Lab 4: More Data Notes and Using Social Explorer

1. Learning Objectives

- Part 1: Aggregating Margins of Error and including them on charts
- Part 2: Become familiar with using Social Explorer for downloading data

2. Aggregating Margins of Error

It often makes sense to congregate census categories. For example, in the Excel sheet “Ages” in the workbook “Lab4Data.xslx” you will find the education information for Census Tract 4004. The information separates the population into 6 age categories. This may be more than we need for the purposes of our analyses (or, simply, easy comprehension), so we are going to consolidate the information into three categories, “Child”, “Working Age Adult”, and “Senior Adult”.

Subject	Census Tract 4004, Alameda County, Total	
	Estimate	Margin of Error
Total civilian noninstitutionalized population	3,986	+/-250
AGE		
Under 5 years	257	+/-101
5 to 17 years	421	+/-125
18 to 34 years	1,241	+/-204
35 to 64 years	1,569	+/-140
65 to 74 years	329	+/-77
75 years and over	169	+/-59

Source: 2013-2017 ACS 5-Year Estimates, Table S1810.

1. Aggregating Categories - Let’s create a column for each of our new categories, and a column for our new categories’ MOEs. Then let’s congregate our categories by summing estimates across categories. “Child” will include “Under 5 years” and “5 to 17 years”. “Working Age Adult” will include “18 to 34 years” and “35 to 64 years”. “Senior Adult” will include “65 to 74 years” and “75 years and over”.

Estimate; Total civilian noninstitutionalized population	MOE; Total civilian noninstitutionalized population	Estimate; Child	MOE; Child	Estimate; Under 5 years	MOE; Under 5 years	Estimate; 5 to 17 years
3986	250	=SUM(E2,G2)		257	101	421

Summing estimates to make up the new “Child” category.

2. Aggregating Margins of Error – To calculate the MOE for aggregated count data:
    a. Obtain the MOE of each individual estimate;
    b. Square the MOE of each estimate;
    c. Sum the squared MOEs; and
    d. Take the square root of the squared MOEs.

    In excel this can be done using the square root function, “=sqrt()”, and square value, “Value^2”. The complete function will read, =SQRT(SUM(Value1^2,..., ValueN^2))

Estimate; Total civilian noninstitutionalized population	MOE; Total civilian noninstitutionalized population	Estimate; Child	MOE; Child	Estimate; Under 5 years	MOE; Under 5 years	Estimate; 5 to 17 years	MOE; 5 to 17 years
3986	250	678	=SQRT(SUM(F2^2,H2^2))		101	421	125

When you are done you should have the following information, if not necessarily this format:

Census Tract 4004 Population Demographics: Age		
Age	**Population Estimate**	**MOE Census**
Child	678	161
Working Age Adult	2810	247
Senior Adult	498	97
Source: 2017 ACS 5-Year Estimates, Table S1810.		
Note: Age summarized from Census table into three categories: "Child" (aged 17 and under), "Working Age Adult" (aged 18-64), and "Senior Adult" (aged 75 years and over).		

3. Adding Error Bars to Tables

For time considerations, Abby has put our age data for Census Tract 4004 into a clean table, with count estimates and the congregated margins of errors in a hidden sheet called “Error_Bars”. Unhide the Sheet. She has also created a column chart with the data. What conclusions can you draw about the ages of Tract 4004 residents (i.e. are there more working age adults, children, seniors residing in the Tract)? Are you sure?

Let’s add error bars for the margins of error to confirm our initial impressions.

1. *PC Instructions:* Double click on the chart. This brings up the chart commands. Click the button at the top left of the screen labeled “Add Chart Element” and select “Error Bar” and then “More Error Bar Options” from the drop down menu. *Mac Instructions:* Double click on a column in the chart. This brings up the “Format Data Series” window. Select Error bars from the left-column menu.

2. We now need to format the error bars. Choose to Display “Both” error bars, and “Cap” the ends.
3. Now we need to specify the error bars amount. Select “Custom” in the “Error amount” menu, and then “Specify Value”. You will then be prompted to enter “Custom Error Bars” Positive and Negative values. You want to select the cells with the margins of error information for BOTH the Positive and Negative values (Why is this the case knowing what you know about Census MOEs?).

Have your conclusions about the resident age composition of Tract 4004 changed after seeing the error bars (and considering where there may be error in the estimated population counts)?

4. Social Explorer

Social Explorer is a great tool for accessing data. Social explorer is free for Berkeley Students and is available on and off campus (off campus you simply sign in using your Berkeley ID). We are going to start by opening social explorer. We suggest routing there through the Berkeley Library Census data page, just to make sure you have access to the Berkeley License. You will need to create a login to use the system and to save your work (click login to get started). If you already have an account click sign in.

The strength of Social Explorer is that it is fairly straightforward and that it is prepopulated with Census data (and a whole lot of other fun data!). In this lab we will focus on the tools for downloading data.

5. Social Explorer: Tables!

Start by clicking on the “Tables” button. At this point you might be thinking “Hey! This looks familiar.” You are right they should look familiar. The first several links are to Census data sets you’ve used before. Let’s see how it works by downloading one of the tables we pulled from FactFinder a few weeks back - the 5-year ACS Survey estimates, Table B03002: Hispanic or Latino by Race.

Steps:

1. Click on the survey source => “American Community Survey (5-Year Estimates)”
2. Choose a year => Select the “Begin Report” under the 2013-2017 survey. (“More Info” links to a detailed explanation of the survey.)
3. Choose a geography => Select census tract, then state, then county, then your tract of interest. Click “Add”. Similar to FactFinder you can add multiple geographies if you want your data summarized for different geographic levels. Now “Proceed to Tables”.
4. Select a Table => You can search for data either by Title or by Keyword. In addition, data is available through both Social Explorer tables and ACS tables. Note that Standard

Errors are only available with the ACS tables (Standard Errors can be used to calculate Margins of Error.)

   a. Try searching for the ACS tables B03002 by title using the “List Table”.
   b. Now try to search for the same table using the “Keyword” search and typing in the table name.

   When you find the table click on “Add” to make a selection and then click “Show results”
5. Download the Data => Click the “Excel” button and chose to download the data with counts and/or percentages.
6. Open the file. What do you notice?

6. Downloading Business Data

Let’s go back to the list of all the data available through Social Explorer. Scroll down a bit and look for the “U.S. Business Patterns” data. These data sets provide economic data about the local economy at the County and ZIP Code level. To start, let’s download some County data.

1. Click on “Begin Report” under “County Business Patterns 2016”.
2. Choose the County designation under geographic type.
3. Select California under the “Select a State”
4. Select “Alameda County” under the “geographic area designation” and “Add” it as a selection.
5. Click “Proceed to Tables” and we will download two tables: **Number of Employees** and **Number of Establishments**. For this example, the simplified tables are really helpful. Go to “Select a Dataset” and choose “Tables Simplified”. Add the two tables mentioned (T1. and T2., respectively).
6. Proceed to “Show Results.”

7. Working with ZIP Codes

As discussed in class, downloading ZIP code data is a little trickier than downloading county data. The Census Bureau helps you to map between ZIP codes and Census Tracts by providing relationship files (https://www.census.gov/geo/maps-data/data/zcta_rel_download.html). Warning, it can be a little scary – so we created a cleaner version of the file and uploaded it to bCourses.

1. Open up the file.
2. Search for your Census Tract. The second column is the ZIP code (or codes) where your tract is located. Column D will tell you what percentage of the Census tract population is in your Census Tract, and column E will tell you what percent of the Census tract is in the ZIP Code.
3. Is your tract in a single ZIP Code? Yes? No?

You can also investigate whether your Tract crosses multiple ZIP codes using usboundary.com. This is, frankly, more accessible than using the Census Relationship Tables.

1. Go to the “Areas” tab near the top of the page and select Census Tract.
2. Enter State and County information, and click on your Tract.
3. Scroll down the page a bit and you’ll see your Tracts “Neighbors,” including “Neighboring 5-Digit ZIP Code Tabulation Area (by Population)” at the bottom. Click on “Neighboring 5-Digit ZIP Code Tabulation Area on the Map.”
    a. You should be able to see which ZIP code(s) neighbor your Tract, as well as nearby ZIP codes. You are likely most interest in those that overlap your Tract.
4. Again, is the Tract you searched in a single ZIP code? Does this result agree with what you observed using the Census Relationship file?

8. On Your Own: Downloading Multiple Tables

One of the big benefits of Social Explorer is that you can download multiple datasets at the same time. This can reduce the total time you need downloading data. As practice, let’s try to download two different data sets on housing stock at the same time. Find one dataset describing Tenure and one describing Median Rent in your Census tract. Add each data set as a selection so that is shows up in the “Current Table Selection” box, and then download the data.