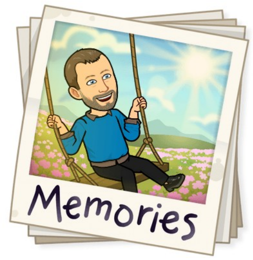
Introduction to Statistics I

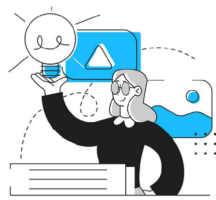
Textbook: Elementary Statistics (4th Edition, by Navidi and Monk), and Workshop Statistics (4th Edition, by Rossman and Chance).

Previous Lecture

- **Clarified**: What's a variable (var)?
- Research Questions (RQ)
- **Variable Types**: Quantitative/Categorical (Binary)



§1.3: Designing Experiments



Experiments (defined below) are good at discovering **cause** & **effect** relationships.

You may have heard: *"Correlation does not imply causation."* What does this mean?

Example RQ: At home games, does our basketball team perform better (win more games) when there's a sellout crowd?
Obv-unit/var(s)/var type?

Obv-unit: Home games

Var(s): Whether or not they won (qual/binary)
Whether or not there was a sellout crowd (qual/binary)

Poll:

bit.ly/introstatsdata

Poll: Sellout Crowds

Data: Oklahoma City Thunder, 2009

- **Sellout Crowds**: 3 wins, 15 losses

Wins $\frac{3}{18} = 0.167$. So, 16.7% win rate.

- **Non-Sellout Crowd**: 12 wins, 11 losses

Wins $\frac{12}{23} = 0.522$. So, 52.2% win rate.

The data is against our hypothesis!

So "not selling out" is *correlated* with winning (they happen together), does this mean "not selling out" *causes* our team to win?

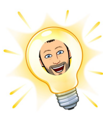
— No, correlation does not imply causation. We need a new idea. —

- **Explanatory Var**: The var we hypothesize is causing an effect.
- **Response Var**: The var we hypothesize is being affected.

Explanatory Var ————> Response Var

Back to Example:

Thunder appears to lose more frequently in sellout crowds. Why?



Thunder was new in 2009. Maybe sellout crowds weren't coming to see *them*?

New Hypothesis

Games sellout more often when *good opponents* play against the Thunder.

The Thunder (obviously) loses more frequently to good opponents.

New Analysis: **Does "Good Opponent" affect "Sellouts"** ??

Good opponents can be defined as teams that win more than ½ their games.

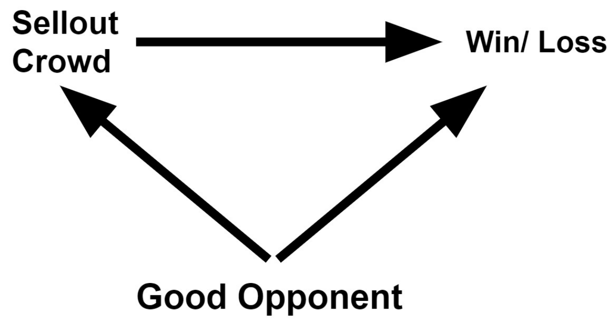
Data: 22 games were against good opponents. 13 of those 22 games were sellouts.

$\frac{13}{22} = 0.59$ or 59%.

19 games were against "bad" opponents. 5 of those 9 were sellouts.

$\frac{5}{19} = 0.26$ or 26%.

Conclusion: "Good opponent" is a **confounding var** in the relationship between sellouts & wins/losses (it affects both response & explanatory).



Var Types

Explanatory Var: A var we hypothesize is causing an effect.

Response Var: A var we hypothesize is being affected.

Confounding Var: A var that could affect both **response** & **explanatory** vars.

Lurking Var: A var we didn't measure that could affect **response** var.

Example: Childhood Obesity and Sleep.

A 2006 article found that children who report sleeping more hours each night were less likely to be obese than children who report sleeping fewer hours.

Explanatory ——————————> Response ...

Sleep Obesity

Potential Lurking Vars? ...

Medication	Economic Standing	Self Esteem
Diet	Family History	Race
Income	Mental Health	Schooling
Exercise	Social Environment	Genetics
Nutrition	Quality of Sleep	Food Availability

Could these also affect the explanatory variable? (i.e., could they be confounding?)

Generalizing: The studies mentioned above are what are called observational studies.

Observational Study: A study in which data is collected and observed, but the explanatory var is not controlled in any way by the researcher.

So, the researcher doesn't randomly assign a "treatment" (e.g., 8 hrs sleep) to different obv-unit subgroups.

Observational studies cannot control for confounding variables.

Thus, observational studies can examine **relationships between vars**, but ...

❗ They can **NEVER** conclude a **cause and effect relationship**!

Researchers **CAN** draw cause-and-effect connection between explanatory and response var when their study **actively and randomly imposes the explanatory var (or treatment) on the obv-units**.

Such a study is called an *experiment*.

Experiments

How do we gather evidence to support a claim?

Suppose you're serious about becoming a better basketball player.
Your friend tells you that plyometric shoes have improved their jumping.

Should you spend $180 on these shoes?

Plyometric Shoe

Anecdotal Evidence: a type of evidence based on personal experiences or accounts.

Anecdotal evidence can help start a scientific inquiry, but can't be used to answer it.

RQ: Do plyometric shoes improve jumping ability?
We take a SRS of students, ask each whether they wear plyometric shoes, and test how high they can jump.

Two vars:

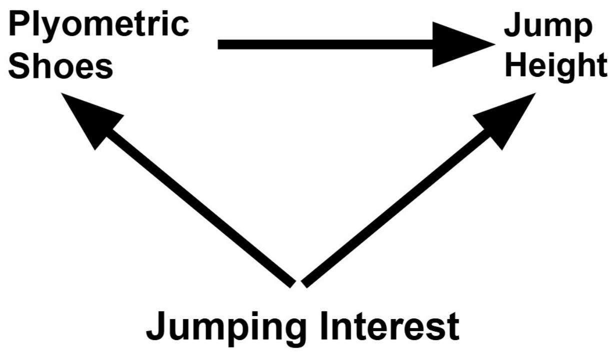
Do they wear plyometric shoes? How high can they jump?

Study Design

Plyometric Shoes → **Jump Height**

Type of study?

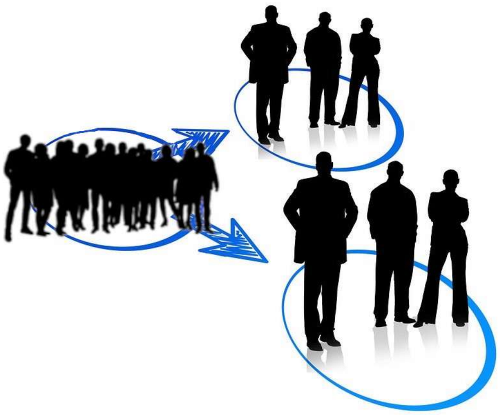
This is an observational study. Also, note that interest in jumping high is a potential confounding var!



Random Assignment

To prevent confounding vars, we need to control the explanatory var.

We do this through **random assignment**.



Randomly assign people in your sample to **wear** plyometric shoes or not.

Then measure their jump height.

What's affecting the explanatory var now? Our random assignment!

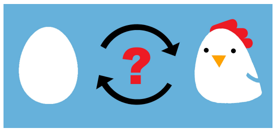
What about other things that affect jumping?

Random assignment will cause any variable affecting jumping (like a person's height or jumping interest), on average, to be split between the two groups evenly. So it won't affect the average jumping height of the groups.

An **experiment** is a study that uses random assignment to create treatment groups.

Because random assignment is the only thing affecting the treatment/explanitory var, there can be no confounding vars.

❗ Therefore, **experiments allow us to make cause-and-effect conclusions**.



Placebo Treatments

Placebo: A treatment that has no medical effect, such as a sugar pill.

Placebo Effect: the fact that some people feel better after taking a placebo (even if they know it's a placebo!).

Activity: 1.3a

Botox for Back Pain

Does botox relieve back pain?

Study 1: Thirty patients with severe back pain were given botox injections. Twenty reported improvement. Can we conclude Botox reduces back pain?

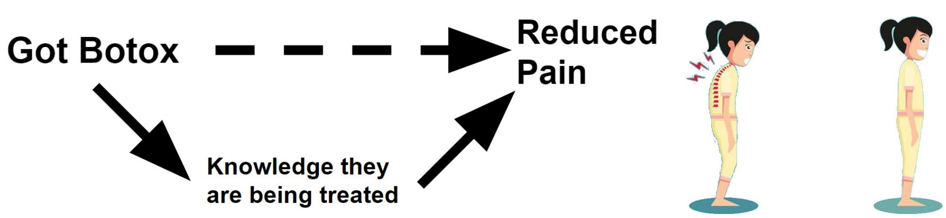
Study 2:

- 15 patients randomly assigned to receive Botox treatment. (**treatment group**)
- 15 patients randomly assigned to receive no treatment. (**control/comparison group**)

Questions:

- Name two ways in which study 2 is an improvement over study 1.

- Can you think of a flaw with study 2?
  (How are medical studies *ideally* designed?)



Placebo Treatment

Sometimes, knowing that you're receiving treatment is enough to improve your condition.

To prevent the **placebo effect**, those in the control group will receive a non-effective treatment, called a **placebo**.

Blindness

When subjects don't know which treatment group they belong to, the study is called **blind**.

We make studies blind by using a placebo, or by hiding the goals of the study.

If the researcher performing the experiment *also* doesn't know which group a subject belongs to, it's called **double-blind**.

Double blind studies are the gold standard, though it can't always be achieved.

Typically, an experiment has a **treatment group** (or groups) and a **control group** (no treatment or placebo treatment).

A **comparison group** refers to either a control group or a group receiving a different treatment than the one primarily being studied.

Memorizing Letters: You have 20 secs (used in 1.3b).

Activity: **1.3b**

What did we learn?

- Explanatory/Response/Confounding/Lurking Vars
- Observational Studies & Cause/Effect
- Anecdotal Evidence
- Experimental design: Random assignment
- Placebo/Blindness
- Control/Treatment groups