

Creating Univariate and Bivariate Tables

Learning Objectives

Skill

We will be analyzing Census data in order to:

- Evaluate hypotheses using both quantitative and qualitative data.
- Use simple statistics to analyze social data.
- Develop skills in using evidence.

I. WebCHIP

For this assignment, our first goal is to gain familiarity with WebCHIP. WebCHIP is an internet based system which includes data access and data analysis. You can access WebCHIP through the SSDAN website. Use these instructions:

1. Go to <http://www.ssdan.net/datacounts/data/>
2. From there, click “Browse” on the left sidebar. Find “**cen1990**” in the drop-down box and select it.
3. Scroll down through the list of data sets until you find “**earn9.dat**” Highlight and click “submit.” This will bring up the data set in the WebCHIP program and it is ready for analysis.
4. You can also click [here](#) to open the dataset in WebCHIP.
5. This data set has 4 variables: income (Earning), race (RaceLat), gender (Gender) and age of worker (WkAge).

II. Variables

Although there are several ways in which these variables may be conceptualized, defined and measured, these are the definitions used by the Census Bureau.

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-Latino White (NLWhite)** – all persons who indicated their race as white and not of Latino origin.
- **Black** – all persons who indicated their race as black.
- **Latino (Hispanic origin)** – 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 (NLOther)** – includes persons who indicated their race as “other” and are not of Latino origin.

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

III. Frequencies

To get a listing of the variables and their frequencies, go to the left top of the screen and use the **“Marginals”** function to get the following output:

RaceLat

NLWhit	Black	Latino	Asian	AmIndi	NLOthe	Total
80.4	9.7	6.5	2.8	.6	.1	= 100.0%

Gender

Male	Female	Total
61.4	38.6	= 100.0%

WkAge

16-24	25-34	35-44	45-54	55-64	65+	Total
9.	31.	28.6	18.9	10.4	2.1	= 100.0%

Earning

<15K	15-25K	25-35K	35-50K	50K+	Total
22.2	30.4	21.1	15.1	11.1	= 100.0%

Print this table and keep it in your notes.

This is a frequency table (also called a univariate table because we are looking at one variable at a time) for all full-time workers in 1990. According to the table, 61.4% of all full-time workers are male as opposed to only 38.6 who are female. 22.2% of full-time workers earn \$15,000 a year or less while 11.1% earn \$50,000 or more. 80.4% of full-time workers are non-Latino white, 9.7% are black, 6.5% are Latino and 2.8% are Asian.

The frequency table allows one to get an overall sense of the distribution of a particular variable or set of variables which 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 typically earn more than women? What racial group typically has the highest earnings? Which group has the lowest earnings? Do Blacks earn more than Latinos? In order to address this question, we will need to cross tabulate the variables of interest.

IV. Crosstab

1. In order to do the crosstabulation of two variables, you want to be sure that you know how the variables are 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 one's 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 is the independent variable (x) and earning is the dependent variable (y). Gender \rightarrow Earning. In other words, earnings to some extent, depends on gender.
2. Create a percent down crosstabulation table in WebCHIP with "Earning" as the "Row Variable" and choosing "Gender" as the "Column Variable". You should get the following table:

	Male	Female	All
<15K	15.9%	32.3%	22.2%
15-25K	25.6%	38.0%	30.4%
25-35K	23.0%	18.1%	21.1%
35-50K	19.4%	8.3%	15.1%
50K+	16.0%	3.3%	11.1%
100%=	45,544,808	28,646,328	N = 74,191,128

In this case, we can see that 16.0% of male full-time workers make \$50,000 or more a year as opposed to only 3.3% of females. On the other hand, 32.32% of female full-time workers earn under \$15,000 a year as opposed to only 15.9% of male full-time workers.

Now examine how race influences earnings, following the same process as step 2 above. (Start by deciding which are your independent and dependent variables.) You should be able to complete the table in order to answer the following questions.

Earnings	NLWhite	Black	Latino	Asian	AmIndian	NLOther	All
<15K							
15-25K							
25-35K							
35-50K							
50K +							
100%=	59660684	7174617	4842169	2056312	414470	42881	74191128

1. What racial group has the highest earnings?
2. What racial group has the lowest earnings?
3. Since these are all full-time workers, what factors might contribute to the observed racial differences in earnings?
4. What, if anything, do you find most surprising about these findings?

(Adapted from Timothy Thornton's "Census Data Assignment")