Chapter 2.4 - Causation, Cautions, and Fallacies

> "If you can't see the causal relation of interest in the reduced form, it's probably not there." - Angrist and Krueger

It's difficult to discuss statistics at length without eventually bringing up gambling because it is *such an incredible tool* for explaining concepts in statistics. So let's play a game of blackjack and talk about one of the most important philosophical concepts in science.

> In blackjack, each player tries to get a hand value as close to 21 as possible without going over. Number cards are worth their face value, face cards (Jack, Queen, King) are worth 10, and Aces can count as 1 or 11. Each player is dealt two cards, and so is the dealer. Players can choose to "hit" (take another card) or "stand" (keep their current hand). If a player's total goes over 21, they "bust" (lose). Once all players finish, the dealer reveals their hidden card and must hit until they reach at least 17.

You and a friend are seated at the same blackjack table. You each have a total of 16 when you both decide to "hit". The dealer deals two cards to you and your friend quickly, your is the first card pulled and your friend's is the second. You receive a 10 and your friend receives a 5, you've lost the game and your friend has won. What *caused* you to lose?

Causality

It's a silly question with a simple answer. You lost because you went over 21, except that's not all of it. You went over 21 because the dealer dealt you a 10. Why was **that card** a 10? Well because that was the order of the deck… but **why**? The dealer stacked the deck. You were destined to lose. The world is inflicting karmic punishment on you.

It's easy to get lost in all of this; there are a plethora of possible "causes". The key point to focus on here is *proving* which *explanation* is the *cause* of the final *outcome* (you losing). Consider the following propositions:

a. Choosing to hit results in a loss, so you lost when you hit.
b. The dealer flipping over the card caused you to lose.
c. The dealer chooses when a loss occurs and chose you to lose.
d. The cards were shuffled in an order where losses are pre-decided. Thus you were always going to lose.

Two of these are easily refuted with intuition. For example, your friend chose to hit and they won so (a) cannot be true. The concept proposed in (c) is absurd since if the dealer can decide who loses they would be acting exclusively against their best interest by allowing your friend to win. Moreover blackjack dealer salaries are primarily held up by tips given to them from excited winners. Casinos are smart businesses and wouldn't let dealers rely heavily on tips if they could choose who wins and who loses. This would mean the dealer always selects a player besides themselves to win, which acts against the casino's interest.

Propositions (b) and (d) are quite interesting, and we should be able to refute simultaneously. The dealer flipping over the card places some assumption of *independence* on you, the player, from the outcome of the game. A pre-determined loss also uses that same assumption of independence. If this is true, we can state:

> Losses occur independently of player decisions.

To make this assumption true we would need to observe that players who make difference decisions (whether to hit vs. stand) have the same loss rates. However, if you had decided to stand your friend would have hit a bust and the dealer would have won due to the order of the cards. So we observe that players who make different decisions have different outcomes, which means they have different loss rates. By this logic:

> Player decisions **must** influence losses, and our original assumption is a contradiction.

What we've done is defeat both (b) and (d) with one statement. More importantly we're left with the conclusion that what **caused** you to lose was your **decision** to hit; everything else that happened was a result of that decision.

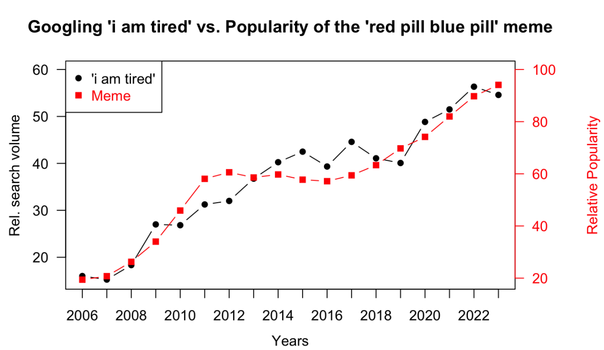
This is the basic premise of causality. Some event, object, or process is considered to have **caused** another event, object, or process to occur/exist. Causation is what we seek to understand in almost all of scientific inquiry. It's worth arguing that the pursuit of correlation is a result of insufficient tools or evidence to obtain causation.

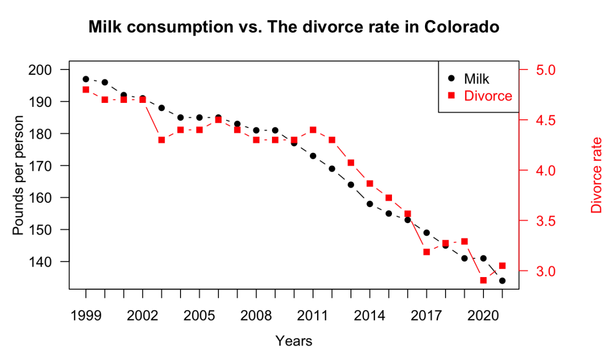
Causation demands **rigorous** proof— which for statisticians comes in the form of strong statistical results. These results can be obtained by tricky mathematics, large quantities of data, or very informative data. If you ever notice a *lack* of these pre-requisites for establishing causation you should approach those causal claims with skepticism. Many reporters, researchers, and business people claim causation because of how *powerful* the statement feels. Yet they are often far from establishing anything besides weak correlation.

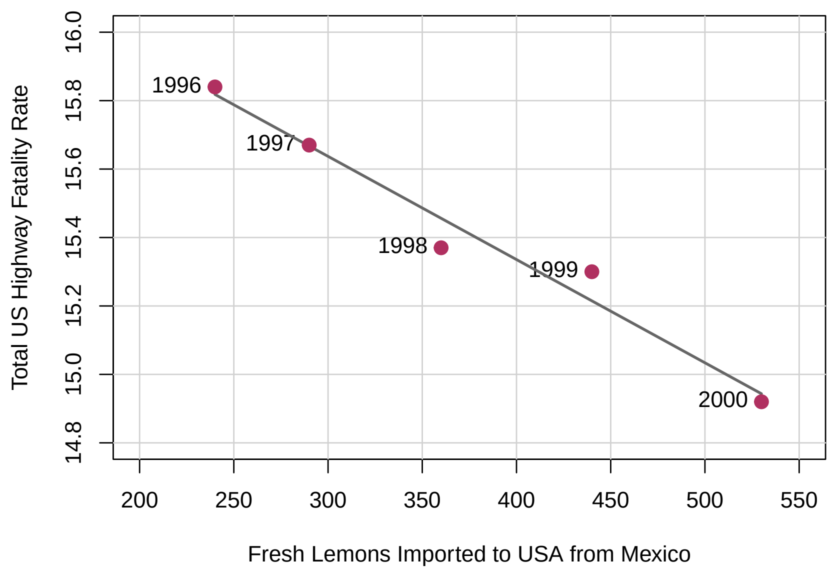
Correlation vs. Causation

It is the *joy* of every scientific educator to enlighten students to the mechanics and wonders of science. But it is their *job* to ensure that students recognize the difference between **correlation** and **causation**.

Recall the discussion on ice cream sales and homicides in Chicago. We established that while these two have a common cause they are not causing one another in any way. Sometimes we can show something that looks like **correlation**, explain it as **causation**, when there was never any true link between the variables. These are great party tricks (assuming statisticians ever attend parties) but they're not making any useful connections. With a investigation, scaling of axes, and searching for r values around -1 or 1, we can make some very ridiculous correlations:







While these may seem ridiculous to connect as causal in any way, sometimes the variables have a *rational* explanation for causation despite there being no direct link.

> The more firefighters that are sent to any given fire in America, the more damage is caused.

This one may feel obvious still, but it's still tricky. The firefighters aren't the reason for increased damage, the causal link is clearly *reversed*. But that's all it took: a change in direction of the relationship still presents a seemingly correct answer.

> On average, student's who get help from tutors receive much *worse* grades than the student's who do not.

This is a problem that's seen *real* and *dedicated* research. Does tutoring disable students from being able to work independently? Should we focus on individual learning rather than encouraging students seek help from peer educators?

Private tutoring companies don't make grade improvements the primary selling point for their services. They tend to focus on the proportion of students who *pass classes* that they receive help with. This is because students and parents of students who seek out private tutors aren't typically worried about getting a *good* grade as much as their worried about *not failing*. This another case of common cause: if a student is receiving help, it's because they're struggling. If they're struggling enough to pay for help, they're probably going to get a lower grade than those who never seek help to begin with *regardless of any effort by the tutor*.

When determining if a relationship is causal or not there are a few simple questions you can ask yourself:

1. Does it make scientific sense that A would cause B?
2. Is B causing A just explainable as A causing B?
3. Is there a distinct time difference between when A happens and B happens?
4. Does the source presenting the connection between A and B benefit more from one direction of the causal link than the other?

Cautions

Experimental design is a complex field due to the fact that it's attempt to capture the complexities of natural phenomenon. The inference we make from any given experiment is typically taken as **absolute**, so we have to be cautious and careful about the mandatory features and pitfalls of design.

Features

Individuals must all be treated identically in every way, except for the treatments being compared. We can't make inference about an effect of treatment if that effect has been muddied up by too many varying elements.

All studies must be replicable. It is absolutely vital that any given experiment CAN be reproduced if provided the same general materials/substrates/environment. Your specific lab bench at your University is not considered an environment in this case. If researchers cannot provide a cookbook with detailed steps that could allow anyone in the world to perform the exact same experiment and get **very similar** results, the experiment could be considered fraudulent.

Fallacies

> "If people do not believe that mathematics is simple, it is only because they do not realize how complicated life is." - John von Neumann

If a study is interpreting the results of an experimental analysis *as if they have captured a real life phenomenon exactly*, you should proceed with caution. The lab is never the same as the real world and the best we can do is try to mimic it poorly.

> About 15 years ago the Food and Drug Administration started requiring both sexes to participate in clinical trials for drugs applicable to both males and females. Before this, drugs were approved at the prescribed dosage for use by both men and women. The clinical trials only had males participating.

Recognize that the only **true** application of our results is in the one setting where we acquired them:

- Beware of lurking variables, bias, and confounding effect.
- If you're seeing a broad generalization of results, stay skeptical.

Problems in Analysis

Statisticians and scientists run into numerous oddities during design and analyses. In the past education around these concepts has focused on how to "handle" or "resolve" them, as if they're unfortunate occurrences. These days scientific fields are running low on interesting research questions— not because we're learning everything there is to science but rather because we're waiting for the next big scientific breakthrough.

A useful perspective to take on with these "problems" is to look at them as opportunities for further scientific inquiry. If you've found something strange in your results, while you may have messed up your entire experiment, you may have also found something messed up in science!

Confounding: Two variables in a study/experiment with relatively indistinguishable effects.

- We struggle to determine which variable caused an effect.
- These can mess with our ability to determine causality.
- Confounding is common in observational studies, hence why they don't generally show causality.
- Outside of exotic methods, designed experiments are the only option for causal inference.
  - We still have to **design** the experiment specifically to determine causality.

Consider the statement: "Drinking wine is linked to better health". Health is confounded with a bunch of messy variables like diet, wealth, and lifestyle. Since we can't control for these naturally our only option would be to construct an experiment. As we can image, this experiment is somewhat impossible.

Collinearity: Two variables are linearly dependent.

- We refer to this as an "extreme form of confounding".
- The variables contain the exact same information to an extent.

There are many fun examples of collinearity, one of which is the prediction of child heights. Upon observing students in an elementary school, what is the single best variable to collect in order to predict their height?

Bias: The design of statistical study is biased when one outcome is systematically preferred to others.

- This is just one form of bias, but the one we should be most afraid of.
- "Systemic bias" as it's called, is **impossible** to adjust for.

Think back to the issue of political polling. If a pollster calls people until 1000 opinions have been gathered, they've **over-represented** a population with middle to upper class socio-economic status. Additionally, the elderly disproportionately answer their phone more than young people. Even the young people who answer the phone are wildly different than the ones who don't.

Ethics

This is not a book on moral philosophy as I (the writer) am not qualified to teach that topic. Most students will take moral philosophy as well as some form of ethics course prior to completing an undergraduate program. The process of learning about those concepts should be primarily left to the subject matter experts. This section just covers a few brief statements and some scenarios that should be considered when thinking about ethics in experimentation.

Incentives: If you're conducting a trial to determine the impact of a social welfare program, should you reward your participants?

- Poor/homeless individuals are easy to exploit
- Paying people for their time and effort is how our economic system functions

Placebos: You acquire 3000 patients for an experimental HIV vaccine trial and provide 100 of them a placebo.

- Is it wrong to administer a placebo?
- You've created a clear effect of treatment and eliminated a lot of confounding variables. Isn't that good?

Intellectual Property: You want to determine the pest resilience of a strain of GMO corn. Your experiment is industry sponsored and is taking place at an industry owned farm, directly adjacent to a family owned farm. Your crop spills over to the family owned farm and changes the genetic makeup of their crop.

- Have you harmed or helped them?
- Do you owe them compensation?
- Are they stealing your intellectual property by selling their crop now?

Case Deletion: You've spent your entire Masters' program working on a project to determine the correlation between a gut microbe and Irritable Bowel Syndrome. After 2 years, your results are "null" (there isn't a correlation) meaning that you can't publish your work. If you remove a few pieces of data you can make the results slightly "significant" (there's a weak correlation). You're also fairly confident that the data are outliers attributable to measurement error.

- Is it right to remove the outliers if you disclose that you've done that in your publication?

A good rule of thumb for all experimentation is to ask yourself if someone has done what you're planning to do and has been hanged in the past for violating rights of some form. This isn't a definitive way to avoid ethical dilemmas but it should filter out the glaring ones.

Examples

Gallup polls considers a proper poll to be a representative sample of 1000 people. A local news station resolves to follow this rule of thumb and call 250 individuals from 4 different age groups as part of the design for a political poll.

Age Group	18 – 26	27 – 40	41 – 52	52+

They resolve to only count calls that are answered, rather than marking unanswered calls as non-respondant. They collect their results, develop their graphics, and run their "Political Statistics" segment as planned.

What are the components of this study?

Are there any problems with the study?

Below is a set of questions in a poll of visitors to a National Park:

1. How much do you appreciate the incredible natural beauty and unique wildlife found in this national park?
2. How satisfied are you with the park's facilities, such as trails, restrooms, and visitor centers?
3. Would you recommend visiting this national park to your friends and family?
4. Do you think this national park does an excellent job preserving its natural beauty and wildlife?
5. How would you rate your overall experience at this national park today?

Each question used a scale of 1-5, where 1 was the most negative relation and 5 was the most positive. The survey was randomly given to 500 park visitors every month for 10 months.

What are the components of this study?

Are there any problems with the study?

A study investigates whether a new type of fertilizer is linked to an increased risk of skin irritation among farmers. The researchers recruit 200 farmers with skin irritation (cases) and 200 farmers without skin irritation (controls). They ask both groups about their fertilizer use over the past five years.

What are the components of this study?

Are there any problems with the study?

A school district examines whether longer lunch breaks improve student test scores. They collect test score data from schools that already have longer lunch breaks and compare them to schools with shorter breaks, adjusting for factors like socioeconomic status, class size, and teacher experience.

What are the components of this study?

Are there any problems with the study?

A 10-year study follows 1,000 people who regularly eat spicy food and 1,000 who do not, examining the incidence of stomach ulcers in each group.

What are the components of this study?

Are there any problems with the study?