corresponding y-values also increase. The scatterplot in Figure 10-2(b) shows a distinct linear pattern. We say that there is a *negative* linear correlation between x and y, since as the x-values increase, the corresponding y-values decrease. The scatterplot in Figure 10-2(c) shows no distinct pattern and suggests that there is no linear correlation between x and y. The scatterplot in Figure 10-2(d) shows a distinct pattern suggesting a correlation between x and y, but the pattern is not that of a straight line.

Measure the Strength of the Linear Correlation

Because conclusions based on visual examinations of scatterplots are largely subjective, we need more objective measures. We use the linear correlation coefficient r, which is a number that measures the strength of the (linear) association between the two variables.

DEFINITION The **linear correlation coefficient *r*** measures the strength of the linear correlation between the paired quantitative x- and y-values in a *sample*. (Its value is computed by using Formula 10-1 or Formula 10-2, included in the following box. [The linear correlation coefficient is sometimes referred to as the **Pearson product moment correlation coefficient** in honor of Karl Pearson (1857–1936), who originally developed it.]

Because the linear correlation coefficient r is calculated using sample data, it is a sample statistic used to measure the strength of the linear correlation between x and y. If we had every pair of x and y values from an entire population, the result of Formula 10-1 or Formula 10-2 would be a population parameter, represented by ρ (Greek letter rho).

Objective

Determine whether there is a linear correlation between two variables.

Notation for the Linear Correlation Coefficient

n	number of pairs of sample data.
Σ	denotes addition of the items indicated.
Σx	sum of all x-values.
Σx^2	indicates that each x-value should be squared and then those squares added.
$(\Sigma x)^2$	indicates that the x-values should be added and the total then squared. Avoid confusing Σx^2 and $(\Sigma x)^2$.
Σxy	indicates that each x-value should first be multiplied by its corresponding y-value. After obtaining all such products, find their sum.
r	linear correlation coefficient for *sample* data.
ρ	linear correlation coefficient for a *population* of paired data.

Requirements

Given any collection of sample paired quantitative data, the linear correlation coefficient r can always be computed, but the following requirements should be satisfied when using the sample data to make a conclusion about correlation in the population.

1. The sample of paired (x, y) data is a simple random sample of quantitative data. (It is important that the sample data have not been collected using some inappropriate method, such as using a voluntary response sample.)