

Source: *Statistics In Plain English*, Harvey J. Brightman, South-Western Publishing Company Cincinnati, Ohio, 1986, pages 265-6.

ASSOCIATION AND CAUSE AND EFFECT

Regression analysis looks for relationships or associations. What does an association mean? Is association the same as cause and effect? Earlier we found a strong relationship between training hours and monthly sales. Can we say that increasing training causes increased monthly sales? Before answering that question, consider the following study. We sample cities throughout the United States and record the number of clergy and the number of alcoholics. We find a strong positive association between the two variables. Can we conclude that the clergy cause people to drink?

A **causal relationship** exists when changes in one variable cause changes in another variable, *holding all other factors constant*. The case for causation is strengthened when:

1. The association between the two variables is found in many studies done by different investigators.
2. You can provide a reasonable argument as to how changes in one variable could cause changes in the other.
3. There are no other plausible factors that explain the connection between the two variables.

In the sales training study, you can make a reasonable argument that additional hours of training was a cause of the increased sales. After all, the purpose of training is to improve product knowledge which should produce more sales. Are there any other plausible factors that might explain the association between training and sales? For example, if stores that received more training had more senior sales forces or carried higher price merchandise, then these two factors might be reasonable alternative explanations to the association. If these could be ruled out, then the argument that more training hours causes more sales is strengthened.

In the clergy example there are no reasonable explanations to the association. There is also an obvious third variable that affects both the number of clergy and the number of alcoholics-population of the city. As cities increase in population, you have more clergy and more alcoholics, but one does not cause the other.

For many years we knew that there was a strong association between amount of smoking and lung cancer. Can we say that smoking is a cause of lung cancer? First, this association has been found by many investigators throughout the world. Second, investigators, in well-controlled experimental studies, have found that substances in cigarettes can produce cancers in animals. So there is a plausible explanation. Finally, the differences in life styles between smokers and nonsmokers (other than smoking) have not been found to be related to the incidence of lung cancer. It appears that smoking is a (not *the*) cause of lung cancer.

Regression focuses on detecting associations. Assessing causation is beyond statistics. Causation requires association and explanation.