

## SORTING AND DISPLAYING THE DATA

After collecting the data, the researcher's first task is to organize and present them in an understandable form. Piles of answer sheets, the collected test scores from 100 secondary school students studying English in Hong Kong, or tape recordings of ten hours of conversation of child language learners are difficult to interpret unless the data are organized.

The statistics used to summarize data are called *descriptive statistics*. Beginning in this chapter, we will discuss the principles of descriptive statistics which allow us to describe *sample* data. The description summarizes only the sample data. However, after you have looked at the data summary, you will undoubtedly want to know whether the results mean anything for other child second language learners or whether the 100 Ss have given you data that is representative of all secondary students in Hong Kong. The principles that let you expand the findings of the sample to predictions about other learners in the *population* are those of *inferential statistics*. These statistics, which are based on theories of probability, tell you how confident you can feel in generalizing the findings from sample subjects back to the total population the sample represents. These principles will be discussed beginning in Chapter 8.

In order to sort and display the data in a meaningful way, most researchers go through three steps: coding the data, doing the numerical computations, and preparing a final display. Once the data have been coded, descriptive statistics will be used to help organize the data. The final data display may be in graph form, table form, arithmetic form, or all three. With the data in a clear, shorthand form, it won't be necessary to listen to the ten hours of tape recordings or look at 100 test papers. If the description is clear and informative, we will know far more about the data from the description than we could get from an examination of the unorganized raw data.

### CODING THE DATA

The way the data are coded will depend, in part, on the scales you have used to measure the variables. Nominal, ordinal, and interval scales are coded in slightly different ways.

**Nominal Data.** If the attribute has only two levels, the data can easily be coded as a series of + or - marks. Suppose you did a survey of foreign students entering the university and wanted to show how many had to take ESL classes and how many were exempt on the basis of the placement exam. The nominal attribute might be called +/- ESL student. You might code the raw scores by going through the list of Ss and their placement test scores:

Raw data	+	-	+	+	-	-	+	...	-	(+ = ESL, - = exempt)	
Coded	1	2	1	1	2	1	2	2	1	...	2

It is possible that the nominal variable being coded will have several levels. For example, if the attribute is +/- native speaker of English, the + group might include students from England, America, Canada, Australia, etc., and the - group might include students from many different countries. In this case, the 1-2 coding system won't work and other numbers (none of which have numerical value but which rather serve to index group membership) will have to be assigned. When coding the data, you may end up with several nationalities for which you have only one or two Ss. Sometimes these can be logically grouped into larger units (e.g., Latin American or Southeast Asian). It is also legitimate to tally infrequent responses as "other" or "none." You may use a "no response" category but, if you collect the data carefully, you should not have to use such categories too frequently.

**Ordinal Data.** To assign the data to an ordinal scale, the Ss (or responses) have to be ranked in some way, and coded for that rank. Let's assume that you want to find out how bilingual people are, how much bilingualism they possess. Assume further that you cannot give a test but rather must rely on some sort of ordinal measurement of bilingualism. How can you identify the amount of bilingualism without a test measure? And how can you code the information? The answer is that you must somehow rank the Ss' degree of bilingualism. There are three traditional ways of obtaining this ranking.

The first method is by judgment; you simply ask someone who should know to serve as a judge. For example, if you want to know how bilingual a group of Ss are, ask a native speaker to listen to taped samples and judge each S as excellent/good/fair/poor/terrible or as nativelylike/intelligible/nonintelligible, or some other set of terms you select for the scale. The problem with this procedure is that it's hard to get reliable judgments, especially when differences are small. It is also hard to make relative judgments over a long listening period. You may not find a good judge; so it is safer to find several and test to make sure they concur in their ratings.

A second method is to count unequal elements. We simply ask the Ss many questions and record their answers. There may be five questions on article usage and one on plurals. No matter. We assume that by asking enough questions, we will get a relative notion of the Ss' language proficiency. We presume that the S who knows the most is the most bilingual, and that the others can be rank-ordered beneath this S. We may be wrong, of course, but the convention does allow us to assign students to an ordinal scale of bilingualism and code them according to their ranks.

A third way of assigning numerical value to how much bilingualism our *Ss* possess is by arranging a series of questions in an order of difficulty. The more extreme the question, the less likely the *S* will be able to answer it. Suppose we want to rank people not on how bilingual they are but rather how multilingual they are. Say that we took a carefully chosen sample of the population of Chicago and asked them these questions:

1. Do you know some words in another language?
2. Can you say "good morning" in three languages?
3. Can you understand three languages?
4. In the languages you understand, can you say "The pen of my aunt is on the table"?
5. Can you translate these six questions into three languages?
6. Can you translate and answer these six questions in six languages?

Most *Ss* will respond "yes" to question 1, but the further along we go in the questions, the lower the number of "yes" answers will become. So it is possible to decide on some sort of degree of multilingualism in our sample on this basis. (Then we can write an article for the Sunday edition of the *Tribune* on multilingualism entitled "How Multilingual are *you*?") This procedure of judging ordinal data within a scale is the basis for the Guttman scale, which we will discuss in Chapter 14 on Implicational Scaling.

Ordinal measurements cannot be compared easily but they can be ranked with respect to one another. The problem is how to show the ranking. The direct way is  $1 < 2 < 3$  or  $3 > 2 > 1$  ( $>$  = greater than, and  $<$  = less than. Just remember that the arrow points toward the smaller one). A whole page of "Mary's pronunciation is better than Gene's pronunciation is worse than Anne's pronunciation" would not be very helpful. Instead, you could list the names in rank order and code each by its rank number. However, the more conventional procedure for coding such ordinal data is not to give a rank order to each student in the group but rather to assign an ordinal judgment scale of A, B, C, etc. Just because we have assigned the ordinal data (rank order) to a grade scale does not mean that we have made the data more exact (not even if we assign scale numbers of 1, 2, 3, 4 instead of A, B, C, D). It only means that we have coded it to a scale.

**Interval Data.** Since experimental research involves at least two variables, often interval scaled, it is unlikely that your data can be coded as simply as those presented thus far. To sort the data, you will need to prepare a data matrix. To do this, it's helpful to have a data sheet, but you can construct your own. The sheet is divided into rows and columns to form the cells of the matrix. The number of rows and columns depends on how many *Ss* and how many variables you have in your project.

For example, suppose you gave a dictation task to your ESL class. Suppose, further, that you scored their compositions using the number of grammatically correct sentences. Finally, suppose that in order to test your hypotheses, you must keep male and female scores separate. You have assigned a 1 to males and

a 2 to females. You have, then, three kinds of information for each *S*. To prepare a data matrix you:

1. Form a rectangle with horizontal and vertical lines.
2. List your *Ss* by number in the first column.
3. Enter each *S*'s scores in the appropriate columns.

	Ss	Sex	Dictation	Error-free sentences
	1	2	25	40
	2	1	36	52
	3	1	20	60
	4	2	18	20
	5	2	40	45

Diagram labels: "Rows" points to the vertical axis on the left; "Columns" points to the horizontal axis at the bottom; "Cells" points to the individual data entries in the rightmost column.

In the above matrix, for instance, we have five (horizontal) rows, 4 (vertical) columns, and 15 cells. Such a table is much easier to understand than going through 25 dictation sheets and 25 compositions. However, data are seldom displayed in data matrices. Instead, the matrices are used as the first step in sorting out the data so that the computations can be done more easily.

### SIMPLE NUMERICAL COMPUTATIONS

Once you have coded the data for your variables, the second step is to look at simple frequencies. For nominal data, this means tallying the 1's and 2's from your coded data. For example, if you were doing an experiment using a sample of Chicano graduate students at San Jose State and wanted to know how many men and women you should draw for the sample, you would first want to know the number of each in the population of Chicano graduate students. You'd check the number of each, coding them as 1-2, and tally them. Suppose the total for males turned out to be 255 and the number for female, 60. The total number of Chicano graduate students (hypothetically), then, is 315.

In addition to these tallies, you could present the information as a *ratio*:

$$\text{Ratio} = \frac{\text{number of females}}{\text{number of males}} = \frac{60}{255} = .24$$

We can say that the ratio of females to males is .24. Since we are talking about people, it is hard to say there is .24 of a woman enrolled for every 1.00 man. So

we usually change this figure by multiplying by 100. This tells us that there are 24 women enrolled at the university to every 100 male Chicanos, and the sample you select for your experiment should reflect that ratio.

It is also possible to present such data using proportions and percentages. Suppose we discover that Anglo parents of children enrolled in a local school project—a project where Anglo children receive their education in Spanish—give predominantly instrumental rather than integrative responses in explaining why they enrolled their children in a Spanish-speaking kindergarten. (Integrative responses indicate the learner wants to identify with or become a part of the community that speaks the language. Instrumental responses are those that indicate the person wants to use the language for advancement, careers, or business purposes.) We want to present these data numerically.

By tallying the frequencies for each type we find that there were 200 integrative responses and 325 instrumental responses (a total of 525). We might present these data as either proportion or percentage figures.

$$\begin{aligned}\text{Proportion} &= \frac{\text{number of instrumental responses}}{\text{total number of responses}} \\ &= \frac{325}{525} \\ &= .618\end{aligned}$$

As you can see, the proportion figure, .618, can easily be changed to a percentage figure by multiplying it by 100:

$$\begin{aligned}\text{Percent} &= (100) \frac{\text{number of instrumental responses}}{\text{total number of responses}} \\ &= (100)(.618) \\ &= 61.8\%\end{aligned}$$

Simple frequencies are useful first ways of reducing the data, but such counts do not always give us a precise picture of the data. This is especially true when the data are obtained from different groups. For example, suppose it was our job to report the frequency of high scores on reading tests for two different schools. Imagine that two students in each school received very high marks. If we know that the frequency of high marks is the same in the two schools, we might feel that they were similar. However, if we knew that there were only 10 students in one school and 2 of them got very high marks, that would be very different from 2 students in 80 students in the second school. To show this difference, we would compute the relative frequency.

To find *relative frequency* we divide the frequency of each score by the total number of *Ss* in the group. Suppose our *Ss* obtained the following scores:

Score	Frequency	Relative frequency	
		Compute	Result
80	5 Ss	$5 \div 60$	.08
70	15	$15 \div 60$	.25
60	20	$20 \div 60$	.34
50	15	$15 \div 60$	.25
40	5	$5 \div 60$	.08
$60 = N$			

In large studies such as language policy surveys, relative frequency may also be presented in terms of *rate*. Rate is used to show how often an event happens compared with how often it might happen. For example, we might use rate to show the number of people who *do* learn a language compared with the number of people who *might* learn the language. The comparison is made by dividing the number who do learn by the number who could have learned. This fraction is multiplied by some standard population number such as 100, or more typically, 1,000.

$$\begin{aligned} \text{Rate} &= (1,000) (\text{relative frequency}) \\ &= (1,000) \frac{\text{number who learn}}{\text{number of potential learners}} \end{aligned}$$

Let's assume we've collected the data displayed in Table 5.1. Look at the "Total ESL speakers" column first. There are more ESL speakers who are in the age group 25–44. However, if we know the population of Galaxy people in each of

Table 5.1. Galaxy people who speak English as a second language

Age	Total ESL speakers	Rate per 1,000 population
0–5	2,905,000	189.5
6–16	11,915,000	314.9 ←
17–24	10,436,000	227.9
25–44	14,905,000 ←	277.8
45–64	7,856,000	218.5
65+	1,400,000	201.4

the age groups, we can show how many English speakers there are in each age range per 1,000 population. While there are more English speakers in the 25–44 age range, the rate column shows you there is a greater proportion of ESL speakers in the 6–16 age group. Therefore, your chances of finding someone to talk to in English is best if you walk up to someone between 6 and 16 years of age.

Rates are often computed for two reasons: (1) to compare different populations with respect to frequency of some variable; and (2) to compare the same population at different times. For example, say that twenty years after we

collected the above data, we conducted another survey of Galaxy people. We might then want to compute the change over time. To do this, we calculate percent change.

$$\text{Percent change} = (100) \frac{n_2 - n_1}{n_1}$$

Notice the letter  $n$  in the formula. It just stands for "number." So the directions tell us that if we want to know the percent change over a period of time, we take our first  $n$  ( $n_1$ ), which is our beginning frequency, and subtract it from our new frequency number ( $n_2$ ). Let's imagine that in our new census 19,742,000 children in the 6 to 16 age group now speak English. So we subtract our earlier census figure from that. The directions also tell us to divide the answer by the beginning census figure. Since we want a percent figure, we then multiply the result by 100.

$$\begin{aligned} \text{Percent change} &= (100) \frac{n_2 - n_1}{n_1} \\ &= (100) \frac{19,742,000 - 11,916,000}{11,916,000} \\ &= (100)(.66) \\ &= 66\% \end{aligned}$$

Once we have found that the percent change is 66%, our next question is why. We might want to attribute the change to Peace Corps intervention, or a desire of parents to be sure their children can qualify for admission to the one and only university in the Galaxy which requires English, or to prizes offered by the government. Note, however, that the "why" answer is *not* in the data so it is *not* accurate to say that the change *is* due to Peace Corps, or *is* due to prizes offered by the government. All these data show is that a change took place. If we wish to show a relationship between these variables, we must make a hypothesis or hypotheses and test them. The above data (if that's all we have) will allow us only to speculate about reasons for the change.

If we want to show the standing of any particular score in a group of scores, we would prepare a *cumulative frequency* distribution. This will show us how many scores fall below that particular point in the distribution. It is also the basis for calculating percentile scores. Let's suppose that we wanted to know the cumulative frequency and percentile figures for student placement in English classes at the university. See Table 5.2. In the table, simple frequency is

Table 5.2. Foreign student placement in English classes based on entrance exam scores

Placement requirement	$f$	$F$	Percentile
0 quarter terms	102	392	100
1 quarter term	130	290	74
2 quarter terms	77	160	41
3 quarter terms	45	83	21
4 quarter terms	38	38	10

symbolized by the letter *f*. To get the cumulative frequency, which is symbolized by *F*, add the frequency of the group required to take four courses to the group required to take three. The total, 83, is the *F* for the second level. Then add the frequency of the next group to get the *F* for that level. So, that's  $38 + 45 = 83$  and then  $83 + 77 = 160$ . Look at the arrows in Table 5.2 if this is not clear. This allows us to see how many *S*s are at or below any particular level.

To get the percentile, divide the *F* of the level you want to check by the sum of the frequencies (*N*) and multiply by 100 (just as we multiply for percentages). The directions are:

$$\text{Percentile} = (100) \frac{F}{N}$$

In Table 5.2, to get the percentile for students at the 1 course level, we look first for the total number of observations (392), the final *F* number obtained by adding all the level frequencies. The directions say to divide the *F* for the level (290 is the *F* for 1 course required) by *N*. So, students who are required to take 1 English course are at the 74th percentile level:

$$\begin{aligned} \text{Percentile} &= (100) \frac{F}{N} \\ &= (100) \frac{290}{392} \\ &= (100)(.739) \\ &= 74\text{th percentile} \end{aligned}$$

The percentile score means that anyone who received a score at the 74th percentile can say that 74% of the students who took the test scored at or lower than that level. If you have taken an ETS (Educational Testing Service) test, the Miller Analogies Test, or the GRE, you probably received a card with your score and a percentile level based on scores of *S*s who took the test previously and established norms for students from your major (education, engineering, etc.). It usually doesn't say how well you did compared with students who took the test the same day that you did, but rather how well you did compared with *S*s from your area of expertise. Some of these tests do give you scores that tell you what percent of the students who took the test at the same sitting as you scored at the same point or below. Always check the "fine print" if you want to know exactly what group your scores have been compared with.

To check your retention of the definitions and computations given in this chapter, see if you can interpret the following:

Score	Frequency ( <i>f</i> )	Relative frequency	Cumulative frequency ( <i>F</i> )	Percentile
80	5	.08	60	100
70	15	.25	55	92
60	20	.34	40	67
50	15	.25	20	33
40	5	.08	5	8



Given the score column and the  $f$  column, would you be able to compute the figures for the remaining columns?

These simple computations for frequencies, ratios, rate, percentages, proportions, and percentiles give us and our readers an initial understanding of the data. For many readers and researchers, a visual display of the data is also useful.

### DISPLAYING THE DATA

The most conventional way to display simple frequency data is in table form. Nominal data are easily presented in this way. For example, we could display the data we coded earlier as shown in Table 5.3. The table could be arranged so that it ran in the other direction, as in Table 5.4, an attribute table on +/- literate frequency. Ribbonlike tables (if they spread across a page) are usually considered unattractive, as are long thin tables. They count as "bad style."

Table 5.3. Distribution of ESL students in the entering foreign student population

Groups	$f$
ESL students	290
Non-ESL students	102
Total	392

Table 5.4. World literacy estimates,\* UNESCO, 1970

Illiterate	Literate	Total
810 mill.	1,525 mill.	2,335 mill.

\*Literacy is equated with 4th grade graduation.

Another way to present the data is to convert the distribution to a visual form. There are two basic types of graphic displays: graphs (either histograms or bar graphs) and polygons (line drawings). The techniques for constructing them are almost the same:

1. Draw two axes (a vertical and a horizontal line).
2. On the horizontal line, enter the scores for the variable.
3. On the vertical line, enter the frequency of each of these scores.
4. Construct the graph or polygon around these frequency points.

To illustrate how this works, let's suppose that we have assigned numerical values to the performance of ESL teacher-trainees according to the ATESL (Anguish in Teaching English as a Second Language) system. The 15 trainees scored:

20, 22, 22, 24, 24, 24, 25, 25, 25, 25, 27, 27, 29, 30, 30

Tallying the data would give us a frequency distribution of the scores as follows:

Scores	$f$	Scores	$f$
30	2	24	3
29	1	22	2
27	2	20	1
25	4		

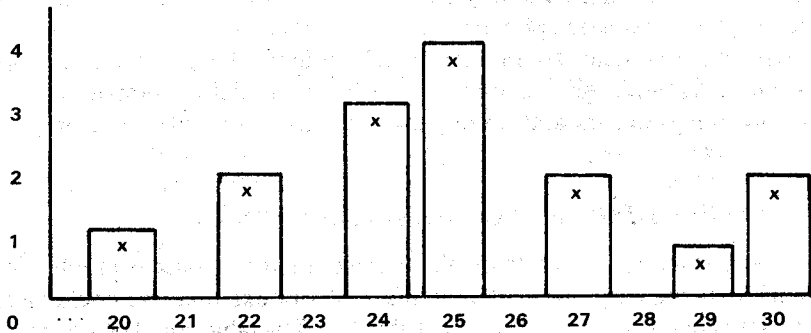


Figure 5-1

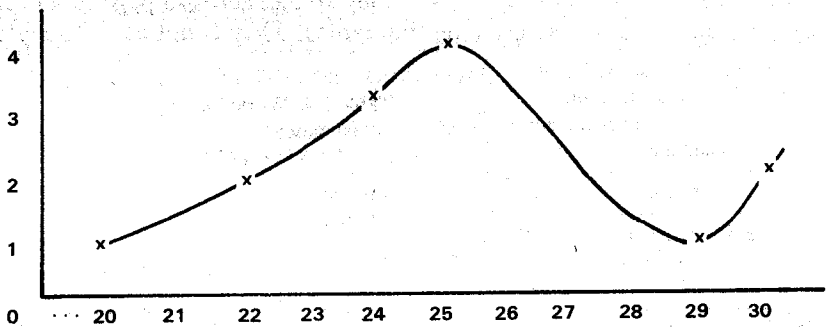
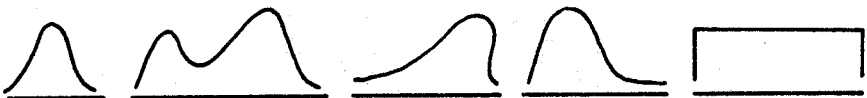


Figure 5-2

Constructing an axis and entering the frequencies would give us a bar graph that looks like Figure 5-1. If we connect the midpoints of the bar graphs by a smooth line, then we have a frequency polygon (Figure 5-2). The shape of the frequency polygon provides useful, easily understood information about the data. By looking at a graph or a polygon, you can quickly get a rough idea of what has happened in the data. It is a picture, in a way, of the data. The shape of the curve, once you learn to read it, is an easy indicator of the information you need. For example, if the distribution of the scores is such that one half of the curve looks like the other half, you know that it is a symmetric distribution. One half is the mirror of the other. This is called the bell-shaped curve (or the beautiful bell-shaped curve because everyone is pleased to get a distribution that approaches this normal distribution of scores). Most of the scores fall in the middle of this curve, the high part. There is a nice gradual slope of scores below and scores above that one midpoint. The frequency distribution (the distribution of the scores) can take very different shapes:



How do these polygons come about, and what do the shapes represent?

**Table 5.5** Distribution of Los Angeles schools by national percentile intervals on 6th grade reading, 1971

National percentile intervals	<i>f</i>	National percentile intervals	<i>f</i>
90 – 100	0	40 – 49	69
80 – 89	3	30 – 39	55
70 – 79	29	20 – 29	56
60 – 69	51	10 – 19	94
50 – 59	75	0 – 9	1

Table 5.5 shows the frequency figures for the standing of each Los Angeles school on a nationally administered reading test. The scores from each school were averaged and then the school was tallied as having a score which placed it at some percentile level of reading scores for the nation. If a school fell in the top 90–100th percentile, the school as a whole did very well indeed. If the school fell in the 0–10th percentile level, they are at the bottom on reading for the country as a whole (as measured by this test). See Table 5.5.

We can change Table 5.5 into a graph such as the histogram in Figure 5-3. (A histogram is simply a bar graph where there is data at each score interval so that the bars can be connected). The frequency polygon keeps the general form of the distribution of the schools but smooths out the line into the general shape shown in Figure 5-4. You will notice that there are two “peaks” in this example. (No, fans of “The Little Prince,” it is not a picture of a boa constrictor swallowing an elephant!) When there are two peaks to the polygon, you would immediately look for a reason for such a distribution. It is *not* a normal distribution. The curve shows us that there are two different patterns in the sample. In our earlier beautiful bell-shaped curve we have a sample where everyone scores about the same, with a gradual spread of scores higher and lower than the average middle

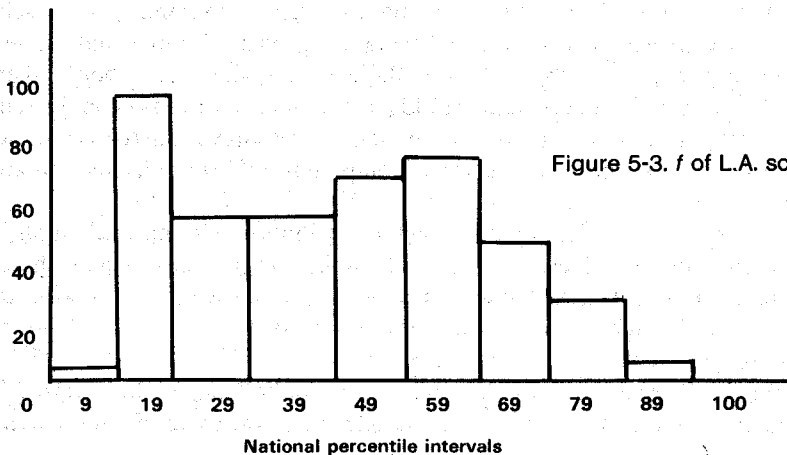


Figure 5-3. *f* of L.A. schools

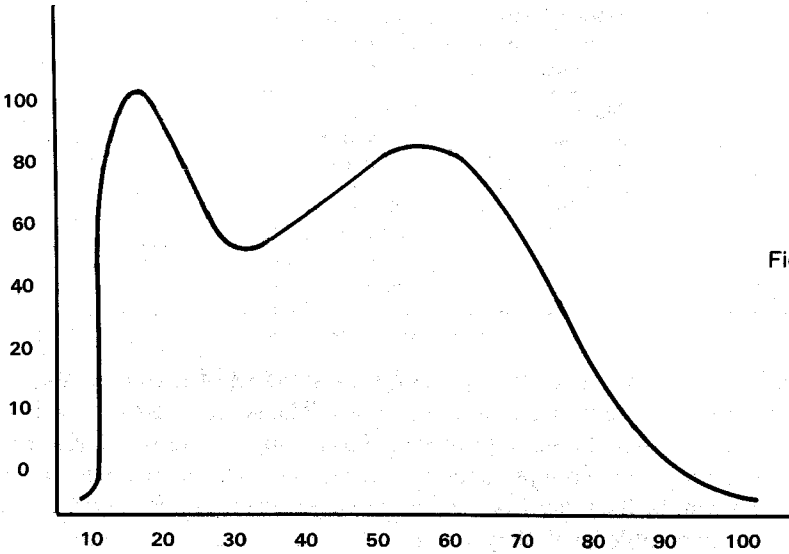


Figure 5-4

score. In this case, we have two peaks with a spread out from each. The bell-shaped curve is called a unimodal distribution because it has only one midpoint. Our double-peaked curve is called a bimodal distribution because it has two. When we have a bimodal distribution of scores it means that something is unusual in the sample. One would expect that the high point in the polygon would be at about the 50th percentile mark; that is, we'd



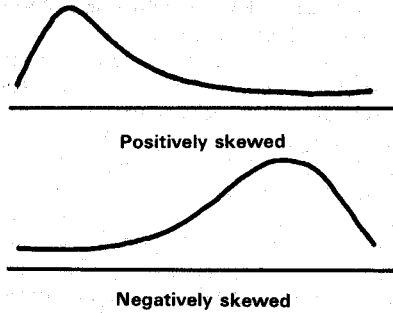
expect that since we have a large number of schools represented (433), they would be representative of the nation as a whole and so fall about in the middle. The second peak looks as if it fits into that pattern pretty well. The first peak, however, is quite different. One explanation might be that many of the schools have newly arrived children whose first language is not English and the tests of reading are, of course, in English. If there were many such schools, then we would expect such a curve. One would have to have information on the schools in the first peak to find out if that were the case. Whatever the reason, it is clear from the polygon that there are two completely different distribution curves within it.

It is also possible that we might get a distribution with no peak at all. This means that all the scores occurred with similar frequencies. Again, this is an unusual situation. It may happen, however, if we have a fairly small sample. That is, two Ss might score 2 points on some test, two score 3, two score 4, etc. If more Ss were tested, a pattern might emerge.

Finally, there are distributions which are not bell-shaped but skewed. This means that most of the scores cluster at one end or the other of the scale. In our

beautiful bell-shaped curve, we said that the lines curve away from the peak on each side. These lines curving away from the peak are called "tails." The shape of the tails in the bell-shaped curve are the same (remember, it's symmetric). However, if the scores are distributed so that most of the scores cluster at one end, the tails will not show the same curve on both sides of the peak. The tail of

the distribution shows us the direction toward the high scores. Most of the scores in the polygon (*top*) were low while a few scores were very high. Those high scores positively skew the curve. A negatively skewed distribution (*bottom*) means that most of the scores were high except that a few very low scores skewed the scores negatively. These



distributions are not symmetric, but asymmetric distributions. Such distributions show you that while most *Ss* in your group scored in similar ways, there are some small portion who behave quite differently from the rest of the *Ss*. That number is small, not enough by any means to make you say that there are different groups within your sample, just that there are a few "outliers" who do not seem to perform in the same way that most of your *Ss* do.

In this chapter we have talked about the first steps in sorting and displaying data. We have discussed ways of coding the data, some of the basic statistics that you might use in describing data, and traditional ways in which data are displayed. We also noted that bar graphs and frequency polygons give us additional information on the form, the distribution of scores in the data. In the following chapter we will consider measures of central tendency (where the middle point in the data falls), and dispersion (how the data spread out from that central point).

### ACTIVITIES

1. To decorate your office you decided to make a frequency table, a bar graph, and a frequency polygon for *The Language Distribution of Students in My ESL Class*. You have 10 Vietnamese, 20 Mexican, 15 Venezuelans, and 5 Cantonese students. Show the results of your handiwork, placing the groups in this order.
2. Your research question is: Do *Ss* from three different language groups (Germanic, Romance, and Malayo-Polynesian) respond differently to a test of English idioms? You've decided to include sex, years of English study, and length of residence in the United States as moderator variables. Draw and label a data matrix that will allow you to include all this information.
3. You believe that *Ss* are moved up through beginning to intermediate to advanced levels too quickly to really make the gains that are necessary in order to succeed in learning the material covered. The placement test is used to place students at these levels as well. So your research question is whether *Ss* who are placed low and then are passed to the higher levels are really at the same level as those students who test in at the high level to start with. You have decided to administer a cloze test, take a composition sample (which you will rate on the basis of error-free *T*-units), and note the final exam grade in the class and the teacher's grade for your two groups:

- the continuing students (those who have worked up to the top level by taking classes) vs. the placed students (those who tested in at the top level). Draw and label a data matrix that will allow you to include all this information. Are there other variables that you think ought to be included as moderator variables in this study? If so, list them.
4. You have found out that there are 110 universities which offer TESL programs. Of these, 90 offer M.A. degrees and 4 offer Ph.D.'s. What percent of the schools offer M.A.'s? What proportion offer Ph.D.'s? Last year, 274 M.A.'s were enrolled and 34 Ph.D.'s. What is the ratio of M.A. to Ph.D. students?
  5. In gathering data on Ss attending a local Community Adult School ESL program, you found that 580 Ss were enrolled; 326 women and 254 men. What is the ratio of men to women? What is the percentage of each in the program?
  6. In problem 2, the scores of your first 20 Ss on the idiom test were: 13, 16, 17, 17, 18, 18, 18, 21, 22, 24, 24, 24, 25, 25, 25, 25, 26, 26, 29, 30. Arrange these to show (1)*f*, (2)*F*, (3) percentile for those who scored 24.
  7. You have collected the following raw data on WPM reading speed for your Ss. Construct a frequency table, histogram, and frequency polygon for the data. You will have to decide on the best interval to use for your reading speed scale (you won't want to mark it off in 1 WPM intervals). Name the polygon shape. 180, 169, 173, 148, 164, 155, 177, 133, 148, 89, 193, 135, 197, 152, 130, 132, 133, 216, 182, 133, 137, 131, 133, 134, 187, 171, 122, 111, 125, 182.
  8. In problem 7, at what percentile would a S be if she scored 133? If she scored 189?
  9. Collect test papers from one of your classes this week. Display the data in terms of a frequency distribution table and a frequency polygon. What is the name of the shape of the polygon?

Suggested further reading for this chapter: Johnson, Shavelson.