Cannlytics

Cannabis Data Science

Saturday Morning Statistics #13

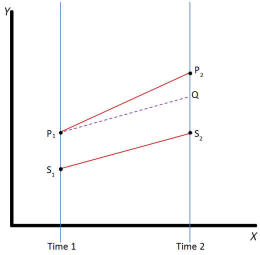
February 19th, 2022

1854–1858 - A map created by Dr. John Snow (1813-1858) depicting cholera cases in the London epidemics of 1854.

Fun fact: Dr. John Snow's research on cholera gave rise to the modern day difference-in-differences model[1].

Designing Difference in Difference Studies: Best Practices for Public Health Policy Research
Coady Wing et al., (2018).
https://doi.org/10.1146/annurev-publhealth-040617-013507

Difference–in–Differences Models



Difference–in–differences techniques can be used to estimate the effects from the causal variable given:

1. Panel data;
2. Certain groups are exposed to a causal variable and others are not.

Advantages of Difference–in–Differences Models

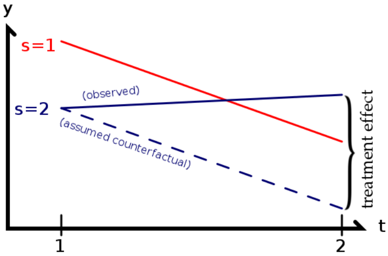
- Can be used to analyze economic conditions and government policy.
- Have been used in hundreds of studies.

Not suitable if:

- The composition of groups pre/post change are not stable.
- The intervention allocation is determined by the baseline outcome.

Disadvantages of Difference–in–Differences Models

- Requires that in the absence of treatment, the difference between the treatment and control group is constant over time.



  - No statistical test for this assumption;
  - Violation of this assumption will lead to biased estimation of the causal effect.

Question of the Day

- Given that we have rich panel data, can we utilize difference–in–differences models to answer any interesting questions? Perhaps...
  - What effect, if any, did the temporary closure of retail in Massachusetts in April of 2020 have on prices?
  - Any of your ideas?

The Classical Linear Regressions

A linear equation with n independent variables:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n + \epsilon$$

In matrix notation:

$$y = \beta^T X + \epsilon$$

With closed form solution:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

The Bayesian Approach to Linear Regressions

The data is assumed to be sampled from a distribution:

$$y \sim N(\beta^T X, \sigma^2 I)$$

The posterior probability is:

$$P(\beta|y, X) = \frac{P(y|\beta, X) \times P(\beta|X)}{P(y|X)}$$

Essentially:

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Normalization}}$$

🙏 Thank you for coming.

Lesson of the Day

- Assumptions matter.