Some thoughts about causality vs. association
Taken from the Little Handbook of Statistics by Gerard E Dallal, Ph.D. Tufts University Chief, Biostatistics Unit, Jean Mayer USDA Human Nutrition Research Center on Aging at Tufts University

"Cause and Effect"! You almost never hear these words in an introductory statistics course. The subject is commonly ignored. Even on this site, all it gets is this one web page. If cause and effect is addressed at all, it is usually by giving the (proper) warning "Association does not imply causation!" along with a few illustrations. For example, in the early part of the twentieth century, it was noticed that when viewed over time, the number of crimes increased with membership in the Church of England. This had nothing to do with criminals finding religion. Rather, both crimes and Church membership increased as the population increased. Association does not imply causation! During WWII it was noticed that bombers were less accurate when the weather was more clear. The reason was that when the weather was clear there was also more opposition from enemy fighter planes. Association does not imply causation, at least not necessarily in the way it appears on the surface!

We laugh at obvious mistakes but often forget how easy it is to make subtle errors any time an attempt is made to use statistics to prove causality. This could have disastrous consequences if the errors form the basis of public policy.

"Cause & effect" must be among the first things that are addressed because this is what most people will use statistics for! Newspapers, radio, television, and the Internet are filled with claims based on some form of statistical analysis. Calcium is good for strong bones. Watching TV is a major cause of childhood and adolescent obesity. Food stamps and WIC improve nutritional status. Coffee consumption is responsible for heavens knows what! All because someone got hold of a dataset from somewhere and looked for associations. Which claims should be believed? Only by understanding what it takes to establish causality do we have any chance of being intelligent consumers of the "truths" the world throws at us.

The rules for claiming causality vary from field to field. The physical sciences seem to have the easiest time of it because it is easy to design experiments in which a single component can be isolated and studied. Fields like history have the hardest time of it. Not only are experiments all but impossible, but observations often play out over generations, making it difficult to collect new data, while much of the existing data is often suspect.

Many outrageous claims can be made because people often do not have the proper foundations in logic (as well as in the subject matter) for making defensible
claims of causality. Two examples that have been offered, (1) Two countries that have McDonalds restaurants have never gone to war. [except for the England and Venezuela!] (2) Before television, two World Wars; after television, no World Wars.

In epidemiology, which relies heavily on observational studies (that is, taking people as you find them), cause and effect is established by observing the same thing in a wide variety of settings until all but the suspected cause can be ruled out.

A modern example is the link between smoking and lung cancer. Because is it impossible to conduct randomized smoking experiments in human populations, it took many decades to collect enough observational data (some free of one types of bias, others free of another) to establish the connection. Much of the observational evidence is compelling. Studies of death rates show lung cancer increasing and lagging behind smoking rates by 20-25 years while other forms of cancer stay flat. Smokers have lung cancer and heart disease at rates greater than the nonsmoking population even after adjusting for whatever potential confounder the tobacco industry might propose. However, when smoking was first suspected of causing lung cancer and heart disease, Sir Ronald Fisher, then the world's greatest living statistician and a smoker, offered the "constitution hypothesis" that people might be genetically disposed to develop the diseases and to smoke, that is, that genetics was confounding the association. This was not an easy claim to put to an experiment. However, the hypothesis was put to rest in a 1989 Finnish study of 22 smoking-discordant monozygotic twins where at least one twin died. There, the smoker died first in 17 cases. In the nine pairs where death was due to coronary heart disease, the smoker died first in every case.

A good statistician will point out that causality can be proven only by demonstrating a mechanism. Statistics alone can never prove causality, but it can show you where to look. Perhaps no example better illustrates this than smoking and cancer/heart disease. Despite all of the statistical evidence, the causal relationship between smoking and disease will not be nailed down by the numbers but by the identification of the substance in tobacco that trigger the diseases.