pits, 4; parts assembled out of sequence, 6; parts under-trimmed, 21; missing holes/slots, 8; parts not lubricated, 5; parts out of contour, 30; and parts not deburred, 3. Construct and interpret a Pareto chart.

6-61. Construct a frequency distribution and histogram for the bridge condition data in Exercise 6-20.

6-62. Construct a frequency distribution and histogram for the acid rain measurements in Exercise 6-21.

6-63. Construct a frequency distribution and histogram for the combined cloud-seeding rain measurements in Exercise 6-22.

6-64. Construct a frequency distribution and histogram for the swim time measurements in Exercise 6-24.

6-4 Box Plots

The stem-and-leaf display and the histogram provide general visual impressions about a data set, but numerical quantities such as $\bar{x}$ or s provide information about only one feature of the data. The **box plot** is a graphical display that simultaneously describes several important features of a data set, such as center, spread, departure from symmetry, and identification of unusual observations or outliers.

A box plot, sometimes called *box-and-whisker plots*, displays the three quartiles, the minimum, and the maximum of the data on a rectangular box, aligned either horizontally or vertically. The box encloses the interquartile range with the left (or lower) edge at the first quartile, q_1, and the right (or upper) edge at the third quartile, q_3. A line is drawn through the box at the second quartile (which is the 50th percentile or the median), $q_2 = \tilde{x}$. A line, or **whisker**, extends from each end of the box. The lower whisker is a line from the first quartile to the smallest data point within 1.5 interquartile ranges from the first quartile. The upper whisker is a line from the third quartile to the largest data point within 1.5 interquartile ranges from the third quartile. Data farther from the box than the whiskers are plotted as individual points. A point beyond a whisker, but less than three interquartile ranges from the box edge, is called an **outlier**. A point more than three interquartile ranges from the box edge is called an **extreme outlier**. See Fig. 6-13. Occasionally, different symbols, such as open and filled circles, are used to identify the two types of outliers.

Figure 6-14 presents a typical computer-generated box plot for the alloy compressive strength data shown in Table 6-2. This box plot indicates that the distribution of compressive strengths is fairly symmetric around the central value because the left and right whiskers and the lengths of the left and right boxes around the median are about the same. There are also two mild outliers at lower strength and one at higher strength. The upper whisker extends to observation 237 because it is the highest observation below the limit for upper outliers. This limit is $q_3 + 1.5\text{IQR} = 181 + 1.5(181 - 143.5) = 237.25$. The lower whisker extends to observation 97 because it is the smallest observation above the limit for lower outliers. This limit is $q_1 - 1.5\text{IQR} = 143.5 - 1.5(181 - 143.5) = 87.25$.

Box plots are very useful in graphical comparisons among data sets because they have high visual impact and are easy to understand. For example, Fig. 6-15 shows the comparative box plots for a manufacturing quality index on semiconductor devices at three manufacturing plants. Inspection of this display reveals that there is too much variability at plant 2 and that plants 2 and 3 need to raise their quality index performance.

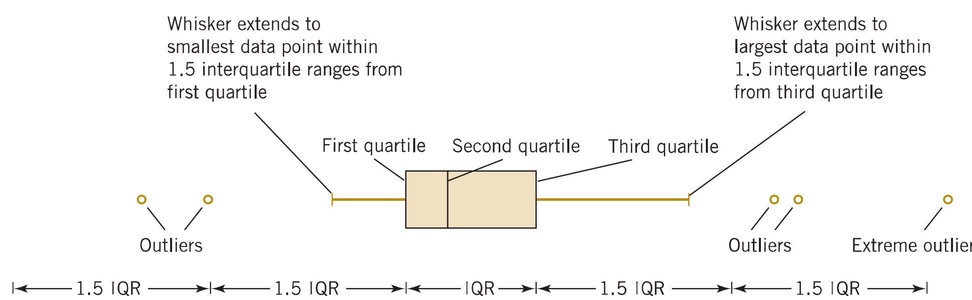

FIGURE 6-13 Description of a box plot.

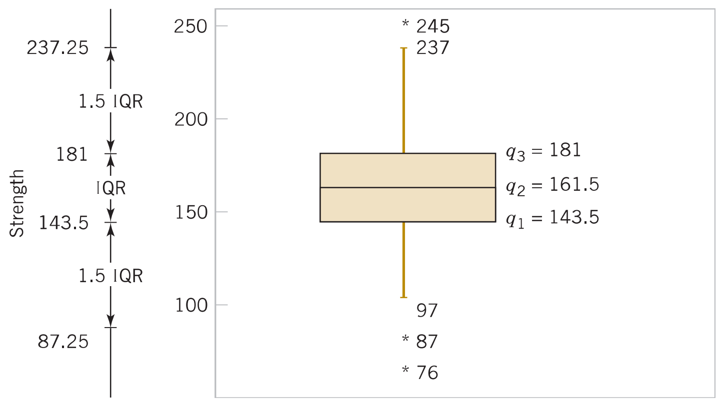

FIGURE 6-14 Box plot for compressive strength data in Table 6-2.

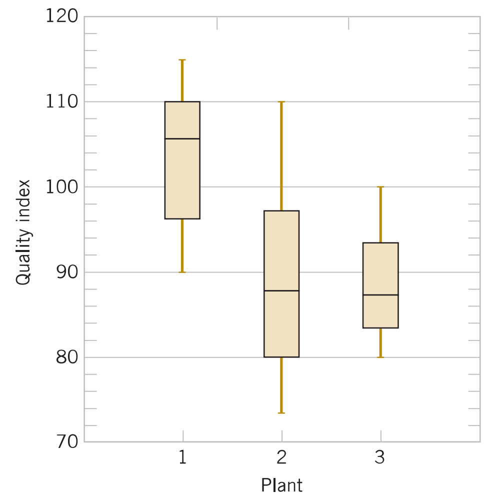

FIGURE 6-15 Comparative box plots of a quality index at three plants.

Exercises FOR SECTION 6-4

⊕ Problem available in *WileyPLUS* at instructor's discretion.

⊕ **Go Tutorial** Tutoring problem available in *WileyPLUS* at instructor's discretion.

6-65. Using the data on bridge conditions from Exercise 6-20,
(a) Find the quartiles and median of the data.
(b) Draw a box plot for the data.
(c) Should any points be considered potential outliers? Compare this to your answer in Exercise 6-20. Explain.

6-66. Using the data on acid rain from Exercise 6-21,
(a) Find the quartiles and median of the data.
(b) Draw a box plot for the data.
(c) Should any points be considered potential outliers? Compare this to your answer in Exercise 6-21. Explain.

6-67. Using the data from Exercise 6-22 on cloud seeding,
(a) Find the median and quartiles for the unseeded cloud data.
(b) Find the median and quartiles for the seeded cloud data.
(c) Make two side-by-side box plots, one for each group on the same plot.
(d) Compare the distributions from what you can see in the side-by-side box plots.

6-68. Using the data from Exercise 6-24 on swim times,
(a) Find the median and quartiles for the data.
(b) Make a box plot of the data.
(c) Repeat (a) and (b) for the data without the extreme outlier and comment.
(d) Compare the distribution of the data with and without the extreme outlier.

6-69. ⊕ **Go Tutorial** The "cold start ignition time" of an automobile engine is being investigated by a gasoline manufacturer. The following times (in seconds) were obtained for a test vehicle: 1.75, 1.92, 2.62, 2.35, 3.09, 3.15, 2.53, 1.91.
(a) Calculate the sample mean, sample variance, and sample standard deviation.
(b) Construct a box plot of the data.

6-70. An article in *Transactions of the Institution of Chemical Engineers* (1956, Vol. 34, pp. 280–293) reported data from an experiment investigating the effect of several process variables on the vapor phase oxidation of naphthalene. A sample of the percentage mole conversion of naphthalene to maleic anhydride follows: 4.2, 4.7, 4.7, 5.0, 3.8, 3.6, 3.0, 5.1, 3.1, 3.8, 4.8, 4.0, 5.2, 4.3, 2.8, 2.0, 2.8, 3.3, 4.8, 5.0.
(a) Calculate the sample mean, sample variance, and sample standard deviation.
(b) Construct a box plot of the data.

6-71. ⊕ The nine measurements that follow are furnace temperatures recorded on successive batches in a semiconductor manufacturing process (units are °F): 953, 950, 948, 955, 951, 949, 957, 954, 955.
(a) Calculate the sample mean, sample variance, and standard deviation.
(b) Find the median. How much could the highest temperature measurement increase without changing the median value?
(c) Construct a box plot of the data.

6-72. Exercise 6-18 presents drag coefficients for the NASA 0012 airfoil. You were asked to calculate the sample mean, sample variance, and sample standard deviation of those coefficients.
(a) Find the median and the upper and lower quartiles of the drag coefficients.
(b) Construct a box plot of the data.
(c) Set aside the highest observation (100) and rework parts (a) and (b). Comment on your findings.

6-73. Exercise 6-19 presented the joint temperatures of the O-rings (°F) for each test firing or actual launch of the space shuttle rocket motor. In that exercise, you were asked to find the sample mean and sample standard deviation of temperature.