

## 5

## Making Controlled Comparisons



Watch screencasts of the guided examples in this chapter. [edge.sagepub.com/pollock](http://edge.sagepub.com/pollock)

**Procedures Covered**

Analyze ► Descriptive Statistics ► Crosstabs (with Layers)  
 Analyze ► Compare Means ► Means (with Layers)  
 Graphs ► Legacy Dialogs ► Line (Multiple)  
 Graphs ► Legacy Dialogs ► Bar (Clustered)

Political analysis often begins by making simple comparisons using cross-tabulation analysis or mean comparison analysis. Simple comparisons allow the researcher to examine the relationship between an independent variable, X, and a dependent variable, Y. However, there is always the possibility that alternative causes—rival explanations—are at work, affecting the observed relationship between X and Y. An alternative cause is symbolized by the letter Z. If the researcher does not control for Z, then he or she may misinterpret the relationship between X and Y.

### CROSS-TABULATION ANALYSIS WITH A CONTROL VARIABLE

To demonstrate how to use SPSS Crosstabs to produce control tables, we will work through an example with the GSS dataset, GSS.sav. Open the GSS dataset.

Consider this hypothesis: In a comparison of individuals, those who attend religious services less frequently will be more likely to favor the legalization of marijuana than will those who attend religious services more frequently. In this hypothesis, attend3, which categorizes respondents' church attendance as "Low," "Moderate," or "High," is the independent variable. The GSS variable grass records respondents' opinions on the legalization of marijuana. (Code 1 is "Legal," and code 2 is "Not legal.")

To stay acquainted with cross-tabulation analysis, we will start by looking at the uncontrolled relationship between attend3 and grass. In addition to considering whether the hypothesis has merit, we will note the tendency of the relationship and we will apply a non-statistical measure of the relationship's strength. By determining tendency and gauging strength, you are better able to interpret relationships involving control variables. To produce an uncontrolled, zero-order cross-tabulation, follow the steps detailed in Chapter 4. Recall that you start by selecting **Analyze ► Descriptive Statistics ► Crosstabs** (refer to Figure 4-1 as necessary).



Screencast

Cross-tabulation  
 Analysis with a Control  
 Variable, Additive  
 relationship

In the Crosstabs window, find the dependent variable, grass, in the left-hand variable list and click it into the Row(s) panel. Find the independent variable, attend3, and click it into the Column(s) panel. Click the Cells button and select the box next to “Column” in the Percentages panel. Click Continue, and then click OK. SPSS reports the results:

**Should Marijuana Be Made Legal \* How Often R Attends Religious Services, 3 Categories Crosstabulation**

		How Often R Attends Religious Services, 3 Categories				
		Low	Moderate	High	Total	
Should Marijuana Be Made Legal	LEGAL	Count	609	295	194	1098
		% within How Often R Attends Religious Services, 3 Categories	74.5%	60.1%	37.2%	60.0%
	NOT LEGAL	Count	208	196	327	731
		% within How Often R Attends Religious Services, 3 Categories	25.5%	39.9%	62.8%	40.0%
Total		Count	817	491	521	1829
		% within How Often R Attends Religious Services, 3 Categories	100.0%	100.0%	100.0%	100.0%

Clearly the hypothesis has merit. Of the low attenders, 74.5 percent favor legalization, compared with 60.1 percent of moderate attenders and 37.2 percent of the highly observant.<sup>1</sup> And note that given the way attend3 is coded—increasing values denote increasing church attendance—a negative relationship exists between religiosity and the percentage favoring legalization. As attendance increases, the percentage favoring legalization declines. (If you interpret the cross-tabulation by examining the “Not legal” row, then the tendency is positive. As attendance increases, the percentage opposing legalization increases.)

How strong is the relationship between church attendance and marijuana legalization opinions? You can arrive at a quick and easy measure of strength by figuring out the percentage-point change in the dependent variable across the full range of the independent variable. At one pole, 74.5 percent of low attenders favor legalization. At the other pole, 37.2 percent of frequent attenders are in favor. Therefore, the percentage favoring legalization drops by  $74.5 - 37.2 = 37.3$ , or about 37 percentage points. By this rudimentary measure, the relationship’s strength is 37 percentage points. (In Chapter 7, we consider statistical measures of strength.)

What other factors, besides church attendance, might account for differing opinions on marijuana legalization? Here’s a plausible answer: whether the respondent has children. Regardless of religiosity, people with children may be less inclined to endorse the legalization of marijuana than are people who do not have children. And here is an interesting (if complicating) fact: People who attend church regularly are substantially more likely to have children than are

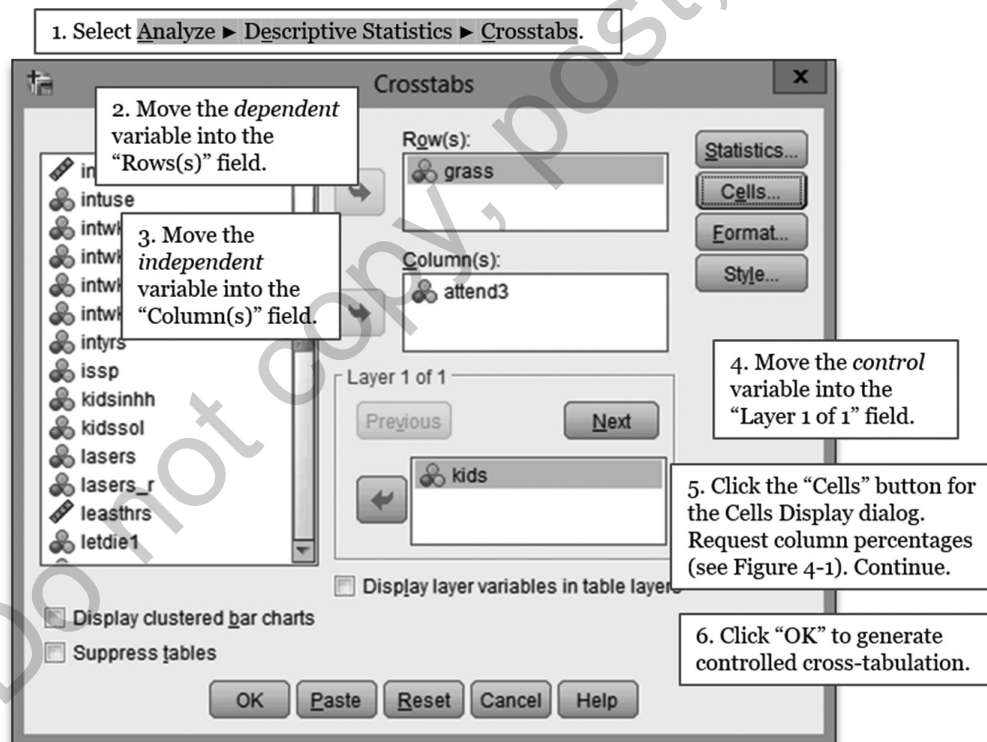
<sup>1</sup>If you get somewhat different numbers, don’t forget to weight observations using the wtss variable.

people who rarely or never attend.<sup>2</sup> Thus, when we compare the marijuana opinions of “High” and “Low” attenders, as we have just done, we are also comparing people who are more likely to have children (“High”) with people who are less likely to have children (“Low”). It could be that secular individuals are more inclined to favor legalization, not because they are less religious but because they are less likely to have children. By the same token, those who go to church more often might oppose legalization for reasons unrelated to their religiosity: They’re more likely to have children.

The only way to estimate the effect of church attendance on marijuana opinions, free from the confounding effect of having children, is to compare low attenders who do not have children with high attenders who do not have children, and to compare low attenders who have children with high attenders who have children. In other words, we need to control for the effect of having children by holding it constant. Crosstabs with layers will perform the controlled comparison we are after.

Let’s run the grass–attend3 analysis again, this time adding kids as a control variable. To produce a controlled cross-tabulation, again select **Analyze** ► **Descriptive Statistics** ► **Crosstabs**, returning to the Crosstabs window. You should find the dependent variable, grass, and the independent variable, attend3, in the Crosstabs window just where you left them. To obtain a controlled comparison—the relationship between grass and attend3, controlling for kids—scroll down the variable list until you find “kids” and move it into the box labeled “Layer 1 of 1,” as shown in Figure 5-1. The GSS variable kids classifies respondents into one of two categories:

**FIGURE 5-1** Generating a Crosstabs with a Control Variable (Layer)



<sup>2</sup>According to the GSS dataset, 80.1 percent of “High” attenders have children, compared with 65.1 percent of “Low” attenders—a 15-percentage-point difference.

those with children (coded 1 and labeled “Yes” on kids) or those without (coded 0 and labeled “No” on kids). Click OK.

SPSS will run a separate cross-tabulation analysis for each value of the variable that appears in the Layer box. And that is precisely what we want: a cross-tabulation of grass and attend3 for respondents without children and a separate analysis for those with children. SPSS returns its version of a control table:

**Should Marijuana Be Made Legal \* How Often R Attends Religious Services, 3 Categories \*  
Does R H ave Children? Crosstabulation**

Does R Have Children?				How Often R Attends Religious Services, 3 Categories			Total
				Low	Moderate	High	
No	Should Marijuana Be Made Legal	LEGAL	Count	257	98	44	399
			% within How Often R Attends Religious Services, 3 Categories	84.3%	72.1%	38.6%	71.9%
	NOT LEGAL	Count	48	38	70	156	
		% within How Often R Attends Religious Services, 3 Categories	15.7%	27.9%	61.4%	28.1%	
	Total	Count	305	136	114	555	
		% within How Often R Attends Religious Services, 3 Categories	100.0%	100.0%	100.0%	100.0%	
Yes	Should Marijuana Be Made Legal	LEGAL	Count	351	196	150	697
			% within How Often R Attends Religious Services, 3 Categories	68.7%	55.4%	37.0%	54.9%
	NOT LEGAL	Count	160	158	255	573	
		% within How Often R Attends Religious Services, 3 Categories	31.3%	44.6%	63.0%	45.1%	
	Total	Count	511	354	405	1270	
		% within How Often R Attends Religious Services, 3 Categories	100.0%	100.0%	100.0%	100.0%	
Total	Should Marijuana Be Made Legal	LEGAL	Count	608	294	194	1096
			% within How Often R Attends Religious Services, 3 Categories	74.5%	60.0%	37.4%	60.1%
	NOT LEGAL	Count	208	196	325	729	
		% within How Often R Attends Religious Services, 3 Categories	25.5%	40.0%	62.6%	39.9%	
	Total	Count	816	490	519	1825	
		% within How Often R Attends Religious Services, 3 Categories	100.0%	100.0%	100.0%	100.0%	

Crosstabs output with layers can be a bit confusing at first, so let's closely consider what SPSS has produced. There are three cross-tabulations, appearing as one table.<sup>3</sup> To the left-hand side of the table you will see the label of the control variable, kids: "Does R Have Children?" The first value of kids (0), labeled "No," appears beneath that heading. So the top cross-tabulation shows the grass–attend3 relationship for people who do not have children. The next cross-tabulation shows the relationship for respondents with children, respondents with the value "Yes" on the control variable. Finally, the bottom cross-tabulation, labeled "Total," shows the overall relationship between grass and attend3.

First, assess the tendency and strength of the relationship between church attendance and support for marijuana legalization among respondents who do not have children. Among people without children, the tendency is negative. As the values of attend3 increase from low to high, support for legalization declines: 84.3 percent of the low attenders favor legalization, compared with 72.1 percent of the middle group and 38.6 percent of the high attenders. How large is the drop? Across the full range of religious attendance, the percentage favoring legalization declines from 84.3 to 38.6—an "attendance effect" of about 46 percentage points!

Next, assess tendency and strength among respondents who have children, the middle section of the controlled cross-tabulation. Note that the tendency, once again, is negative: 68.7 percent of the low attenders favor legalization, compared with 55.4 percent of moderate attenders and 37.0 percent of high attenders. Note also that the strength of the relationship is also very strong for people who have children. Among respondents with kids, the percentage who favor legalization drops from 68.7 among low attenders to 37.0 among high attenders—an "attendance effect" of about 32 percentage points.

To help you make correct interpretations of controlled comparisons, it is a good idea to evaluate the relationship between the control variable and the dependent variable, within each value of the independent variable. In the current example, we would determine the tendency and strength of the relationship between the control variable, kids, and marijuana attitudes, controlling for church attendance. This is accomplished by jumping between the "No" cross-tabulation and the "Yes" cross-tabulation, comparing marijuana opinions of people who share the same level of attendance but who differ on the control variable, kids. Among low attenders, those without kids are more likely to favor legalization than those with kids. When the control variable switches from "No" to "Yes," the percentage of marijuana supporters drops from 84.3 percent to 68.7 percent—a "kid effect" of about 16 percentage points. How about moderate attenders? As with low attenders, the kid effect is negative: 72.1 percent compared with 55.4 percent—about 17 points. Among high attenders, the effect is in the same direction but not as strong: 38.6 percent versus 37.0 percent—a kid effect of about 2 percentage points.

How would you characterize this set of relationships? Does a spurious relationship exist between grass and attend3? Or are these additive relationships, with attend3 explaining legalization opinions and kids adding to the explanation? Or is interaction going on? If the grass–attend3 relationship were spurious, then the relationship would weaken or disappear after controlling for kids. Among respondents without children, low, moderate, and high attenders would all hold the same opinion about marijuana legalization. Ditto for people with children: Attendance would not play a role in explaining the dependent variable. Because the relationship persists after controlling for kids, we can rule out spuriousness.

Now, it is sometimes difficult to distinguish between additive relationships and interaction relationships. In additive relationships, the effect of the independent variable on the dependent variable is the same or quite similar for each value of the control variable. In interaction

<sup>3</sup>If you don't like the look of SPSS's default cross-tabulation, remember you can easily edit an SPSS table. See the section on "How to Format an SPSS Table" in Chapter 1.



Screencast

Cross-tabulation  
Analysis with a Control  
Variable, Interactive  
relationship

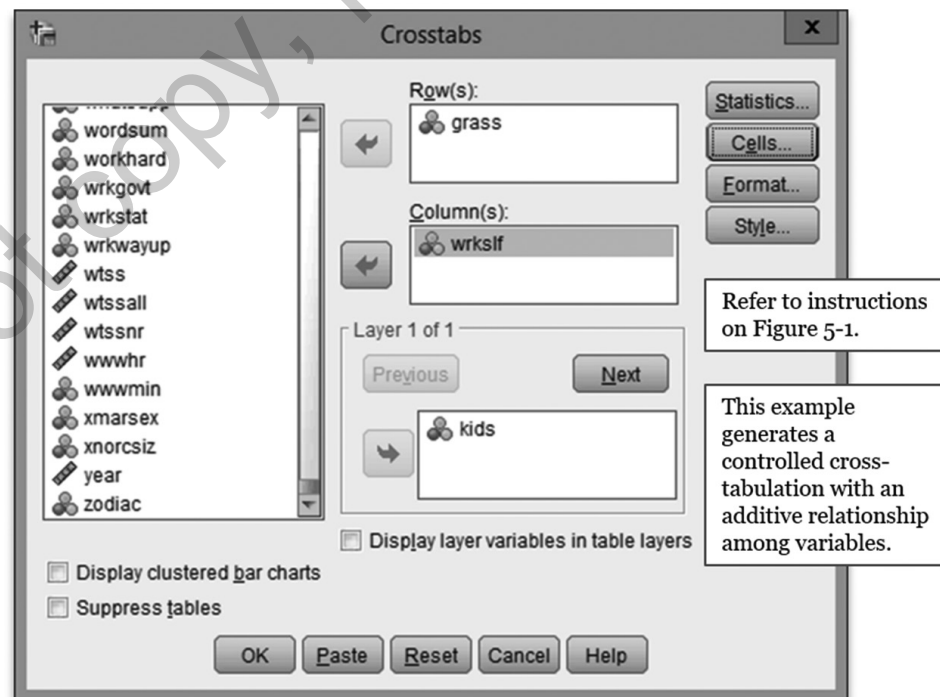
relationships, by contrast, the effect of the independent variable on the dependent variable varies in tendency or strength for different values of the control variable. In this example, we found that the “attendance effect” for non-parents was a whopping 46 percentage points, compared to a 32-percentage-point difference for parents. Similarly, the “kid effect” was about the same for low and moderate attenders (16 and 17 percentage points, respectively) but just 2 percentage points for those who attend church frequently. So, the grass–attend3 relationship is interactive with parent status.

Let’s do one more controlled cross-tabulation to see an example of an additive relationship. Rather than looking at the effect of church attendance on opinions about marijuana legalization, let’s consider individuals’ employment status. The wrkslf variable in the GSS dataset distinguishes between people who are self-employed and those who work for someone else. Those who are self-employed face fewer consequences from workplace drug testing and probably value individual choice more highly than those who work for others, which may increase their support for marijuana legalization.

To analyze the relationship between employment status and marijuana opinions, controlling for having children, again select **Analyze** ► **Descriptive Statistics** ► **Crosstabs**, returning to the Crosstabs window. You should find the variables grass, attend3, and kids in the Crosstabs window just where you left them. Move the old column variable, attend3, back to the variable list and replace it with wrkslf (Figure 5-2). Assuming you’ve requested column percentages in the cells, click the OK button.

Now, if someone were to ask, “What is the effect of employer status on marijuana opinions?” we would not be misrepresenting the results to reply, “People who work for themselves are about 8 percentage points more likely to favor legalization than those who work for someone else.” (This corresponds to the average of effects of employer status for parents, 9.4 percentage points, and non-parents, 7.4 percentage points.) If asked about the role of children, we would be well

**FIGURE 5-2** Controlled Cross-tabulation of an Additive Relationship



**Should Marijuana Be Made Legal \* R Self-Emp Or Works For Somebody \* Does R Have Children? Crosstabulation**

Does R Have Children?			R Self-Emp Or Works For Somebody		Total	
			SELF-EMPLOYED	SOMEONE ELSE		
No	Should Marijuana Be Made Legal	LEGAL	Count	30	342	372
			% within R Self-Emp Or Works For Somebody	78.9%	71.5%	72.1%
	NOT LEGAL	Count	8	136	144	
		% within R Self-Emp Or Works For Somebody	21.1%	28.5%	27.9%	
	Total	Count	38	478	516	
		% within R Self-Emp Or Works For Somebody	100.0%	100.0%	100.0%	
Yes	Should Marijuana Be Made Legal	LEGAL	Count	87	611	698
			% within R Self-Emp Or Works For Somebody	64.0%	54.6%	55.6%
	NOT LEGAL	Count	49	509	558	
		% within R Self-Emp Or Works For Somebody	36.0%	45.4%	44.4%	
	Total	Count	136	1120	1256	
		% within R Self-Emp Or Works For Somebody	100.0%	100.0%	100.0%	
Total	Should Marijuana Be Made Legal	LEGAL	Count	117	953	1070
			% within R Self-Emp Or Works For Somebody	67.2%	59.6%	60.4%
	NOT LEGAL	Count	57	645	702	
		% within R Self-Emp Or Works For Somebody	32.8%	40.4%	39.6%	
	Total	Count	174	1598	1772	
		% within R Self-Emp Or Works For Somebody	100.0%	100.0%	100.0%	

within the boundaries of the data to say, “People without kids are about 16 percentage points more likely to favor legalization than are people who have children.” (This corresponds to the average of effects of kids for those self-employed, 14.9 percentage points, and those who work for others, 16.9 percentage points.) All additive relationships share this straightforward simplicity: same tendency, same or similar strengths, at all values of the control variable.

## GRAPHING CONTROLLED COMPARISONS WITH CATEGORICAL DEPENDENT VARIABLES

In Chapter 4 you learned how to create bar charts and line charts to show the relationship between a categorical independent variable and a dependent variable. Building on these methods, you can

use SPSS to produce two types of graphs for controlled comparisons: clustered bar charts and multiple line charts.

For clarifying controlled comparisons, multiple line charts are preferred, particularly when the independent variable is measured at the ordinal level. Compared with bar charts, line charts are simpler and more elegant, and they have a more favorable data/ink ratio, defined as “the proportion of a graphic’s ink devoted to the non-redundant display of data-information.”<sup>4</sup> In other words, if one were to “add up” all the ink used in a graph, line charts tend to devote a larger proportion of the total ink to the essential communication of the data.<sup>5</sup>

In the following guided example, we produce a multiple line chart for the additive relationship analyzed at the beginning of the chapter, the relationship between grass and attend3 (an ordinal independent variable), controlling for kids. To create this multiple line chart, select **Graphs** ► **Legacy Dialogs** ► **Line**. When the Line Charts dialog opens, click Multiple and make sure that the Summaries for Groups of Cases radio button is selected (the default). Click Define. The Define Multiple Line: Summaries for Groups of Cases window appears (Figure 5-3).

What do we want the line chart to depict? We want to see the percentage of respondents who think marijuana should be legal (code 1 on grass) for each value of the independent variable (attend3). Furthermore, we want to see the grass–attend3 relationship separately for each value of the control variable, kids. In all SPSS charts, the values of the independent variable appear along the axis labeled “Category Axis.” Move the independent variable, attend3, into the Category Axis field, as shown in Figure 5-3. For this controlled comparison, we want to see the attend3–grass relationship for each value of the control variable, kids. In a multiple line chart, the values of the control variable “define” the lines. The variable named “kids” is the control variable, so click kids into the “Define Lines by” field.

Now we need to make sure that the lines will represent the percentages of respondents saying “legal.” In the Lines Represent panel, select the “Other statistic” radio button, which activates the Variable box. Find grass in the variable list and then move it into the Variable box. By default, SPSS will display the mean value of grass, “MEAN(grass),” which does not suit our purpose so we’ll click the Change Statistic button. In the Statistic window, click the “Percentage inside” radio button. Type “1” in the Low box and “1” in the High box. As in Figure 5-3, these instructions tell SPSS to display the percentage of respondents in one category of the dependent variable, the percentage coded 1 on grass. Click Continue, returning to the Define Multiple Line window. The Variable box should now read “PIN(1 1)(grass),” meaning “The lines will display the percentages of respondents inside the value of 1 on grass at the low end and the value of 1 on grass at the high end.” Once, you’ve defined the right multiple line chart settings, click OK.

The multiple line chart, constructed to our specifications, appears in the Viewer (Figure 5-4).

This multiple line chart greatly facilitates interpretation of the controlled comparison. The upper line shows the relationship between grass and attend3 for people without children, and the lower line depicts the relationship for people with children. Trace the effect of the independent variable by moving from left to right along each line, across the values of attend3. As we learned from the controlled cross-tabulation, the “attendance effect” is more pronounced for respondents without kids than it is for those with kids. Viewed another way, as the vertical distance between lines,

<sup>4</sup>Edward R. Tufte, *The Visual Display of Quantitative Information*, 2nd ed. (Cheshire, Conn.: Graphics Press, 2001), 93.

<sup>5</sup>The **Graphs** ► **Legacy Dialogs** ► **Bar (Clustered)** interface is comparable to the multiple line chart interface, except the control variable is used to define the category axis (rather than lines) and the independent variable is used to define clustered. The chart should separate observations according to values of the control variable and then facilitate controlled comparisons. If you prefer clustered bar charts, which may be more appropriate for nominal-level independent variables, you can directly apply the skills you will learn in this chapter.

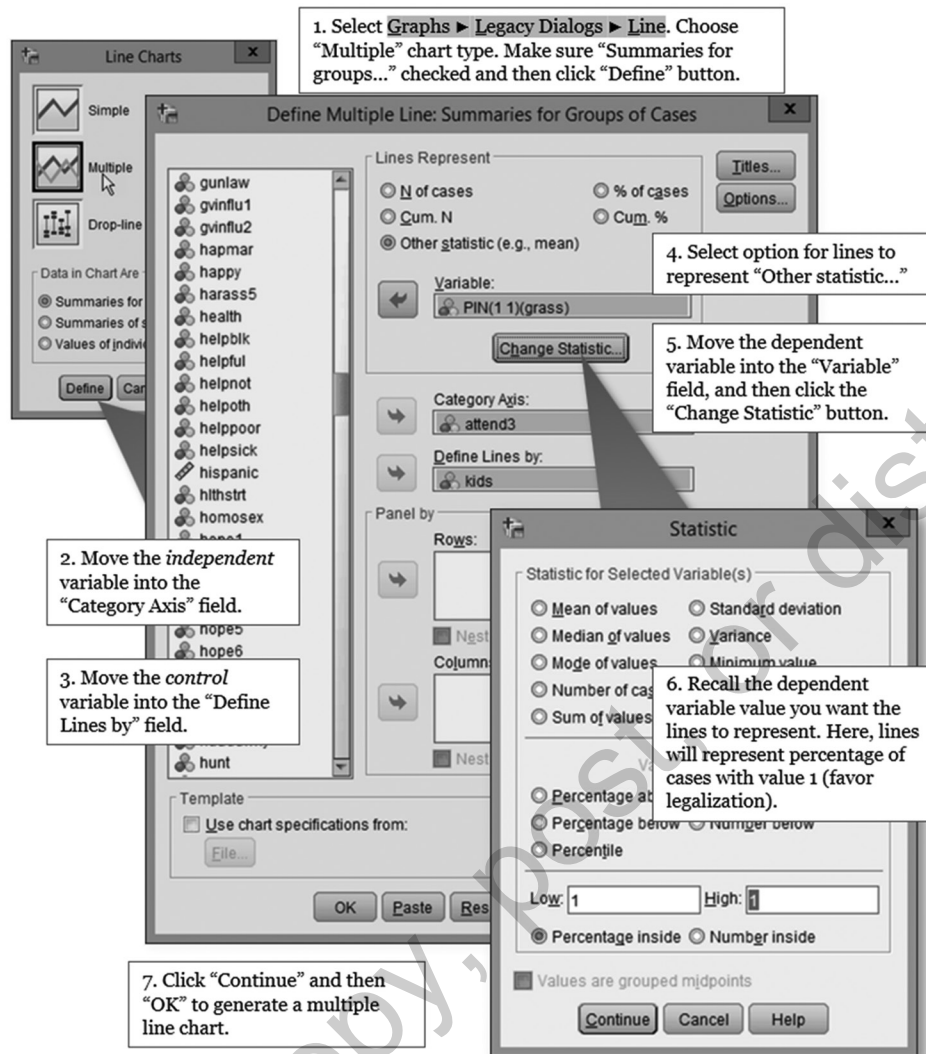


Screencast

Multiple Line Chart,  
Additive



FIGURE 5-3 Creating a Multiple Line Chart

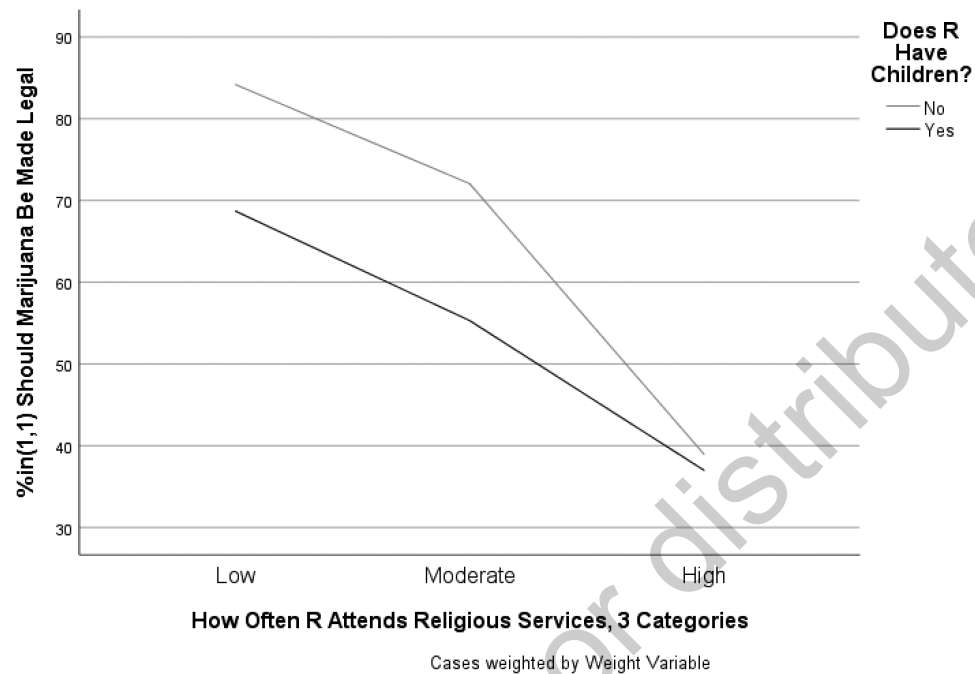


we see that the "kid effect" is quite similar for low and moderate attenders but is relatively minor among those with high church attendance. This is a very informative graphic. Before moving on, let's spruce up this multiple line chart using the Chart Editor.

We will make three changes to the chart: First, we will change the title on the *y*-axis. Second, we will make the lines thicker. Finally, we will change the style of one of the lines, so that the legend clearly communicates the categories of the control variable, kids. (If you print graphics in black and white, as we do in this book, it is sometimes difficult to distinguish subtle differences in the colors of the lines.)

Recall how to start SPSS's Chart Editor. Either double-click the chart in the Viewer or right-click it and select "Edit Content ► In Separate Window." You'll see the multiple line chart in the Chart Editor.

To change the title on the *y*-axis, single-click the axis title. SPSS will highlight the *y*-axis title. After the title is highlighted, single-click it again. (It's like a slow-motion double-click.) You'll see the original title displayed in a text box, ready for editing (see Figure 5-5). Replace the current title with

**FIGURE 5-4** Multiple Line Chart of an Interactive Relationship

this new title: “Percent Who Think Marijuana Should Be Legal.” (Clicking anywhere else on the chart returns the axis title to its proper position.)

Next, let’s make the lines thicker. Double-click on one of the lines, as shown in Figure 5-6. The Chart Editor selects both lines and opens the Properties window.

Select the Lines tab of the Chart Editor’s Properties dialog. In the Lines section of this panel, click the Weight drop-down and select a heavier weight such as 1.5 or 2. (If you’re using SPSS Version 25, you might want to make the lines lighter.) Click Apply. The Editor makes both lines thicker.

Finally, let’s change the style of one of the lines to make it easier to tell them apart. To edit the properties of a single line, single-click on the line. SPSS highlights both lines. Now single-click the line again (another slow-motion double-click). The Chart Editor will select/highlight only that line. Now double-click the one highlighted line to call up the Properties dialog again (see Figure 5-7).

To edit the line’s style, click the Style drop-down from the Lines tab and choose from among the many dashed patterns. When you click Apply, the Editor modifies the line’s style and makes a corresponding change in the legend. Close the Properties window and exit the Chart Editor. A newly edited multiple line chart appears in the Viewer (Figure 5-8).

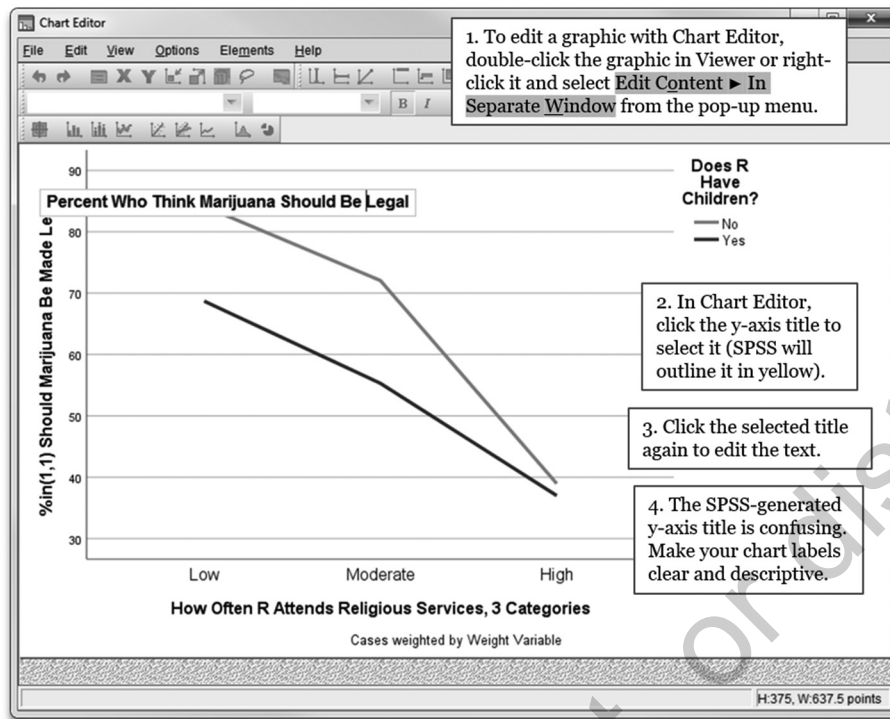
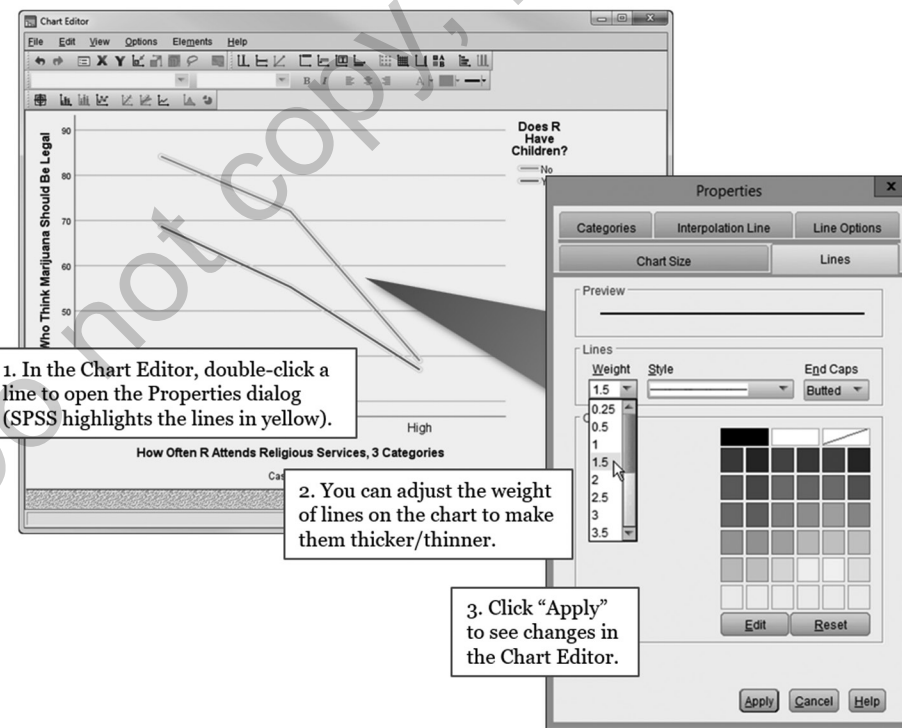
## MEAN COMPARISON ANALYSIS WITH A CONTROL VARIABLE

Mean comparison analysis is used when the dependent variable is interval level and the independent variable and the control variable are nominal or ordinal level. In most ways, the procedure for using Compare Means with layers to obtain controlled comparisons is similar to that for using Crosstabs. However, the two procedures differ in one important way. We will work through two guided examples using the NES dataset (NES.sav). The first example shows an interesting pattern of interaction. The second example gives you a chance to identify a set of additive relationships. Open the NES dataset and let’s begin the first guided example.



Screencast

Mean Comparison  
Analysis with a Control  
Variable

**FIGURE 5-5** Changing the y-Axis Title with the Chart Editor**FIGURE 5-6** Changing Line Weights with the Chart Editor

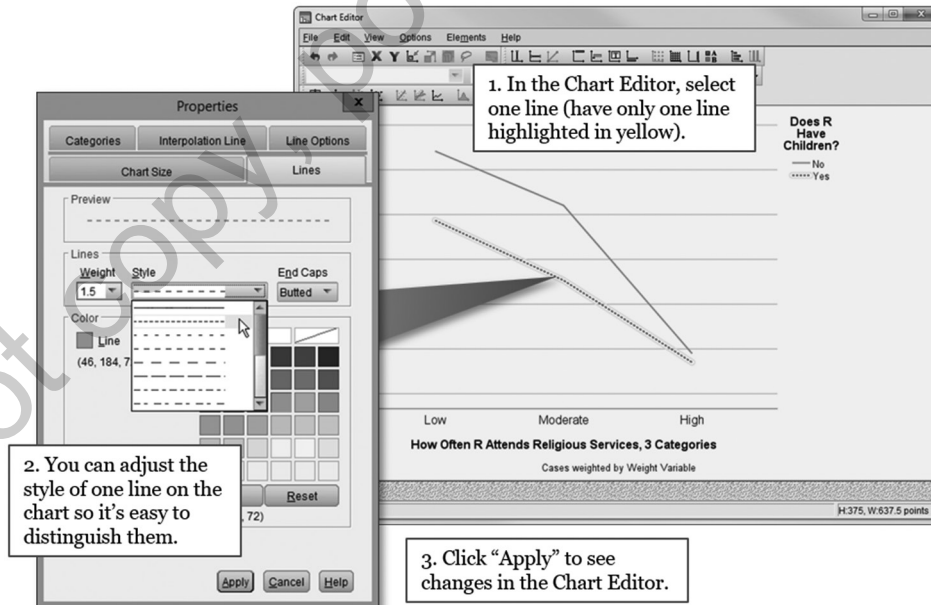
## Example of an Interaction Relationship

In Chapter 4, we looked at the relationship between party identification and feeling thermometer ratings of President Donald Trump. We found, not too surprisingly, that individuals who identify more strongly with the Republican Party give Trump higher mean scores than do those who identify less strongly with the Republican Party. In this chapter, we'll control for other variables that may explain the partisan difference in mean Trump ratings.

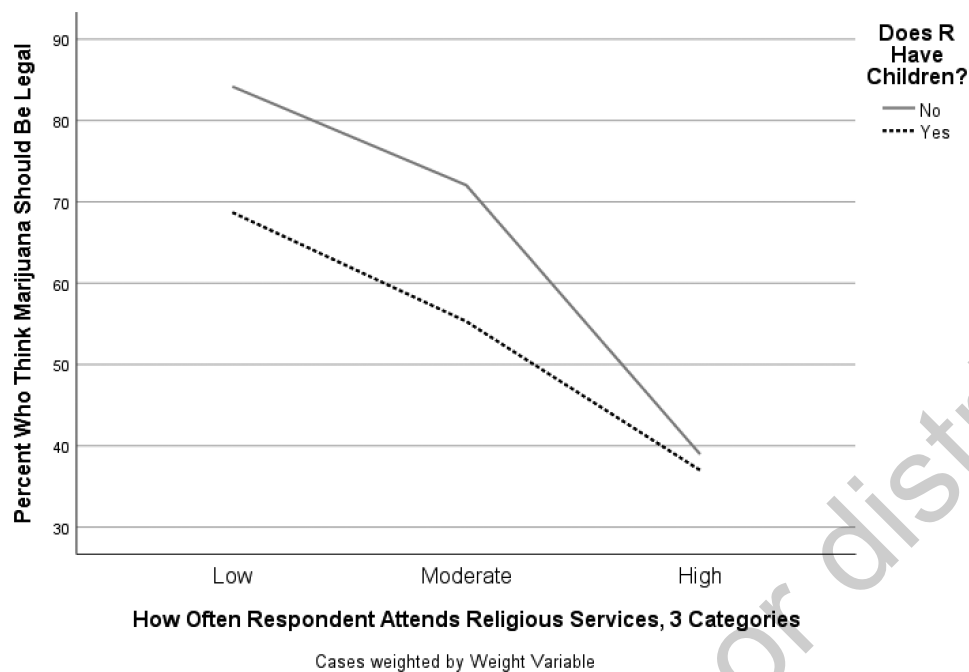
It has become an article of faith that African Americans are more strongly attracted to the Democratic Party than whites are. Indeed, on NES's Democratic Party feeling thermometer (ft\_Dem), African Americans' mean rating is over 30 degrees warmer than whites' mean rating: 74 degrees versus 43 degrees. It may be the case, then, that the partisan differences we observed initially are, to some extent, the result of racial differences in partisan identification. To compare party differences independent of racial differences, let's investigate the relationship between party identification and Trump ratings, controlling for race (NES variable Race2).

As you know, the NES dataset contains many feeling thermometer variables, which record respondents' ratings of different political groups and personalities on a scale from 0 (cold or negative) to 100 (warm or positive). The ft\_Trump\_pre variable, which gauges feelings toward President Donald Trump, will be the dependent variable in the current example. The independent variable is partyid7, a seven-category ordinal measure. For the control variable we will use Race2.<sup>6</sup> We will use **Analyze** ► **Compare Means** ► **Means** to produce mean values of ft\_Trump\_pre for each value of partyid7, controlling for race.

**FIGURE 5-7** Changing a Line Style with the Chart Editor



<sup>6</sup>The Race2 variable identifies white and black respondents. It makes for a simpler demonstration of controlled mean comparison analysis, but you are welcome to work with the Race3 variable, which also identifies Hispanics, if you prefer.

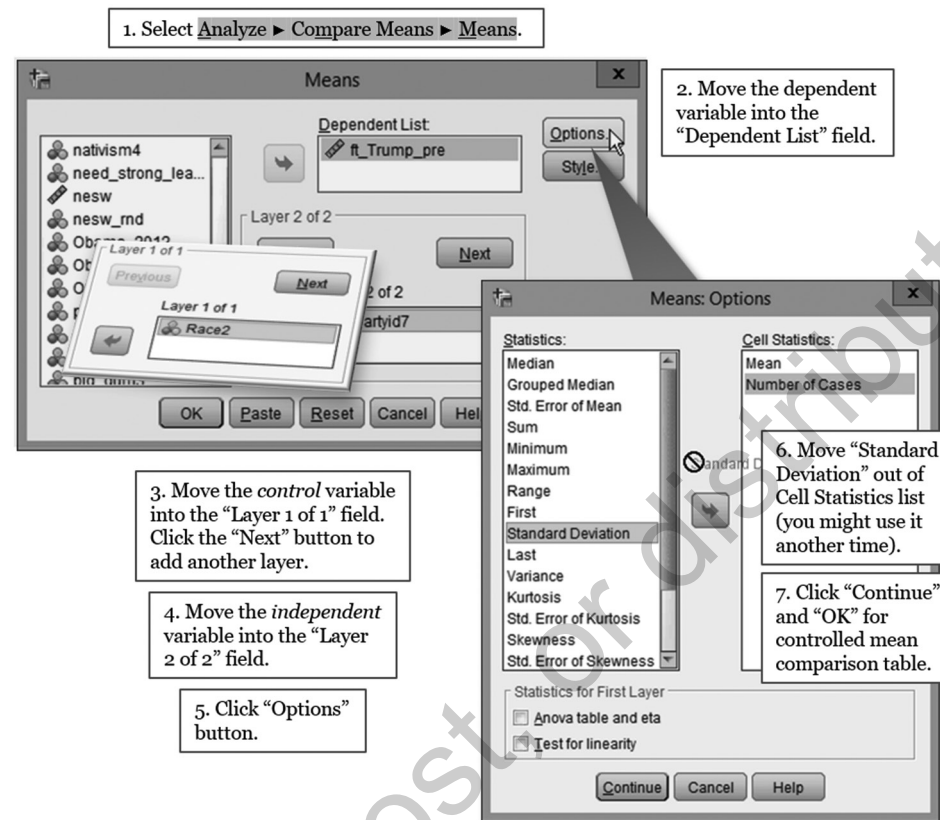
**FIGURE 5-8** Edited Multiple Line Chart of an Interactive Relationship

To conduct a controlled mean comparison analysis with SPSS, select **Analyze** ► **Compare Means** ► **Means**. Find `ft_Trump_pre` in the variable list and click it into the Dependent List box. Now we want SPSS to proceed as follows. First, we want it to separate respondents into two groups, whites and African Americans, based on the control variable, `Race2`. Second, we want SPSS to calculate mean values of `ft_Trump_pre` for each category of the independent variable, `partyid7`. SPSS handles mean comparisons by first separating cases on the variable named in the first Layer field. It then calculates means of the dependent variable for variables named in the subsequent Layer field. For this reason, it is best to put the control variable in the first layer and to put the independent variable in the second layer. Because `Race2` is the control variable, locate it in the variable list and move it into the “Layer 1 of 1” box (Figure 5-9). Click Next.

The next Layer field, labeled “Layer 2 of 2,” opens. The independent variable, `partyid7`, goes in this field. Move `partyid7` into the “Layer 2 of 2” field (Figure 5-9). One last thing: Click Options. In Cell Statistics, click Standard Deviation back into the left-hand Statistics box, and then click Continue. You are ready to go. Click OK.

The following controlled mean comparison table appears in the Viewer.

The values of the control variable, `Race2`, appear along the left-hand side of the table. The top-most set of mean comparisons shows the mean Trump thermometer ratings for white respondents, the next set is for black respondents, and the bottom set (labeled “Total”) shows the uncontrolled relationship between `ft_Trump_pre` and `partyid7`—for white and black respondents combined. This is a compact table, yet it contains a wealth of information. From the “Total” table, we can retrieve the overall relationship between party ID and the feeling thermometer: 8.16 for Strong Democrats all the way to 74.07 for Strong Republicans, about a 66-point swing. (The “Total” row of the “Total” table tells us the overall mean for the entire sample: 39.02 degrees.) Notice, too, that we have obtained the overall means for white (42.93) and black (13.84) respondents—a nearly 30-point divide.

**FIGURE 5-9** Procedure for Controlled Mean Comparison Analysis

Now, to evaluate the controlled effect of the independent variable, partyid7, on Trump ratings, we would compare the mean ratings of Trump at varying party identifications separated by race. What do these comparisons reveal? Consider white respondents. The mean Trump rating among white Strong Democrats is 8.83. This mean warms all the way to 75.02 among Strong Republicans. So, for white respondents, the thermometer gets 66 degrees warmer as we move across the party ID scale. Now shift your attention to African American respondents. The mean Trump rating among black Strong Democrats is 6.74. This mean warms up to 40.06 among Strong Republicans. So, for black respondents, the thermometer gets 33 degrees warmer as we move across the party ID scale (half the amount of warming observed among whites).

Does the partyid7–Trump thermometer relationship have the same tendency/direction at both values of the control? Yes, for both white and black respondents, those who identify more strongly with the Republican Party have more positive feelings about Trump than do those with weaker Republican identifications. Do the relationships have the same (or similar) strengths at both values of the control? No, the relationship between party ID and Trump ratings is stronger among whites than it is among blacks; the swing from one end of the party scale to the other is about twice as large among whites than it is among blacks. Again, a situation such as this—same tendency, different strengths—is one form of interaction.

Confirm the interaction interpretation by determining how the control variable, Race2, affects Trump ratings for each value of party identification. For Strong Democrats, for instance, there is a difference of about 2 degrees: 8.83 for white Strong Democrats compared with 6.74 for black Strong Democrats. Among Weak Democrats, the white-black difference is about 13 degrees, however; for Independent Democrats, 5 degrees; Independents, 18 degrees; Independent-Republicans,

**Report**

Feeling therm: Donald Trump (pre, v161087)

White/Black (v161310x)	Party ID (v161158x)	Mean	N
White	StrngDem	8.83	412
	WkDem	26.67	290
	IndDem	18.75	264
	Indep	39.82	321
	IndRep	59.22	344
	WkRep	55.11	360
	StrngRep	75.02	505
	Total	42.93	2496
	Black	StrngDem	6.74
WkDem		13.56	74
IndDem		16.88	41
Indep		22.20	47
IndRep		67.28	5
WkRep		40.74	11
StrngRep		40.06	14
Total		13.84	388
Total		StrngDem	8.16
	WkDem	24.00	364
	IndDem	18.50	306
	Indep	37.59	368
	IndRep	59.33	348
	WkRep	54.70	371
	StrngRep	74.07	519
	Total	39.02	2884

–8 degrees; Weak Republicans, 14 degrees; and Strong Republicans, 35 degrees.<sup>7</sup> Again, the relationship has the same tendency (whites tend to rate Trump higher than do blacks), but different strengths (the race difference is substantially larger among Republican identifiers than it is among other party identifications).

### Example of an Additive Relationship

Compared with the protean complexity of interaction, additive relationships are the soul of symmetry. In a set of additive relationships, both the independent and the control variables help to explain the dependent variable. More than this, the effect of the independent variable is the same or very similar—same tendency, same strength—for all values of the control variable. Interaction relationships assume several forms. Additive relationships assume only one.

To demonstrate, we'll stick with our familiar comparison of mean Trump ratings for varying party identifications. This time, however, we'll control for gun ownership, suspecting that gun

<sup>7</sup>The exceptional 67.28 mean Trump rating among African American Independent-Republicans should be read with caution, as it is based on only five respondents.

ownership is correlated with both Trump ratings and party identification. To compare party differences independent of differences in gun ownership, let's investigate the relationship between party identification and Trump ratings, controlling for gun ownership (NES variable `gun_own`). Start by selecting **Analyze** ► **Compare Means** ► **Means**.

If settings from our `ft_Trump_pre`–`partyid7`–`Race2` analysis are still in place, you can simply move `gun_own` into the “Layer 1 of 2” field (in place of `Race2`). If you're starting with a reset Means window, move `ft_Trump_pre` into the Dependent list. Click the control variable, `gun_own`, into the first Layer box. Click Next. Click the independent variable, `partyid7`, into the second Layer box. Click Options and remove the Standard Deviation from Cell Statistics. Click OK (refer to Figure 5-9, keeping in mind we're using `gun_own` in place of `Race2` in this analysis). Another information-rich table is at hand:

**Report**

Feeling therm: Donald Trump (pre, v161087)

Does R own a gun?  
(v161496)

	Party ID (v161158x)	Mean	N
No	StrngDem	6.55	564
	WkDem	20.02	358
	IndDem	16.91	287
	Indep	33.95	313
	IndRep	57.17	212
	WkRep	52.84	221
	StrngRep	71.67	275
	Total	31.32	2229
	Yes	StrngDem	10.71
WkDem		22.93	129
IndDem		21.71	102
Indep		37.18	159
IndRep		57.39	159
WkRep		54.95	193
StrngRep		73.47	249
Total		43.39	1178
Total		StrngDem	7.59
	WkDem	20.79	487
	IndDem	18.17	390
	Indep	35.04	472
	IndRep	57.26	371
	WkRep	53.82	413
	StrngRep	72.53	524
	Total	35.49	3407

Does gun ownership add to our explanation of varying Trump feeling thermometer scores? Yes. For both non-gun owners and gun owners, mean values of `ft_Trump_pre` ascend as we move from “Strong Democrat” to “Independent” to “Strong Republican.” Indeed, the magnitude of the end-to-end increase is virtually identical for both groups: 65 degrees for non-gun owners (71.67 – 6.55) and 63 degrees for gun owners (73.47 – 10.71). And notice the consistent effects of gun ownership.



At each value of the independent variable, gun owners rate Trump a few degrees warmer than do non-gun owners: 4 degrees separate the Strong Democrats, 3 degrees separate the Weak Democrats, and so forth. (The Independent-Republican category is a bit of a hiccup in the analysis.) Thus, regardless of party identification, the “gun ownership effect” is about 3 degrees. And, controlling for gun ownership, the “party effect” is about 64 degrees.

## VISUALIZING CONTROLLED MEAN COMPARISONS

There are a few different ways to graphically depict controlled mean comparisons. If the independent variable is measured at the ordinal level, a line chart is often the best choice. In this section, we will show how to make a line chart of mean Trump sentiment with an ordinal measure of party identification serving as the independent variable. If you make a controlled mean comparison with a nominal-level independent variable, connecting mean values with a sloping line is misleading because the order of  $x$ -axis values is arbitrary; in this situation, consider producing a clustered bar chart with bar height representing the mean dependent value in each category defined by the independent and control variables.<sup>8</sup>

A line chart can illuminate the interactive and additive relationships we identified in the previous section. The steps for obtaining a multiple line chart for an interval-level dependent variable are the same steps you learned earlier, with one labor-saving exception.

To create a line chart that compares mean Trump feeling thermometer ratings by party ID, controlling for race, complete the following steps:

1. Select **Graphs** ► **Legacy Dialogs** ► **Line**. This opens the Line Charts window.
2. In the Line Charts window, click **Multiple**, and then click **Define**. Keep the default summaries for groups of cases option selected. This opens the Define Multiple Line window.
3. Click the independent variable, `partyid7`, into the Category Axis box.
4. Click the control variable, `Race2`, into the “Define Lines by” box.
5. In the Lines Represent panel of the Define Multiple Line window, select the “Other statistic” radio button. This activates the Variable box.
6. Click the dependent variable, `ft_Trump_pre`, into the Variable box. SPSS moves `ft_Trump_pre` into the Variable box with its default designation, “MEAN(`ft_Trump_pre`).” For an interval-level dependent variable, this default is precisely what you want. (There is no need to change the statistic now.)
7. Click **OK**.

We now have a tailor-made line chart of the `partyid7`–`ft_Trump_pre` relationship, controlling for race (Figure 5-10).

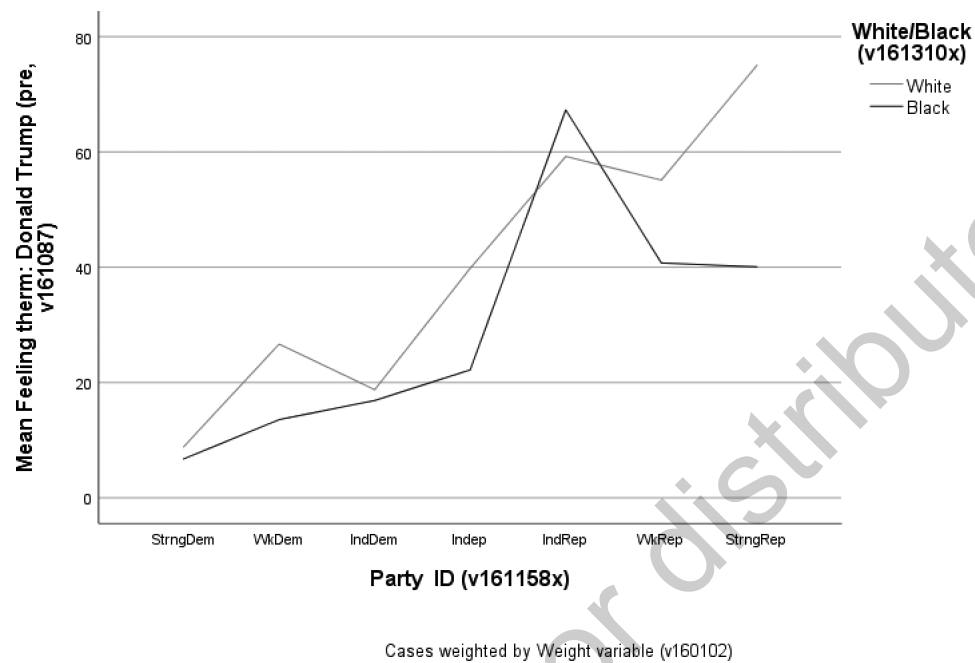
You can see why line charts are essential for correctly interpreting controlled comparisons. By tracing along each line, from Strong Democrat to Strong Republican, you can see the effect of party ID on Trump thermometer ratings. Among white respondents, the line slopes up sharply; there is a huge difference in mean Trump ratings from one end of the political spectrum to the other. Among black respondents, by contrast, the line rises less sharply. (The high mean score for African American

<sup>8</sup>For example, if we were comparing mean Trump sentiment between gun owners and non-gun owners (instead of party identification) controlling for race, we would suggest a clustered bar chart instead of a line chart. See Figure 7-7 for an example of a clustered bar chart that depicts a controlled comparison with a nominal-level independent variable. The bar height in that graph represents the proportion of cases with a specific dependent variable value, rather than the mean of the dependent variable in each subcategory, but the general idea is the same.



Screencast

Multiple Line Chart,  
Interaction

**FIGURE 5-10** Multiple Line Chart of Interaction Relationships

Independent-Republicans throws things off, but remember that is based on only five respondents.) Notice, too, that the relationship between race and the dependent variable varies from one side of the graph to the other: although the mean for whites is higher, the gap is wider on the right than it is on the left.

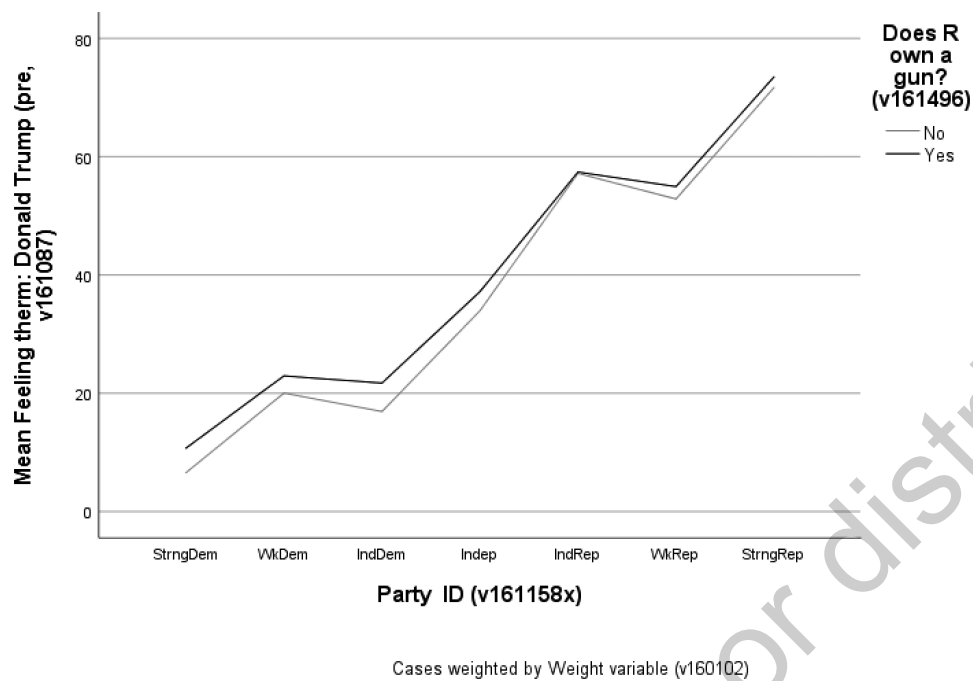
The relationship between the independent and dependent variables has the same tendency for both values of the control variable, but the relationship is stronger for one value than for the other value. Interaction can assume other forms, too. Interaction has two field marks, though, that will give it away. First, when you examine the relationship between the independent variable and the dependent variable at different values of the control variable, you may find that the relationship varies in tendency or direction, perhaps positive for one value of the control variable or zero or negative for other control values. Second, the relationship may have the same tendency for all control values but differ in strength, such as negative-weak versus negative-strong or positive-weak versus positive-strong. In identifying interaction, practice makes perfect. And, believe it or not, statistics can help (see Chapter 9).

Let's create a multiple line chart of the partyid7-ft\_Trump\_pre relationship, controlling for gun ownership. By now, creating this graph should be a straightforward exercise. Follow the seven steps outlined above, but specify a different control variable in step 4:

- 1 – 3. Same as above.
4. Click the control variable, gun\_own, into the “Define Lines by” box.
- 5 – 7. Same as above.

You can see how this line chart (Figure 5-11) communicates the additive relationship.

Moving from left to right, from strong identification with the Democratic Party to strong identification with the Republican Party, each line rises by about 64 units. That's the effect of the independent variable in this analysis, partyid7. The effect of gun ownership is conveyed by

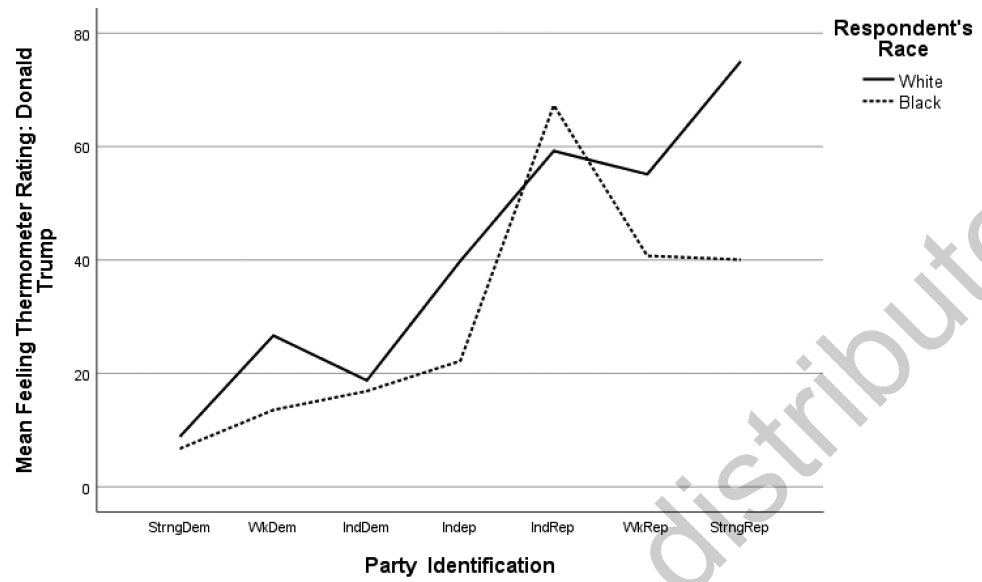
**FIGURE 5-11** Multiple Line Chart of Additive Relationships with Positive Tendency

the vertical distance between the two lines. Despite a narrowing in the Independent-Republican category, the difference between gun owners and non-gun owners is quite consistent.

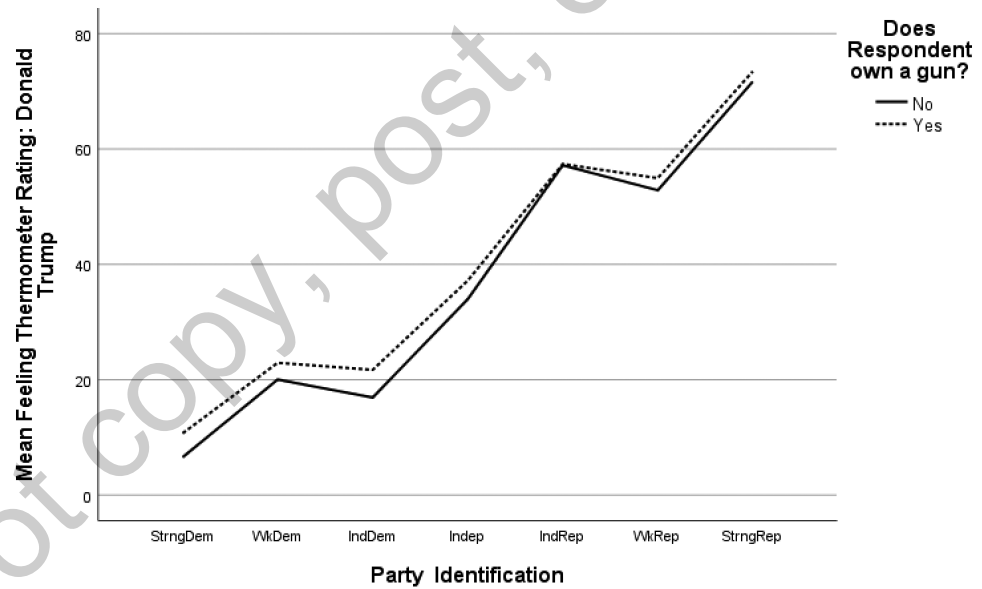
Now, you might encounter additive relationships in which the lines slope downward, imparting a negative relationship between the independent and dependent variables. And the lines might “float” closer together, suggesting a consistent but weaker effect of the control variable on the dependent variable, controlling for the independent variable. But you will always see symmetry in the relationships. The effect of the independent variable on the dependent variable will be the same or very similar for all values of the control variable, and the effect of the control variable on the dependent variable will be the same or very similar for all values of the independent variable.

The multiple line charts produced from default settings (like Figures 5-10 and 5-11) leave plenty of room for improvement. Invoke the Chart Editor and make some improvements to this multiple line chart. Experiment with the Chart Editor. Ask the Properties window to do new things. Figure 5-12 may serve as an example. (We’ll demonstrate more chart editing techniques in Chapter 8.)

**FIGURE 5-12** Multiple Line Charts Edited with the Chart Editor



Cases weighted by Weight variable (nesw)



Cases weighted by Weight variable (nesw)

Name: _____	Date: _____
E-mail: _____	Section: _____

## CHAPTER 5 EXERCISES

- (Dataset: NES. Variables: polknow2, political\_trust, ft\_Tea, nesw.) Given the Tea Party movement’s deep skepticism of government activism, it seems plausible to hypothesize that individuals who distrust the government would have warmer feelings toward the Tea Party than would those who trust the government to do what’s right. Of course, persons would need to be reasonably well informed about politics to make the connection between their assessment of government and their evaluation of the Tea Party. When we control for political knowledge (control variable), we may find that the relationship between Tea Party ratings (dependent variable) and trust of government (independent variable) gets stronger as knowledge increases. In other words, interaction could be occurring in this set of relationships. Consider two propositions and an ancillary hypothesis.

*Proposition 1:* At all levels of political knowledge (NES variable, polknow2), individuals who distrust the government (political\_trust) will give the Tea Party higher ratings (ft\_Tea) than will people who trust the government.

*Proposition 2:* The relationship between political trust and Tea Party ratings will be weaker for lower-knowledge respondents than for those with higher knowledge.

*Ancillary hypothesis:* In a comparison of individuals, those with higher levels of political knowledge are more likely to trust the government than are those with lower levels of political knowledge.

The dependent variable is the Tea Party feeling thermometer (ft\_Tea), which runs from 0 (cold or negative feelings) to 100 (warm or positive feelings). The independent variable is political\_trust, which captures how often respondents trust in government to do what’s right, with five ordinal values—never, some, about half, most, or always.

- Use the **Analyze** ► **Compare Means** ► **Means** procedure (with layers) to do a controlled mean comparison analysis of ft\_Tea for each combination of political\_trust and polknow2. Record the means next to the question marks in the following table. Use weights so your results are nationally representative.

How often can you trust federal government to do what’s right?	Political Knowledge		Total
	Low	High	
Never	?	?	?
Some	?	?	?
About half	?	?	?
Most	?	?	?
Always	?	?	?
Total	?	?	?

- Create a presentation-quality multiple line chart of the relationship between the Tea Party thermometer and political\_trust, controlling for polknow2. There should be one line for high-knowledge respondents and another for low-knowledge respondents. Print the chart.

- Consider the numeric table and the graph. Do the results support Proposition 1? Answer yes or no, and explain.

---

---

---

---

---

---

- Do the results support Proposition 2? Answer yes or no, and explain.

---

---

---

---

---

---

- Test the ancillary hypothesis by producing a cross-tabulation with political\_trust as the dependent variable and polknow2 as the independent variable.

Obtain column percentages. Do the results support the hypothesis? Answer yes or no, and explain, making specific reference to the cross-tabulation percentages.

---



---



---



---

2. (Dataset: World. Variables: women13, womyear2, pr\_sys.) In the Chapter 2 exercises, you analyzed the distribution of the variable women13, the percentage of women in the lower house of the legislatures in a number of countries. In this exercise you will analyze the relationship between women13 and two variables that could have an impact on the number of women serving in national legislatures.

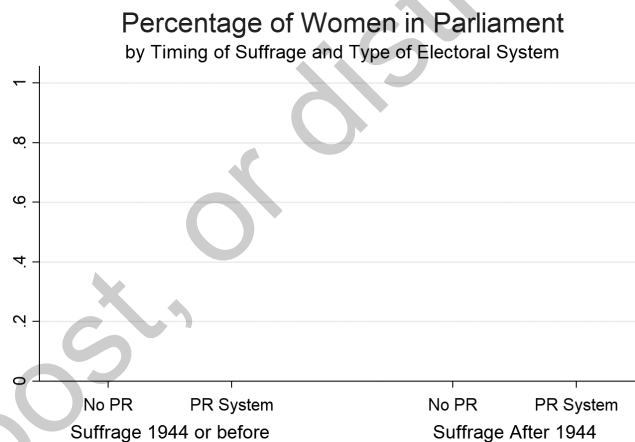
First, consider the role of the type of electoral system. Many democracies have proportional representation (PR) systems. PR systems foster multiple parties having diverse ideological positions—and, perhaps, having diverse demographic compositions as well. Non-PR systems, like the system used in U.S. elections, militate in favor of fewer and more homogeneous parties. Thus, you might expect that non-PR countries will have fewer women in their national legislatures than will countries with PR-based electoral systems.

Next, consider the role of history and tradition. In some countries, women have had a long history of political empowerment. New Zealand, for example, gave women the right to vote in 1893. In other countries, such as Switzerland (where women were not enfranchised until 1971), women have had less experience in the electoral arena. Thus, it seems reasonable to hypothesize that countries with longer histories of women’s suffrage (say, that enfranchised women before 1944) will have higher percentages of women in their national legislatures than will countries in which women’s suffrage is a more recent development (since 1944). In this exercise you will isolate the effect of the type of electoral system on the percentage of women in parliament, controlling for the timing of women’s suffrage. However, before you run any analyses, you will graphically depict different possible scenarios for the relationships you might discover.

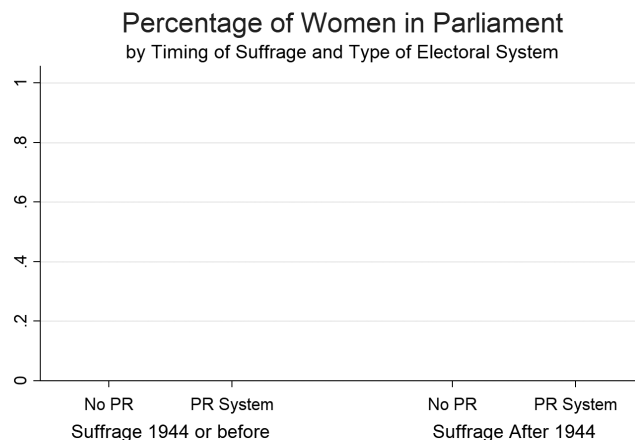
Parts A and B contain graphic shells showing the percentage of women in parliament along the vertical axis and the type of electoral system along the horizontal axis. Countries without PR systems are represented by “No,” and countries with PR systems by “Yes.” For each shell, you will draw four bars within the graphic space: a bar for “1944 or before” countries without PR systems (“No”),

a bar for “1944 or before” countries with PR systems (“Yes”), a bar for “After 1944” countries without PR systems (“No”), and a bar for “After 1944” countries with PR systems (“Yes”).

- A. Draw an *additive* relationship fitting this description: Countries with PR systems have higher percentages of women in parliament than do countries with non-PR systems, and countries with a longer history of women’s suffrage have higher percentages of women in parliament than do countries with a shorter history of women’s suffrage. (*Hint: In additive relationships, the strength and tendency of the relationship is the same or very similar for all values of the control variable.*)



- B. Draw a set of *interaction* relationships fitting this description: For countries with a longer history of women’s suffrage, those with PR systems have higher percentages of women in parliament than do countries with non-PR systems. For countries with a shorter history of women’s suffrage, the type of electoral system has no effect on the percentage of women in parliament.



In addition to the dependent variable, `women13`, the World dataset contains `pr_sys`, labeled “No” and coded 0 for countries with non-PR systems and labeled “Yes” and coded 1 for those having PR systems. Use `pr_sys` as the independent variable. World also contains `womyear2`, which measures the timing of women’s suffrage by two values: “1944 or before” (coded 0) and “After 1944” (coded 1). Use `womyear2` as the control variable.

- C. Use the **Analyze** ► **Compare Means** ► **Means** procedure (with layers) to produce a controlled mean comparison table showing mean values of `women13` for each value of `pr_sys`, controlling for `womyear2`. In the table below, record the mean values of `women13` next to each question mark.

PR system?	Women’s Suffrage		Total
	1944 or before	After 1944	
No	?	?	?
Yes	?	?	?
Total	?	?	?

- D. Produce a clustered bar chart of the relationship between `women13` and `pr_sys`, controlling for `womyear2`. Make desired changes to enhance appearance and readability. Print the chart you created.
- E. Examine the numeric data (part C) and the chart (part D). Consider the `women13–pr_sys` relationship for countries that enfranchised women in 1944 or before. Examine the difference between the means for non-PR countries and PR countries. This difference shows that the mean for PR countries is \_\_\_\_\_ points (circle one)

lower than    higher than  
the mean for non-PR countries.

Now consider the `women13–pr_sys` relationship for countries that enfranchised women after 1944. Examine the difference between the means for non-PR countries and PR countries. This difference shows that the mean for PR countries is \_\_\_\_\_ points (circle one)

lower than    higher than  
the mean for non-PR countries.

- F. Which of the following best characterizes the `women13–pr_sys` relationship, controlling for `womyear2`? (check one)
- The `women13–pr_sys` relationships have the same tendency and very similar strengths at both values of `womyear2`.

- The `women13–pr_sys` relationships have the same tendency but very different strengths at each value of `womyear2`.
- The `women13–pr_sys` relationships have different tendencies at each value of `womyear2`.

- G. Review your artistic work in parts A and B. Examine the table (part C) and the chart (part D) as well as your conclusions in parts E and F. Which possible scenario—the bar chart you drew in part A or B—resembles more closely the pattern shown in the data? (circle one)

The chart in part A    The chart in part B

3. (Dataset: NES. Variables: `ft_Rep`, `region`, `Race2`, `nesw`.) Two political analysts are discussing reputed partisan differences between southern and non-southern states.

*Political Analyst 1:* “Media pundits and confused academics tend to exaggerate the South’s reputation as a stronghold of Republican sentiment. In fact, people who live outside the South and people who live in the South don’t differ that much in their ratings of the Republican Party. Look at my latest SPSS analysis. I found this: Non-southerners rated the Republicans at 40.85 on the feeling thermometer; this compared with a slightly warmer 46.05 for southerners. That’s a paltry 5-point difference on the 100-point scale!”

*Political Analyst 2:* “Hmmm. . . that’s interesting. But did you control for race? I wonder what happens to the relationship between region and Republican ratings after you take race into account. After all, blacks are less strongly attracted to the Republicans than are whites. And since southern states have a higher proportion of blacks than do non-southern states, racial differences in Republican ratings could affect regional differences in Republican ratings.”

- A. According to Political Analyst 2, why did Political Analyst 1’s analysis find a relatively small difference between the Republican ratings of non-southerners and the Republican ratings of southerners? (check two)
- Because southern respondents are more likely to be white than are non-southern respondents
- Because southern respondents are more likely to be black than are non-southern respondents
- Because blacks give the Republican Party lower ratings than do whites
- Because blacks give the Republican Party higher ratings than do whites
- B. Create a dummy variable to identify NES respondents from the South. You can use the **Transform** ► **Recode into Different Variables** procedure to generate a

new dummy variable based on the existing region variable. Check your work by running the **Analyze ► Descriptive Statistics ► Frequencies** procedure on your new dummy variable. According to this analysis, \_\_\_ percent of respondents are from the South and \_\_\_ percent of respondents are not from the South. (Make sure you weight observations and your two entries sum to 100.)

- C. Obtain a breakdown table showing mean thermometer ratings of the Republican Party (ft\_Rep) by non-southern/southern region of residence (the dummy variable you created in part B), controlling for race (Race2). The independent variable is the South region dummy variable. Use Race2 as the control variable: “White” is coded 1, and “Black” is coded 2. Fill in the mean values (Mean) and weighted frequencies (rounded to two decimal places) next to each question mark in the table below.

Means and frequencies of feeling thermometer: Republican Party	Race of Respondent		Total <sup>9</sup>
	White	Black	
Non-South mean	?	?	41.68
Non-South weighted frequency	1,627	159	1,786
South mean	?	?	46.91
South weighted frequency	853	236	1,089
Total mean	?	?	43.66
Total weighted frequency	2,480	395	2,874.46

- D. Now figure out if the group of southern respondents has proportionately more blacks than the group of non-southern respondents. According to the table in part C, there are about 1,089 southern respondents in the data set. About what percentage of these 1,089 southern respondents are black? (circle one)

about 20 percent    about 30 percent  
about 40 percent

About what percentage of the 1,786 non-southern respondents are black? (circle one)

about 10 percent    about 20 percent  
about 30 percent

<sup>9</sup>The totals reported here are a little different than those stated by Political Analyst 1 because not all respondents have values white or black on the Race2 variable.

- E. Evaluate the relationship between ft\_Rep and the South/Non-South, controlling for race. When you compare southern whites with non-southern whites, you find that southern whites are \_\_\_\_\_ points (circle one)

cooler                      warmer

toward the Republican Party than are non-southern whites.

When you compare southern blacks with non-southern blacks, you find that southern blacks are \_\_\_\_\_ points (circle one)

cooler                      warmer

toward the Republican Party than are non-southern blacks.

- F. Use the **Graphs ► Legacy Dialogs ► Bar (Clustered)** procedure to obtain a presentation-quality clustered bar chart of the relationship between ft\_Rep and south, controlling for Race2. The category axis for this graphic should be the dummy variable you created to identify southern respondents and the clusters should be defined by Race2. Print the chart.

- G. Consider all of the evidence you have obtained in this exercise. How would you characterize this set of relationships? (circle one)

spurious                  additive                  interaction

Explain your reasoning:

---



---



---



---



---

4. (Dataset: GSS. Variables: fepol, natchld, natenvir, intsex, sex, wtss.) For an exercise in Chapter 4, you tested for the presence of *preference falsification*, the tendency for respondents to offer false opinions that they nonetheless believe to be socially desirable under the circumstances. You evaluated the hypothesis that respondents are more likely to express support for feminist viewpoints when questioned by a female interviewer than when questioned by a male interviewer. But you did not control for the respondent’s gender. That is, you did not look to see whether men are more (or less) likely than women to misrepresent their support for feminist viewpoints, depending on the gender of the interviewer.



Furthermore, it may be that men, and perhaps women as well, will engage in the same preference-falsifying behavior for policies that do not explicitly reference race but that may *symbolize* gendered issues, such as “government spending for child care.” Although “child care” does not mention “women,” it may be that respondents see “child care” through a gendered lens and will respond *as if* the question refers to a gendered issue. Of course, some policies, such as “government spending for improving and protecting the environment,” do not evoke such symbolic connections. Questions about these gender-neutral policies should not show the same gender-of-interviewer effects as questions that make explicit—or implicit—reference to gender.

In this exercise you will extend your Chapter 4 analysis in two ways. First, you will analyze the relationship between interviewer gender (intsex, the independent variable) and three dependent variables: opinions on an explicitly gendered issue (fepol, which asks if females are suited for politics), a symbolically gendered policy (natchld, opinions on spending for child care), and a gender-neutral policy (natenvir, spending for improving and protecting the environment). Second, you will perform these analyses while controlling for respondents’ gender (sex).

Based on previous research in this area, what might you expect to find? Here are two plausible expectations:

*Expectation 1:* For both male and female respondents, the gender-of-interviewer effect will be strongest for the explicitly gendered issue (fepol), weaker for the symbolically gendered policy (natchld), and weakest for the gender-neutral policy (natenvir).

*Expectation 2:* For the explicitly gendered issue (fepol) and for the symbolically gendered policy (natchld), the gender-of-interviewer effect will be greater for male respondents than for female respondents. For the gender-neutral policy (natenvir), the gender-of-interviewer effect will be the same (or close to 0) for both male respondents and female respondents (see Expectation 1).

A. Use the **Analyze** ► **Descriptive Statistics** ► **Crosstabs** procedure (with layers) to conduct the appropriate controlled cross-tabulation analyses. You will need to execute this procedure at least three times (one time per dependent variable). In the table that follows, record the *percentages* of respondents who express the following liberal opinions: disagree with statement women unsuited for politics (fepol), believe that national government spends too little on childcare (natfare), and think that national government spends too little on improving and protecting the environment (natenvir).

	Gender of Respondent				Total
	Male		Female		
	Male	Female	Male	Female	
Disagree with statement women unsuited for politics	?	?	?	?	80.97%
Government spends too little on childcare	?	?	?	?	57.66%
Government spends too little on improving and protecting the environment	?	?	?	?	63.28%

B. Examine the data closely. Among male respondents, would you say that Expectation 1 is or is not supported by the evidence? (circle one)

Expectation 1 is not supported.  
 Expectation 1 is supported.

Explain your reasoning.

---



---



---



---



---

Among female respondents, would you say that Expectation 1 is or is not supported by the evidence? (circle one)

Expectation 1 is not supported.  
 Expectation 1 is supported.

Explain your reasoning.

---



---



---



---



---

- C. Now compare the gender-of-interviewer effects between respondents of different genders. That is, compare the gender-of-interviewer effect on fepol among male respondents with the gender-of-interviewer effect on fepol among female respondents. Do the same for natchld and natenvir. Generally speaking, would you say that Expectation 2 is supported or is not supported by the evidence? (circle one)

Expectation 2 is not supported.

Expectation 2 is supported.

Explain your reasoning.

---



---



---



---



---

- D. Use the **Graphs** ► **Legacy Dialogs** ► **Bar (Clustered)** procedure to obtain a presentation-quality clustered bar chart of the relationship between natchld and intsex, controlling for sex. The category axis for this graphic should be the respondents' gender (sex) and the clusters should be defined by the interviewers' gender (intsex). Print the chart.

5. (Dataset: World. Variables: world.dta, democ, gdpcao8\_2, frac\_eth3.) Some countries have democratic regimes, and other countries do not. What factors help to explain this difference? One idea is that the type of government is shaped by the ethnic and religious diversity in a country's population. Countries that are relatively homogeneous, with most people sharing the same language and religious beliefs, are more likely to develop democratic systems than are countries having more linguistic conflicts and religious differences. Consider the ethnic heterogeneity hypothesis: In a comparison of countries, countries with lower levels of ethnic heterogeneity will be more likely to be democracies than will countries with higher levels of ethnic heterogeneity.

- A. According to the ethnic heterogeneity hypothesis, if you were to compare countries having lower heterogeneity with countries having higher heterogeneity, you should find (check one)
- a lower percentage of democracies among countries having lower heterogeneity.
  - a higher percentage of democracies among countries having lower heterogeneity.

- no difference between the percentage of democracies among countries having lower heterogeneity and the percentage of democracies among countries with higher heterogeneity.

- B. The World dataset contains the variable democ ("Democracy?"), which classifies each country as a democracy ("Yes," coded 2) or a dictatorship ("No," coded 1). The variable democ is the dependent variable in this analysis. The dataset also contains frac\_eth3, which classifies countries according to their level of ethnic heterogeneity: "Low" (coded 1), "Medium" (coded 2), or "High" (coded 3). This is the independent variable. Use the **Analyze** ► **Descriptive Statistics** ► **Crosstabs** procedure to test the ethnic heterogeneity hypothesis. Fill in the percentages of democracies in the following table.

	Ethnic Heterogeneity			Total
	Low	Medium	High	
Democracy	?	?	?	59.3%
Not democracy	?	?	?	40.7%
Total	?	?	?	100.0%

- C. Based on these results, you could say that (check one)
- as ethnic heterogeneity increases, the percentage of democracies increases.
  - as ethnic heterogeneity increases, the percentage of democracies decreases.
  - as ethnic heterogeneity increases, there is little change in the percentage of democracies.
- D. A country's level of economic development also might be linked to its type of government. According to this perspective, countries with higher levels of economic development will be more likely to be democracies than will countries with lower levels. The World dataset contains the variable gdpcao8\_2. This variable, based on gross domestic product per capita, is an indicator of economic development. Countries are classified as "Low" (coded 0) or "High" (coded 1). Use the **Analyze** ► **Descriptive Statistics** ► **Crosstabs** procedure (with layers) to obtain a cross-tabulation analysis of the democ-frac\_eth3 relationship, controlling for gdpcao8\_2. Fill in the percentages of democracies in the table.

	Economic Development in Country						Total
	Low GDP per capita			High GDP per capita			
	Ethnic Heterogeneity			Ethnic Heterogeneity			
	Low	Medium	High	Low	Medium	High	
Democracy	?	?	?	?	?	?	59.33%
Not democracy	?	?	?	?	?	?	40.67%
Total	?	?	?	?	?	?	100.00%

E. Examine the relationship between ethnic heterogeneity and democracy in high-GDP countries and low-GDP countries.

Consider the democ-frac\_eth3 relationship for low-GDP countries. Are ethnic heterogeneity and democracy related? Answer yes or no and briefly explain.

---

---

---

---

---

---

Now consider the democ\_regime-frac\_eth3 relationship for high-GDP countries. Are ethnic heterogeneity and democracy related? Answer yes or no and briefly explain.

---

---

---

---

---

---

F. Obtain and print a presentation-quality multiple line chart of the democ-frac\_eth3 relationship, controlling for gdpcapo8\_2. (Remember that democracies are coded 1 on democ.) In the Chart Editor, give the scale axis this new title: "Percentage of Democracies." Edit the line weights. Change the style of one of the lines. Make other desired changes to enhance appearance and readability. Print the chart you created.

G. Think about the set of relationships you just analyzed. Consider all the numeric and graphic evidence. How would you describe the relationship between ethnolinguistic heterogeneity and democracy, controlling for GDP per capita? (circle one)

spurious    additive    interaction

Explain your reasoning.

---

---

---

---

---

That concludes the exercises for this chapter. Before exiting SPSS, be sure to save your datasets.