The ACS Statistical Analyzer

National Center for Transit Research
University of South Florida, Tampa
March 2010

Final Report

http://ebr.eller.arizona.edu/images/Oct09_photos/crowd_jumping.jpg

Contract Number: BDK85 977-02

TECHNICAL REPORT DOCUMETATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle The ACS Statistical Analyzer		5. Report Date March 2010
		6. Performing Organization Code
7. Author(s) Xuehao Chu		8. Performing Organization Report No. NCTR778-02, FDOT BDK85 977-02
9. Performing Organization Name and Address National Center for Transit Research (NCTR) University of South Florida 4202 E Fowler Ave., CUT100, Tampa, FL 33620-5375		10. Work Unit No.
		11. Contract or Grant No. USDOT DTRS98-G-0032
12. Sponsoring Agency Name and Address Office of Research and Special Programs U.S. Department of Transportation, Washington, D.C. 20590 Florida Department of Transportation 605 Suwannee Street, MS 30, Tallahassee, FL 32399		13. Type of Report and Period Covered
		14. Sponsoring Agency Code
15. Supplementary Notes Supported by a grant from the Florida Department of Transportation and the U.S. Department of Transportation		

16. Abstract

This document provides guidance for using the *ACS Statistical Analyzer*. It is an Excel-based template for users of estimates from the American Community Survey (ACS) to assess the precision of individual estimates and to compare pairs of estimates for their statistical differences. The *ACS Statistical Analyzer* covers the following four functions and fifteen sub-functions (not listed):

- To derive other precision measures for published ACS estimates at American FactFinder or from the Census Transportation Planning Products (CTPP), which already have a margin of error (MOE).
- To derive the precision measures for estimates that do not already have an MOE.
- To derive the precision measures of new estimates obtained from two or more original estimates that already have an MOE.
- To compare pairs of two estimates that already have an MOE.

Measures of precision for an estimate include its MOE, relative reliability, and confidence interval. The implementation of the *ACS Statistical Analyzer* is expected to reduce the agency cost of, and to lessen the technical barriers to, dealing with the precision of ACS estimates when agencies use these estimates. These direct benefits in turn can lead to wider and more effective usage of ACS data.

17. Key Words	18. Distribution Statement		
ACS, Census 2000, Estimates, Precision, ACS Statistical Analyzer, Averages, Medians, Ratios, Percentages, Frequencies, Totals	Available to the public through the National Technical Information Service (NTIS), 5285 Port Royal Road, Springfield, VA 22161, 703-487-465, and through the NCTR website at http://www.nctr.usf.edu		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of pages 32	22. Price

DISCLAIMER

ACKNOWLEDGMENT

The *ACS Statistical Analyzer* was inspired by the Statistical Significance Calculator (http://www.trbcensus.com/notes/StatisticalCalculationsMenu.xls), which was developed by the staff of the New York State Data Center at Empire State Development. The author wants to thank Patria Ball, Ed Hillsman, Steve Polzin, and Joel Volinski of the Center for Urban Transportation Research, Liang Long of Cambridge Systematics, Inc., and Michael Starsinic of the U.S. Census Bureau for their comments. These reviews do not represent an official approval by these organizations or their staff of the formulas in the *ACS Statistical Analyzer*, the accompanying documentation, or the results obtained by users from the *ACS Statistical Analyzer*. All remaining errors and shortcomings are the author's responsibility.

Executive Summary

Background

Transportation planning in general and transit planning in particular in Florida and throughout the nation have relied heavily on the commuting and socio-demographic data from the long-form survey of the decennial census at various levels of geography. While the short-form count will continue every 10 years, the long-form survey has been replaced by the American Community Survey (ACS). While providing more current information, ACS data represent serious challenges for transportation planning professionals to use them effectively. One of these challenges results from the fact that the precision of estimates from the ACS is significantly lower than the precision of estimates from the traditional decennial census long-form survey. This requires transportation planning professionals to explicitly take into account the precision of estimates from the ACS when they use these estimates either individually or for comparisons. Transportation planning professionals, however, face difficulties in overcoming this challenge:

- Estimates in published ACS tables at American FactFinder come with a margin of error (MOE) but without other measures of precision. This makes it difficult for transportation planning professionals to judge the usability of these estimates.
- Some estimates do not come with any measure of precision. While necessary statistical procedures and formulas are available in various documents from the U.S. Census Bureau, they are not easily accessible to many transportation planning professionals.
- When the procedures and formulas are accessible, they typically involve statistical procedures and formulas that many transportation planning professionals do not feel comfortable working with.

Objectives

The objective of this project was to develop a tool that helps transportation planning professionals overcome these difficulties in using ACS data. The target users are those who are familiar with the statistical concepts involved, are familiar with the measures of precision and their use, and are even capable of following the statistical procedures and formulas, but do not want to go through these procedures and formulas by themselves.

Findings and Conclusions

The resulting tool from the research project is the *ACS Statistical Analyzer*. Transportation planning professionals can use it to assess the precision of individual estimates in terms of several measures of precision without the need to work directly with the statistical procedures and formulas involved. They also can use this tool to compare pairs of estimates in terms of their statistical differences without the need to work directly with the statistical procedures and formulas involved. The tool is comprehensive and covers a full range of functions and sub-functions for transportation planning professionals to derive measures of precision in individual estimates and to compare estimates:

A. To derive other precision measures for published ACS estimates at American FactFinder or estimates in the Census Transportation Planning Products (CTPP) for ACS data. Estimates from these two sources come with a margin of error (MOE):
   01. For up to 200 ACS estimates from the same table.

B. To derive the precision measures for individual estimates that do not already have an MOE. These include published Census 2000 estimates, CTPP 2000 estimates, individual user-derived estimates from an ACS Public Use Microdata Sample (PUMS), and user-derived estimates from a Census 2000 PUMS.
   02. For frequencies, totals, averages, or medians from an ACS PUMS using replicate estimates.
   03. For averages from Census 2000 or a PUMS using a distribution table.
   04. For medians from Census 2000 or a PUMS using a distribution table.
   05. For frequencies from Census 2000 or a PUMS using design factors.
   06. For percentages from 2000 Census or a PUMS using design factors.

C. To derive the precision measures for new estimates obtained from two or more original estimates that already have an MOE. These estimates can be published ACS estimates, CTPP estimates, estimates whose precision measures are derived using Function B, or estimates whose precision measures are derived using another sub-function of this function. This function covers estimates obtained using one of the following six operations:
   07. Sum of two or more estimates.
   08. Difference of two estimates.
   09. Percent difference of two estimates.
   10. Ratio of one estimate over another.
   11. Percentage of one estimate in another.
   12. Product of two estimates.

D. To compare pairs of two estimates that already have an MOE. The estimates to be compared may be published ACS estimates, CTPP estimates, estimates whose precision measures are derived using Function B, or estimates that are derived along with their precision measures using Function C. This function covers three types of comparisons:
   13. One ACS estimate with another.
   14. One ACS estimate with a Census 2000 estimate using actual MOE.
   15. One ACS estimate with a Census 2000 estimate using assumed MOE.

Benefits

The implementation of the *ACS Statistical Analyzer* is expected to reduce agencies' cost and lessen the technical barriers to dealing with the precision of ACS estimates when agencies use these estimates for transportation planning. These direct benefits, in turn, can lead to wider and more effective use of ACS data for transportation planning.

This project was conducted by Xuehao Chu at the University of South Florida. For more information, contact Daniel Harris, Transit Planning Project Manager, at (850) 414-4532, daniel.harris@dot.state.fl.us.

TABLE OF CONTENTS

Sections

EXECUTIVE SUMMARY V
Background v
Objectives v
Findings and Conclusions v
Benefits vi
INTRODUCTION 1
FEATURES OF ESTIMATES AND COMPARISONS 3
The Source of an Estimate 3
The Form of an Estimate 3
Measures of Precision 4
Statistical Significance 6
FUNCTIONS AND SUB-FUNCTIONS 7
Description 7
Applicability 8
Joint and Alternative Uses 9
DATA NEEDS 13
Obtaining Data from American FactFinder 13
Obtaining Data from the CTPP 15
Summary of Data Needs for Function B 16
Obtaining Required Data for Function B 18
REFERENCES 23
APPENDIX 25

List of Figures

Figure 1. U.S. Census Bureau's Web Page for Census 2000 Information 14
Figure 2. ACS Web Page for Available Data Sets 15
Figure 3. An Example of the Base of a Percentage 20

List of Tables

Table 1. Unique Data Requirements for Function B 16
Table 2. Finite Population Correction Factor by Data Source 16
Table 3. American FactFinder "Detailed Tables" for Size of the Geography 18
Table 4. A Distribution Table of Household Income, Miami, Florida 19
Table 5. Detailed Table P31, Miami City, Florida 21
Table 6. American FactFinder "Detailed Tables" for Percent in Sample Values 21

This page intentionally left blank.

INTRODUCTION

The American Community Survey (ACS) annually collects socio-economic and commuting data from a sample of housing units that is much smaller than the long-form survey during the decennial census. Estimates from the ACS may be used individually for indicating the condition of a geographic area or of a sub-population in the area. The percentage of the households under poverty would be a condition of a geographic area, while the average household income among households under poverty would be a condition of a sub-population group. Estimates from the ACS may also be used jointly for indicating differences in the conditions of different geographic areas or different sub-population groups or for indicating changes in the condition of a given geographic area or for a given sub-population over time.

Estimates from all surveys include some amount of error due to sampling, and the amount of error with estimates from the ACS is significantly greater than that with estimates from the traditional long-form survey. The U.S. Census Bureau has the following guidance to using estimates from the ACS for both of the above purposes:

> "As the ACS estimates are based on a sample survey of the U.S. population, information about the sampling error associated with the estimates must be taken into account when analyzing individual estimates or comparing pairs of estimates across areas, population groups, or time periods."

This *ACS Statistical Analyzer* helps users of ACS estimates follow this guidance of the U.S. Census Bureau. An ACS estimate can come from one of several sources. These include published ACS tables at American FactFinder, estimates from the Census Transportation Planning Products (CTPP), those derived from these published estimates or CTPP estimates, and those derived from any ACS Public Use Microdata Sample (PUMS).

Although the *ACS Statistical Analyzer* focuses on ACS estimates, it also deals with Census 2000 estimates. Users of ACS estimates often want to compare current conditions as reflected in ACS estimates with conditions in 2000 as reflected in Census 2000 estimates. Census 2000 estimates also include not only those in the published tables at American FactFinder, but also those from CTPP tables, those derived from the published tables or CTPP tables, and those derived from a 2000 PUMS.

This document is a guide to using the *ACS Statistical Analyzer*. It combines the various pieces of the guidance to using the *ACS Statistical Analyzer* that is already contained in the *ACS Statistical Analyzer*. The remainder of this document is divided into three sections:

- *Features of Estimates and Comparisons* defines the essential features of an estimate and a key feature of statistical comparisons used throughout the *ACS Statistical Analyzer* and this document. These features include the source of an estimate, the different forms of an estimate, measures of precision for an estimate, and statistical significance.

- *Functions and Sub-Functions* summarizes the 4 functions and 15 sub-functions that the *ACS Statistical Analyzer* serves. It also discusses the applicability of these functions and sub-functions and the joint use of more than one sub-function for certain analyses.

- *Data Needs* details data needs for using the *ACS Statistical Analyzer* and how the user may obtain them.

In addition, an appendix provides the statistical procedures and formulas used for each function and sub-function for those curious users.

FEATURES OF ESTIMATES AND COMPARISONS

This chapter focuses on describing the essential features of an estimate and a comparison. The primary objective is to familiarize the user with a few basic concepts and terms used in the *ACS Statistical Analyzer* and this guide. The key feature of a comparison is statistical significance, and the essential features of an estimate are grouped into three categories:

- The *source* of an estimate.
- The *form* of an estimate.
- The *measures of precision* for an estimate.

The Source of an Estimate

The original source of any estimate that can be analyzed by the *ACS Statistical Analyzer* is either the ACS or the Census 2000 long-form survey. The direct sources of estimates that can be analyzed with the *ACS Statistical Analyzer* are:

- ACS estimates in published tables at American FactFinder.
- Census 2000 estimates from Summary File 3 in published tables at American FactFinder.
- ACS estimates from the CTPP.
- Census 2000 estimates from the CTPP.
- User-derived ACS estimates from two or more published ACS estimates.
- User-derived estimates from two or more published Census 2000 estimates.
- User-derived estimates from two or more CTPP estimates.
- User-derived estimates from an ACS PUMS.
- User-derived estimates from a Census 2000 PUMS.

American FactFinder is the web site that the U.S. Census Bureau has developed as its primary vehicle for distributing Census data, including data from Census 2000 and the ACS. The CTPP contains tables specially designed and tabulated for the transportation community. For both Census 2000 and the ACS, the CTPP contains three sets of tables that show residence characteristics, workplace characteristics, and commuter-flow characteristics. A PUMS is a public use microdata sample that represents a sub-sample of the Census 2000 long-form survey or the ACS. Census 2000 has two PUMS datasets representing1% and 5% of the population. The ACS has one PUMS dataset representing approximately 1% of the population.

The user is assumed to be aware of these sources. As a result, these sources for using the *ACS Statistical Analyzer* are not described here. Some aspects of obtaining estimates from these sources, however, are described in the section on data needs.

The Form of an Estimate

An estimate of a characteristic of a population is a numerical value obtained from a statistical sample of the population to represent the true value of the characteristic that would have been obtained from the entire population. Estimates from both the ACS and the Census 2000 long-form survey are based on interview data collected from samples of all housing units. Estimates may take different forms, including frequencies, totals, ratios, percentages, averages, and medians.

Frequencies. An estimate of a frequency represents the size of a population in terms of persons, households, families, or housing units.

Totals. A total represents the total amount of some characteristic of a population. The U.S. Census Bureau uses the word "aggregate" to describe a total. Examples include household income, vehicles available, rent paid, wage earnings, commuting time, etc.

Ratios. A ratio is one estimate (i.e., the numerator) divided by another estimate (i.e., the denominator), with the numerator not being part of the denominator. The average number of workers per household is one example where the numerator is the total number of workers and the denominator is the total number of households in a geographic area. The number of workers who commute to work by public transportation divided by the total number of workers would not be a ratio because the numerator, in this case, is part of the denominator.

Percentages. A percentage is a ratio multiplied by 100 with the numerator being part of the denominator. One example would be the percentage of workers who usually commute to work by public transportation.

The *ACS Statistical Analyzer* distinguishes two types of percentages: frequency-based and total-based. A frequency-based percentage uses frequencies for their numerator and denominator. The percentage of workers who usually commute to work by a particular mode is an example of a frequency-based percentage. A total-based percentage, on the other hand, uses totals for the numerator and denominator. Consider vehicles available in households. The percentage of vehicles available in households under poverty versus vehicles available in all households is an example of total-based percentages.

Averages. An average may be estimated as the ratio between an estimate of a total as the numerator and an estimate of a frequency as the denominator. The average commute time per worker is an example where the total is the total amount of commuting time for a sub-population in a geographic area and the frequency is the total number of workers for the same sub-population and geographic area.

Medians. The median divides the total population into two equal parts: one-half of the cases fall below the median and one-half of the cases exceed the median. One example would be median household income.

The *ACS Statistical Analyzer* distinguishes these different forms of an estimate because the statistical procedures and formulas for deriving measures of precision differ for these different forms of estimates.

Measures of Precision

The *ACS Statistical Analyzer* focuses on three measures of precision, including relative reliability, confidence interval, and margin of error. Although it is not directly used in the *ACS Statistical Analyzer*, the concept of standard error is described first because it is the centerpiece to all three measures of precision. In addition, standard error and relative reliability are features of the estimates, but confidence interval and margin of error also depend on the so-called level of confidence.

Standard Error. There are many possible samples that could be observed from a given population. In practice, we observe only one of these samples, and one estimate of the true value is produced from that sample. Consider the average of estimates from all possible samples from the same population. The standard error is an estimate of how much the possible estimates from these many possible samples differ from their average. Smaller standard errors suggest that the many possible estimates, including the one for the observed sample, tend to be close to each other.

Relative Reliability. The standard error of an estimate gives an absolute measure of the estimate's precision. A relative measure of precision, however, often is more effective in determining the usability of an estimate. The relative reliability of an estimate is the ratio of its standard error to the estimate, expressed in terms of a percentage. The lower the ratio, the higher the relative reliability of the estimate.

In "ACS Advanced Applications and Issues: Technical Appendix to ACS User Handbook, unpublished draft," Leonard Gaines has suggested the following criteria in using relative reliability to determine the usability of an estimate:

- In many applications, a level of relative reliability of 10% or less is desirable.
- The user should be cautious before using estimates with a level of relative reliability greater than 10% but not greater than 50%.
- Avoid using estimates with a relative reliability level greater than 50%.

These criteria may be used not only to determine the usability of individual estimates, but also to increase the precision in estimates used in applications, especially for ACS data. The precision of ACS data can be improved by using estimates for a longer period or by aggregating across geographic areas and population groups. If the estimate of interest is for a specific geographic area and a specific population group, multiyear estimates should be used where possible when the single-year estimate has a level of relative reliability that is greater than 10%.

Level of Confidence. A level of confidence shows how likely a finding is due to chance. The most commonly used level is 95%. This means that a finding has a 95% chance of being true. A level of confidence is used in constructing a confidence interval and a margin of error, and in determining whether the difference between two estimates is statistically significant.

Confidence Interval. An estimate and its standard error permit the construction of a confidence interval that represents the degree of precision about the estimate at a particular level of confidence. A confidence interval consists of a lower bound and an upper bound. The center of the interval is the estimate, and both the standard error and the level of confidence determine the two bounds. The level of confidence plays a role in determining the two bounds through a confidence multiplier. This multiplier is 1.645, 1.96, and 2.576 for confidence levels of 90%, 95%, and 99%, respectively. Specifically, the lower and upper bounds of a 90% confidence interval for an estimate are:

Lower bound = estimate – 1.645 * standard error.
Upper bound = estimate – 1.645 * standard error.

Over all possible samples of a population, 90% of the intervals produced in this way contain the true value.

Margin of Error. The standard error of an estimate can be used to construct the margin of error at a specific level of confidence. It is one-half of the width of a 90% confidence interval of the estimate for the cases where the confidence interval is symmetric around the estimate. Specifically, the margin of error at the 90% confidence level is given by 1.645 * standard error.

Statistical Significance

The concept of statistical significance is used in determining whether the difference between two estimates is not likely to be from random chance alone. This determination is based on the two estimates, their measures of precision, and a desired confidence level. Two estimates found to be significantly different at the 90% confidence level means that one can be 90% certain that the difference truly exists or that there is a less than 10% chance that the difference came entirely from random chance.

It is important for users to be aware of several issues with the concept of statistical significance:

- In non-statistical terms, being significant often means being important. In statistical terms, however, being statistically significant means being probably true (not due to chance). For a difference being important requires the magnitude of the difference being large enough. Users can use the *ACS Statistical Analyzer* to determine if the difference between two estimates is statistically significant, but users need to determine if the difference is important outside the *ACS Statistical Analyzer*.

- If many comparisons are done, falsely significant results are a problem. A 95% chance of something being true means that there is a 5% chance of it being false. This means that of every 100 comparisons that show results significant at the 95% level, the odds are that 5 of them do so falsely. If you took a totally random, meaningless set of data and did 100 tests of significance, the odds are that 5 tests would be falsely reported to be significant.

- If the impact of an incorrect conclusion reached from a comparison is substantial, the user should consider increasing the level of confidence for the comparison.

- Failing to find evidence that there is a statistically significant difference between two estimates does not constitute evidence that there is no difference between them.

FUNCTIONS AND SUB-FUNCTIONS

This section describes the individual functions and sub-functions of the *ACS Statistical Analyzer*, discusses their applicability, and presents several examples of their joint and alternative uses.

Description

The *ACS Statistical Analyzer* serves 4 functions (A, B, C, D) and 15 sub-functions:

A. To derive other precision measures for ACS estimates in published tables at American FactFinder or estimates in CTPP ACS tables. Estimates from these two sources come with a margin of error (MOE):

   01. For up to 200 ACS estimates from the same table (**A01-ACS**).

B. To derive the precision measures for individual estimates that do not already have an MOE. These include published Census 2000 estimates, CTPP 2000 estimates, individual user-derived estimates from an ACS PUMS, and user-derived estimates from a Census 2000 PUMS.

   02. For frequencies, totals, averages, or medians from an ACS PUMS using replicate estimates (**B02-ACS Direct**).
   03. For averages from Census 2000 or a PUMS using a distribution table (**B03-Average**).
   04. For medians from Census 2000 or a PUMS using a distribution table (**B04-Median**).
   05. For frequencies from Census 2000 or a PUMS using design factors (**B05-Frequency**).
   06. For percentages from 2000 Census or a PUMS using design factors (**B06-Percentage**).

C. To derive the precision measures for new estimates obtained from two or more original estimates that already have an MOE. These estimates can be published ACS estimates, CTPP estimates, estimates whose precision measures are derived using Function B, or estimates whose precision measures are derived using another sub-function of this function. This function covers estimates obtained using one of the following six operations:

   07. Sum of two or more estimates (**C07-Sum**).
   08. Difference of two estimates (**C08-Diff**).
   09. Percent difference of two estimates (**C09-%Diff**).
   10. Ratio of one estimate over another (**C10-Ratio**).
   11. Percentage of one estimate in another (**C11-Percentage**).
   12. Product of two estimates (**C12-Product**).

D. To compare pairs of two estimates that already have an MOE. The estimates to be compared may be published ACS estimates, CTPP ACS estimates, estimates whose precision measures are derived using Function B, or estimates that are derived along with their precision measures using Function C. This function covers three types of comparisons:

   13. One ACS estimate with another (**D13-ACS&ACS**).
   14. One ACS estimate with a Census 2000 estimate using an actual MOE for the 2000 estimate (**D14-ACS&2000 Actual**).

15. One ACS estimate with a Census 2000 estimate using an assumed MOE for the 2000 estimate (**D15-ACS&2000 Assumed**)

Applicability

The applicability of a function or sub-function refers to the conditions under which it may be used or the conditions under which it must not be used. There are two types of conditions that determine the applicability of these functions and sub-functions. One type has to do with the design of the template. The other type is statistical in nature and has to do with the applicability of the statistical procedures and formulas involved. Both of these conditions are explicitly stated in the worksheet for each sub-function. The design-related conditions are straightforward and are not repeated here. The statistical conditions require some discussion here.

Place of Work Characteristics

Users of the *ACS Statistical Analyzer* must not use sub-functions **B03 - B06** to derive measures of precision for estimates of place of work characteristics for Census 2000, including published estimates at American FactFinder, CTPP estimates, or user-derived estimates from a PUMS. Data on place of work in both ACS and Census 2000 refer to the characteristics of workers that are tabulated for the geographic location at which these workers carried out their occupational activities during the week prior to the date of data collection. Data for the ACS and the Census 2000 long-form survey were collected with sampling done in terms of residences. As a result, certain necessary input data required for some of the sub-functions for deriving measures of precision are available for residential places but not available for work places.

However, sub-function **B02 - ACS Direct** can be used to derive measures of precision for estimates of place of work characteristics using replace weights from an ACS PUMS.

Measures of Precision for Totals and Total-Based Percentages

Users must not use sub-function **B05 – Frequency** to derive measures of precision for estimates of totals such as total household income, total commuting time by workers, total number of vehicles available, total rent, etc. in a geographic area. This sub-function applies only to estimates of frequencies such as the number of persons, households, families, or housing units. Refer to the next sub-section on joint and alternative uses of the *ACS Statistical Analyzer* for suggestions on deriving measures of precision for estimates of totals.

Users must not use sub-function **B06 – Percentage** to derive measures of precision for estimates of total-based percentages. The percentage of vehicles available in households under poverty versus vehicles available in all households in an area would be an example of total-related percentages. This sub-function applies only to estimates of frequency-based percentages. The percentage of workers who usually commuted to work by bus in a state would be an example of frequency-related percentages. Refer to the next sub-section on joint and alternative uses of the *ACS Statistical Analyzer* for suggestions on deriving measures of precision for estimates of total-related percentages.

Comparing ACS Estimates

Sub-function **D13 – ACS&ACS** can be used to compare ACS estimates, but it must not be used for comparing ACS estimates that overlap. For comparing estimates from single-year data, the two areas being compared must not overlap if the comparison is between two areas (i.e., they do not share a common area). In addition, the two population groups must not overlap if the comparison is between two population groups (i.e., they do not include a common group).

For multi-year estimates, avoid comparing them with single-year estimates(e.g., three-year vs. one-year); avoid comparing two multi-year periods of different lengths (e.g., three-year vs. five-year); avoid comparing one pre-2006 and one post-2006 (inclusive of 2006) period for an area with a substantial group quarters population; and avoid comparing two multi-year periods that overlap (e.g., 2005-2009 vs. 2006-2010).

Comparing ACS and 2000 Estimates

Sub-Functions **D14 – ACS&2000 Actual** and **D15 – ACS&2000 Assumed** can be used to compare ACS and Census 2000 estimates, but must not be used to compare estimates between the ACS and Census 2000 that may be affected by these key differences between ACS and 2000 data: 1) residence rules (usual residence for Census 2000 but two-month residence for the ACS); 2) reference period (prior calendar year for Census 2000 but prior 12 months for the ACS for income and school attendance); and 3) seasonal variation (April 1 for Census 2000 but continuous for the ACS).

Joint and Alternative Uses

For some statistical analyses of ACS or 2000 estimates, users need to use more than one sub-function. Such joint use of sub-functions can occur both for comparing estimates and for deriving measures of precision for individual estimates. In addition, sub-functions can be used for alternative purposes other than their original design purposes. These joint and alternative uses of the sub-functions increase the flexibility of the *ACS Statistical Analyzer* flexible for its users.

Joint Use

Comparing Estimates
In many cases, users may get two ACS estimates and their MOE directly from American FactFinder or the CTPP and conduct a statistical test to determine whether these two estimates are statistically different at a desired level of confidence. In many other cases, however, one or both estimates to be compared and their measures of precision may be derived from using one of the other sub-functions before Function D is used for the comparison.

Deriving Measures of Precision
Three joint uses of more than one sub-function are discussed for deriving measures of precision:

1. For certain forms of estimates, users may need to use more than one sub-function to derive their measures of precision. For example, users would need to use two sub-functions in two sequential steps to derive measures of precision for estimates of ratios:

   - If the estimates are from published Census 2000 tables or a Census 2000 PUMS, these two steps are:

     1) Estimating the numerator and denominator separately and their measures of precision using one of the last four sub-functions of Function B. The particular sub-function to use depends on the form of the estimates, including averages, medians, frequencies, or percentages.
     2) Deriving measures of precision for the ratios using the results from the first step and sub-function **C10 – Ratio**.

   - If the estimates are from an ACS PUMS, the second step would remain the same, but the first step would be to estimate the numerator and denominator separately and to derive their measures of precision using sub-function **B02 – ACS Direct** or one of the sub-functions of Function B for frequencies, averages, or medians.

2. Deriving measures of precision for estimates of totals, such as total household income for a sub-population in a geographic area, would involve using three sub-functions sequentially:

   1) Using sub-function **B03 – Average** to derive measures of precision for both the relevant average. For an estimate of total household income for a sub-population in an area, for example, the average would be household income per household in the sub-population, and the total frequency would be the total number of households in the sub-population of the area.
   2) Using **B05 – Frequency** to derive measures of precision for the total frequency. In the example of household income, the total frequency would be the number of households in the sub-population.
   3) Using sub-function **C12 – Product** to derive the measures of precision for the estimate of a total. The two estimates for the product are the average and the total frequency.

3. Deriving measures of precision for estimates of total-related percentages would involve using four sub-functions sequentially. Consider the percentage of vehicles available in households under poverty versus vehicles available in all households as an example:

   1) Using sub-function **B03 – Average** to derive measures of precision for the relevant average. This step needs to be done for the average vehicles available per household for households under poverty and for all households, respectively.
   2) Using **B05 – Frequency** to derive measures of precision for the total frequency. This step needs to be done both for the number of households under poverty and the number of households for all income levels.
   3) Using sub-function **C12 – Product** to derive the measures of precision for the estimate of a total. The two estimates for the product are the average from 1) and the total frequency

from 2). This step needs to be done for both households under poverty and all households.

4) Use sub-function **C10 – Ratio** to derive measures of precision for the percentage of vehicles available in households under poverty versus all households. The numerator would be the number of vehicles available in households under poverty, and the denominator would be the number of vehicles available in all households.

Alternative Use

Sub-functions **C07 – Sum**, **C08 – Diff**, and **C10 – Ratio** may be used to derive measures of precision at an alternative confidence level for an estimate that already has an MOE at a different confidence level. Sub-function **A01 – ACS** is designed to derive measures of precision for an alternative confidence level, but it is limited to ACS estimates from published tables at American FactFinder. For deriving measures of precision at an alternative confidence level for other estimates that already have an MOE and a different confidence level, users may do the following:

- **C07 – Sum**: Enter the required input data for a single estimate.
- **C08 – Diff**: Enter the required input data for the minuend but enter 0 for both the subtrahend and its MOE.
- **C10 – Ratio**: Enter the required input data for the numerator but enter 1 for the denominator and 0 for its MOE.

This page intentionally left blank.

DATA NEEDS

The worksheet for each sub-function lists the data needs and how they should be entered in detail. These details are not repeated here. Instead, this section focuses on the following issues:

- Obtaining data from American FactFinder.
- Obtaining data from the CTPP.
- Summary of data needs for Function B.
- Obtaining required data for Function B.

Unlike for other functions, the sub-functions of Function B require certain input data that are unique to them. These unique data items require some explanation on what each required data item is and how it may be obtained.

Obtaining Data from American FactFinder

Census 2000

American FactFinder provides Census 2000 estimates in published tables but no measure of precision. These estimates were derived from Census 2000 Summary File 3. The *ACS Statistical Analyzer* can be used to derive all measures of precision for these estimates.

All required input data from Census 2000 for using the *ACS Statistical Analyzer* can be obtained directly or indirectly through the Census Bureau's web page, as shown in Figure 1. Published tables and related technical documentation are located on the right-hand side once the appropriate summary file is chosen from the middle of the page. Most relevant for the *ACS Statistical Analyzer* would be Census 2000 Summary File 3. PUMS files and related technical documentation are located at web pages accessible through the link at "Download data sets via FTP" located at the upper right-hand corner under "Other Resources."

ACS

American FactFinder provides both estimates and their MOE at the 90% confidence level in published tables that are derived from the full ACS sample. The *ACS Statistical Analyzer* can be used to derive other measures of precision for these estimates and all measures of precision for an alternative confidence level. American FactFinder also provides ACS PUMS datasets for users to derive estimates for geographical areas or population groups that are not available in published ACS tables. The *ACS Statistical Analyzer* can be used to derive all measures of precision for these PUMS estimates.

Users can easily obtain the required input data from the ACS at the Data Sets web page of the ACS, as shown in Figure 2. The middle of the page lists the available single-year and multi-year ACS datasets. Once a particular dataset is selected, as illustrated in Figure 2 for the 2006-2008 three-year dataset, the list of available types of published tables of ACS estimates appears on the right-hand side. This list can vary across the different datasets, particularly between single-year and multi-year datasets. Also shown in this list is a link to download the PUMS file for the selected ACS dataset. For details about data products from the ACS produced by the U.S. Census Bureau, users are referred to the Quick Guide to the American Community Survey (ACS) Products in American FactFinder at http://factfinder.census.gov/home/saff/aff_acs2008_quickguide.pdf.

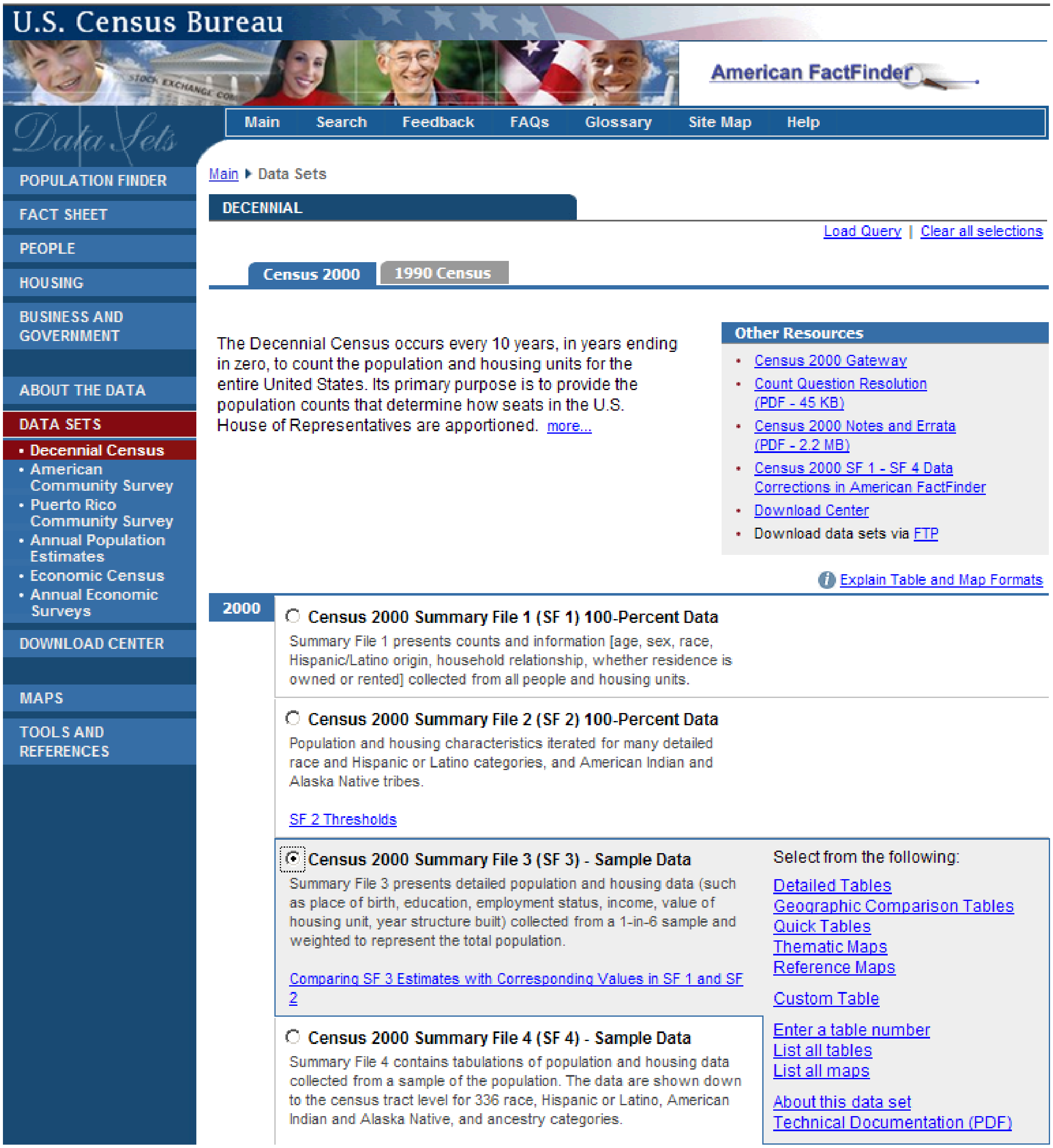

Figure 1. U.S. Census Bureau's Web Page for Census 2000 Information

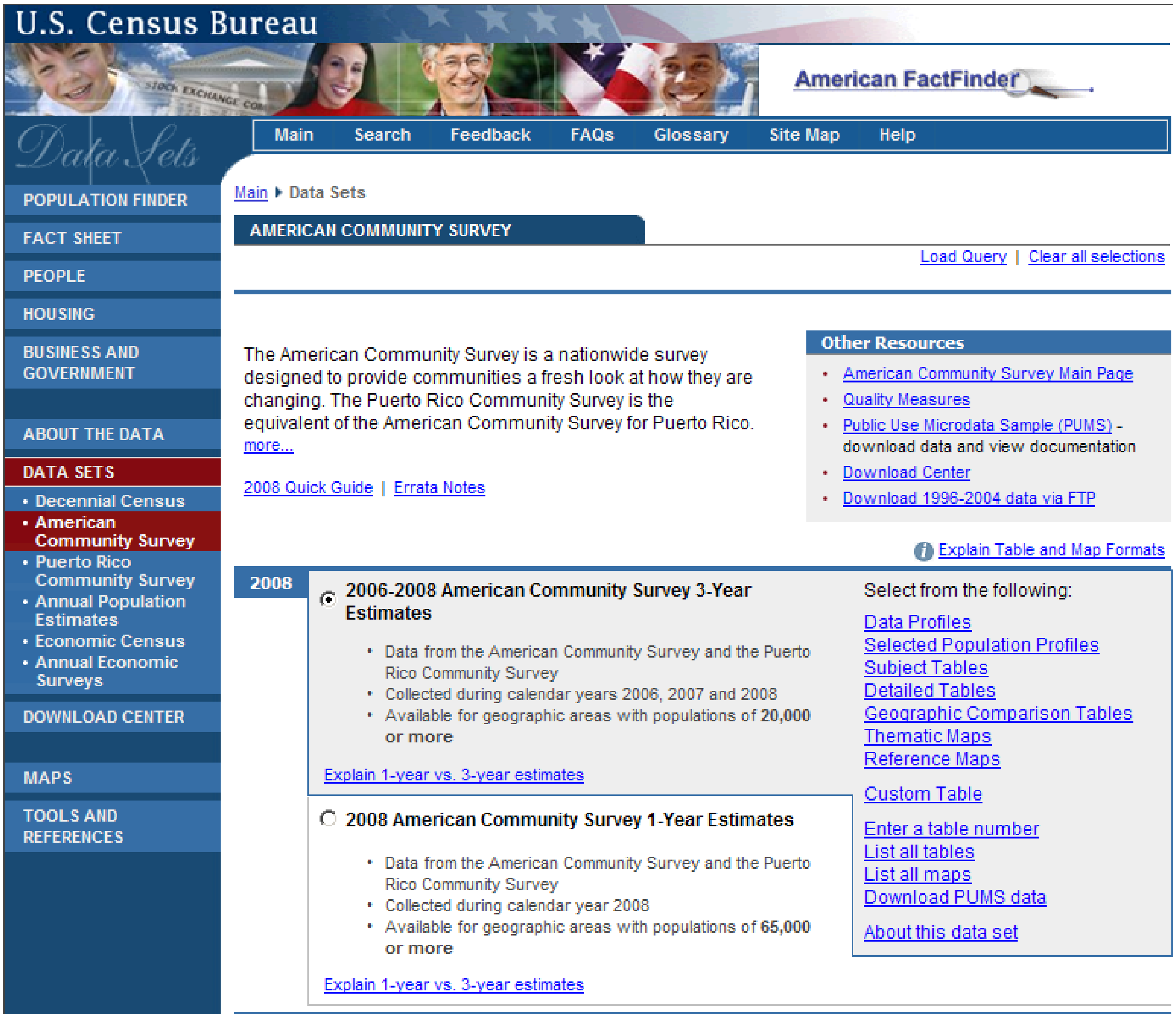

Figure 2. ACS Web Page for Available Data Sets

Obtaining Data from the CTPP

Accessible at http://www.fhwa.dot.gov/ctpp/dataprod.htm, the CTPP provides estimates in three sets of pre-specified tables, including residence tables, place of work tables, and journey-to-work flow tables, for both the ACS and Census 2000. Most of the estimates in the CTPP ACS tables come with their 90% MOE as well; their MOE information is not provided for some estimates, particularly for ratios and percentages. Estimates in the CTPP 2000 do not come with any measure of precision. The *ACS Statistical Analyzer* can be used to derive measures of precision for estimates from the residence tables or journey-to-work flow tables, but not from the place of work tables.

Summary of Data Needs for Function B

The unique data requirements for Function B are described. Table 1 lists each required data item and which sub-function uses it. Each of the first six data items is directly used in the *ACS Statistical Analyzer*. The last two data items are used only to define a design factor.

Table 1. Unique Data Requirements for Function B

Data Items	B02–ACS Direct	B03-Average	B04-Median	B05-Frequency	B06-Percentage
1. Source of Data		X	X	X	X
2. Replicate Estimates	X				
3. Size of the Geography				X	
4. Base of a Percentage					X
5. Distribution Table		X	X		
6. Design Factor		X	X	X	X
7. Percent in Sample		X	X	X	X
8. State of the Geography		X	X	X	X

1. *Source of Data* refers to the original source for the other data items. There are six original sources – the full 2000 long-form survey for published 2000 estimates, 1% 2000 PUMS, 5% 2000 PUMS, 1-year ACS, 3-year ACS, and 5-year ACS. The information on these sources determines the finite population correction factor. This correction factor is determined by the total number of housing units sampled as a percent of all housing units at the national level for each of these six sources. This correction factor is used by most of the sub-functions of Function B to derive precision measures.

   While it is not necessary for the user to know the specific value of this correction factor for each source, Table 2 shows the national sample size and the specific value of the finite population correction factor for each source for those users who are interested in them.

Table 2. Finite Population Correction Factor by Data Source

Source	National Sample	Finite Population Correction Factor
2000 Long-Form Survey	16.67%	5
2000 1% PUMS	1%	99
2000 5% PUMS	5%	19
1-Year ACS	1%	99
3-Year ACS	3%	97/3
5-Year ACS	5%	19

2. *Replicate Estimates* are supplemental estimates to any ACS estimate that a user can obtain from an ACS PUMS. The ACS estimate and a total of 80 replicate estimates can be used to determine the precision measures for the ACS estimate. Every ACS PUMS file comes with the full weight for obtaining the ACS estimate of interest. Every ACS PUMS file also comes with 80 sets of replicate weights for obtaining the replicate estimates:

   - PWGTP1 through PWGTP80 should be used for characteristics related to persons
   - WGTP1 through WGTP80 should be used for characteristics related to households, families, or housing units

3. *Size of the Geography* refers to the total number of persons, households, families, or housing units in the geography for which the estimate of interest is measured. If the estimate is of the number of persons, use the number of total population; if the estimate is of families or households, use the number of households; otherwise, use the number of housing units.

4. *Base of a Percentage* refers to the denominator used in computing the percentage. If the percentage is the number of workers who usually commuted to work by public transportation as a percent of all workers during 2008 in Miami City, for example, the base would be the total number of workers during 2008 in Miami City. These estimates of percentages may be directly obtained from American FactFinder, CTPP 2000, CTPP ACS, or derived by the user.

5. A *Distribution Table* shows the number of persons, workers, households, families, or housing units by ranges of a characteristic on a numerical scale. Examples of characteristics on a numerical scale include income, number of vehicles available, age, travel time, etc. The specific distribution table to be used for a specific analysis depends on the specific average and median. For selection purposes, the focus should be on the characteristic and whether the average or median is about persons, workers, households, families, or housing units:

   - If the average of interest is average commuting time per worker, the distribution table should show the number of workers for each range of commuting time.
   - If the median of interest is median household income, the distribution table should show the number of households for each range of household income.

6. *Design Factor* refers to an adjustment factor by most of the sub-functions of Function B. For Census 2000, design factors vary by the state of the geography, by the characteristic for the estimate, and a percent in sample value. For the ACS, design factors vary by the state of the geography and by the characteristic for the estimate. A design factor reflects the effect of how the original survey data were collected and how the estimate of interest was derived.

   If an estimate is a combination of two or more characteristics, use the largest design factor for this combination of characteristics. The only exception to this is for items crossed with race or Hispanic Origin. For an item crossed with race or Hispanic Origin, use the largest design factor not including the race or Hispanic Origin design factor.

7. A *Percent in Sample* value refers to the number of units in the sample as a percent of the total number of units in the geography of interest for the estimate in question. The units are in terms of persons or housing units (including households or families). A percent in sample value is unique to both the geography and the characteristic for the estimate in question. There are two types of percent in sample values. The percent of the population in sample is used for selecting the design factor for a population characteristic. The percent of the total housing units in sample is used for selecting the design factor for a housing characteristic. It is relevant to design factors for Census 2000.

8. *State of the Geography* refers to the state in which the geography of the estimate of interest is located. If the user is interested in the share of workers who usually commuted to work by public transportation during 2008 in Miami City, for example, the state of the geography would be Florida. If the geography for the estimate of interest covers more than one state, the state of the geography would be the United States. This information is used in selecting the design factor, and design factors are available for the United States as a whole as well as for individual states.

Obtaining Required Data for Function B

A user's interest in a particular estimate of a characteristic determines the required source of data and the state of the geography for the estimate. A user derives replicate estimates from an ACS PUMS rather than obtaining them directly from American FactFinder. The following describes how each of the other required data items for Function B may be obtained from American FactFinder.

Size of the Geography

The size of the geography for a given characteristic (i.e., persons, households, etc.) is available in many of the "Detailed Tables" at American FactFinder for both Census 2000 and the ACS. Table 3 suggests a set of these possible tables and the actual sizes for Miami City:

Table 3. American FactFinder "Detailed Tables" for Size of the Geography

Source	Table	Description	Measure of Size	Size (Miami City, Florida)
ACS (2008)	B01003	Total Population	Persons	343,142
	B11012	Household Type by Tenure	Households	138,786
	B11003	Family Type by Presence and Age of Own Children under 18 Years	Families	77,218
	B25001	Housing Units	Housing Units	168,252
Census 2000	P1	Total Population	Persons	362,563
	P10	Household Size by Household Type by Presence Of Own Children under 18 Years	Households	134,344
	P15	Family Type by Presence of Own Children under 18 Years by Age of Own Children	Families	84,195
	H1	Housing Units	Housing Units	148,554

Base of a Percentage

An example is used to illustrate obtaining data for the base of a percentage. The percentage of interest is the percent of workers in Miami City who usually carpooled to work in 2000. The base of the percentage depends on whether workers who worked at home were included as a mode of commuting in computing the percentage. If working at home is included as a mode, the base would be the total number of workers. If working at home is not included as a mode, the base would be the total number of workers minus the number of workers who worked at home.

Figure 3 shows Detailed Table QT-P23 from Census 2000 Summary File 3. The base of this percentage is 126,536 if this percentage is computed with working at home as a mode, but is 123,931 if working at home is not included as a mode.

Distribution Table

The specific distribution table to be used for a specific analysis depends on the specific average or median. For selection purposes, the focus should be on the characteristic and whether the average or median is about persons, workers, households, families, or housing units. Two examples are considered.

Example 1*.* According to Detailed Table P53, Median Household Income in1999 (dollars), for Census 2000 Summary File 3 (SF 3), the median household income in Miami City was $23,483 in 1999. In this example, the metric is household income, and the median relates to households. As a result, the distribution table should show the number of households by ranges of household income. Detailed Table P52, MEDIAN HOUSEHOLD INCOME IN 1999 (DOLLARS), is such a table for Miami City:

Table 4. A Distribution Table of Household Income, Miami, Florida

Range of Household Income	Households
Less than $10,000	32,558
$10,000 to $14,999	14,370
$15,000 to $19,999	12,080
$20,000 to $24,999	11,007
$25,000 to $29,999	9,128
$30,000 to $34,999	8,152
$35,000 to $39,999	6,763
$40,000 to $44,999	5,737
$45,000 to $49,999	4,536
$50,000 to $59,999	7,360
$60,000 to $74,999	7,124
$75,000 to $99,999	6,458
$100,000 to $124,999	3,319
$125,000 to $149,999	1,510
$150,000 to $199,999	1,581
$200,000 or more	2,661

Quick Tables

You are here: Main ▸ Data Sets ▸ Data Sets with Quick Tables ▸ Geography ▸ Tables ▸ Results

Use the links above to change your results | Options | Print / Dow

QT-P23. Journey to Work: 2000
Data Set: Census 2000 Summary File 3 (SF 3) - Sample Data
Geographic Area: **Miami city, Florida**

NOTE: Data based on a sample except in P3, P4, H3, and H4. For information on confidentiality protection, sampling error, nonsampling error, definitions, and count corrections see http://factfinder.census.gov/home/en/datanotes/expsf3.htm.

Subject	Number	Percent
MEANS OF TRANSPORTATION AND CARPOOLING		
Workers 16 and over	**126,539**	**100.0**
Car, truck, or van	102,183	80.8
Drove alone	81,561	64.5
Carpooled	20,622	16.3
In 2-person carpool	15,311	12.1
In 3-person carpool	2,945	2.3
In 4-person carpool	1,185	0.9
In 5- or 6-person carpool	770	0.6
In 7-or-more-person carpool	411	0.3
Workers per car, truck, or van	1.13	(X)
Public transportation	14,382	11.4
Bus or trolley bus	12,676	10.0
Streetcar or trolley car (público in Puerto Rico)	125	0.1
Subway or elevated	948	0.7
Railroad	338	0.3
Ferryboat	0	0.0
Taxicab	295	0.2
Motorcycle	68	0.1
Bicycle	700	0.6
Walked	4,646	3.7
Other means	1,952	1.5
Worked at home	2,608	2.1

Figure 3. An Example of the Base of a Percentage

Example 2. This example illustrates the selection of a distribution table for deriving measures of precision for workers' average travel time to work. "Detailed Tables" for Census 2000 Summary File 3 do not have any published table that shows average travel time to work. However, Quick Table QT-P23 shows 28.1 minutes as the average travel time to work for Miami City. In this case, the metric is travel time to work and the average is about workers. As a result, we need a distribution table that shows the number of workers who did not work at home by ranges of travel time to work.

Detailed Table P31 has information for such a distribution table. Table 5 shows the information from Detailed Table P31 for Miami City. Only the rows in italic would be the distribution table to derive measures of precision for the average travel time to work.

Table 5. Detailed Table P31, Miami City, Florida

Range of Travel Time	Workers
Total:	126,539
Did not work at home	123,931
Less than 5 minutes	*1,832*
5 to 9 minutes	*8,094*
10 to 14 minutes	*14,858*
15 to 19 minutes	*21,137*
20 to 24 minutes	*20,206*
25 to 29 minutes	*6,890*
30 to 34 minutes	*24,268*
35 to 39 minutes	*2,428*
40 to 44 minutes	*4,059*
45 to 59 minutes	*8,211*
60 to 89 minutes	*7,703*
90 or more minutes	*4,245*
Worked at home	2,608

Percent in Sample

Table 6 shows the location of these percent-in-sample values for Census 2000 at American FactFinder and actual values for Miami City.

Table 6. American FactFinder "Detailed Tables" for Percent in Sample Values

Table	Description	Type of Characteristics	Miami City, Florida
P4	PERCENT OF THE POPULATION IN SAMPLE	Population	12.0
H4	PERCENT OF HOUSING UNITS IN SAMPLE BY OCCUPANCY STATUS	Occupied Housing	12.3

Design Factors

The design factors for Census 2000 are in Table C, Chapter 8 of *Census 2000 Summary File 3 Technical Documentation* at http://www.census.gov/prod/cen2000/doc/sf3.pdf. This PDF document also can be accessed by clicking "Technical Documentation (PDF)" located at the lower right-hand corner of Figure 1. Once this PDF file is open, these tables are assessable via its bookmarks. Since this set of design factors will not change in the future, they have been included in a worksheet, **2000 DF**, of the *ACS Statistical Analyzer* for easy access by the user.

The design factors for an ACS PUMS, however, need to be obtained from American FactFinder. These are contained in the PUMS Accuracy of the Data report. This report may be obtained by clicking “Public Use Micro Data Sample (PUMS)” located at the upper right-hand corner of Figure 2. These design factors for up to 2008 were calculated using 2005 ACS data. They may change, however, for future PUMS data and will be updated periodically.

REFERENCES

These references may not be directly cited in this document, but information from them was used in developing the *ACS Statistical Analyzer* and this document.

Gaines, Leonard, *ACS Advanced Applications and Issues*, Technical Appendix to the ACS User Handbook, unpublished draft, not dated. (https://ctools.*umich*.edu/access/content/group/34a72eab-daa4-4d14-80e0-9150727aed6c/Technical%20-%20Statistical/ ACS%20Technical%20Appendices_Gaines.pdf).

U.S. Census Bureau, *PUMS Accuracy of the Data (2008)*, Section 6.1, not dated. http://www.census.gov/acs/www/Downloads/2008/AccuracyPUMS.pdf.

U.S. Census Bureau, Census 2000, *Public Use Microdata Sample, (PUMS), United States, Technical Documentation*, *Chapter* 4 - Accuracy of the Microdata Sample Estimates, U.S. Government Printing Office, Washington, D.C., 2003.

U.S. Census Bureau, *Census 2000 Summary File 3 Technical Documentation*, Chapter 8 - Accuracy of the Data and Chapter 9 - Technical Documentation Notes, U.S. Government Printing Office, Washington, DC, 2007. (http://www.census.gov/prod/cen2000/doc/sf3.pdf)

U.S. Census Bureau, *A Compass for Understanding and Using American Community Survey Data: What Researchers Need to Know*, Appendix 3, U.S. Government Printing Office, Washington, DC, 2009. (www.census.gov/acs/www/Downloads/ACSResearch.pdf)

This page intentionally left blank.

APPENDIX

This appendix provides the statistical procedures and formulas involved in the calculations for each of the 15 sub-functions of the *ACS Statistical Analyzer*. The steps, symbols, formulas, and related descriptions for each sub-function are adopted from the references.

Function A

A01 – ACS derives other precision measures for individual published ACS estimates that already have an MOE at the 90% confidence level. This function derives an estimate's relative reliability and its confidence interval at the current confidence level. At an alternative confidence level, this function also derives a new MOE and a new confidence interval. These calculations require deriving the standard error of an estimate from its MOE.

Deriving the Standard Error from the MOE

For an MOE at a given confidence level α, the associated standard error is expressed as:

$$\text{Standard Error} = \frac{\text{MOE}_\alpha}{\text{M}_\alpha}$$

In this formula M_α is the confidence multiplier, which is 1.645 for the 90% confidence level, 1.96 for the 95% confidence level, and 2.576 for the 99% confidence level. If working with published ACS 1-year estimates for 2005 or earlier, use the value 1.65 rather than 1.645 as the confidence multiplier.

Deriving the Relative Reliability

The relative reliability is traditionally called the coefficient of variation for an estimate. It is given by the ratio of the estimate's standard error over the estimate. It is stated in percent terms. Specifically, the relative reliability of an estimate is:

$$\text{Relative Reliability} = \frac{\text{SE}}{\text{X}} \cdot 100$$

where X is the estimate and SE is its standard error.

Deriving the Confidence Interval from the MOE

For an estimate X and its MOE at a given confidence level α, the lower and upper bounds of its confidence interval at the same confidence level are given by:

$$L_\alpha = X - \text{MOE}_\alpha$$

$$U_\alpha = X + \text{MOE}_\alpha$$

Users should consider logical boundaries when creating confidence intervals from an estimate's MOE:

- The lower boundary may be less than zero as calculated above for an estimate that is small; a negative number does not make sense in most cases of using ACS and Census 2000 data. So the lower boundary should be set to zero. One exception to this consideration would be a confidence interval for an estimate of a difference between two existing estimates.

- The upper boundary may be greater than 100% as calculated using the formulas above for an estimate of a percentage that is close to 100%. A percentage greater than 100% does not make sense; so the upper bound in this case should be set to 100%.

Deriving the MOE for an Alternative Confidence Level

For an MOE_α at a given confidence level α, the MOE_β at an alternative confidence level β can be expressed as:

$$MOE_\beta = \frac{M_\beta}{M_\alpha} \cdot MOE_\alpha$$

In this formula, M_α and M_β are the confidence multipliers for confidence levels α and β, respectively. If working with published ACS 1-year estimates for 2005 or earlier for the base MOE, use the value 1.65 rather than 1.645 for the confidence multiplier.

Reference

These procedures are from the following document:

U.S. Census Bureau, *A Compass for Understanding and Using American Community Survey Data: What Researchers Need to Know*, Appendix 3, pp. A-12, A-13, U.S. Government Printing Office, Washington, DC, 2009. (www.census.gov/acs/www/Downloads/ACSResearch.pdf)

Function B

B02 - ACS Direct

This sub-function derives measures of precision for frequencies, totals, averages, or medians from an ACS PUMS using replicate estimates. Replicate estimates can be used to calculate what is referred to as direct standard errors. Standard errors for the published ACS estimates are calculated using replicate estimates. Direct standard errors will often be more accurate than those standard errors derived from the other sub-functions of Function B. The advantage of using replicate estimates is that a single formula is used to calculate the standard error of many forms of estimates.

Each housing unit and person record contains a full weight and 80 replicate weights. The full weight is used to derive the estimate of interest X. For this discussion, X is referred to as the full sample estimate. The 80 replicate weights are used to derive 80 replicate estimates. The first replicate estimate X_1 is computed using the first replicate weight, the second replicate estimate X_2 is computed using the second replicate weight, and so on. Each replicate estimate is computed using the replicate weights in the same way that the full sample estimate X is computed.

The standard error of X is estimated using the sum of the squared differences between each replicate estimate X_r and the full sample estimate X. The standard error formula is:

$$SE(X) = \sqrt{\frac{4}{80}\sum_{r=1}^{80}(X_r - X)^2}$$

This formula fails to provide correct measures of precision for the following conditions:

- The estimate of X is zero
- The estimated MOE for a median is zero
- The estimate of a frequency is controlled at some aggregate level, such as total male or total female persons living in households in a state

Switch to using the other sub-functions of Function B in these cases:

- **B03 - Average** if the estimate of interest is an average
- **B04 - Median** if the estimate of interest is a median.
- **B05 - Frequency** if the estimate of interest is a frequency.

These procedures are from the following:

U.S. Census Bureau, *PUMS Accuracy of the Data (2008)*, Section 6.1, not dated. http://www.census.gov/acs/www/Downloads/2008/AccuracyPUMS.pdf.

B03 – Average

This sub-function derives measures of precision for averages using a distribution table. The formula for estimating the standard error of an average, x, is

$$\mathrm{SE}(x) = \sqrt{\left(\frac{f}{\text{base}} \cdot s^2\right)} \cdot \text{Design Factor}$$

In this equation, f is the finite population correction factor, the base is the sum of the frequencies in the distribution table, and s^2 may be estimated using data in the distribution table. Use Table 2 in the section on data needs and the source of data for the distribution table to determine the finite population correction factor.

For this method, the value for the characteristic is divided into c ranges, where the lower and upper boundaries of range j are L_j and U_j, respectively. Each unit is placed into one of the c ranges such that the value of the characteristic is between L_j and U_j. The estimated s^2 is then given by:

$$s^2 = \sum_{j=1}^{c} p_j m_j^2 - x^2$$

where p_j is the estimated proportion of units in range j (based on weighted data) and m_j is the midpoint of the j^{th} range, calculated as:

$$m_j = \frac{L_j + U_j}{2}$$

If the c^{th} range is open-ended, (i.e., no upper range boundary exists) then approximate m_c by:

$$m_c = \left(\frac{3}{2}\right) L_c$$

The above procedures are applicable to both Census 2000 and ACS PUMS data. For Census 2000 data, they are the recommended method by the U.S. Census Bureau in the following document:

> U.S. Census Bureau, Census 2000, *Public Use Microdata Sample, (PUMS), United States, Technical Documentation*, *Chapter* 4 - Accuracy of the Microdata Sample Estimates, p. 4-6, U.S. Government Printing Office, Washington, D.C., 2003.

For ACS PUMS data, these procedures will give an approximation for the measures of precision of an average. The U.S. Census Bureau, however, has recommended a different method to approximate the measures of precision for an average. For 2004 or earlier ACS data, the recommended method is in the following document:

> U.S. Census *Bureau*, PUMS Accuracy of the Data (2004).
> http://www.census.gov/*acs*/www/Downloads/2004/AccuracyPUMS.pdf.

For 2005 or later ACS data, the recommended method is in the following document:

> U.S. Census Bureau, *PUMS Accuracy of the Data (2008)*, Section 6.1, not dated.
> http://www.census.gov/acs/www/Downloads/2008/AccuracyPUMS.pdf.

The Census-recommended method for 2005 or later ACS data requires data at the individual household level and is not included in the current tool. If users want exact measures of precision of an average, they should use sub-function **B02 - ACS Direct.**

B04 – Median

This sub-function derives measures of precision for medians using a distribution table. It involves the following steps:

1. From Table 2 determine the finite population correction factor f.
2. Obtain the appropriate (person or housing unit) observed percent in sample value for the specific geographic area. Use this value to locate the design factor for the characteristic of interest.
3. Obtain the distribution table for the selected variable. Cumulate these frequencies to yield the base.
4. Determine the standard error of the estimate of 50 percent from the distribution using the formula:

$$\text{SE(50 percent)} = \sqrt{\left(\frac{f}{\text{base}} \cdot 50^2\right)} \cdot \text{Design Factor}$$

5. Subtract from and add to 50 percent the standard error determined in step 4.

$$\text{p\_lower} = 50 - \text{SE(50 percent)}$$

$$\text{p\_upper} = 50 + \text{SE(50 percent)}$$

6. Determine the range in the distribution table containing p_lower and the range in the distribution table containing p_upper. If p_lower and p_upper fall in the same range, follow the steps below. If p_lower and p_upper fall in different ranges, go to step 9.

    - Define A1 as the smallest value in that range.
    - Define A2 to be the smallest value in the next (higher) range.
    - Define C1 as the cumulative percent of units strictly less than A1.
    - Define C2 as the cumulative percent of units strictly less than A2.

7. Use the following formulas with p_lower, p_upper, A1, A2, C1, and C2 to determine a lower bound and an upper bound:

$$\text{Lower Bound} = \left[\frac{\text{p\_lower} - \text{C1}}{\text{C2} - \text{C1}}\right] \cdot (\text{A2} - \text{A1}) + \text{A1}$$

$$\text{Upper Bound} = \left[\frac{\text{p\_upper} - \text{C1}}{\text{C2} - \text{C1}}\right] \cdot (\text{A2} - \text{A1}) + \text{A1}$$

8. Divide the difference between the lower and upper bounds, determined in step 7 above, by two to obtain the estimated standard error of the median:

$$\text{SE(median)} = \frac{\text{Upper Bound} - \text{Lower Bound}}{2}$$

9. For the range:

    a. containing p_lower, define the values A1, A2, C1, and C2 as described in step 6 above. Use these values and the formula in step 7 to obtain the Lower Bound.
    b. containing p_upper, define a new set of values for A1, A2, C1, and C2 as described in step 6. Use these values and the formula in step 7 to obtain the Upper Bound.

10. Use the Lower Bound and Upper Bound obtained in step 8 and the formula in step 7 to calculate the standard error of the estimated median.

The above procedures are from the following:

> U.S. Census Bureau, Census 2000, *Public Use Microdata Sample, (PUMS), United States, Technical Documentation*, *Chapter* 4 - Accuracy of the Microdata Sample Estimates, pp. 4-4, 4-5, U.S. Government Printing Office, Washington, D.C., 2003.

B05 – Frequency

This sub-function derives measures of precision for frequencies from Census 2000 or a PUMS using design factors. The formula for calculating the standard error for an estimate of a frequency is given by:

$$\text{SE(Y)} = \sqrt{\left[\text{f} \cdot \text{Y}\left(1 - \frac{\text{Y}}{\text{N}}\right)\right]} \cdot \text{Design Factor}$$

In this formula, f is the finite population correction factor, Y is the estimate of a frequency, and N is the size of the geography in the same unit as the estimate of the frequency. If an estimated frequency is less than 425 or within 425 of the total size of the tabulation area, use a basic standard error of 246 multiplied by the design factor for the estimate.

The above procedures are from the following:

U.S. Census Bureau, Census 2000, *Public Use Microdata Sample, (PUMS), United States, Technical Documentation*, *Chapter* 4 - Accuracy of the Microdata Sample Estimates, p. 4-8, U.S. Government Printing Office, Washington, D.C., 2003.

While this document or any other document from the U.S. Census Bureau containing this formula does not explicitly make it clear, this formula applies to estimates of the number of persons, households, families, or housing units only. It does not apply to estimates of totals such as total household income, total commuting time by workers, total number of available household vehicles, total rent, etc. in a geographic area.

B06 – Percentage

This sub-function derives measures of precision for frequency-based percentages using design factors. The formula for calculating the standard error for an estimate of a frequency-based percentage is given by:

$$SE(p) = \sqrt{\left[\frac{f}{B} \cdot p(1-p)\right]} \cdot \text{Design Factor}$$

In this formula, f is the finite population correction factor, p is the estimate of a percentage, and B is the base of the percentage. If an estimate of a percentage is smaller than 2% (greater than 98%), the value of 2% (98%) is used when using this formula.

The above procedures are from the following:

U.S. Census Bureau, Census 2000, *Public Use Microdata Sample, (PUMS), United States, Technical Documentation*, *Chapter* 4 - Accuracy of the Microdata Sample Estimates, p. 4-9, U.S. Government Printing Office, Washington, D.C., 2003.

While this document or any other document from the U.S. Census Bureau containing this formula does not explicitly make it clear, this formula applies to estimates of percentages derived from estimates of frequencies such as persons, households, families, or housing units only. The percentage of workers who usually commuted to work by bus in a state would be an example. It does not apply to estimates of percentages that are derived in terms of totals such as total household income, total commuting time by workers, total number of vehicles available, total rent, etc. in a geographic area. The percentage of vehicles available in households under poverty versus all households in an area would be an example of these percentages.

Function C

This function derives new estimates and their precision measures from two or more original estimates that already have an MOE. These new estimates may be derived from one of the following six operations:

C07 – Sum derives the sum of two or more estimates and its measures of precision. The MOE for a sum is calculated as:

$$\text{MOE}_{\text{sum}} = \sqrt{\sum_{c} \text{MOE}_c^2}$$

where MOE_c is the MOE of component estimate c.

C08 – Diff derives the difference of two estimates and its measures of precision. The procedure to calculate the MOE of a sum discussed previously should be used here to obtain the MOE of a difference.

C09 - %Diff derives the percent difference of two estimates and its measures of precision. The procedure to calculate the MOE of a ratio discussed next should be used to obtain the MOE of the percent difference.

C10 – Ratio derives a ratio from two estimates and its measures of precision with the numerator not being part of the denominator. The MOE for a ratio is calculated as:

$$\text{MOE}_{\text{ratio}} = \frac{\sqrt{\text{MOE}_{\text{num}}^2 + \left(\text{R}^2 \cdot \text{MOE}_{\text{den}}^2\right)}}{\text{X}_{\text{den}}}$$

where:

MOE_{num} of the MOE of the numerator.
MOE_{den} is the MOE of the denominator.
$\text{R} = \text{X}_{\text{num}}/\text{X}_{\text{den}}$ is the derived ratio.
X_{num} is the estimate used as the numerator of the derived ratio.
X_{den} is the estimate used as the denominator of the derived ratio.

C11 – Percentage derives a percentage from two estimates and its measures of precision with the numerator being part of the denominator. The MOE for a derived percentage is calculated as:

$$\text{MOE}_{\text{percentage}} = \frac{\sqrt{\text{MOE}_{\text{num}}^2 - \left(\text{p}^2 \cdot \text{MOE}_{\text{den}}^2\right)}}{\text{X}_{\text{den}}}$$

where:

MOE_{num} of the MOE of the numerator.
MOE_{den} is the MOE of the denominator.
$\text{p} = \text{X}_{\text{num}}/\text{X}_{\text{den}}$ is the derived percentage.
X_{num} is the estimate used as the numerator of the derived percentage.
X_{den} is the estimate used as the denominator of the derived percentage.

There are rare instances where the value under the square root will be negative. If that happens, use the formula for derived ratios which will provide a conservative estimate of the MOE.

C12 – Product derives the product of two estimates and its measures of precision. The MOE for a product is calculated as:

$$\text{MOE}_{\text{product}} = \sqrt{(\text{A}^2 \cdot \text{MOE}_{\text{B}}^2) + (\text{B}^2 \cdot \text{MOE}_{\text{A}}^2)}$$

where A and B are the first and second estimates, respectively, MOE_A is the MOE of the first estimate, and MOE_B is the MOE of the second estimate.

These procedures and formulas are from the following:

U.S. Census Bureau, *A Compass for Understanding and Using American Community Survey Data: What Researchers Need to Know*, Appendix 3, pp. A-14 through A-17, U.S. Government Printing Office, Washington, DC, 2009. (www.census.gov/acs/www/Downloads/ACSResearch.pdf)

Function D

Function D compares pairs of two estimates that already have an MOE. There are three sub-functions:

- One ACS estimate with another (**D13 - ACS&ACS**)
- One ACS estimate with a Census 2000 estimate using an actual MOE for the 2000 estimate (**D14 - ACS&2000 Actual**)
- One ACS estimate with a Census 2000 estimate assuming the MOE for the Census 2000 estimate = the MOE for the ACS estimate (**D15 - ACS&2000 Assumed**). This sub-function assumes that the MOE for a Census 2000 estimate is smaller than the ACS estimate. This assumption may be violated for some cases, such as controlled population.

For all three sub-functions, the test of significance can be expressed as follows. If the following is true

$$\left|\frac{X_1 - X_2}{\sqrt{SE_1^2 + SE_2^2}}\right| > Z_{CL}$$

then the difference between estimates X_1 and X_2 is statistically significant at the specified confidence level, CL. In the above equation:

- X_i is estimate i (=1,2)
- SE_i is the standard error for estimate i (=1,2)
- Z_{CL} is the critical value for the desired confidence level (=1.645 for 90 percent, 1.960 for 95 percent, 2.576 for 99 percent).

These procedures are from the following:

U.S. Census Bureau, *A Compass for Understanding and Using American Community Survey Data: What Researchers Need to Know*, Appendix 3, pp. A-18, U.S. Government Printing Office, Washington, DC, 2009. (www.census.gov/acs/www/Downloads/ACSResearch.pdf)