

# SOC 210 – Social Problems

## Data Analysis Assignment

### Poverty in the United States – DUE XXX



For this assignment we will examine poverty using individual income in 2000. The purpose of this assignment is to give you additional practice with a data analysis software program (WebCHIP), to develop some familiarity working with Census data, to experience analyzing data, and to link course material to explain poverty.

#### ***I. OBJECTIVES***

Social science research is often categorized in terms of two objectives—to **describe** and to **analyze** the social world. The first objective presents data to build a portrait of the way things are (or were) while the latter brings conceptual frameworks to bear on the data to explain **why** things are the way they are. For this assignment, our objective is to both describe and analyze poverty in the United States. We will answer the following questions:

- **What is the distribution of poverty, how has it changed over time, and how is it defined?**
- **What social variables are related to poverty?**
- **What are some possible explanations for poverty?**

Other learning outcomes for this assignment include honing analytical skills associated with generating, reading, and interpreting rudimentary forms of data analysis. You will bring in a third variable to bring more depth to your attempt to understand poverty. Our goal is to interpret the information for a lay audience (e.g., write a summary that any newspaper reader could understand).

#### ***II. CENSUS DATA*** (This description draws from W. Frey's *Investigating Change in American Society*, Second Edition.)

As you're probably aware, one of the government's biggest efforts is the decennial Census (mandated by our Constitution). The Census's original purpose was for apportioning our congressional representatives every ten years. The first one was done in 1790; the most recent, of course, was the 2000 Census. Although originally meant to count every single person, it collects a lot of other information. Recently, for instance, the Census Bureau sent out two different surveys, a "short" and a "long" form. The "short" form asked just for basic socio-demographic information (e.g., age, gender, race/ethnicity etc.), while the "long" form added many more items on a variety of social characteristics (e.g., occupation, education, language proficiency). These characteristics represent variables (things that vary across individuals in the United States). The information on these characteristics is of immense importance for both planners at all governmental levels (federal, state, and local) and social science researchers. For this class, we can examine variables related to poverty.

Before we get to the actual assignment, one thing to remember is that no research effort is perfect, including that done by the Census Bureau. They are never able to count everyone (e.g., homeless), and some people may be more likely to answer the different forms than others. We refer to this potential problem as **response bias**. The sample sizes are so large, however, that it still represents an accurate picture of American people.

Although there are several ways in which the following terms may be conceptualized, defined and measured, these are the definitions used by the U.S. Census Bureau and the Bureau of Labor Statistics:

**Poverty** – Poverty relative to “poverty cutoff”

- **Poor** – Family income below the poverty cutoff.
- **Near-Poor** – Family income between 1 and 1-1/2 times the poverty cutoff.
- **Middle** – Family income between 1-1/2 and 5 times the poverty cutoff.
- **Comfortable** – Family income more than 5 times the poverty cutoff.

**Race (RaceLat)** – individual’s self-identification as:

- **Non-Hispanic White (NHwhite)** – all persons who indicated their race as white and not of Hispanic origin.
- **Black** – all persons who indicated their race as black.
- **Hispanic** – persons of white or “other” races who identified themselves as Mexican, Puerto Rican, Cuban, or Other Spanish/Hispanic.
- **Asian (or Pacific Islander)** – includes all persons who indicated their race or ethnicity as Chinese, Filipino, Japanese, Asian Indian, Korean, Vietnamese, Cambodian, Hmong, Laotian, Thai, or other Asian as well as Hawaiian, Samoan, Guamanian or other Pacific Islander.
- **American Indian (AmIndian)** – all persons who classified themselves as American Indian, Eskimo or Aleut.
- **Other Race (NL Other)** – all persons who classified themselves as “Other” and who are not of Latino origin.
- **Lmulti** – all persons who listed multiple racial or ethnic categories (includes those of Latino origin).

**Gender (Gender)** - individual’s self-identification as either male or female.

**Age (Age)** – age in years, grouped into 10-year intervals, starting at age 25

**Education (Educ)** – age in years, grouped into 10-year intervals, starting at age 25

- **0-9 Years** – persons who completed less than 10 years of schooling.
- **9-12 Years** – persons who completed 9-12 years of schooling but did not graduate from high school.
- **HSGrad** – persons who graduated from high school, includes GED.
- **SomeColl** – persons who completed some years in college or received an Associate degree, but did not receive a Bachelor’s degree.
- **CollGrad** – persons who graduated from college with a Bachelor’s degree.
- **Masters** – persons who completed a Master’s degree, i.e., MA, MS, MSW, MBA.
- **PhD-Prof** – persons who completed a terminal degree in their field (i.e., PhD or EdD) or a professional school degree (i.e., MD, JD, DDS)

### **III. Using WebCHIP**

You will access data from the Census 2000. Instructions are as follows:

1. Using Internet Explorer (Netscape doesn’t work as well), go to the Social Science Data Analysis Network at [http://www.ssdan.net/ida\\_resources.shtml](http://www.ssdan.net/ida_resources.shtml) and click on the “WebCHIP” launcher link. This is where you pick your data set. At the next screen, use the “WebCHIP Launcher.” When you get to this screen, using the scroll-down feature highlight “cen2000” and click on “Select Files.” Scroll down through the list of data sets until you find “ppovedu2k.dat” and click again on “Select Files.”

2. Now, simply select the "Marginals" button for univariate analyses and you will get descriptive data for each variable in the file (univariate = descriptive distribution of one variable).
3. For your Crosstab analysis, put your dependent variable in the row variable and your independent variable in the column. For example, if you are hypothesizing that race predicts poverty, then race is your independent variable and poverty your dependent variable. Make sure your tables are constructed as "% Down."
4. Output can be copied from WebCHIP into a word processor by highlighting it (left-click and drag your mouse) and hitting the "Ctrl" and "C" keys simultaneously. Once in your word processor, right click your mouse and select "Paste."

I will demonstrate the project in class using poverty and gender. Note that we are only using a sample of the Year 2000 Census, but it is a VERY large sample. We will discuss these findings in class and you will go on to address other questions using this data set and then you will use another data set to test hypotheses you develop about household poverty.

First, we will do our univariate analysis (using marginals in the ppovedu2k.dat).

<b>RaceLat</b>								
<b>NLWhite</b>	<b>Black</b>	<b>Asian</b>	<b>Latino</b>	<b>AmIndian</b>	<b>NLOther</b>	<b>Lmulti</b>	<b>Total</b>	
72.7%	10.6%	3.7%	10.7%	0.7%	0.1%	1.3%	= 100.0%	
<b>Gender</b>								
	<b>Male</b>	<b>Female</b>	<b>Total</b>					
	47.6%	52.4%	= 100.0%					
<b>Pov</b>								
	<b>Poverty</b>	<b>NearPoor</b>	<b>Middle</b>	<b>Comf</b>	<b>Total</b>			
	9.9%	7.5%	51.5%	31.1%	= 100.0%			
<b>Age</b>								
	<b>25-34</b>	<b>35-44</b>	<b>45-54</b>	<b>55-64</b>	<b>65+</b>	<b>Total</b>		
	21.6%	25.3%	20.9%	13.4%	18.7%	= 100.0%		
<b>Educ</b>								
	<b>0-9Yrs</b>	<b>10-12Yrs</b>	<b>HSGrad</b>	<b>SomeColl</b>	<b>CollGrad</b>	<b>Masters</b>	<b>PhD-Prof</b>	<b>Total</b>
	9.9%	9.4%	28.6%	27.5%	15.7%	6.0%	3.0%	= 100.0%

Look at these data carefully, what percentage of this Census sample are:

White \_\_\_\_\_ Black \_\_\_\_\_ Asian \_\_\_\_\_ Latino \_\_\_\_\_?  
 American Indian \_\_\_\_\_ Biracial \_\_\_\_\_?  
 Male \_\_\_\_\_ Female \_\_\_\_\_?  
 Are poor or near poor \_\_\_\_\_? Are comfortable \_\_\_\_\_?  
 Have not graduated from high school \_\_\_\_\_?

Getting the idea?

Now, let's do a bivariate analysis of this data. First, we'll develop a hypothesis, that men earn more than women. That makes sense given what we've been studying. Following the directions above, we'll get a crosstab for these two variables.

Table 1: Poverty Status of Individuals by Gender, from Census 2000

	<b><u>Male</u></b>	<b><u>Female</u></b>	<b><u>Total</u></b>
Poverty	8.3% (7,166,593)	11.4% (10,792,240)	9.9%
NearPoor	6.6% (5,698,976)	8.2% (7,813,589)	7.5%
Middle	51.8% (44,545,602)	51.2% (48,594,942)	51.5%
Comf	33.2% (28,591,021)	29.1% (27,618,797)	31.1%
Total	86,002,192	94,819,568	N = 180,821,760

---

Notice how this table is arranged. Our independent variable is in the columns, and the percentages are down the columns so we can compare across differences in our dependent variable. (Rounding error sometimes accounts for small differences in the total.) Now, we can ask the question, are males or females more likely to live in poverty?

**IV. THE ASSIGNMENT**

Name \_\_\_\_\_

Now, on to your assignment! There are three parts to this assignment. First, you need to learn a little more about how poverty is defined and how it has changed over the years. To answer these questions you will go to the U.S. Census website. Then you will go to the Census data that we used in class and develop a question/hypothesis to test. Lastly, you will go to a file that uses household rather than individual data and develop another hypothesis to analyze poverty. Answer all questions associated with each task

**TASK #1:** Go to the U.S. Census website ([www.Census.gov](http://www.Census.gov)) and get the following information. NOTE: THIS COULD BE MODELED IN CLASS OR DONE PRIOR TO THE REST OF THE ASSIGNMENT.

1. What was the poverty threshold for a family of four with two children under 18 years old in 2000? \_\_\_\_\_

2. Look at poverty trends over time using data from the Census by filling in the following table.

Fill in the matrix below based on your data.

Year	1960	1970	1980	1990	2000
% of individuals in poverty					

What changes have occurred in the number of people in poverty?

What do you think might explain why these changes occurred?

**TASK #2:** Time to go back to WebCHIP and test some hypotheses. First, we'll use ppovedu2k.dat (the file we used in class). Since we've already done a frequency distribution, you don't have to do another one (just refer back to the 3<sup>rd</sup> and 4<sup>th</sup> pages of this assignment).

1. Develop a hypothesis that might explain poverty and test this hypothesis by producing a cross tabulation. Remember, you need to specify your independent and dependent variables and construct your table so you can compare across categories of your dependent variable, therefore, you must percentage to 100% across categories of your independent variable. Attach a copy of your table. List your hypothesis and describe your findings here.

2. Okay, time to bring in a control variable. This is a third variable that you think will affect the relationship between the two variables in your hypothesis. You do this in WebCHIP by clicking on the variable you want to use in the control box and then clicking on the “Control by” box. You then click on the “Crosstab” box. Attach a copy of your tables. List your revised hypothesis, with your control variable and describe your findings here.

TASK #3: Time to look at the issue of poverty from a slightly different perspective. Go to the Census data for head of households (hhpov2k.dat) and complete the following tasks. Note that the variables have changed slightly and poverty now has only two values in this data set. In addition, a new variable was added – HHType. This new variable looks at household type: married couple families (MrrdCpl), single mother families (FemlFam), single father families (MaleFam), and non-family households with male (MaleNonF) or female (FemaleNonF) heads. It is a good idea to do a univariate analysis to look at how individual variables are distributed (use the Marginals button).

1. Let’s begin by looking at the relationship between poverty and head of household status? Which is the dependent \_\_\_\_\_ and the independent \_\_\_\_\_ variable?

Attach a copy of your table and interpret your findings.

2. Now, switch variables and make your dependent variable independent and vice versa. Which is the dependent \_\_\_\_\_ and the independent \_\_\_\_\_ variable?

Attach a copy of your table and interpret your findings.

Notice that it does make a difference how you specify your variables and percentage your tables in terms of how you describe your findings.

3. Time to try a control variable. Use race/ethnicity or age as a control variable. Attach your tables and describe what you have found. Is the pattern consistent across values of your control variable? Does the relationship between poverty and household type vary based upon differences between groups in your control variable?