9.1 Notes (2)

Causal Inference (2)

Always be mindful of

(1) Corr w/o caus

(2) Simpson's Paradox

Individual Treatment Effect (ITE)

$Y_i(1) - Y_i(0)$

Average Treatment Effect (ATE)

$E[Y|T=1] - E[Y|T=0]$

Randomized Control Trials

require us to randomize subjects into treatment & control groups

9.2 Notes

Randomized experiments are the gold standard

Completely Randomized Design (CRD) is when you have 2 treatment gps., and you assign n people into each gp. so that each gp. has $\frac{n}{2}$ ppl.

↳ This is an **AB test**

E.g. Google's Blue study where they tried 41 marginally different shades of blue for their link to test click-thru rates

Via LOTP: $E[Y] = E_c[Y|t,C]$

$= \sum E_c[Y|t,c]P(c)$

How to do a causal analysis when you weren't allowed to randomize?

- Maybe do a kitchen-sink regression
- Isolate similar data w/ same background characteristic
  - ↳ creates a new T and C gp. that would take w/ replacement
  - ↳ then take the difference between each response (D, D?)

"Similarity" is easy for few, qualitative predictors. How do you solve the matching problem?

Propensity scores consider predictors and map this to a singular value. Then, it matches on the 1D features

- $e(X_i) = P(T_i = 1|X_i)$
- basically do regression w/ treatment as the outcome, then the PS is the predicted prob. of being treated

The Problem

Unmeasured confounders can't be controlled for whereas randomization does do that