

CHAPTER 4

GRAPHING DATA

OBJECTIVES

After completing this chapter, you should

- understand the following graphing conventions: the three-quarters rule and X-axis values should begin with 0 and reflect reasonable deviations in the data.
- be able to construct the following types of graphs: the frequency polygon, the cumulative frequency curve, the histogram, the bar graph, the stem-and-leaf plot, and the line graph.
- appreciate the differences between the normal curve and skewed curves.

CHAPTER REVIEW

Before discussing specific types of graphs, two graphing conventions are presented. The first convention is the *three-quarters rule*, which states that the *Y* axis should be approximately three-fourths as long as the *X* axis. The second convention is that the values on the *Y* axis generally should begin with 0, and the units on the axis should reflect reasonable deviations in the data.

The *frequency polygon* is constructed by plotting the scores on the baseline, or *X* axis, and plotting the frequency of each score along the *Y* axis. The axes of the frequency polygon are labeled with the word “Score” below the *X* axis and the word “Frequency” to the left of the *Y* axis. In addition, each graph should have a caption describing the source of the data. If you want to compare two or more sets of data on a single set of axes and the sample sizes differ, the frequencies must be converted to percentages before plotting.

One common and very important curve is the *normal* or bell-shaped curve. The normal curve is symmetrical, in contrast to *skewed curves*, in which a large number of scores are piled up at one end or the other end, with a tail at the opposite end. *Positively skewed curves* have a tail toward the upper end of the *X* axis, and *negatively skewed curves* have a tail toward the lower end of the *X* axis.

The *cumulative frequency* or *cumulative percentage polygon* can be plotted from a frequency distribution. In this curve, cumulative frequency (or percentage) is plotted over the scores. The cumulative frequency curve is useful for determining the relative position of any individual.

The only difference between the frequency polygon and the *histogram* is that dots are plotted over the score values in the polygon, whereas rectangular bars are plotted over the values in the histogram. The *bar graph* is a type of histogram used to plot categorical or nominal scale data. A bar is drawn over each category, with its height indicating frequency; the bar's width is arbitrary.

The *stem-and-leaf plot* has features of both the frequency distribution and the frequency histogram. To construct it, each data point is split into two parts: a stem (the first digit[s]) and a leaf (the last digit[s]). For example, a score of 155 has a stem of 15 and a leaf of 5. If a stem-and-leaf plot is rotated 90° counterclockwise, it becomes a histogram with digits over the scores (stems) rather than bars. Another useful feature of the stem-and-leaf plot is that it can be used to compare two groups of data.

In a *line graph*, an independent variable is recorded on the *X* axis, and some measure of the dependent variable is shown on the *Y* axis. The independent variable is assumed to be continuous, and a line is used to connect the plotted points.

TERMS TO DEFINE AND/OR IDENTIFY

abscissa

ordinate

frequency polygon

normal curve

skewed curve

positively skewed curve

negatively skewed curve

cumulative frequency curve

histogram

bar graph

stem-and-leaf plot

line graph

FILL-IN-THE-BLANK ITEMS

Introduction

An old adage states that a picture is worth a (1) _____ words, and recognition of the truth of this leads to the discussion of (2) _____ as a descriptive technique. The graphs discussed in the chapter are the frequency polygon, the (3) _____ frequency or percentage curve, the (4) _____, the bar graph, the stem-and-leaf plot, and the (5) _____ graph.

Rules for Graphing

To help prevent misrepresentation of data, a graphing convention is often used that states that the (6) _____ axis should be approximately (7) _____ as long as the X axis. This convention is called the (8) _____ rule. It is also important to begin the values on the Y axis with (9) _____ and to be sure that the Y axis units reflect reasonable (10) _____ in the data.

The Frequency Polygon

In a frequency polygon, the (11) _____ are plotted on the baseline or X axis, and the (12) _____ are plotted on the ordinate or Y axis. Labeling of the axes is very important: the word (13) “ _____ ” appears below the X axis, and the word (14) “ _____ ” appears to the left of the Y axis. The graph should also have a (15) _____ describing the origin of the data.

A comparison of distributions using the percentage or relative frequency polygon

If you want to compare distributions with unequal N s on the same axes, you must first convert frequencies to (16) _____. Once you have done this, you can plot a percentage or (17) _____ frequency polygon.

Shapes of frequency polygons

The (18) _____ or bell-shaped curve is an important symmetrical curve. Unimodal curves with a high peak at one end and a long tail at the other end are called (19) _____ curves. If the tail is to the right, the curve is (20) _____; if the tail is to the left, the curve is (21) _____.

The cumulative frequency (or cumulative percentage) polygon

In the cumulative frequency or percentage polygon, the cumulative frequencies or percentages are plotted over the (22) _____. The relative position of an individual may be determined from the cumulative polygon by drawing a (23) _____ line from that person's score on the X axis to the curve and then drawing a (24) _____ line from that point on the curve to the Y axis. The point at which the (25) _____ line meets the Y axis gives an approximate number (or percentage) of individuals scoring at or below the score being considered.

The Histogram

The histogram is like the frequency polygon except that a rectangular (26) _____ is drawn over each score value on the X axis, with its height determined by the score's (27) _____. Each bar is centered above a score value and extends (28) _____ between adjacent scores.

The Bar Graph

The bar graph is a type of histogram used to graph (29) _____ scale data. The bars don't have to touch, and the spacing between them is (30) _____.

The Stem-and-Leaf Plot

To construct a stem-and-leaf plot, each score is divided into two parts: a (31) _____ and a (32) _____. The first digit(s) is the (33) _____ and the last digit(s) is the (34) _____. For example, a score of 133 would have a stem of (35) _____ and a leaf of (36) _____. Each stem is listed from lowest to highest, a (37) _____ line is drawn to the right of the column of stems, and then the

(38) _____ are put beside the stems. If the plot is rotated so that the stems become
 (39) _____ on the baseline of a graph, the result is a (40) _____, with digits
 over the scores rather than bars.

The Line Graph

In a line graph, an (41) _____ variable is recorded on the X axis, and some measure of the
 dependent variable appears on the Y axis. Because the independent variable is assumed to be
 (42) _____, a line is used to connect the plotted points.

PROBLEMS

1. You have gathered some test score data from a 50-item test in an introductory psychology class. In order to “curve” the grades, you decide to graph the data. The frequency distribution is shown here.

X	f
48	1
47	1
46	2
42	2
41	3
39	2
38	5
37	4
36	3
35	2
33	2
32	1
30	<u>1</u>
$N = 29$	

- a. Construct a frequency polygon.

b. Make a histogram.

c. Construct a cumulative frequency curve.

- d. Construct a cumulative percentage curve and use it to find the approximate percentage of scores at or below a score of 35.

2. Before using behavior modification to reduce smoking, a confirmed smoker records the number of cigarettes smoked per day over a 2-week period. Use the appropriate graph to illustrate the data.

<u>Day</u>	<u><i>f</i></u>	<u>Day</u>	<u><i>f</i></u>
1	23	8	25
2	27	9	22
3	18	10	24
4	19	11	29
5	25	12	25
6	26	13	19
7	33	14	22

5. The mean quality point average (QPA) has been determined by the semester after joining for the members of a campus social organization. Plot the result.

<u>Semesters</u> <u>After Joining</u>	<u>Mean QPA</u>
1	2.97
2	2.65
3	2.58
4	2.45
5	2.30
6	2.19

6. In a recent study, patients suffering from depression received one of three types of treatment: counseling alone, antidepressant medication alone, or a combination of counseling and medication. Their average level of depression (rounded to the nearest whole number) is assessed each month for 6 months. The data are given here. Plot the three groups on the same graph for comparison.

<u>Month</u>	<u>Counseling</u>	<u>Medication</u>	<u>Counseling and Medication</u>
1	17	16	18
2	15	10	12
3	14	9	11
4	13	11	9
5	10	10	6
6	8	9	4

7. Compare frequency distributions A and B on the same set of axes.

<u>Distribution A</u>		<u>Distribution B</u>	
<u>X</u>	<u>f</u>	<u>X</u>	<u>f</u>
25	1	25	6
24	3	24	7
23	3	23	8
22	4	22	10
21	5	21	7
20	6	20	4
19	8	19	3
18	7	18	2
17	7	17	2
16	6	16	1

- a. Construct frequency polygons. The shape of Distribution A is _____ skewed; the shape of Distribution B is _____ skewed.

b. Plot the distributions as cumulative frequency curves.

8. The following numbers are first-lap speeds in miles per hour of 35 cars in the Indianapolis 500 race. Make a stem-and-leaf plot of the speeds. Then rotate the plot 90° counterclockwise so that it looks like a frequency histogram.

183	177	159	172	210
166	167	199	185	188
203	207	197	188	193
168	175	192	201	204
169	177	185	188	189
199	193	195	205	195
188	185	192	199	198

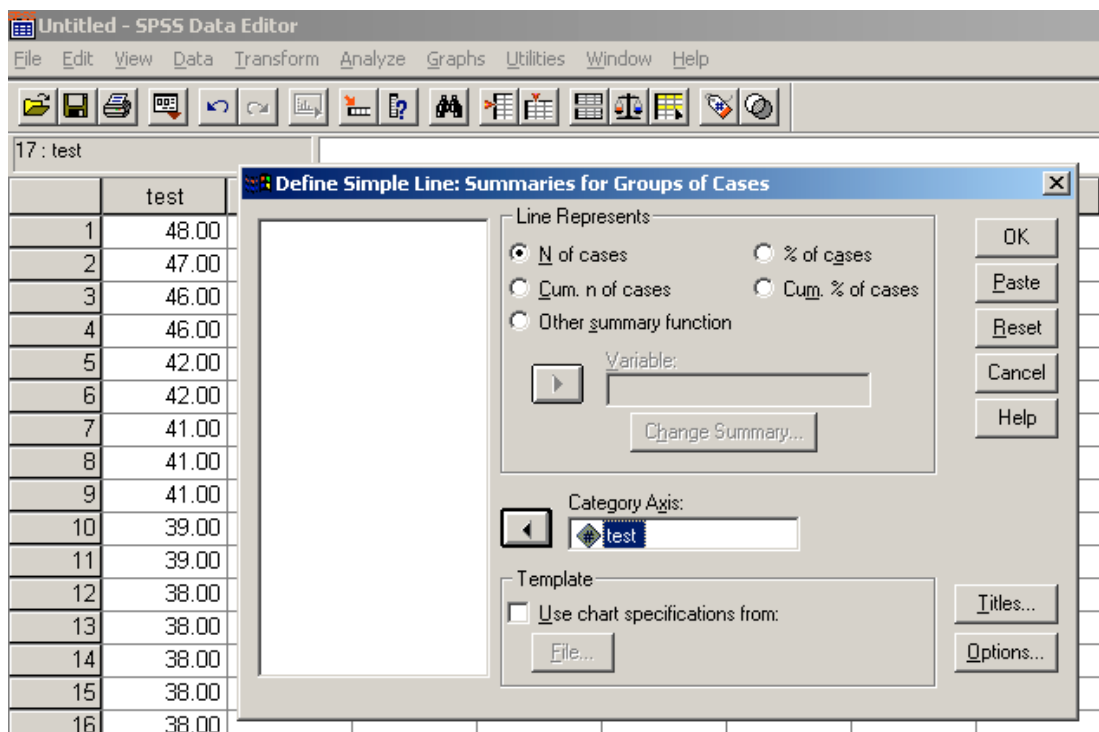
USING SPSS—EXAMPLE AND EXERCISES

If you haven't already done so, read Appendix 4 before proceeding with these exercises.

SPSS has a variety of graphing capabilities, both within certain procedures and independently under the Graphs menu. Unfortunately, the available choices for graphs may not correspond exactly to the ones requested in this chapter and described in the textbook. Thus, you may have to choose a graph that is an approximation to the one requested or one that may require you to draw a few lines to achieve the desired result.

Example: We will use SPSS to produce graphs for Problem 1. In this problem, part **a** requests a frequency polygon, and a histogram is requested in part **b**. SPSS won't easily produce either of these graphs exactly as we have defined them in the text. However, a good approximation for the frequency polygon is provided by an SPSS line graph constructed as follows:

1. Start SPSS, enter the data, and name the variable **test**. Again, remember that when we enter data from a frequency distribution, each score must be entered the number of times indicated by its frequency.
2. Click *Graphs>Line>Simple>Summaries for groups of cases>Define*.
3. Highlight and move the variable **test** into the Category Axis box and be sure "N of cases" is indicated in the Line Represents box. The Define Simple Line dialog box should appear as follows:



4. Click OK and the graph should appear in the output Viewer window.

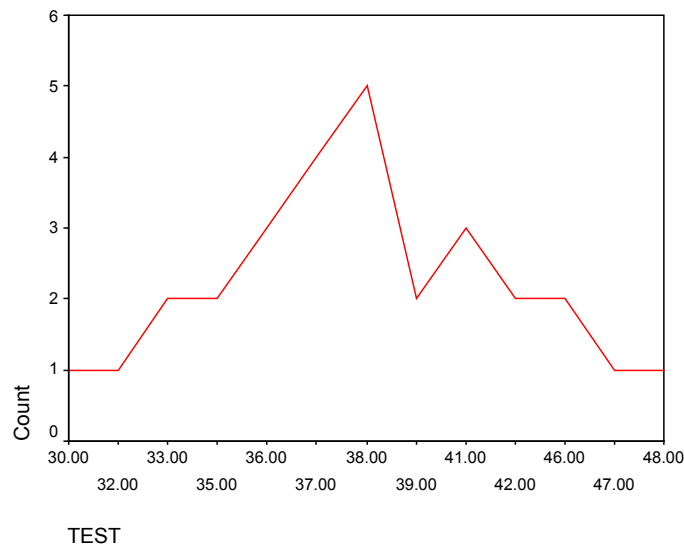
5. To approximate a histogram for part **b**, click *Graphs>Bar>Simple>Summaries for groups of cases>Define*.
6. Move **test** into the Category Axis box and be sure “N of cases” is indicated. The dialog box should appear as shown in step 3 except that the box is labeled as *Define Simple Bar* rather than *Define Simple Line*.
7. Click *OK* and the graph appears in the output Viewer window.
8. To construct a cumulative frequency curve for part **c**, repeat steps 2–4 except choose “Cum. N of cases” rather than “N of cases” in the Line Represents box.
9. To construct a cumulative percentage curve for part **d**, repeat steps 2–4 except choose “Cum. % of cases.”

Note that SPSS does not graph values with 0 frequency. For example, in the first graph in which we used an SPSS line graph to approximate the frequency polygon, you can see that the horizontal axis does not show the score of 31 and the line does not drop back to the *X* axis to indicate a frequency of 0 for that score. The same is true of the other graphs, and this is one way that SPSS approximates the graphs as we defined them in the text.

The output showing the four graphs for Problem 1 should appear as follows. (Again, note that SPSS does not usually plot scores with a frequency of zero.)

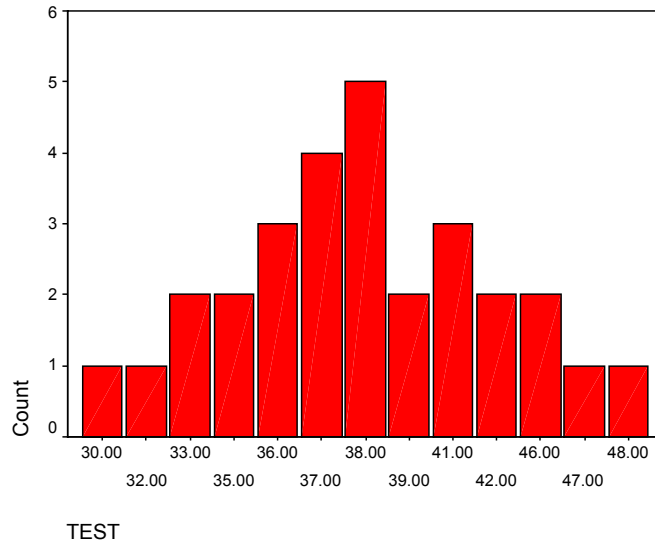
```
GRAPH
  /LINE (SIMPLE)=COUNT BY test
  /MISSING=REPORT.
```

Graph



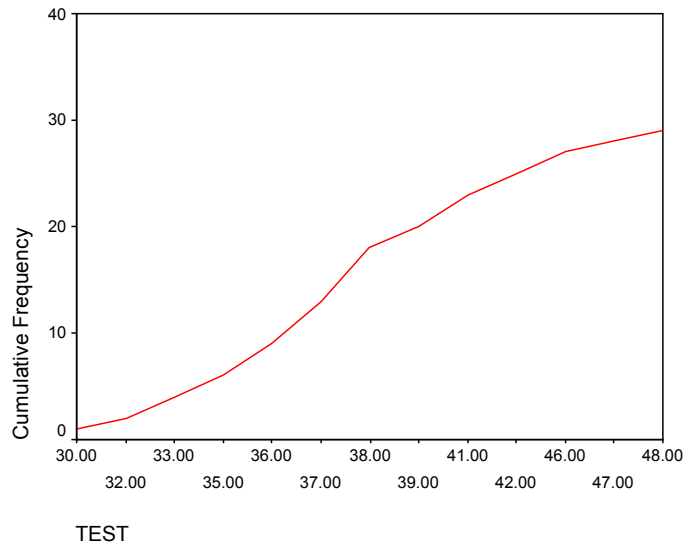
```
GRAPH
  /BAR(SIMPLE)=COUNT BY test
  /MISSING=REPORT.
```

Graph



```
GRAPH
  /LINE(SIMPLE)=CUFREQ BY test
  /MISSING=REPORT.
```

Graph

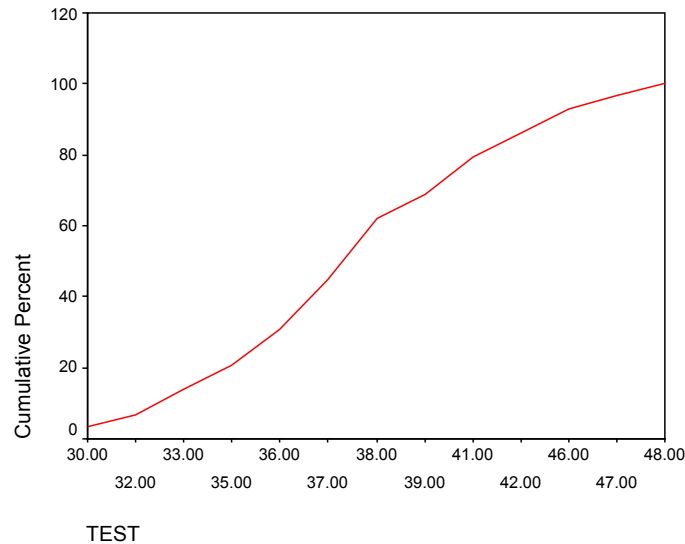


```

GRAPH
  /LINE(SIMPLE)=CUPCT BY test
  /MISSING=REPORT.

```

Graph



Exercises Using SPSS

1. Create a line graph using the data in Problem 2. Because of the way the data are provided, we have some special instructions for you to use to “trick” SPSS into giving the graph we want you to produce. Enter the data as a separate variable for **day** and as another separate variable for the frequency that you should call **freqcig**. The steps are as follows:
 - a. Click *Graphs>Line>Simple>Values of individual cases>Define*.
 - b. In the Define Simple Line dialog box, **freqcig** should be in the Line Represents box and **day** should be in the Category Labels Variable box.
 - c. Click *OK* and the graph should appear in the output Viewer window.
2. Create a stem-and-leaf plot of the speeds of cars given in Problem 8. Again, we need to provide some special instructions for where to look for the stem-and-leaf plot in SPSS. The instructions are as follows:
 - a. Start SPSS, enter the data, and name the variable **speeds**.
 - b. The stem-and-leaf plot is a plot option under the Explore procedure. Click as follows:
Analyze>Descriptive Statistics>Explore.
 In the Explore dialog box, move **speeds** to the Dependent List and click *Plots* in the Display box (lower left).

Click *Plots* (lower right) to open the Explore: Plots dialog box. Be sure “stem-and-leaf” is checked and None is checked under Boxplots (it won’t hurt if you leave it), then click *Continue*.

In the Explore dialog box, click *OK* and the output should appear.

CHECKING YOUR PROGRESS: A SELF-TEST

1. According to a graphing convention, which of the following is true?
 - a. The X axis should be three-fourths of the Y axis.
 - b. The Y axis should be three-fourths of the X axis.
 - c. The two axes should be approximately equal.
 - d. None of the above; there is no convention about the relative length of the axes.
2. True or False: Positively skewed curves have a tail toward the upper end of the X axis.
3. A histogram used to plot nominal scale data is called a
 - a. line graph.
 - b. stem-and-leaf plot.
 - c. bar graph.
 - d. frequency polygon.
4. The following are scores on a test in statistics. Make a stem-and-leaf plot.

66	88	89
86	91	88
76	89	76
74	55	95
93	54	80
80	90	29
65	92	69
86	92	84

5. The latency for rats to leave a lighted platform was timed on four trials. Their average latency in seconds over the four trials was as follows: Trial 1, 10.5; Trial 2, 4.6; Trial 3, 3.8; Trial 4, 4.1. Graph the data.

6. A marketing student tabulates all the cars that pass through an intersection according to country of manufacture. The data are as follows: United States, 43; Japan, 27; Germany, 10; Sweden, 3; Korea, 6; unknown, 4. Graph the data.

7. On the final exam in a typing class, students were required to type as many words as possible in a 1-minute test.
- Using the following frequency distribution, make a frequency polygon.
 - Plot a cumulative percentage curve and use it to show about what percentage typed 65 or fewer words per minute.

X	f
75	1
74	1
73	2
72	2
70	3
69	2
68	4
65	4
64	3
63	2
62	2
57	1
55	1