Investigation of the relationship between two sets of scores proceeds through the plotting of scatter diagrams and through calculation of the correlation coefficient.

Scatter Diagrams

Assume two sets of scores for a group of people—scores on a test and scores on a measure of job performance. A scatter diagram is simply the plot of the joint distribution of the two sets of scores. Inspection of the plot provides a visual representation of the type of relationship that exists between the two sets of scores. Exhibit 7.3 provides three different scatter diagrams for the two sets of scores. Each X represents a test score (predictor) and job performance (criterion, or outcome) score combination for an individual.

Example A in Exhibit 7.3 suggests very little relationship between the two sets of scores. Example B shows a modest relationship between the scores, and example C shows a somewhat strong relationship between the two sets of scores.

Correlation Coefficient

The relationship between two sets of scores may also be investigated through calculation of the correlation coefficient. The symbol for the correlation coefficient is r. Numerically, r values can range from $r = -1.0$ to $r = 1.0$. An r value of 0 indicates there is no relationship between the two sets of scores. The larger the absolute value of r, the stronger the relationship. So, $r = -.55$ is stronger than $r = .30$. When an r value is shown without a sign (plus or minus), the value is assumed to be positive.

The sign of the correlation coefficient is arbitrary. One could say that job satisfaction is positively related to job performance (a positive r), just as one could say that job *dis*satisfaction is negatively related to job performance (a negative r). Thus, whether r is positive or negative simply depends on how the relationship is worded (and how the measures are scored).

Naturally, the value of r bears a close resemblance to the scatter diagram. To demonstrate this, Exhibit 7.3 also shows the approximate r value for each of the three scatter diagrams. The r in example A is low ($r = .10$), the r in example B is moderate ($r = .25$), and the r in example C is high ($r = .60$).

Calculation of the correlation coefficient is straightforward, even by hand. An example of this calculation and the formula for r are shown in Exhibit 7.4. In the exhibit are two sets of scores for 20 people. The first set is the test scores for the 20 individuals in Exhibit 7.2. The second set of scores is an overall job performance rating (on a 1–5 rating scale) for these people. As can be seen from the calculation, there is a correlation of $r = .58$ between the two sets of scores. The resultant value of r succinctly summarizes both the strength of the relationship between the two sets of scores and the direction of the relationship.

While the formula for the correlation is good to know, correlations are easily calculated by a variety of computer programs, including Microsoft Excel. There, scores on each variable are placed in adjacent columns, and Excel's correlate function "(=Correl)" calculates r.