SocSci200: Social Science Perspectives on the Family
Lab Exercise: Week 8
Spring 2000
Professors Brines, Cauce, Lundberg
May 18/19 2000

First task: Start up

1. Click the icon on the very lower left-hand side of your screen, labeled "Start" (with the Microsoft logo).

2. This will throw a menu on the screen. If the previous user of your computer did not logoff, you will want to shut down the connection and start again. In this case, move the cursor to "shut down". This will throw up another set of options. Select "close all programs/start as different user," and click.

3. The screen will now ask you for a password to the login name "public." There is no password for this account, so just leave the password field empty and hit return.

4. The computer will take a few seconds to log you in. Once the hourglass disappears from the screen, you can proceed.

5. Click "Start" again, locate "Run" on the menu of options and click. Windows will ask you to type in the name of the folder or file you want to run.

Type the following: o:classdata ssci200. This tells your computer where to go to get the data you'll analyze with SPSS.

6. Locate the "SPSS for Windows 9.0" icon on the desktop and click.

7. Now the computer is entering SPSS, otherwise known as the Statistical Package for the Social Sciences.

Second Task: Loading the Data Set:

1. Your screen might be titled "SPSS for Windows Viewer" and contain a warning. Don't be alarmed; it simply means that SPSS is looking for data. Ignore this warning and move to the 'SPSS Data Editor' window by clicking on the "Untitled-SPSS for Windows" button at the bottom of the screen. The window that pops up should be empty and look like a spreadsheet. What we need to do is retrieve some data and load it into this window.

2. In the upper-left of the Data Editor screen, click on 'File'.

3. The 'File" menu should pop down.

4. Click on 'Open'.

5. The 'Open File' window should pop up.

6. The prompt will be "Look in:" Find the C: drive and click.

7. Find the folder labeled 'Temp' and click.

8. Find the folder labeled 'ssci200' and click.

9. Now locate a file titled 'family.sav' and click

10. Click the Open button.

On the screen, you will see data from the file "family.sav." This dataset contains information for 1500 adults in the U.S. The data were gathered in 1996.

We are going to use these data to investigate the question:

Is there a relationship between the type of the family you live in and your happiness?

Our investigation will focus on how two variables are related to each other. The first variable is "famtype," which contains information on the family type of each person in the sample. The second variable, "happy," measures each person's self-reported happiness.

But there are many other variables contained in the dataset "family.sav." Before looking at the relationship between marital status and happiness, take a few minutes to familiarize yourself with some of the other variables in "family.sav." You'll be using some of them for your second paper assignment. Attached at the back of this exercise is a codebook that describes in detail the codes for selected variables in "family.sav."

Third Task: Inspect the variables "famtype" and "happy."

Before researchers attempt to analyze the relationship between two variables, they need to be thoroughly familiar with each variable. That is, they need to know what the codes mean, how the codes for each variable are distributed in the sample, and what the "typical" value is for each variable.

1. Check on how "famtype" and "happy" are coded by highlighting each variable and clicking "Data," then "Define Variable" (or refer to the attached codebook).

2. Find out how the codes are distributed in the sample by creating **frequency distributions** for each variable. Click on the "Analyze" menu at the top of the Data Editor Screen.

3. Click "Descriptive Statistics"

4. Click "Frequencies"

5. Find the variable 'famtype' on the shelf of variables and highlight it. Then click on the arrow between the shelf of variables and the "Variables" box.

6. Repeat this operation for the variable "happy."

7. The next step will give you measures of central tendency, or the "typical" value for each variable. Click on the "Statistics" button at the bottom of the "Frequencies" window.

8. Because of the way these variables are measured, only the "mode" is an appropriate measure of central tendency. Check it under the box labeled "Central Tendency."

9. Click "Continue"

10. Now click the button "Charts." One nice feature of SPSS is that it will transform your frequency tables into eye-catching bar charts or pie charts if you ask for these. Click on "bar charts" (preferred by social researchers over "pie charts," which are favored in business/marketing).

11. Click "Continue"

12. Click "OK"

13. In a new window, you should see: the modes for "famtype" and "happy"; the frequency tables for the two variables, and bar charts that graph the contents of the frequency tables. Use this output to answer the following questions (You can either write your answers on a sheet of paper OR if you have a printer card, you can type your answers directly onto the output by clicking on Insert (toolbar) and then New Text. Print your output at the end of today's lab).

(a) What is the least common household type observed in this sample? How do you know?

(b) Interpret the "mode" for the variable "famtype" (i.e., what does this number mean? Refer to last week's lab handout if you need help)

(c) In general, would you say that members of this sample were happy or unhappy at the time of the survey? Briefly describe the evidence.

Fourth Task: Crosstabulate "famtype" and "happy" to see if happiness depends on the type of family a person lives in.

1. At the toolbar, click "Analyze"

2. Click "Crosstabs"

3. In the "Crosstabs" window, you'll see the shelf of variables and to the right, a "row" box and a "column" box. One of the variables you will click into the "row" box, and the other, into the "column" box. But which is which?

The Rule: The **independent** variable goes into the **column** box
The **dependent** variable goes into the **row** box

When we look at the relationship between two variables, we usually assume that one of the variables has an impact or effect on the other, but not vice-versa. The variable that has an impact is called the independent variable, and the variable that is affected by it is called a dependent variable.

So now consider: is it most reasonable to think that your family type affects your happiness? Or does your happiness affect the type of family you're in?

4. Click the independent variable into the "column" box and the dependent variable into the "row" box.

5. Now click "Statistics"

6. In thinking about the impact of one variable on another, we'd like to have one statistic that tells us something about the strength of that impact. Looking at all the data in a table, especially if it's a large table with many rows and columns, can get complicated—it can be difficult to summarize the overall pattern or trend in the table. What we need is a single measure that might summarize the effect of the independent variable on the dependent variable in the table.

   Fortunately, SPSS provides us with these types of statistics; they're called **measures of association (MOA)**.

7. Look in the Statistics window for the "Nominal" and "Ordinal" boxes. Each item listed there is a measure of association; each is based on a slightly different mathematical formula. We won't concern ourselves with the mathematical details here. All you need to know is the following:

For each crosstab, SPSS will compute requested measures of association. Each measure gives a different view of the strength of the relationship between (in this case) "famtype" and "happy."

All measures of association (Cramer's V, Gamma, Somer's D, etc.) take on values within a fixed range: between –1.0 and +1.0.

Computed MOAs that have an **absolute value close to 1.0 give evidence of a very strong relationship** between the independent and dependent variable.

Computed MOAs that are **close to 0.0 give evidence of a very weak connection** between the two variables. Use the following guidelines to interpret the values for MOAs (disregard the +/- signs):

Absolute value of MOA	Interpretation
0.00 to .04	No; nearly no relationship
0.05 to .09	Very weak relationship
0.10 to .14	Weak relationship
0.15 to .24	Moderate relationship
0.25 to .34	Substantial
0.35 to .49	Strong
0.50 to .89	Very strong
0.90 to .99	Near-perfect
1.00	Perfect

Choosing among MOAs listed under "Nominal" vs. "Ordinal" depends on the types of variables analyzed. See p. 9 for how to tell the difference between "Nominal" and "Ordinal" variables.

The Rule: Choose one of the "Nominal" MOAs **unless** both variables are of the "ordinal" type. To simply things, always use "Phi and Cramer's V" for a Nominal MOA; always use "Gamma" whenever choosing an Ordinal MOA.

Because the variable "famtype" is nominal, we must choose from among the "Nominal" MOAs.

8. Click "Phi and Cramer's V" under "Nominal."

9. Click "Continue"

10. Click "Cells"

11. In looking for a relationship between famtype and happy, you'll be looking for evidence that the percentage of people who say they are very happy (or not at all happy) is different for different family types. When comparing the happiness of people in different types of families, DO NOT compare the NUMBER of cohabiting people vs. single parents vs. etc. who say they're very happy, pretty happy, or not at all happy. Many people made the mistake of using raw numbers as evidence instead of percentages/proportions in their first papers. Instead, compare the PERCENTAGES of cohabiting people vs. single parents vs. etc. who report different levels of happiness.

12. To do this, click "column" in the box labeled "Percentages." This will give you the appropriate percentages in your crosstab output. ALWAYS refer to these percentages, not the actual counts of people, in discussing your results.

13. Click "continue"

14. Click "OK"

15. After a few seconds, three pieces of output will appear. The first will be labeled "Case Processing Summary"; this you can disregard. The second piece is your crosstabulation of happiness by famtype. The small boxes or cells in the table contain the actual number of people in the sample with that column's code for famtype and that row's code for happy, and the percentage of people of that column's family type with that row's code for happy. Take for example the upper left-hand cell. The number '159' means that there were 159 people (out of the total 1500) who lived in childless couples AND said that they were very happy. You'll also see the figure "40.5%." This means that 40.5% of childless couples said they were "very happy." Notice that only 15.3% of single parents said they were "very happy."

Focus on these percentage comparisons when answering the question: **Is there a relationship between family type and happiness?**

(d) In 3-5 sentences, use the crosstabulation to compare the happiness of people living in different types of families. Which group is happiest? Least happy? Do married people seem to be happier than singles?

16. Now move on to the third piece of output, labeled "Symmetric measures." SPSS has computed two measures of association from the crosstabulated data: Phi and Cramer's V. In large tables (i.e., more than four cells), Phi tends to give bad information, so use Cramer's V. The two columns give two pieces of information about Cramer's V: its value and its statistical significance.

The guidelines on the previous page tell whether the value for Cramer's V indicates a strong relationship, a moderate one, or a weak one between family type and happiness.

Statistical significance has to do with whether the results from this sample of 1500 came about by chance. It's about whether we can take much stock in them as a reflection of what is true for all adults in childless couples, single parent households, etc. in the U.S. (more than 100 million people). The sample of 1500 was carefully selected to reproduce important characteristics of the 100+ million population so that the patterns and relationships we observe in the sample will reflect what's true for the population.

But it's still possible that, by chance, the Cramer's V observed for this small group of 1500 is unique to this group and that in fact, there is no relationship between family type and happiness for the adult population as a whole. The number under "Approx Sig" tells you how likely it is that this happened—that the sample of 1500 gave you a nonzero Cramer's V when the truth for the 100+ million would be "no relationship between famtype and happy" (Cramer's V = 0.0).

The Rule: A small number under "Approx Sig"—anything less than .05—tells you that a "chance" result from your sample is not likely. This means that you can believe in your sample results if they give you a nonzero MOA —they're a good reflection of what you would find if you had data for all 100+ million U.S. adults. We would call your sample evidence showing a connection between family type and happiness "statistically significant."

Answer the following questions:

(e) Based on the value for Cramer's V, describe the strength of the association between family type and happiness: how much of an impact does the former have on the latter?

(f) Can you be confident that your answer to (e) applies not only to this sample of 1500 people but to the entire population of adults in the U.S.? Why or why not?

<u>Fifth Task (Optional): Printing a copy of your output</u>

1. Print a copy of your output/answers by clicking the printer icon. Return to the tree diagram of your SPSS output (left-hand side of your screen). Click on Frequencies and Crosstabs. This will select all components of your output for printing. Then click the printer icon on the toolbar.

2. The Print window will pop up. Click "OK"

3. Put away your output by double-clicking the button in the upper left-hand corners. Answer 'no' when SPSS asks if you want to save your output.

Sixth Task: Saying Goodbye

1. Call up the Data Editor window if it is not already on the screen by clicking on the "family.sav" button at the bottom of your desktop.

2. Click on "File" at the top of the screen.

3. Click on "Exit."

4. You're done!

SAVE THIS HANDOUT!!!! You'll want to refer to it as you conduct your data analysis for the second paper assignment.

Nominal and Ordinal Variables

Choosing an appropriate measure of association for a crosstabulation depends on the type of variables you're working with. If at least one of the two variables is of nominal type, then you must select from the "Nominal" menu of possibilities. If both of the variables are of ordinal type, then you should select from the "Ordinal" menu. For the purposes of this course, always choose "Phi and Cramer's V" under Nominal, and "Gamma" under Ordinal.

Nominal variables are variables that measure things in terms of distinct, qualitatively different categories. These categories cannot be ranked or ordered in value in any meaningful way. Even though the codes for the variable might be numeric—i.e., 1, 2, 3, 4—the categories they signify represent different social groups or attributes that cannot be ordered reasonably from "low" to "high." Examples of nominal variables include race or ethnicity, gender, family/household type, religious affiliation, political party affiliation, and most survey items that elicit either a "yes" or "no" response.

Ordinal variables are made up of categories that can be ranked in a meaningful way. The numbers that "code" the categories specify the rank of that category amid all the other categories for the variable. It is meaningful to think of the categories as ranging from "low" to "high" or "high" to "low." Examples of ordinal (i.e., "ordered") variables are: highest educational degree (1=less than high school, 2=high school diploma, 3=2year college degree, 4=4yr college degree, 5=postgraduate degree), and a person's social class (1=upper class, 2=middle class, 3=working class).

All of the variables in family.sav are either nominal or ordinal.