All hypothesis testing in political research follows a common logic of comparison. The researcher separates subjects into categories of the independent variable and then compares these groups on the dependent variable.

Suppose I think that gender (independent variable) affects opinions about gun control (dependent variable) and that women are less likely than men to oppose gun control. I would divide subjects into two groups based on gender, women and men, and then compare the percentage of women who oppose gun control with the percentage of men who oppose gun control. Similarly, if I hypothesize that Republicans have higher incomes than do Democrats, I would divide subjects into partisanship groups (independent variable), Republicans and Democrats, and compare the average income (dependent variable) of Republicans with that of Democrats.

Although the logic of comparison is always the same, the appropriate method depends on the levels of measurement of the independent and dependent variables. In this chapter, you will learn how to address two common hypothesis-testing situations. First, we will cover situations in which the independent and the dependent variables are both categorical (nominal or ordinal). Then we will consider instances in which the independent variable is categorical and the dependent variable is interval. In this chapter, you also will learn to create box plots, bar charts, and strip charts, all of which can greatly assist you in interpreting relationships.

CROSS-TABULATION ANALYSIS

Cross-tabulations are the workhorse vehicles for testing hypotheses for categorical variables. A cross-tabulation is a table that compares the value(s) of a dependent variable observed in observations with different values of an independent variable. For example, one would use a cross-tabulation to compare the party identifications observed in people with different levels of income.

Consider this hypothesis: In a comparison of individuals, people who have lower incomes will be more likely to identify with the Democratic Party than those who have higher incomes. The NES dataset contains

1For future discussion of the methods used to make comparisons and how to interpret cross-tabulation and mean comparison tables, see Philip Pollock III, The Essentials of Political Analysis, 5th ed. (Thousand Oaks, CA: CQ Press, an imprint of SAGE Publications, 2016), 58–63. In later chapters, we’ll discuss how to perform comparisons when the independent variable is measured at the interval level and the dependent variable is also interval level (regression analysis) or the dependent variable is binary (logistic regression). The logic is the same in all cases: we want to study the effect of different values of independent variables.
a variable named partyid3, which measures respondents' partisanship in three categories: Democrat ("Dem"), Independent ("Ind"), and Republican ("Rep"). This will serve as the dependent variable. The variable, income3, is the independent variable. Recall that income3 classifies individuals by terciles of income: the lowest third ("Low"), the middle third ("Mid"), and the highest third ("High"). Let's ask Stata to test the party identification–income hypothesis.

When setting up a cross-tabulation, you must observe the following three rules. First, values of the independent variable should define the columns and values of the dependent variable should define the rows. Second, always obtain percentages of the independent variable, not the dependent variable. Third, interpret the cross-tabulation by comparing the percentages of subjects who fall into the same category of the dependent variable.

Stata's omnipresent tabulate command—the command that produces frequency distributions and generates indicator variables—also produces cross-tabulations. The general syntax for a bivariate cross-tabulation that heeds the three rules of cross-tab construction is as follows:

tabulate dep_var indep_var [aw=weightvar], column

This syntax is simple and direct. Type "tabulate" or simply "tab", followed by the name of the dependent variable and the name of the independent variable. To ensure that Stata will report percentages of the independent variable, type a comma and then "col" (an acceptable abbreviation for the column option).

Applying the syntax to the analysis at hand, we would run the following command:

* use "nes.dta"

```
tab partyid3 income3 [aw=nesw], col
```

Stata treats the first-named variable as the row variable and the second-named variable as the column variable. Under this standard setup, the column option will produce percentages of the independent variable. In situations in which the independent variable has many more categories than the dependent variable, the standard setup might create formatting problems, so you may want to create a cross-tabulation having the dependent variable on the columns and the independent variable on the rows. The following syntax would produce the desired result: tab independent_variable dependent_variable, row. The row option instructs Stata to calculate percentages of the row variable, which in this case is the independent variable.

---

2Stata treats the first-named variable as the row variable and the second-named variable as the column variable. Under this standard setup, the column option will produce percentages of the independent variable. In situations in which the independent variable has many more categories than the dependent variable, the standard setup might create formatting problems, so you may want to create a cross-tabulation having the dependent variable on the columns and the independent variable on the rows. The following syntax would produce the desired result: tab independent_variable dependent_variable, row. The row option instructs Stata to calculate percentages of the row variable, which in this case is the independent variable.
The presence of weighted frequencies creates a confusing clutter of digits. Before we rerun the cross-tabulation and request a cleaner display, let’s consider the table at hand. The values of the dependent variable, partyid3, define the rows of the table, and the values of the independent variable, income3, define the columns. In fact, Stata has given us a set of side-by-side frequency distributions of the dependent variable—one for each category of the independent variable—plus an overall frequency distribution for all analyzed cases. Accordingly, the table has four columns of numbers. The first column shows the distribution for approximately 1,236 respondents in the low-income category. The middle column shows how the 1,196 individuals in the medium-income group are distributed across partyid3, and the next column depicts the 1,050 people in the high-income group. The “Total” column shows the distribution of all 3,482 valid cases across the dependent variable. Each cell reports the column percentage and the weighted frequency of cases in the cell.

To get a clearer view of the column percentages, which will allow us to test the hypothesis that people who have lower incomes will be more likely to identify as Democrats, we can rerun the command using the “nofreq” option to suppress frequencies. We’ll also include the “nokey” option, which suppresses the space-consuming key box.

```
tab partyid3 income3 [aw=nesw], col nofreq nokey
```

What do you think? Does the cross-tabulation fit the hypothesis? The third rule of cross-tabulation analysis is easily applied. Focusing on the “Dem” value of the dependent variable, we see a pattern in the hypothesized direction. A comparison of respondents in the “Low” column with those in the “Med” column reveals a decline from 39.71 to 33.63 in the percentage who are Democrats, a drop of about 6 percentage points. Moving from the “Med” column to the “High” column, we find a slight decrease, from 33.63 percent to 33.28 percent. Are lower-income people more likely to be Democrats than middle-income and higher-income people? Yes. This cross-tabulation analysis supports the hypothesis.

**VISUALIZING COMPARISONS WITH NOMINAL OR ORDINAL DEPENDENT VARIABLES**

As we have just seen, a cross-tabulation summarizes the relationship between two variables measured at the nominal or ordinal level. In conjunction with cross-tabulation analysis, you may wish to create a bar chart to visualize the relationship between two categorical variables, such as the relationship between income (income3) and party identification (pid_3).

Let’s take another look at the partyid3-income3 cross-tabulation we produced in the last section to figure out what a bar chart of the relationship should look like. Suppose we would like the bar chart’s y-axis to display the percentage of respondents who are Republicans. The categories of income3—“Low,” “Med,” and “High”—will appear along the horizontal axis. Based on the cross-tabulation analysis, we can anticipate what the bar chart will look like. The bar for the low-income group will stand at 19.61 percent, the medium-income group at 31.61 percent, and the high-income group at 33.35 percent. We can easily state our graphic goal in human language: “I want to graph the percentage of Republicans across different values of income3.”
As is sometimes the case with Stata, however, this human language cannot be translated directly into a single Stata command.

The simple truth is that Stata does not like to graph categorical dependent variables—not in their naturally coded state, anyway. But Stata does like to graph indicator variables, transformations of categorical variables. In Chapter 3 you learned how to run tabulate (with the generate option) to compute a set of indicator variables from a categorical variable. For any categorical variable having k categories, Stata will generate k indicator variables, named sequentially based on the name stem that you supply in the generate option. Each indicator variable will be coded 1 for cases falling into that category of the original variable and coded 0 for cases not falling into that category.

The first step in graphing a categorical dependent variable is to generate a set of indicator variables. Let's get to the example at hand. Execute the following command to generate a set of party identification indicator variables.

```
* use "nes.dta"

tab partyid3 [aw=nesw], gen(pid_dum)
```

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dem</td>
<td>1,288.18</td>
<td>35.48</td>
<td>35.48</td>
</tr>
<tr>
<td>Ind</td>
<td>1,325.8454</td>
<td>36.51</td>
<td>71.99</td>
</tr>
<tr>
<td>Rep</td>
<td>1,016.9746</td>
<td>28.01</td>
<td>100.00</td>
</tr>
<tr>
<td>Total</td>
<td>3,631</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>

Stata produces a frequency distribution table in the Results window. Stata also creates three indicator variables—pid_dum1, pid_dum2, and pid_dum3—and lists them at the bottom of the Variables window. (We're using the suffix "dum" in the indicator variable name stem because we are generating dummy variables from partyid3.)

Focus your attention on pid_dum3. Because "Dem" is the lowest coded value of partyid3 (code 1), the first indicator variable that Stata generates, pid_dum1, is coded 1 for respondents who are Democrats and coded 0 for respondents who are Independents or Republicans. The variable pid_dum3 is coded 1 for Republicans and 0 for Democrats and Independents. Now let's see why Stata likes to work with indicator variables. The command below will return the mean of pid_dum3 (dependent variable) for each value of income3 (independent variable). The mean of pid_dum3 is equal to the proportion of observations coded 1 (respondents who identify with the Republican Party).

```
tab income3 [aw=nesw], sum(pid_dum3) nost noref noobs
```

<table>
<thead>
<tr>
<th>Income (v16136i)</th>
<th>Summary of pid_dum3==Rep</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>.19606226</td>
<td></td>
</tr>
<tr>
<td>Med</td>
<td>.31611435</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>.33350273</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>.27874678</td>
<td></td>
</tr>
</tbody>
</table>
What is the mean of pid_dum3 for low-income respondents (“Low”)? For this group, the mean of pid_dum3 is .1961 (that is, .19606226 rounded to four decimal places). Refer back to the cross-tabulation analysis. What percentage of this income group are Democrats? 19.61 percent. You can see that the means of our newly created indicator variable, pid_dum3, are identical to the proportions of respondents in each income group who fall into the “Rep” category of the original variable, partyid3. A proportion is just a percentage that’s been divided by 100. In the cross-tabulation we found that 31.61 percent of respondents in the medium-income group are Republicans, which appears as a proportion equal to .3161 on pid_dum3. For the high-income group, the numbers are equivalent as well: 33.35 percent and .3335.

More generally, the mean of an indicator variable is equal to the proportion of cases coded 1 on the indicator variable. Because a code of 1 on pid_dum3 denotes “Rep,” then the means .1961, .3161, and .3335 report the proportions of respondents in each value of the independent variable, income3, who are Republicans.

Having created the necessary indicator variable, let’s use the bar graph command we introduced in Chapter 2 to visualize the proportion of respondents in each income group that identify as Republicans. We’ll demonstrate the basic version of the bar chart shown in Figure 4-1 first and then fine-tune the graphic to communicate the relationship more effectively.

```
graph bar (mean) pid_dum3 [aw=nesw], over(income3)
```

**Figure 4-1** Proportion Identifying as Republican by Income Category

Graphically, percentages are more attractive than proportions. Using Stata’s replace command, we can convert pid_dum3 into percentages. Why use the replace command instead of generate? (See “A Closer Look” for a discussion of the replace command.)

```
replace pid_dum3 = pid_dum3*100
```
Because our enhanced bar chart of the relationship between income3 and pid_dum3 will take more than one line of code, use the Do-file editor. The heart of the modified bar chart is the basic bar chart to which we set additional graphical parameters, such as the color of the bars, the background color, the x- and y-axis tiles, and a note regarding the source of the data. A bar chart of the pid_dum3-income3 relationship appears in the Graph window (Figure 4-2).

```stata
#delimit ;
graph bar (mean) pid_dum3 [aweight = nesw], over(income3)
    bar(1, fcolor(gs10))
    graphregion(fcolor(white))
    ytitle("Proportion Republican", size(medsmall))
    title("Proportion Republican, by Income", size(medlarge))
    note("Source: 2016 ANES")
#delimit /
```

Figure 4-2  Proportion Identifying as Republican by Income Category, Enhanced With Options

SOURCE: 2016 ANES

A Closer Look  The replace Command

Stata’s replace command, a special instance of the generate command, can be used to change or replace the values of a variable that already exists. Unlike Stata’s devil-may-care attitude toward the recode command—if the user doesn’t consciously specify the generate option, the recode command will destroy and replace the original variable—Stata will not permit the generate command to overwrite an existing variable.

Suppose you sought to convert pid_dum1 from a proportion to a percentage; that is, you wanted to multiply pid_dum1 times 100. If you entered “generate pid_dum1 = pid_dum1 * 100”, Stata would refuse, saying “pid_dum1 already defined”. However, if you typed “replace pid_dum1 = pid_dum1 * 100”, Stata would perform as requested.
MEAN COMPARISON ANALYSIS

We now turn to another common hypothesis-testing situation: a categorical independent variable and an interval-level dependent variable. The logic of comparison still applies—divide cases on the independent variable and compare values of the dependent variable—but the method is different. Instead of comparing percentages, we now compare means.

To illustrate, let’s say that you are interested in explaining this dependent variable: attitudes toward Donald Trump. Why do some people give him positive ratings while others rate him lower? Here is a plausible (if not self-evident) idea: Partisanship (independent variable) will have a strong effect on attitudes toward Donald Trump (dependent variable). The hypothesis: In a comparison of individuals, those who are Republicans will have more favorable attitudes toward Donald Trump than will those who are Democrats.

The NES dataset contains a variable named ft_Trump_pre, a 100-point feeling thermometer (conducted before the 2016 election). In the survey, each respondent was asked to rate candidate Trump on this scale, from 0 (cold or negative) to 100 (warm or positive). This is the dependent variable. The dataset also has partyid7, which measures partisanship in seven ordinal categories, from Strong Democrat (coded 1) to Strong Republican (coded 7). (The intervening codes capture gradations between these poles: Weak Democrat, Independent-Democrat, Independent, Independent-Republican, and Weak Republican.) This is the independent variable.

If the hypothesis is correct, we should find that Strong Republicans have the highest mean scores on ft_Trump_pre, and that mean scores vary systematically across categories of partyid7, hitting bottom among respondents who are Strong Democrats. Is this what happens?

Yet again, Stata’s tabulate command will provide the answer. In using tabulate to perform bivariate mean comparison analyses, the general syntax is as follows:

```
tabulate indep_var [aw=weightvar], summarize(dep_var)
```

The syntax to the left of the comma says that you want Stata to divide cases according to the values of the independent variable. In the current example, partyid7 is the independent variable, so we would begin the command by typing “tab partyid7”. The summarize option (which can be abbreviated “sum”) tells Stata to calculate summary statistics of the dependent variable for each category of the independent variable. So our dependent variable, ft_Trump_pre, would go inside the parentheses. The completed command, including NES's weight variable, for the Donald Trump thermometer is as follows:

```
* use "nes.dta"

tab partyid7 [aw=nesw], sum(ft_Trump_pre)
```

Stata responds:

<table>
<thead>
<tr>
<th>Party ID (v161158x)</th>
<th>Summary of Feeling therm: Donald Trump (pre, v161087)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>StrongDem</td>
<td>7.6824527</td>
</tr>
<tr>
<td>WkDem</td>
<td>21.24444</td>
</tr>
<tr>
<td>IndDem</td>
<td>18.130297</td>
</tr>
<tr>
<td>Indep</td>
<td>35.304794</td>
</tr>
<tr>
<td>IndRep</td>
<td>57.456165</td>
</tr>
<tr>
<td>WkRep</td>
<td>53.808955</td>
</tr>
<tr>
<td>StrongRep</td>
<td>73.049062</td>
</tr>
</tbody>
</table>

| Total                | 36.407078 | 34.99848 | 3,586 | 3,600 |
By default, the `summarize` option will display the mean of the dependent variable (“Mean”), its standard deviation (“Std. Dev.”), plus frequencies (“Freq.”). For weighted data, Stata also displays the unweighted number of cases in each category of the independent variable (“Obs.”). Despite this bumper crop of output, the table remains eminently readable.

If you prefer to focus your attention on the mean values of Trump thermometer values, you may suppress the standard deviation (by adding the “nost” option), the weighted frequencies (“nofreq”), and the number of unweighted cases (“noobs”):

```
  tab partyid7 [aw=nesw], sum(ft_Trump_pre) nost nofreq noobs
```

A more concise table appears in the Results window:

```
<table>
<thead>
<tr>
<th>Party ID (v161158x)</th>
<th>Summary of Feeling therm: Donald Trump (pre, v161087)</th>
</tr>
</thead>
<tbody>
<tr>
<td>StringDem</td>
<td>7.6824527</td>
</tr>
<tr>
<td>WkDem</td>
<td>21.24444</td>
</tr>
<tr>
<td>IndDem</td>
<td>18.1230297</td>
</tr>
<tr>
<td>Indep</td>
<td>35.304794</td>
</tr>
<tr>
<td>IndRep</td>
<td>57.455160</td>
</tr>
<tr>
<td>WkRep</td>
<td>53.608955</td>
</tr>
<tr>
<td>StringRep</td>
<td>73.049062</td>
</tr>
<tr>
<td>Total</td>
<td>36.407078</td>
</tr>
</tbody>
</table>
```

Compared with a cross-tabulation, a mean comparison table is the soul of simplicity. The label of the dependent variable, “Summary of Feeling therm: Donald Trump,” appears along the top of the table. The label for the independent variable, “Party ID,” defines the left-most column, which shows all seven categories, from the lowest-coded value, Strong Democrat, at the top, to the highest-coded value, Strong Republican, at the bottom. Beside each partisan category, Stata has calculated the mean of ft_Trump_pre, displayed at six-decimal-point precision, as in “73.049062” for the Strong Republican mean. (Sometimes you will need to change the number of decimal places that Stata displays in the Results window or in a graphic. The format command is designed for this purpose. See “A Closer Look” for a discussion of the format command.) The bottom row, “Total,” gives the mean for the whole sample.

Among Strong Republicans, the mean of ft_Trump_pre is very high—73.05 degrees. Does the mean decline as attachment to the Republican Party weakens and identification with the Democratic Party strengthens? Yes, but take note of some interesting anomalies. To be sure, the mean drops sharply among Weak Republicans (who average 53.81 degrees), but it momentarily increases among Independent-Republican leaners (57.46). Trump’s ratings then continue to decline predictably as one proceeds to the left on the political spectrum, although they bump up slightly between Independent-Democrat leaners (18.13) and Weak Democrats (21.24). Strong Democrats, who average 7.68 degrees on the thermometer, have the coldest response to Trump. On the whole, then, the data support the hypothesis.
There are two main visual accompaniments for mean comparison analysis: box plots and bar charts. Box plots describe a numeric variable by graphing a five-number summary: minimum, lower quartile, median, upper quartile, and maximum. Box plots also reveal outliers, observations with very unusual values. Bar charts graph the means of a dependent variable across the values of an independent variable. We will consider both kinds of graphs, beginning with box plots.

What is depicted by a box plot? The box plot communicates a number of important qualities about a variable: the lower quartile (the value below which 25 percent of the cases fall), the median (the value that splits the cases into two equal-size groups), and the upper quartile (the value below which 75 percent of the cases fall). Thus, the distance between the bottom and top of the box defines the interquartile range (IQR), the range of a variable that encompasses the “middle half” of a distribution. The IQR is used to locate the upper and lower hinges; the upper hinge is located 1.5*IQR above the upper quartile and the lower hinge is located 1.5*IQR below the lower quartile. Any observations above the upper hinge or below the lower hinge are considered outliers.

The general syntax for creating a box plot with Stata is:

```
graph box dep_var [aweight = weightvar], over (indep_var)
```
For the Trump thermometer and party identification example, we would use the command below. Go ahead and run the basic command. We can add options to improve the graphic later.

**Figure 4-3** Box Plot of a Trump Feeling Thermometer by Partisan Identification

Consider the box plot shown in Figure 4-3. For example, notice how spread out Independents are in their feelings about Trump: Their median rating is 40, but half of them rated Trump in the long interval between 0 (lower quartile) and 60 (upper quartile), an interquartile range of about 60 points. Contrast this to the cohesiveness of the Strong Democrats: Their median Trump rating is 0 and the 75% quartile score isn’t much higher.

Notice in Figure 4-3 that Trump sentiments are so varied among Weak Democrats, Independents, Independent-Republicans, and Weak Republicans that all possible feeling thermometer values from 0 to 100 are between the upper and lower hinges. Box plots convey a lot of information for comparative purposes in a compact display. Among Strong Democrats and Strong Republicans we can see outlier observations: Strong Democrats with very positive sentiments about Trump and Strong Republicans who give Trump low scores.

Before we move on, we will rerun the graph box command, adding labels, color, and other enhancements. The finished product appears in the Graph window (Figure 4-4).

---

3The “over” option only works for variables with numeric value codes. If the variable to be plotted “over” has only text values (strings), the command will not work. See Chapter 3 for more information on assigning/modifying numeric values for variables.
A box plot favors the display of dispersion over central tendency, providing a valuable complement to mean comparison analysis. A bar chart, which plots the mean of a dependent variable across the values of an independent variable, is aimed more squarely at visualizing mean comparisons by expressing them in graphic form.

The general syntax of the basic bar chart command is very similar to the box plot syntax. Note that you specify “bar” instead of “box,” and tell Stata you want to graph the mean values (“mean”) of the dependent variable across values of the independent variable.

```
graph bar (mean) dep_var [aw = weightvar], over(indep_var)
```

Sticking with our Trump thermometer and party identification example, we execute the following command to produce the bar chart shown in Figure 4-5.

```
graph bar (mean) ft_Trump_pre [aw=nesw], over(partyid7)
```

**Figure 4-5**  Bar Chart of Mean Trump Feeling Thermometer Scores by Partisan Identification
The simplified syntax produced a highly readable chart. This graphic is at once simple and informative. The y-axis records mean values of the dependent variable, the Donald Trump feeling thermometer. The bars trace the mean ratings across the values of party identification: approximately 8 degrees for Strong Democrats, 21 degrees for Weak Democrats, and so on, all the way to 73 degrees for Strong Republicans. One can immediately see the relationship between the independent and dependent variables—as partyid7 increases from low codes to high codes (from Strong Democrat to Strong Republican), mean Trump ratings increase. At the same time, one sees the curious anomalies between the independent leaners and weak partisans of each party.

As useful as the bar graph of Trump support is, you can always use options to add content and style. With one exception (the bar options take the place of box options), the graph options used to create Figure 4-6 are identical to those you applied to the box plot of the same relationship:

```
#delimit ;
graph bar ft_Trump pre [aw=nesw], over(partyid7)
   bar(1, fcolor(gs8) lcolor(black))
   graphregion(fcolor(white))
   ytitle("Trump Rating", size(medsmall))
   title("Trump Rating, by Party Identification", size(medlarge))
   note("Source: 2016 ANES")
;
```

Another graphics success appears in the Graph window.

**Figure 4-6**  Bar Chart of Mean Trump Feeling Thermometer Scores by Partisan Identification, Enhanced With Options

![Bar Chart](image-url)
STRIP CHARTS: GRAPHS FOR SMALL-N DATASETS

The graphic types discussed thus far—box plots and bar charts—excel at displaying summary statistics for large datasets. Strip charts display case-level information, making them particularly appropriate for smaller datasets, such as states or world. Like box plots, strip charts show how the cases are distributed within each value of an independent variable. Thus, the analyst might create a strip chart to complement a mean comparison analysis.

Suppose we are interested in comparing levels of unionization in different regions of the United States. For this analysis, we can use the states dataset. We would run tabulate with the summary option, obtaining the mean of union_2016 (percentage of the workforce belonging to a union) by region (of the country):

```
* use "states.dta"
tab region, sum(union_2016)
```

<table>
<thead>
<tr>
<th>Census region</th>
<th>Summary of Percent of workers who are union members</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Northeast</td>
<td>14.355556</td>
<td>4.3491699</td>
</tr>
<tr>
<td>Midwest</td>
<td>9.9416666</td>
<td>3.3035678</td>
</tr>
<tr>
<td>South</td>
<td>6.35</td>
<td>3.3119984</td>
</tr>
<tr>
<td>West</td>
<td>11.3</td>
<td>5.4703746</td>
</tr>
<tr>
<td>Total</td>
<td>9.94</td>
<td>4.9502422</td>
</tr>
</tbody>
</table>

The Northeast has the highest average level of unionization (14.36 percent), followed by the West (11.30 percent), the Midwest (9.94 percent), and the South (6.35 percent). But examine the corresponding standard deviations, which measure variation. Notice that the standard deviation for the West (5.47) is comparatively large, nearly twice that for the South (3.31). This suggests that the thirteen western states are widely dispersed, with some having high levels of unionization and others having lower levels.

A strip chart, produced with the scatter command, can provide valuable insights on questions of dispersion and skewness. The basic syntax about as straightforward as it gets in Stata:

```
scatter dep_var indep_var
```

To apply this basic syntax to the union_2016-region example, we can enter the following command to make a graphic like Figure 4-7.
The strip chart created by this terse, three-word command is austere but interpretable. We can see why the mean comparison analysis returned interregional differences in the level of unionization. And notice the differences between southern states (code 3 on region) and western states (code 4). A group of western states has low levels of unionization, much like states in the South.

Now that we've produced a basic strip chart, let's adjust some optional strip chart settings to make Figure 4-7 more attractive and informative. The following is a command for a well-optioned strip chart, adapted to the example at hand. Some of these options should look familiar. The xlabel(1 2 3 4, valuelabel) option replaces the numeric codes on the basic strip chart axis with the value labels “Northeast”, “Midwest”, “South”, and “West”. The jitter(7) option randomly varies the point placement a small amount so they don't appear as overlapping points on a straight vertical line.

The strip chart created by this terse, three-word command is austere but interpretable. We can see why the mean comparison analysis returned interregional differences in the level of unionization. And notice the differences between southern states (code 3 on region) and western states (code 4). A group of western states has low levels of unionization, much like states in the South.

Now that we've produced a basic strip chart, let's adjust some optional strip chart settings to make Figure 4-7 more attractive and informative. The following is a command for a well-optioned strip chart, adapted to the example at hand. Some of these options should look familiar. The xlabel(1 2 3 4, valuelabel) option replaces the numeric codes on the basic strip chart axis with the value labels “Northeast”, “Midwest”, “South”, and “West”. The jitter(7) option randomly varies the point placement a small amount so they don't appear as overlapping points on a straight vertical line.

The resulting graph (Figure 4-8) is quite presentable.
Making Comparisons

EXERCISES

1. (Dataset: NES. Variables: Who_2016, better_worse_past_econ, [aw=nesw].) What factors determine how people vote in presidential elections? Political scientists have investigated and debated this question for many years. A particularly powerful and elegant perspective emphasizes voters’ retrospective evaluations. According to this view, for example, voters who think the country’s economy has gotten better during the year preceding the election are likely to reward the candidate of the incumbent party. Voters who believe the economy has worsened, by contrast, are likely to punish the incumbent party by voting for the candidate of the party not currently in power. As political scientist V. O. Key once famously put it, the electorate plays the role of “rational god of vengeance and reward.” Does Key’s idea help explain how people voted in the 2016 election?

A. Test this hypothesis: In a comparison of individuals, those who thought the economy had improved during the year preceding the 2016 election were more likely to vote for the candidate of the incumbent party, Hillary Clinton, than were individuals who thought the economy had not improved. Use these two variables from the NES dataset: Who_2016 (dependent variable), better_worse_past_econ (independent variable), and nesw (weight variable). Obtain a cross-tabulation of the relationship. Be sure to use the “col” option for column percentages. Record the percentages voting for Clinton and Trump in the table that follows.

<table>
<thead>
<tr>
<th>Did the economy get better/worse in the last year?</th>
<th>H. Clinton</th>
<th>Trump</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Much better</td>
<td>Somewhat better</td>
<td>About same</td>
</tr>
<tr>
<td>Respondent’s vote, 2016</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Total</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

B. What do you think? Are the data consistent with V.O. Key’s retrospective evaluation hypothesis? Write a paragraph explaining your reasoning.

C. Loss aversion is an interesting psychological phenomenon that can shape the choices people make.\(^5\) One idea behind loss aversion is that losses loom larger than commensurate gains. According to this theory, for example, the psychological pain felt from losing $100 is greater than the pleasure felt from gaining $100. Applied to retrospective voting, loss aversion might suggest that the “vengeance” impulse is stronger than the “reward” impulse—that the anti-incumbent party motivation among those who say the economy has worsened will be stronger than the pro-incumbent party motivation among those who think it has improved.

With this idea in mind, examine the percentages in the table in part A. What do you think? Do the data suggest that Key’s rational god of vengeance is stronger than his rational god of reward? Answer yes or no, and write a few sentences explaining your reasoning.

2. (Dataset: nes2012. Variables: libcon7_Dem, libcon7_Rep, partyid7, [aw=nesw].) Partisan polarization can create some interesting perceptual distortions. Do partisans tend to view themselves as more moderate than they view the opposing party? For example, do Democrats think Republicans are ideologically extreme, yet see themselves as more moderate? By the same token, do Republicans perceive Democrats as liberal extremists, but perceive themselves as purveyors of middle-of-the-road politics? Where do Independents place the Democrats and Republicans on the left-right continuum? A Pew survey found this thought-provoking asymmetry: All partisans—Democrats, Independents, and Republicans—placed the Republicans at practically the same conservative position on the liberal-conservative scale. However, the placement of the Democrats varied widely: Republicans placed Democrats well toward the liberal side, Independents saw Democrats as somewhat left-of-center, and Democrats placed themselves squarely at the moderate position.\(^6\)

In this exercise, you will see if you can replicate the Pew report’s findings using the NES dataset.

The NES variable libcon7_Dem measures respondents’ perceptions of the ideological position of the Democratic Party using the standard 7-point scale: 1 (“Extremely liberal”), 2 (“Liberal”), 3 (“Slightly liberal”), 4 (“Moderate”), 5 (“Slightly conservative”), 6 (“Conservative”), and 7 (“Extremely conservative”). Another variable, libcon7_Rep, asks respondents to place the Republican Party along the same 7-point metric. For the purposes of this exercise, you will treat these two measures as interval-level variables. Thus, lower mean values denote higher perceived liberalism, values around 4 denote perceived moderation, and higher mean values denote higher levels of perceived conservatism. For this exercise, libcon7_Dem and libcon7_Rep are


Making Comparisons

The dependent variables. Our old reliable variable, partyid7, is the independent variable. If the Pew results are correct, you should find that all partisan groups, from Strong Democrats to Strong Republicans, share very similar conservative perceptions of the Republican Party—but hold very different perceptions of the Democratic Party.

A. Run two mean comparison analyses, one for the relationship between libcon7_Dem and partyid7 relationship, and one for the relationship between libcon7_Rep and partyid7. Fill in the means in the following table.

<table>
<thead>
<tr>
<th>Respondent's party ID</th>
<th>Mean ideological placement of Democrats</th>
<th>Mean ideological placement of Republicans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong Democrat</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Weak Democrat</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Independent Democrat</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Independent</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Independent Republican</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Weak Republican</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Strong Republican</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

B. Examine your findings. Are the Pew findings borne out by the NES data? Explain your reasoning.

3. (Dataset: GSS. Variables: polviews, femrole [aw=wtss].) Why do some people hold more traditional views about the role of women in society and politics, whereas others take a less traditional stance? General ideological orientations, liberalism versus conservatism, may play an important role in shaping individuals’ opinions on this cultural question. Thus, it seems plausible to suggest that ideology (independent variable) will affect opinions about appropriate female roles (dependent variable). The hypothesis: In a comparison of individuals, liberals will be more likely than conservatives to approve of nontraditional female roles.

The GSS dataset (file name: gss.dta) contains femrole, a scale that measures opinions about the appropriate role of women. The numeric values of femrole range from 0 (women belong in traditional roles) to 9 (women belong in nontraditional roles). That is, higher scores denote less traditional beliefs. This is the dependent variable. The dataset has another familiar variable, polviews, a 7-point ordinal scale measuring ideology. Scores on polviews can range from 1 (extremely liberal) to 7 (extremely conservative). This is the independent variable.

A. According to the hypothesis, as the values of polviews increase, from 1 through 7, mean values of femrole should (circle one)

- decrease
- neither decrease nor increase
- increase

B. Run tabulate with the summarize option to obtain mean values of femrole across values of polviews. Specify the “nost” and “noobs” options. Fill in the table that follows.
Chapter 4

Summary of female role by respondent's ideological self-placement

<table>
<thead>
<tr>
<th>Respondent's ideological self-placement:</th>
<th>Mean</th>
<th>Frequency*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extremely liberal</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Liberal</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Slightly liberal</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Moderate</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Slightly conservative</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Conservative</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Extremely conservative</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Total</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

*Round weighted frequencies to two decimal places.

C. Do the results support the hypothesis? Write a few sentences explaining your reasoning.

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

D. Create and print a bar chart showing mean values of femrole (y-axis) over values of polviews. Provide descriptive titles and fonts. Add a source note. For assistance, refer to this chapter's examples. Print the chart.

4. (Dataset: NES. Variables: egalit_3, educ4, [aw=nesw].) Pedantic Pontificator is offering a group of students his thoughts about the relationship between educational attainment and egalitarianism, the belief that government should do more to make sure resources are more equitably distributed in society: "Educated people have a humanistic world view that is sorely lacking among the self-seeking, less-educated classes. They see inequality . . . and want to rectify it! Plus, most colleges and universities are populated with liberal faculty, who indoctrinate their students into left-wing ideologies at every opportunity. Thus, it's really quite simple: As education goes up, egalitarianism increases."

The NES dataset contains egalit_3, which measures egalitarian beliefs in three categories: "Low," "Medium," and "High." The NES dataset also has educ4, which records educational attainment in four categories: high school or less ("HS or less"), some college or associate's degree ("SmColl/Assoc"), bachelor's degree ("BA"), and graduate degree ("Grad").

A. Run a cross-tabulation analysis that tests Pedantic Pontificator's idea about the relationship between education and egalitarianism. Obtain column percentages. Use the "col" and "nofreq" options to request column percentages and uncluttered results. Be sure to use weights to produce nationally representative results. Use the table below to record the percentages you obtained.
### Making Comparisons

<table>
<thead>
<tr>
<th></th>
<th>HS or less</th>
<th>SmColl/Assoc</th>
<th>BA</th>
<th>Grad</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total</strong></td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

B. Based on your analysis, would it appear that Pedantic Pontificator is correct? Answer yes or no, and explain.

C. Run tabulate with the generate option to obtain a set of indicator variables for egalit_3. Use “egal” and the variable stem. Obtain a nicely optioned bar chart showing the percentage in the “More egalitarian” category across values of educ4. Print the chart.

5. (Dataset: GSS. Variables: intsex, fepol, fefam, [aw=wtss].) Untruthful answers by survey respondents can create big headaches for public opinion researchers. Why might a respondent not tell an interviewer the truth? Certain types of questions, combined with particular characteristics of the interviewer, can trigger a phenomenon called preference falsification: “the act of misrepresenting one's genuine wants under perceived social pressures.” For example, consider the difficulty in gauging opinions on the role of women in society. One might reasonably expect people questioned by a female interviewer to express greater support for feminist views than those questioned by a male pollster. Someone who supports traditional gender roles, not wanting to appear insensitive to a female questioner, might instead offer a false pro-feminist opinion.

The GSS dataset contains intsex, coded 1 and labeled “Male” for respondents questioned by a male interviewer, and coded 2 and labeled “Female” for those questioned by a female interviewer. This is the independent variable that will allow you to test two preference falsification hypotheses:

**Hypothesis 1:** In a comparison of individuals, those questioned by a female interviewer will be more likely to disagree with the proposition that women are unsuited for politics than will those questioned by a male interviewer. (The dependent variable is fepol, coded 1 for “Agree” and 2 for “Disagree.”)

**Hypothesis 2:** In a comparison of individuals, those questioned by a female interviewer will be more likely to disagree with the statement that it's better for men to work and women to tend home than will those questioned by a male interviewer. (The dependent variable is fefam, which is coded 1 for “Strongly Agree,” 2 for “Agree,” 3 for “Disagree,” and 4 for “Strongly Disagree.”)

---


8It may have occurred to you that this effect might be greater for white respondents than for black respondents, with white subjects more likely to hide their true preferences in the presence of a black interviewer. An exercise in Chapter 5 will give you a chance to investigate this possibility.
A. Run a cross-tabulation analysis analyzing the relationship between intsex and fepol. Make sure to request column percentages and use weights. Complete the cross-tabulations that follow.

<table>
<thead>
<tr>
<th>Statement: Women are not suited for politics.</th>
<th>Interviewer's gender</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Percentage who &quot;Agree&quot;</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Percentage who &quot;Disagree&quot;</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Total</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

These findings (circle one)

support Hypothesis 1.  
do not support Hypothesis 1.

Briefly explain your reasoning. ______________________________________________________________

B. Run a cross-tabulation analysis analyzing the relationship between intsex and fefam. Make sure to request column percentages and use weights. Complete the cross-tabulations that follow.

<table>
<thead>
<tr>
<th>Statement: It's better for men to work and women to tend home.</th>
<th>Interviewer's gender</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Percentage who &quot;Strongly Agree&quot;</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Percentage who &quot;Agree&quot;</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Percentage who &quot;Disagree&quot;</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Percentage who &quot;Strongly Disagree&quot;</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Total</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

These findings (circle one)

support Hypothesis 2.  
do not support Hypothesis 2.

Briefly explain your reasoning. ______________________________________________________________

______________________________________________________________
Perform two tasks before proceeding to part C. (i) Generate a set of indicator variables from `fepol`, using the variable handle “fempolitics”. The new variable `fempolitics2` will have value 1 for respondents who disagree with the proposition that women are unsuited for politics. (ii) Run the replace command to convert `fempolitics2` from a proportion to a percentage (in other words, replace the 1s with 100s).

C. Create and print a nicely optioned bar chart of the relationship between interviewer gender and opinions about whether women are suited for politics. Be sure to provide a y-axis title, a chart title, and source note.

6. (Dataset: states. Variables: ProLife, region, state.) In this exercise, you will (i) create and print a box plot of the relationship between ProLife (the percentage of the public holding a “pro-life” position on abortion) and region; (ii) identify outliers within regions; and (iii) create and print a strip chart of the relationship.

A. Obtain and print a box plot of the relationship between ProLife and region. You will want to apply the options discussed in this chapter. However, you will want to add an option (not discussed previously) that will label outliers using the variable `state`, a string variable that identifies each state by name. In the box plot options, add the following, just as you see it here. The first character inside the left parenthesis is the number one.

```
mark(1, mlab(state))
```

B. The box plot you produced in part A identifies one outlier in the Northeast and two outliers in the West. (Fill in the blanks that follow.) Which state is the outlier in the Northeast? _________________. Which two states are outliers in the West? ________________ and ________________.

C. Consider this claim: “As measured by the interquartile range, Southern states are less spread out—that is, have less variation in pro-life opinions—than states in the Midwest.” Is this claim correct? Answer yes or no, and briefly explain.

```

```

D. Consider this claim: “Ignoring their outliers, the Northeastern states are more cohesive in their pro-life opinions than are the Western states.” Is this claim correct? Answer yes or no, and briefly explain.

```

```

E. Use the scatter command to obtain a strip chart of the ProLife-region relationship. Make the chart presentable by specifying appropriate options. Make sure to include the jitter option, so the markers do not overlap. Print the chart.

7. (Dataset: states. Variables: suicide_rate, Gun_rank3.) Two policy researchers are debating whether gun control can prevent suicides:
Policy Researcher 1: “People who are determined to commit suicide will find a way to do it. Restricting access to guns will have no effect on suicide rates.”

Policy Researcher 2: “Look, any behavior that’s made more difficult will tend to decrease. If state governments want to discourage suicides, then restricting access to guns will certainly have the desired effect. More restrictions mean fewer suicides.”

Imagine a bar chart of the relationship between restrictions on guns and suicide rates. The horizontal axis measures access to guns in three categories, from more restrictive on the left to less restrictive on the right. The vertical axis records the suicide rate per 100,000 population. Below are two graphic shells, A and B. In shell A, you will sketch a bar chart depicting what the relationship would look like if Policy Researcher 1 is correct. In shell B, you will sketch a bar chart depicting what the relationship would look like if Policy Researcher 2 is correct.

A. If Policy Researcher 1 were correct, what would the bar chart look like? Sketch three bars inside the graphic space, depicting the relationship proposed by Policy Researcher 1.

B. If Policy Researcher 2 were correct, what would the bar chart look like? Sketch three bars inside the graphic space, depicting the relationship proposed by Policy Researcher 2.
C. The states dataset contains the variables suicide_rate and Gun_rank3. Run a mean comparison analysis, using suicide_rate as the dependent variable and Gun_rank3 as the independent variable. Record your results in the table that follows.

<table>
<thead>
<tr>
<th>Access to guns</th>
<th>Suicide rate per 100,000 population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>More restrictions</td>
<td>?</td>
</tr>
<tr>
<td>Mid-level restrictions</td>
<td>?</td>
</tr>
<tr>
<td>Fewer restrictions</td>
<td>?</td>
</tr>
<tr>
<td>Total</td>
<td>15.43</td>
</tr>
</tbody>
</table>

D. Create and print a nicely optioned bar chart of the relationship you just analyzed. Be sure to include a descriptive chart title, y-axis title, and source note.

E. Examine the mean comparison table and the line chart. Which policy researcher is more correct? (check one)

- Policy Researcher 1 is more correct.
- Policy Researcher 2 is more correct.

F. Write a paragraph explaining your reasoning in part E.

8. (Dataset: world. Variables: regime_type3, durable.) Three comparative politics scholars are trying to figure out what sort of institutional arrangement produces the longest lasting, most stable political system.

Scholar 1: “Presidential democracies, like the United States, are going to be more stable than are any other type of system. In presidential democracies, the executive and the legislature have separate electoral constituencies and separate but overlapping domains of responsibility. The people's political interests are represented both by the president's national constituency and by legislators’ or parliament members’ more localized constituencies. If one branch does something that's unpopular, it can be blocked by the other branch. The result: political stability.”

Scholar 2: “Parliamentary democracies are by far more stable than presidential democracies. In presidential systems, the executive and legislature can be controlled by different political parties, a situation that produces deadlock. Since the leaders of parliament can't remove the president and install a more compliant or agreeable executive, they are liable to resort to a coup, toppling the whole system. Parliamentary democracies avoid these pitfalls. In parliamentary democracies, all legitimacy and accountability resides with the legislature.”
Chapter 4

The parliament organizes the government and chooses the executive (the prime minister) from among its own leaders. The prime minister and members of parliament have strong incentives to cooperate and keep things running smoothly and efficiently. The result: political stability.

Scholar 3: “You two have made such compelling—if incorrect—arguments that I almost hesitate to point this out: Democracies of any species, presidential or parliamentary, are inherently unstable. Any system that permits the clamor of competing parties or dissident viewpoints is surely bound to fail. If it's stability that you value above all else, then dictatorships will deliver. Strong executives, feckless or nonexistent legislatures, powerful armies, social control. The result: political stability.”

The world dataset contains the variable durable, which measures the number of years since the last regime transition. The more years that have passed since the system last failed (higher values on durable), the more stable a country’s political system. The variable regime_type3 captures system type: dictatorship, parliamentary democracy, or presidential democracy.

A. Perform a mean comparison analysis of the relationship between durable and regime_type3. Based on a comparison of means, which is the apparently correct ranking of regime types, from most stable to least stable?

- parliamentary democracies (most stable), presidential democracies, dictatorships (least stable)
- parliamentary democracies (most stable), dictatorships, presidential democracies (least stable)

B. Create a box plot of the relationship. Closely examine the box plot. In what way does the graphic evidence support the ranking you chose in part A?

C. In what way does the graphic evidence NOT support the ranking you chose in part A?

D. Print the box plot you created in part B.
9. (Dataset: world. Variables: enpp3_democ, district_size3, frac_eth3.) Two scholars of comparative politics are discussing possible reasons why some democracies have many political parties and other democracies have only a few.

**Scholar 1:** “It all has to do with the rules of the election game. Some countries, such as the United Kingdom, have single-member electoral districts. Voters in each district elect only one representative. This militates in favor of fewer and larger parties, since small parties have less chance of winning enough votes to gain the seat. Other countries, like Switzerland, have multimember districts. Because voters choose more than one representative per district, a larger number of smaller parties have a chance to win representation. It doesn’t surprise me in the least, then, that the UK has fewer political parties than Switzerland.”

**Scholar 2:** “I notice that your explanation fails to mention the single most important determinant of the number of political parties: social structural heterogeneity. Homogeneous societies, those with few linguistic or religious differences, have fewer conflicts and thus fewer parties. Heterogeneous polities, by the same logic, are more contentious and will produce more parties. By the way, the examples you picked to support your case also support mine: The UK is relatively homogeneous and Switzerland relatively heterogeneous. It doesn’t surprise me in the least, then, that the UK has fewer political parties than Switzerland.”

A. Scholar 1’s hypothesis: In a comparison of democracies, those having single-member districts will have (circle one)

- fewer political parties
- more political parties

than democracies electing multiple members from each district.

B. State Scholar 2’s hypothesis:

The world dataset variable enpp3_democ measures, for each democracy, the number of effective parliamentary parties: “1–3 parties,” “4–5 parties,” or “6–11 parties.” Use enpp3_democ as the dependent variable to test each hypothesis. For independent variables, test Scholar 1’s hypothesis using district_size3, which measures, for each democracy, the number of members per district: “single-member” districts, more than one but fewer than six members (“>1 to 5”), and countries with “6 or more members” per district. Test Scholar 2’s hypothesis using frac_eth3, which classifies each country’s level of ethnic/linguistic fractionalization as “Low,” “Medium,” or “High.”

C. In the table that follows, record the percentages of cases falling into the lowest code of the dependent variable, 1–3 parties.

<table>
<thead>
<tr>
<th>Average number of members per district</th>
<th>Single member</th>
<th>2 to 5 members</th>
<th>6 or more members</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage having 1–3 parties</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
D. Which of the following statements best summarizes your findings? (check one)

- Scholar 1’s hypothesis is supported by the analysis, but Scholar 2’s hypothesis is not supported by the analysis.

- Scholar 2’s hypothesis is supported by the analysis, but Scholar 1’s hypothesis is not supported by the analysis.

- Both hypotheses are supported by the analysis.

- Neither hypothesis is supported by the analysis.

E. Making specific reference to your findings, write a paragraph explaining your choice in part D.