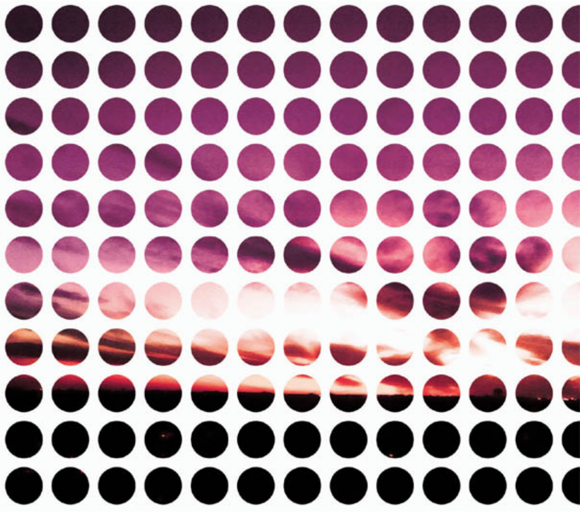

Pearson New International Edition

Research Methods
A Process of Inquiry

Anthony M. Graziano Michael L. Raulin

Eighth Edition

PEARSON®

Pearson New International Edition

Research Methods
A Process of Inquiry
Anthony M. Graziano Michael L. Raulin
Eighth Edition

PEARSON®

TABLE 2 Sample Data From 24 Participants

PERSON	AGE	INCOME ($)	NUMBER OF TIMES VOTED IN LAST 5 YEARS	GENDER	POLITICAL AFFILIATION
1	28	32,000	6	M	R
2	46	50,000	4	M	D
3	33	44,000	0	F	D
4	40	45,000	5	M	R
5	21	30,000	1	M	R
6	26	35,000	0	F	O
7	39	42,000	6	M	O
8	23	34,000	0	F	D
9	20	27,000	1	M	O
10	26	31,000	2	M	R
11	29	39,000	6	F	R
12	24	34,000	2	M	D
13	34	44,000	2	M	O
14	35	45,000	3	M	O
15	52	46,000	8	M	O
16	31	39,000	4	F	D
17	30	43,000	6	M	R
18	45	47,000	7	F	D
19	18	28,000	0	M	O
20	29	44,000	7	M	R
21	26	38,000	6	F	D
22	23	37,000	3	M	O
23	47	48,000	7	M	D
24	53	51,000	8	M	D

Note: R, Republican; D, Democrat; O, other.

What type of data does each of these variables generate? Consider the variables age, income, and the number of times the participant voted. Each of these measures has the property of magnitude; 34 is older than 25, $35,000 is more than $28,000, and so on. All three measures have the property of equal intervals; the difference in age between 25 and 20 is the same as the difference between 38 and 33. The variables also have a true zero point; a person whose income is zero does not earn anything; a person who has voted zero times in the last 5 years has not voted in that time. These variables are measured on ratio scales and therefore generate score data. The other two variables, gender and political affiliation, are measured on nominal scales producing nominal or categorical data; there is no meaningful way of ordering the categories in a nominal scale.

Frequency Distributions

Nominal and Ordered Data. For most nominal and ordered data, statistical simplification involves computing **frequencies**: the number of participants who fall into each category. We organize the frequencies into **frequency distributions**, which show the frequency in each category. Table 3 shows the frequency distribution of gender for the data from Table 2. In any frequency distribution, when we sum across all categories, the total should equal the total number of participants. It is helpful to convert frequencies to percentages by dividing the frequency in each cell by the total number of participants and multiplying each of these proportions by 100, as was done in Table 3.

Cross-tabulation is a useful way to categorize participants based on more than one variable at the same time, such as categorizing participants based on gender and political affiliation. Cross-tabulation can help the researcher to see relationships between nominal measures. In this example, there are two levels of the variable gender (male and female) and three levels of the variable political affiliation (Democrat, Republican, and other), giving a total of six (2 × 3) possible joint categories. We arranged the data in a 2 × 3 matrix in Table 4, in which the numbers in the matrix are the frequency of people in each of the joint categories. For example, the first cell represents the number of male Democrats. Note that the sum of all the frequencies in the six cells equals the total number of participants. Also, note that the row and column totals represent the **univariate** (one-variable) frequency distribution for the political affiliation and gender variables, respectively. For example, the column totals in Table 4 of 17 males and 7 females represent the frequency distribution for the single variable of gender and, not surprisingly, are the same numbers that appear in Table 3.

Score Data. The simplest way to organize a set of score data is to create a frequency distribution. It is difficult to visualize all 24 scores at a glance for the voting variable shown in Table 2. Some of the participants have not voted at all during that time, and two participants voted eight times, but where do the rest of the participants tend to fall? A frequency distribution organizes the data to answer such questions at a glance. There may be no participants for some of the scores, in which case the frequency would be zero. Table 5 shows the frequency distribution for this voting variable.

TABLE 3 Frequency of Males and Females in Sample in Table 2

	MALES	FEMALES	TOTAL
Frequency	17	7	24
Percentage	71	29	100

TABLE 4 Cross-Tabulation by Gender and Political Affiliation

	MALES	FEMALES	TOTAL
Democrats	4	5	9
Republicans	6	1	7
Other	7	1	8
Totals	17	7	24

TABLE 5 Frequency Distribution of Voting Behavior in Last 5 Years

NUMBER OF TIMES VOTED	FREQUENCY
8	2
7	3
6	5
5	1
4	2
3	2
2	3
1	2
0	4

If there are many possible scores, then the frequency distribution will be long and almost as difficult to read as the original data. In this situation, we use a **grouped frequency distribution**, which reduces the table to a more manageable size by grouping the scores into intervals. A grouped frequency distribution is required with a **continuous variable**, in which there are theoretically an infinite number of possible scores between the lowest and the highest score. Table 6 shows a grouped frequency distribution for the continuous variable of income, which ranges from $27,000 to $51,000. Grouping salary into $2,000 intervals yields 13 intervals.

TABLE 6 Grouped Frequency Distribution for Income

INTERVAL NUMBER	ANNUAL INCOME ($)	FREQUENCY
1	50,000–51,999	2
2	48,000–49,999	1
3	46,000–47,999	2
4	44,000–45,999	5
5	42,000–43,999	2
6	40,000–41,999	0
7	38,000–39,999	3
8	36,000–37,999	1
9	34,000–35,999	3
10	32,000–33,999	1
11	30,000–31,999	2
12	28,000–29,999	1
13	26,000–27,999	1

Graphical Representation of Data

A Chinese proverb states, "one picture is worth a thousand words" (Bartlett, 1980), and this is especially true with statistical information. **Graphs** can clarify a data set by presenting the data visually. Most people find graphic representations easier to understand than other statistical procedures. Graphs and tables are excellent supplements to statistical analyses.

We can represent frequency or grouped frequency distributions graphically by using either a **histogram** or a **frequency polygon**. Figure 1 shows a histogram and a frequency polygon representing the voting data summarized in Table 5. (We generated these graphs in just a few seconds using the *PASW for Windows* data analysis program.) Both the histogram and the frequency polygon represent data on a two-dimensional graph, in which the horizontal axis (***x*-axis** or **abscissa**) represents the range of scores for the variable and the vertical axis (***y*-axis** or **ordinate**) represents the frequency of the scores. In a histogram, the frequency of a score is represented by the height of a bar above that score, as shown in Figure 1(a). In the frequency polygon, the frequency is represented by the height of a point above each score on the abscissa. Connecting the adjacent points, as shown in Figure 1(b), completes the frequency polygon. To aid in the interpretation of histograms and frequency polygons, it is important to label both axes carefully.

It is possible to display two or more frequency distributions on the same graph so that one can compare the distributions. Each distribution is graphed independently with different colors or different types of lines to distinguish one distribution from the other. Figure 2 shows the distribution for the voting variable, graphed separately for males and females.

When group size is small, a frequency polygon or histogram is usually jagged, but will have an overall shape, like those graphed in Figures 1 and 2. As group size increases, the frequency polygon looks more like a smooth curve. We often describe data by drawing smooth curves, even though such curves are seen only when the group sizes are extremely large.

Figure 3 shows several smooth-curve drawings of various distribution shapes frequently found in psychology. Figure 3(a) shows a common shape for a **symmetric distribution**: a bell-shaped curve. Most of the participants are near the middle of the distribution. In symmetric distributions,

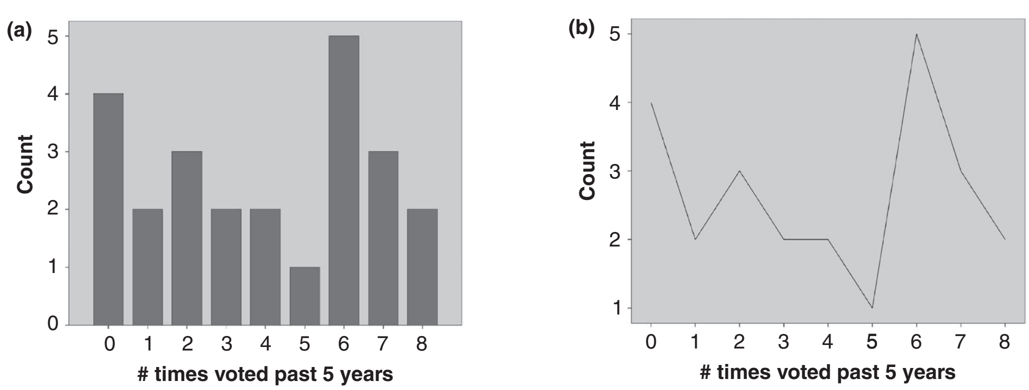

FIGURE 1 Histograms and Frequency Polygons. Graphing the distribution of scores with either a histogram or a frequency polygon helps the researcher to visualize the data.

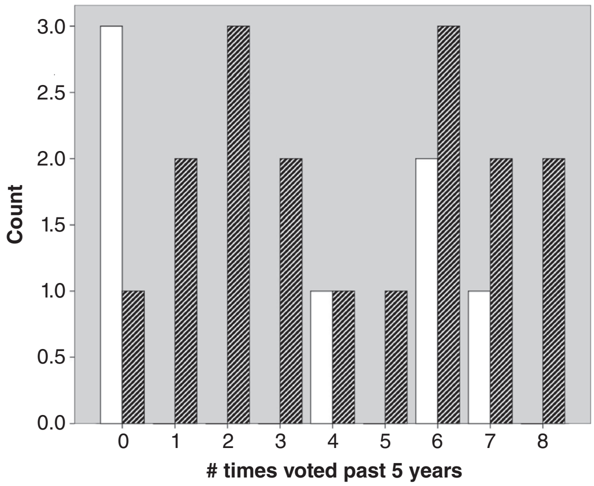

FIGURE 2 Comparing Two Distributions. Graphing frequency data from two or more groups on the same histogram or frequency polygon gives a visual representation of how the groups compare.

the right and left sides of the distribution are mirror images. Distributions with this bell-shape are **normal distributions**. Many variables in psychology form normal distributions, including measures of most human characteristics, such as height, weight, and intelligence.

In **skewed distributions**, the scores pile up on one end of the distribution [see Figures 3(b) and 3(c)]. The tail of the curve indicates the direction of the skew. In Figure 3(b), the curve is **positively skewed**, with most of the scores piled up near the bottom (the tail points toward the high or positive end of the scale). Figure 3(c) is **negatively skewed**. We might see a negatively skewed distribution on an easy classroom test, on which almost everyone does well and only a few people do poorly.

In addition to the shape of the curve, we also describe distributions in terms of the location of the middle of the distribution on the *x*-axis, which is called the **central tendency** of the distribution. We can also quantify the horizontal spread of the distribution, which is called the **variability** of the distribution. *The visual display of quantitative information* (Tufte, 2001) is an excellent book on the graphical presentation of data.

Quick-Check Review 2: Organizing Data

1. What are frequency distributions? With what kind of data can we use frequency distributions?
2. Define “cross-tabulation.”
3. What is the difference between frequency and grouped frequency distributions?
4. What type of variable requires a grouped frequency distribution?
5. What are the basic shapes of distributions found in psychology?

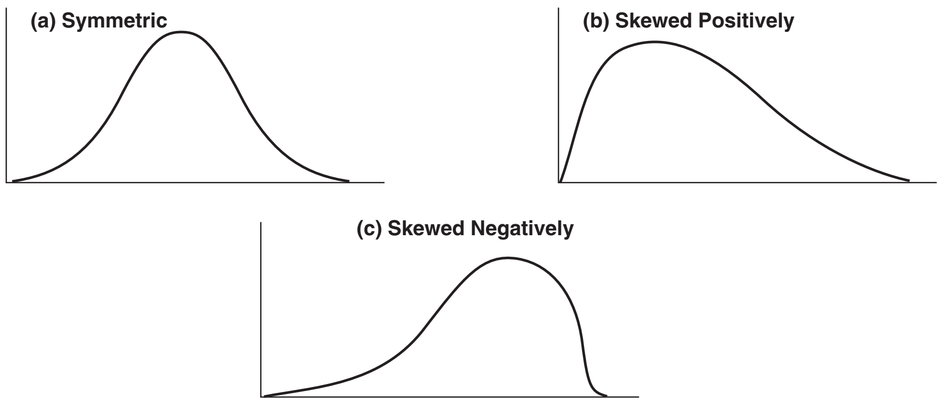

FIGURE 3 Symmetric and Skewed Distributions. Many measures yield the classic bell-shaped distribution shown in (a). When scores bunch up at either the bottom (b) or top (c) of the distribution, the distributions are skewed.

DESCRIPTIVE STATISTICS

Descriptive statistics serve two purposes. The first is to describe data with just one or two numbers, which makes it easier to compare groups. The second is to provide a basis for later analyses using inferential statistics. This section covers measures of central tendency, variability, and relationship and introduces the concept of standard scores.

Measures of Central Tendency

Measures of central tendency describe the typical or average score. They indicate the center of the distribution, where most of the scores cluster. Table 7 summarizes three measures of central tendency: mode, median, and mean.

The **mode** is the most frequently occurring score in the distribution. In the example shown in Table 1, the modes are 87 and 72 for Groups A and B, respectively. In a frequency distribution like the one in Table 5, we can determine the mode by finding the largest number in the frequency column and noting the score with that frequency. In Table 5, the mode is 6. A distribution may have more than one mode. If there are two, then the distribution is **bimodal**; if there are three, it is **trimodal**.

The mode has the advantage of being easy to compute, but it has the disadvantage of being unstable, which means that it can be affected by a change in only a few scores. We can use the mode with all scales of measurement.

TABLE 7 Measures of Central Tendency

Mode	Most frequently occurring score in a distribution
Median	Middle score in a distribution; the score at the 50th percentile
Mean	Arithmetic average of the scores in a distribution; computed by summing the scores and dividing by the number of scores