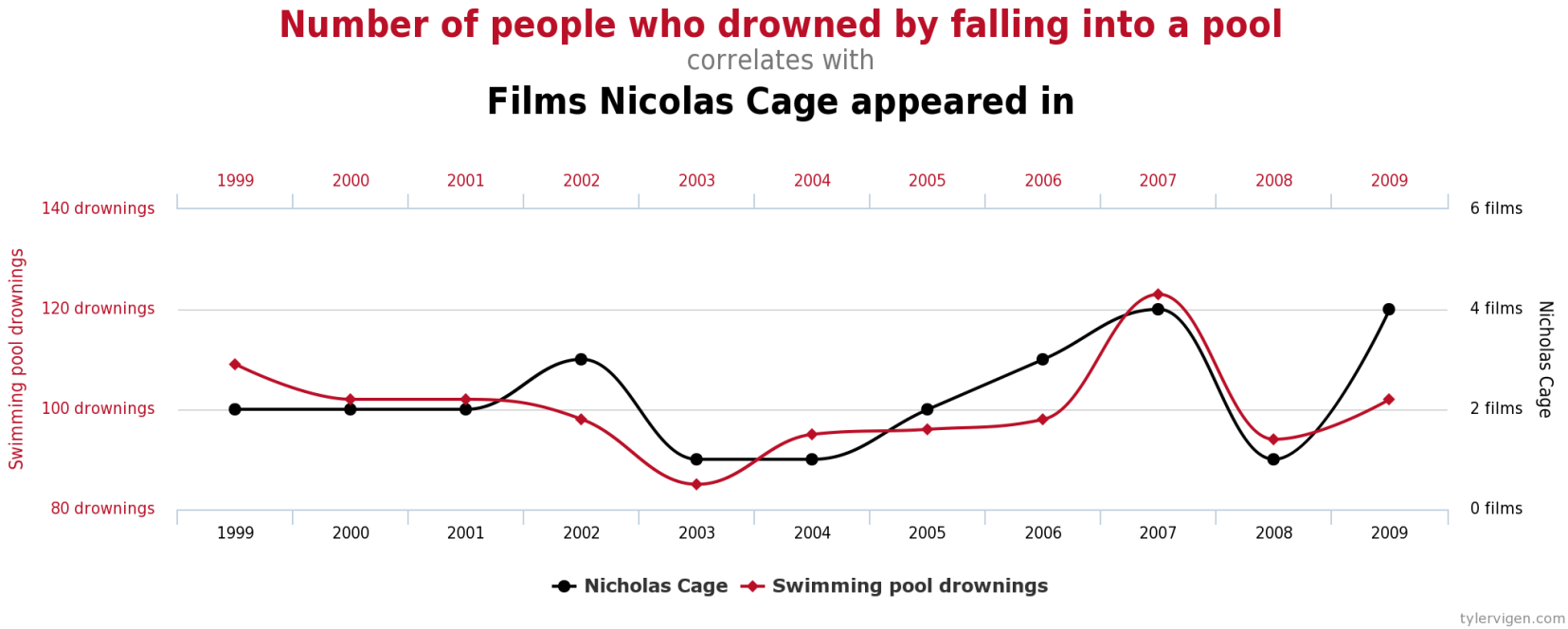
Causal Inference Introduction

Paul English

November 19, 2016

Correlation does not imply causation



http://www.tylervigen.com/spurious-correlations

Causal Effect given access to a Time Machine

How we would measure causal effect if we had a time machine:

subject	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
Joe	-5	5	-10
Mary	-10	-5	-5
Sally	0	10	-10
Bob	-20	-5	-15
Mean			-10

Causal Effect since we don't have a time machine

But we haven't found any time machines yet, so we're usually stuck with this:

subject	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
Joe	?	5	?
Mary	-10	?	?
Sally	?	10	?
Bob	-20	?	?
Mean			?

Causal Effect since we don't have a time machine

What if we look at the average of each group?

subject	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
Joe	?	5	?
Mary	-10	?	?
Sally	?	10	?
Bob	-20	?	?
Mean	-15	7.5	?

Average treatment effect: $\bar{Y}(1) - \bar{Y}(0) = -22.5$

Causal Effect since we don't have a time machine

- In general we cannot observe the causal effect directly.
- Estimating causal effects requires:
  - substitutes for the potential outcome,
  - randomization,
  - or statistical adjustment.

Substitues to Potential Outcome

Examples:

- Maybe you can repeat treatment, e.g. drinking tea before bed.
- Dividing up a piece of plastic and exposing it to a corrosive chemical.
- The effect of a diet over time by measuring weight.

These tend to carry strong assumptions that may be implicit in the choice of substitution.

The Gold Standard: The Randomized Controlled Trial

- We cannot compare treatment and control on the same units, so we compare similar units.
- Selection bias is avoided through randomization.
- Well-proven methodology and typically one of the best ways to design a study.

The Gold Standard: The Randomized Controlled Trial

Problems:

- It's not always possible to conduct an experiment.
- It could be cost prohibitive.
- Participants could self-select into the treatment group, e.g. company wellness programs.
- You may not be involved in the study design, and only receive data post-hoc.
- It might be unethical to control treatment.

Statistical Adjustment

- Usually attempts to approximate what a random experiment can achieve.
- Attempts to create similar units.
- Regression estimate of the outcome.
- Matching to achieve balance.

The Rubin Causal Model

For the moment let's assume randomization and revisit our example:

subject	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
Joe	?	5	?
Mary	-10	?	?
Sally	?	10	?
Bob	-20	?	?
Mean	?	?	?

What if we can estimate the unknown values which we can't observe in practice?

The Rubin Causal Model

Building a linear model on treatment alone,

$$\hat{Y}_i = \alpha + \tau D_i + \epsilon_i$$

we have the least squares estimator of,

$$(\hat{\tau}, \hat{\alpha}) = \arg\min_{\tau,\alpha} \sum_{i=1}^{N} (Y - \alpha - \tau D_i)^2$$

The Rubin Causal Model

Solving for $\hat{\tau}$ we end up with,

$$\hat{\tau} = \frac{\sum_{i=1}^{N}(D_i - \bar{D})(\hat{Y} - \bar{Y})}{\sum_{i=1}^{N}(D_i - \bar{D})^2}$$

This can be shown to be equal to,

$$\hat{\tau} = \bar{Y}(1) - \bar{Y}(0)$$

That is our estimator τ is identical to the difference in average outcomes of treatment status.

The Rubin Causal Model

In general we can estimate our treatment effect with regression

- Controlling for covariates:

$$\hat{Y}_i = \alpha + \tau D_i + \beta X_i + \epsilon_i$$

- Interaction with the treatment:

$$\hat{Y}_i = \alpha + \tau D_i + \beta X_i + \gamma D_i(X_i - \bar{X}) + \epsilon_i$$

Causal Inference Assumptions

What are we assuming here?
Treatment is *Strongly Ignorable*:

- Unconfoundedness: D is independent of $(Y(0), Y(1))$ conditional on $X = x$, e.g. the treatment of one group does not affect the other group.
- Overlap: $c < \mathbb{P}(D = 1|X = x) < 1 - c$, for $c > 0$.

Matching

- Matching tries to avoid and remove selection bias from datasets.
- The goal is to approximate a randomized or controlled trial.
- There are various ways to do this, the most commonly seen is propensity matching.
- Can also be good for unbalanced groups in any trial, where the size of one group is significantly smaller than another (Maybe an observational study).

Propensity Score Matching

A propensity score is an approximate model of how likely a subject is to have been in the treatment group.

$$P(D_i = 1|X_i)$$

Solve this using a regression,

$$D_i = \text{logistic}(\alpha + \beta X_i + \epsilon_i)$$

We can then select the control group from the non-treatment population by matching members based on minimizing the difference in this score, e.g. pairing.

Distance Matching

Additionally we could minimize the distance between covariates,

$$m(i) = \arg \min_{j: W_j \neq W_i} ||X_j - X_i||$$

Where we define $||X_j - X_i||$ as a distance between the covariate vectors X_j and X_i as follows:

$$||X_j - X_i|| := (X_j - X_i)' W (X_j - X_i).$$

where W is defined as

$$W = \text{diag}\{\hat{\sigma}_1^{-2}, \ldots, \hat{\sigma}_K^{-2}\}$$

also known as Mahalanobis distance.

Examples

Let's try some of this out.