To Do

Read Chapter 8

Do Chapter 8 Problems 1-6

Assignment 5 is due Monday December 5.

Detailed information about Final Exam is posted on Learn.

My exam office hours are posted on Learn.

Last Class

1) Two-Way Tables and Multinomial Models

2) Two-Way Tables and Testing for Independence of Two Variates

Today’s Class

1) What does the statement "X causes Y " mean?

2) Relationships and Causation (Confounding and Lurking Variates)

3) How to establish causation in experimental studies - the importance of randomization.

4) Observational studies and establishing causation.

Causality or Relationships Can Be Deceiving

Example 1

A strong correlation has been found in a certain city in Northern Ontario between weekly sales of hot chocolate and weekly sales of facial tissues.

Does this mean hot chocolate causes people to use facial tissues?

Example 2

Researchers have shown that there is a positive correlation between the average dietary fat intake and the breast cancer rate across countries. In other words, countries with higher intake tend to have higher breast cancer rates.

Does this prove that dietary fat **causes** breast cancer?

Example 3

If you were to draw a scatterplot of *number of women in the workforce* versus *number of Christmas trees sold* in Canada between 1930 and the present, you would find a very strong correlation.

Does more women in the workforce cause more Christmas trees to be sold?

Cause and Effect

What does the statement “X causes Y” mean?

Does your STAT 230 grade cause your STAT 231 grade?

Does your country of hometown cause your program?

Does your handspan cause your foot length?

Cause and Effect

Gravity causes dropped objects to fall to the ground.

Does smoking cause lung cancer?

Does everyone who smokes get lung cancer?

In most cases causation is difficult to define.

Cause and Effect

Consider the following definition:

Definition:

Let *y* be a response variate and let *x* be an explanatory variate associated with units in a population or process. Then, if all other factors that affect *y* are held constant, let us change *x* (or observe different values of *x*) and see if y changes. If it does we say that *x* has a causal effect on *y*.

Cause and Effect

The definition is not broad enough since a change in x may only lead to a change in the distribution of y.

For example, giving a person, who is at risk of a stroke, a small daily dose of aspirin instead of a placebo may not necessarily lower their risk. (Not everyone is helped by this medication.) However, on average the effect is to lower the risk of stroke in a population of people taking a daily aspirin.

Cause and Effect

A better definition of causation is to say that changing *x* should result in a change in some property of the distribution of the random variable *Y.*

For example:

(1) E(Y)

(2) P($Y > c$)

An example of (2) would be the probability that a randomly selected person has a stroke within 3 years if they are taking a daily dose of aspirin versus taking a daily dose of a placebo.

Cause and Effect

Improved Definition:

x **has a causal effect on *Y* if, when all other factors that affect *Y* are held constant, a change in *x* induces a change in a property of the distribution of *Y*.**

Unfortunately this definition is impractical since we cannot hold all other factors that affect *y* constant. (We may not even know what all the factors are!)

The definition serves as an ideal that should be used to conduct studies in order to show that a causal relationship exists.

Designing studies to show causation

Studies should be designed so that alternative explanations (to the variate *x*) of what causes changes in the distribution of *y* can be ruled out, leaving *x* as the causal agent.

This is much easier to do in experimental studies, where explanatory variates may be controlled, than in observational studies.

More on this shortly.

Some reasons two variates can be related:

1) The explanatory variate is the direct cause of the response variate.
2) The response variate is causing a change in the explanatory variate.
3) The explanatory variate is a contributing but not sole cause of the response variate.
4) Both variates are changing with time.
5) The association may be due to coincidence.
6) Confounding variates may exist.
7) Both variates may result from a common cause.

Reason 1: The explanatory variate is the direct cause of the response variate.

Occasionally, a change in the explanatory variate is the direct cause of a change in the response variate. For example, if we were to measure amount of food consumed in the past hour and level of hunger, we would find a relationship.

We would probably agree that the differences in the amount of food consumed were responsible for the difference in levels of hunger.

Unfortunately, even if one variate is the direct cause of another, we may not see a strong association.

For example, even though intercourse is the direct cause of pregnancy, the relationship between having intercourse and getting pregnant is not strong; most occurrences of intercourse do not result in pregnancy.

Reason 2: The response variate is the direct cause of the explanatory variate.

Sometimes the causal connection is the opposite of what might expect.

For example, suppose

response variate = hotel occupancy rate

and

explanatory variate = advertising sales (in dollars) per room.

You would probably expect that higher advertising expenditures would cause higher occupancy rates.

Instead, it turns out that the relationship is reversed because, when occupancy rates are low, hotels spend more money on advertising to raise them.

Thus, although we might expect higher advertising dollars to cause higher occupancy rates, if they are measured at the same point in time, we instead find that low occupancy rates cause higher advertising revenue.

Reason 3: The explanatory variate is a contributing but not the only cause of the response variate.

The complex kinds of phenomena most often studied by researchers are likely to have multiple causes.

Even if there was a causal connection between diet and a type of cancer, for instance, it would be highly unlikely that the cancer was caused solely by eating that certain type of diet.

It is particularly easy to be misled into thinking you have found a sole cause for a particular outcome, when you have actually found a *necessary contributor* to the outcome.

For example, scientists generally agree that in order to have AIDS, you must be infected with HIV. In other words, HIV is *necessary* to develop AIDS. But it does not follow that HIV is the *only* cause of AIDS, and there has been some controversy over whether that is actually the case.