11.5 CORRELATION ANALYSIS

In many instances, when people talk about regression analysis they also talk about *correlation analysis.* The two topics are related, but not interchangeable. Both regression and correlation analyses deal with bivariate quantitative data and the relationship between the two variables. The main purpose of regression analysis is to find an equation or model that allows the decision maker to predict the value of the dependent variable. On the other hand, *correlation analysis* simply *measures the strength* of the linear relationship between two quantitative variables. The output of the analysis is a single number. In correlation analysis, there is no need to identify which variable is dependent and which is independent, since prediction is not the end result.

11.5.1 The Correlation Coefficient

We have talked about the types of linear relationships that can exist between two variables. In Figure 11.14 you see three types of relationships: perfect negative, none, and perfect positive.

If we simply want to measure the strength of a relationship between two variables, we use the **correlation coefficient.**

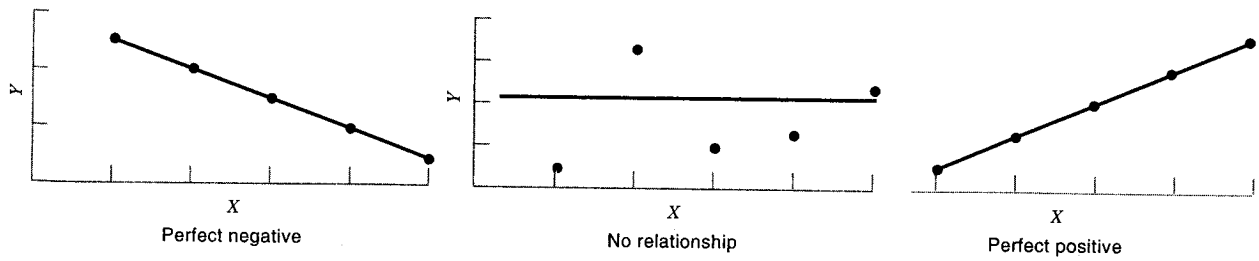

FIGURE 11.14 Three types of relationships: perfect negative, no relationship, and perfect positive

EXAMPLE 11.19 HMO Health

Calculating the Correlation Coefficient

The analysts looking at the relationship between revenues and number of members for various HMOs are pretty certain that the relationship between the two variables will be a strong, positive one, so they decide to calculate the correlation coefficient to check their assumption. The numbers that they need to perform the calculation are

$$n = 10 \quad \Sigma x = 21.09 \quad \Sigma y = 31.97 \quad \Sigma xy = 77.6371$$
$$\Sigma x^2 = 53.4063 \quad \Sigma y^2 = 117.3013$$

The calculation is

$$\frac{77.6371 - (21.09)(31.97)/10}{\sqrt{53.4063 - (21.09)^2/10}\sqrt{117.3013 - (31.97)^2/10}} = \frac{10.2124}{11.60795\ldots} = 0.8798$$

Looking at the value, the analysts feel that the number does indicate the strong positive correlation that they expected to find.

Increasing Capacity ***Calculating the Correlation Coefficient***

The relevant data to calculate the correlation coefficient for the oil company problem are

$$n = 10 \quad \Sigma x = 78 \quad \Sigma y = 463.64 \quad \Sigma xy = 4121.86 \quad \Sigma x^2 = 714$$
$$\Sigma y^2 = 25{,}359.3224$$

Find the correlation coefficient for the data.

The correlation coefficient is also related to one of the quantities that we looked at in regression analysis, the coefficient of determination, R^2. The value of r is equal to the square root of R^2. The sign of r is the same as the sign of the slope coefficient.

$$r = \pm\sqrt{R^2}$$