The Statistical Process Summary

KEY TERMS

- variability
- data
- statistical question
- statistical process
- categorical data
- quantitative data
- population
- sample
- survey
- observational study
- experiment
- bar graph
- circle graph
- frequency
- mode
- dot plot
- distribution
- symmetric
- skewed right
- skewed left
- clusters
- gaps
- peaks
- outliers
- stem-and-leaf plot
- histogram
- grouped frequency table

LESSON 1 What's Your Question?

Statistical problem solving begins with a statistical question. A **statistical question** is a question that anticipates an answer based on data that vary. **Data** are categories, numbers, or observations gathered in response to a statistical question. Statistics is a problem-solving process because it is about determining a possible answer to a question that has variability. In statistics, **variability** means that the value of the attribute being studied can change from one person or thing to another.

The **statistical process** has four components:

1. Formulating a statistical question.

   The statistical question posed should anticipate answers that will vary.
   Example: How many members do the clubs at my school have?
   Non-example: How many students are in the Chess Club?

2. Collecting appropriate data.

   Two types of variable data that can be collected are categorical and quantitative data. **Categorical data**, or qualitative data, are data for which each piece of data fits into exactly one of several different groups or categories. **Quantitative data**, or numerical data, are data for which each piece of data can be placed on a numerical scale and compared.

   A statistical question can be answered by collecting data from an entire population or, more commonly, from a sample of the population. A **population** is an entire set of items from which data are collected. A **sample** is a selection from a population.

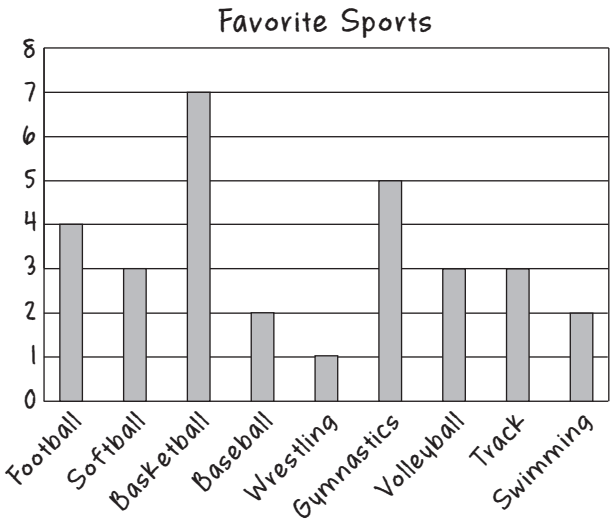
   Three common methods of data collection are surveys, observational studies, and experiments. In a **survey,** people are asked one or more questions. Similarly, in an **observational study**, the researcher collects data by observing the variable of interest. In an **experiment**, the researcher imposes a condition and observes the results.

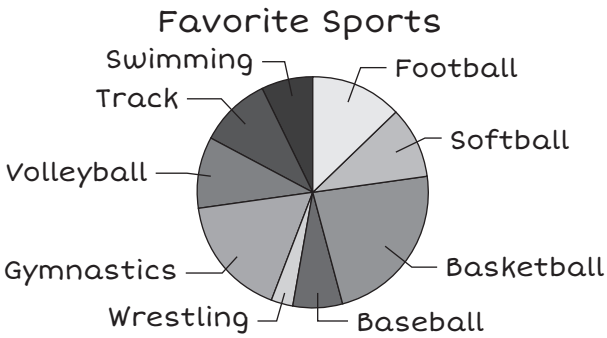
3. Analyzing the data graphically and numerically.

   After you collect the data, it is time to analyze and interpret the results. Analysis includes selecting the most appropriate graphical display and numerical summaries for your question and your method of data collection.

   A **bar graph** displays categorical data using either horizontal or vertical bars on a graph. The height or length of each bar indicates the value for that category.

   A **circle graph**, often called a pie chart, displays categorical data using sectors, or "wedges," of a circle. It shows how parts of the whole relate to the whole and how parts of the whole relate to the other parts. The area of each sector corresponds to the percentage of the part in relation to the whole.





To create graphs, you can determine the frequency of each response to a statistical question and record the frequencies in a frequency table. A **frequency** is the number of times an item or number occurs in a data set. Once the frequency is known, you can determine the mode. The **mode** is the value or values that occur most frequently in a data set.

4. Interpreting the results of the analysis.

You can use your analysis to make conclusions about the data. For example, from the graphs and table above you can conclude that basketball is the most popular sport among those that were surveyed.

Sport	Frequency (*f*)
Football	4
Softball	3
Basketball	7
Baseball	2
Wrestling	1
Gymnastics	5
Volleyball	3
Track	3
Swimming	2

LESSON 2

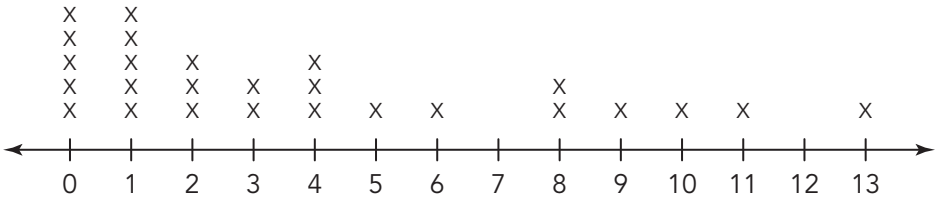
Get in Shape

One way to describe a set of quantitative data is by drawing a graphical display of the data.

A **dot plot** is a data display that shows discrete data on a number line with dots, Xs, or other symbols. Dot plots help organize and display a small number of data points.

In this example of a dot plot, the number line represents the number of gold medals won by countries in the 2014 Winter Olympics. Each X above a number represents a country that won that many gold medals.

2014 Winter Olympics Gold Medal Wins by
Medal-Winning Countries



Number of Gold Medals Won

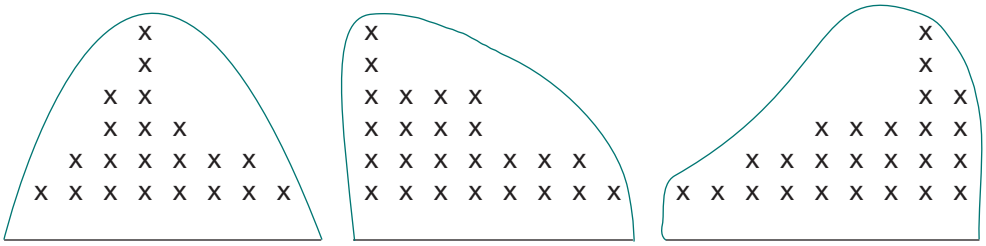
When you analyze a graphical representation of numeric data, you can look at its shape, center, and spread to draw conclusions.

The overall shape of a graph is called the distribution of data. A **distribution** is the way in which the data are spread out. The shape of the distribution can reveal a lot of information about data. There are many different distributions, but the most common are **symmetric**, **skewed right**, and **skewed left**.

When analyzing a graphical display of data, you can also look for any interesting patterns. Some of these patterns include:

- **clusters**—areas where data are grouped close together
- **gaps**—areas where there are no data
- **peaks**—values that contain more data points than the values on either side of it
- **outliers**—data values that lie a large distance from the other data. Outliers usually accompany gaps in data.

Shapes of Typical Distribution of Graphical Displays of Data



symmetric

- The left and right halves of the graph are mirror images of each other.
- The peak is in the middle, because there are many data values in the center.

skewed right

- The peak of the data is to the left side of the graph.
- There are only a few data points to the right side of the graph.

skewed left

- The peak of the data is to the right side of the graph.
- There are only a few data points to the left side of the graph.

Total Medals Won by Countries
2016 Summer Olympics

0	5 5 5 6 6 6 7 7 7 8 8 8 8 8 9 9
1	0 0 0 1 1 1 1 1 3 3 5 5 7 8 8 8 9 9
2	2 8 9
3	
4	1 2 2
5	6
6	7
7	0
8	
9	
10	
11	
12	1

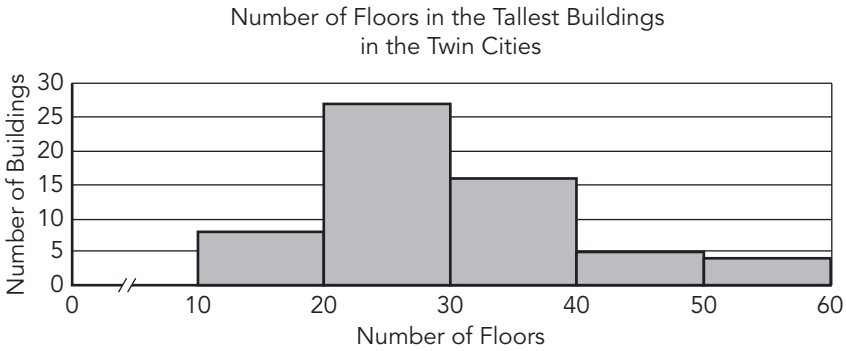
Key: 4|1 = 41 medals won.

A **stem-and-leaf plot** is a graphical method used to represent ordered numerical data sets with a larger range of data values. Once the data is ordered, the stem and leaves are determined. Typically, the stem is all the digits in a number except the rightmost digit, which is the leaf.

LESSON 3

Skyscrapers

A **histogram** is a graphical way to display quantitative or numerical data using vertical bars. The numerical data are represented continuously with intervals. The intervals in a histogram must all be the same size. The width of a bar in a histogram represents the interval. The height of the bar indicates the frequency, or the number of data values, in the interval.



Dot plots show individual data values. Histograms display grouped data. For example, you cannot determine from the histogram how many buildings have 21 floors, or more than 45 floors.

Floor Intervals	Frequency (*f*)
10–20	8
20–30	27
30–40	16
40–50	5
50–60	4

To create a histogram, data is usually organized into a grouped frequency table. A **grouped frequency table** is a table used to organize data according to how many times data values within a given range of values occur.

For example, this grouped frequency table displays the data represented by the histogram above.