Causality

By Saurav

Let's start with an example and a question!

20 students in a college trek got ill. All were given one Herb in some village. Everyone became well after 4 days.

Now, can we say “The Herb caused the recovery”?

Correlation != Causation

But if A and B are correlated (let's say strongly and positively), what could be the likely scenarios?

A Causes B (A -> B)

B Causes A (B -> A)

C Causes A and B (C -> A, B)

Just a random coincidence or a fluke

20 students in a college trek got ill. All were given one Herb in some village. Everyone became well after 4 days.

What would you do in this scenario to know the effectiveness or ‘causal’ effect of Herb?

Experiment!

Divide them into 2 groups, give Herb to one, no Herb to another, and then check the result.

Control Groups & Treatment Groups

Concept of Randomization. Why Randomize?

Randomized Control Trial (RCT) Experiments

What if you want to know the effect of a vaccine on non-smokers and smokers?

Can we create Treatment & Control groups?

Experiments* – *We conduct experiments or tests to see the causal effect in data analysis.

Two types of Experiments.

- ***True Experimental Research***

- ***Quasi - Experimental Research***

True Experimental Research

- ***Experiments Type* – Like the experiments we do in lab.**

- ***Two variables***

- ***Control Group***

- ***Treatment Group***

Some other ways of doing experiment :

- *Before-after design without the control group*

- *Example of productivity of soil before and after the fertilizer.*

- *Flaw of this experiment : Other factors could have changed in between. Like temperature, rain etc..*

Quasi Experimental Research

- ***Like comparing old with young. Vegetarians with meat eaters.***

- ***Like older people have less lung capacity than younger people. So, is it because of age?***

- ***Flaw : Other factors could have caused the effect too.***

- ***Like old people meant more exposure to pollution. Generational gap leading to older people smoking earlier in their generations.***

Why is Causality difficult to know??

True Experiments are infeasible or expensive.

Quasi Experiments will have some bias one way or another.

Examples and Applications

- *In Marketing: Eg - Advertisement Campaigns*
- *In Policy: Eg - Introducing new rules in traffic*
- *In Medical Science and Medicines: Eg - New Medicine*

Application in Software/IT World

A/B Testing

How do you know if the change seen in two groups are significant?

Statistical Tests that we have done before!

By creating a Hypothesis with Null as the default!

Time for a story. On Causality.

Smoking Tobacco Causes Lung Cancer.

Do you support this statement? Or, do you believe this statement to be statistically true?

Great Statistical/Causality Debate of 1950s

There were people who believed one could not statistically prove Causality between Tobacco Smoking and Lung Cancer.

Including one person called Ronald Fisher.

The father of modern statistics believed one could not prove Causality between Tobacco Smoking and Lung Cancer.

Historical Background

Spike in number of Lung Cancer cases around 1940s.

Spike in the commercial use of Cigarette around same time.

Spike in use of automobiles, automobile smoke and roads tar.

Fisher and Others' arguments

Maybe early stage of cancer made people smoke instead.

Some gene related to both were causing both instead.

People who were disciplined enough to resist smoking were disciplined to be overall health conscious and avoid cancer.

Final Conclusion

We all know what is the conclusion now. Tobacco Smoking does cause Cancer.

Strong evidence of numerous experiments and experiments done with animals also supported the Causality hypothesis.

Identification of Carcinogenic Chemicals in Tobacco helped the conclusion too.

Take away from story

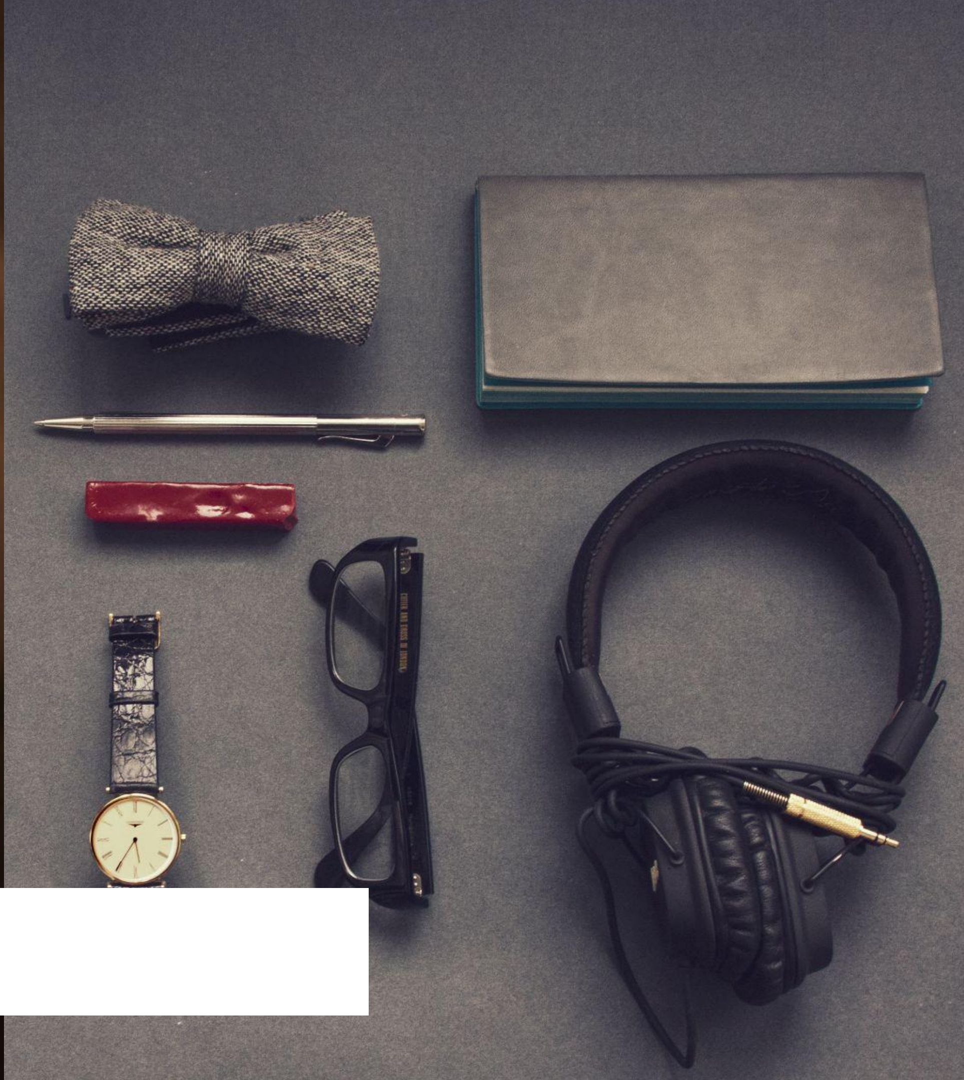
Sometimes it might be difficult to prove Causality from Data or Statistics alone.

Still, that does not mean Correlation is not important!

Domain Knowledge / Subject Matter Knowledge is very important.

Where to draw the line of Correlation vs Causation?

Do we even need to care about Causation always?



Thank You!