differ from the official 1999 population estimates. For 2001, we used the Census 2000 counts projected forward. They are the official 2001 population estimates for the household population. The ACS uses population controls at the county level by age, race, hispanic origin and sex. Even at the county level, the ACS does some collapsing of control categories because of small sample sizes.

The census long form estimates were controlled to the hundred percent census counts at a weighting area level. The weighting area level is roughly equivalent to a census tract. The weighting controls were at a much smaller geographic level than for the ACS estimates. The census used race, Hispanic origin, age, sex, family type, and group quarter residents to control the census population sample estimates. For the housing items, the census used occupied and vacant housing units as controls. Because of small sample sizes, there was often the need to collapse some categories or variables.

2.3 Variance Estimation

The ACS used replication methods to compute the standard errors for all estimates. The ACS population and housing counts that were control variables may have zero or very small standard errors. The standard errors at the county level were set to zero for the total population and total number of housing units. Technical documentation (Navarro and Starsinic 2003) describes standard error calculations for some special cases (for example small sample sizes and when there is no sample in one or two years).

For Census 2000, we used the generalized variance formula for all percents. This method uses a simple random sample formula multiplied by a design effect. For some means, we used the ratio method to calculate the variances. For other means and all medians, we used the ACS standard error and ratio adjusted it to account for the larger sample sizes for Census 2000. This is probably an over-estimate of the true standard error. The standard errors for the total population and total number of housing units at the county were automatically set to zero.

2.4 Testing for Differences Between the ACS and the Census

Because the ACS and the Census 2000 sample data are both estimates derived from a sample, we expect the estimates to differ. We use the Z-score to determine

whether the differences are due to sampling variability or with a 90 percent certainty, the U.S. Census Bureau standard for testing, are due to issues other than sampling variability. At the county level we use a Bonferroni adjustment to state whether the whole group of counties are statistically significantly different or whether one or more are statistically significantly different. The Bonferroni adjustment accounts for the fact that we are testing 36 estimates, one for each county, and that we will be 90 percent confident that all of them are not statistically significantly different.

2.5 Analysis

Our major means of analysis of the differences between the ACS and the Census 2000 estimates is through graphical techniques. We use a dot plot showing the percent differences for all counties for a specified characteristic, with a darkened dot for all counties that tested to be statistically significant. We also include a table with each plot showing the ACS and Census 2000 average percent for each category and the number of ACS and census counties with differences that are statistically significant. The ACS and Census 2000 averages are the unweighted average of the percents, so all counties are equally weighted. The plot points are ordered from smallest to largest in population size.

3. Analysis of Selected Characteristics

3.1 Race

The race data are tabulated for the total household population of the county. Table 1 shows the average percent for each of the eight categories for the 36 counties and a count of the number of counties with statistically significant higher ACS and Census 2000 estimates. The one race category and two or more races equal the total universe. The individual race categories are tabulated only for those who responded using a single race. Separate tables in the full report (Diffendal, Petroni and Williams 2004) also address the specified Asian subcategories and the Native Hawaiian and Other Pacific Islanders subcategories.

The figure depicts the differences between the ACS and the Census 2000 race categories for the 36 ACS counties. One or more county estimates are statistically significantly different between the ACS and the Census 2000 estimates for all categories except for Native Hawaiian and other Pacific Islander. There are large numbers of differences for the One race category, Two-or-more races category, White, and Some other race. Similar results have been documented at the national level by Bennett and Griffin (2002). Their analysis indicates that the Hispanic population were much more likely to respond as Some other race in Census 2000 than in the Census 2000 Supplementary

Survey (C2SS) (42 percent vs 29 percent respectively). The C2SS is a national survey using the ACS methodology. Also the Hispanic population were more likely to report Two-or-more races in Census 2000 than in the C2SS (6.3 percent vs 4.8 percent).

Race is used as a population control variable, but does not correspond to the categories used here because of collapsing of multiple race responses into a single race category for weighting purposes. Also, collapsing occurs frequently because of small numbers of sample cases in some race categories in some areas.

The results for the White and Some other race categories generally follow the results from the Bennett and Griffin (2002) paper where counties with large Hispanic populations show statistically significant differences. However the results for Bronx, NY; Starr, TX; and Yakima, WA are of the opposite direction from the national level differences shown in Bennett and Griffin (2002).

In summary, many of the differences between the ACS and Census 2000 estimates for race are traced to the differences observed for the different responses for the Hispanic population.

3.2 School Enrollment

The school enrollment data are tabulated for the population over three years of age and attending school. Table 2 shows the average percent for each of the five categories for the 36 counties.

The figure depicts the differences between the ACS and the Census 2000 school enrollment for the 36 ACS counties. Differences for one or more counties are statistically significant between the ACS and the Census 2000 for preschool, elementary, and college attendance. No differences are significant for kindergarten and high school attendance estimates. The Census 2000 preschool estimates seem to be consistently higher than the responses from the ACS. The elementary estimates average about the same. The college estimates are slightly higher for the ACS than for the Census 2000 estimates. The differences for kindergarten and high school are clustered around zero.

An error occurred in the ACS processing of large households (over 5 persons) that mailed their forms back (Raglin et al 2004). These cases went to a telephone follow-up operation to try to complete the information for persons who were not recorded on the ACS questionnaire. The error occurred when no contact was made (a telephone noninterview). For these cases, the ACS processing incorrectly converted the housing unit to only five persons, instead of the number recorded on the questionnaire. We would expect many of these persons removed would be the youngest children and at least some of them would have been nursery or preschool children artificially lowering the ACS estimate. This ACS processing error has since been fixed.

The larger ACS estimate for the percent estimate attending college may be due to removal of the group quarter population from the estimates. Census 2000 counts all college students at their college address and not at their parents address. For those living in dormitories, Census 2000 would classify them as living in a group quarter facility. In the ACS, college students living in a dormitory while attending college would not be counted since the ACS did not interview in group quarter facilities. However if they were living at home in the summer months, they would be included in the ACS, following the ACS residency rule including all persons living at the housing unit for at least two months.

3.3 Disability Status

The Disability Status data are tabulated for the household population 5 years of age and older. Table 3 shows the average percent for each of the four categories for the 36 ACS counties. We limited the scale on the figure to be consistent with our other graphs, only showing differences up to 10 percent even though many are higher than this amount. They are all shown as having values of 10 percent.

The figure depicts the differences between the ACS estimate and the Census 2000 estimate for the 36 ACS counties. There is a clear difference between the ACS and Census 2000 for these estimates. For the percent with a disability age 21-64, the ACS estimates are 5.0 smaller than the census and two-thirds of the counties' differences are statistically significant. The percent without a disability age 21-64 is the opposite of the percent with a disability age 21-64 with the ACS higher than the census for two-thirds of the counties. There are also a large number of statistically significant differences for the disability estimates for the persons age 5-20 with the ACS being generally smaller than the census. Interesting is the fact that for persons 65+, the ACS has a higher estimate of persons with a disability than the census.

Stern (2003) examined the differences between the Census 2000 and the Census 2000 Supplementary Survey (C2SS) using the ACS instrument for the disability data. She examined the six disability questions that go into the disability data item reported here, by mail, followup, and allocation (imputing for missing values) of the item. There were no significant differences due to the allocation of disability between the census and the ACS. The data show consistent responses for four of the six disability questions. For the other two items, the census mail, census followup, and ACS mail were all reasonably consistent.

However the ACS telephone and personal visit data showed a very large difference versus the other data. The census mail, census followup, and ACS mail all use a qualifier for the two disability questions by only asking persons 16 years of age or over to answer

this question. The ACS telephone and personal visit data are collected using a computer instrument which already knows whether the person is 16 years of age or older from their earlier response to the age question. Stern interprets these differences as the respondent incorrectly interpreting the age qualifier to the two disability questions and responding to being just over 16 years of age. This would explain the differences we see for disability between the census and the ACS for the ages 21-64 and the smaller differences for disability for ages 5-21, since the effect noted would only apply to persons 16 to 21 years of age in this group.

Clearly, there are large and statistically significant differences between the ACS and Census 2000 for the disability status tabulations. These differences are attributed to the improvement from using computer-assisted interviewing, but large error is still expected in both the ACS mail responses and any Census 2000 responses. Stern's analysis tells us that with fixes to the ACS mail questionnaire, the ACS should be able to obtain accurate disability data.

3.4 Household Income

Household income data are tabulated using the household population over 14 and tabulated for all persons in a household. Table 4 shows the average percent for each of the ten categories for the 36 ACS counties.

The figure depicts the differences between the ACS household income and the census household income for the 36 ACS counties. Differences for one or more counties are statistically significant between the ACS and the census for seven categories and no differences are statistically significant for three categories. The ACS estimates are generally higher for the \$10,000-\$14,999 and \$15,000-\$24,999 categories while the Census 2000 estimates are generally higher for the \$50,000-\$74,999 category. Except for the minor differences noted, the ACS estimates seem to be in general agreement with the Census 2000 estimates.

Even though there are a number of statistically significant differences between the ACS and Census 2000, almost all of these differences are fairly small (less than 2.0) and there does appear to be general agreement between the ACS and Census 2000 for the household income categories.

3.5 Value of Owner-Occupied Units

Value data are tabulated for all specified owner-occupied housing units. Table 5 shows the average percent for each of the eight categories for the 36 ACS counties.

The figure depicts the differences between the ACS estimate and the Census 2000 estimate for the 36

ACS counties. There are one or more statistically significant differences for six categories and no differences are statistically significant for two categories. There is general agreement between the ACS and the Census 2000 estimates with a few outliers in the \$200,000 to \$299,999, \$300,000 to \$499,999, and \$500,000 to \$999,999 categories.

4. Conclusions

This paper covers a few of the comparable variables available for the ACS 3-year average estimates and the Census 2000 estimates for the 36 counties. Graphs were used to summarize the data and the graphs clearly show the percent differences at the county level. In addition, much higher variability is observed for the small counties than for the large counties, while smaller differences are statistically significant for the large counties. Many of these differences are less than 1.0 but are statistically significant. Care should be used in interpreting small differences which may not be substantively meaningful. *Most of the variables show small differences between the ACS and Census 2000.* Many of these are not shown here, but can be found in the main report listed in the references.

However, the results also identify a few variables where the ACS estimates differed from the Census 2000 estimates, especially for race and disability. Comparisons between the ACS and the Census 2000 estimates at the national level may verify whether these differences exist at the national level or whether some of these differences are particular to the counties chosen for this study. The paper by Stern (2003) argues that the disability differences are due to questionnaire wording which is corrected using the ACS computer-assisted instruments for the telephone and personal visit ACS interviews. Other differences may be due to other differences between the ACS and the census processes such as questionnaire design issues, editing, and imputation procedures, differences due to mode of data collection, differences in weighting, differences in residence rules, using multiple years of data for the ACS, and perhaps other reasons. Analysts in the Population, and Housing and Household Economics Statistics Divisions are examining some of these differences to provide a clearer understanding of why these differences exist. In addition, a series of reports have been released to examine these differences at the national level using the Census 2000 Supplementary Survey data which uses the ACS methodology.

10. References

Bennett, Claudette and Griffin, Deborah (2002) "Race and Hispanic Origin Data: A Comparison of Results From the Census 2000 Supplementary Survey and Census 2000," paper presented at the 2002 Joint

Statistical Meetings.

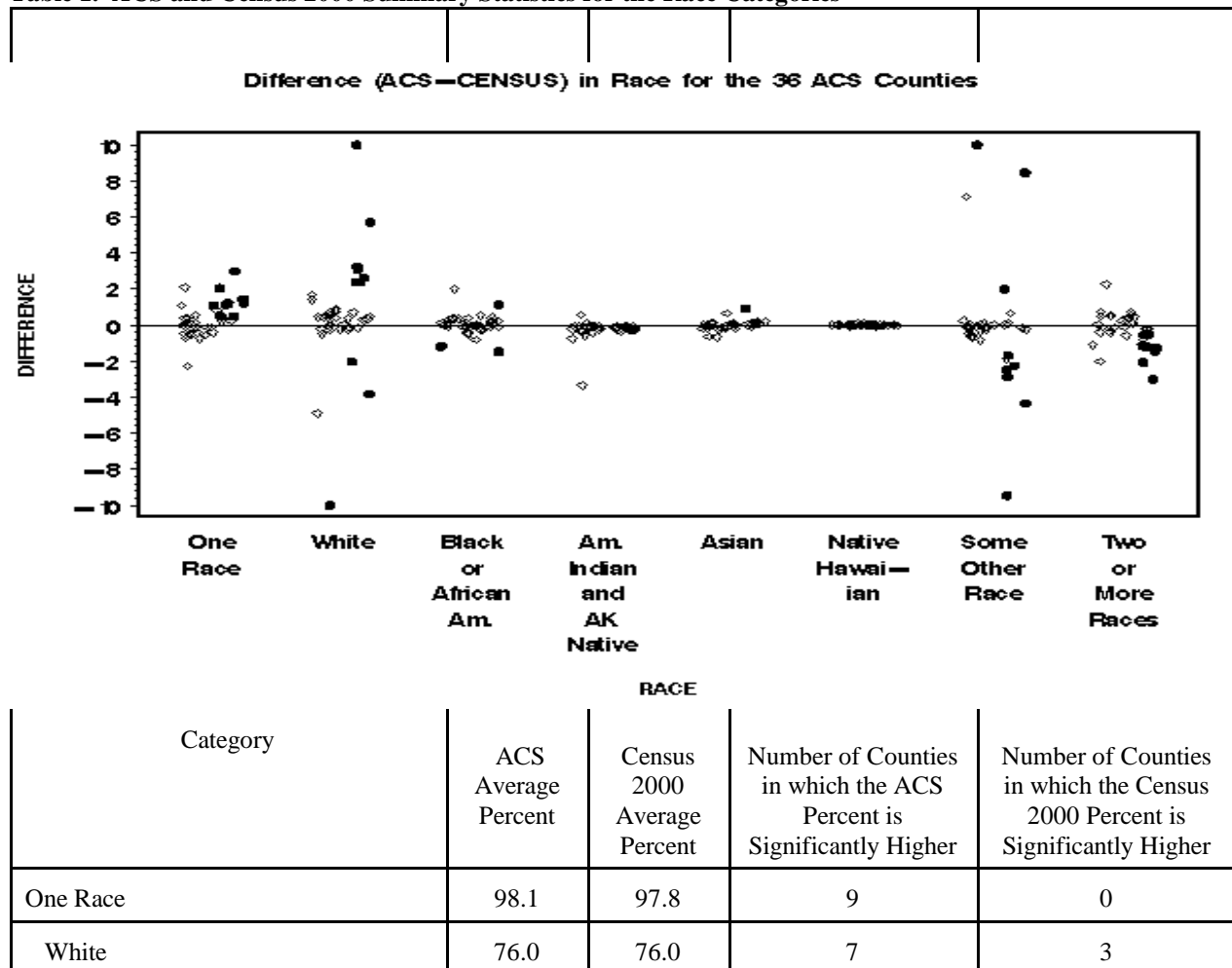
Diffendal, Gregg, Petroni, Rita and Williams, Andre (2004) "Meeting 21st Century Data Needs-Implementing the American Community Survey - Report 8: Comparison of the American Community Survey Three-Year Averages and the Census Sample for a Sample of Counties and Tracts" U.S. Census Bureau Report.

Navarro, Alfredo and Starsinic, Michael (2003) "American Community Survey 1999-2001 Three Year Average and Census 2000 Sample Comparison Profile Technical Documentation," available on the Census web page at www2.census.gov/pub/outgoing/acsc2k/docs/acs_census_technical_documentation.doc

Raglin, David, Broadwater, Joan, Griffin, Deborah Leslie, Theresa, and Love, Susan (2004) "Meeting 21st Century Data Needs-Implementing the American Community Survey - Report 9: Comparing Social Characteristics with Census 2000" U.S. Census Bureau Report.

Stern, Sharon (2003) "Counting people with disabilities: A Comparison of Estimates in Census 2000 and the Census 2000 Supplementary Survey," paper presented at the 2003 Joint Statistical Meetings.

Table 1. ACS and Census 2000 Summary Statistics for the Race Categories



Black or African American	12.0	12.0	1	2
American Indian and Alaska Native	1.4	1.7	0	2
Asian	2.6	2.6	1	0
Native Hawaiian and Other Pacific Islander	0.06	0.05	0	0
Some Other Race	6.0	5.5	3	6
Two or More Races	1.9	2.2	0	9

Difference (ACS – CENSUS) in School Enrollment for the 36 ACS Counties

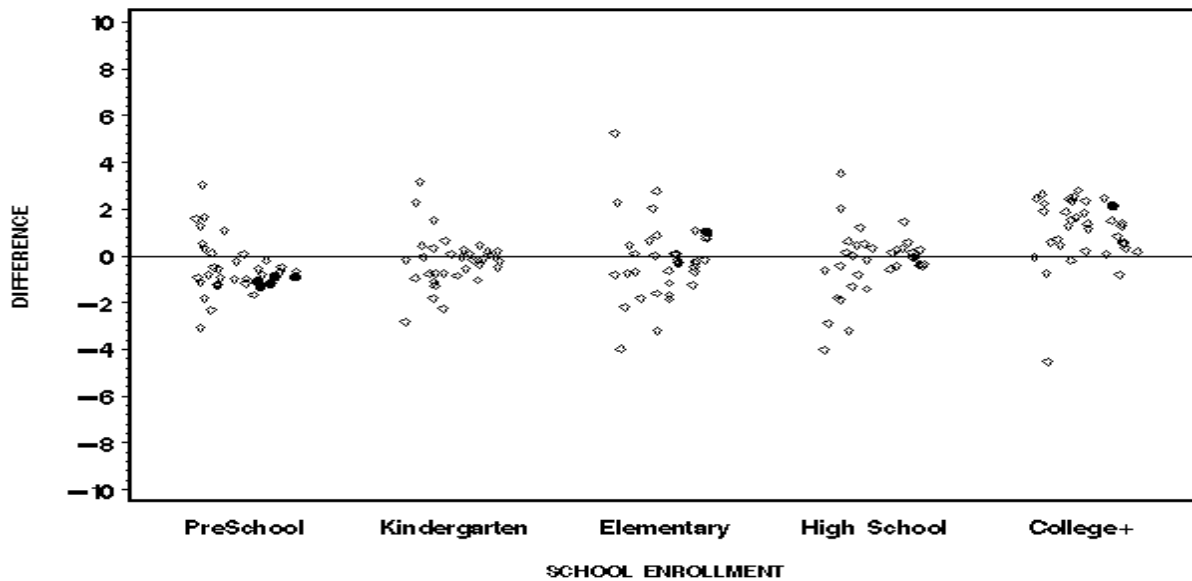


Table 2. ACS and Census 2000 Summary Statistics for the School Enrollment Categories

Category	ACS Average Percent	Census 2000 Average Percent	Number of Counties in which the ACS Percent is Significantly Higher	Number of Counties in which the Census 2000 Percent is Significantly Higher
Nursery School, Preschool	5.9	6.4	0	5
Kindergarten	5.5	5.7	0	0
Elementary (Grades 1-8)	47.3	47.5	1	0
High School (Grades 9-12)	22.7	22.9	0	0
College or Graduate School	18.6	17.5	1	0

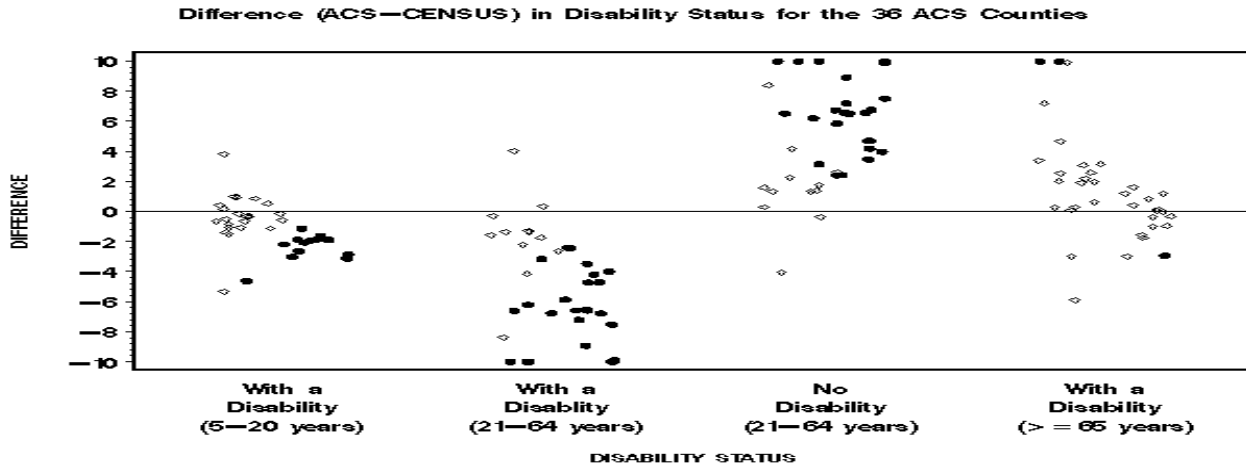


Table 3. ACS and Census 2000 Summary Statistics for the Disability Status Categories

Category	ACS Average Percent	Census 2000 Average Percent	Number of Counties in which the ACS Percent is Significantly Higher	Number of Counties in which the Census 2000 Percent is Significantly Higher
With a Disability 5-20 Years	7.1	8.2	0	14
With a Disability 21-64 Years	16.4	21.5	0	24
No Disability 21-64 Years	83.6	78.5	24	0
With a Disability 65+ Years	45.6	43.9	2	1

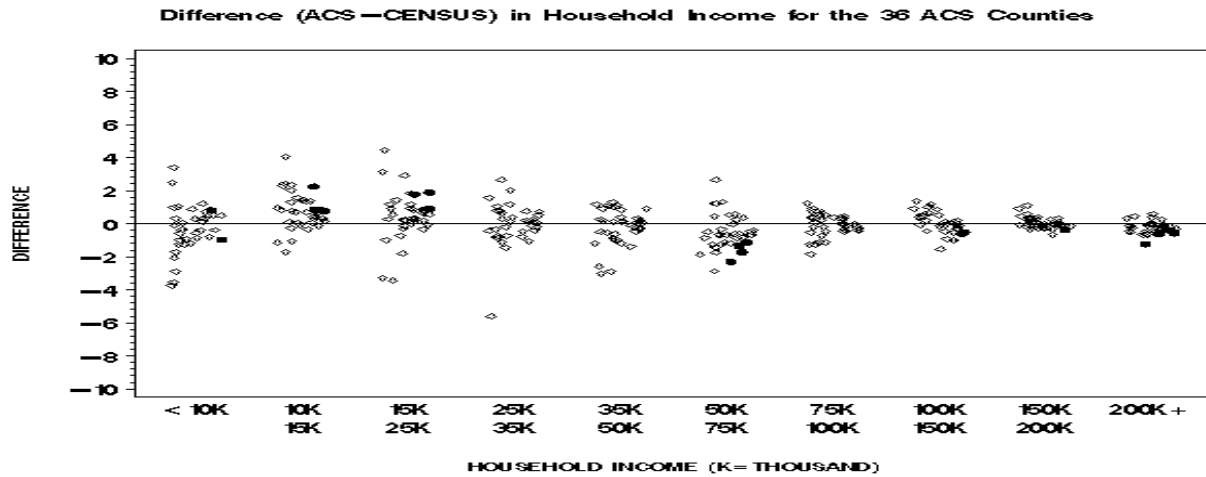


Table 4. ACS and Census 2000 Summary Statistics for the Household Income Categories

Category	ACS Average Percent	Census 2000 Average Percent	Number of Counties in which the ACS Percent is Significantly Higher	Number of Counties in which the Census 2000 Percent is Significantly Higher
Less than \$10,000	11.9	12.2	1	1

\$10,000 - \$14,999	8.5	7.8	5	0
\$15,000 - \$24,999	15.3	14.8	3	0
\$25,000 - \$34,999	13.8	13.8	0	0
\$35,000 - \$49,999	16.3	16.5	0	0
\$50,000 - \$74,999	16.9	17.4	0	4
\$75,000 - \$99,999	8.2	8.3	0	0
\$100,000 - \$149,999	5.9	5.9	0	2
\$150,000 - \$199,999	1.7	1.7	0	1
\$200,000 or More	1.6	1.8	0	5

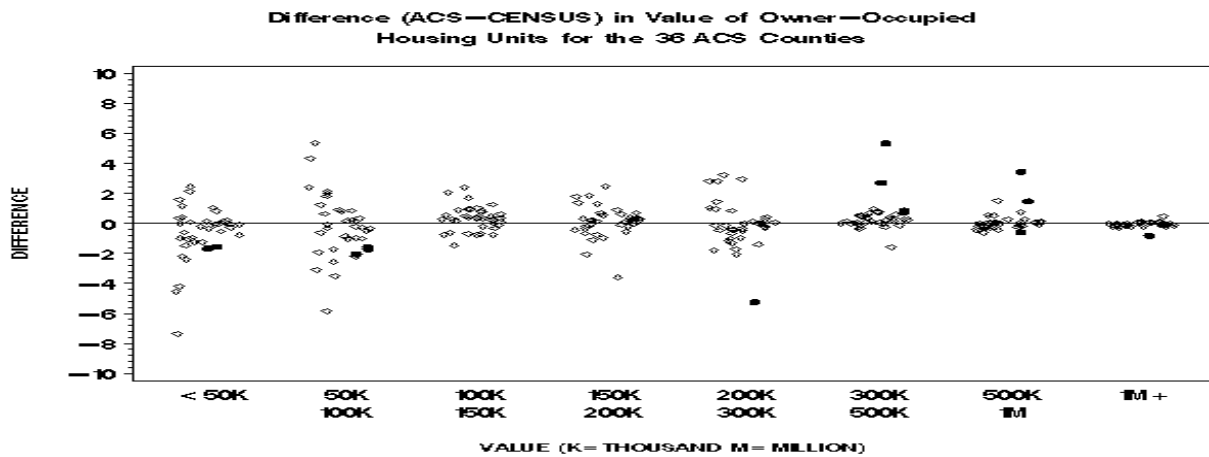


Table 5. ACS and Census 2000 Summary Statistics for the Value of Owner-Occupied Housing Unit Categories

Category	ACS Average Percent	Census 2000 Average Percent	Number of Counties in which the ACS Percent is Significantly Higher	Number of Counties in which the Census 2000 Percent is Significantly Higher
Less than \$50,000	17.4	18.0	0	2
\$50,000 to \$99,999	34.6	34.8	0	3
\$100,000 to \$149,999	20.2	19.9	0	0
\$150,000 to \$199,999	11.3	11.2	0	0
\$200,000 to \$299,999	9.2	9.2	0	1
\$300,000 to \$499,999	5.2	4.7	4	0
\$500,000 to \$999,999	1.8	1.7	2	1
\$1,000,000 or More	0.4	0.5	0	1