The American Community Survey (ACS) Statistical Analyzer

- Xuehao Chu
- Center for Urban Transportation Research

What do we cover today?

- ACS
- Precision
- Usage
- An Excel tool
- Demonstration
- Guidance

What is the ACS?

- Designed & administered by the U.S. Census Bureau
- Instrument is similar to the traditional long form
  - Socio-demographic and housing
  - Economic and journey-to-work
- Administration is different from the Census long-form
  - Residence rule - current vs. usual
  - Measurement - continuous vs. snapshot (April 1st)
  - Reference period for some characteristics
  - Frequency - annual vs. every 10 years
  - Sample size - 1 in 40 (2.5%) vs. 1 in 6 (16.7%)

What data are available?

- Type of data products
  - Pre-derived estimates
  - Sub-samples
- Format of pre-derived estimates
  - Published tables (at American FactFinder)
  - ACS Summary File (at ACS FTP site)
  - CTPP tables (FHWA's CTPP site)
- Type of pre-derived estimates
  - Place of residence characteristics
  - Place of work characteristics
- Sub-samples
  - 1% PUMS by state (at American FactFinder)

How do the data depend on their period and geography?

Population in the Geography of Data	Period of Data		
	1-Year	3-Year	5-Year
Large: 65,000+	X	X	X
Medium: 20,000-65,000		X	X
Small: under 20,000 (include block groups)			X

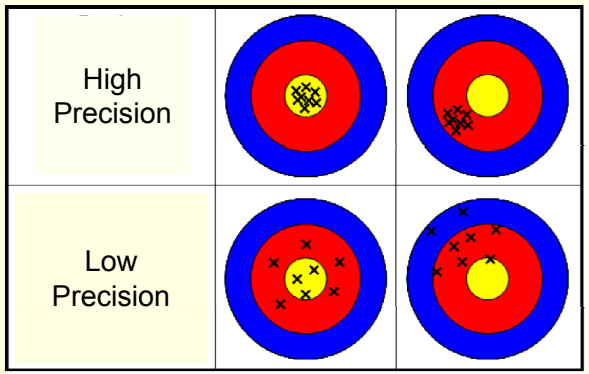
What forms do estimates take?

- Frequencies (i.e., persons, households, housing units)
- Totals (i.e., total income, commute time, etc.)
- Averages
- Medians
- Ratios
- Percentages
  - Frequency-based percentages
  - Total-based percentages

What guidance do experts offer?

- "As the ACS estimates are based on a [relatively small] sample survey of the U.S. population, information about the sampling error [i.e., precision] 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."

What does the precision of an estimate indicate?



What metrics can be used to measure precision?

- Independent of confidence:
  - **Standard Error** (SE)
  - **Relative Reliability** (RR)

$$RR = 100 * SE / Estimate$$

- Dependent on confidence:
  - **Margin of Error** (MOE)

$$MOE_{90\%} = SE * 1.645$$

  - **Confidence Interval** (CI) =

$$(Estimate - MOE_{90\%}, Estimate + MOE_{90\%})$$

How can measures of precision be obtained?

- Published ACS estimates come with an MOE but not other measures of precision
- Most CTPP ACS estimates come with an MOE but not other measures of precision
- Users need to derive measures of precision
  - Census 2000 published
  - CTPP 2000 estimates
  - Census 2000 PUMS data
  - User-derived estimates from published ACS estimates
  - User-derived estimates from an ACS PUMS

What the data may be used for?

- Indicating conditions of a geography or a population group using individual estimates, e.g.:
  - Average commuting time
  - Median household income
  - % workers commuting to work by transit
- Indicating differences by comparing estimates
  - Over time
  - Across population groups
  - Across geographic areas
- *Input data for multivariate analyses*
- *Input data for modeling*

Are my estimates usable?

- MOE is an absolute indicator of precision & ineffective
- RR is a relative indicator & more effective
- General guidelines for using RR
  - Usable if RR < 10%
  - Cautious to use if 10% < RR ≤ 50%
  - Avoid if RR > 50%
- Consider the importance of an estimate
  - Use lower RR for more important usage (e.g., funding)

How can precision be improved?

- For a given period
  - Combine smaller geographies into a larger one
  - Combine smaller population groups into a larger one
- For a given geography or population group
  - Use estimates for a longer period

How can two estimates be compared statistically?

- Conduct a test of significance
- Select a confidence level for the test
- Calculate a test statistic based on the estimates and their precision levels
- Determine the critical value for the selected confidence
- If test statistic > critical value, the difference is statistically significant at the selected confidence level
- If test statistic ≤ critical value, the difference is not statistically significant
- Use higher confidence if
  - The conclusion is important (e.g., funding related)
  - Multiple pairs are compared

How does the *ACS Statistical Analyzer* help?

- It derives measures of precision
- It tests if the difference between two estimates is statistically significant at a given confidence level
- It does these for six different forms of estimates
- It focuses on ACS but deals with Census 2000 too
- It is a 2003 Excel-based template
- It has 4 functions (A, B, C, D)
- It has 15 sub-functions
- Users just need to enter the required data
- It takes care of the statistical procedures and formulas involved

What does Function A do?

- Focuses on estimates with 90% MOE but no other measures of precision
- One sub-function **A01 – ACS**:
  - Derives other precision measures at 90% confidence
  - Derives all precision measures at an alternative confidence level
  - Allows up to 100 estimates, preferably from the same published ACS or CTPP table

What does Function B do?

- Derives measures of precision for estimates without any measure of precision
- Five sub-functions
  - **B02 – ACS Direct**: ACS PUMS estimates using replicate estimates
  - **B03 – Average**: averages using a distribution table
  - **B04 – Median**: medians using a distribution table
  - **B05 – Frequency**: frequencies using a design factor
  - **B06 – Percentage**: frequency-based percentages using a design factor

What are replicate estimates?

- Supplemental estimates to any ACS estimate that a user can obtain from an ACS PUMS
- Every ACS PUMS file comes with
  - A full weight for obtaining the ACS estimate
  - 80 replicate weights for obtaining the replicate estimates
- The ACS estimate and a total of 80 replicate estimates from an ACS PUMS can be used to determine the precision measures for the ACS estimate
- Used for sub-function **B02 – ACS Direct**
- Users need have these replicate estimates derived before using **B02**

What is a distribution table?

- Number of persons, workers, households, families, or housing units by ranges of a characteristic on a numerical scale
- Used by sub-functions **B03 – Average** and **B04 – Median**

Range of Travel Time to Work	Workers
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

What is a design factor?

- Factor to adjust estimates to reflect how much a survey deviates from simple random sampling
- For Census 2000, design factors vary by the state of the geography, by the characteristic for the estimate, and the sample size (provided in the Analyzer)
- For ACS, design factors vary by the state of the geography and by the characteristic for the estimate
- If an estimate is for a combination of two or more characteristics, use the largest design factor for this combination of characteristics
- Used by sub-functions **B03-B06**

What does Function C do?

- Derives measures of precision for estimates obtained from two or more other estimates with an MOE
- Covers six operations:
  - **C07 – Sum**: sum of up to 200 estimates
  - **C08 – Diff**: difference of two estimates
  - **C09 – %Diff**: percent difference of two estimates
  - **C10 – Ratio**: ratio of one estimate over another
  - **C11 – Percentage**: percent of one estimate in another
  - **C12 – Product**: product of two estimates

What does Function D do?

- Tests whether a difference between two estimates truly exists at a given confidence level
  - **D13 – ACS&ACS**: Two ACS estimates with an MOE
  - **D14 – ACS&2000 Actual**: ACS and Census 2000 estimates with an MOE
  - **D15 – ACS&2000 Assumed**: ACS and Census 2000 estimates with 2000 MOE = ACS MOE
    - Valid if the difference is found statistically significant
    - Invalid otherwise

How do I use the *ACS Statistical Analyzer*?

- We briefly demonstrate for six sub-functions:
  - **A01 – ACS**
  - **B01 – ACS Direct**
  - **B02 – Average**
  - **B03 – Percentage**
  - **C07 – Sum**
  - **D14 – ACS&2000 Actual**

What should you avoid when comparing ACS estimates?

- Avoid ACS estimates that overlap
  - Between geographic areas
  - Between population groups
  - Between multiyear periods
- Avoid these comparisons for multiyear estimates
  - A multiyear period with a single-year period
  - Two multiyear periods with different lengths
  - A pre-2006 period with a post-2006 one (including 2006)

What should you consider when comparing ACS and 2000 estimates?

- Effects of different residence rules
  - Usual residence for Census 2000
  - Current residence of two months+ for the ACS
- Effects of seasonal variation
  - April 1 for Census 2000
  - Continuous for the ACS
- Effects of different reference periods for income and school attendance
  - Prior calendar year for Census 2000
  - Prior 12 months for the ACS

May I use the *Statistical Analyzer* for place of work characteristics?

- **Must not** use these sub-functions for estimating measures of precision for place of work characteristics from 2000 Census
  - **B03 – Average**: averages using a distribution table
  - **B04 – Median**: medians using a distribution table
  - **B05 – Frequency**: frequencies using a design factor
  - **B06 – Percentage**: percentages using a design factor
- May use **B2 – ACS Direct** for estimating measures of precision for place of work characteristics by using replicate estimates from an ACS PUMS

May I use the *Statistical Analyzer* for totals and total-based percentages?

- **Must not** use **B05 – Frequency** for totals
- **Must not** use **B06 – Percentage** for total-based percentages
- You may use several sub-functions for estimating measures of precision for total household income:
  - Use **B03 – Average** for income per household
  - Use **B05 – Frequency** for number of households
  - Use **C12 – Product** for total household income
- Similarly you may use multiple sub-functions for total-based percentages

Do you have any questions?

THANK YOU!

- xchu@cutr.usf.edu
- 813-974-9831