Amy E. Traver
Professor of Sociology
Queensborough Community College, The City University of New York

The Relationship Between Group/Categorical-Level Variables and Marital Status in the United States: A Data Analysis Assignment for Sociology of the Family[1]

Learning Objectives - Skills

- Students will construct hypotheses
- Students will access existing quantitative datasets from the United States Census
- Students will conduct secondary analyses of existing quantitative data using the online WebCHIP program
- Students will use the online WebCHIP program to produce bivariate tables with existing quantitative data
- Students will read and interpret bivariate crosstabulation tables
- Students will read and interpret pie, bar, and stacked-bar charts
- Students will evaluate hypotheses using quantitative data
- Students will write about their quantitative findings

Learning Objectives - Substance

- Students will learn how sociologists use existing data and statistical analytic techniques to describe, analyze, and understand the social world.
- Students will learn about the relationship between group-level/categorical variables and marital status in the United States.

In this assignment, we will be accessing 2000 United States Census data to better understand the relationships between group/categorical-level variables and marital status in the United States. In the process, we will develop our hypothesis construction/testing, data representation and analysis, and quantitative-data-based writing skills. We will also come to better understand the social realities of marriage in the United States. We will work on research question #1 in class today; you will complete research question #2 from home for collection the following week. The complete exercise is due XX, and it is worth XX% of your final course grade.

Research Preparation and Warm-Up

Go to the Social Science Data Analysis Network website at https://ssdan.net/datacounts. Then click on WebCHIP. In this assignment, we will be using the Census 2000 collection (from "Collections"), and the Marital2k dataset (from "Dataset"). Select each from their respective dropdown menus. Then follow the next five steps.

First, let's answer some questions about the collection and dataset.

1. What is the Census 2000 collection?

2. What is the Marital 2k dataset?

Second, let's get a sense of the American realities reflected in/by the overall dataset. Click on "Compute Marginals" (a blue box). The Marginals Workbook provides a general demographic overview of the dataset. To become comfortable with the dataset, let's look at the univariate (one variable) statistics that constitute the dataset.

1. How many variables are in the Marital 2k dataset, and what are they? How many categories exist for each variable, and what are those categories?

Third, let's remember how to read percentages.

1. What percentage of the Census 2000 population of Americans is male?
2. What percentage of the Census 2000 population of Americans is female?
3. What percentage of the Census 2000 population of Americans is 15-24 years old?
4. What percentage of the Census 2000 population of Americans is 45-54 years old?
5. What percentage of the Census 2000 population of Americans is over 65 years old?

Fourth, let's remember how to "translate" quantitative data into words.

1. Describe the race/ethnicity data in words.
2. Describe the marital data in words.

Fifth, let's remember how to think critically about data. What two percentages about Americans in this 2000 Census dataset most surprise you, and why?

1. Surprising percentage one, and why
2. Surprising percentage two, and why

Now, let's turn to the primary objective of this assignment, which is to better understand the social realities of marriage in America. More specifically, we want to use this dataset to better understand the relationship between marital status and big group/categorical-level social variables like gender and race/ethnicity. To do this, we need to move from our review of univariate statistics to conducting bivariate (two variable) statistical analyses.

Research Question #1:
According to marital status, how do American men and women compare?

Research Hypotheses

A hypothesis is an educated guess about what you expect to find in your research. In response to the research question above, let's construct four now.

Hypothesis 1. Of the 2000 Census population of Americans, do you think that a greater percentage of American men will be married, that a greater percentage of American women will be married, or that there will be no difference between the percentages? Why? Please provide some rationale or justification for your hypothesis.

Hypothesis 2. Which category of Americans (men or women) do you think will have a higher percentage of widows? Why? Please provide some rationale or justification for your hypothesis.

Hypothesis 3. Which category of Americans (men or women) do you think will have a higher percentage of divorcees? Why? Please provide some rationale or justification for your hypothesis.

Hypothesis 4. Finally, which category of Americans (men or women) do you think will have a higher percentage of never-marrieds? Why? Please provide some rationale or justification for your hypothesis.

Research Analysis: Cross Tabulation and Hypothesis Testing

To analyze the relationship between marital status and gender, let's construct a cross tabulation that puts these two variables into relationship. Scroll down to "Choose Variables." For row, choose "Marital." For column, choose "Gender." Then scroll down to "Generate Table" and click "Percent Down." Copy the data from WebCHIP into the chart below.

	Male	Female	Total
Currently Married			
Widowed			
Divorced			
Separated			
Never Married			
Total			

First, let's practice how to read the chart, paying attention to the difference between row and column percentages by answering the following 4 questions in turn and in relationship to each other.

1. Of American males, what percentage is currently married?
2. Of American females, what percentage is currently married?
3. Of the entire population of Americans, what percentage is currently married?
4. This means that a higher percentage of _____ are currently married than are Americans in general.

Second, given your reading of the chart, let's construct a chart title.

Third, let's take each of your four hypotheses in turn.

Hypothesis 1. What was your hypothesis and was it correct? Confirm or refute with data.

Hypothesis 2. What was your hypothesis and was it correct? Confirm or refute with data.

Hypothesis 3. What was your hypothesis and was it correct? Confirm or refute with data.

Hypothesis 4. What was your hypothesis and was it correct? Confirm or refute with data.

<u>Research Representation: Graphic Data</u>

Does representing your data graphically change your understanding of that data? To understand the power of graphical representation, let's construct two different graphs.

First, let's look at the data in pie-chart form. Go to "Generate Chart" and click on "Pie." Copy the chart into the assignment using the following shortcut: shift+window+s.

What this representation of the data does well is showcase how we are comparing two distinct categories (men and women) within the population of Americans. It also showcases, using brilliant colors, how the 5 marital statuses (currently married, widowed, divorced, etc.) *together* make up the whole of each gender category.

What do you now realize about the data by looking at it in pie-chart form?

Second, let's look at the data in bar-chart form. Again, go to "Generate Chart" but now click on "Bar." Copy the chart into the assignment using the following shortcut: shift+window+s.

What this representation of the data does well is showcase how we are comparing two distinct categories of Americans (men and women) *across the same* 5 marital statuses (currently married, widowed, divorced, etc.).

What do you now realize about the data by looking at it in bar-chart form?

Final Note: Direction Matters

What happens when we change the direction of percentage calculation for the marital status and gender data? Scroll back up to choose variables. Maintain "Marital" for row and "Gender" for column. However, now click "Percent Across" and copy the data from WebCHIP into the chart below.

	Male	Female	Total
Currently Married			
Widowed			
Divorced			
Separated			
Never Married			
Total			

Note how the computational direction of 100% has changed. What does this mean? How and why is this data different from what you collected and analyzed before?

Given that difference, how would you title *this* chart?

Pick a data point from this new set of data. Describe it in words.

Does analyzing the data in this way indicate something new about the relationship between marital status and gender in America?

Does analyzing the data in this way yield new or different concerns about the relationship between marital status and gender in America?

Research Question #2:
According to marital status, how do Americans of different racial/ethnic groups compare?

Research Hypotheses

As a reminder, a hypothesis is an educated guess about what you expect to find in your research. In response to the research question above, let's construct four now.

Hypothesis 1. Of the 2000 Census population of Americans, which racial/ethnic group do you think will have the highest percentage of currently married people? Why? Please provide some rationale or justification for your hypothesis.

Hypothesis 2. Which racial/ethnic group do you think will have the lowest percentage of divorced Americans? Why? Please provide some rationale or justification for your hypothesis.

Hypothesis 3. Which racial/ethnic group do you think will have the highest percentage of never married Americans? Why? Please provide some rationale or justification for your hypothesis.

Hypothesis 4. Finally, which racial/ethnic group do you think will have the highest percentage of widowed Americans? Why? Please provide some rationale or justification for your hypothesis.

Research Analysis: Cross Tabulation and Hypothesis Testing

To analyze the relationship between marital status and race/ethnicity, let's construct a cross tabulation that puts these two variables into relationship. Scroll down to "Choose Variables." For row, choose "Marital." For column, choose "RaceLat." Then scroll down to "Generate Table" and click "Percent Down." Copy the data from WebCHIP into the chart below.

	Non-Latino White	Black	Asian	Latino	Amer. Indian	Non-Latino Other	Non-Latino Multi-Racial	Other
Currently Married								
Widowed								
Divorced								
Separated								
Never Married								
Total								

First, let's practice how to read the chart, paying attention to the difference between row and column percentages by answering the following 5 questions in turn and in relationship to each other.

1. Of the population of Latino Americans, what percentage is separated?
2. Of the population of non-Latino-other Americans, what percentage is separated?
3. Of the population of Asian Americans, what percentage is separated?
4. Of the entire population of Americans, what percentage is separated?
5. This means that a lower percentage of _____ are separated than are Americans in general.

Second, given your reading of the chart, let's construct a chart title.

Third, let's take each of your four hypotheses in turn.

Hypothesis 1. What was your hypothesis and was it correct? Confirm or refute with data.

Hypothesis 2. What was your hypothesis and was it correct? Confirm or refute with data.

Hypothesis 3. What was your hypothesis and was it correct? Confirm or refute with data.

Hypothesis 4. What was your hypothesis and was it correct? Confirm or refute with data.

Research Representation: Graphic Data

Does representing your data graphically change your understanding of that data? To understand the power of graphical representation, let's construct a new type of graph.

More specifically, let's look at the data in stacked bar-chart form. Go to "Generate Chart" but now click on "Stacked Bar." Copy the chart into the assignment using the following shortcut: shift+window+s.

What this representation of the data does well is showcase how we are comparing seven distinct categories of Americans (Racial/Ethnic group) *across the same* 5 marital statuses (currently married, widowed, divorced, etc.).

What do you now realize about the data by looking at it in stacked bar-chart form?

Final Note: Direction Matters

What happens when we change the direction of percentage calculation for the marital status and race/ethnicity data? Scroll back up to choose variables. Maintain "Marital" for row and "RaceLat" for column. However, now click "Percent Across" and copy the data from WebCHIP into the chart below.

	Non-Latino White	Black	Asian	Latino	American Indian	Non-Latino Other	Non-Latino Multiracial	Other
Currently Married								
Widowed								
Divorced								
Separated								
Never Married								
Total								

Note how the computational direction of 100% has changed. What does this mean? How and why is this data different from what you collected and analyzed before?

Given that difference, how would you title *this* chart?

Pick a data point from this new set of data. Describe it in words.

Does analyzing the data in this way indicate something new about the relationship between marital status and race/ethnicity in America?

Does analyzing the data in this way yield new or different concerns about the relationship between marital status and race/ethnicity in America?