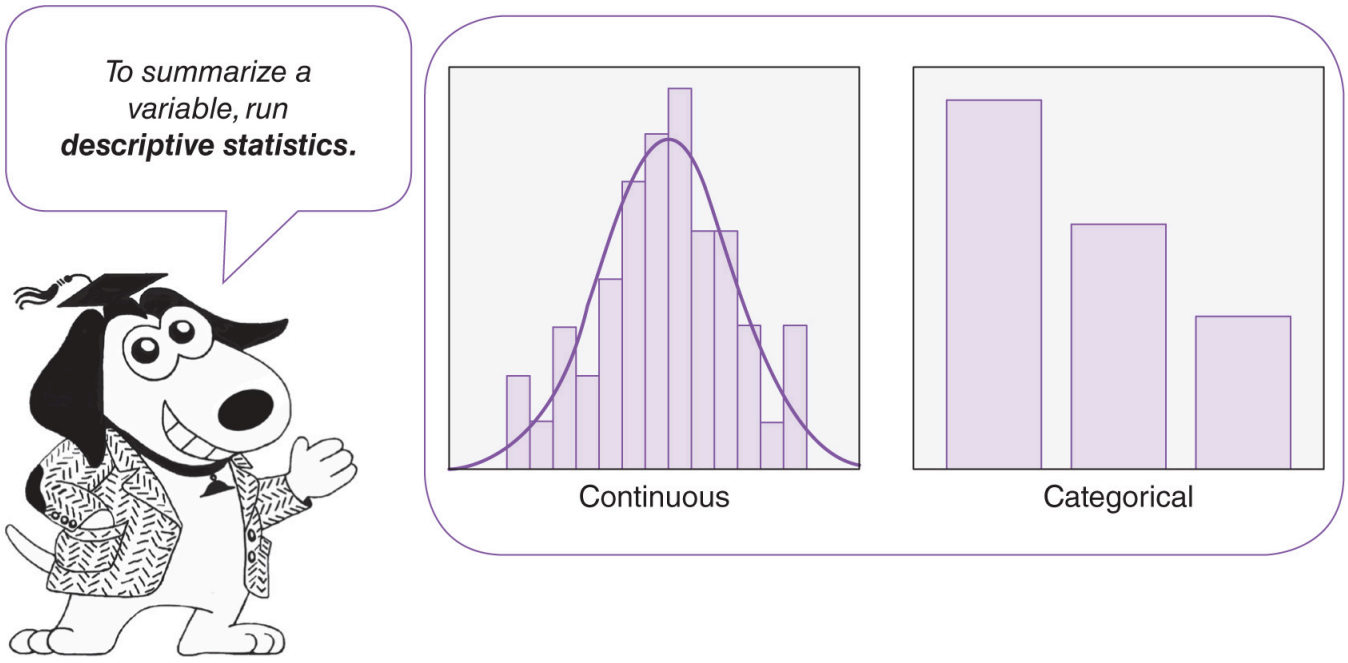
Sage Research Methods

Introductory Statistics Using SPSS

For the most optimal reading experience we recommend using our website. A free-to-view version of this content is available by clicking on this link, which includes an easy-to-navigate-and-search-entry, and may also include videos, embedded datasets, downloadable datasets, interactive questions, audio content, and downloadable tables and resources.

Author: Herschel Knapp
Pub. Date: 2022
Product: Sage Research Methods
DOI: https://doi.org/10.4135/9781071878910
Methods: T-test, Descriptive statistics, Normal distribution
Disciplines: Psychology, Mathematics
Access Date: July 19, 2024
Publisher: SAGE Publications, Inc
City: Thousand Oaks
Online ISBN: 9781071878910

Descriptive Statistics



Whenever you can, count.

—Sir Francis Galton

Learning Objectives

Upon completing this chapter, you will be able to:

- Comprehend the meaning of each descriptive statistic: number (*n*), mean (μ), median, mode, standard deviation (SD), variance, minimum, maximum, and range
- Understand central tendency
- Load an SPSS data file
- Order and interpret descriptive statistics for continuous variables: statistics table, histogram with normal curve, and skewed distribution
- Order and interpret descriptive statistics for categorical variables: statistics table, bar chart, and pie chart
- Select records to process or rule out

- Select all records to process

Videos

The videos for this chapter are **Ch 04 - Descriptive Statistics - Continuous.mp4** and **Ch 04 - Descriptive Statistics - Categorical.mp4**. These videos provide guidance on setting up, processing, and interpreting descriptive (summary) statistics for continuous and categorical variables using the data set **Ch 04 - Example 01 - Descriptive Statistics.sav**.

Overview—Descriptive Statistics

Statistics is about understanding data, which could consist of one or many groups and involve data sets of virtually any size. Take a look at the data set in Table 4.1; if someone were to ask you to describe this data set, you might say, “It looks like a list of about half women and half men, mostly in their 20s,” which would not be wrong, but we can be more precise than that. This list consists of (only) 50 records and 100 data items, but how would you make a viable summary statement if the list consisted of 100, 1,000, or even 100,000 records? Such data would span multiple pages and defy cursory visual inspection.

To better comprehend and communicate the nature of a data set, we use *descriptive statistics*, sometimes

referred to as *summary statistics*. Descriptive statistics enable us to concisely understand a data set of any size using a handful of figures and simple graphs that serve to summarize the contents of a variable.

Continuous variables can be summarized using nine descriptive statistics: number (*n*), mean, median, mode, standard deviation, variance, minimum, maximum, and range. Graphically, continuous variables can be depicted using a histogram (a kind of bar chart) with a normal curve.

Table 4.1 Data Set Containing 50 Records With Two Variables per Record: gender and age.

Male	24	Male	30	Female	22	Male	26	Female	25
Male	25	Male	25	Male	25	Male	25	Female	25
Male	31	Female	26	Male	25	Female	24	Female	22
Female	19	Male	27	Female	24	Male	29	Male	22
Female	27	Male	24	Female	23	Male	23	Female	20
Male	20	Female	25	Male	25	Female	21	Female	22
Male	28	Male	24	Female	19	Female	22	Male	18
Female	23	Female	26	Female	23	Female	21	Female	18
Male	26	Female	23	Female	28	Female	24	Female	23
Male	24	Male	22	Female	24	Male	26	Female	27

Categorical variables can be summarized using number (*n*) and percentage. Graphically, categorical variables are depicted using a simple bar chart or pie chart. You can see examples of a histogram with a normal curve for a continuous variable and a bar chart for a categorical variable on the first page of this chapter.

We will begin with an explanation of the nine summary statistics used to analyze continuous variables. For

simplicity, we will work with a small data set—the first 10 ages drawn from the second column of Table 4.1: 24, 25, 31, 19, 27, 20, 28, 23, 26, and 24.

Descriptive Statistics

Number (*n*)

The most basic descriptive statistic is the number, represented by the letter *n*. To compute the *n*, simply count the number of elements (numbers) in the sample; in this case, there are 10 elements: 24 is the first, 25 is the second, 31 is the third, and so on through 24 (at the end), the tenth, so *n* = 10.

A lowercase *n* denotes the number of elements in a sample, whereas an uppercase *N* denotes the number of elements in the (whole) population. SPSS output reports always use the capital *N*. Since it is rare to be processing a data set consisting of an entire population, it is considered good practice to use the lowercase *n* in your documentation, as such "*n* (age) = 10."

Mean (μ)

In statistical language, the *average* is referred to as the *mean*. The calculation for the mean is the same as for the average: Add up all the numbers and then divide that figure by the total number of values involved (*n* = 10):

Mean (*age*) = (24 + 25 + 31 + 19 + 27 + 20 + 28 + 23 + 26 + 24) ÷ 10

Mean (*age*) = 247 ÷ 10

Mean (*age*) = 24.7

μ (*age*) = 24.7

The mean can be written using the lowercase Greek letter μ (pronounced *m-you*) or as the variable name with a horizontal bar over it; hence, the mean may be documented as such:

μ (*age*) = 24.7

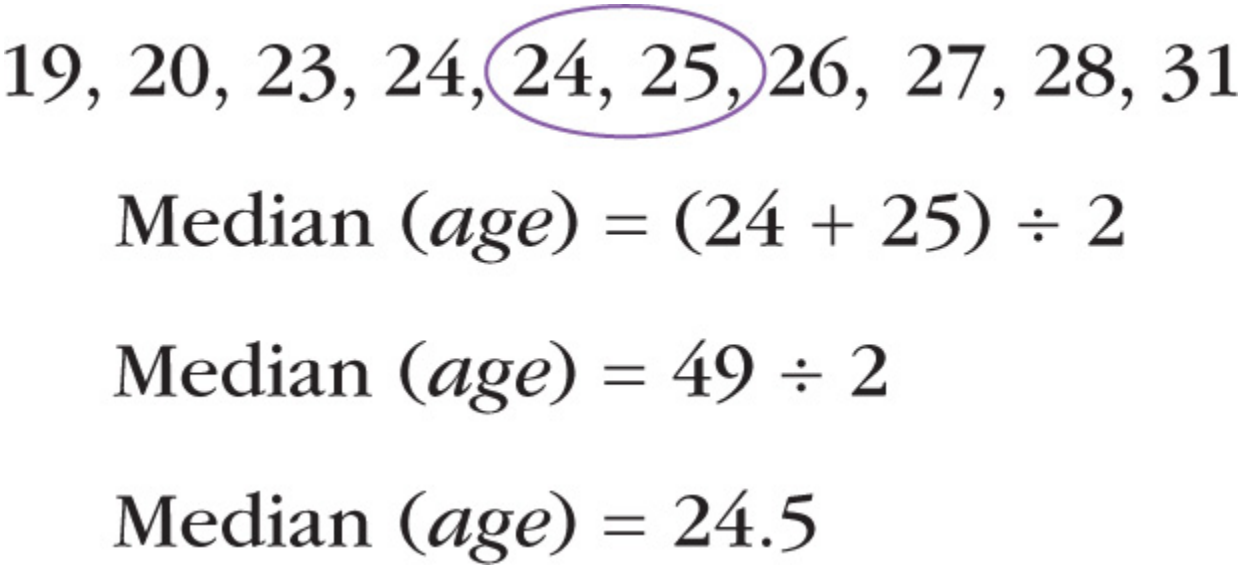
$\overline{age}$ = 24.7

For consistency throughout the rest of this text, the mean will be documented using the more common "μ (*age*)" style.

Median

The median is the middle value of a variable. Think of the term *median* in terms of a street—the median is in the middle; it splits the street in half. To find the median, arrange the data in the variable from lowest to highest and then select the middle value(s). In smaller data sets, the mean can be altered substantially by outlier scores—scores that are unexpectedly high or low. In such instances, the median can provide a more stable indicator of the central value than the mean.

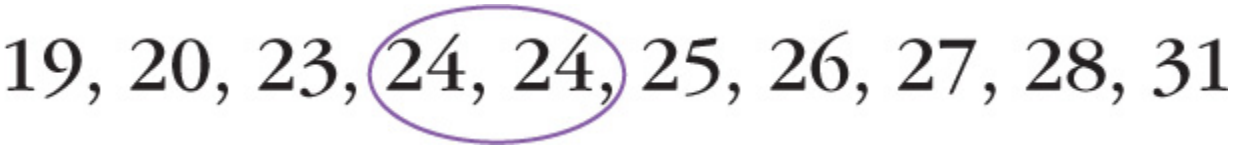
When the *n* is even, as in the data set below (*n* = 10), there are two middle numbers: 24 and 25. The median is the mean of these two middle numbers:



When the *n* is odd, as in the small data set below (*n* = 5), there is (only) one middle number—hence, the median is simply the (one) middle number: 86.

Mode

The mode is the most common number in the data set. Notice that ***mo**de* and ***mo**st* share their first two letters. In this case, we see that each number in this data set is present only once, except for 24, which occurs twice; hence the mode is 24.



It is possible for a data set to have more than one mode; the example below has two modes, 24 and 31, since both occur most frequently (there are two 24's and two 31's; all the other numbers appear just once). Such a variable would be referred to as *bimodal*, meaning two modes.

Although it is relatively rare, a variable may have more than two modes, which would be referred to as *multimodal*.

When SPSS detects more than one mode within a variable, it reports only the lowest one and provides a footnote indicating that there is more than one mode.

The mean, median, and mode are referred to as *measures of central tendency*, as they suggest the *center point* of a variable, where most of the data typically reside. Usually, the mean is exhibited as the best representation of the central values within a variable. In cases where the *n* is relatively small, the mean can be unstable, in that one or several outliers (extremely low or high values) can radically influence the mean, in which case the median is typically considered more stable, as it is less vulnerable to the presence of such outliers. The mode is generally of interest, as it indicates the most frequent value (score) within a variable.

Standard Deviation (SD)

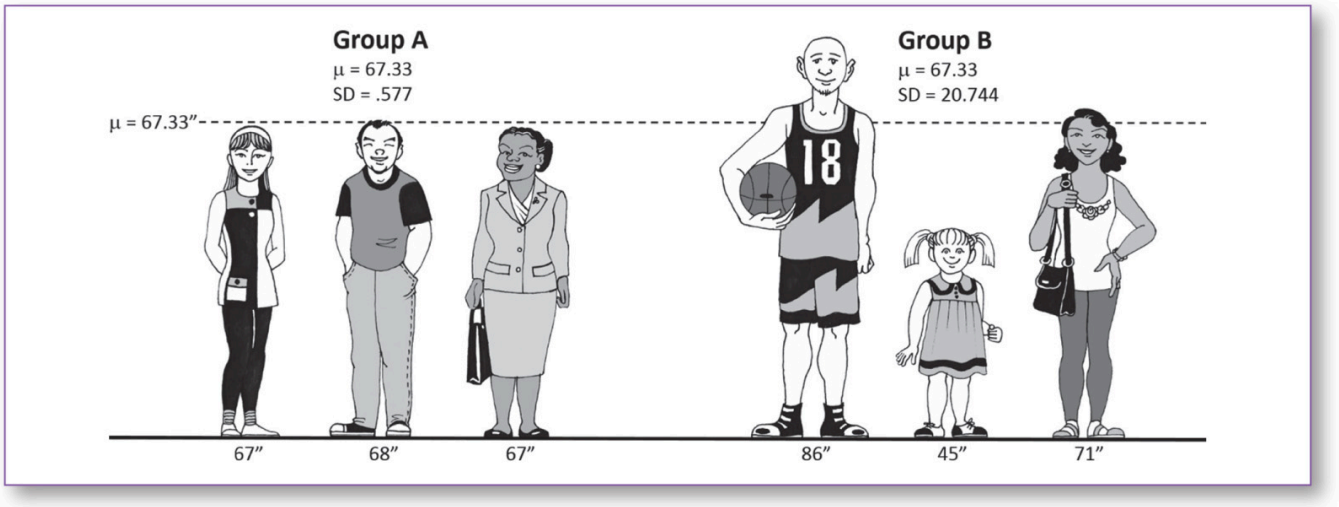
The standard deviation (SD) indicates the dispersion of the numbers within a variable. If a variable contains numbers that are fairly similar to one another, this produces a low(er) standard deviation; conversely, if there is more variety in the numbers, this renders a high(er) standard deviation.

Take a look at Figure 4.1. First, notice that the three people in Group A are all about the same height (67", 68", and 67"); their heights are just slightly above or below the mean (μ = 67.33). Statistically speaking, their heights do not *deviate* much from the mean; hence, this group produces a low(er) standard deviation (SD = .577).

Now, focus on the three people in Group B. Note that they have very diverse heights (86", 45", and 71"); their heights are fairly far apart from one another—substantially above or below the mean (μ = 67.33). Statistically speaking, their heights *deviate* much more from the mean, producing a high(er) standard deviation (SD = 20.744), more than 35 times the standard deviation for Group A (SD = .577).

For clarity, the heights for Groups A and B have been set to produce the same mean (μ = 67.33). The point is, if all we had was the mean for each group, we might get the (wrong) impression that the heights in Group A are just like the heights in Group B. The standard deviation statistic tells us if the numbers contained within a variable deviate a little from the mean, as in Group A, wherein the heights are fairly similar to the mean (and to one another), or if the numbers deviate more from the mean (and from one another), as in Group B, wherein the heights are more different from one another.

Figure 4.1 Standard deviation (SD) illustrated: Low(er) diversity (Group A) renders a lower standard deviation, and higher diversity (Group B) renders a high(er) standard deviation.



In statistical reporting, the standard deviation is often presented with the mean (e.g., "μ = 67.33 [SD = .577]").

Variance

The variance is simply the standard deviation squared. For example, we will calculate the variance of the heights for Group B, where SD = 20.774.

Variance (*height*) = [standard deviation (*height*)]^2

Variance (*height*) = 20.774^2

Variance (*height*) = 20.774 × 20.774

Variance (*height*) = 431.559

The variance is seldom included in statistical reports; it is used primarily as a term within other statistical formulas.

Minimum

The minimum is the smallest number in a variable. In the data set below, the minimum is 19.

19, 20, 23, 24, 24, 25, 26, 27, 28, 31

Maximum

The maximum is the largest number in a variable. In the data set below, the maximum is 31.

19, 20, 23, 24, 24, 25, 26, 27, 28, 31

Identifying the minimum and maximum values has some utility, but try not to bring inappropriate suppositions to your interpretation of such figures—bigger is not necessarily better. The meaning of the minimum and maximum values depends on the nature of the variable. For example, high bowling scores are good, while low golf scores are good, and high (or low) phone numbers are neither good nor bad.

Range

The range is the span of a data set; the formula for the range is **maximum** − **minimum**. In the data set below, we would calculate 31 − 19 = 12; the range is thus 12 (years).

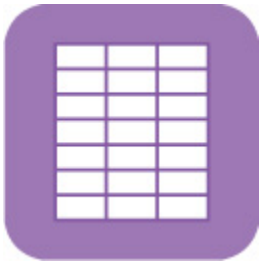
19, 20, 23, 24, 24, 25, 26, 27, 28, 31

SPSS—Loading an Spss Data File

For clarity, the examples used thus far have involved only 10 data items (n = 10). Now it is time to use SPSS to process descriptive statistics using the entire data set consisting of 50 records and both variables (*gender* and *age*).

Run SPSS

Data Set

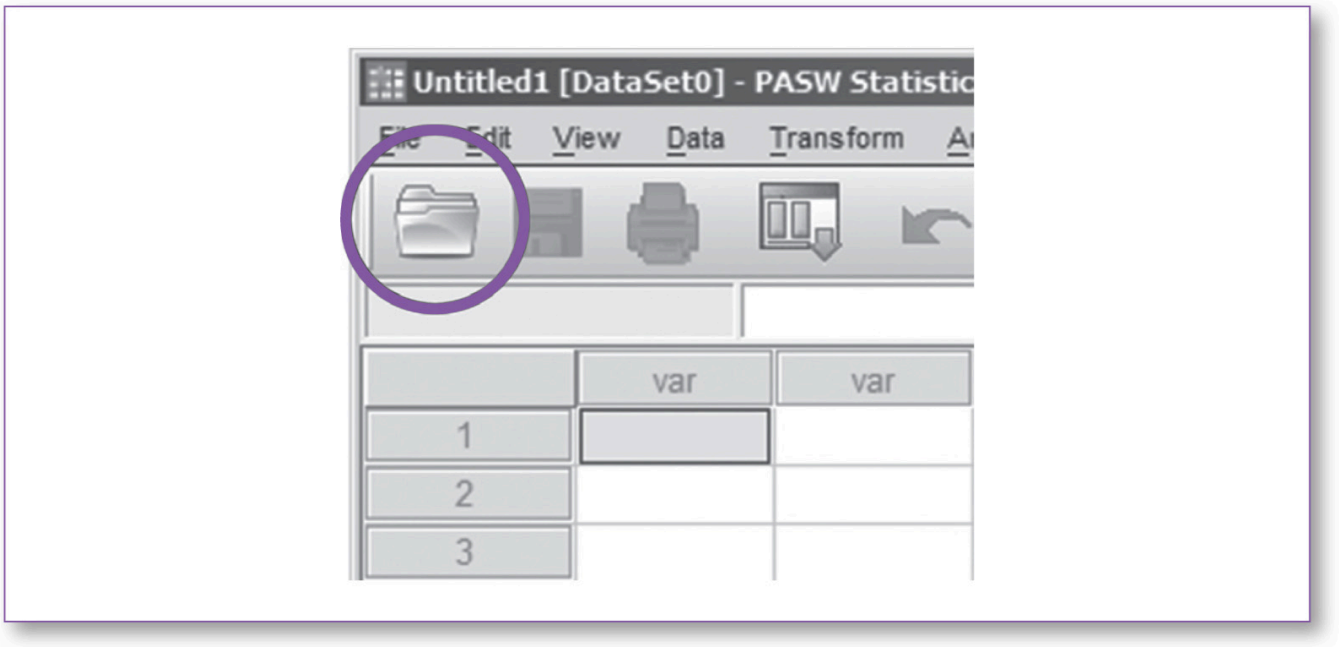


Use the *Open Data Document* icon (Figure 4.2) to load the SPSS data file **Ch 04 - Example 01 - Descriptive Statistics.sav**.

Codebook

Variable: gender
Definition: Gender of participant
Type: Categorical (1 = Female, 2 = Male)
Variable: age
Definition: Age of participant
Type: Continuous

Figure 4.2 Open Data Document icon.



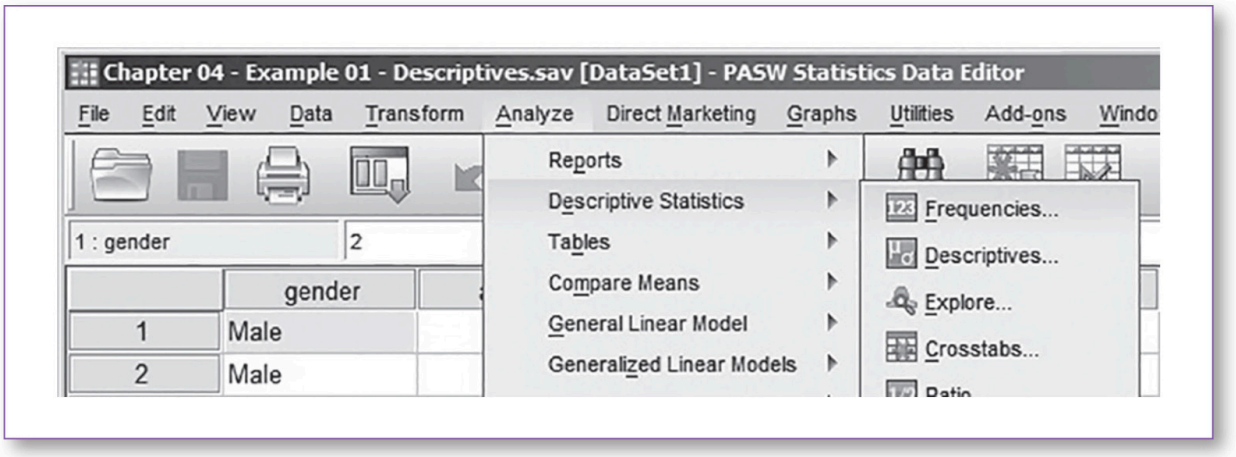
Test Run

SPSS—Descriptive Statistics: Continuous Variables (*Age*)

There are two types of variables in this SPSS file: *age* is a continuous variable, and *gender* is a categorical variable. In this section, we will process descriptive statistics for the continuous variable (*age*); later in this chapter, we will process the categorical variable (*gender*).

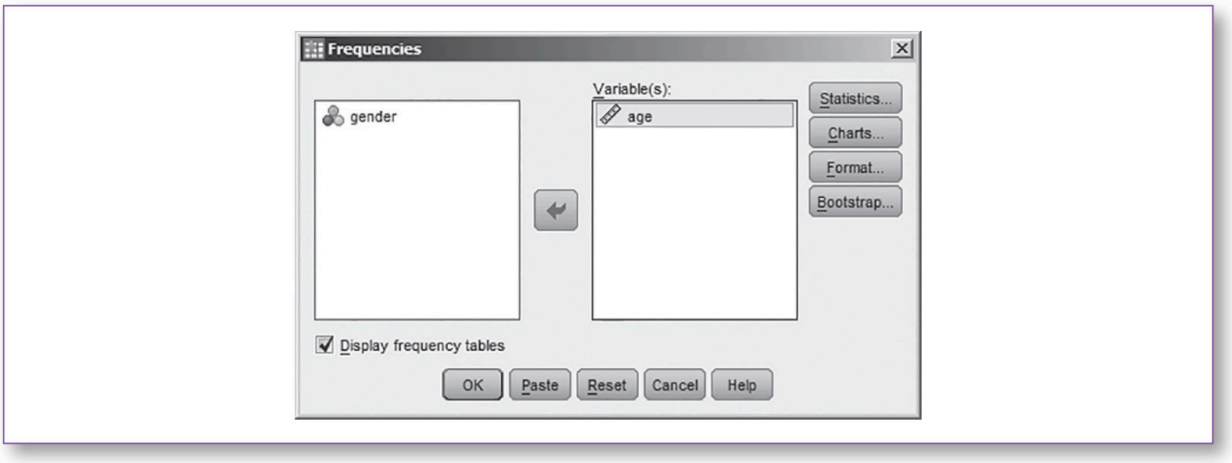
1. After loading the data, click on *Analyze*, *Descriptive Statistics*, *Frequencies* (Figure 4.3).

Figure 4.3 Running a descriptive statistics report; click on Analyze, Descriptive Statistics, Frequencies.



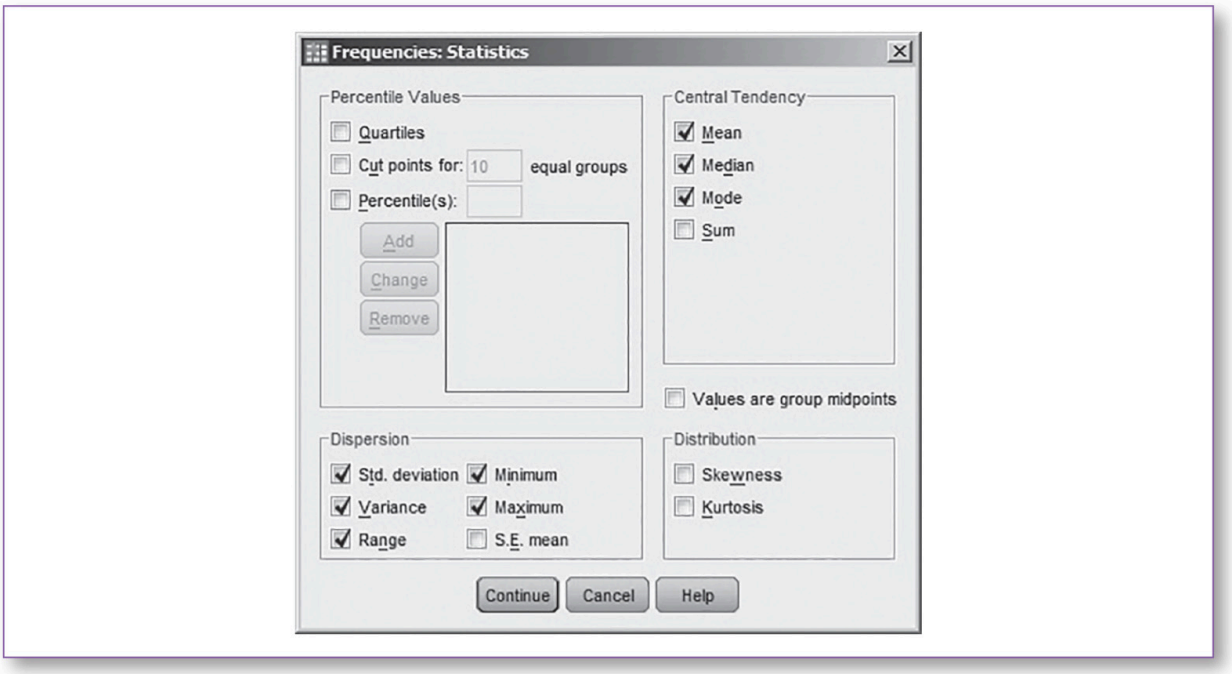
2. SPSS will then prompt you to select the variable(s) you would like to process. Move the *age* variable to the *Variable(s)* panel (Figure 4.4).

Figure 4.4 Move the variable(s) to be analyzed (age) from the left panel to the right Variable(s) panel.



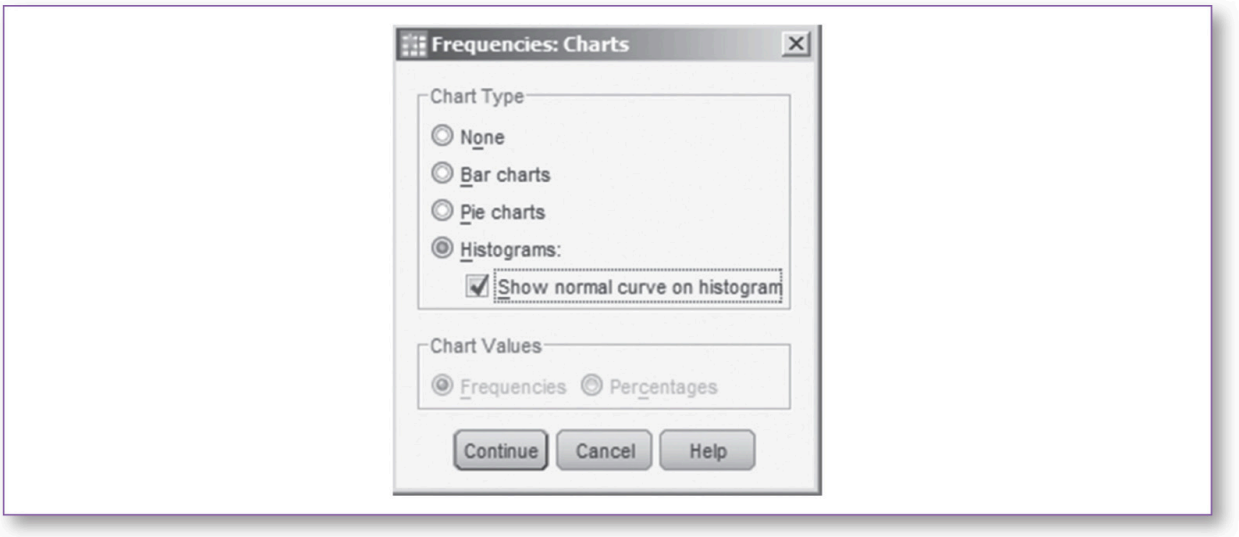
3. Click on the *Statistics* button.

Figure 4.5 Frequencies: Statistics window.



4. Select the descriptive statistics indicated by the checkboxes (Figure 4.5): *Mean*, *Median*, *Mode*, *Std. deviation*, *Variance*, *Range*, *Minimum*, and *Maximum*.
5. Click on the *Continue* button. This will take you back to the *Frequencies* window (Figure 4.4).

Figure 4.6 Frequencies: Charts window; select Histograms and Show normal curve on histogram.



6. In the *Frequencies* window (Figure 4.4), click on the *Charts* button.
7. Select *Histograms* and *Show normal curve on histogram* (Figure 4.6).
8. Click on the *Continue* button. This will take you back to the *Frequencies* window (Figure 4.4).
9. Click on the *OK* button in the *Frequencies* window; this tells SPSS to process the Frequencies report based on the parameters you just specified. SPSS should produce this report in under a minute.

Statistics Tables

The *Statistics* table (Table 4.2) shows the summary statistical results, as discussed earlier.

Table 4.2 Statistics Table Showing Summary Statistics for age.

Table 4.2 *Statistics* Table Showing Summary Statistics for *age*.

Statistics

age

N	Valid	50
	Missing	0
Mean		24.00
Median		24.00
Mode		25
Std. Deviation		2.857
Variance		8.163
Range		13
Minimum		18
Maximum		31

The report also includes the frequency of each value in the *age* variable (Table 4.3). Focus on columns 1 and 2 of this table, which show that the numbers 18, 19, 20, and 21 each occur twice in the data set; 22 and 23 each occur six times; 24 occurs eight times; 25 occurs nine times; and so on.

Table 4.3 Statistics Table Showing the Frequency of Each Value in the age Variable.

Table 4.3 Statistics Table Showing the Frequency of Each Value in the *age* Variable.

age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	18	2	4.0	4.0	4.0
	19	2	4.0	4.0	8.0
	20	2	4.0	4.0	12.0
	21	2	4.0	4.0	16.0
	22	6	12.0	12.0	28.0
	23	6	12.0	12.0	40.0
	24	8	16.0	16.0	56.0
	25	9	18.0	18.0	74.0
	26	5	10.0	10.0	84.0
	27	3	6.0	6.0	90.0
	28	2	4.0	4.0	94.0
	29	1	2.0	2.0	96.0
	30	1	2.0	2.0	98.0
	31	1	2.0	2.0	100.0
	Total	50	100.0	100.0	

Histogram With Normal Curve

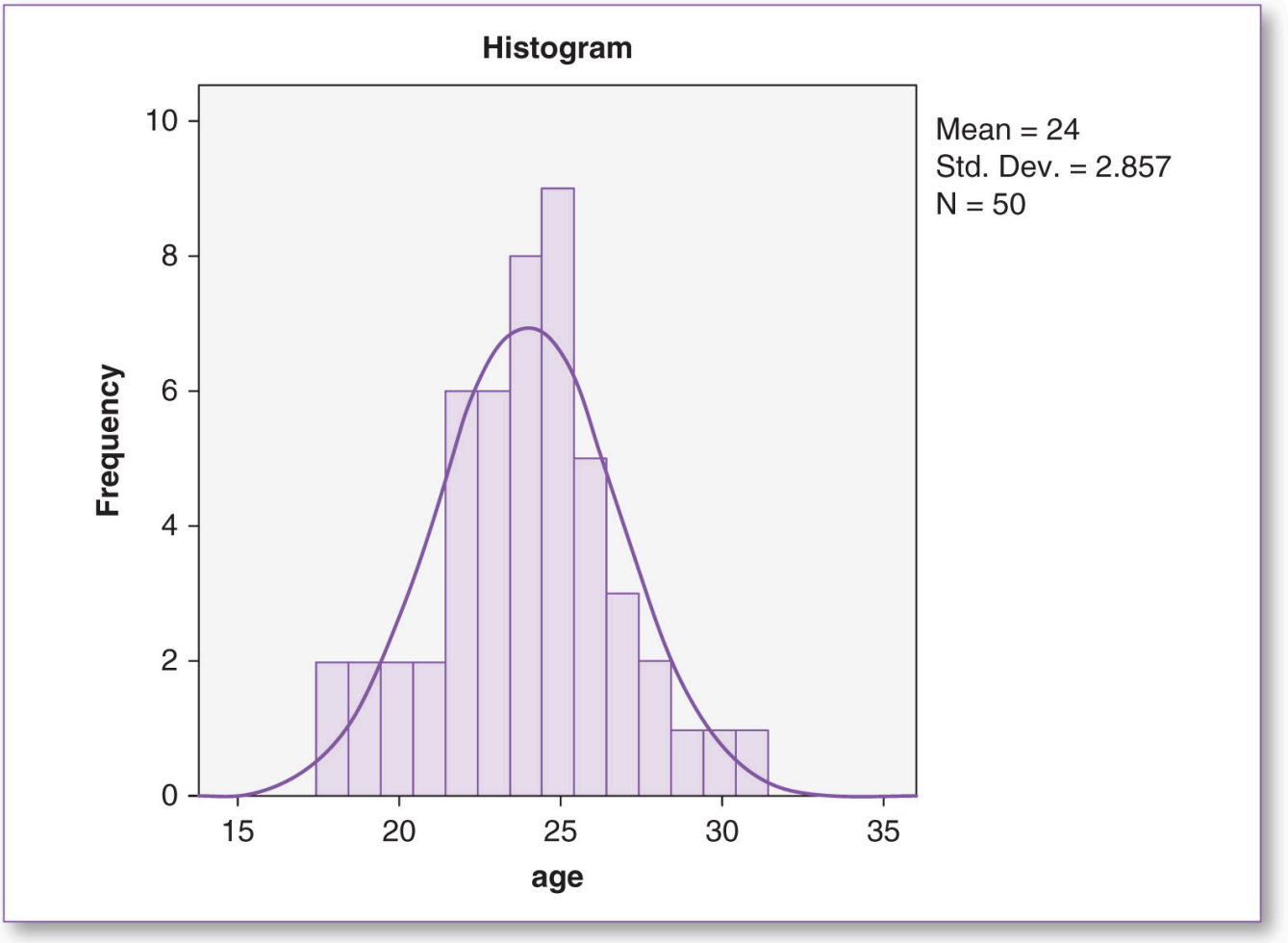
The next part of this report is the histogram of the *age* variable. The histogram is simply a graphical representation of the frequency statistics. Basically, Figure 4.7 is a picture of the data in Table 4.3. Notice that the first four bars are each two units tall; this is because the first four numbers in the table (18, 19, 20, and 21) each occur two times in the data set. Notice that the tallest bar is nine units tall; this is because the number 25 occurs nine times in the data set.

The histogram provides further insight into the (descriptive) characteristics of a continuous variable—a picture

is indeed worth a thousand words.

In addition to the bars, which constitute the histogram, it is also traditional to include a normal curve—sometimes referred to as a “bell curve” because of its shape. The normal curve is derived from the same source data as the bar chart; you can think of this normal curve as the smoothed-out version of the bar chart. More often than not, we see this sort of symmetrical (bell-shaped) distribution in continuous variables—most of the values are gathered toward the middle, with the frequencies progressively dropping off as the values depart above or below the mean.

Figure 4.7 Histogram of the age variable.



For example, if we were to measure the heights of 100 randomly selected people, we would expect to find that most people are of moderate height (which would constitute the tallness in the middle of the normal curve).

We would also expect to find a few exceptionally short people and about the same number of exceptionally tall people, which would account for the tapering off seen on the left and right tails of the normal curve. This phenomenon of the bell-shaped distribution is so common that it is referred to as the *normal distribution*, as represented by the normal curve. When inspecting a histogram for normality, it is expected that the bars may have some jagged steps from bar to bar; however, to properly assess a variable for normality, our focus is on the symmetry of the normal curve more than the bars. If we were to slice a proper normal curve vertically down the middle, the left half of the normal curve should resemble a mirror image of the right half.

Skewed Distribution

As with any rule, there are exceptions; not all histograms produce normally shaped curves. Depending on the distribution of the data within the variable, the histogram may be skewed, meaning that the distribution is shifted to one side or the other, as shown in Figures 4.8 and 4.9.

In Figure 4.8, we see that most of the data are on the right, between about 150 and 300, but there is a small scattering of lower values (under 100), forcing the left tail of the curve to be extended out. These few low values that substantially depart from the majority of the data are referred to as *outliers*. Typically, outliers become apparent when graphing the data. We would say that the histogram in Figure 4.8 has outliers to the left—hence, it is *skewed left*, or *negatively skewed*.

Outliers can also be positive. Figure 4.9, which is a virtual mirror image of Figure 4.8, shows outliers scattered to the right; this distribution would be referred to as being *skewed right*, or *positively skewed*. The notion of normality of the data distribution will be discussed further in future chapters.

Figure 4.8 Negative (left) skew.

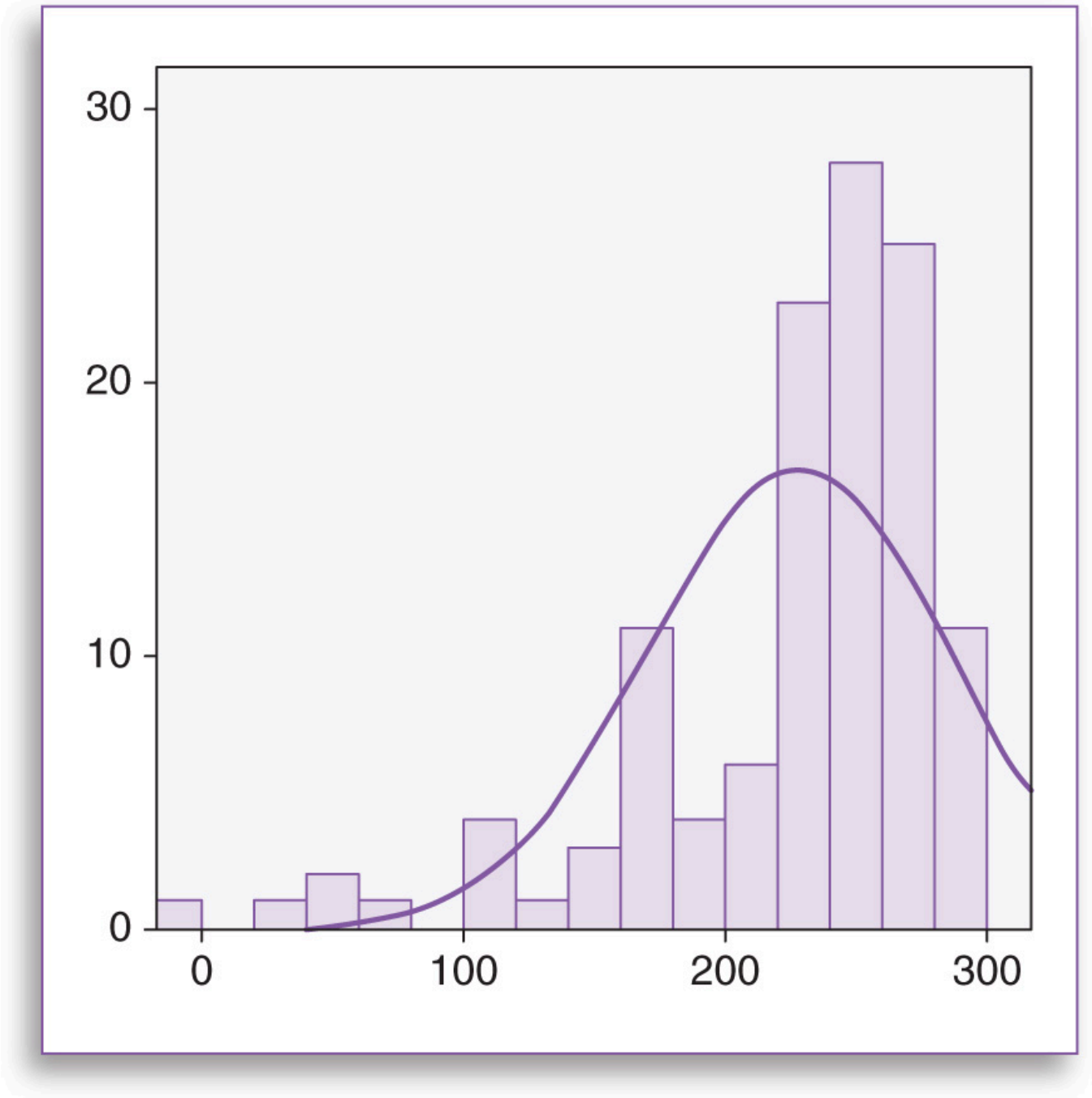
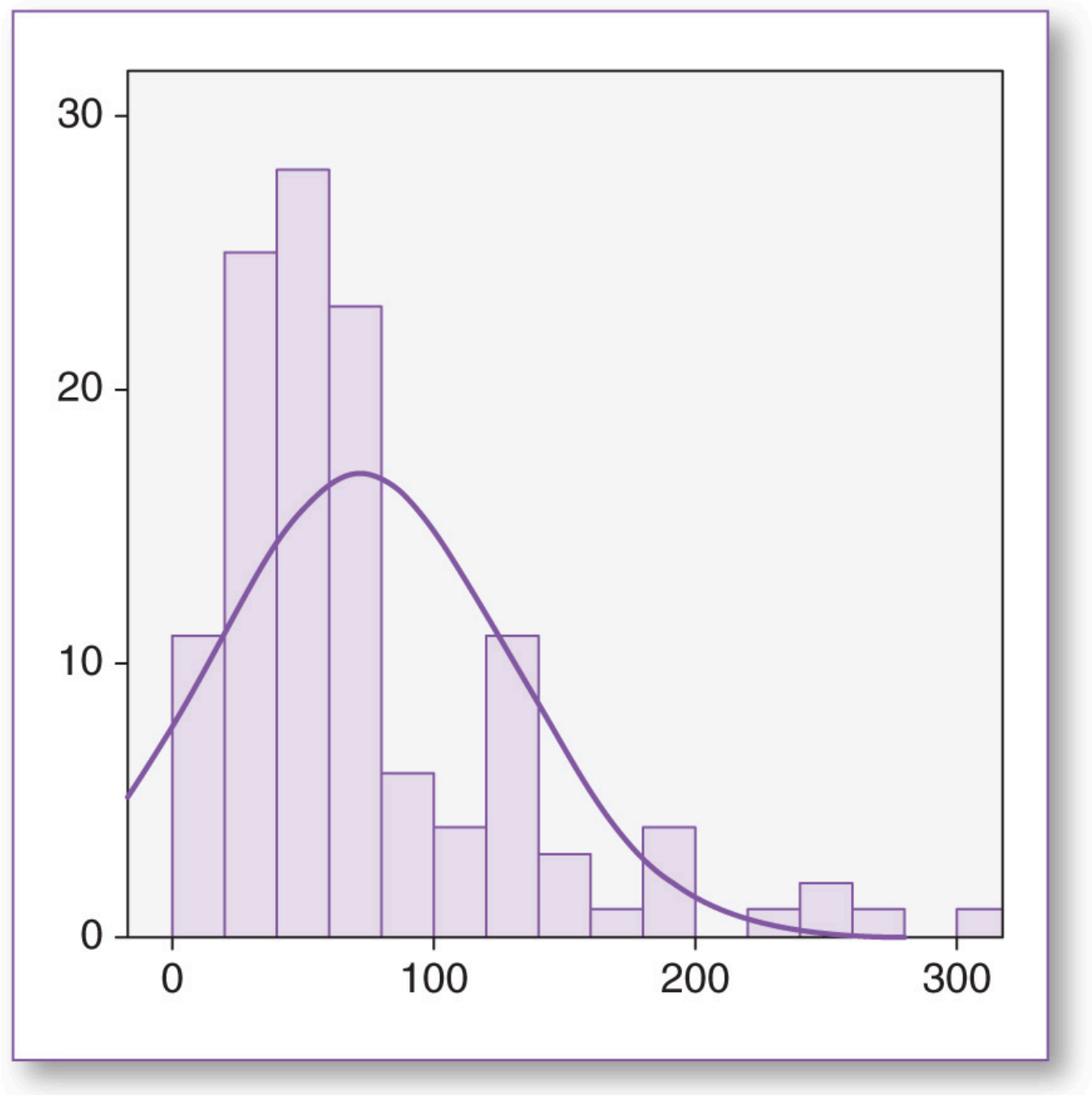


Figure 4.9 Positive (right) skew.

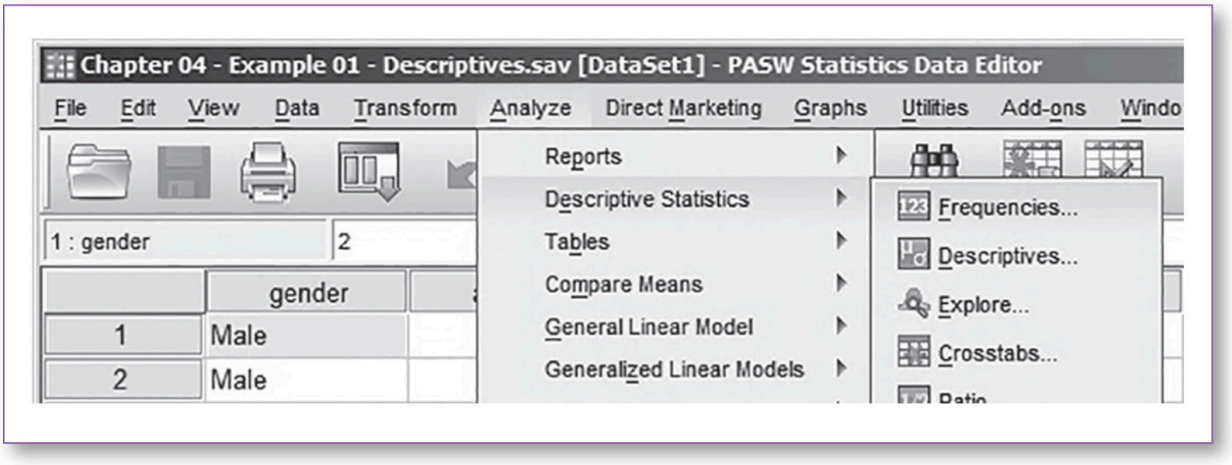


SPSS—Descriptive Statistics: Categorical Variables (*Gender*)

Descriptive statistics for categorical variables are derived using the same ordering menus as for continuous variables, except you will be specifying different options. Although it is plausible to compute the mode for a categorical variable to determine which is the largest category, it would be inappropriate to compute statistics such as mean, median, maximum, minimum, range, standard deviation, and variance. For example, in this data set, the *gender* variable is coded as 1 for Female and 2 for Male; if we requested the mean for *gender*, the result would be 1.46, which is essentially meaningless. Furthermore, since the coding designation for this categorical variable is fairly arbitrary, we could have coded the categories for *gender* as 14 for Female and 83 for Male, in which case, instead of 1.46, the mean for *gender* would now be 45.74, which is equally meaningless.

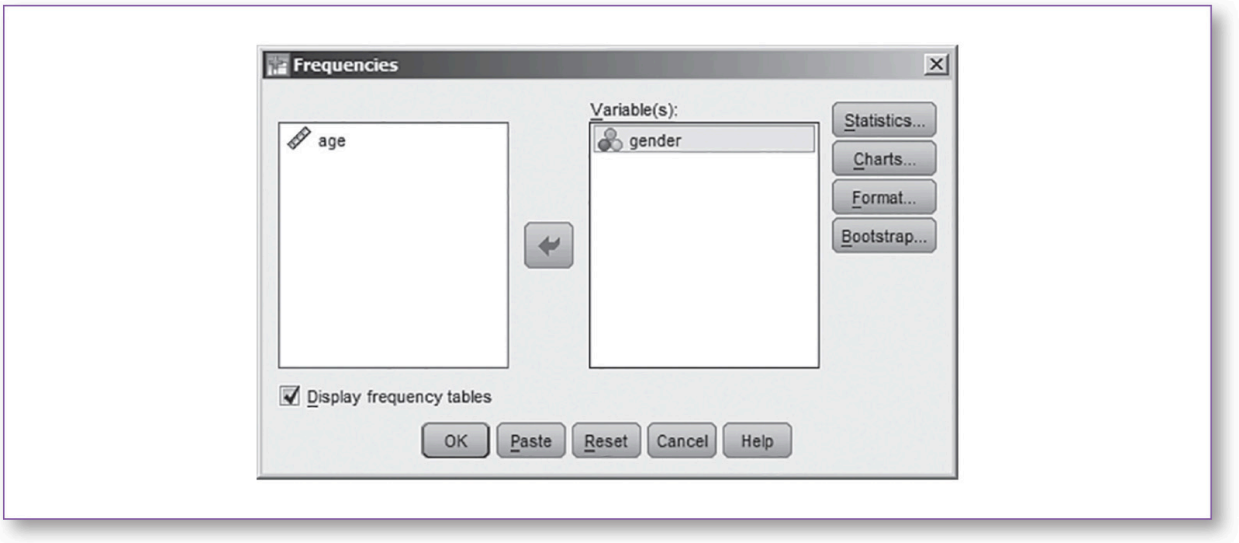
1. Click on *Analyze*, *Descriptive Statistics*, *Frequencies* (Figure 4.10).

Figure 4.10 Running descriptive statistics report; click on Analyze, Descriptive Statistics, Frequencies.



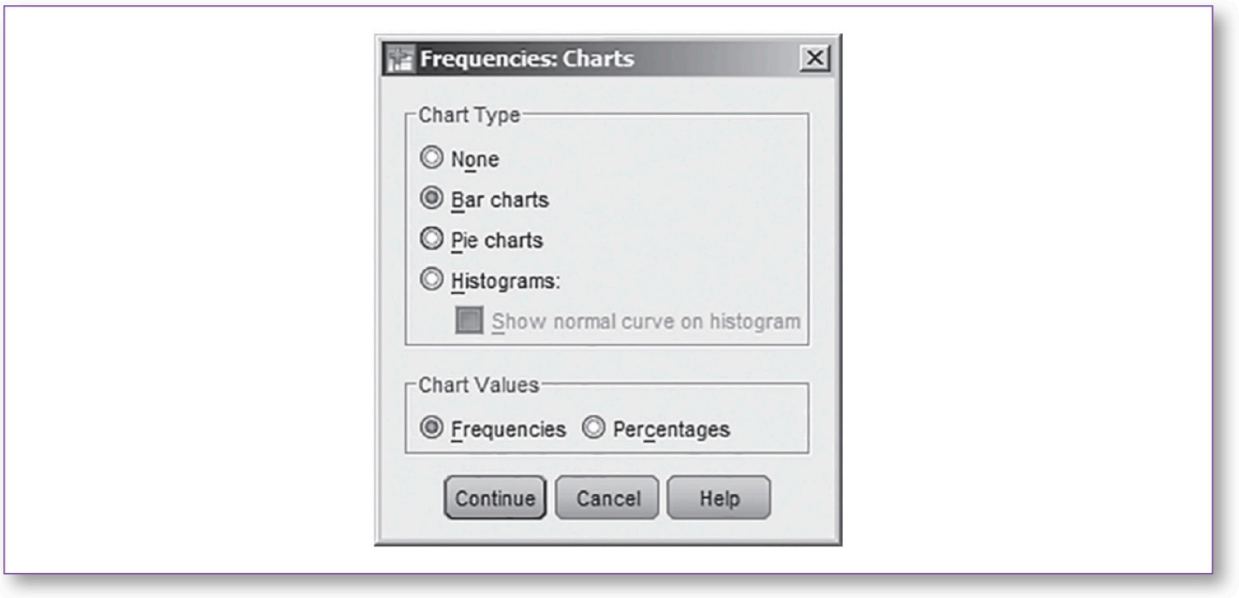
2. First, click on *Reset*; this will clear the parameters on the subwindows that you specified when running the summary statistics for *age*.
3. Next, move the *gender* variable from the left panel to the right *Variable(s)* panel (Figure 4.11).
4. Click on the *Charts* button.

Figure 4.11 Click on the Reset button to clear the prior options, then move the variable(s) to be analyzed (gender) from the left panel to the right Variable(s) panel.



5. In the *Frequencies: Charts* window, there are two viable options for categorical variables: *Bar charts* and *Pie charts* (Figure 4.12). In statistics, bar charts are used more often than pie charts. You can also choose to represent the numbers as frequencies (the actual counts) or percentages. For this example, select *Frequencies*.

Figure 4.12 Frequencies: Charts window; select Bar charts and Frequencies. After running this analysis as specified, feel free to return to this window to rerun this analysis using different settings (e.g., Bar charts with Percentages, Pie charts).



6. Click on the *Continue* button; this will return you to the *Frequencies* window.
7. Click on the *OK* button in the *Frequencies* window; this tells SPSS to process the *Frequencies* report based on the parameters you just specified.

Statistics Tables

The first part of this frequency report shows the overall *n* (*N* in SPSS)—the total number of entries (records) in the variable: 50 valid records and 0 missing, as shown (Table 4.4).

Table 4.4 Descriptive Statistics for gender: n, Valid, and Missing.

Table 4.4 Descriptive Statistics for *gender*: *n*, Valid, and Missing.

Statistics

gender

N	Valid	50
	Missing	0

The next part of the report provides more detailed information regarding the *n*, indicating the frequency (actual number) and percentage for each category within the *gender* variable (Female and Male), as shown in Table 4.5.

Table 4.5 Descriptive Statistics for gender: Frequency and Percentage.

Table 4.5 Descriptive Statistics for *gender*: *Frequency* and *Percentage*.

gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	27	54.0	54.0	54.0
	Male	23	46.0	46.0	100.0
	Total	50	100.0	100.0	

Incidentally, to calculate the percentage, divide the frequency for the category by the (valid) *n* and multiply by 100; for Female, it would be (27 ÷ 50) × 100 = 54%.

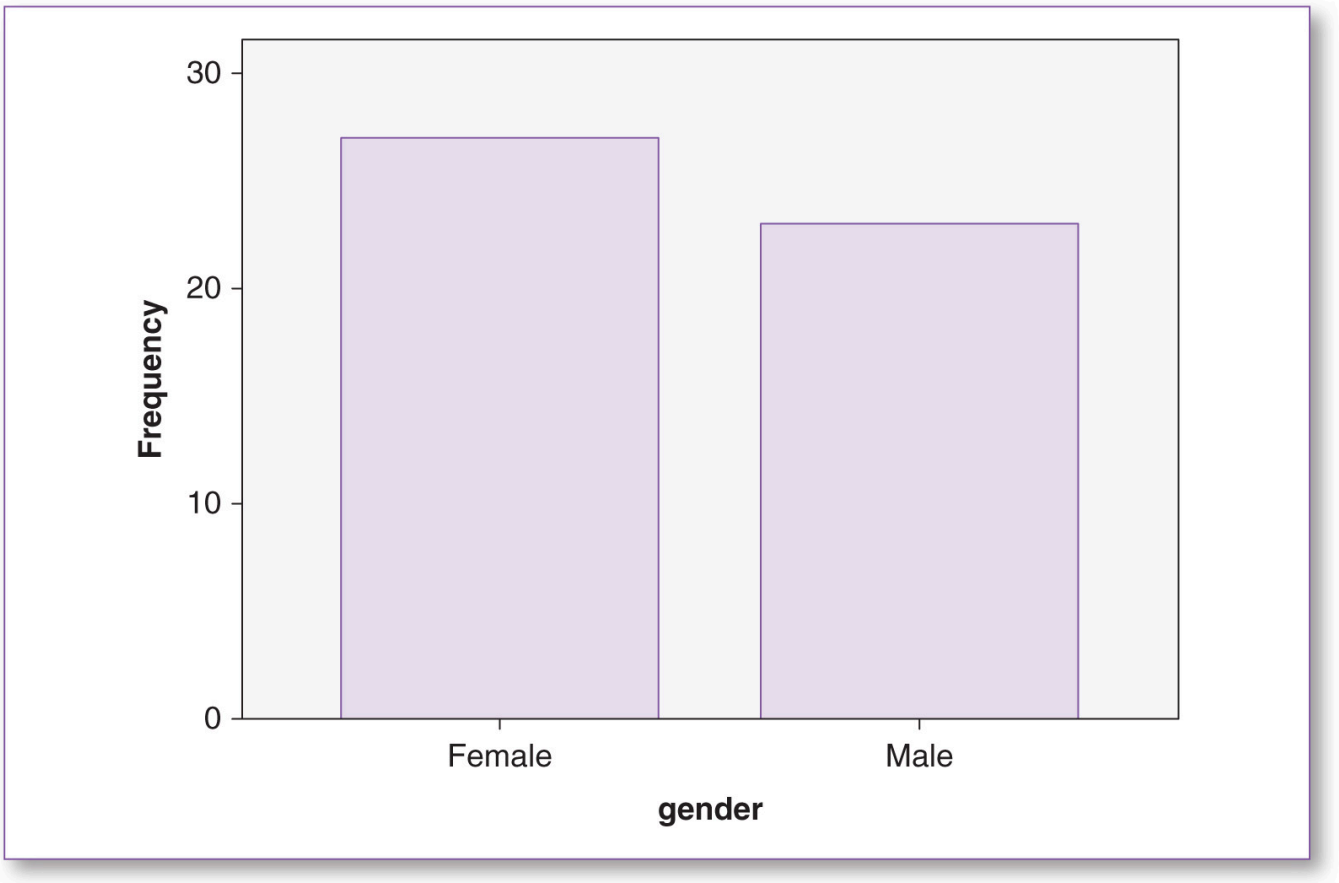
Percentage Formula

(Part ÷ Total) × 100
(27 ÷ 50) × 100
.54 × 100
54%

Bar Chart

Last, the report provides a bar chart representing the two *gender* categories (Female and Male) (Figure 4.13).

Figure 4.13 Bar chart of the gender variable.



SPSS—Descriptive Statistics: Continuous Variable (*Age*) Select by Categorical Variable (*Gender*)—Female or Male Only

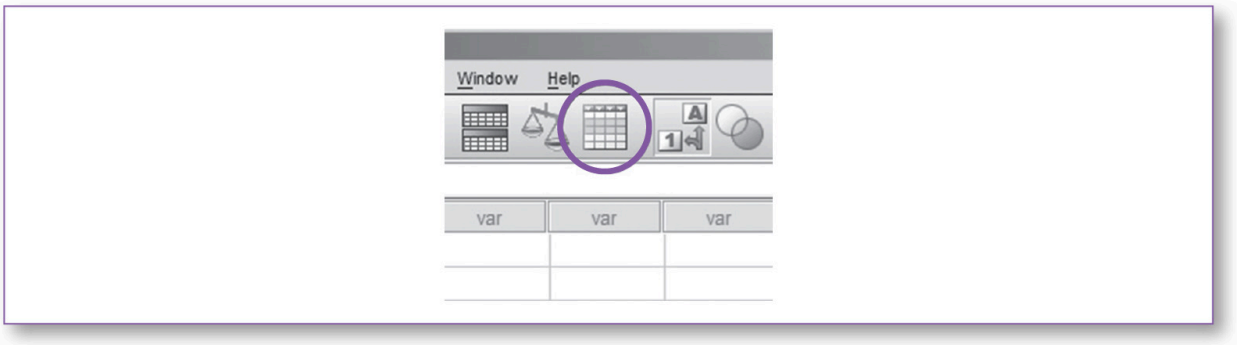
So far, we have processed the continuous variable, *age*, with both genders combined, but it is also possible to produce separate reports for women only and men only, showing the summary statistics and histograms for each. This technique not only satisfies curiosity about what is going on in each category but it will also be essential for running the *pretest checklist* reports that will be covered in future chapters.

We will begin with processing the *age* summary statistics for women only, and then we will repeat the process selecting data for men only. The *Select Cases* option allows you to efficiently specify which cases (rows of data, also known as *records*) you would like to process; SPSS will temporarily ignore all other cases—they will not be included in any statistical computations until you choose to reselect them.

The following procedure will select only the cases where *gender* = 1 (Female).

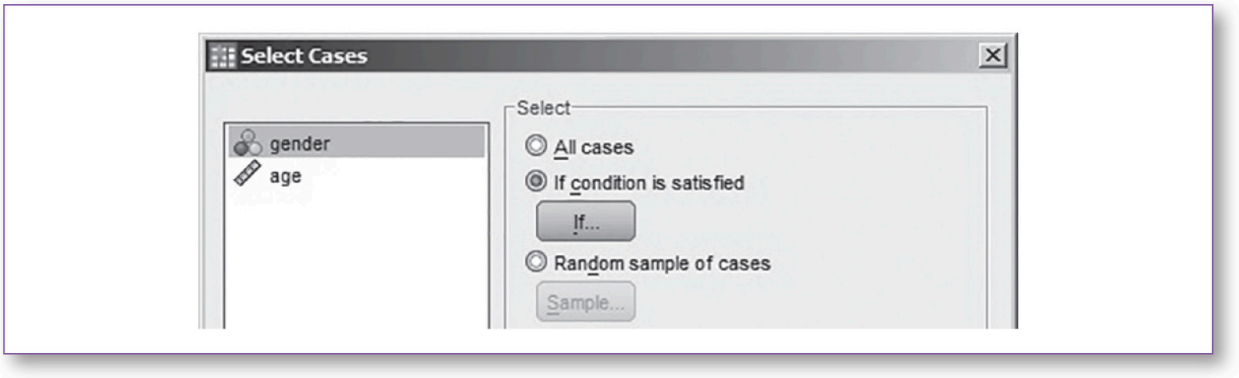
1. Click on the *Select Cases* icon (Figure 4.14).

Figure 4.14 The Select Cases icon.



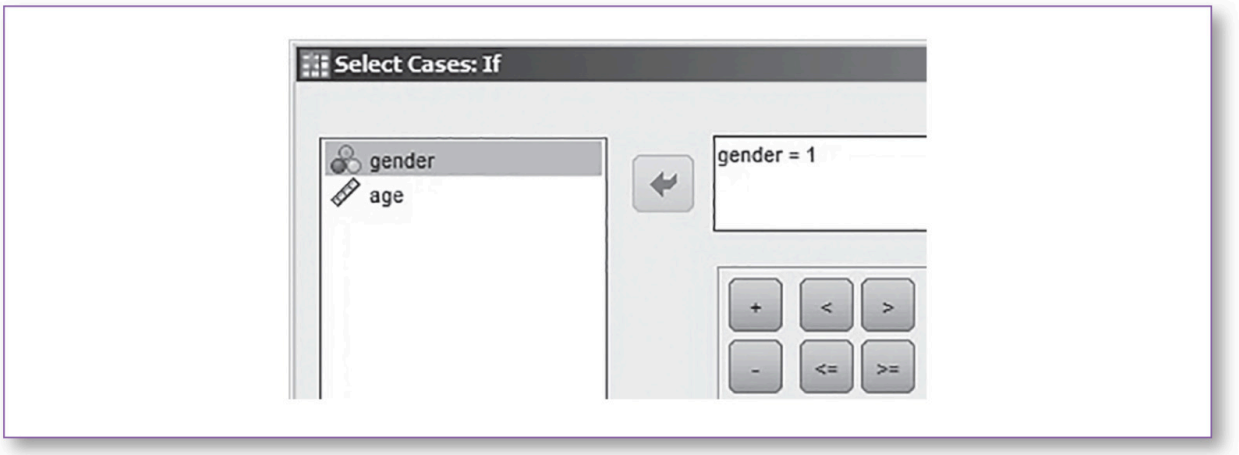
2. In the *Select Cases* window (Figure 4.15), the default selection is *All cases*. Click on *If condition is satisfied*, then click on the *If* button.

Figure 4.15 The Select Cases window (top only).



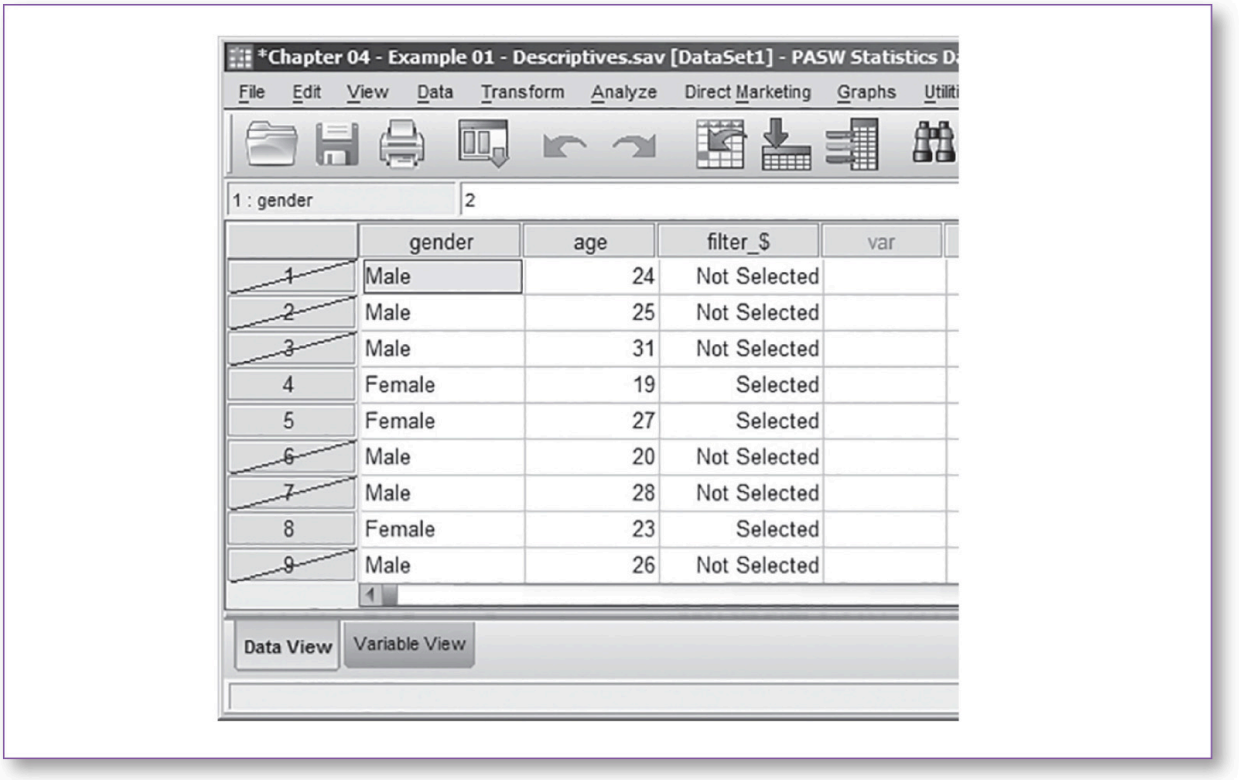
3. This will bring you to the *Select Cases: If* window (Figure 4.16).

Figure 4.16 The Select Cases: If window (top only).



4. Remember: SPSS handles categorical variables as numbers; earlier, we established that for the categorical variable *gender*, 1 = Female and 2 = Male. Since we want statistical reports for women only, we enter the inclusion criteria, *gender* = 1, in the big box at the top of the window. Then click on the *Continue* button.
5. This will return you to the *Select Cases* window. Click on the *OK* button, and the system will process your selection criteria.
6. Switch back to the *Data View* screen. First, notice that the record (row) number for each male participant is slashed out (Figure 4.17). You can think of all the data in the slashed-out rows as being in a sort of penalty box—they are still part of the data set, but they cannot play in any of the rounds of analyses until they are reselected; however, you can still edit such data.

Figure 4.17 The Data View screen after Select Cases (gender = 1) has been executed.



Notice that SPSS has created the temporary variable *filter_$* in the last column, which corresponds to the slashes in each row. If you click on the *Value Labels* icon or go to the *Variable View* screen, you will see that the *filter_$* variable contains two categories: 0 = Not Selected and 1 = Selected.

Since we selected only cases where *gender* = 1, this means that if we were to (re)run the descriptive statistics, the summary statistics and histogram would reflect women only, as opposed to the earlier report that combined women and men.

7. Rerun the analysis for the *age* variable using the procedure (go to the ★ icon on page 72). The resulting statistical report should resemble the data shown in Table 4.6.

Table 4.6 Statistics Table Showing Summary Statistics for age for Women Only.

Table 4.6 *Statistics* Table Showing Summary Statistics for *age* for Women Only.

Statistics

age

N	Valid	27
	Missing	0
Mean		23.19
Median		23.00
Mode		23
Std. Deviation		2.543
Variance		6.464
Range		10
Minimum		18
Maximum		28

8. Notice that the *n* has changed from 50, which included both women and men, to 27, which is women only. Compared with the first report, all of the other statistics have changed as well. Continuing our analysis of women only, observe the frequency statistics (Table 4.7) and corresponding histogram (Figure 4.18).

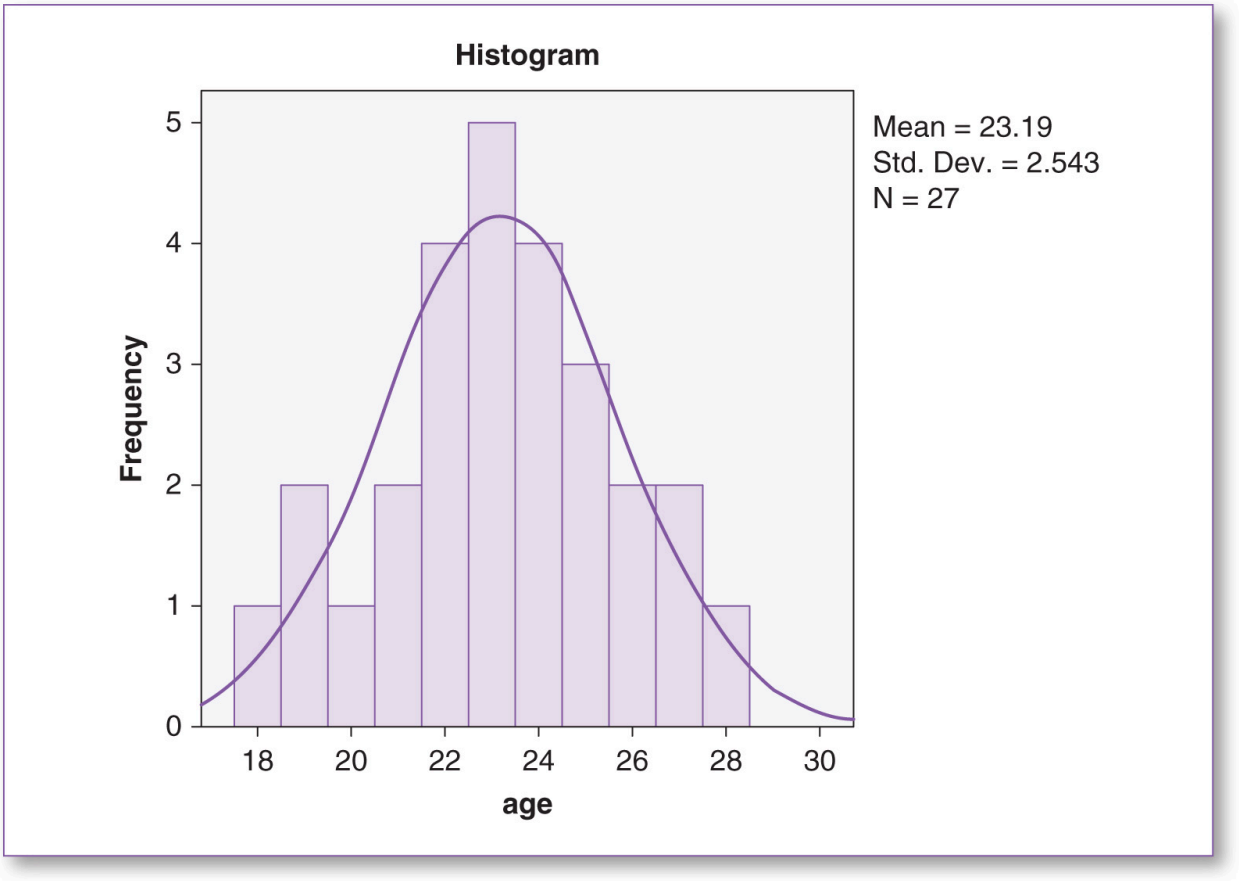
Table 4.7 Statistics Table Showing the Frequency of Each Value in the age Variable for Women Only.

Table 4.7 *Statistics* Table Showing the Frequency of Each Value in the *age* Variable for Women Only.

age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	18	1	3.7	3.7	3.7
	19	2	7.4	7.4	11.1
	20	1	3.7	3.7	14.8
	21	2	7.4	7.4	22.2
	22	4	14.8	14.8	37.0
	23	5	18.5	18.5	55.6
	24	4	14.8	14.8	70.4
	25	3	11.1	11.1	81.5
	26	2	7.4	7.4	88.9
	27	2	7.4	7.4	96.3
	28	1	3.7	3.7	100.0

Figure 4.18 Histogram of the age variable for women only.



9. As you can see, there is a lot to be learned by selecting the data and examining statistics pertaining to women only. The next step is to run the same reports for men only.
10. Begin the men-only analysis by selecting only the cases that pertain to men; go to the ★ icon on page 81, except when you get to **Step 4**, instead of specifying *gender* = 1, change that to *gender* = 2 (remember, we established *gender* as 1 for Female and 2 for Male).
11. Upon rerunning the data for men, notice that the slashes, *filter_$*, and output reports have all changed to reflect men only (see Table 4.8, Table 4.9, and Figure 4.19).

SPSS—(Re)Selecting All Variables

At this point, we have run three sets of descriptive statistics on *age*: (a) all records, (b) women only, and (c) men only. Now, suppose we want to perform further analyses using the entire data set again; there are several ways to reactivate all of the slashed-out records:

Table 4.8 Statistics Table Showing Summary Statistics for age for Men Only.

Table 4.8 *Statistics* Table Showing Summary Statistics for *age* for Men Only.

Statistics

age

N	Valid	23
	Missing	0
Mean		24.96
Median		25.00
Mode		25
Std. Deviation		2.962
Variance		8.771
Range		13
Minimum		18
Maximum		31

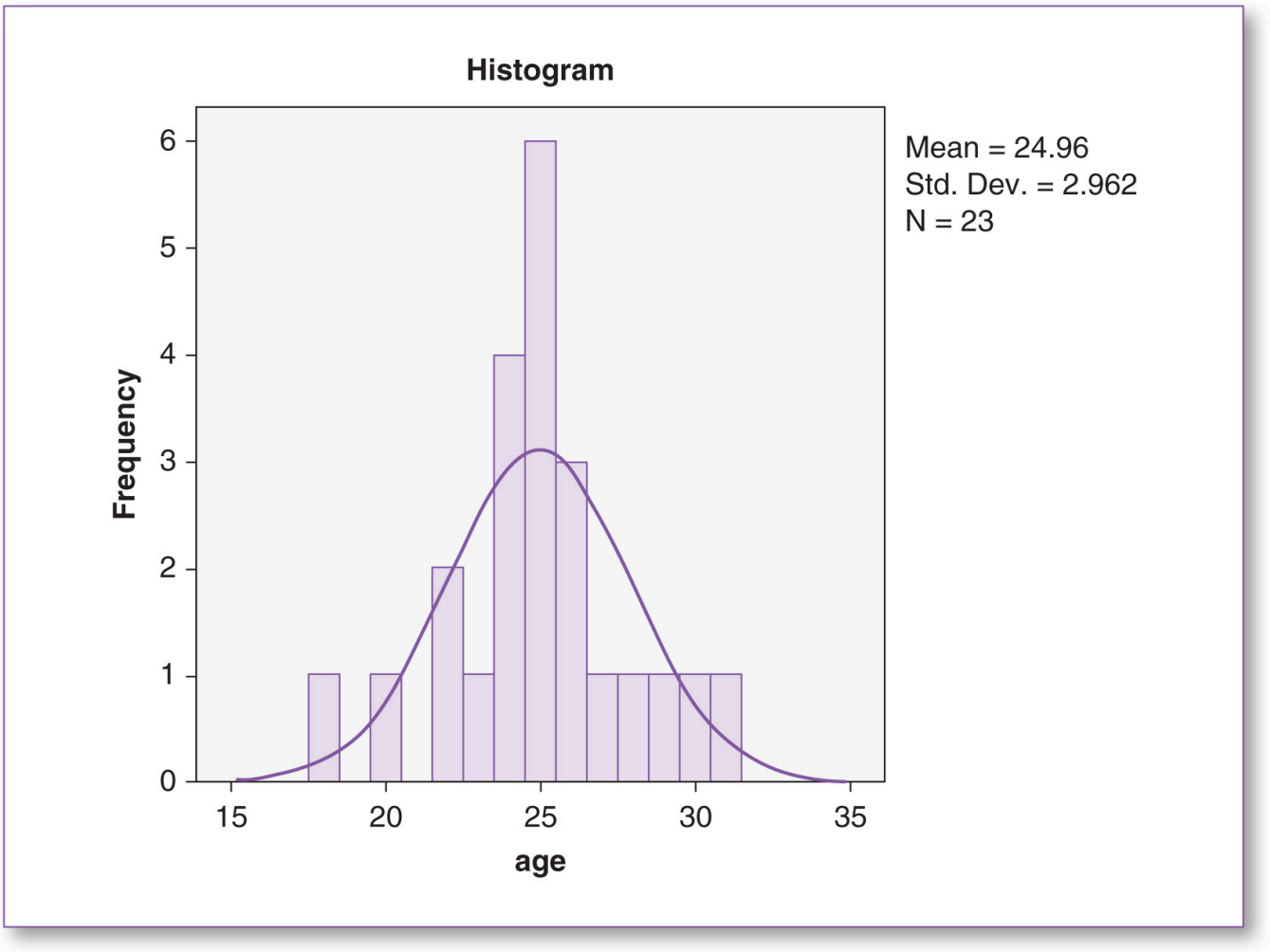
Table 4.9 Statistics Table Showing the Frequency of Each Value in the age Variable for Men Only.

Table 4.9 *Statistics* Table Showing the Frequency of Each Value in the *age* Variable for Men Only.

age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	18	1	4.3	4.3	4.3
	20	1	4.3	4.3	8.7
	22	2	8.7	8.7	17.4
	23	1	4.3	4.3	21.7
	24	4	17.4	17.4	39.1
	25	6	26.1	26.1	65.2
	26	3	13.0	13.0	78.3
	27	1	4.3	4.3	82.6
	28	1	4.3	4.3	87.0
	29	1	4.3	4.3	91.3
	30	1	4.3	4.3	95.7
	31	1	4.3	4.3	100.0
	Total	23	100.0	100.0	

Figure 4.19 Histogram of the age variable for men only.



- On the *Data View* screen, click on the column header *filter_$* (which will highlight the whole column) and press the Delete key.
- On the *Variable View* screen, click on the corresponding row number—in this case, row 3 (which will highlight the whole row)—and press the *Delete* key.
- Click on the *Select Cases* icon, then click on the *All Cases* button. Finally, click on the *OK* button.

Good Common Sense

Although SPSS is proficient at processing statistical data, keep in mind that the program has no real intelligence per se—it just pushes the numbers through the behind-the-scenes formulas using the parameters you

specify. As such, it is up to you to enter the data accurately and make intelligent processing selections.

In the examples detailed in this chapter, we requested that SPSS perform a variety of statistical analyses on the *age* variable, which makes sense; knowing the mean age can be useful. Larger databases are likely to contain other numeric variables, such as patient ID numbers, street addresses, phone numbers, serial numbers, license numbers, and so on. Technically, you could instruct SPSS to compute descriptive statistics on such variables; SPSS is not bright enough to know that computing the mean for a series of phone numbers makes no sense; SPSS, or any other statistical processing software, would mindlessly process the variable and provide you with an average phone number, which would be useless. In summary, it is up to you to proceed mindfully when using any such software.

Key Concepts

- Descriptive statistics
    - Number (*n*)
    - Mean (μ)
    - Median
    - Mode
    - Standard deviation (SD)
    - Variance
    - Minimum
    - Maximum
    - Range
- Central tendency
- Loading SPSS data files
    - Histogram
    - Normal curve
- Skew
    - Negative (left) skew
    - Positive (right) skew
- Outliers
- Bar chart
- Pie chart
- Select cases

- Good common sense

Practice Exercises

Use the prepared SPSS data sets (download from **study.sagepub.com/knappstats2e**).

Load the specified data sets, then process and document your findings for each exercise.

Exercise 4.1

A survey was conducted in Professor Lamm's class and Professor Milner's class. The question that students responded to is, "How many siblings do you have?"

Data set: Ch 04 - Exercise 01.sav
Codebook
Variable: class
Definition: Class designation
Type: Categorical (1 = Prof. Lamm, 2 = Prof. Milner)
Variable: siblings
Definition: Number of siblings
Type: Continuous

a. Run descriptive statistics and a histogram with a normal curve for *siblings* for the whole data set.

b. Run descriptive statistics and a bar chart for *class* for the whole data set.

c. Run descriptive statistics and a histogram with a normal curve for *siblings* for members of Professor Lamm's class only.

d. Run descriptive statistics and a histogram with a normal curve for *siblings* for members of Professor Milner's class only.

Exercise 4.2

While waiting in line to donate blood, donors were asked, "How many times have you donated before?" The

researcher recorded their gender and number of prior donations.

Data set: **Ch 04 - Exercise 02.sav**

Codebook

Variable: gender
Definition: Gender
Type: Categorical (1 = Female, 2 = Male)
Variable: donated
Definition: Total number of blood donations given before today
Type: Continuous

a. Run descriptive statistics and a histogram with a normal curve for *donated* for the whole data set.
b. Run descriptive statistics and a bar chart for *gender* for the whole data set.
c. Run descriptive statistics and a histogram with a normal curve for *donated* for women only.
d. Run descriptive statistics and a histogram with a normal curve for *donated* for men only.

Exercise 4.3

You want to know if typing proficiency is associated with better spelling skills. You administer a spelling test consisting of 20 words to the students in a classroom. At the bottom of the sheet, there is a question: “Can you type accurately without looking at the keyboard?”

Data set: **Ch 04 - Exercise 03.sav**

Codebook

Variable: looker
Definition: Does the student look at the keyboard to type?
Type: Categorical (1 = Looks at keyboard, 2 = Doesn’t look at keyboard)

Variable: spelling
Definition: Score on spelling test
Type: Continuous

a. Run descriptive statistics and a histogram with a normal curve for *spelling* for the whole data set.

b. Run descriptive statistics and a bar chart for *looker* for the whole data set.

c. Run descriptive statistics and a histogram with a normal curve for *spelling* for "looks at keyboard" only.

d. Run descriptive statistics and a histogram with a normal curve for *spelling* for "doesn't look at keyboard" only.

Exercise 4.4

You are interested in the length of time it takes for individuals to complete their transaction(s) at an ATM. You use a stopwatch to record your unobtrusive observations and gather two pieces of information on each person: gender and the length of his or her ATM session (in seconds).

Data set: **Ch 04 - Exercise 04.sav**

Codebook

Variable: gender
Definition: Gender
Type: Categorical (1 = Female, 2 = Male)
Variable: atmsec
Definition: Number of seconds spent at ATM
Type: Continuous

a. Run descriptive statistics and a histogram with a normal curve for *atmsec* for the whole data set.

b. Run descriptive statistics and a bar chart for *gender* for the whole data set.

c. Run descriptive statistics and a histogram with a normal curve for *atmsec* for women only.

d. Run descriptive statistics and a histogram with a normal curve for *atmsec* for men only.

Exercise 4.5

You are interested in finding out how many units students are enrolled in. You conduct a survey of 40 students and record two pieces of information: the degree (level) the students are working on (bachelor's, master's, doctorate) and the total number of units they are taking this term.

Data set: **Ch 04 - Exercise 05.sav**

Codebook

Variable: degree
Definition: Highest degree the person has
Type: Categorical (1 = Bachelor's, 2 = Master's, 3 = Doctorate)
Variable: units
Definition: Current number of enrolled units
Type: Continuous

a. Run descriptive statistics and a histogram with a normal curve for *units* for the whole data set.
b. Run descriptive statistics and a bar chart for *degree* for the whole data set.
c. Run descriptive statistics and a histogram with a normal curve for *units* for bachelor's degree only.
d. Run descriptive statistics and a histogram with a normal curve for *units* for master's degree only.
e. Run descriptive statistics and a histogram with a normal curve for *units* for doctorate only.

Exercise 4.6

You stand at a register in a hospital cafeteria; for each patron, you gather two pieces of information: professional role (nurse, doctor, other), as indicated on his or her badge, and the amount of the person's bill (as shown on the register).

Data set: **Ch 04 - Exercise 06.sav**

Codebook

Variable: profrole
Definition: Professional role
Type: Categorical (1 = Nurse, 2 = Doctor, 3 = Other)
Variable: bill
Definition: Total as shown on the register
Type: Continuous

a. Run descriptive statistics and a histogram with a normal curve for *bill* for the whole data set.
b. Run descriptive statistics and a bar chart for *profrole* for the whole data set.
c. Run descriptive statistics and a histogram with a normal curve for *bill* for nurse only.
d. Run descriptive statistics and a histogram with a normal curve for *bill* for doctor only.
e. Run descriptive statistics and a histogram with a normal curve for *bill* for other only.

Exercise 4.7

You recruit a group of people who agree to report their total e-mail counts (sent + received) for 30 days. Each participant also completed a survey regarding his or her employment status (full-time, part-time, unemployed).

Data set: **Ch 04 - Exercise 07.sav**

Codebook

Variable: employ

Definition: Employment status
Type: Categorical (1 = Full-time, 2 = Part-time, 3 = Unemployed)
Variable: emails
Definition: Total number of e-mails sent and received for 30 days
Type: Continuous

a. Run descriptive statistics and a histogram with a normal curve for *emails* for the whole data set.
b. Run descriptive statistics and a bar chart for *employ* for the whole data set.
c. Run descriptive statistics and a histogram with a normal curve for *emails* for full-time only.
d. Run descriptive statistics and a histogram with a normal curve for *emails* for part-time only.
e. Run descriptive statistics and a histogram with a normal curve for *emails* for unemployed only.

Exercise 4.8

The members of an exercise walking group agree to partake in your study; you randomly give half of the group walking music in a major key, and the others are given walking music in a minor key. Each participant can walk as often and for as long as he or she likes. The participants will record and submit the total number of minutes that they walked in a week.

Data set: **Ch 04 - Exercise 08.sav**

Codebook

Variable: musickey
Definition: Music key
Type: Categorical (1 = Major, 2 = Minor)
Variable: minwalk
Definition: Total number of minutes walked
Type: Continuous

a. Run descriptive statistics and a histogram with a normal curve for *minwalk* for the whole data set.

b. Run descriptive statistics and a bar chart for *musickey* for the whole data set.

c. Run descriptive statistics and a histogram with a normal curve for *minwalk* for major only.

d. Run descriptive statistics and a histogram with a normal curve for *minwalk* for minor only.

Exercise 4.9

The administrator of a two-ward hospital randomly selects one ward wherein the nurses will be assigned to tend to two patients each; nurses in the other ward will tend to four patients each. Over the course of a month, upon discharge, each patient will complete a nursing care satisfaction survey, which renders a score ranging from 1 to 100 (1 = *very unsatisfied*, 100 = *very satisfied*).

Data set: **Ch 04 - Exercise 09.sav**

Codebook

Variable: ward
Definition: Ward number
Type: Categorical variable (1 = 2 patients per nurse, 2 = 4 patients per nurse)
Variable: nsatisfy
Definition: Patient's nurse satisfaction score
Type: Continuous

a. Run descriptive statistics and a histogram with a normal curve for *nsatisfy* for the whole data set.

b. Run descriptive statistics and a bar chart for *ward* for the whole data set.

c. Run descriptive statistics and a histogram with normal curve for *nsatisfy* for 2 patients per nurse ward only.

d. Run descriptive statistics and a histogram with a normal curve for *nsatisfy* for 4 patients per nurse ward only.

Exercise 4.10

To determine if dancing enhances mood, you recruit 100 voluntary participants. You randomly select 50 and give them seven free dance lessons; the other 50 get no dance lessons. After the seventh class, you administer the Acme Happiness Scale Survey (AHSS) to all 100 individuals; this survey renders a score ranging from 1 to 30 (1 = *extremely unhappy*, 30 = *extremely happy*).

Data set: **Ch 04 - Exercise 10.sav**

Codebook

Variable: dance
Definition: Dance class membership status
Type: Categorical (1 = Dancer, 2 = Nondancer)
Variable: ahss
Definition: Score on Acme Happiness Scale Survey
Type: Continuous

a. Run descriptive statistics and a histogram with a normal curve for *ahss* for the whole data set.
b. Run descriptive statistics and a bar chart for *dance* for the whole data set.
c. Run descriptive statistics and a histogram with a normal curve for *ahss* for dancers only.
d. Run descriptive statistics and a histogram with a normal curve for *ahss* for nondancers only.

https://doi.org/10.4135/9781071878910