Overview of PUMS Econ 29000 Kevin R Foster, CCNY Fall 2011

We will use data from the Census Bureau's "Public Use Microdata Survey," or PUMS. This is collected in the American Community Survey; every ten years since 1790 the Census has made a complete enumeration of the US population as required by the Constitution.

We will work on this data using SPSS. For an overview of the basics of how to use that program, find the separate class document online.

The dataset is ready to use in SPSS. Download it from the class's Blackboard page onto your computer desktop. If it is zipped, then unzip it. Remember that if you're in the computer lab, just double-clicking on the SPSS file may not automatically start up SPSS; you'll get some error code. So use the Start bar to find SPSS and start it that way. Then open up your dataset once the program has loaded.

SPSS has two views of the dataset: Variable View and Data View. Usually we use the Variable View; this lists all of the different information available.

The dataset has information on 315,771 people in 133,043 households. If there is a family living together in an apartment, say a mother and two kids, then each person has a row of data telling about him/her (age, gender, education, etc) but only the head of household (in this case, the mother) would have information about the household (how much is spent on rent, utilities, etc.). Depending on what analysis is to be made, the researcher might want to look at all the people or all of the households (or subsets of either). If you look at the "Data View" tab you can see the difference. (Note that the "head of household" is defined by the person interviewed so it could be the man or woman, if there are both.)

The first column of data is a serial number, shared by each person in the household. After that you can see that some variables are filled in for every person (age, female, education levels) but other variables are only filled in for one person in the household (has\_kids, kids\_under6, kids\_under17).

## Basics of government race/ethnicity classification

The US government asks questions about people's race and ethnicity. These categories are social constructs, which is a fancy way of pointing out that they are not based on hard science but on people's own views of themselves (influenced by how people think that other people think of them...). Currently the standard classification asks people separately about their "race" and "ethnicity" where people can pick labels from each category in any combination.

The "race" categories that are listed are: "White alone," "Black or African-American alone," "American Indian alone," "Alaska Native alone," "American Indian and Alaska Native tribes specified; or American Indian or Alaska native, not specified and no other race," "Asian alone," "Native Hawaiian and other Pacific Islander alone," "Some other race alone," or "Two or more major race groups." (Then the supplemental race categories offer more detail.)

These are a peculiar combination of very general (well over 40% of the world's population is "Asian") and very specific ("Alaska Native alone") representing a peculiar history of popular attitudes in the US. Only in the 2000 Census did they start to classify people in mixed races. If you were to go back to historical US Censuses from more than a century ago, you would find that the category "race" included separate entries for Irish and French and various other nationalities. Stephen J Gould has a great book, *The Mismeasure of Man*, discussing how early scientific classifications of humans tried to "prove" which nationalities/races/groups were the smartest.

Note that "Hispanic" is not "race" but rather ethnicity (includes various other labels such as Spanish, Latino, etc.). So a respondent could choose "Hispanic" and any race category – some choose "White," some choose "Black," some might be combined with any other of those complicated racial categories.

What that means, specifically for us reporting statistics on a dataset like this, is that we can easily find that, of the 315,771 people in the PUMS dataset who live in the five boroughs of New York City, 48.2% report their race as "White alone" and 24.2% as "Black alone," 12.3% report as "Asian alone," 12.7% report "some other race alone," 2.2% report multiple races, and less than 1% report any Native American category. Then 24.1% classify their ethnicity as Hispanic. Can we just take the 48.2% White, subtract the 24.1% Hispanic to say that 24.1% are "non-Hispanic White" (a category commonly used in other government classifications)? NO! Because that assumes that all of the people who self-classified as Hispanic were also selfclassified as "White only" which is not true. We would have to create a new variable for non-Hispanic White to find that proportion.

How can we do that with SPSS? On the drop-down menu find "Transform" then "Compute Variable" then in the dialog box, give the new variable (it calls it the "Target Variable") a name (e.g. "nonHispWhite") and a Numeric Expression, for example here " (RAC1P = 1) & (HISP = 1) ". The first expression evaluates, for each case, whether the variable, "RAC1P" which is the variable coding race, has a value of 1 (which corresponds to the label "White alone"). If it equals 1 then the expression is True, which is coded as 1; if RAC1P does not equal 1 then the expression is False, coded as zero. The second expression evaluates if "HISP" equals 1 or not. The "&" sign in the middle evaluates if both expressions are true or not. When we run this classification, we find that 38.9% are non-Hispanic white, a big difference!

All of these racial categories might make some people uneasy: is the government somehow encouraging racism by recognizing these classifications? Some other governments might choose not to collect race data. But that doesn't mean that there are no differences, only that

the government doesn't choose to measure any of these differences. In the US, government agencies such as the Census and BLS don't collect data on religion, which means that we can't answer certain questions.