

This module is updated based on one created by Tim Thorton of SUNY Brockport to include more recent data and for ease of use with WebCHIP 4.0.

# 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 2008. The purpose of this assignment is to introduce you to some basic data analysis software (WebCHIP), 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.

### LEARNING OBJECTIVES

Skill

After using this module, students will gain skills in:

- Using software to access and analyze census data
- Identifying independent and dependent variables
- Employing control variables
- Forming testable hypotheses using quantitative data
- Quantitative writing
- Learning how to construct, read, and interpret bivariate tables displaying frequencies and percentages
- Using real world data to enhance and support key course concepts

### I. DATA SETS

You will need to use the **Earn** and **Earn\_NY**, plus one other data set from a state of your choice. You can access WebCHIP through the SSDAN website. Use these instructions:

- <http://ssdan.net/datacounts/webchip>
- From there, under "collections" find "**geo2008**" in the drop-down box and select it.
- Next under "Datasets" find the data set in the WebCHIP program you wish to use, either **Earn.NY** or another state and it is ready for analysis.

This module is updated based on one created by Tim Thorton of SUNY Brockport to include more recent data and for ease of use with WebCHIP 4.0.

## 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 (RaceEth)** – 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.
- **Non-Hispanic Others** – all persons who indicated their race as other and not of Hispanic origin
- **Non-Hispanic Multiple** – all persons who indicated themselves as being multiple races and not of Hispanic Origin

**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 WEBCHIP

A. Frequency Distributions or Marginals

- To get a listing of the variables and their frequencies, use the "**Marginals**" function. You should get the following output:

This module is updated based on one created by Tim Thorton of SUNY Brockport to include more recent data and for ease of use with WebCHIP 4.0.

### RaceEth

NHWhite	Black	Asian	Hispanic	AmIndian	NHOther	NHMulti	Total
67.67%	11.2%	5.31%	13.85%	0.65%	0.2%	1.12%	100%

### Gender

Male	Female	Total
57.09%	42.91%	100%

### Earning

<25k	25-34k	35-49k	50-69k	70-99k	100k+	Total
22.52%	17.59%	21.01%	17.34%	11.41%	10.13%	100%

### WkAge

16-24	25-34	35-44	45-54	55-64	65+	Total
7.22%	22.36%	25.19%	26.74%	15.72%	2.77%	100%

This is a frequency table for the CPS sample of full-time, year-round workers in 2010 ("acs2010" under collections "Earn" dataset). According to the table, 57.09% of all full-time workers are male as opposed to 42.91% female. 22.52% of full-time workers earn less than \$25,000 a year while 10.13% earn more than \$100,000. 67.67% of full-time workers are non-Hispanic white, 11.2% are black, 13.85% are Hispanic, 5.31% are Asian, and only 0.65% are American Indian. Workers between the ages of 35-54 account for 51.93% of all full-time workers. Only 2.77% 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.

This module is updated based on one created by Tim Thorton of SUNY Brockport to include more recent data and for ease of use with WebCHIP 4.0.

#### IV. CROSSTABS

A. 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 Earnings

B. To create a crosstabulation table in WebCHIP, you need to tell the program which is the independent and which is the dependent variable. Create a percent down crosstab with the dependent (y) or row variable "Earning" and the independent (x) or column variable "Gender." You should get the following table:

	<b>Male</b>	<b>Female</b>	<b>Total</b>
<b>&lt;25k</b>	18.5%	27.9%	22.5%
<b>25-34k</b>	15.3%	20.6%	17.6%
<b>35-49k</b>	20.2%	22.1%	21%
<b>50-69k</b>	18.7%	15.5%	17.3%
<b>70-99k</b>	13.5%	8.6%	11.4%
<b>100k+</b>	13.7%	5.3%	10.1%
<b>Total</b>	100%	100%	100%

In this case, we can see that 45.9% of male full-time workers make \$50,000 or more a year as opposed to only 29.4% of females. On the other hand, 27.9% of female full-time workers earn under \$25,000 a year as opposed to only 18.5% of male full-time workers.

This module is updated based on one created by Tim Thorton of SUNY Brockport to include more recent data and for ease of use with WebCHIP 4.0.

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 2008?

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>AmIndian</b>	<b>NHOther</b>	<b>NHMulti</b>	<b>All</b>
<b>100K+</b>								
<b>70-99K</b>								
<b>50-69K</b>								
<b>35-49K</b>								
<b>25-35K</b>								
<b>&lt;25K</b>								
<b>100%=</b>	45,394,805	7,509,805	3,560,795	9,291,319	434,753	137,416	750,423	N=67,079,316

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?

### **V. FINISHING UP**

A. Repeat Part III using the Earn\_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?

This module is updated based on one created by Tim Thorton of SUNY Brockport to include more recent data and for ease of use with WebCHIP 4.0.

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 **EARN\_WI.DAT** 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:

- A cover page with your name, course, section number, and title.
- **ALL** of the output from your WebCHIP 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 WebCHIP 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.
- A 3-4 page typed (word-processed) analysis of your findings addressing:
  - o 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?
  - o The six questions posed in Part IVB and the four questions in Part VA. You should refer to the specific tables from your WebCHIP output and the web resources provided at the end of this assignment to guide you in understanding and interpreting your findings.

## VI. USEFUL WEB RESOURCES

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

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

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

This module is updated based on one created by Tim Thorton of SUNY Brockport to include more recent data and for ease of use with WebCHIP 4.0.

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

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

apply what you have learned in the course to try to explain differences in earnings based on race and gender.