

ANALYSIS PAPER 2

QUANTITATIVE DATA ANALYSIS

The purpose of this assignment is twofold. In the first place, you will learn how to analyze quantitative data and interpret the results. In the second place, you will learn about the relevance of education and family type to earnings, how it has changed over time, and the relevance of race in understanding these relationships.

Learning Objectives

The learning objectives regarding quantitative data analysis include:

- To gain exposure to working with numerical data.
- To learn how to read and interpret univariate, bivariate, and multivariate tables.

The learning objectives regarding the content of sociology include:

- To recognize the earnings consequences of one's level of education.
- To understand how the relationship between education and earnings have changed over time.
- To understand how the relationship between education and earnings changes when taking race into account.
- To understand how the relationship between education, earnings, and race has changed over time.

Data: The data you will be using comes from the 1950, 1960, 1970, 1980, and 1990 Census and includes five variables: year, race, family type (Famtype), level of education (Educ), and family income (FmIncome). The categories for the five variables we will be using are as follows:

Year	Race	Family Type	Level of Education	Family Income
1950	Nonblack	MrrdCpl = Married Couple	LTHS = Less than high school	<15K = Less than \$15,000
1960	Black	MaleFam = single male head of family	HSGrad = High school graduate	15-25K = \$15,000 - \$24,999
1970		FemlFam = single female head of family	SomeColl = some college, no degree	25-35K = \$25,000 - \$34,999
1980			CollGrad = college graduate	35-50K = \$35,000 - \$49,999
1990				50K+ = \$50,000 or more

Task 1: Present the univariate results for the five variables

Step One: Opening the data set. Open your web browser (NetScape, Internet Explorer, etc.) and go to the following website: http://www.ssdan.net/ida_resources.shtml. Click on WebCHIP Launcher. Click on WebCHIP Launcher again! Under the heading “Menu-Enabled Launcher,” there is a box with a down arrow. Click on the down arrow and select “custom.” Then click on the “Select Files” button. Next, click on the down arrow again and select “fmin5090.dat.” Finally, click on the “Select Files” button one more time. It will take a moment, but you have now accessed the data in WebCHIP.

Across the top of the page, you will see the command buttons you will use for your analysis. Below that is the name of the data set: “1950-90 Family Income by Family Type for Family Heads Age 25+.” Below that, from left to right are the five variables together with the number of categories for each variable: Year (5), Race (2), Famtype (3), Educ (4), and FmIncome (5). On the last line you will see “N=.” This number represents the number of individuals in the U.S. population included in this data set. (**Note:** the number represents all those 25 and older living in households of two or more people who are relatives in 1950 *plus* all those meeting the same criteria in 1960, 1970, 1980, & 1990.)

Step Two: Displaying variable frequencies. To display the frequency distributions for each variable, all you need to do is click on the “marginals” button. Copy and paste (you may need to “Ctrl + C” to copy) the frequency distributions for race, education, and family income into your assignment under question number three. **Important note: make sure you have one-inch margins on the left and the right sides of your document. After pasting the frequency distribution into your document, highlight it and change the font size to 8 point type. Do this every time you copy and paste a table into your assignment.**

Task 2: Look at the trends in education and family income over time (bivariate analyses).

Step One: Education over time. On the WebCHIP command bar, find “row variable” and select “Year.” Find “column variable” and select “Education.” Click on the button for “% Across.” Click on “crosstab.” You will now see a table that displays distribution of educational attainment of the population in each time period. Copy and paste this table into your assignment under question number four.

Step Two: Family income over time. Repeat the directions in step one substituting family income for education. Copy and paste this table into your assignment under question number five.

Task 3: Analyze the relationship between education and family income.

Step One: Identify the variables. On the WebCHIP command bar, find “row variable” and select the dependent variable (is it education or family income?!). Find “column variable” and select the independent variable.

Step Two: Display the cross tabulation. Click on the button for “% Down.” Click on “crosstab.” You will now see a table that displays the relationship between education and family income. Copy and paste this table into your assignment under question number six.

Step Three: Control for the year (a multivariate analysis). On the WebCHIP command bar, in the “Control by” box, click on “year.” Then click on “control by.” Click on “crosstab.” You will now see *five* tables that display the relationship between education and family income, one for each year of the Census. Copy and paste the tables for the years 1950 and 1990 into your assignment under question number seven.

Step Four: Control for race (a multivariate analysis). On the WebCHIP command bar, in the “Control by” box, click on “race.” Then click on “control by.” Click on “crosstab.” You will now see *two* tables that display the relationship between education and family income, one for nonblacks and one for blacks. Copy and paste the tables into your assignment under question number eight.

Step Five: Control for race *and* year (a multivariate analysis). On the WebCHIP command bar, in the “Control by” box, click on “race.” Then click on “control by.” Then, again in the “Control by” box, click on “year.” Then click on “control by” again. Click on “crosstab.” You will now see *ten* tables that display the relationship between education and family income, controlling for race and year. Copy and paste the tables for nonblacks *and* black for the years 1950 *and* 1990 into your assignment under question number nine.

Step Six: You are now ready to answer question number ten.

Task 4: Analyze the relationship between three (or more) variables of your choice.

Step One: Create a table representing a *multivariate* analysis using a *different* combination of variables than we have done so far. Copy and paste the tables into your assignment under question number eleven.

Step Two: You are now ready to answer the remaining questions below.

Analysis Questions

(Type the following questions and your responses into your paper):

1. How many individuals are included in this data set?
2. What is the population included in this data set? What are the strengths and weaknesses of restricting the data to this population?
3. Copy and paste the table representing the frequency distributions for race, education, and family income. Write a title for the table that concisely describes what the table represents. What is your reaction to the actual percentages as compared to what you might have predicted?

4. Copy and paste the table representing the distribution of educational attainment within the population in each time period. (For this and each of the following questions, make sure you are copying *only* the table(s) asked for in the question.) Write a title for the table that concisely describes what the table represents. Provide a written, lay interpretation of the data (i.e. written for a newspaper-reading audience) as represented by this table, citing key percentages from the table which support your interpretation and making any relevant comparisons: What is the overall trend in the data? What might explain these findings?
5. Copy and paste the table representing the distribution of family income within the population in each time period. Write a title for the table that concisely describes what the table represents. Provide a written, lay interpretation of the data (i.e. written for a newspaper-reading audience) as represented by this table, citing key percentages from the table which support your interpretation and making any relevant comparisons: What is the overall trend in the data? What might explain these findings?
6. Copy and paste the table representing the relationship between education and family income. Write a title for the table that concisely describes what the table represents. Provide a written, lay interpretation of the data (i.e. written for a newspaper-reading audience) as represented by this table, citing key percentages from the table which support your interpretation and making any relevant comparisons: What is the overall trend in the data?
7. Copy and paste the tables representing the relationship between education and family income *controlled by year* for 1950 and 1990. Write a title for the table that concisely describes what the table represents. Provide a written, lay interpretation of the data (i.e. written for a newspaper-reading audience), citing key percentages from the table which support your interpretation and specifically comparing the relationship between education and family income between 1950 and 1990. What is the overall trend in the data? How has this relationship changed over time? What do you make of these changes?
8. Copy and paste the tables representing the relationship between education and family income *controlled by race*. Write a title for the table that concisely describes what the table represents. Provide a written, lay interpretation of the data (i.e. written for a newspaper-reading audience), citing key percentages from the table which support your interpretation and specifically comparing the relationship between education and family income blacks and nonblacks. How does race affect the relationship between education and level of income? What do you make of these differences?
9. Copy and paste the tables representing the relationship between education and family income *controlled by race AND by year* for 1950 and 1990. Write a title for the table that concisely describes what the table represents. Provide a written, lay interpretation of the data (i.e. written for a newspaper-reading audience), citing key percentages from the table which support your interpretation and specifically comparing the relationship between education and family income for blacks and nonblacks in 1950 and 1990. How does race affect the relationship between education and level of income in these two different time periods? Given that in 1950 the U.S. was still segregated and that in 1990 we were about 25 years beyond the Civil Rights Movement, what do you make of these differences?
10. Given what you have seen regarding the relationship between education and family income, and how this has varied over time and between races, what have you learned

about the relevance of education? What have you learned about the relevance of race? What might explain these findings?

11. Create a table representing a *multivariate* analysis using a *different* combination of variables than we have done so far. Title the table. What question does this table answer? What is the answer to that question as represented in the table you have created? Again, be sure to cite key percentages from the table which support your interpretation and make any relevant comparisons that highlight the trend apparent in the table.
12. Overall, what are the limitations of an analysis of data based on crosstabs like all the tables you created for this assignment?
13. The data you've just analyzed covers the entire United States over a five-decade period. As such, this represents a quantitative, macro-level analysis of the relationship between education, race, and family income. In the book, *Ain't No Makin' It*, the author presents a qualitative, micro-level analysis looking at these same basic variables. How well *and in what ways* does the quantitative data from this analysis fit together with the qualitative data from the book? What does the quantitative data tell us that the qualitative data can't? What does the qualitative data tell us that the quantitative data can't?
14. Write a "curiosity question" (i.e. a question you're curious about) regarding the relationship between the variables used in this assignment given what you've seen in your analysis of the data. What question (or questions!) arises from what you have seen in the data? What makes this an interesting, sociological question? How might it be possible to answer the question(s)?