RCT - Randomized Control tests

Goal: Does Sending email increase purchase Conversion

Select users at random

Control grp | Treatment grp

→ When we can't run RCT

ex:- Setting BillBoards

→ Experiment takes to long.

Causal Inference

Confounders

goal: Does elixir make you feel better

Flu

Treatment grp - Users who use elixir
Avg age : 35

Control Group - " " don't
Avg age : 65

Treatment grp recovery >>

So, it doesn't prove elixir worked

Here Age is Confounder

2. Selection BIAS

Here treatment represents young people so

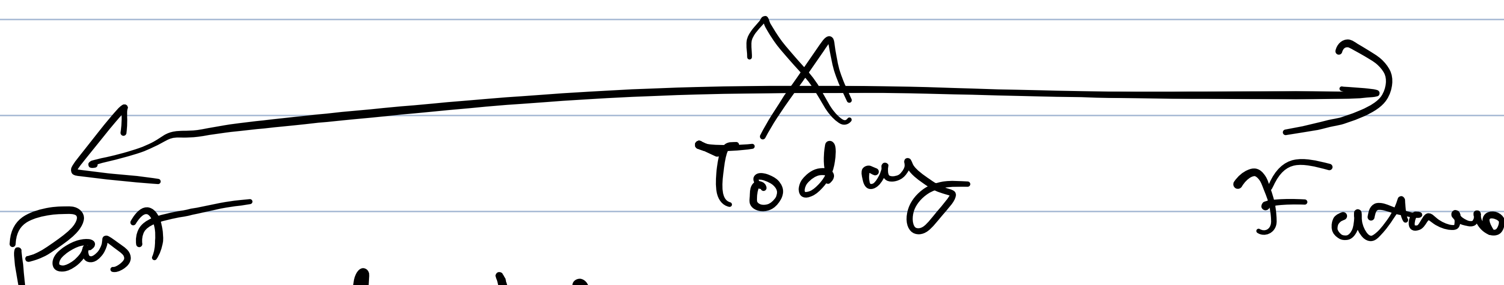
3. Counterfactual

what would have been
case if the person has
not recieved the elexia

C.I

Assumptions

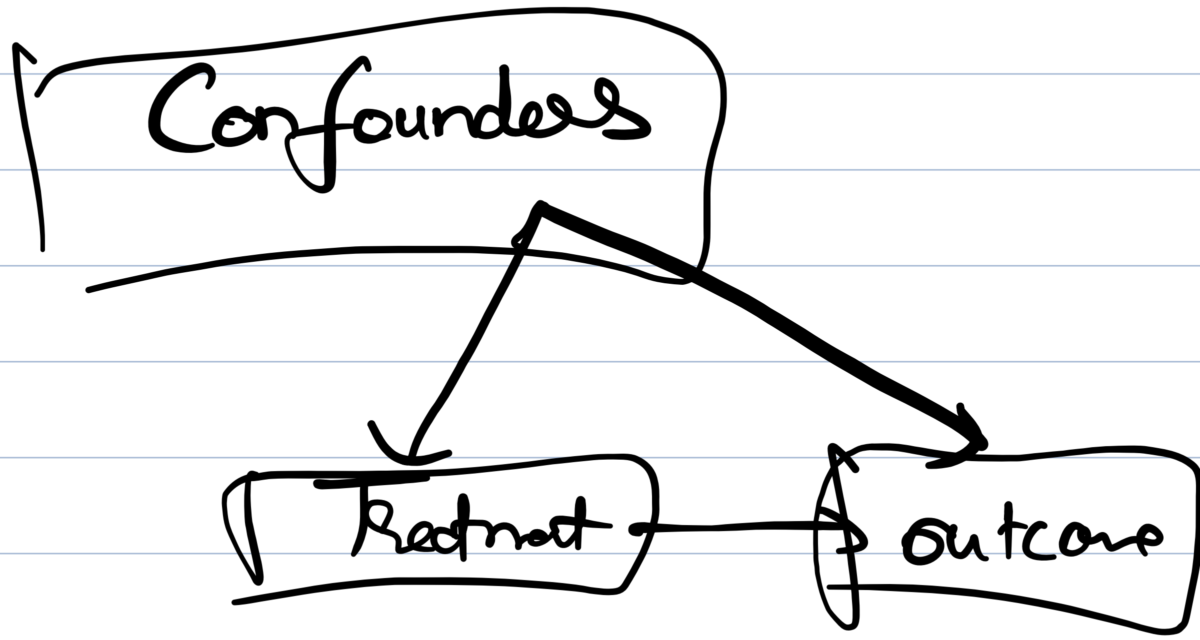
Data



CI from here
needs to be RCT

1. Causal Markov Condition:

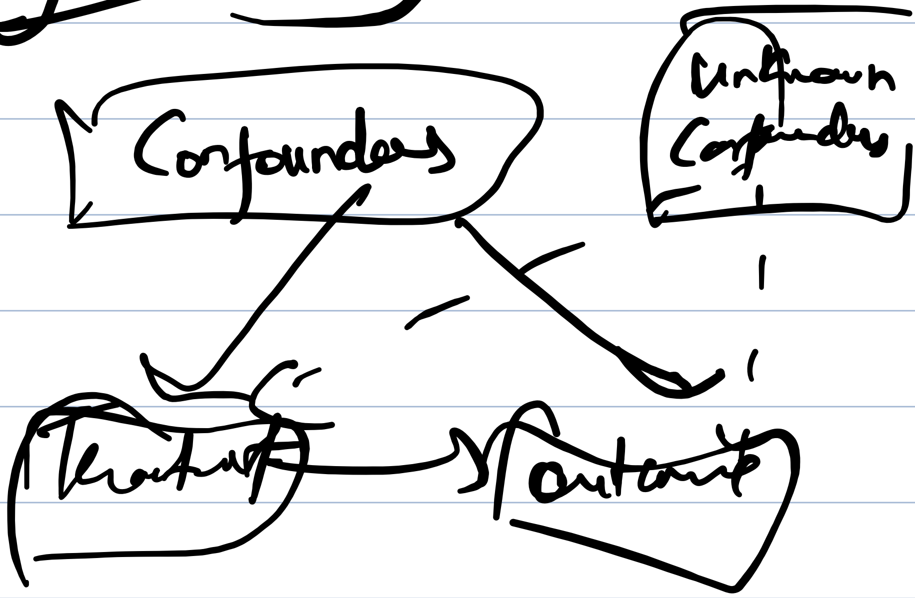
Causal graph { DAG }



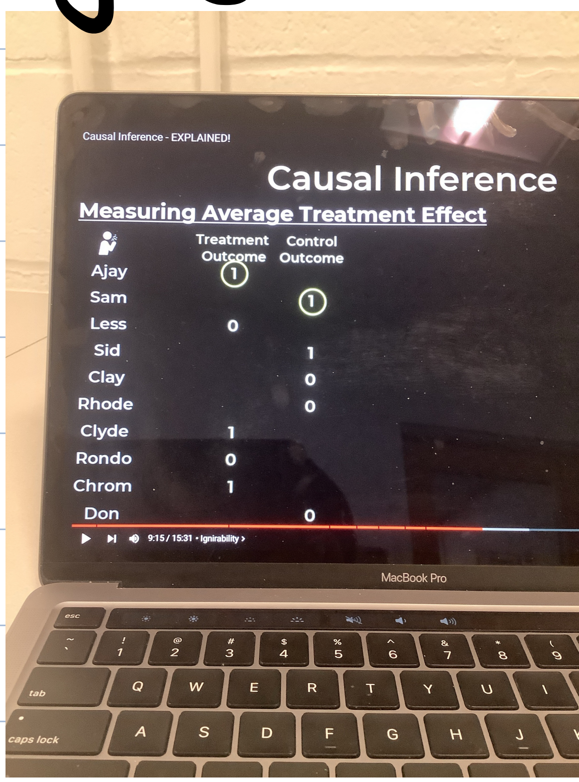
2. SUTVA (Stable Unit Treatment Value Assumption)

ex: people who recieved elixir influencing people who don't have elixir

3. Ignorability



Measuring Avg Treatment effect



1 Means got Better
0 means didn't
Treatment got elixir
outcome didn't get elixir

Does elixir make people Better

Count people who got elixir & Better

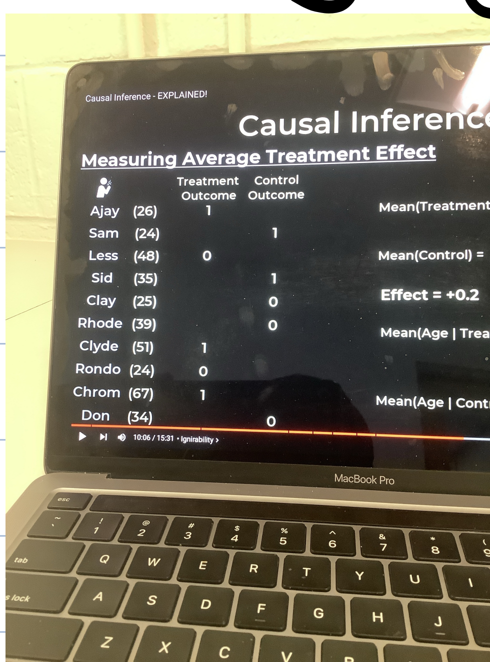
$$\text{Mean Treatment} = \frac{1+0+1+0+1}{5} = +0.6$$

$$\text{Mean Control} = \frac{1+1+0+0+0}{5} = +0.4$$

↳ $\dfrac{\text{people who didn't get elixir \& Better}}{\text{people who didn't get elixir}}$

Effect = +0.2

Now Adding Age (Average)



$$\text{Mean}(\text{Age} \mid \text{Treatment}) = \frac{26+51+67}{3} = 48$$

$$\text{Mean}(\text{Age} \mid \text{Control}) = \frac{24+35}{2} = 29.5$$

Avg age who recieved elixir's 48 & who didn't is 29.5

$\Rightarrow$ So, here Age is Confounding Variable

Counterfactuals

- Matchings
- ML Imputing

In the Above Example

Sam (24) & Rondo (24)
are Same age & recieved
diff treatments

	Treat out	Control out
Sam (24)	0	1
Rondo (24)	0	1

Matching ways means
Basically finding another
person in Same age who
recieved different treatment

Blue - Original

Red - Counterfactuals
Based on Vice Versa

Another Approach
is
ML

ML Approach

Confounders → $\boxed{\text{Model}}$ → outcome

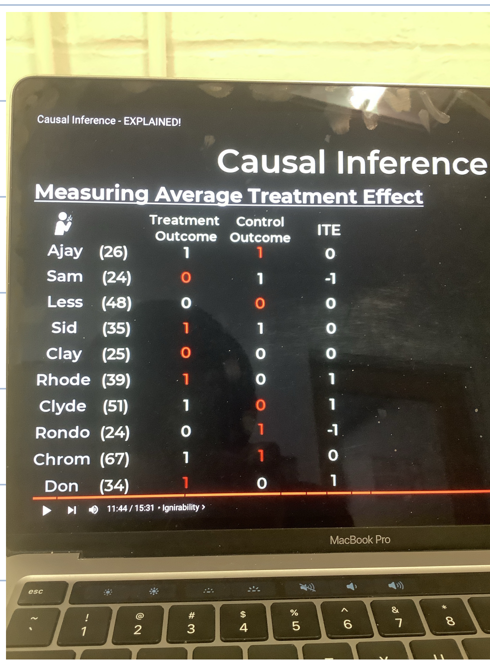
↑

Treatment

We Build Model Takes Age & treatment as input and then predict output. we will train factual data & try to predict Counterfactuals

---

Consider the Red one' are Counterfactuals



To determine Avg treatment effect = we Subtract the Case where person got (or) would have gotten treatment with Case where they had not gotten (or) would not have gotten elixir treatment

ITE → Avg Treat effect
we calculate for every single individual

Avg of Individual Treatment effect

$$ATE = \frac{0+(-1)+0+0+0+1+1+(-1)+0+1}{10}$$

$$= +0.1$$

So, it looks like elixir does help even when accounting the age ⇒ +0.1

Now Considering Above, let's say everyone who has flu let's just give elixir

Since, here Age is Confounding variable

Let's determined ATE Conditioned on age $\rightarrow$ Conditional ATE

$$CATE(Age >= 35) = \frac{0+0+1+1+0}{5}$$

ITE Values

$$= +0.4$$

$$CATE(Age < 35) = \frac{0+(-1)+0+0+(-1)?!}{5}$$

$$= 0.2$$

Average of ITE Values of those Values of Age

↳ These are Called Treatment Heterogeneity

Based on this we can Conclude

⇒ it can help older people But doesn't seem to have positive effect on younger people

Here we didn't use RCT

1.) we used Causal Analysis / Inference that Simulates like RCT.

2.) Confounding, Selection Bias makes Causal inference on past data Challenging

3.) Assumptions are required for Causal Inferencing.

4.) ATE useful when deciding on population, CATE on subsets of population