Example 6 Explained Variation

In Example 5 we noted that for the 40 pairs of shoe print lengths and heights listed in Data Set 2 from Appendix B, the linear correlation coefficient is $r = 0.813$. What proportion of the variation in height can be explained by the variation in the length of shoe print?

Solution

With $r = 0.813$, we get $r^2 = 0.661$.

Interpretation

We conclude that 0.661 (or about 66%) of the variation in height can be explained by the linear relationship between lengths of shoe prints and heights. This implies that about 34% of the variation in heights cannot be explained by lengths of shoe prints.

Common Errors Involving Correlation

Here are three of the most common errors made in interpreting results involving correlation:

1. *A common error is to assume that correlation implies causality.* One classic example involves paired data consisting of the stork population in Copenhagen and the number of human births. For several years, the data suggested a linear correlation. *Bulletin:* Storks do not actually cause births, and births do not cause storks. Both variables were affected by some other variable lurking in the background. (A **lurking variable** is one that affects the variables being studied but is not included in the study.)
2. *Another error arises with data based on averages.* Averages suppress individual variation and may inflate the correlation coefficient. One study produced a 0.4 linear correlation coefficient for paired data relating income and education among individuals, but the linear correlation coefficient became 0.7 when regional averages were used.
3. *A third error involves the property of linearity.* If there is no linear correlation, there might be some other correlation that is not linear, as in Figure 10-2(d). (Figure 10-2(d) is a scatterplot that depicts the relationship between distance above ground and time elapsed for an object thrown upward.)

CAUTION Know that *correlation does not imply causality.*

Part 2: Formal Hypothesis Test (Requires Coverage of Chapter 8)

Here in Part 2 we describe formal procedures for conducting hypothesis tests to determine whether there is a significant linear correlation between two variables. The following box contains key elements of the hypothesis test.