Part 2: Outliers and Modified Boxplots

Outliers

We noted that the description of outliers is somewhat vague, but for the purposes of constructing *modified boxplots*, we can consider outliers to be data values meeting specific criteria based on quartiles and the interquartile range. (The interquartile range is often denoted by IQR, and IQR $= Q_3 - Q_1$.)

1. Find the quartiles Q_1, Q_2, and Q_3.

2. Find the interquartile range (IQR), where IQR $= Q_3 - Q_1$.

3. Evaluate $1.5 \times$ IQR.

4. **In a modified boxplot, a data value is an *outlier* if it is**

    above Q_3, by an amount greater than $1.5 \times$ IQR

    or   below Q_1, by an amount greater than $1.5 \times$ IQR

Modified Boxplots

The boxplots described earlier are called **skeletal** (or **regular**) **boxplots,** but some statistical software packages provide modified boxplots, which represent outliers as special points. A **modified boxplot** is a regular boxplot constructed with these modifications: (1) A special symbol (such as an asterisk or point) is used to identify outliers as defined above, and (2) the solid horizontal line extends only as far as the minimum data value that is not an outlier and the maximum data value that is not an outlier. (*Note: Exercises involving modified boxplots are found in the "Beyond the Basics" exercises only.*)

Example 10   Modified Boxplot

Use the Hannaford chocolate chip counts from Table 3-1 to construct a modified boxplot.

Solution

Let's begin with the above four steps for identifying outliers in a modified boxplot.

1. Using the Hannaford chocolate chip counts from Table 3-1, the three quartiles are $Q_1 = 13.0$, $Q_2 = 14.0$, and $Q_3 = 15.5$.

2. The interquartile range is IQR $= Q_3 - Q_1 = 15.5 - 13.0 = 2.5$.

3. $1.5 \times$ IQR $= 1.5 \times 2.5 = 3.75$.

4. Any outliers are above $Q_3 = 15.5$ by more than 3.75, or below $Q_1 = 13.0$ by more than 3.75. This means that any outliers are greater than 19.25 or less than 9.25.

We can now examine the original Hannaford chocolate chip counts to see that 21 is the only value greater than 19.25, and there are no values less than 9.25. Therefore, 21 is the only outlier.

We can now construct the modified boxplot shown in Figure 3-8. In Figure 3-8, the outlier is identified as a special point, the three quartiles are shown as in a regular
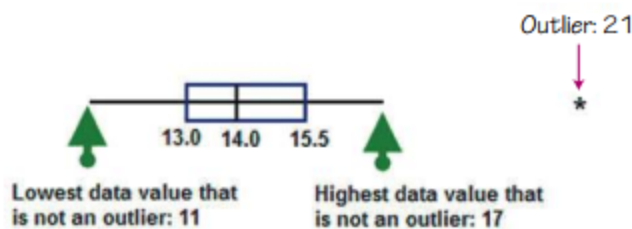

Outlier: 21

13.0 14.0 15.5

Lowest data value that is not an outlier: 11

Highest data value that is not an outlier: 17

Figure 3-8
Modified Boxplots of Hannaford Chocolate Chip Counts