

**Social Structure—Personality: What is the relationship  
between social class and child-rearing values?**

**INTRODUCTION**

The sociologist, Melvin Kohn, argued that people's locations in social structures, particularly the occupational structure, influenced the values they would stress for their children because variations in structural locations exposed them to different experiences. His research, however, was done some time ago. Much has changed in our social world since then. Does the original relationship still hold?

**LEARNING OBJECTIVES**

*Skills*

**This assignment is asking you to generate statistical results, as well as to analyze and interpret the data produced. Other skills you will learn include:**

- Using software to access and analyze data
- Identifying independent and dependent variables
- Forming testable hypotheses using quantitative data
- How to construct, read, and interpret bivariate tables displaying frequencies and percentages
- Identifying population trends over time

*Substance*

**Analytically, we will answer the following 3 questions:**

- How important to people is stressing to their children that they think for themselves?
- Does social class influence this child-rearing value of thinking for yourself?
- Has this influence changed over time?
- Other learning outcomes for this assignment include honing analytical skills associated with generating, reading, and interpreting rudimentary forms of data analysis.

**A PRIMER ON THE LOGIC OF SOME BASIC FORMS OF DATA ANALYSIS**

*Introduction*

Most science is predicated on the belief that the world is not just a random, chaotic mess, but rather things happen in certain ways and for certain reasons. Therefore, scientific research is often categorized by two major objectives—**to describe**

and **to analyze**. Scientists will document the order or patterns in the world (our descriptive objective) and the reasons behind such patterns (our analytical objective). Social scientists are no different. We're interested in whether or not there are patterns or relationships among the social phenomena on which we collect information. Analyzing this information, whether quantitative or qualitative in form, involves a conversation between you and the data. As researchers, we pose questions guided by our conceptual understandings of the subject matter and then use various statistical strategies to see what the data suggest in the form of answers. We have a variety of tools for collecting information or data; surveys are one major conduit for such information. Survey research translates our phenomena of interest into **variables**. A variable is anything that can vary, i.e., take different values, and it represents our way of measuring the concepts in which we are interested. For example, gender is a variable because people can be male or female. Information on variables like gender from surveys can be entered into a computer program by assigning numerical values to people's answers. Then the data can be analyzed in the search for patterns.

### *Univariate Analysis*

Now, to look for patterns, I might examine the variation on just this one variable by looking at the distribution of cases [respondents] across the variable's response set (we call this doing "**univariate** analysis" by using a statistical program to generate a "**frequency distribution**"). In reference to gender, doing that would tell me how many (or what proportion in %) of my survey respondents classified themselves as male or female. Thus, I get a portrait of how gender is distributed in my sample. While it doesn't take me real far, this type of analysis is still useful in its own right.

Since we're interested in beliefs, behaviors, and feelings, here's another real example from some actual data taken from the same survey you'll be analyzing. It's called the General Social Survey (GSS). Further below you'll find a brief description of the GSS. The table below presents the distribution of people's answers to the question, 'in general, do you find life exciting, pretty routine, or dull?'

Table 1: Respondents' Self-Reported Level of Excitement in Life—1998.

---

<i>Level of Excitement</i>	<i>% of Sample</i>	<i>N</i>
Exciting	45.1%	838
Routine	49.4%	919
Dull	5.5%	103
Total	100%	1860

---

### *Trend Analysis*

Another useful way to look at frequency distributions is to examine how they change over time. We call this **trend** analysis. For instance, keeping with my above example, if I had surveys from different years that asked the same question on how exciting people find their lives, I could document whether or not there's been any change. In fact, over the years, the GSS has asked people this question. Here's an example of looking at some trend data on life excitement using their results (Note: GSS aggregated their results in the first time-period):

Table 2: Trends in How Exciting Respondents Find Their Lives: 1988-2000.

	<u>1988-1991</u>	<u>1998</u>	<u>2000</u>
<i>Level of Excitement</i>			
Exciting	44.7%	45.1%	52.1%
Routine	50.4%	49.4%	44.2%
Dull	4.9%	5.5%	3.7%

### *Interpreting Data*

Statistical software generates the results like those shown above. However, I've taken those results and dressed them up a little in a couple professional-looking tables (when you do your actual analysis, you'll see how the output is different from these tables). But while the computer does the *grunt* work, you must still interpret what the numbers are saying. **Being able to clearly and concisely interpret results for an audience is an important skill to develop for many occupations.** For these simple tables, probably 2-4 sentences would suffice. You won't see this in a textbook, but **here are a couple guidelines for interpreting this kind of information. First, always provide the audience with some of the contextual detail (i.e., what is the data, where does it come from, when, etc.). Second, construct your interpretation from the "general to the specific."** In other words, start out with a general statement and then move to including some specific figures. Interpreting this kind of frequency distribution often makes use of a particular statistic called the mode. A **mode** "is the most frequent, most typical, or most common value in a distribution" (Levin & Fox, 1994, p. 77). In other words, for this data, it's the category with the largest percentage of respondents. Using these guidelines, interpreting Table 1 might go something like this:

*Respondents from the 1998 General Social Survey are nearly split on how exciting they find life. A slight majority of GSS respondents, based on their self-reports, find life either routine (49.4%) or dull (5.5%). On the other hand, nearly half (45.1%) report that they find life "exciting."*

Here's what we might say in looking at the trends in Table 2:

*The percentage of GSS respondents that report life is “exciting” increased from 1988-2000. While less than half of those surveyed in the 1988-1991 period (44.7%) reported life is “exciting,” slightly over half (52.1%) were excited in 2000. The greatest increase happened between 1998 and 2000. Overall, those who found life either “routine” or “dull” declined in this same time period.*

#### **BIVARIATE ANALYSIS: LOOKING FOR RELATIONSHIPS BETWEEN TWO VARIABLES**

That’s a brief introduction to some basic types of data analysis, but if we want to see how variables are related to each other, we can use a technique called cross-tabulation. A **cross-tab** is a table that presents the distribution (in frequencies and/or percents) of one variable across the categories of another variable(s) (e.g., what percentage of men find life exciting compared to women?). Since it lets us look at two variables, we label this a **bivariate** analysis.

Typically, in crosstabs and many other statistical techniques, we conceptualize the relationship between the two variables in terms of one influencing the other. The language we use to capture such relationships is to call one variable an **independent variable (IV--it's doing the influencing)** and the second variable a **dependent variable (DV--it's the one that is being influenced)**. To run a crosstab you decide on two variables that you think might be related to one another. Then, drawing from a conceptual framework, one next states a **hypothesis** for the relationship between your chosen variables. A hypothesis is an educated guess about what you think you will find. Hypotheses should always state a specific relationship and specify the comparison (e.g., Women are more likely than men to find life exciting.).

In order to understand the logic of crosstabulation, think of a psychologist’s experiment. Say she wanted to examine how a particular studying strategy affects students’ test performance. The previous sentence casts study strategy as my IV (it’s doing the influencing) and test scores as the DV (it’s being influenced). She might hypothesize that students using the special study strategy are more likely to do better on the test compared to those students who didn’t use the strategy. So she sets up a situation where one group is given the special study strategy and the other group is not. She then tests the students. How does she figure out whether the study strategy makes any difference? She needs to compare the two groups, of course. For example, she might calculate the mean test score for each group to see if one is higher than the other. Or she might divide up test scores into “low,” “medium,” and “high” categories and see what percentage of students falls into each one for the two groups. This latter strategy is what crosstabulation analysis is all about.

For this assignment, we’re interested in assessing whether social class makes a difference in childrearing values. Kohn originally stressed occupational location, but for

this assignment we'll use another rough structural proxy—social class location. While it's not perfect, it's reasonable to think that people in higher social classes are more likely to be in the kinds of jobs that stress autonomy and self-direction. If that's the case, who will be more likely to stress “thinking for him/herself” as an important value for kids to develop? Stating an answer to this question is formulating a hypothesis. But you need to be clear on thinking about what is the **independent variable** and the **dependent variable** (hint: is social class doing the influencing or is it the thing being influenced?).

To show you what a crosstab looks like and how it's interpreted, I'll use another topic related to childrearing and which also draws on the social structure—personality approach. Sociologists know that social class is connected to many things. An analytical question might be, does social class make a difference in how parents discipline their children? Drawing from Kohn, middle class parents, because they tend to work in less regimented environments, might actually be less regimented in their approach to discipline. A concrete variable could be whether or not someone favors physical forms of punishment. Let's say that the actual measures for these two concepts are educational level as a proxy for social class and the degree to which a parent believes in spanking as a form of discipline (measured with a response set of “strongly agree,” “agree,” etc.). (Before going any further, let me stress that these measures are not “perfect” indications of the concepts we're interested in. There is no such thing as a “perfect” indicator. But they are decent approximations and social scientists work with these types of indicators all the time.) I hypothesize that parents with higher levels of education will be LESS likely to favor spanking as a disciplinary technique compared to parents with a lower level of education. Here, education is the IV and attitude towards spanking is the DV. Below are results generated from the GSS for their 1993 survey.

**Table 3: The Relationship Between Educational Level and Attitudes Toward Spanking for Discipline—1993 GSS**

	<b>Educational Level</b>	
	<u>High School Grad Or Less 4-yr.</u>	<u>College Degree or Higher</u>
<b>Favor Spanking to Discipline A Child</b>		
Agree	76.1% (542)	66.3% (189)
Disagree	23.9% (170)	33.7% (96)
	100%	100%

Interpreting the results. Here again, the computer does the number-crunching for you, but you still have to interpret what the tables are “saying.” Remember what you’re after: you want to see if the IV (education) makes any difference in the DV (spanking attitudes). To do that, you want to compare percentage differences **between the categories of the IV** (e.g., compare those with lower levels of education against those with higher). *[Note: to do this properly, you want to make sure each category of the IV adds up to 100%. Analytically, you’re asking what % of people with a high school education or less favor spanking versus what % of people with a college education?]*

In this example, if you compare DV categories, you see that a higher percentage of people approve of spanking than don’t. I’m not saying that’s not interesting, but it doesn’t address the conceptual question we began with: does social class make a difference in childrearing attitudes? To answer that, we look at differences *between* educational categories. Besides, you could have already noted the differences in those who approve or disapprove spanking in a univariate analysis (frequency distribution).

There are three basic questions we can answer from this type of crosstabs output. You should frame your interpretations around these 3 questions:

- Is there a relationship between the IV and DV? (You’re looking to see if there is enough of a difference in percentages between the columns to matter. Some guidelines are found in the third bullet here.)
- What is the relationship? (Here you want to put in sentence form what it is that the table is telling you. Be careful how you state it because you may end up saying something that the results don’t show. You always state the % of people in the IV [from below, “...two-thirds (66.3%) **of those with a 4-year college degree or higher...**”]. Also, please be aware that I’ve given a simple example here. Many of our variables of interest have more than two categories. If your IV has 3 or more categories, the logic is still the same. You want to just compare across the IV categories. If the categories on each end seem to represent opposite ends of whatever variable you’re looking at, then you can look at those.
- How strong is the relationship? (For this one, you want to look at the % differential. In the example, it’s roughly 10%, which, in the social sciences, is a moderate relationship. Here’s my rule of thumb for gauging strength: If it’s 1-3%, then I’m hesitant to say there’s much of a relationship. If it’s 4-7%, then it’s a slight relationship. An 8-15% difference is a moderate relationship, while anything approaching 20% is a substantial relationship. This is only meant as a rough guideline. These aren’t technical rules.)

Here’s a summary of Table 3 that addresses these questions (if you were writing a summary, use the same guidelines that I mentioned earlier. Give some contextual details and move from the general to the specific in your summary):

*Table 3 shows a moderate relationship between education and whether or not people favor spanking as a means for disciplining children. Respondents who have more education are less likely to favor spanking as a way of disciplining children. According to GSS respondents in 1993, two-thirds (66.3%) of those with a 4-year college degree or higher approved of spanking compared to three-quarters (76.1%) of those with a high school degree or less.*

**YOUR DATA** (Most of the following information is adopted from the NORC website [www.norc.uchicago.edu] and Davis and Smith's, *The NORC General Social Survey: A User's Guide*.)

For this assignment, I'm sending you to a University of Michigan website for the Interuniversity Consortium for Political and Social Research (its address is provided further below). This outfit's basic function is to facilitate social science research by maintaining and providing access to literally hundreds of data sets. We'll be accessing data from the General Social Survey. The General Social Survey has been administered since 1972. Originally done annually, the survey became biennial in 1994. It is one of the most significant data sets used by sociologists. According to the National Opinion Research Center's description of the GSS, its "basic purposes ... are to gather data on contemporary American society in order to monitor and explain trends and constants in attitudes, behaviors, and attributes; to examine the structure and functioning of society in general as well as the role played by relevant subgroups; to compare the United States to other societies in order to place American society in comparative perspective and develop cross-national models of human society; and to make high-quality data easily accessible to scholars, students, policy makers, and others, with minimal cost and waiting. Since 1988, the GSS has also collected data on number of sex partners, frequency of intercourse, extramarital relationships, and sex with prostitutes" (quoted from the NORC website).

The GSS is a survey, but the data is actually collected through an **in-person interview** (rather than a paper questionnaire) which lasts roughly 90 minutes. Sample sizes ranged from roughly 1500 in the earlier years to currently about 3,000 since it's been done biennially. To select respondents, the GSS uses representative sampling techniques. In order to find people in households, they employ what's called, in research language, a **multi-stage, probability design**. In plain English, this means that in order to find people, they sample through a succession of geographical areas, continually narrowing the lens until they get to a house that they can knock on the door (e.g., it might go something like this: Northeast New York Western NY Rochester Greece Flynn Rd House #497). At each stage, they are randomly selecting the unit (i.e., think pulling names out of a hat). Thus, it's meant to be a representative sample of noninstitutionalized adults in America. Practically, this means that certain groups might be slightly underrepresented (e.g., those 18-25 who are more likely to live in college dorms or military quarters or elderly who might be in nursing homes). The response rate for the survey typically hovers around 70%. That means that of all the doors they knock

on and the particular persons they ask to interview, around 30% might say no or the interviewers simply could never find the person at home even though the researchers came back a couple times. In research, this is considered to be a pretty good response rate. The survey is also only done in English, so Hispanic people are likely to be underrepresented.

**Now on to your assignment!!**

It used to be that researchers would have to obtain the whole data set from the Consortium, but computer technology now enables us to actually analyze data from their website. Way Cool!! Using the data analysis skills introduced here, and with instructions for doing the statistical analysis provided below, we will test Kohn's social structure—personality theory with the GSS data plus examine whether the relationship has changed over time.



**Task #1:** Produce and interpret frequency distributions for the two variables—**thnkself** and **class**.

Procedures:

1. Go to the website <http://www.norc.org/GSS+Website/>
2. Click on the button for “Data Analysis.”
3. On the left sidebar, click to expand the “NORC Public Use Data Catalog.”
4. Click to expand “GSS.”
5. Click to expand “General Social Surveys, 1972-2006 [Cumulative File.]”
6. Click to expand “Variable Description.”
7. Click to expand “Mnemonic.” Select “T” from the alphabet and scroll down to find **thnkself**. Click on that to see the exact question wording (there is also additional information there). Do the same for **class**.
8. Go back to the main GSS screen, click on “analyze”; select “frequencies or crosstabulations” and click “start.”
9. In the “row” box, type in “thnkself, class” then in the “selection filters” box, type “year(2002)” [skip the quotation marks]. Make sure the “column” box by “percentaging” is checked. Ignore the “other options” and hit “run the table.”
10. Take the resulting output and fill in the table I’ve provided at the end of the handout. As you are doing this, think about how you would interpret the table by addressing these questions, ‘how important is thinking for oneself for the survey respondents?’ and ‘how are the respondents distributed by social class?’

**Task #2:** Produce and interpret two separate crosstabulations of **thnkself** by **class**. Do this for the years, 1983-87 (aggregated) and 2002.

Procedures:

1. First, based on Kohn’s theory, you must formulate a hypothesis. Which social classes do you think will be more likely to emphasize thinking for one’s self—the lower or higher social classes? (This will be a quiz question.)
2. Follow same instructions from above, 1-4.
3. Put **thnkself** in the row box. Now, we’re going to modify it slightly. From the frequency distributions, you’ve seen that this variable has 5 categories and class had 4. When we put them together into a crosstab, that produces a table with 20 cells (5x4). That can be difficult to interpret so we want to make it a little simpler. This website allows us to temporarily recode variables to create a smaller number of categories. With **NO** space after **thnkself**, type in “(r:1-2; 3-5)” [again, forget the quotation marks but do everything else exactly as I’ve done it here. If you get confused, you can find instructions by clicking on the “recoding variables” label above in this screen.] This instruction is telling GSS to add together the first two categories (“most” and “2nd most” important) and the last 3 (“3rd” through “least” important) categories. We’ll call the first one “more important” and the other “less important.”

4. Put **class** in the column box and in the selection/filters box put “year(1983-1987)”. Be sure the “columns” box is checked next to “percentaging.” Ignore “other options” and hit “run the table.” Then, following these instructions, do another crosstab for 2002.
5. Take the resulting output and fill in the table I’ve provided at the end of the handout. Use my earlier one as a guideline. You need to fill in the %s and the actual numbers (in parentheses) from each cell. Things to consider when filling out the table: for each table, the analytical question is, ‘which social class values thinking for oneself more?’ But then you want to address a final question, has the relationship between social class and thinking for oneself changed over time? Are the classes getting further apart or closer together on this value? Please speculate on reasons why you think you’re seeing what you’re seeing. Relate to class discussion.

**IMPORTANT SUGGESTION:** I strongly suggest writing out interpretations of these tables for yourself (it would only take a couple paragraphs).

Table 1: 2002 GSS Respondents' Emphasis on the Importance of a Child Thinking for Oneself

How Important is it To Think for Oneself?	% of Sample	N
Most Important		
2nd most Important		
3rd Most Important		
4th Most Important		
Least Important		
Total	100%	

Table 2: 2002 GSS Respondents' Self-Reported Social Class Location

Class Location	% of Sample	N
Lower Class		
Working Class		
Middle Class		
Upper Class		
Total	100%	

Table 3: Importance of a Child Thinking for Oneself by Social Class: Comparison between 1983-87 and 2002.

<b>1983-1987</b>				
	<i>Lower Class</i>	<i>Working Class</i>	<i>Middle Class</i>	<i>Upper Class</i>
<u>Importance</u>				
More Important	% ( )			
Less Important				
Total	100%	100%	100%	100%
<b>2002</b>				
	<i>Lower Class</i>	<i>Working Class</i>	<i>Middle Class</i>	<i>Upper Class</i>
<u>Importance</u>				
More Important	% ( )			
Less Important				
Total	100%	100%	100%	100%