

**SSRIC Teaching Resources Depository**  
**Public Opinion on Social Issues -- 1975-2010**  
**Elizabeth N. Nelson and Edward E. Nelson, California State**  
**University, Fresno**

**Chapter 3**  
**Introducing a Control Variable (Multivariate Analysis)**

*© The Authors, 2011; Last Modified 25 June 2011*

Human behavior is usually too complicated to be studied with only two variables. Often we will want to consider sets of three or more variables (called **multivariate analysis**). We will want to consider three or more variables when we have discovered a relationship between two variables and want to find out (1) if this relationship might be due to some other factor, (2) how or why these variables are related, or (3) if the relationship is the same for different types of individuals.

In each situation, we identify a third variable that we want to consider. This is called the **control** or the **test variable**. (Although it is possible to use several control variables simultaneously, we will limit ourselves to one control variable at a time.) To introduce a third variable, we identify the control variable and separate the cases in our sample by the categories of the control variable. For example, if the control variable is age divided into two categories--younger and older, we would separate the cases into two groups. One group would consist of individuals who are younger and the other group would be those who are older. We would then obtain the crosstabulation of the independent and dependent variables for each of these age groups. Since there are two categories in this control variable, we obtain two **partial tables**, each containing part of the original sample. (If there were three categories in our control variable, for example, young, middle aged, and old, we would have three partial tables.)

The process of using a control variable in the analysis is called **elaboration** and was developed at Columbia University by Paul Lazarsfeld, Patricia Kendall, and their associates. There are several different types of outcomes to the elaboration process. We will discuss each briefly.

Table 2.3 showed that females were more likely than males to say they were willing to vote for a woman. Let's introduce a control variable and see what happens. In this example we are going to use age as the control variable.

Table 3.1 is the three-variable table with voting preference as the dependent variable, sex as the independent variable, and age as the control variable. When we look at the older respondents (the left-hand partial table), we discover that this partial table is very similar to the original two-variable table (Table 2.3). The same is true for the younger respondents (the right-hand partial table). Each partial table is very similar to the original two-variable table. This is often referred to as **replication** because the partial tables repeat the original two-variable table (see Babbie, 2009). It is not necessary that they be identical; just that each partial table be basically the same as the original two-variable table. Our conclusion is that age is not affecting the relationship between sex and voting preference. In other words, the difference between males and females in voting preference is not due to age.

**Table 3.1 -- Voting Preference by Sex Controlling for Age**

	<b>Older</b>			<b>Younger</b>		
	Male %	Female %	Total %	Male %	Female %	Total %
<b>Voting Preference</b>						
Willing to Vote for a Woman	43.8	56.1	49.0	44.2	55.8	52.9
Not Willing to Vote for a Woman	56.2	43.9	51.0	55.8	44.2	47.1
	100.0	100.0	100.0	100.0	100.0	100.0
	(240)	(180)	(420)	(120)	(360)	(480)

Since this is a hypothetical example, imagine a different outcome. Suppose we introduce age as a control variable and instead of getting Table 3.1, we get Table 3.2. How do these two tables differ? In Table 3.2, the percentage difference between males and females has disappeared in both of the partial tables. This is called **explanation** because the control variable, age, has explained away the original relationship between sex and voting preference. (We often say that the relationship between the two variables is **spurious**, not causal.) When age is held constant, the difference between males and

females disappears. The difference in the relationship does not have to disappear entirely, only be reduced substantially in each of the partial tables. This can only occur when there is a relationship between the control variable (age) and each of the other two variables (sex and voting preference).

Next, we are interested in how or why the two variables are related. Suppose females are more likely than males to vote for a woman and that this difference cannot be explained away by age or by any other variable we have considered. We need to think about why there might be such a difference in the preferences of males and females. Perhaps females are more

**Table 3.2 -- Voting Preference by Sex Controlling for Age**

	Older			Younger		
	Male %	Female %	Total %	Male %	Female %	Total %
<b>Voting Preference</b>						
Willing to Vote for a Woman	32.9	33.9	33.3	65.8	66.9	66.7
Not Willing to Vote for a Woman	67.1	66.1	66.7	34.2	33.1	33.3
	100.0	100.0	100.0	100.0	100.0	100.0
	(240)	(180)	(420)	(120)	(360)	(480)

liberal than males, and liberals are more likely to say they would vote for a woman. So we introduce liberalism/conservatism as a control variable in our analysis. If females are more likely to support a woman because they are more liberal, then the difference between the preferences of men and women should disappear or be substantially reduced when liberalism/conservatism is held constant. This process is called **interpretation** because we are interpreting how one variable is related to another variable. Table 3.3 shows what we would expect to find if females supported the woman because they were more liberal. Notice that in both partial tables, the differences in the percentages between men and women has disappeared. (It is not necessary that it disappears entirely, but

only that it is substantially reduced in each of the partial tables.) The crucial difference between explanation and interpretation is that in interpretation the control variable is an intervening variable. The independent variable is causally prior to the control variable and the control variable is causally prior to the dependent variable. The control variable provides the link between the independent and dependent variables. It helps us interpret how the independent variable leads to the dependent variable.

**Table 3.3 -- Voting Preference by Sex Controlling for Liberalism/Conservatism**

	Conservative			Liberal		
	Male %	Female %	Total %	Male %	Female %	Total %
<b>Voting Preference</b>						
Willing to Vote for a Woman	32.9	33.9	33.3	65.8	66.9	66.7
Not Willing to Vote for a Woman	67.1	66.1	66.7	34.2	33.1	33.3
	100.0	100.0	100.0	100.0	100.0	100.0
	(240)	(180)	(420)	(120)	(360)	(480)

Finally, let's focus on the third of the situations outlined at the beginning of this section--whether the relationship is the same for different types of individuals. Perhaps the relationship between sex and voter preference varies with other characteristics of the individuals. Maybe among whites, females are more likely to prefer women candidates than the males are, but among African-Americans, there is little difference between males and females in terms of voter preference. This is the outcome shown in Table 3.4. This process is called **specification** because it specifies the conditions under which the relationship between sex and voter preference varies.

In the earlier section on bivariate analysis, we discussed the use of chi square. Remember that chi square is a test of independence used to determine if there is a relationship between two variables. Chi square is used in multivariate

analysis the same way it is in bivariate analysis. There will be a separate value of chi square for each partial table in the multivariate analysis. You should keep a number of warnings in mind. Chi square assumes that the expected frequencies for each cell are five or larger. As long as 80% of these expected frequencies are five or larger and no single expected frequency is very small, we don't have to worry. However, the expected frequencies often drop below five when the number of cases in a column or row gets too small. If this should occur, you will have to either **recode** (i.e., combine columns or rows) or eliminate a column or row from the table.

**Table 3.4 -- Voting Preference by Sex Controlling for Race**

	White			African American		
	Male %	Female %	Total %	Male %	Female %	Total %
<b>Voting Preference</b>						
Willing to Vote for a Woman	42.9	56.5	51.2	50.0	50.0	50.0
Not Willing to vote for a Woman	57.1	43.5	48.8	50.0	50.0	50.0
	100.00	100.00	100.00	100.00	100.00	100.00
	(310)	(490)	(800)	(50)	(50)	(100)

Another point to keep in mind is that chi square is affected by the number of cases in the table. With a lot of cases it is easy to reject the null hypothesis of no relationship. With a few cases, it can be quite hard to reject the null hypothesis. Also, consider the percentages within the table. Look for patterns. Do not rely on any single piece of information. Look at the whole picture.

We have concentrated on crosstabulation and chi square. There are other types of statistical analysis such as regression. When you have mastered these techniques, look at some other types of analysis.

- Earl Babbie, 2009, *The Practice of Social Research* (12<sup>th</sup> edition), Wadsworth.
- Kenneth R. Hoover and Todd Donovan, 2011, *The Elements of Social Scientific Thinking* (10<sup>th</sup> edition), Wadsworth.

### **Survey Research and Sampling**

- Earl Babbie, 2009, *The Practice of Social Research* (12<sup>th</sup> edition), Wadsworth.
- Arlene Fink, Linda Bourque, Eve P. Fieldler, Sabrina Mertens Oishi, Mark S. Litwin, 2003, *The Survey Kit* (2<sup>nd</sup> edition), Sage.
- Don A. Dillman, Jolene D. Smyth and Leah Melani Christian, 2009. *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method* (3<sup>rd</sup> edition), John Wiley.

### **Statistical Analysis**

- Herman J. Loether and Donald G. McTavish, 1993, *Descriptive and Inferential Statistics An Introduction* (4<sup>th</sup> edition), Allyn and Bacon.
- J. Richard Kendrick, Jr., 2004, *Social Statistics: An Introduction Using SPSS* (2<sup>nd</sup> edition), Mayfield.
- Marija J. Norusis, 2011, *IBM SPSS Statistics 19 Guide to Data Analysis*, Prentice Hall.

### **Elaboration and Causal Analysis**

- Hirschi, Travis and Hanan C. Selvin, 1967, *Delinquency Research--An Appraisal of Analytic Methods*, Free Press.
- Rosenberg, Morris, 1968, *The Logic of Survey Analysis*, Basic Books.