

Using DataFerrett to Create Custom Tables from ACS PUMS

1. Informational Query:
 - a. How many college graduates aged 25 to 34 years are living at home with their parents?
 - b. Answer is not a standard tabulation in the ACS summary tables on American FactFinder.
 - c. Need to create a custom table from the ACS Public Use Microdata Sample (PUMS).
2. Custom Table
 - a. Universe: Persons with college education aged 25 – 34 years who are living at home with their parents and are the son or daughter of the householder.
 - b. Stratifiers: Sex, Employment Status, Marital Status
3. Go to <http://dataferrett.census.gov>
 - a. Launch DataFerrett
4. Ferrett Login screen
 - a. enter email address
5. DataFerrett application window
 - a. click on “Get Data Now”

Select Dataset and Variables

6. Select Dataset
 - a. Open the “American Community Survey” folder
 - b. Open the “5-Year Estimates – Public” folder
 - c. Select 2008-2012
7. Select Variables
 - a. Select “View Variables”
 - b. Filtering records to restrict a universe
 - i. Age – 25 to 34
 1. Use Search to find the variable “Age” among the Population Records.
 2. Double click on AGEP so see the categories available for the variable
 3. Check “Select ACS AGEP Age” and then click on “Deselect all values”
 4. Check ages 25 through 34 and then click on “OK”, then getting confirmation that you’ve added 1 variable to your DataBasket, and clicking on “OK”
 - ii. Educational Attainment – Bachelor’s degree or higher
 1. Repeat steps above for the variable “Educational Attainment”
 - iii. Relationship to Householder – Son or daughter – biological, adopted, or step
 1. Repeat steps above for the variable “relationship”
 2. Note: This is the step in which persons in group quarters are eliminated from the universe.
 - c. Stratifiers
 - i. Sex
 1. Repeat steps above
 2. Except after checking “Select ACS SEX Sex” and both categories—Male and Female—are checked, DO NOT Deselect but click “OK”
 - ii. Employment Status
 1. Repeat steps above

2. Except click on the radio button for “match ALL words”
- iii. Marital Status
 1. Repeat steps above

Make a Table

8. Select the tab at the top for “Step 2: DataBasket/Download/Make a Table”
 - a. Verify that you have the variables you will need for your table
9. Display data on the screen of the Person records meeting your selection of file, variables, and categories from the PUMS
 - a. NOTE: This is not a necessary step and is included in this demo to illustrate what the PUMS files contain.
10. Click on the “Make A Table” box containing the icon of a Ferrett looking over a blank table
 - a. Produces an informational window with a Ferrett Tab Message describing the steps for creating a table. Click OK
 - b. Ferrett Tabulation window opens with rows and columns and the variables on the right are from your DataBasket
 - c. Table 1 – simple tabulation of a single variable
 - i. Drag RELP to the cell R1C1
 - ii. The “Total” category is listed first followed by biological, adopted, and step sons and daughters
 - iii. Click on the box containing “GO Get Data” and watch the ferret scurry to retrieve your data and perform the tabulation [animation in lower right corner of the screen]
 - iv. Default is for the US. This is an estimate from the PUMS of the number of persons fitting our query.
 - d. Table 2 – cross tabulation of Relationship with Sex, Employment status, and Marital Status
 - i. Drag SEX to R1C1
 - ii. Leave a row and drag ESR to C1
 - iii. Repeat with MAR
 - iv. Drag RELP R1C2
 - v. Click on “GO Get Data”
11. Options includes expressing as percent
 - a. Express the frequencies in Table 2 as percentages
 - b. In the Menu bar at the top, click on “Options” and then “Percent” to see the options
 - c. Select “Show percent of first data cell”
 - d. Experiment with the other options
12. Save or Print from the File menu

Part 2: Selecting Geographies

13. Most detailed geographic areas in the PUMS are PUMAs which have approximately 100,000 persons
14. Determine which PUMAs are to be included

- a. Go to PAD website [<http://pad.human.cornell.edu>]
 - b. Maps > Reference maps > PUMAs
 - c. PUMAs in the Capital District Economic Development Region – 2010 PUMA ID (2000 PUMA ID)
 - i. Albany County 02001 (02401); 02002 (02402)
 - ii. Columbia County 02100 (02500)
 - iii. Greene County 02100 (02500)
 - iv. Rensselaer County 01900 (02300)
 - v. Saratoga County 01801 (02202); 1802 (02201)
 - vi. Schenectady County 01700 (02100)
 - vii. Warren County 00300 (00300)
 - viii. Washington County 00300 (00300)
 - d. Total Pop
 - i. PUMS with above PUMAs 1,079,301
 - ii. AFF Summary File 1,079,084
15. Select the tab at the top for “Step 1: Select Dataset & Variable”
16. Select same dataset and click on “View Variables”
- a. Check the box next to “Selectable Geographies”
 - b. and “Geographic Entities”
 - c. Click “Search Variables”
 - d. Double click on the variable named “Geography”
 - i. Click on “State”
 - ii. Then in the panel labeled “Hierarchies” click on “State Code” and then click on “Use Hierarchy.”
 - iii. Drag states of interest from the left most panel to the right most panel, and when done, click “Finish”
 - e. Return to the listing of the three geographic variables and double click on the variable named “PUMA10” which are the codes for PUMAs used in the 2010 Census and for vintage 2012 ACS records. All of the PUMA ID codes are listed.
 - i. Check “Select ACS PUMS10 Public use microdata area codes” and then click on “Deselect all values”
 - 1. Check the first category, “-9) Code classification is Not Applicable because data are 2012 vintage.” This is necessary because the older records for prior years are coded -9 on this variable.
 - 2. Check the all the codes for PUMAs of interest to create a multi-PUMA area. Use the Control key to select more than one PUMA.
 - 3. Click OK and add to DataBasket.
 - f. Return to the listing of the three geographic variables and double click on the variable named “PUMA00” which are the codes for PUMAs used in the 2000 Census and for ACS records prior to vintage 2012. All of the PUMA ID codes are listed.
 - i. Check “Select ACS PUMS00 Public use microdata area codes” and then click on “Deselect all values”

1. Check the first category, “-9) Code classification is Not Applicable because data are 2012 vintage.” This is necessary because the older records for prior years are coded -9 on this variable.
2. Check the all the codes for PUMAs of interest to create a multi-PUMA area. Use the Control key to select more than one PUMA.
3. Click OK and add to DataBasket.

Make a Table

17. Select the tab at the top for “Step 2: DataBasket/Download/Make a Table”
 - a. Verify that you have the variables you will need for your table
18. Repeat steps 10 and 11 from above. Instead of producing a table for the whole US, this table will for the area that was based on selected PUMAs.
19. Count of respondents in each cell.
 - a. Note that some of the cells in the regional table are very sparse or even 0.
 - b. Go to “Options” and then Weighting. Select “Unweighted to see actual number of respondents that the estimate is based on. Avoid using any cell with fewer than 30 respondents.

Presenter:

Warren Brown, Cornell Program on Applied Demographics, wab4@cornell.edu