SYD 3410 Urban Sociology Fall, 2001

> Data Analysis Module One The Social Structure of the Cities

In class, we discussed the US Census definitions of cities, metropolitan areas, and urban areas. We focused in particular on the census definitions of metropolitan, suburban, and non-metropolitan areas. Our objectives in this first data exercise are:

- To discover how the present-day US population is distributed across these various census categories;
- To discover how the distribution has changed over time; and
- To see how some of the social characteristics of people who live in cities, suburbs and non-metro areas differ.

1. How is the US population distributed across the Census categories of metro, suburban, and non-metro places?

In order to answer this first question, you need to open WebChip and open a data set called **POPGEO9.DAT** 

## Instructions:

1. Go to the web site, <u>http://www.ssdan.net/ida\_resources.shtml</u> Scroll down the page until you find <u>WebCHIP Launcher</u>. Double click on the Launcher and you will get the following screen:

## WebCHIP Launcher

To use, complete the dataset filename. For example, entering "datasets/pc/cen1990/asnusa9.dat" will open the asnusa9.dat dataset in WebCHIP. Enter all dataset information in LOWERCASE and make sure to include the ".dat" or state abbreviation (e.g., ".mi") as appropriate. Direct any problems with this system to <u>ssdan-</u> <u>tech@umich.edu</u>.

datasets/pc/

2. In the box where it says **datasets/pc**/ enter "cen1990/popgeo9.dat" (without quote marks, all lower case), then hit ENTER or Submit Query. If all has gone well, you will get a screen with a gray bar at the top and a screen that looks like this:

[Cut and Paste from WebCHIP Printout]

The gray bar at the top is what you use to set up your cross-tabulation. The text at the bottom tells you about the data set you have opened. This one, **POPGEO9.DAT**, contains 1990 data on the whole US population, which numbered 243,878,788 in 1990. The data set contains four variables: RaceLat (racial and ethnic status in six categories), region of the country (in four categories), "geo," which is the city size variable in three categories, and age (in eight categories).

3. In the gray bar, click on Marginals and you will get the marginal frequencies for all the variables contained in the file, i.e., the percentage of the US population that falls into each of those racial, regional, city-size, and age categories.

4. **STOP** and answer the following questions:

In 1990, what percentage of the US population was Latino? \_\_\_\_\_\_ What percent was Asian? \_\_\_\_\_\_ What percentage lived in the South? \_\_\_\_\_\_ What percentage was between the ages of 15 and 24? \_\_\_\_\_\_

5. Now look closely at the marginals for the variable GEO. What is the percentage distribution of persons across these three Census city-type categories? Are these percentages about what you expected to find? How do they differ from your expectations? In the space below, write a short paragraph about what you learn from the marginals on GEO:

## 2. How has the distribution of persons across city types changed over time?

1. You will notice that **POPGEO9.DAT** only has data for 1990. To answer Question #2, we need a data set that has GEO for an earlier decade. (Alas, the 2000 data for this variable have not yet been released, so the best we can do is compare 1980 to 1990 and make some predictions about what 2000 will look like.) The data set you need can be accessed in WebCHIP with the following path: **datasets/pc/custom/hpovgeo8090.dat** So as a next step, pull that data file into WebChip and answer the following questions:

How many cases are contained in **HPOVGEO8090.DAT?** 

This file has data for both 1980 and 1990; **POPGEO9.DAT** only has cases for 1990. Why does HPOVGEO8090 have **fewer** cases than POPGEO9.DAT? (Hint: Look carefully at the description of each data set as it first comes onto the screen.)

How many variables are contained in the new data set?

2. Now, get the marginal frequencies for all the variables, and answer the following questions:

What percentage of cases in this file are from 1980? 1990? Why isn't this an exact 50:50 split?

Combining 1980 and 1990, what percentage of persons in this file were below, at, or near the federally defined poverty line?

3. Next, prepare a crosstab that shows how GEO has changed from 1980 to 1990. Put GEO in the rows, YEAR in the column, click Percentage Down, then click on Crosstab and answer the following questions:

Did the percentage of the population living in cities increase or decrease between 1980 and 1990? How about the suburban population? The non-metro population? What, if anything, do these trends say about urbanization as a "master trend" of the late 20<sup>th</sup> century?

3. What are some of the social differences between people who live in the cities, those who live in the suburbs, and those who live in the non-metropolitan areas?

One of the readings we will discuss later in the course is a paper by Farley and others entitled "Chocolate Cities, Vanilla Suburbs." The title suggests that different racial and ethnic groups are distributed differently among city-type categories. How pronounced are these differences?

1. To begin, get **POPGEO9.DAT** back on your screen . (Go back to the Launch page and hit Reset.) Then prepare a crosstab showing the racial and ethnic composition of the cities, the suburbs, and the non-metro areas. Put RaceLat in as the Row Variable, Geo as the column variable, percentage down, hit Crosstab, then answer the following questions:

Do cities, suburbs and non-metro areas have different racial and ethnic compositions? How pronounced are these differences?

Which of the three categories has the highest percentage of whites?

The highest percentage of blacks?

Hispanics?

Where is the Native American population concentrated?

Based on the analysis you have just done, is there anything wrong with the depiction, "Chocolate Cities, Vanilla Suburbs?" What?

Focusing just on racial composition, are the suburbs "more like" the cities or "more like" the non-metro areas?

2. Cities, of course, differ from suburbs and non-metro areas in ways other than racial and ethnic composition. Another variable in **HPOVGEO8090** is one called FamType, or "family type." The categories are as follows:

MrrdCpl = a married couple, with or without children MaleFam = an adult male with children and no wife FemlFam = an adult female with children and no husband MaleNonF = a single male with no children and no wife FemaleNonF = a single female with no children and no husband

Get HPOVGEO8090 back on the screen. Prepare a crosstab that shows how the metro, suburban, and non-metro populations are distributed across family types. Put FamType in as the row variable, Geo as the column variable, percent down, hit Crosstab, then answer the following questions:

1. Which of the three city-size types has the lowest percentage of married couples?

2. Which has the highest percentage of female-headed households?

3. Do the cities, suburbs and non-metro areas differ in family type as well as racial and ethnic composition. And in respect to family type, are the suburbs "more like" the cities or "more like" the non-metro areas?

## Extra Credit

Yet another variable in the file is Pov, or "poverty status." The categories are:

Poverty = the household income is at or beneath the federal poverty line. NearPoor = the household income is between 100% and 150% of the poverty line. Other = the household income is more than 150% of the poverty line.

How do the cities, suburbs, and non-metro areas differ according to the poverty status of their respective populations? Are there any surprises in the table? Can you offer any explanation for the "surprise?"

This exercise is due on September 3.