Graphs That Enlighten

Scatterplots

A **scatterplot** (or **scatter diagram**) is a plot of paired (x, y) quantitative data with a horizontal x-axis and a vertical y-axis. The horizontal axis is used for the first (x) variable, and the vertical axis is used for the second variable. The pattern of the plotted points is often helpful in determining whether there is a correlation (or relationship) between the two variables. (This issue is discussed at length when the topic of correlation is considered in Section 10-2.)

Example 1 **Correlation: Waist and Arm Circumference**

Data Set 1 in Appendix B includes the waist circumferences (cm) and arm circumferences (cm) of randomly selected males. Figure 2-6 is a scatterplot of the paired waist/arm measurements. The points show a pattern of increasing values from left to right. This pattern suggests that there is a correlation, or relationship, between waist circumference and arm circumference in males.

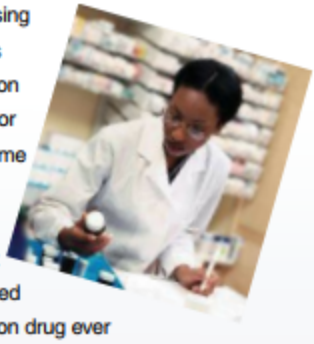
Example 2 **No Correlation: Weight and Pulse Rate**

Data Set 1 in Appendix B includes weights (kg) and pulse rates (beats per minute) of randomly selected males. Figure 2-7 is a scatterplot of the paired weight/pulse rate measurements. The points in Figure 2-7 do not show any obvious pattern, and this lack of a pattern suggests that there is no correlation, or relationship, between the weight and pulse rate of males.

Correlation Coefficient Examples 1 and 2 involve making decisions about a correlation based on subjective judgments of scatterplots, but Chapter 10 introduces more objective methods. Those methods involve calculating a value of a *linear correlation coefficient r*, which is a value between -1 and 1. If r is close to -1 or close to 1, there appears to be a correlation, but if r is close to 0, there does not appear to be a correlation. For the data depicted in the scatterplot of Figure 2-6, $r = 0.788$, and the data in the scatterplot of Figure 2-7 result in $r = -0.213$. Section 10-2 describes the calculation and interpretation of a value of r, so that a decision about correlation is much more objective. Even though the methods based on calculations of r are much more objective than the subjective interpretation of a scatterplot, it is always wise to construct a scatterplot first, so that we can see characteristics that cannot be seen by examining the list of paired data values.

The Power of a Graph

With annual sales around $13 billion and with roughly 50 million people using it, Pfizer's prescription drug Lipitor has become the most profitable and most widely used prescription drug ever marketed. In the early stages of its development, Lipitor was compared to other drugs (Zocor, Mevacor, Lescol, and Pravachol) in a process that involved controlled trials. The summary report included a graph showing a Lipitor curve that had a steeper rise than the curves for the other drugs, visually showing that Lipitor was more effective in reducing cholesterol than the other drugs. Pat Kelly, who was then a senior marketing executive for Pfizer, said "I will never forget seeing that chart. . . . It was like 'Aha!' Now I know what this is about. We can communicate this!" The Food and Drug Administration approved Lipitor and allowed Pfizer to include the graph with each prescription. Pfizer sales personnel also distributed the graph to physicians.



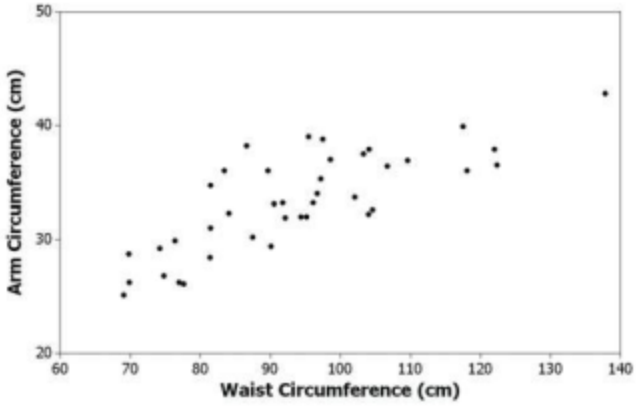


Figure 2-6 **Waist Circumference and Arm Circumference in Males**

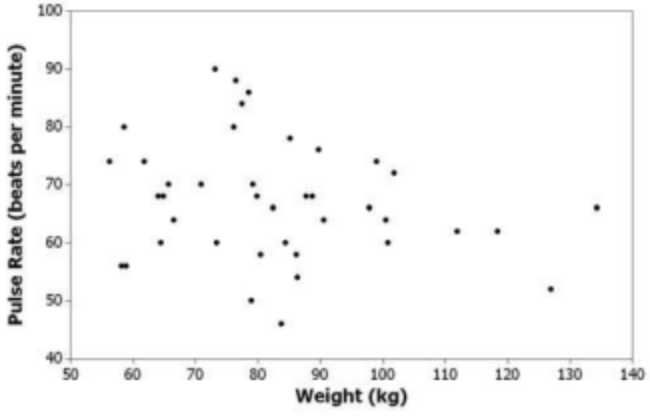


Figure 2-7 **Weight and Pulse Rate in Males**