variation, the correlation will be $r = .00$ (technically, it is undefined). That is, for there to be a correlation, there must be variation in both sets of scores. The lack of variation in scores is called the problem of restriction of range.

Third, the formula used to calculate the correlation in Exhibit 7.4 is based on the assumption that there is a linear relationship between the two sets of scores. This may not always be a good assumption; something other than a straight line may best capture the true nature of the relationship between scores. For example, a new training program may initially be associated with a rapid increase in trainee performance, but repeated exposure to that program may begin to provide diminished returns. To the extent that two sets of scores are not related in a linear fashion, use of the formula for calculating the correlation will yield a value of r that understates the actual strength of the relationship.

Finally, the correlation between two variables does not imply causation between them. A correlation simply states how two variables covary or relate to each other; it says nothing about one variable necessarily causing the other one.

Significance of the Correlation Coefficient

The statistical significance refers to the likelihood that a correlation exists in a population, based on knowledge of the actual value of r in a sample from that population. Concluding that a correlation is indeed statistically significant means that there is most likely a correlation in the population. That means if the organization were to use a selection measure based on a statistically significant correlation, the correlation is likely to be significant when used again to select another sample (e.g., future job applicants).

More formally, r is calculated in an initial group, called a sample. From this piece of information, the question arises whether to infer that there is also a correlation in the *population*. For example, one might want to know whether a correlation calculated in a sample of 200 information technology analysts would also be found in the population (all information technology analysts). To answer this, compute the t value of our correlation using the following formula,

$$t = \frac{r}{\sqrt{(1-r^2)/n-2}}$$

where r is the value of the correlation, and n is the size of the sample.

A t distribution table, which can be found in any elementary statistics book as well as online, shows the significance level of r.[7] The significance level is expressed as $p <$ some value, for example, $p < .05$. This p level tells the probability of concluding that there is a correlation in the population when in fact there is not a relationship. Thus, a correlation with $p < .05$ means there are fewer than 5 chances in 100 of concluding that there is a relationship in the population when in fact there is not. This is a relatively small probability and usually leads to the conclusion that a correlation is indeed statistically significant.