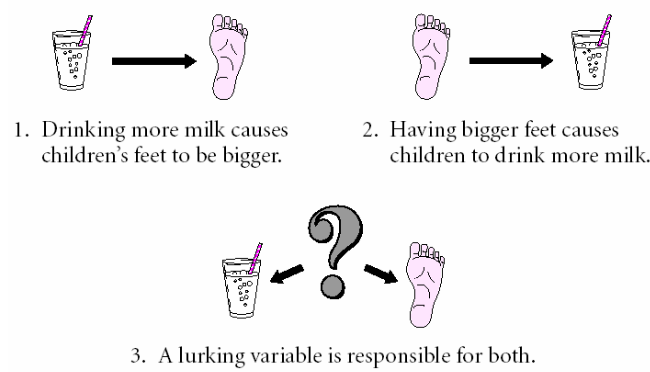
Math 140
Introductory Statistics

Professor Bernardo Ábrego
Lecture 10
Section 4.3, 5.1

4.3 Experiments and Inference about Cause

- Cause and Effect



4.3 Experiments and Inference about Cause

- Cause and Effect
- Lurking Variable:

  A variable in the background that could explain a pattern between the variables investigated.
- How to establish cause and effect?

Answer: Conduct an experiment.

Experiments

- Goal: To establish cause and effect by comparing two or more ***conditions*** (called treatments) using an ***outcome variable*** (called the response).
- To be a real experiment, the subjects must be randomly assigned to their treatments. To make this distinction sometimes we call these Randomized Experiments.

Example: Kelly's Hamsters

- Assumptions
  - Golden Hamsters hibernate.
  - Hamsters rely on the amount of daylight to trigger hibernation.
  - An animal's capacity to transmit nerve impulses depends in part on an enzyme called Na^+K^+ ATP-ase.
- Question: If you reduce the amount of light a hamster gets, from 16 hours to 8 hours per day, what happens to the concentration of Na^+K^+ ATP-ase.
- Subjects: Eight golden hamsters.
- Treatments: Raised in long days (16 hours) or short days (8 hours) of daylight.
- Random Assignment of Treatments: Kelly randomly assigns four of the hamsters to short days, and four to long days.
- Replication: Each treatment was given to four hamsters.
- Response Variable: Enzyme concentration.

Kelly's Hamsters (Results)

- Results

Enzyme concentrations in milligrams per 100 milliliters.

Short Days	12.500	11.625	18.275	13.225
Long Days	6.625	10.375	9.900	8.800

Kelly's defense of her design

Kelly: I claim that the observed difference in enzyme concentrations between the two groups of hamsters is due to the difference in daylight.

Skeptic: Wait a minute. As you can see, the concentration varies from one hamster to another. Some just naturally have higher concentrations. If you happened to assign all the high-enzyme hamsters to the group that got short days, you'd get results like the ones you got.

Kelly: I agree, and I was concerned about that possibility. In fact, that's precisely why I assigned day lengths to hamsters by using random numbers. The random assignment makes it **extremely unlikely** that all the high-enzyme hamsters would get assigned to the same group. If you have the time, I can show you how to compute the probability.

Skeptic: (*Hastily*) That's OK for now. I'll take your word for it. But maybe you can catch me in Chapter 6.

Discussion D21 (page 245)

- **D21**. Kelly has shown that hamsters raised in less daylight have higher hormone concentration than hamsters raised with more daylight. In order for Kelly to show that less daylight *causes* an increase in the hormone concentration, she must convince us that there is no other explanation. Has she done that?

Confounding in Observational Studies

- ***Confounded***: mixed-up, confused, at a dead end.
- Two possible influences on an observed outcome are said to be confounded if they are mixed together in a way that makes it impossible to separate their effects.

Confounding in Observational Studies

Imagine yourself in this situation:

- You know that many infants are dying of what seem to be respiratory obstructions.
- You begin to do autopsies on infants who die with respiratory symptoms.
- the infants all have thymus glands that look too big in comparison to body size. Aha! That must be it:
- The respiratory problems are caused by an enlarged thymus.
- It became quite common in the early 1900s for surgeons to treat respiratory problems in children by removing the thymus. Even though a third of the children who were operated on died.

		Age	
		Child	**Adult**
Thymus Size	**Large**	Problems	No evidence
	Small	No evidence	No problems

Experiments vs Observational Studies

- The best solution to guard against confounding: To randomize.
- Observational Study: No treatment gets assigned to the subjects by the experimenter.
- (Randomized) Experiment: Comparing results of treatments assigned to subjects at random.
- Clinical Trial: Randomized experiment comparing medical treatments.

- For observational studies the conditions are called factors, (not treatments)

Factors and Levels

- The term factor is also used for experiments when there are many characteristics that want to be compared.
- The different values that a factor may take are called levels.
- Example. If Kelly added the type of diet to her experiment.

		Factor 1 Type of Diet	
		Light	Heavy
Factor 2 Length of Day	Short	Light-Short	Heavy-Short
	Long	Light-Long	Heavy-Long

Why randomization makes inference possible?

- By assigning treatments to units at random, there are only **two possible causes** for a difference in the responses to the treatments: **chance** or **the treatments**.
- If the probability is small that chance alone will give such a difference in the responses, then we can **infer** that the cause of the difference was **the treatment**.

Control or Comparison Group

- Anecdotal evidence is not proof. Why?
- Placebo Effect: When people believe they are getting special treatment they tend to improve.
- Control Group: A group of people given a placebo.
- Comparison Group: A group of people given the standard treatment (when comparing against a new treatment).
- Blind Experiment: People do not know which treatment they are given.
- Double Blind Experiment: patients and doctors do not know which treatment they are assigned.

5.1 Models of Random Behavior

- Outcome: Result or answer obtained from a chance process.
- Event: Collection of outcomes.
- Probability: Number between 0 and 1 (0% and 100%). It tells how likely it is for an outcome or event to happen.
  - $P = 0$ The event cannot happen.
  - $P = 1$ The event is certain to happen.

Where do Probabilities come from?

- *Observed data* (long-run relative frequencies).
  For example, observation of thousands of births has shown that about 51% of newborns are boys. You can use these data to say that the probability of the next newborn being a boy is about 0.51.
- *Symmetry* (equally likely outcomes).
  If you flip a fair coin, there is nothing about the two sides of the coin to suggest that one side is more likely than the other to land facing up. Relying on symmetry, it is reasonable to think that heads and tails are equally likely. So the probability of heads is .5.
- *Subjective estimates*.
  What's the probability that you'll get an A in this statistics class? That's a reasonable, everyday kind of question, and the use of probability is meaningful, but you can't gather data or list equally likely outcomes. However, you *can* make a subjective judgment.