

# SOC 100 – Introduction to Sociology

## Income Inequality in the United States

### Data Analysis Assignment



For this assignment we will explore the impact of gender and race on the earnings of full-time workers in 2000. The purpose of this assignment is to introduce you to some basic data analysis software (StudentCHIP), to develop some familiarity with working with data from the Current Population Survey, and to apply what you have learned in the course to try to explain differences in earnings based on race and gender.

## I. Data Sets

You will need to use the EARN2KC.DAT and EARN2KC.NY plus one other data set from a state of your choice.

- A.** If you are using StudentCHIP from the computers in Dailey Hall you will need to download the datasets you are using onto your disk from Angel.
1. Log on to the SOC 100.03 site in Angel and click on the “Lessons” tab.
  2. Under the “Lessons” tab in Angel you will find a folder called “StudentCHIP Data Files” which contains data sets for all 50 states and the entire United States. The file name extension indicates which state, for example EARN2KC.NY is the 2000 earnings information for New York State, EARN2KC.DAT is the dataset for the entire country.
  3. To download the data, click on the dataset of your choice, for example we will start with EARN2KC.DAT. Angel will take you to a new page stating: “The requested file can be accessed using the link below.”
  4. Right click on the file and highlight the “Save Target As...” option.
  5. In the “Save In” window, select “3 ½ Floppy (A:)” or the appropriate drive or folder where you want to save the dataset and click on the “Save” button.
  6. Repeat the process for the EARN2KC.NY dataset and the third dataset that is the state of your choice.
- B.** If you are using WebCHIP, 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.
1. Using the “Menu-Enabled Launcher” highlight “geography” and click “Select Target.”
  2. Next highlight “earn2kc” and then click “Select Target.”
  3. Finally, highlight “EARN2KC.DAT” or another data set of your choice and click on “Select Target.” WebCHIP will then launch using the dataset you selected.

## II. Variables

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:

**Income (Earning)** – the money a person makes from working, as wages, salary, or a form of self-employment, expressed as an annual amount.

**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.

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

**Work Age (WkAge)** – age in years, grouped into 9 – 10 year intervals, starting at age 16

## III. Using StudentCHIP

### A. Frequency Distributions or Marginals

1. From the **“Start”** button find **Programs → StudentCHIP**
2. Click On **File → Open** and double click on EARN2KC.DAT, StudentCHIP will list a short description of the data file you opened. You may need to change the path to find your file by going to the “Look In” window and selecting “3 ½ Floppy (A:)” or whatever drive or directory you have saved the files to in Part IA above.
3. To get a listing of the variables and their frequencies, go to the **“Command”** menu and click on **“Marginals,”** you should get the following output:

#### RaceLat

NHwhite	Black	Hispanic	Asian	AmIndian	Total
73.6	11.7	9.9	4.0	.7	= 100.0%

#### Gender

Male	Female	Total
58.7	41.3	= 100.0%

## Earning

<15K	15-25K	25-35K	35-50K	50K-75K	75K+	Total
12.3	22.3	20.9	20.4	14.7	9.4	= 100.0%

## WkAge

16-24	25-34	35-44	45-54	55-64	65+	Total
7.8	24.4	30.2	25.0	10.7	1.9	= 100.0%

This is a frequency table for the CPS sample of full-time, year-round workers in 2000. According to the table, 58.7% of all full-time workers are male as opposed to 41.3% female. 12.3% of full-time workers earn less than \$15,000 a year while 9.4% earn more than \$75,000. 73.6% of full-time workers are non-Hispanic white, 11.7% are black, 9.9% are Hispanic, 4.0% are Asian, and only 0.7% are American Indian. Workers between the ages of 35-54 account for 55.2% of all full-time workers. Only 1.9% of full-time workers are over the age of 65.

The frequency table allows one to get an overall sense of the distribution of a particular variable or set of variables that is an important place to start. However, what we are interested in exploring further is the impact of gender and race on earnings. That is, do men still generally earn more than women? What racial group typically has the highest earnings? Which group has the lowest earnings? Do blacks earn more than Hispanics? In order to address these questions, we will need to crosstabulate the variables of interest.

## B. CrossTabs

1. In order to do the crosstabulation of the two variables to explore, you want to be sure that you know how the variables might be associated. It makes sense to say that one's gender may influence his or her earnings, but it does not make sense to say that earnings influences whether one is male or female. The variable that influences or affects another variable is known as the independent variable (x) and the variable that is influenced or affected by another variable is called the dependent (y) variable. You can write this as:  $x \rightarrow y$ . In this case, we are interested in how gender influences earnings. Gender would be the independent variable (x) and earnings would be the dependent variable (y). In other words, earnings to some extent, depends on gender:  
**Gender  $\rightarrow$  Earnings.**
2. To create a crosstabulation table in StudentCHIP, you need to tell the program which is the independent and which is the dependent variable. From the "**Command**" **Menu**, click on "**CrossTab.**" A "**Table Variables**" window opens up with a list of variables. The dependent (y) or row variable is entered first, which in this case is "Earning." The independent (x) or column variable is entered next, which in our example is "Gender." From the "**Table**" **menu**, select the "**Percentage Down**" option. This is the best option to chose when your table is constructed with the dependent variable as the row and the independent variable as the column. You should get the following table:

	<u>Male</u>	<u>Female</u>	<u>All</u>
<b>75K+</b>	13.3	3.9	9.4
<b>50-75K</b>	18.6	9.3	14.7
<b>35-50K</b>	21.9	18.3	20.4
<b>25-35K</b>	19.1	23.4	20.9
<b>15-25K</b>	17.9	28.5	22.3
<b>&lt;15K</b>	9.2	16.7	12.3
<b>100%=</b>	56212417	39610636	N = 95823053

In this case, we can see that 31.9% of male full-time workers make \$50,000 or more a year as opposed to only 13.2% of females. On the other hand, 45.2% of female full-time workers earn under \$25,000 a year as opposed to only 27.1% of male full-time workers.

1. Do males or female full-time workers typically earn more money? Report some of the specific findings (as I did above).
2. What does this suggest about the current state of income inequality between men and women? Does the gender earning gap appear to be any smaller in the year 2000?

Now examine how race influences earnings by following the same process as step two above but substituting the “RaceLat” variable in place of “Gender.” You should be able to complete the table below and answer the following questions:

<b>Earnings</b>	<b>NHwhite</b>	<b>Black</b>	<b>Hispanic</b>	<b>Asian</b>	<b>Amlndian</b>	<b>All</b>
<b>75K+</b>						
<b>50-75K</b>						
<b>35-50K</b>						
<b>25-35K</b>						
<b>15-25K</b>						
<b>&lt;15K</b>						
<b>100%=</b>	70500890	11220490	9510429	3880288	710956	N = 95823053

3. What racial group has the highest earnings? Report some specific percentages from the table.
4. What racial group has the lowest earnings? Again, report some specifics.
5. Since these are all full-time workers, what factors might contribute to the observed racial differences in earnings?
6. What do you find most interesting and/or surprising about these findings?

## IV. Finishing Up

**A.** Repeat Part III using the EARN2KC.NY dataset for New York State. After calculating the marginals and crosstabulations for New York, answer the following:

1. What differences in income do you observe in NY state in comparison to the entire United States?
2. Is the gender gap in wage earnings higher or lower in New York?

3. How do the racial differences in earnings in New York compare to the national data?
4. What might account for the differences observed in New York? Be sure to summarize your findings using the marginal frequencies as well as the two crosstabulation tables using gender and race. Report specific findings from your tables, just as I did in my earlier analysis.

**B.** Repeat Part III again for a different state of your choosing. Remember that the suffix for each data set indicates the state it represents (for example EARN2KC.WI is the 2000 earnings data for Wisconsin). Answer the four questions for this state as you did in part A above. How does the gender and racial income inequality in this state compare to New York and nationwide? Why do you think the income situation in this state might be any different?

**C.** Your completed assignment should include:

1. A cover page with your name, course, section number, and title.
2. **ALL** of the output from your StudentCHIP analysis, which includes the marginals, earnings by gender and earnings by race for the national, New York and the additional state data. Note that the output from StudentCHIP is easily copied by left clicking and dragging your mouse on the output and then pasting it (right click) into your word processor. Please make sure that you have cleaned up your output, so you don't have four copies of the same table before you print it! If you copy the output into a word-processor, it would be helpful to label each table so that you can refer to it in your analysis.
3. A 3-4 page typed (word-processed) analysis of your findings addressing:
  - a) the marginals from each of the three data sets. Describe the distribution of gender, race and earnings of full-time workers at the national and state level. Is there much difference in the New York frequencies in comparison to the other state you selected? How does New York compare to the national level data? Refer to my description in PartIIIA3 for an idea of how to interpret the marginals.
  - b) the six questions posed in Part IIIB2 and the four questions in Part IVA. You should refer to the specific tables from your StudentCHIP output and draw on Chapters 9,10, and 11 as well as class discussions and the web resources provided at the end of this assignment to guide you in understanding and interpreting your findings.
  - c) the inequality trend models discussed in Marvin Olsen's "The Affluent Prosper While Everyone Else Struggles" from your reader. Does your analysis support the "Comfortable Class Convergence," "Affluent Class Convergence," "Class Stability" or one of the "Class Divergence" models? Provide evidence from your results to support your argument. In other words, given the data he discusses from 1970 and 1987, has income inequality in 2000 appeared to have stayed about the same, decreased, or increased?

## V. Using WebCHIP

The functionality and output from WebCHIP differs slightly from the StudentCHIP version. If you use WebCHIP, note the following changes in the instructions:

1. There is no "Command" menu, simply select the "Marginals" or "Crosstab" button.
2. For your Crosstab analysis, the row variable will always be "Earning," the column variable will be either "Gender" or "RaceLat."
3. Make sure your tables are constructed as "% Down."
4. DO NOT select any variables in the "Control by:" window.
5. If you want start over, click on the "Clear text" button. CAUTION: This will wipe out ALL your output.
6. 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."
7. To start analyzing a new data set, click on the "Back" button in your browser to return to the "WebCHIP Launcher" and follow the instructions from Part IB.

***This assignment is due at the start of class, Thursday December 5, 2002.***

### Useful Web Resources

Bureau of Labor Statistics: <http://www.bls.gov/>

CensusScope: <http://www.censuscope.org/>

Current Population Survey: <http://www.bls.census.gov/cps/cpsmain.htm>

The Social Science Data Analysis Network: <http://www.ssdan.net/index.shtml>

U.S. Census Bureau: <http://www.census.gov/>

WebCHIP Launcher: [http://www.ssdan.net/ida\\_resources.shtml](http://www.ssdan.net/ida_resources.shtml)