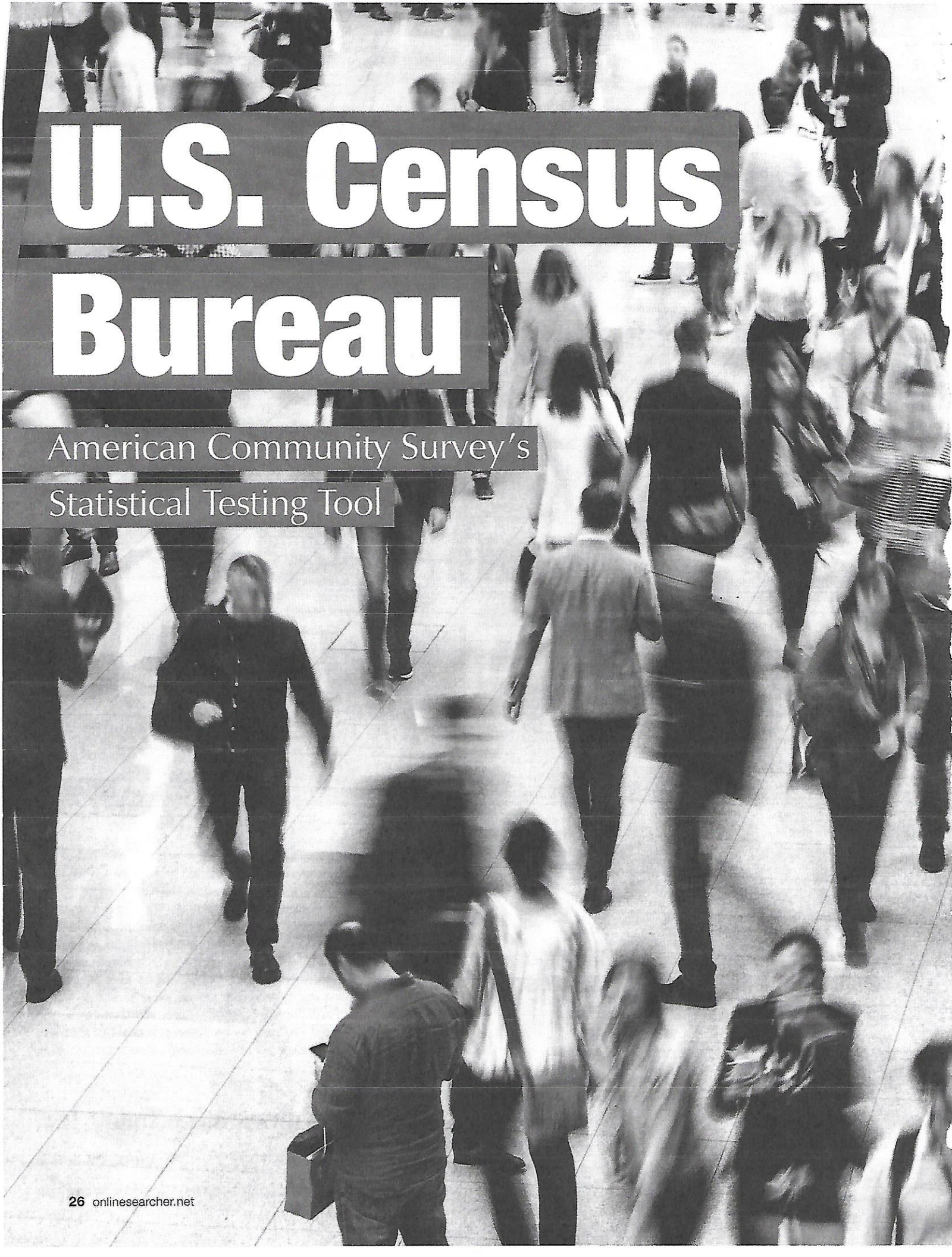
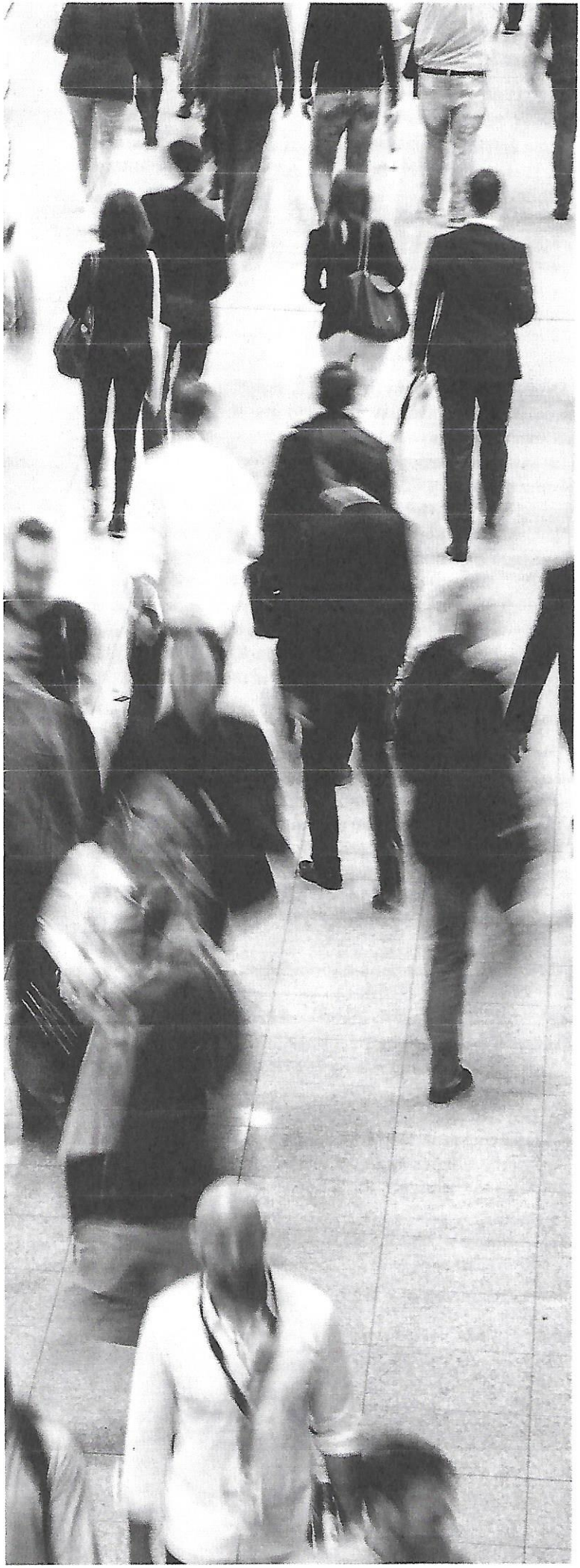


U.S. Census Bureau

American Community Survey's Statistical Testing Tool



By Roger Magnus

You can criticize governments for many things, but extensive data collection is not one of them. The U.S. government, in particular, has raised the art and science of data gathering to a new level. The U.S. Census Bureau's American Community Survey (ACS) is a case in point, containing an overwhelming amount of information. To its credit, it also provides the Statistical Testing Tool that can save researchers time and allow them to better use the data.

Unlike the U.S. Census,

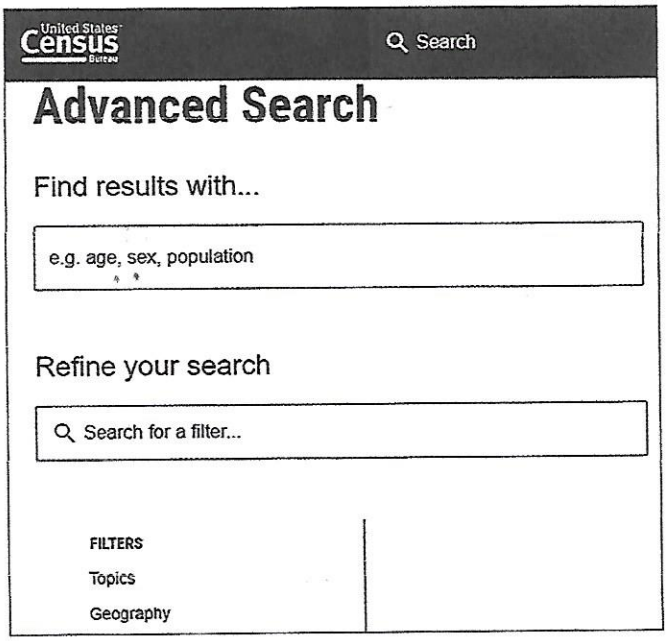
which counts the U.S. population every 10 years, the ACS is a rolling monthly survey that publishes 1-year and 5-year average data depending on the population size of the geography reported. It asks more questions than does the Census, and the answers form the basis of decisions about education, health services, transportation, and other government services. Additionally, entrepreneurs and established businesses use the information to decide on new products, markets, jobs, and locations. Journalists rely on the data to create stories, such as how many children were enrolled in kindergarten or how many veterans from World War II are still alive.

To get maximum advantage from ACS data, it's important to understand how to use its statistical tool. It is not a database nor a website but a specially programmed Excel file that came out in 2016. You can download the tool at the Census website (census.gov/programs-surveys/acs/guidance/statistical-testing-tool.html), which also lists its features. These features include being able to compare more than 3,200 estimate pairs or a matrix of up to 150 multiple estimates at one time.

ACS ESTIMATES

The ACS contains estimates with not just one datapoint but three. This is because most estimates contain a Margin of Error (MOE) and are in the middle of a range between a lower limit and upper limit with a 90% confidence interval or 90% certainty that the real number falls within. These

>



The Statistical Testing Tool's Advanced Search page allows for more precise searching than its basic search.

MOEs can be sizeable because the ACS sample size surveyed in 2017 was quite small—only 3.52 million households out of 137.40 million. That's only about one in 40 (2.5%). While the MOEs need to be mentioned in any kind of data reporting (i.e., estimate +/- MOE), this is particularly so whenever two or more estimates are computed (such as being added together for age ranges in a particular geography) or compared between geographies or years.

You do not just add up or otherwise compute these estimates, ignore the MOEs, and consider the sum or other mathematical solution to be correct. Also, you do not compare two estimates and consider one to be different than another without including the MOEs; the MOEs will determine the actual statistical significance between these estimates.

Unfortunately, both these scenarios occur frequently because many ACS users either are not aware that they need to include the MOEs, do not have the time to do so, or do not know how. While there are numerous guides and presentation on the Census website that detail these computations, these materials are not always easy to find or follow. Fortunately, at least in the case of common ACS data comparisons, the Statistical Testing Tool is a helpful option.

NAVIGATING AROUND THE TOOL

Once you have downloaded the tool, note that there are several tabs at the bottom of this spreadsheet, including the following:

- Overview
- Instructions
- Stat_Test_for_Two_Est
- Stat_Test_for_Multiple_Est
- Worked_Example
- Contact_Us

Overview justifies why the MOE is important, explains that the datapoints need to be entered for the sheet to work, and lists Special Cases involving MOE.

Instructions provides step-by-step directions for using the two and multiple estimate tabs of the spreadsheet and how to interpret the results for each.

Worked_Example displays a real example for multiple estimates showing a matrix of cells displaying Yes and No for statistically significant differences at a 90% confidence interval.

Contact_Us provides phone numbers, an email address, and a list of Additional Resources with links to other ACS training guides, including Handbooks, Presentations, and links to Definitions, Statistical Testing, and Accuracy of the Data.

WORKING WITH TWO ESTIMATES

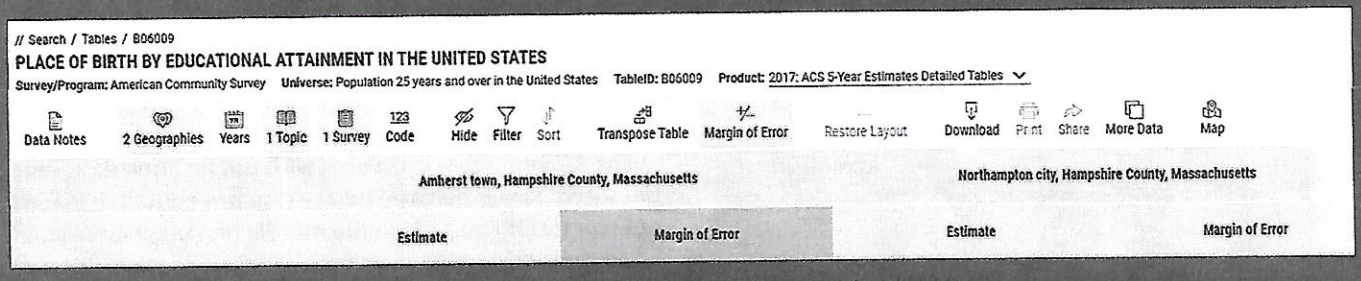
The Stat_Test_for_Two_Est tab contains a split screen. In the top half, there are explanations for Purpose and Results, a seven-step list of instructions, and links to the other tabs. In the bottom half, an actual example is shown.

Below is an example I created following the instructions provided in the tool. For this search, I compared Educational Attainment for Amherst and Northampton, Mass., from the 2012–2016 5-Year ACS Estimate.

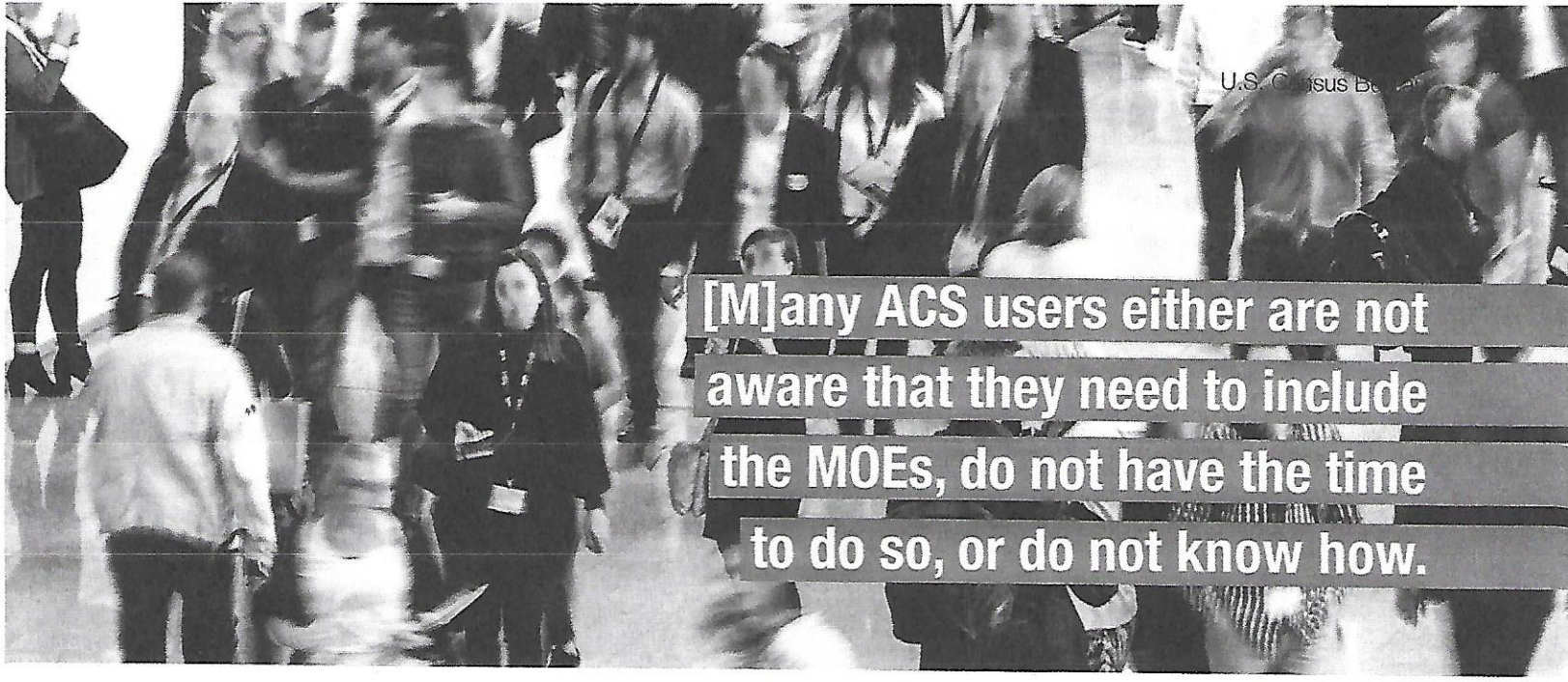
1. **Download ACS data into a spreadsheet.**

Currently, there are two main ways to obtain these data:

- The American FactFinder page (factfinder.census.gov), which will be retired in Spring 2020, had no new data releases after July 1, 2019.
- The Data.census.gov page (data.census.gov) is replacing American FactFinder, but note that the interface is evolving, so directions for accessing the data may change in the future.



The Customize Table provides the ability to manipulate data tables.



Note that not one, but two, CSV files for Microsoft Excel will be downloaded. (There is also a text file which shows the table title.) One contains the data, but the field names are IDs that have corresponding field names in the rows immediately below (if the column width is extended) or, in an easier-to-view listing, in the metadata file. For the variable field names that are being compared for each municipality, ensure that they are clearly displayed in the data file and then copy and paste them into the Statistical Testing Tool file.

To start the search, enter data.census.gov and then go to the Advanced Search page (data.census.gov/cedsci/advanced), which is a more precise way to search this dataset than the Basic search. From there, go to Filters.

- Select Surveys > ACS 5-Year Estimates Detailed Tables. (One-year estimates are only available for areas with populations higher than 65,000. Since Amherst and Northampton have smaller populations, they require 5-year data.)
- Click on Geography > County Subdivision > under County Subdivision, Massachusetts > under Massachusetts Counties, check Hampshire > under Hampshire County, check Amherst and Northampton.
- Select Topics > Education > Educational Attainment, and click on the View All Results button (lower right corner).
- There are three tables available. Click on PLACE OF BIRTH BY EDUCATONAL ATTAINMENT IN THE UNITED STATES Table BO6009.
- Select Customize Table button (upper right corner) > the down arrows (upper right) > and the down arrow to the right of Product to change the year from 2017 to 2016. Select Customize Table again (if needed) to see a screen with more ways to manipulate the data table. The data is averaged across 5 years because both municipalities have populations less than 65,000. Check Show Margins of Error if it is not selected. Download the file into a CSV format by clicking on the Download link, selecting year(s), file type .csv, and Download button in the bottom right corner.

2. **Input statistical variable and geography (and year) keyword(s) into the "Label" column (B).**

Be clear for both in this cell (line 15), as the headers in columns C–F (First Estimate, First MOE, Second Estimate, and Second MOE) are protected and cannot be overwritten. Alternatively, there could be a cell label for statistical variable (line 15) and one below (line 16) listing the geographic variable pair for the two datapoints.

3. **Insert the first number or percentage estimate into the "First Estimate" column (C).**

4. **Insert the corresponding MOE into "First Margin of Error (MOE)" column (D).**

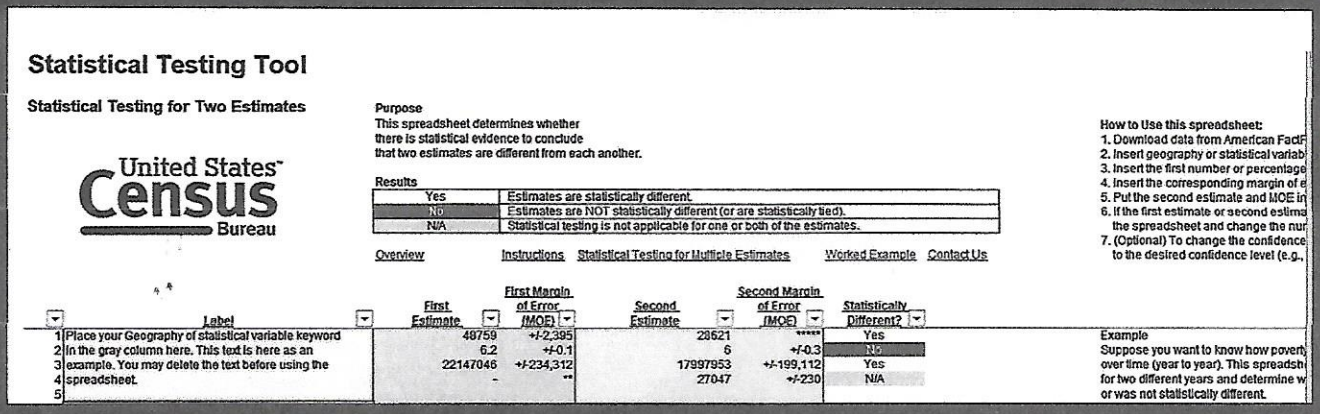
5. **Insert the second estimate and MOE into the "Second Estimate" and "Second Margin of Error (MOE)" columns (E and F), respectively**

You can follow steps 2–5 for up to 3,200 variables comparing two geographies or time frames for one geography. For each one, column R will determine if these two estimates are Statistically Different (Yes, No, or NA).

Note: It is unnecessary to format the variables as they are entered.

There are also options at the very bottom of the sheet to adjust Confidence level (for example, change to 95%) and directions explaining how to handle an estimate containing a Standard Error (SE) rather than an MOE.

For this example, I compared persons in Amherst and Northampton who have Graduate or Professional degrees to see if there is a statistically significant difference. In Amherst, the estimate is 5,144 +/- 437, and in Northampton, the estimate >



Statistical Testing Tool

Statistical Testing for Two Estimates

United States Census Bureau

Purpose
This spreadsheet determines whether there is statistical evidence to conclude that two estimates are different from each another.

Results

Yes	Estimates are statistically different.
No	Estimates are NOT statistically different (or are statistically tied).
N/A	Statistical testing is not applicable for one or both of the estimates.

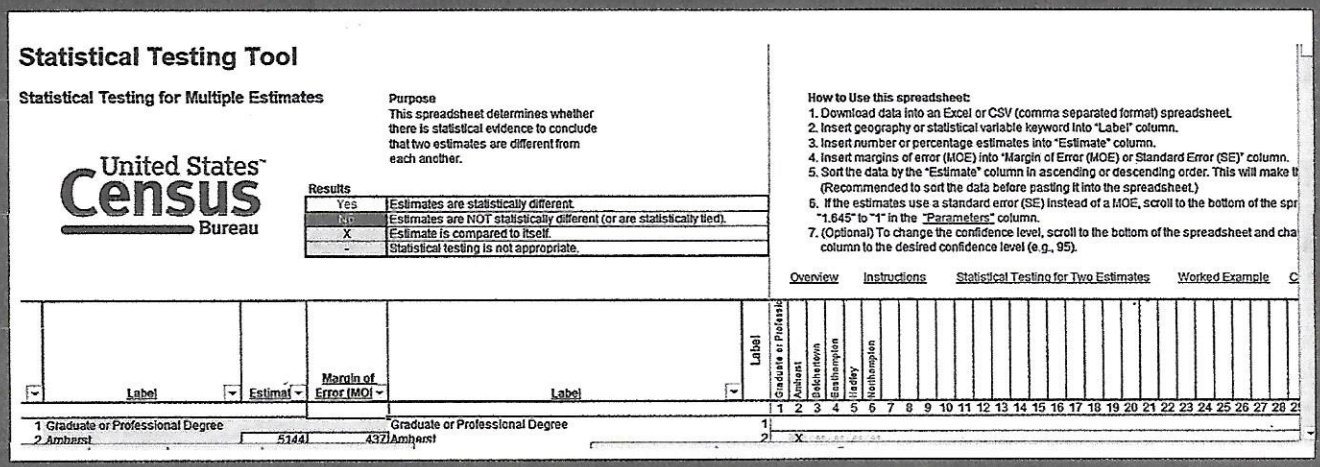
Overview | Instructions | Statistical Testing for Multiple Estimates | Worked Example | Contact Us

How to Use this spreadsheet:
1. Download data from American FactF
2. Insert geography or statistical variab
3. Insert the first number or percentage
4. Insert the corresponding margin of e
5. Put the second estimate and MOE in
6. If the first estimate or second estima the spreadsheet and change the nur
7. (Optional) To change the confidence to the desired confidence level (e.g.,

	Label	First Estimate	First Margin of Error (MOE)	Second Estimate	Second Margin of Error (MOE)	Statistically Different?
1	Place your Geography of statistical variable keyword	48759	+/-2,395	28621	*****	Yes
2	In the gray column here. This text is here as an	6.2	+/-0.1	6	+/-0.3	No
3	example. You may delete the text before using the	22147046	+/-234,312	17997953	+/-199,112	Yes
4	spreadsheet.	-	**	27047	+/-230	N/A
5						

Example
Suppose you want to know how poverty over time (year to year). This spreadsh for two different years and determine w or was not statistically different.

Compare two estimates using the Statistical Testing Tool.



Statistical Testing Tool

Statistical Testing for Multiple Estimates

United States Census Bureau

Purpose
This spreadsheet determines whether there is statistical evidence to conclude that two estimates are different from each another.

Results

Yes	Estimates are statistically different.
No	Estimates are NOT statistically different (or are statistically tied).
X	Estimate is compared to itself.
-	Statistical testing is not appropriate.

How to Use this spreadsheet:
1. Download data into an Excel or CSV (comma separated format) spreadsheet.
2. Insert geography or statistical variable keyword into "Label" column.
3. Insert number or percentage estimates into "Estimate" column.
4. Insert margins of error (MOE) into "Margin of Error (MOE) or Standard Error (SE)" column.
5. Sort the data by the "Estimate" column in ascending or descending order. This will make t (Recommended to sort the data before pasting it into the spreadsheet.)
6. If the estimates use a standard error (SE) instead of a MOE, scroll to the bottom of the spr "1.645" to "1" in the "Parameters" column.
7. (Optional) To change the confidence level, scroll to the bottom of the spreadsheet and cha column to the desired confidence level (e.g., 95).

Overview | Instructions | Statistical Testing for Two Estimates | Worked Example | C

	Label	Estimat	Margin of Error (MO	Label
1	Graduate or Professional Degree			Graduate or Professional Degree
2	Amherst	5144	437	Amherst

	Label	Graduate or Professi	Amherst	Belchertown	Easthampton	Hadley	Northampton
1							
2		X					

The ACS contains estimates with not just one, but three, datapoints.

is 6310 +/- 422. The indicator displays "Yes," so there is a difference. (To test the interface, I changed the MOE for Amherst to 1,437 and, in that case, the indicator changed to "No," no difference.) Another variable I used with much smaller estimates is Born outside the United States with a Graduate or Professional Degree. In Amherst, the estimate is 69+/-58, and in Northampton, 137+/-61. The indicator read "No" in this case.

WORKING WITH MULTIPLE ESTIMATES

This tab is similar to working with two estimates except on a larger scale, as a matrix of multiple estimates (up to 150) are compared to every other one These comparisons could be done in the Two_ Est. sheet, but it would be extremely unwieldy and time-consuming for more than a few variables.

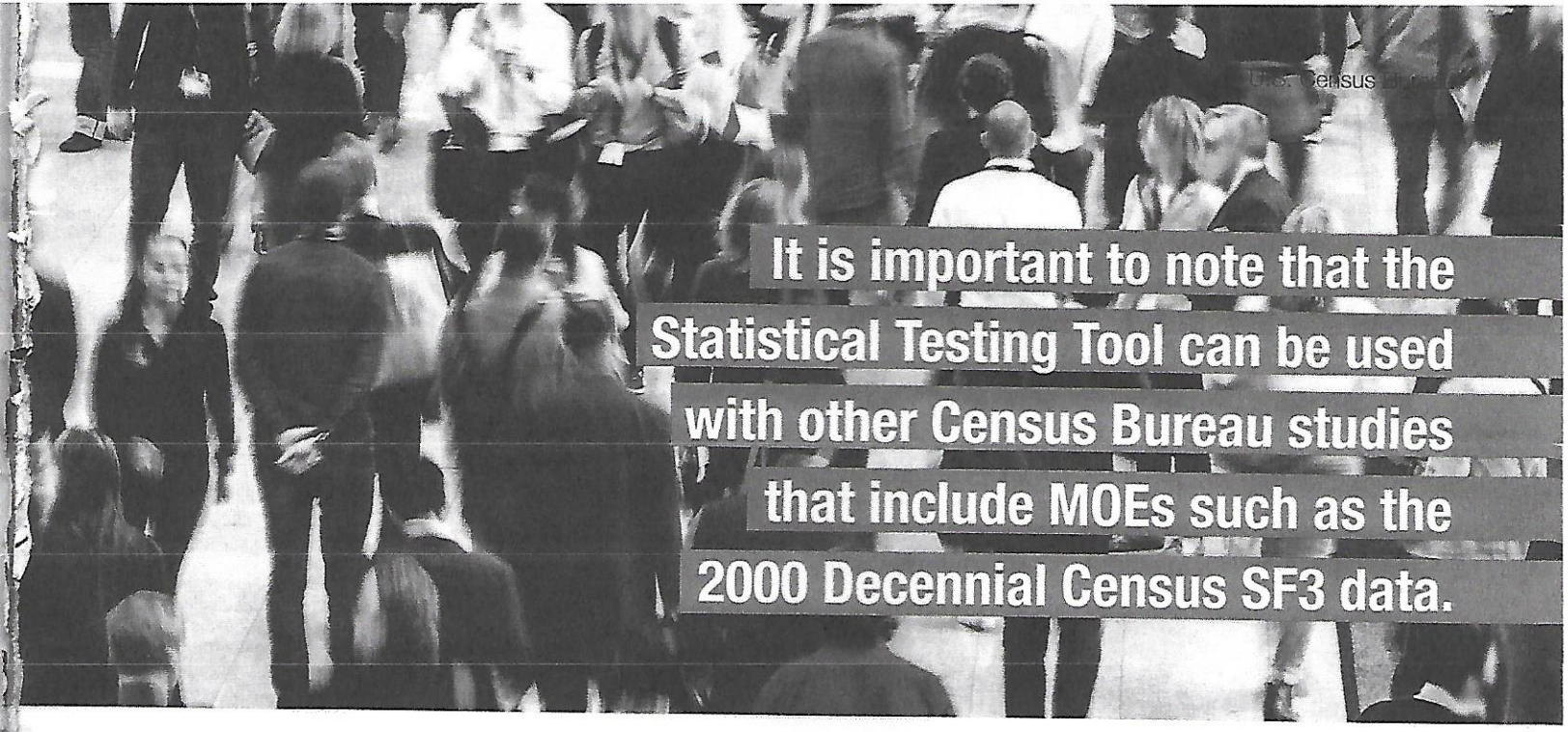
I compared the same table as above for those with Graduate or Professional degrees across not two, but five, towns in Hampshire County, Mass. (Amherst and Northampton as well as Belchertown, Easthampton, and Hadley).

The first four steps are identical to working with two estimates.

However, for step 2 in the Two_Est tab above, there is a second "Label" column that copies the first (columns B and K). To stay organized, it is best to put the data variable in the first cell (column B, line 18) followed by the various geographic names in subsequent cells (column B, lines 19, etc.)

There is also an extra recommended sorting step (n this case I did not do it because I wanted to preserve the alphabetic order of the town names):

> "Sort the data by the 'Estimate' column in ascending or descending order. This will make the results easier to read. (Recommended to sort the data before pasting them into the spreadsheet.)"



In most cases, the comparison used will be geographic (not time frames) since any geographies with having less than a 65,000 population will have 5-year average estimates that should only be compared against distinct 5-year periods with no overlapping years. To the right will be a matrix comparing every geography or time frame to every other one with Yes, No, X (estimate is itself), or – (statistical testing is not appropriate).

For this example, the labels for data variable and towns as well as estimates and MOEs for each of the latter are entered in columns B, C, and D in lines 19–24. It turns out that all of the towns have statistically significant differences among their graduate and professionally educated residents.

Note: The data variable and town labels are all displayed vertically in line 17 starting with column M.

WHAT THE STATISTICAL TESTING TOOL DOES NOT CALCULATE

While this tool can save the researcher lots of time in doing comparisons, it does not cover all calculations that need to be done for proper usage of ACS data. These include the following:

1. Adding (aggregating) estimates—this includes MOEs for each estimate to create a new MOE for the final estimates sum (census.gov/content/dam/Census/library/publications/2018/acs/acs_general_handbook_2018_ch08.pdf).
2. Comparing estimates in overlapping time spans of 3 or 5 years—not recommended but it can be done (census.gov/content/dam/Census/library/publications/2018/acs/acs_general_handbook_2018_ch07.pdf; see page 6).
3. Variance Replicate Estimate Tables for 5-year data—This is a more exact way to add estimates and MOEs, because these tables account for co-variance (census.gov/programs-surveys/acs/data/variance-tables.html).
4. Adjusting for inflation with dollar-denominated variables over time (census.gov/content/dam/Census/library/publications/2018/acs/acs_general_handbook_2018_ch10.pdf)

MAKING THE TOOL MORE OBVIOUS

It is important to note that the Statistical Testing Tool can be used with other Census Bureau studies that include MOEs such as the 2000 Decennial Census SF3 data. While the Statistical Testing Tool does not cover every application of an estimate's MOE, it is a huge time-saver when doing comparisons. Before this tool came about to help automate the process, it was an arduous, multi-step procedure for figuring just one comparison. In addition, extra time was spent finding and reading guides or presentations that explained how to do these comparisons.

Moreover, the Tool should, in theory, increase the number of users who will use an estimate's MOE. However, the problem with the tool is that many users still may not know it exists. An obvious link to it is nowhere to be found on the ACS homepage (census.gov/programs-surveys/acs) or the ACS Data page (census.gov/programs-surveys/acs/data.html). I believe it should be listed prominently in the same place on every page related to ACS data to encourage its use as much as possible. Microsoft Excel is a software package most data users have on their computer and know how to use at least at a basic level. Without better marketing, these data unfortunately still run a significant risk of being not very precise or helpful if only estimates are reported and calculated.

Of course, my ultimate hope will be that the Statistical Testing Tool will become not only ubiquitously used but also the first of many time-saving automations that the Census Bureau develops to increase the proper use of American Community Survey data that are unrivaled in their distinctiveness, currency, and value.

Roger Magnus *(magnurog@yahoo.com) is the owner of Roger Magnus Research based in Amherst, Mass.*

Comments? Email the editor-in-chief (marydee@xmission.com).