2 Descriptive Statistics

Watch a screencast of the guided examples in this chapter.
edge.sagepub.com/pollock

Procedures Covered

Analyze → Descriptive Statistics → Frequencies

Analyze → Reports → Case Summaries

Analyzing descriptive statistics is the most basic—and sometimes the most informative—form of analysis you will do. Descriptive statistics reveal two attributes of a variable: its typical value (central tendency) and its spread (degree of dispersion or variation). The precision with which you can describe central tendency for any given variable depends on the variable's level of measurement. For nominal-level variables you can identify the *mode,* the most common value of the variable. For ordinal-level variables, those whose categories can be ranked, you can find the mode and the *median*—the value of the variable that divides the cases into two equal-size groups. For interval-level variables you can obtain the mode, median, and arithmetic *mean,* the sum of all values divided by the number of cases.

In this chapter you will use Analyze → Descriptive Statistics → Frequencies to obtain appropriate measures of central tendency, and you will learn to make informed judgments about variation. With the correct prompts, the Frequencies procedure also provides valuable graphic support—bar charts and (for interval variables) histograms. These tools are essential for distilling useful information from datasets having hundreds of anonymous cases, such as the American National Election Study (NES2012) or the General Social Survey (GSS2012). For smaller datasets with aggregated units, such as the States and World datasets, SPSS offers an additional procedure: Analyze → Reports → Case Summaries. Case Summaries lets you see firsthand how specific cases are distributed across a variable that you find especially interesting.

INTERPRETING MEASURES OF CENTRAL TENDENCY AND VARIATION

Finding a variable's central tendency is ordinarily a straightforward exercise. Simply read the computer output and report the numbers. Describing a variable's degree of dispersion or variation, however, often requires informed judgment.[1] Here is a general rule that applies to any variable at any level of measurement: A variable has no dispersion if all the cases—states, countries, people, or whatever—fall into the same value of the variable. A variable has maximum dispersion if the cases are spread evenly across all values of the variable. In other words, the number of cases in one category equals the number of cases in every other category.

Central tendency and variation work together in providing a complete description of any variable. Some variables have an easily identified typical value and show little dispersion. For example, suppose you were to ask a large number of U.S. citizens what sort of economic system they believe to be the best: capitalism, communism, or socialism. What would be the modal response, or the economic system preferred by most people? Capitalism. Would there be a great deal of dispersion, with large numbers of people choosing the alternatives, communism or socialism? Probably not.

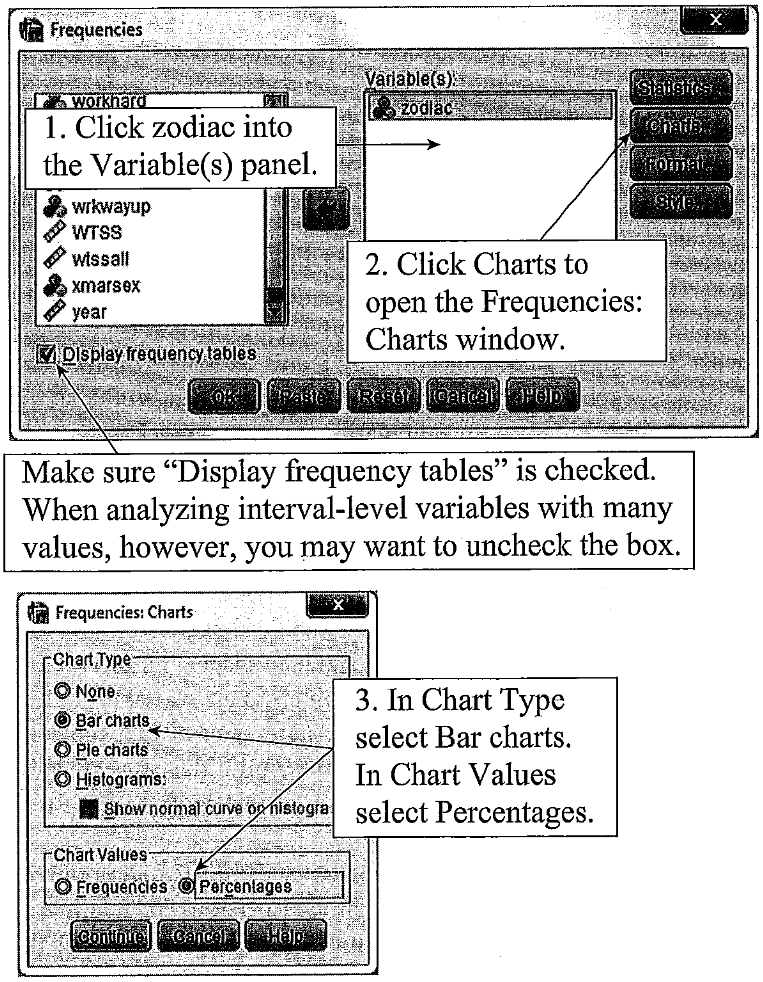
In other instances, however, you may find that one value of a variable has a more tenuous grasp on the label *typical.* And the variable may exhibit more dispersion, with the cases spread out more evenly across the variable's other values. For example, suppose a large sample of voting-age adults were asked, in the weeks preceding a presidential election, how interested they are in the campaign: very interested, somewhat interested, or not very interested. Among your own acquaintances you probably know a number of people who fit into each category. So even if one category, such as "somewhat interested," is the median, many people will likely be found at the extremes of "very interested" and "not very interested." In this instance, the amount of dispersion in a variable—its degree of spread—is essential to understanding and describing it.

These and other points are best understood by working through some guided examples. For the next several analyses, you will use GSS2012. Open the dataset by double-clicking the GSS2012 icon. (If you are using SPSS Student Version, open GSS2012_Student_A.) In the Data Editor, click Edit → Options and then click on the General tab. Just as you did with NES2012 in Chapter 1, make sure that the radio buttons in the Variable Lists area are set for Display names and Alphabetical. (If these options are already set, click Cancel. If they are not set, select them, click Apply, and then click OK. Now you are ready to go.)

DESCRIBING NOMINAL VARIABLES

First, you will obtain a frequency distribution and bar chart for a nominal-level variable, zodiac, which records respondents' astrological signs. Click Analyze → Descriptive Statistics → Frequencies. Scroll down to the bottom

Figure 2-1 Obtaining Frequencies and a Bar Chart (nominal variable)



☑ Display frequency tables

1. Click zodiac into the Variable(s) panel.

2. Click Charts to open the Frequencies: Charts window.

Make sure "Display frequency tables" is checked. When analyzing interval-level variables with many values, however, you may want to uncheck the box.

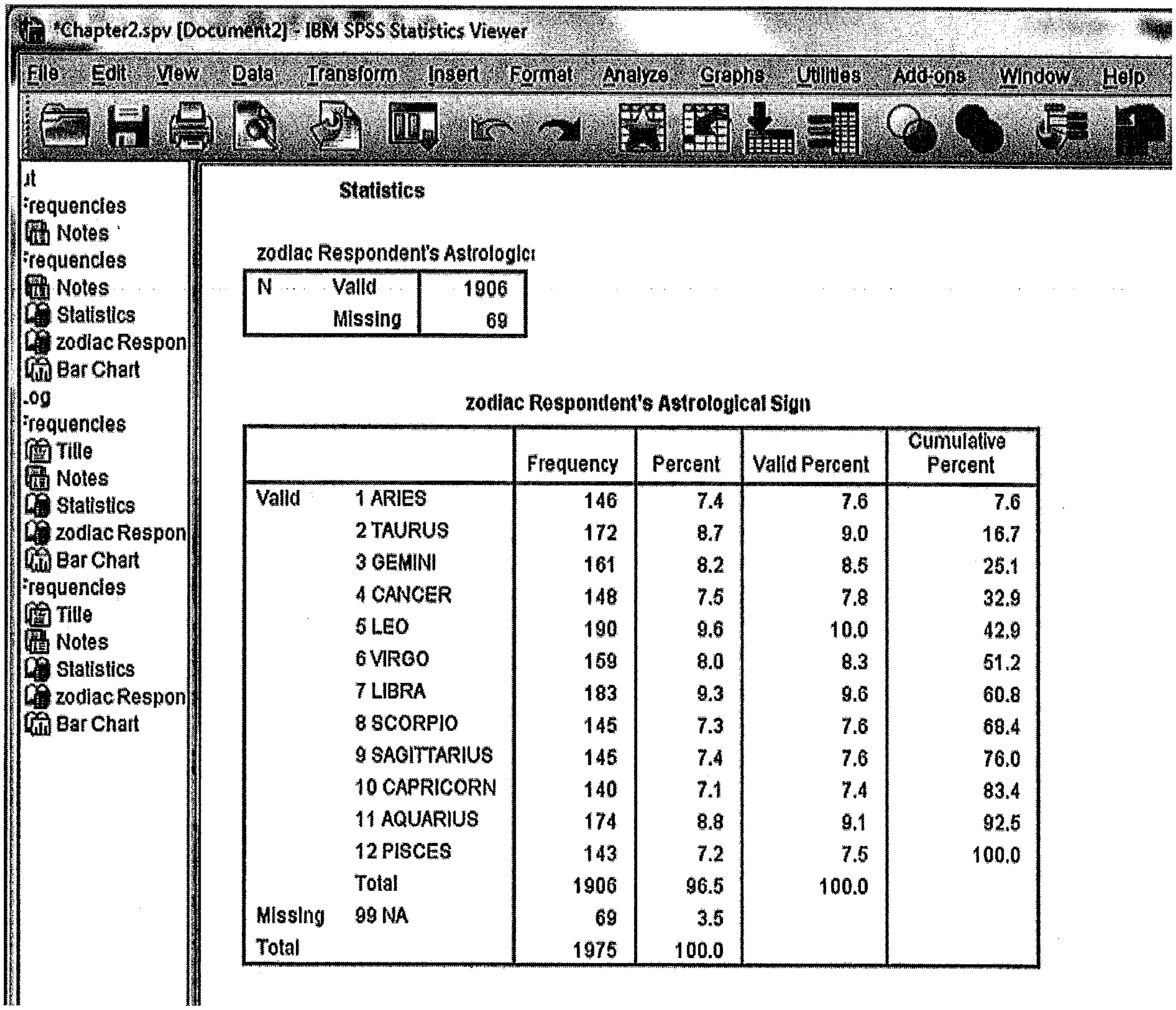
3. In Chart Type select Bar charts. In Chart Values select Percentages.

of the left-hand list until you find zodiac. Click zodiac into the Variable(s) panel. To the right of the Variable(s) panel, click the Charts button (Figure 2-1). The Frequencies: Charts window appears. In Chart Type, select Bar charts. In Chart Values, select Percentages. Click Continue, which returns you to the main Frequencies window. Click OK. SPSS runs the analysis.

SPSS has produced two items of interest in the Viewer: a frequency distribution of respondents' astrological signs and a bar chart of the same information. Examine the frequency distribution (Figure 2-2). The value labels for each astrological code appear in the leftmost column, with Aries occupying the top row of numbers and Pisces occupying the bottom row. There are four numeric columns: Frequency, Percent, Valid Percent, and Cumulative Percent. The Frequency column shows raw frequencies, the actual number of respondents having each zodiac sign. Percent is the percentage of *all* respondents, including missing cases, in each category of the variable. Ordinarily the Percent column can be ignored, because we generally are not interested in including missing cases in our description of a variable. Valid Percent is the column to focus on. Valid Percent tells us the percentage of nonmissing responses in each value of zodiac. Finally, Cumulative Percent reports the percentage of cases that fall in *or below* each value of the variable. For ordinal or interval variables, as you will see, the Cumulative Percent column can provide valuable clues about how a variable is distributed. But for nominal variables, which cannot be ranked, the Cumulative Percent column provides no information of value.

Now consider the Valid Percent column more closely. Scroll between the frequency distribution and the bar chart, which depicts the zodiac variable in graphic form (Figure 2-3). What is the mode, the most common astrological sign? For nominal variables, the answer to this question is (almost) always an easy call: Simply find the value with the highest percentage of responses. Leo is the mode. Does this variable have little dispersion or a

Figure 2-2 Frequencies Output (nominal variable)



Statistics

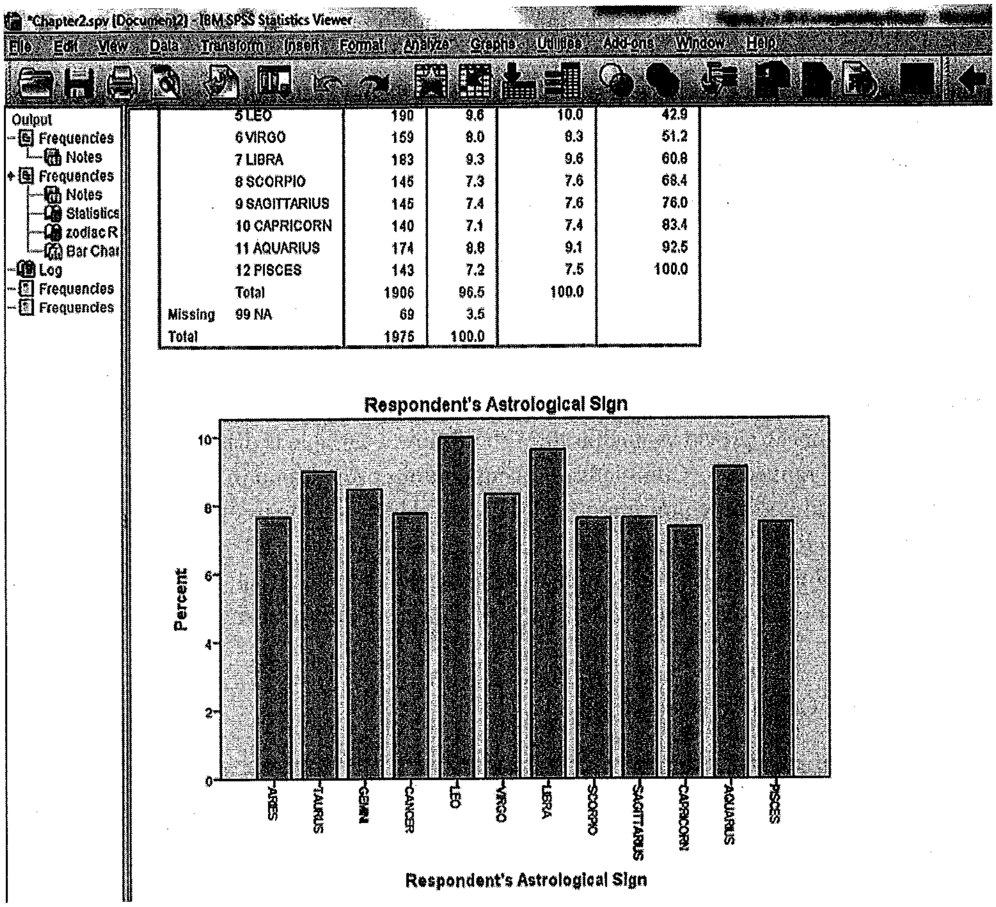
zodiac Respondent's Astrologic

N	Valid	1906
	Missing	69

zodiac Respondent's Astrological Sign

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 ARIES	146	7.4	7.6	7.6
	2 TAURUS	172	8.7	9.0	16.7
	3 GEMINI	161	8.2	8.5	25.1
	4 CANCER	148	7.5	7.8	32.9
	5 LEO	190	9.6	10.0	42.9
	6 VIRGO	159	8.0	8.3	51.2
	7 LIBRA	183	9.3	9.6	60.8
	8 SCORPIO	145	7.3	7.6	68.4
	9 SAGITTARIUS	146	7.4	7.6	76.0
	10 CAPRICORN	140	7.1	7.4	83.4
	11 AQUARIUS	174	8.8	9.1	92.5
	12 PISCES	143	7.2	7.5	100.0
	Total	1906	96.5	100.0	
Missing	99 NA	69	3.5		
Total		1975	100.0		

Figure 2-3 Bar Chart (nominal variable)


*Chapter2.spv [Document2] - IBM SPSS Statistics Viewer
File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help
Output
Frequencies
Notes
Frequencies
Notes
Statistics
zodiacR
Bar Char
Log
Frequencies
Frequencies
Respondent's Astrological Sign
Percent
10
8
6
4
2
0
ARIES
TAURUS
GEMINI
CANCER
LEO
VIRGO
LIBRA
SCORPIO
SAGITTARIUS
CAPRICORN
AQUARIUS
PISCES
Respondent's Astrological Sign

	5 LEO	190	9.6	10.0	42.9
	6 VIRGO	159	8.0	8.3	51.2
	7 LIBRA	183	9.3	9.6	60.8
	8 SCORPIO	145	7.3	7.6	68.4
	9 SAGITTARIUS	145	7.4	7.6	76.0
	10 CAPRICORN	140	7.1	7.4	83.4
	11 AQUARIUS	174	8.8	9.1	92.5
	12 PISCES	143	7.2	7.5	100.0
	Total	1906	96.5	100.0	
Missing	99 NA	69	3.5		
Total		1975	100.0		

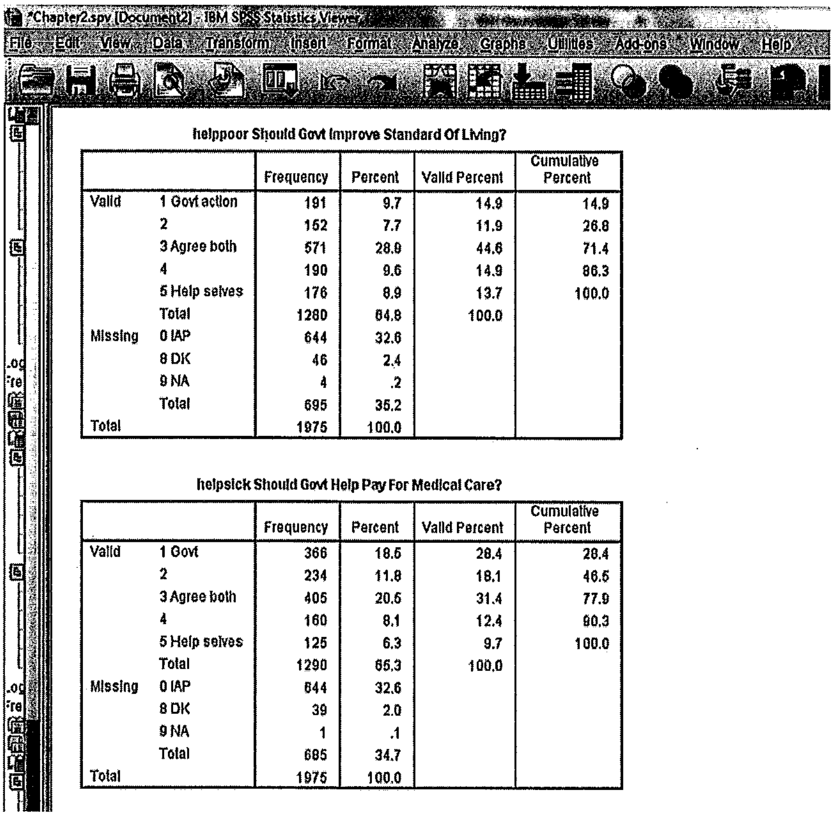
lot of dispersion? Again study the Valid Percent column and the bar chart. Apply the following rule: *A variable has no dispersion if the cases are concentrated in one value of the variable; a variable has maximum dispersion if the cases are spread evenly across all values of the variable.* Are most of the cases concentrated in Leo, or are there many cases in each value of zodiac? Because respondents are spread out—all astrological signs are about equally represented—you would conclude that zodiac has a high level of dispersion.

DESCRIBING ORDINAL VARIABLES

Next, you will analyze and describe two ordinal-level variables, one of which has little variation and the other of which is more spread out. Along the top menu bar of the Viewer, click Analyze → Descriptive Statistics → Frequencies. SPSS remembers the preceding analysis, so zodiac is still in the Variable(s) list. Click zodiac back into the left-hand list. Scroll through the list until you find these variables: helppoor and helpsick. Each of these is a 5-point ordinal scale. Helppoor asks respondents to place themselves on a scale between 1 ("The government should take action to help poor people") and 5 ("People should help themselves"). Helpsick, using a similar 5-point scale, asks respondents about government responsibility or individual responsibility for medical care. Click helppoor and helpsick into the Variable(s) list. SPSS retained your earlier settings for Charts, so accompanying bar charts will appear in the Viewer. Click OK.

SPSS runs the analysis for each variable and produces two frequency distributions, one for helppoor and one for helpsick, followed by two bar charts of the same information. First, focus on helppoor. Scroll back and forth between the frequency distribution (Figure 2-4) and the bar chart (Figure 2-5). Because helppoor is an ordinal variable, you can report both its mode and its median. Its mode, clearly enough, is the response "Agree with both," which contains 44.6 percent of the cases. What about the median? This is where the Cumulative Percent column of the frequency distribution comes into play. *The median for any ordinal (or interval) variable is the*

Figure 2-4 Frequencies Output (ordinal variables)


*Chapter2.spv [Document2] - IBM SPSS Statistics Viewer
File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help

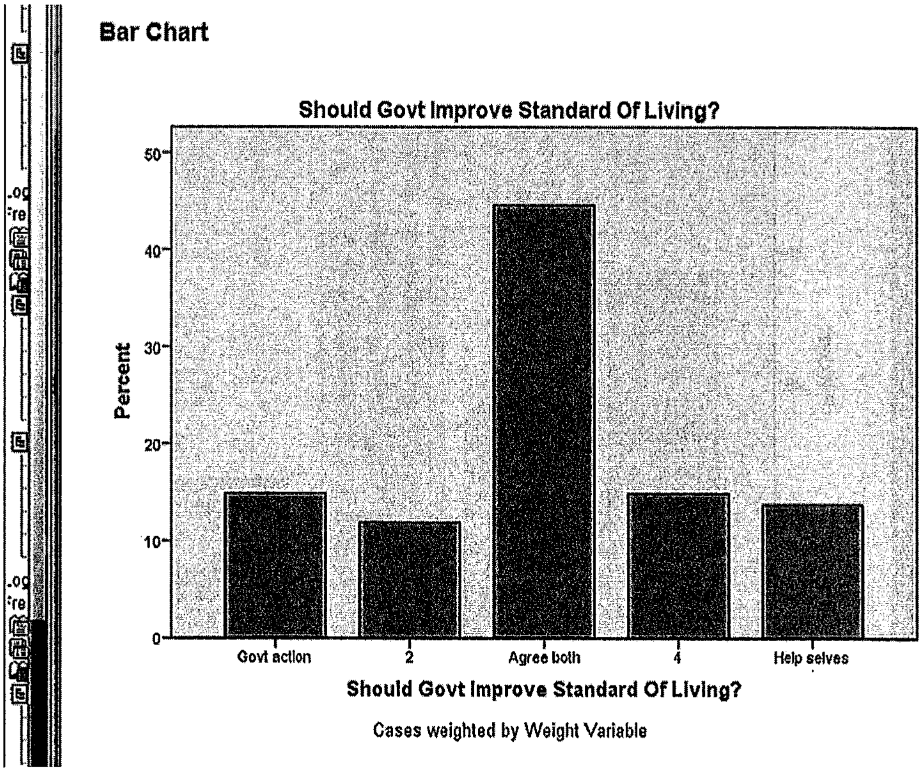
helppoor Should Govt Improve Standard Of Living?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Govt action	191	9.7	14.9	14.9
	2	152	7.7	11.9	26.8
	3 Agree both	571	28.9	44.6	71.4
	4	190	9.6	14.9	86.3
	5 Help selves	176	8.9	13.7	100.0
	Total	1280	64.8	100.0	
Missing	0 IAP	644	32.6		
	8 DK	46	2.4		
	9 NA	4	.2		
	Total	695	35.2		
Total		1975	100.0		

helpsick Should Govt Help Pay For Medical Care?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Govt	366	18.5	28.4	28.4
	2	234	11.8	18.1	46.5
	3 Agree both	405	20.5	31.4	77.9
	4	160	8.1	12.4	90.3
	5 Help selves	125	6.3	9.7	100.0
	Total	1290	65.3	100.0	
Missing	0 IAP	644	32.6		
	8 DK	39	2.0		
	9 NA	1	.1		
	Total	685	34.7		
Total		1975	100.0		

Figure 2-5 Bar Chart (ordinal variable with low dispersion)

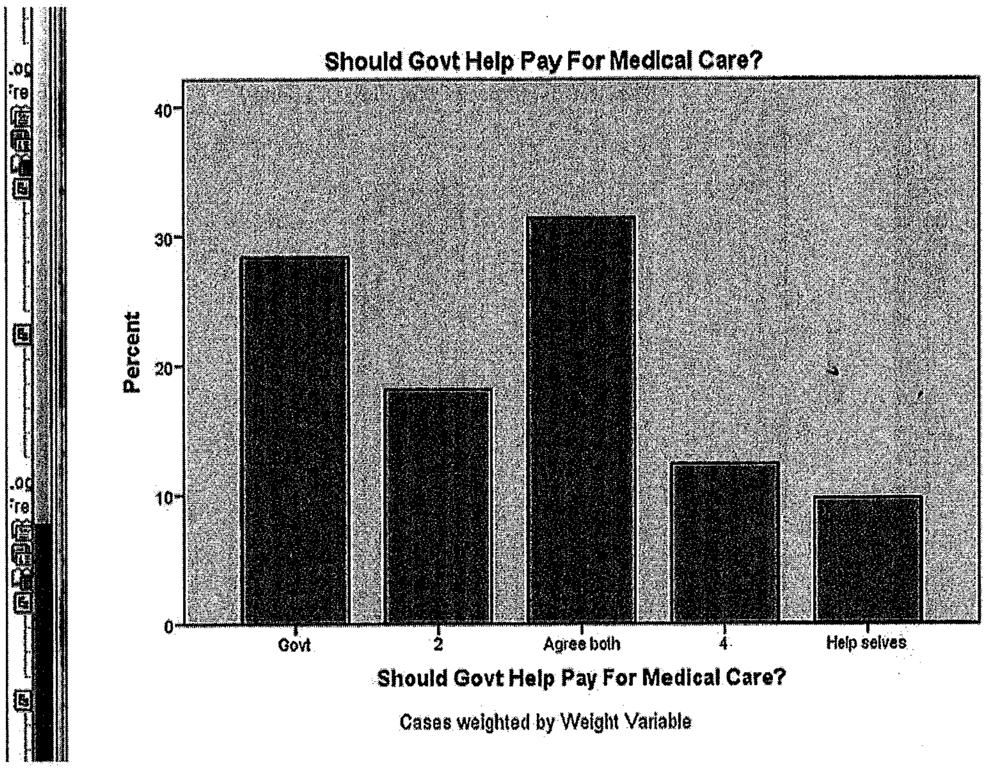


category below which 50 percent of the cases lie. Is the first category, "Govt action," the median? Nó, this code contains fewer than half of the cases. How about the next higher category? No, again. The Cumulative Percent column still has not reached 50 percent. The median occurs in the "Agree with both" category (cumulative percentage, 71.4).

Now consider the question of whether helppoor has a high degree of dispersion or a low degree of dispersion. If helppoor had a high level of variation, then the percentages of respondents in each response category would be roughly equal, much like the zodiac variable that you analyzed earlier. So, roughly one-fifth of the cases would fall into each of the five response categories: 20 percent in "Govt," 20 percent in response category "2," 20 percent in "Agree with both," 20 percent in response category "4," and 20 percent in "Help selves." If helppoor had no dispersion, then all the cases would fall into one value. That is, one value would have 100 percent of the cases, and each of the other categories would have 0 percent. Which of these two scenarios comes closest to describing the actual distribution of respondents across the values of helppoor? The equal-percentages-in-each-category, high variation scenario? Or the 100-percent-in-one-category, low variation scenario? It seems clear that helppoor is a variable with a relatively low degree of dispersion. "Agree with both," with 44.6 percent of the cases, contains nearly three times as many cases as its nearest rival ("Govt"), and more than three times as many cases as any of the other response categories.

Now contrast helppoor's distribution with the distribution of helpsick (Figure 2-6). Interestingly, helpsick has the same mode as helppoor ("Agree with both," with 31.4 percent of the cases), and the same median (again, "Agree with both," where the cumulative percentage exceeds 50.0). Yet, with helppoor it seemed reasonable to say that "Agree with both" was the typical response. Would it be reasonable to say that "Agree with both" is helppoor's typical response? No, it would not. Notice that, unlike helppoor, respondents' values on helpsick are more spread out, with sizable numbers of cases falling in the first value ("Govt," with 28.4 percent), making it a close rival to "Agree with both" for the distinction of being the modal opinion on this issue. Clearly, the public is more divided—more widely dispersed—on the question of medical assistance than on the question of assistance to the poor.

Figure 2-6 Bar Chart (ordinal variable with high dispersion)



DESCRIBING INTERVAL VARIABLES

Let's now turn to the descriptive analysis of interval-level variables. An interval-level variable represents the most precise level of measurement. Unlike nominal variables, whose values stand for categories, and ordinal variables, whose values can be ranked, the values of an interval variable *tell you the exact quantity of the characteristic being measured.* For example, age qualifies as an interval-level variable because its values impart each respondent's age in years.

Because interval variables have the most precision, they can be described more completely than can nominal or ordinal variables. For any interval-level variable, you can report its mode, median, and arithmetic average, or *mean.* In addition to these measures of central tendency, you can make more sophisticated judgments about variation. Specifically, you can determine if an interval-level distribution is *skewed.*

Skewness refers to the symmetry of a distribution. If a distribution is not skewed, the cases tend to cluster symmetrically around the mean of the distribution, and they taper off evenly for values above and below the mean. If a distribution is skewed, by contrast, one tail of the distribution is longer and skinnier than the other tail. Distributions in which some cases occupy the higher values of an interval variable—distributions with a skinnier right-hand tail—have a *positive skew.* By the same token, if the distribution has some cases at the extreme lower end—the distribution has a skinnier left-hand tail—then the distribution has a *negative skew.* Skewness affects the mean of the distribution. A positive skew tends to "pull" the mean upward; a negative skew pulls it downward. However, skewness has less effect on the median. Because the median reports the middlemost value of a distribution, it is not tugged upward or downward by extreme values. *For badly skewed distributions, it is a good practice to use the median instead of the mean in describing central tendency.*

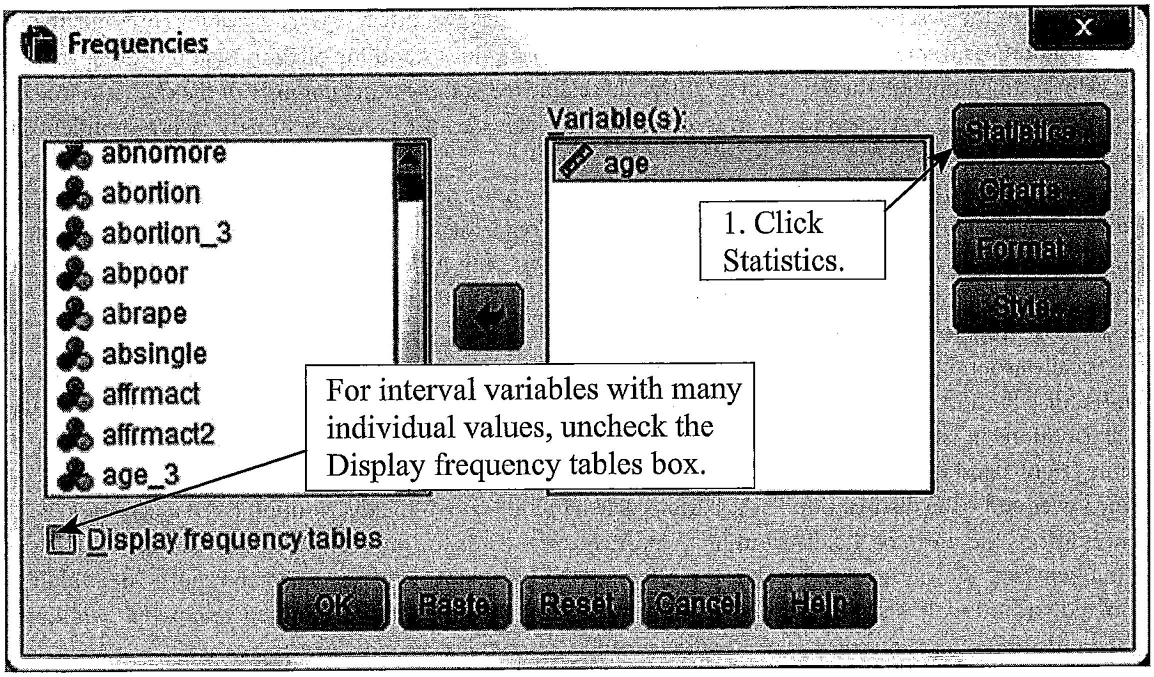
A step-by-step analysis of a GSS2012 variable, age, will clarify these points. Click Analyze → Descriptive Statistics → Frequencies. If helppoor and helpsick are still in the Variable(s) list, click them back into the left-hand list. Click age into the Variable(s) list. Click the Charts button. Make sure that Bar charts (under Chart Type) and Percentages (under Chart Values) are selected. Click Continue, which returns you to the main Frequencies window (Figure 2-7).

So far, this procedure is the same as in your analysis of zodiac, helppoor, and helpsick. When running a frequencies analysis of an interval-level variable, however, you need to do two additional things. One of these is a must-do. The other is a may-want-to-do. The must-do: Click the Statistics button in the Frequencies window, as shown in Figure 2-7. The Frequencies: Statistics window appears. In the Central Tendency panel, click the boxes next to Mean, Median, and Mode. In the Distribution panel, click Skewness. Click Continue, returning to the main Frequencies window. The may-want-to-do: *Un*check the box next to Display frequency tables, appearing at the foot of the left-hand list.[2] Click OK.

SPSS runs the analysis of age and dumps the requested statistics and bar chart into the Viewer (Figure 2-8). Most of the entries in the Statistics table are familiar to you: valid number of cases; number of missing cases; and mean, median, and mode. In addition, SPSS reports values for skewness and a statistic called standard error of skewness. When a distribution is perfectly symmetrical—no skew—it has skewness equal to 0. If the distribution has a skinnier right-hand tail—positive skew—then skewness will be a positive number. A skinnier left-hand tail, logically enough, returns a negative number for skewness. For the age variable, the skewness statistic is positive (.338). This suggests that the distribution has a skinnier right-hand tail—a feature that is confirmed by the shape of the bar chart. Note also that the mean (46.1 years) is higher than the median (45 years), a situation that often—although not always—indicates a positive skew.[3] Even so, the mean and median are only about 1 year apart. You have to exercise judgment, but in this case it would not be a distortion of reality to use the mean instead of the median to describe the central tendency of the distribution.[4]

All the guided examples thus far have used bar charts for graphic support. For nominal and ordinal variables, a bar chart should always be your choice. For interval variables, however, you may want to ask SPSS to produce a histogram instead. What is the difference between a bar chart and a histogram? A bar chart displays each value of a variable and shows you the percentage (alternatively, the raw number) of cases that fall into each category. A histogram is similar, but instead of displaying each discrete value, it collapses categories into ranges (called bins), resulting in a compact display. Histograms are sometimes more readable and elegant than bar charts. Most of the time a histogram will work just as well as a bar chart in summarizing an interval-level variable. For interval variables with a large number of values, a histogram is the graphic of choice. (Remember: For nominal or ordinal variables, you always want a bar chart.)

Figure 2-7 Requesting Statistics for an Interval Variable

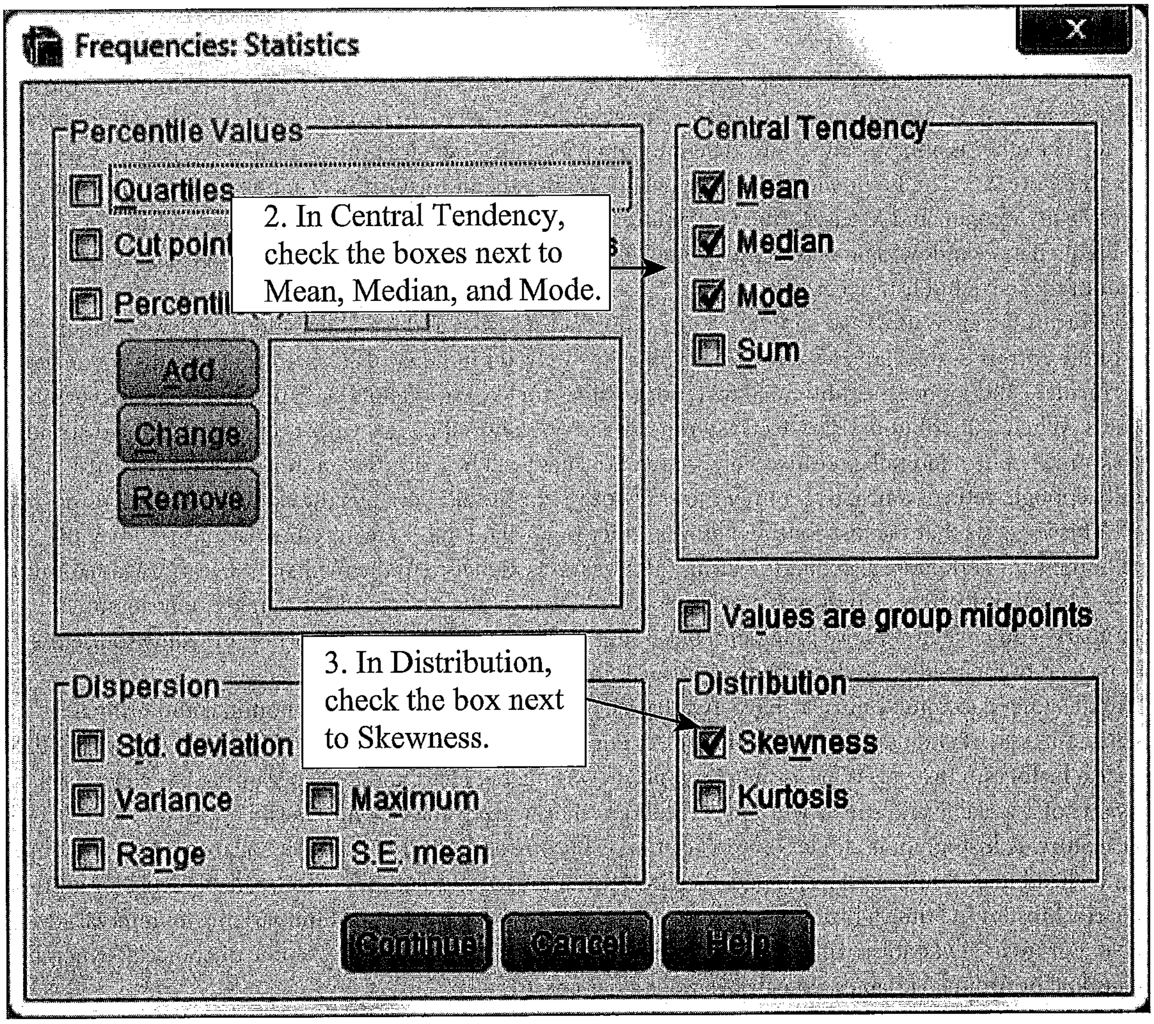


Variable(s): age

□ Display frequency tables

1. Click Statistics.

For interval variables with many individual values, uncheck the Display frequency tables box.



□ Quartiles

□ Cut point

□ Percentil

☑ Mean

☑ Median

☑ Mode

□ Sum

□ Values are group midpoints

□ Std. deviation

□ Variance

□ Range

□ Maximum

□ S.E. mean

☑ Skewness

□ Kurtosis

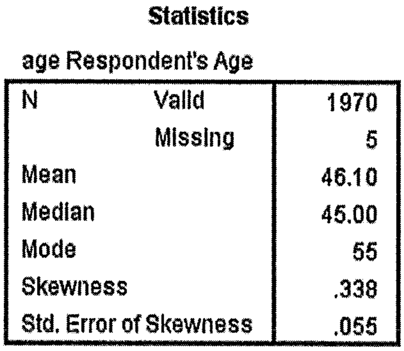
2. In Central Tendency, check the boxes next to Mean, Median, and Mode.

3. In Distribution, check the box next to Skewness.

So that you can become familiar with histograms, run the analysis of age once again—only this time ask SPSS to produce a histogram instead of a bar chart. Click Analyze → Descriptive Statistics → Frequencies. Make sure age is still in the Variable(s) list. Click Statistics, and then uncheck all the boxes: Mean, Median, Mode, and Skewness. Click Continue. Click Charts, and then select the Histograms radio button in Chart Type. Click Continue. For this analysis, we do not need a frequency table. In the Frequencies window, uncheck the Display frequency tables box. (Refer to Figure 2-7.) Click OK.

This is a bare-bones run. SPSS reports its obligatory count of valid and missing cases, plus a histogram for age (Figure 2-9). On the histogram's horizontal axis, notice the hash marks, which are spaced at 20-year intervals. SPSS has compressed the data so that each bar represents about 2 years of age rather than 1 year of age. Now scroll up the Viewer to the bar chart of age, which you produced in the preceding analysis. Notice that the histogram has smoothed out the nuance and choppiness of the bar chart, though it still captures the essential qualities of the age variable.

Figure 2-8 Statistics and Bar Chart (interval variable)



Statistics

age Respondent's Age

N	Valid	1970
	Missing	5
Mean		46.10
Median		45.00
Mode		55
Skewness		.338
Std. Error of Skewness		.055

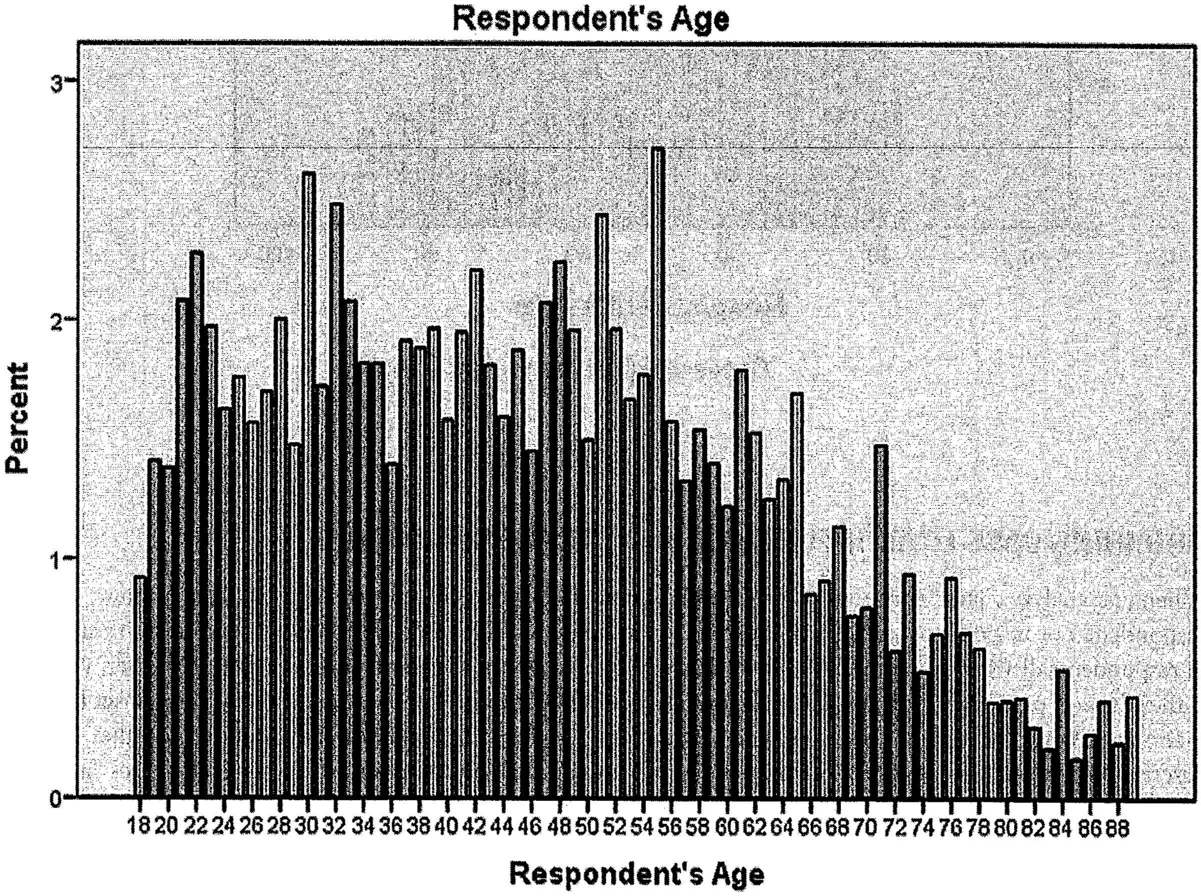
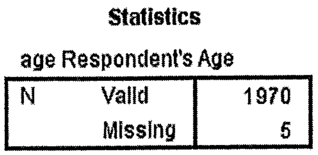


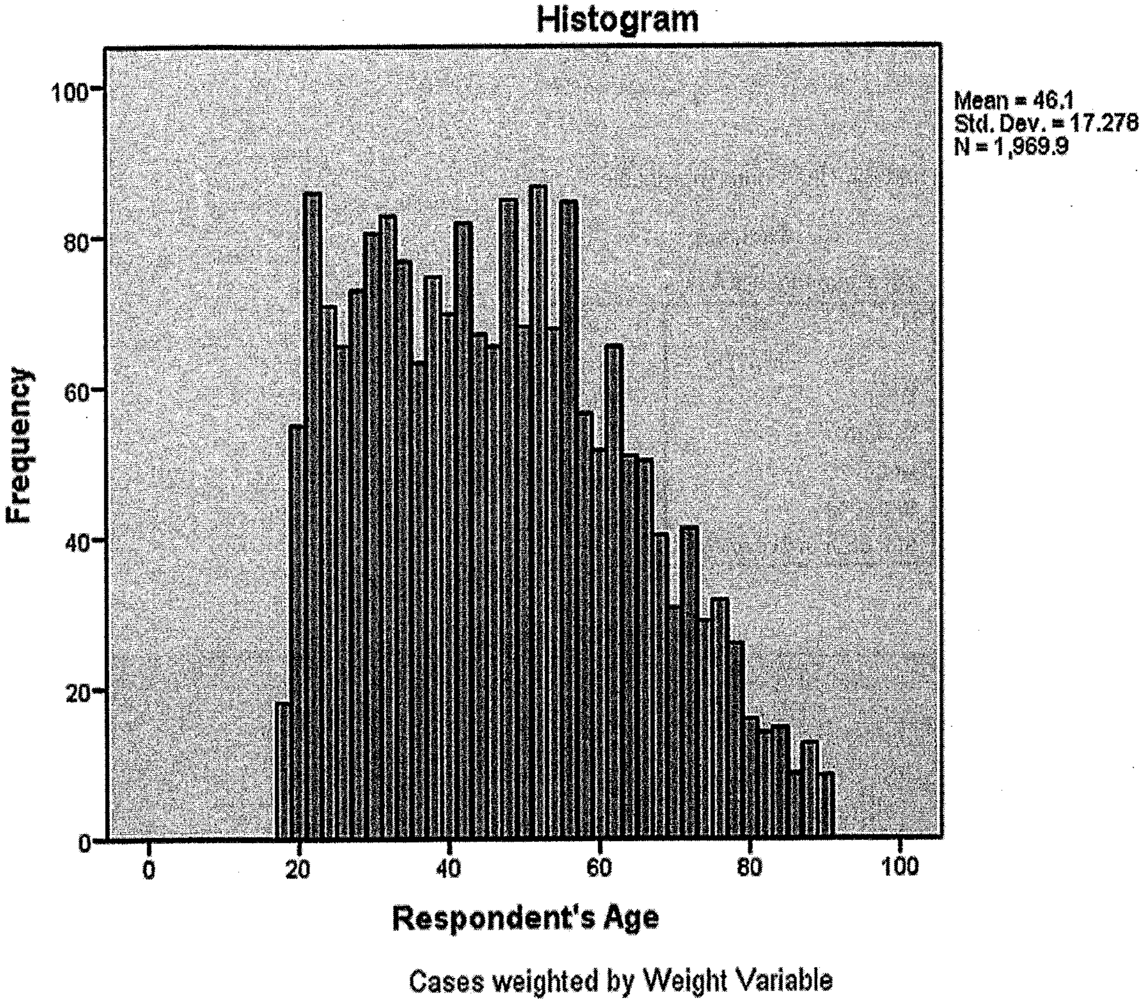
Figure 2-9 Histogram (interval variable)



Statistics

age Respondent's Age

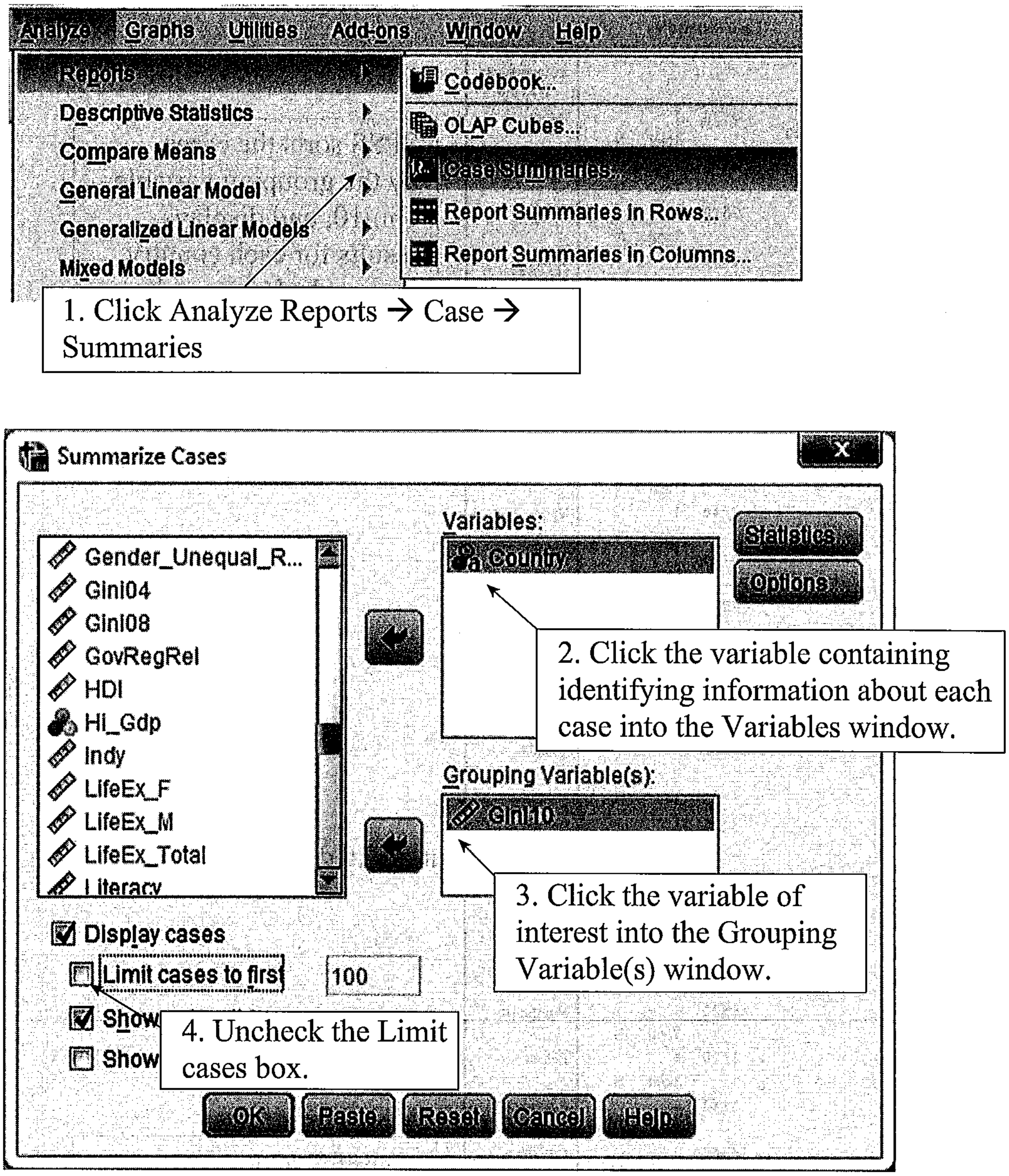
N	Valid	1970
	Missing	5



OBTAINING CASE-LEVEL INFORMATION WITH CASE SUMMARIES

When you analyze a large survey dataset, as you have just done, you generally are not interested in how respondent x or respondent y answered a particular question. Rather, you want to know how the entire sample of respondents distributed themselves across the response categories of a variable. Sometimes, however, you gather data on particular cases because the cases are themselves inherently important. The States dataset (50 cases) and World dataset (167 cases) are good examples. With these datasets, you may want to push the descriptions beyond the relative anonymity of Frequencies analysis and find out where particular cases "are" on an interesting variable. Analyze → Reports → Case Summaries is readymade for such elemental insights. Before beginning this guided example, close GSS2012 and open World.

Figure 2-10 Obtaining Case Summaries



1. Click Analyze Reports → Case → Summaries

2. Click the variable containing identifying information about each case into the Variables window.

3. Click the variable of interest into the Grouping Variable(s) window.

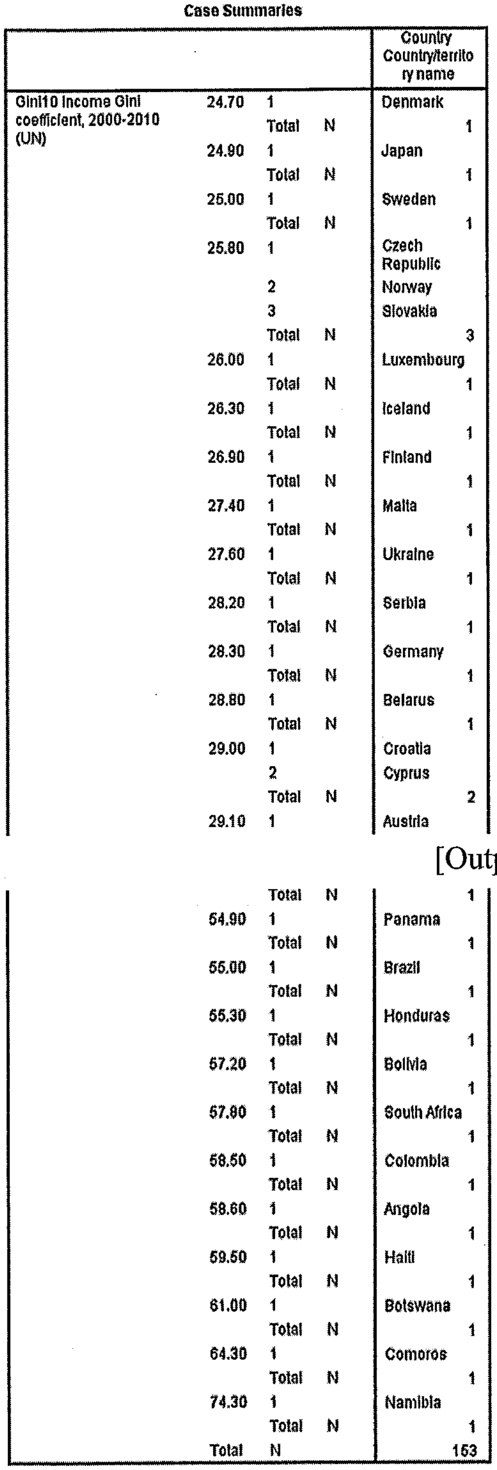
4. Uncheck the Limit cases box.

Suppose you are interested in identifying the countries that have the most equitable distribution of wealth, as well as those in which wealth is more concentrated in the hands of a few. The World dataset contains Gini10, the Gini coefficient for each country. The Gini coefficient measures wealth distribution on a scale that ranges from 0 (wealth is distributed equitably) to 100 (wealth is distributed inequitably). Exactly which countries are the most equitable? Which are the least equitable? Where does the United States fall on the list? Case Summaries can quickly answer questions like these. SPSS will sort the countries on the basis of a "grouping variable" (in this example, Gini10) and then produce a report telling you which countries are in each group.

With the World dataset open, click Analyze → Reports → Case Summaries. The Summarize Cases window opens (Figure 2-10). You need to do three things here:

1. Click the variable containing the cases' identities into the Variables window. In the World dataset, this variable is named Country, an alphabetic descriptor of each country's name.

Figure 2-11 Case Summaries Output



```
                              Case Summaries
                                                     Country
                                                     Country/territo
                                                     ry name
Gini10 Income Gini      24.70   1                    Denmark
coefficient, 2000-2010          Total   N                   1
(UN)                    24.90   1                    Japan
                                Total   N                   1
                        25.00   1                    Sweden
                                Total   N                   1
                        25.80   1                    Czech
                                                     Republic
                                2                    Norway
                                3                    Slovakia
                                Total   N                   3
                        26.00   1                    Luxembourg
                                Total   N                   1
                        26.30   1                    Iceland
                                Total   N                   1
                        26.90   1                    Finland
                                Total   N                   1
                        27.40   1                    Malta
                                Total   N                   1
                        27.60   1                    Ukraine
                                Total   N                   1
                        28.20   1                    Serbia
                                Total   N                   1
                        28.30   1                    Germany
                                Total   N                   1
                        28.80   1                    Belarus
                                Total   N                   1
                        29.00   1                    Croatia
                                2                    Cyprus
                                Total   N                   2
                        29.10   1                    Austria
```

[Output omitted]

```
                                Total   N                   1
                        54.90   1                    Panama
                                Total   N                   1
                        55.00   1                    Brazil
                                Total   N                   1
                        55.30   1                    Honduras
                                Total   N                   1
                        57.20   1                    Bolivia
                                Total   N                   1
                        57.80   1                    South Africa
                                Total   N                   1
                        58.50   1                    Colombia
                                Total   N                   1
                        58.60   1                    Angola
                                Total   N                   1
                        59.50   1                    Haiti
                                Total   N                   1
                        61.00   1                    Botswana
                                Total   N                   1
                        64.30   1                    Comoros
                                Total   N                   1
                        74.30   1                    Namibia
                                Total   N                   1
                        Total   N                         153
```

SPSS sorts the cases by the grouping variable, Gini10, and displays results for each country. Denmark, Japan, and Sweden are the three countries with the lowest values on Gini10.

Namibia, Comoros, and Botswana are the three countries with the highest values on Gini10.

2. Click the variable you are interested in analyzing, Gini10, into the Grouping Variable(s) window.
3. Uncheck the Limit cases box. This is important. If this box is left checked, SPSS will limit the analysis to the first 100 cases, which in many instances, such as the World dataset, will produce an incomplete analysis.

Click OK and consider the output (Figure 2-11). SPSS sorts the cases on the grouping variable, Gini10, and tells us which country is associated with each value of Gini10. For example, Denmark, with a Gini coefficient of

24.70, is the country having the most equitable wealth distribution. Which countries rank highest on Gini10? Scroll to the bottom of the tabular output. With a Gini coefficient of 74.30, Namibia is the country having the least equitable distribution of wealth.

EXERCISES

1. (Dataset: World. Variables: Women13, Country.) What percentage of members of the U.S. House of Representatives are women? In 2013 the number was 17.8 perçent, according to the Inter-Parliamentary Union, an international organization of parliaments.[5] How does the United States compare to other democratic countries? Is 17.8 percent comparatively low, comparatively high, or average for a typical national legislature? World contains Women13, the percentage of women in the lower house of the legislature in each of 90 democracies. Perform a frequencies analysis on Women13. In Statistics, obtain the mean and the median. In Charts, Chart Type, select Histogram. In the main Frequencies window, make sure that the Display frequency tables box is checked.

   A. The mean of Women13 is equal to (fill in the blank) ___________. The median is equal to ____________.

   B. Analysts generally prefer to use the mean to summarize a variable's central tendency, except in cases where the mean gives a misleading indication of the true center of the distribution. Make a considered judgment. For Women13, can the mean be used, or should the median be used instead? (circle your answer)

   Mean Median

   Explain your answer. ____

   ____

   ____

   C. Recall that 17.8 percent of U.S. House members are women. Suppose a women's advocacy organization vows to support female congressional candidates so that the U.S. House might someday "be ranked among the top one-fourth of democracies in the percentage of female members." According to the frequencies analysis, to meet this goal, women would need to constitute what percentage of the House? (circle one)

   About 21 percent About 25 percent About 28 percent

   D. Print the histogram. Basing your answer on the shape of the histogram, would you say that Women13 has a negative skew or a positive skew? (circle your answer)

   Negative skew Positive skew

   Briefly explain your answer.

   ____

   ____

   E. Run Analyze → Reports → Case Summaries. Click Country into the Variables box and Women13 into the Grouping Variable(s) box. Make sure to uncheck the box next to Limit cases to first 100. Examine the output. Which five countries have the lowest percentages of women in their legislatures?

   1.______________ 2. ______________ 3. ______________ 4.______________ 5.______________

   Which five countries have the highest percentages of women in their legislatures?

   1.______________ 2. ______________ 3. ______________ 4.______________ 5.______________

2. (Dataset: GSS2012. Variables: science_quiz, wordsum.) The late Carl Sagan once lamented: "We live in a society exquisitely dependent on science and technology, in which hardly anyone knows anything about

science and technology." Do the data support Sagan's pessimistic assessment? How does the public's grasp of scientific facts compare with other skills, such as word recognition and vocabulary?

GSS2012 contains science_quiz, which was created from 10 questions testing respondents' knowledge of basic scientific facts. Values on science_quiz range from 0 (the respondent did not answer any of the questions correctly) to 10 (the respondent correctly answered all 10).[6] GSS2012 also contains wordsum, which measures respondents' knowledge of the meanings of 10 words. Like science_quiz, wordsum ranged from 0 (the respondent did not know any of the words) to 10 (the respondent knew all 10 words).

A. Obtain frequency distributions and bar charts for science_quiz and wordsum. In Statistics, request mean, median, and mode. In Charts, request bar charts with percentages. Fill in the following table:

	science_quiz	wordsum
Mean	?	?
Median	?	?
Mode	?	?

B. Consider the following Sagan-esque statement: "The public knows more about words than about science." Based on your results in part A, is this statement correct or incorrect? (circle one)

Correct Incorrect

Explain your reasoning, making specific reference to the statistics you reported in A.

____

____

____

____

C. Examine the frequency distributions. According to conventional academic standards, scores of 9 or 10 on a 10-point quiz would be A's. What percentage of respondents would receive a grade of A on science_quiz? (fill in the blank) ____________. What percentage of respondents would receive a grade of A on wordsum? (fill in the blank) ____________.

D. Now turn your attention to the bar charts. Compare the science_quiz chart with the wordsum chart and think about the variation—how respondents are dispersed across the values of each variable. Consider this statement: "Science_quiz has a greater degree of dispersion than wordsum." Is this statement correct or incorrect? (circle one)

Correct Incorrect

Explain your reasoning, making specific reference to the bar charts.

____

____

____

____

E. Print the bar charts that you created for this exercise.

3. (Dataset: GSS2012. Variable: femrole.) Two pundits are arguing about how the general public views the role of women in the home and in politics.