2. Visual examination of the scatterplot must confirm that the points approximate a straight-line pattern.
3. Because results can be strongly affected by the presence of outliers, any outliers must be removed if they are known to be errors. The effects of any other outliers should be considered by calculating $r$ with and without the outliers included.

*Note:* Requirements 2 and 3 above are simplified attempts at checking this formal requirement: The pairs of $(x, y)$ data must have a **bivariate normal distribution.** Normal distributions are discussed in Chapter 6, but this assumption basically requires that for any fixed value of $x$, the corresponding values of $y$ have a distribution that is approximately normal, and for any fixed value of $y$, the values of $x$ have a distribution that is approximately normal. This requirement is usually difficult to check, so for now, we will use Requirements 2 and 3 as listed above.

Formulas for Calculating $r$

**Formula 10-1**

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$$

Formula 10-1 simplifies manual calculations, but $r$ is usually calculated with computer software or a calculator. Formula 10-2 is better for understanding the reasoning behind $r$; see the rationale discussed later in this section.

**Formula 10-2**

$$r = \frac{\Sigma(z_x z_y)}{n - 1}$$

where $z_x$ denotes the $z$ score for an individual sample value $x$ and $z_y$ is the $z$ score for the corresponding sample value $y$.

Interpreting the Linear Correlation Coefficient $r$

- ***Using Computer Software to Interpret r:*** If the $P$-value computed from $r$ is less than or equal to the significance level, conclude that there is sufficient evidence to support a claim of a linear correlation. Otherwise, there is not sufficient evidence to support a claim of a linear correlation.
- ***Using Table A-6 to Interpret r:*** Consider critical values from Table A-6 as being both positive and negative, and draw a graph similar to Figure 10-3 that accompanies Example 4.

  ***Correlation*** If the computed linear correlation coefficient $r$ lies in the left tail beyond the leftmost critical value or if it lies in the right tail beyond the rightmost critical value, conclude that there is sufficient evidence to support the claim of a linear correlation.

  ***No Correlation*** If the computed linear correlation coefficient lies *between* the two critical values, conclude that there is not sufficient evidence to support the claim of a linear correlation.

(Here are equivalent criteria: If the absolute value of $r$, denoted $|r|$, exceeds the value in Table A-6, conclude that there is a linear correlation. Otherwise, there is not sufficient evidence to support the conclusion of a linear correlation.)

**CAUTION** Know that the methods of this section apply to a *linear* correlation. If you conclude that there does not appear to be linear correlation, it is possible that there might be some other association that is not linear, as in Figure 10-2(d).

## Rounding the Linear Correlation Coefficient $r$

Round the linear correlation coefficient $r$ to three decimal places (so that its value can be directly compared to critical values in Table A-6 ). If manually calculating $r$ and other statistics in this chapter, rounding in the middle of a calculation often creates substantial errors, so try to round only the final result.