XLSTAT

Correlation matrix (Pearson):

Variables	Shoe Print	Height
Shoe Print	1	**0.8129**
Height	**0.8129**	1

Values in bold are different from 0 with a significance level alpha=0.05

p-values:

Variables	Shoe Print	Height
Shoe Print	0	**< 0.0001**
Height	**< 0.0001**	0

Values in bold are different from 0 with a significance level alpha=0.05

Solution

Requirement check (1) The sample is a simple random sample of quantitative data. (2) The points in the scatterplot of Figure 10-1(b) roughly approximate a straight-line pattern. (3) There are no outliers that are far away from almost all of the other pairs of data. ✓

Using Software: The P-value obtained from XLSTAT is less than 0.0001. Because the P-value is less than or equal to 0.05, we conclude that there is sufficient evidence to support a claim of a linear correlation between the lengths of shoe prints and heights.

Using Table A-6: If we refer to Table A-6 with $n = 40$ pairs of sample data, we obtain the critical values of -0.312 and 0.312 for $\alpha = 0.05$. Because the computed value of $r = 0.813$ *does* exceed the critical value of 0.312 from Table A-6, the computed value of r lies in the right tail beyond the rightmost critical value of 0.312, so we conclude that there is sufficient evidence to support a claim of a linear correlation between the lengths of shoe prints and heights.

Example 4 used only five pairs of data and we concluded that there is *not* sufficient evidence to support the conclusion that there is a linear correlation between shoe print lengths and heights of people, but this example uses 40 pairs of data and here we conclude that there *is* sufficient evidence to support the conclusion that there is a linear correlation between shoe print lengths and heights of people. This larger data set provided the additional evidence that enabled us to support the presence of a linear correlation. Such is the power of larger data sets.

Interpreting r: Explained Variation

If we conclude that there is a linear correlation between x and y, we can find a linear equation that expresses y in terms of x, and that equation can be used to predict values of y for given values of x. In Section 10-3 we will describe a procedure for finding such equations and show how to predict values of y when given values of x. But a predicted value of y will not necessarily be the exact result that occurs because in addition to x, there are other factors affecting y, such as random variation and other characteristics not included in the study. In Section 10-4 we will present a rationale and more details about this principle:

The value of r^2 is the proportion of the variation in y that is explained by the linear relationship between x and y.