

Intervention and prediction

StatPREP Class Activity

Orientation

The appropriate choice of one variable as the response and other variables as the explanatory depends on the purpose for which you want to describe the relationship.

Consider these two distinct purposes:

1. "I plan to intervene to change the value of variable B, and I want to know how A **will change** in response." (Example: "I'm going to turn down the thermostat setting, and I want to know whether water will condense on my windows.")
2. "I know the value of variable B, but not yet the value of variable A. I want to make a *prediction* of A from what I know about B." (Example: "I just noticed condensation on my windows and I want to figure out from that whether someone turned down the thermostat.")

In the case of (1), there's good reason to think that turning down the thermostat *causes* changes in the window condensation situation. The cause (thermostat) should be among the explanatory variables, the effect (condensation) should be the response variable.

In (2), it is still the case that thermostat causes condensation. But I know something about condensation and I want to figure out about the thermostat. So condensation should be the explanatory variable and thermostat the response.

Only when variable B *causes* A can we reasonably expect to anticipate the result on A of an *intervention* to change B. On the other hand, all we need is an *association* between A and B (regardless of the direction of causation) in order to be able to use knowledge of one variable to predict the other.

Statistical methods can be very powerful in detecting an *association* between two variables and exploiting that association to make a prediction of the value of one variable given the other. But in order to anticipate the result on A of a change in B, we need to bring into play our knowledge of how the world works.

Sometimes causation is a matter of common sense (the rising sun causes the rooster to crow), and sometimes it can be subtle or is a matter of your beliefs about how things work in the world. No matter, if there is an association between two variables, the causal possibilities are always one of a small set. The simplest to understand are these three:

- a. response causes explanatory.
- b. explanatory causes response.
- c. response & explanatory are both caused in common by another factor, C. (This is called a "common cause")

More subtle, and harder to understand even for professionals, is this one:

d. response & explanatory both cause another factor C, and the data include only some of the possible ra

Here are nine pairs of variables from data sets appearing in the Little Apps:

- Births_2014
 1. Whether the mother is covered by the WIC program and the age of mother.
 2. The age of the mother and of the father.
 3. The length of gestation and the baby's weight at birth.
- NHANES
 4. Systolic blood pressure and sex.
 5. Diabetes and age
 6. Weight and body mass index (BMI)
 7. Income and home_type
- diamonds
 8. The price of a diamond and its weight in carats.
 9. The weight of a diamond and its clarity.

For each of these pairs, consider A to be the first variable mentioned and B to be the second. For instance, in (7), A will be income and B will be home_type.

Activity

1. For each of the nine pairs of variables listed above choose an appropriate Little App to display variables A and B with A being the explanatory variable and B being the response variable.

*Viewing the data in the Little App, decide if there is any evidence for an **association** between the two variables. Write down your conclusions for the nine pairs of variables here: . . .*

2. Go back through the nine pairs of variable and decide which of the four causal arrangements common sense dictates as the relationship between A and B. For example, in looking at income and home type, it's fair to say that simply changing the type of home someone lives in does not change their

income, but changing someone's income may eventually change the type of home they live in.

For each of the nine pairs, select the causal mechanism (a), (b), (c), (d), or "none", that best accords with your understanding of how the world works. For each one, give a short explanation of your reasons for your decision.

. . .