often great for comparing two or more data sets. When using two or more boxplots for comparing different data sets, graph the boxplots on the same scale so that comparisons can be easily made.

Example 9 Do the Different Brands of Cookies Have Different Chocolate Chip Counts?

The Chapter Problem includes Table 3-1, which lists counts of chocolate chips in cookies from different brands. Use the same scale to construct the five corresponding boxplots; then compare the results.

Solution

The STATDISK-generated boxplots shown in Figure 3-7 suggest that the numbers of chocolate chips in the different brands of cookies are very different. In particular, the counts from the Keebler and Hannaford cookies appear to be very different; the boxplots show that there isn't any overlap, and all of the Hannaford cookies have lower counts than any of the Keebler cookies. It might seem that the Hannaford brand is stingy with its chocolate chips, but the Hannaford brand had many chips that were substantially larger than those in any of the other brands.

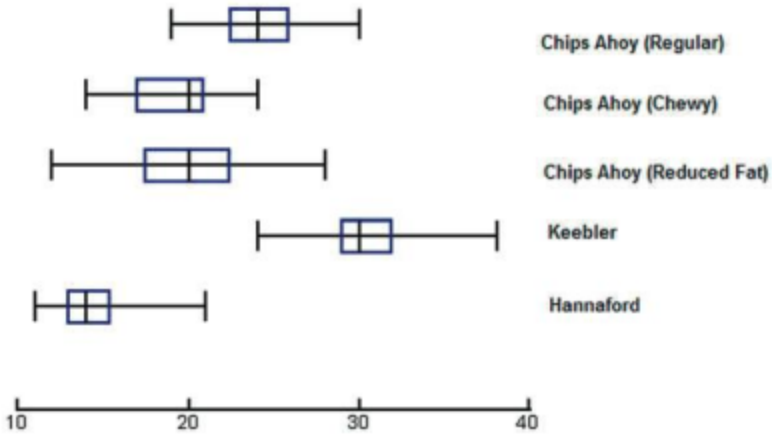

Figure 3-7 Boxplots of Counts of Chocolate Chips in Cookies

Methods discussed later in this book allow us to analyze this issue more formally. It is always wise to construct suitable graphs, such as histograms, dotplots, and boxplots, but we should not rely solely on subjective judgments based on graphs.

Outliers

When analyzing data, it is important to identify and consider outliers because they can strongly affect values of some important statistics (such as the mean and standard deviation), and they can also strongly affect important methods discussed later in this book. In Section 2-1 we described outliers as sample values that lie very far away from the vast majority of the other values in a set of data, but that description is vague and it does not provide specific objective criteria. Part 2 of this section includes a description of *modified boxplots* along with a specific definition of outliers used in the context of creating modified boxplots.

CAUTION When analyzing data, always identify outliers and consider their effects, which can be substantial.