Causality

Identifying causal relationship

Quick Review of Regression

Consider simple linear regression: $Y = a + bX$.

Y is your dependent variable and X is your independent variable.

We want to know how X affects the value of Y.

A Motivating Example

Suppose that you want to answer the following question: "Does eating ice-creams increase a person's weight?"

To answer this question, which regression should you run? What data should be collected?

The Next Problem

A key issue in economics is estimating the return to education.

More specifically, how much more money can you make by taking one extra year’s education?

How would you answer this question?

The Simple Idea

One simple idea is to collect data from different individuals. For example, suppose that we collect data from N individuals. For each one, we know his or her income (which we call Y) as well as his or her year of education (which we call X).

Then, we can run the regression $Y = a + bX$.

Here, if you take one extra year education, your income will increase by b. We are done!

The Simple Idea

Then, we can run the regression $Y = a + bX$.

Here, if you take one extra year education, your income will increase by b. We are done!

Any issues here? Discuss it with your classmates!

Omitted Variable Bias

More on Fixed Effects

Common types of fixed effects:

Year / Month fixed effects: 2021 and 2020 may be different.

Weekday/ Weekends fixed effects: Monday may be different from Sunday.

Individual fixed effects: Alice is different from Bob.

Location fixed effects: China may be different from Japan.

A Crowdfunding Example

Suppose that you want to investigate how the target of a crowdfunding campaign affects the total funding raised. Which types of fixed effects may come into play?

Individual fixed effects --- some people are more successful than others.

Location fixed effects --- The state can make a difference.

Subtype fixed effects --- Smartwatches may be different from software.

A Crowdfunding Example

Question: In the Kickstarter dataset, is it possible for you to control for these fixed effects? Why?

Individual fixed effects --- some people are more successful than others.

Location fixed effects --- The state can make a difference.

Subtype fixed effects --- Smartwatches may be different from software.

A Crowdfunding Example

Question: In the Kickstarter dataset, is it possible for you to control for these fixed effects? Why?

Individual fixed effects --- some people are more successful than others.

To be able to control for the individual fixed effect, we must have at least two observations for each entrepreneur, so we can calculate the difference. However, in the crowdfunding dataset, each entrepreneur typically only has one project.

A Crowdfunding Example

Now consider the following regression:

Here, we do not control for any fixed effects.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	3.43693	0.27031	12.71	<2e-16	***
LogTarget	0.30421	0.02726	11.16	<2e-16	***

A Crowdfunding Example

Now consider the following regression:

Here, we control for fixed effects.

```
result = lm(LogFunding ~ LogTarget +
factor(Subtype) + factor(Location), data = mydata)
summary(result)
```

A Crowdfunding Example

Now consider the following regression: Here, we control for fixed effects.

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.66186	0.40466	16.463	< 2e-16	***
LogTarget	0.20560	0.02403	8.556	< 2e-16	***
factor(Subtype)Apps	-4.19174	0.33504	-12.511	< 2e-16	***
factor(Subtype)CameraEquipment	1.30956	0.45434	2.882	0.003960	**
factor(Subtype)DIYElectronics	-0.20024	0.40840	-0.490	0.623940	
factor(Subtype)FabricationTools	-2.53178	0.56363	-4.492	7.17e-06	***
factor(Subtype)Flight	-1.50850	0.47699	-3.163	0.001571	**
factor(Subtype)Gadgets	-0.33102	0.34636	-0.956	0.339246	
factor(Subtype)Hardware	0.18633	0.33815	0.551	0.581627	
factor(Subtype)Makerspaces	-1.07444	0.56385	-1.906	0.056752	.
factor(Subtype)Robots	-0.10786	0.43235	-0.249	0.803008	
factor(Subtype)Software	-3.07865	0.34560	-8.908	< 2e-16	***
factor(Subtype)Sound	0.05697	0.42930	0.133	0.894436	
factor(Subtype)SpaceExploration	-0.88965	0.50991	-1.745	0.081076	.
factor(Subtype)Technology	-1.70381	0.33201	-5.132	2.95e-07	***
factor(Subtype)Wearables	0.19453	0.38017	0.512	0.608882	
factor(Subtype)Web	-4.43896	0.34505	-12.865	< 2e-16	***
factor(Location)IL	-0.96675	0.16151	-5.986	2.26e-09	***
factor(Location)MA	0.07324	0.15776	0.464	0.642465	
factor(Location)NY	-0.36623	0.11099	-3.300	0.000973	***
factor(Location)TX	-1.31772	0.11932	-11.043	< 2e-16	***
factor(Location)WA	-0.63908	0.16752	-3.815	0.000137	***

Difference-in-Difference Analysis

David Card
The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2021

Born: 1956, Guelph, Canada

Affiliation at the time of the award: University of California, Berkeley, CA, USA

Prize motivation: "for his empirical contributions to labour economics."

Prize share: 1/2

Diff-in-Diff

David Card and Alan Krueger wanted to investigate the effect of minimum wage on unemployment rate. They consider two states in the US: New Jersey (NJ) and Pennsylvania (PA).

Minimum Wage	March 1992	December 1992
New Jersey	4.05	5.05
Pennsylvania	4.05	4.05

Diff-in-Diff

Minimum Wage	March 1992	December 1992
New Jersey	4.05	5.05
Pennsylvania	4.05	4.05

Unemployment Rate	March 1992	December 1992
New Jersey	A	B
Pennsylvania	C	D

Diff-in-Diff

Note that unemployment rates are determined not only by the minimum wage, but also by the following factors:

The location: New Jersey may have better economy and lower unemployment rate

The time: The economy may be stronger in December 1992 so that unemployment rate is lower then.

Diff-in-Diff

Economics Model:

$$\mathrm{Unemployment}_{\mathrm{it}} = \mathrm{a} + \mathrm{State}_{\mathrm{i}} + \mathrm{Month}_{\mathrm{t}} + \mathrm{bWage}_{\mathrm{it}}$$

$$A = a + NJ + Mar + b \times 4.05$$
$$B = a + NJ + Dec + b \times 5.05$$
$$C = a + PA + Mar + b \times 4.05$$
$$D = a + PA + Dec + b \times 4.05$$

Diff-in-Diff

$$A = a + NJ + Mar + b \times 4.05$$
$$B = a + NJ + Dec + b \times 5.05$$
$$C = a + PA + Mar + b \times 4.05$$
$$D = a + PA + Dec + b \times 4.05$$

$$B - A = (Dec - Mar) + b \times 1.0$$
$$D - C = (Dec - Mar)$$

$$(B - A) - (D - C) = b \times 1.0$$

Diff-in-Diff

Example

We all know that smoking is bad for health, and quitting smoking is the best thing that a smoker can do.

Example

As data scientists, we want to estimate the effect of quit smoking. For instance, we want to answer the following question: Can a lifetime smoker reduce the chance of developing lung cancer by quitting smoking?

What would you do to answer this question?

Example

Consider another example. In US, people want to answer the following question: “Does the police reduce the crime rate?”

Here, our dependent variable Y is the crime rate, and our independent variable X is the size of the police force. By running regression, we find that Y increases with X.

We should defund the police!

OUR TASK

We want to know how the change of X affects the value of Y. We want to exlude other factors and reserved causality.

Two Common Causes

Omitted Variable

Reserve Causality

Two Common Solutions

Experiments

Instrumental Variable

Hypothetical Example

Which of the following schools should you join if you want to make more money in the future?

Hypothetical Example

Which of the following schools can help you make more money in the future?

Suppose that you analyze the data and find that the average salary for the graduates of the three schools are (HKD):

Harvard:	1M
HKU:	500K
Lanxiang (蓝翔):	200K

Can we say that Harvard is the best school that helps you make more money?

Hypothetical Example

Of course not.

Reason: Harvard/HKU is a famous school, and it has very admission standard. You must have a high IQ to be admitted to Harvard/HKU.

It is hard to say whether Harvard/HKU graduates are making more money because of their Harvard degree or because of their high IQ.

Hypothetical Example

We are more interested in the following question:

Holding all other things equal (e.g., IQ, family background, education background, gender, …) would a person make a higher salary if he/she chooses Harvard instead of HKU or Lanxiang?

American Business Schools Sorted by Average Starting Salary and Bonus

School	Average Starting Salary and Bonus	Percent Employed at Graduation	Average GMAT Score (full-time)	Acceptance Rate (full-time)
University of Pennsylvania (Wharton)	$159,815	82.3%	730	19.2%
Stanford Univeristy	$159,440	63.9%	737	5.7%
Harvard University	$158,049	78.9%	731	9.9%
University of Virginia (Darden)	$153,576	83.4%	713	24.5%
Dartmouth College (Tuck)	$152,805	80.2%	722	23.0%
Cornell University (Johnson)	$152,207	80.3%	700	29.9%
Columbia University	$151,849	69.9%	727	14.0%
University of Chicago (Booth)	$151,085	88.0%	730	23.5%
University of Michigan, Ann Arbor (Ross)	$150,052	89.7%	716	25.3%

Hypothetical Example

Note that now, students are randomly allocated to the three schools. In this case, which school you enter is independent of your personal information: The school you enter has nothing to do with your IQ, your family background, …

Harvard will get both high IQ and low IQ students. Similarly, Lanxiang will also get both high IQ and low IQ students. Their students' backgrounds are comparable.

Then, if their graduates make different salaries, it is the school that makes the difference and we can rule out all other factors!

One Solution

Of course, this example is more or less impractical…

You cannot force Harvard to admit a student, or force a student to join Lanxiang.

This type of experiment is also illegal…

But the idea is great! We can use it elsewhere.

One Solution

Similarly, we can do this for the smoking problem.

Recall that we want to know if quitting smoking reduces the chance of developing cancer.

One Solution

First, we recruit many smokers to do the experiment.

Second, we randomly assign them into two groups: a control group that continues smoking, and a treatment group.

Individuals in the treatment group are forced to quit smoking.

Later, we examine whether each individual develops cancer in each group.

Pharmaceutical

In the pharmaceutical industry, researchers need to prove that a new drug (X) is effective in curing or alleviating a certain disease (Y). To establish a credible causal relationship, they use "randomized controlled trials (RCTs)"

Pharmaceutical

- We randomly assign participants into two groups, a treatment group and a control group.
- The treatment group takes the medication (X=1) while the control group takes nothing (X=0).
- Finally, we calculate the value of Y (seriousness of disease) for each participant in the study.

AB Testing

The most commonly used experimental method

“Whenever we launch a new APP, we run many AB tests in different platforms to find a best name for it. Even though you are 99.9% sure one name is the best, it does not hurt running an AB test.”

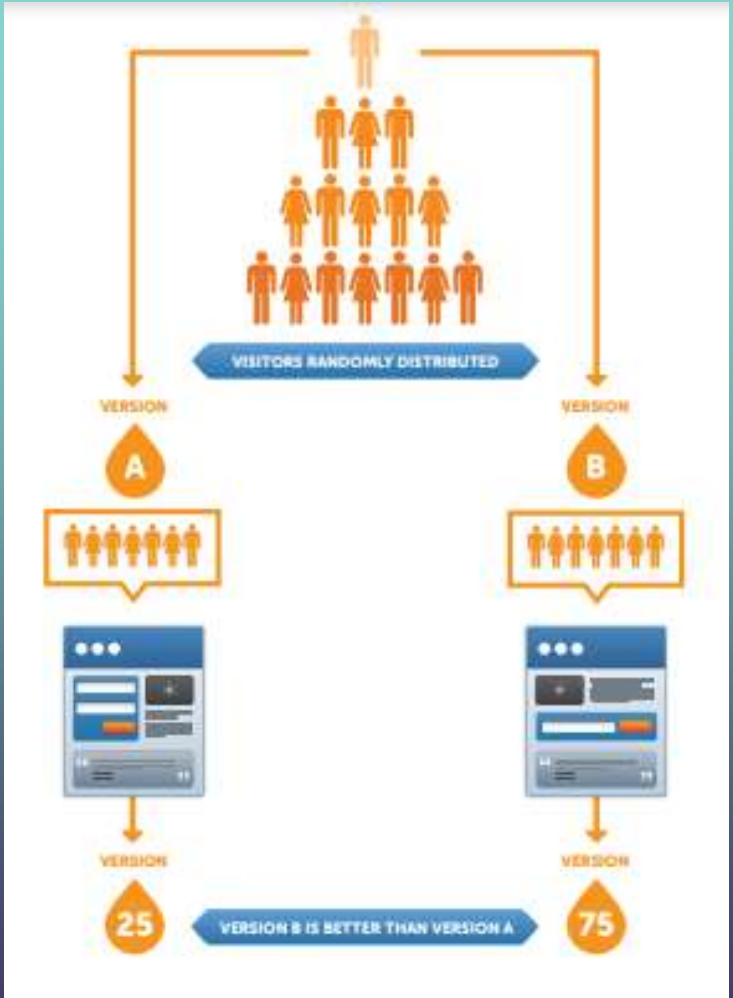
---Zhang Yiming, ByteDance CEO
(字节跳动张一鸣)

AB Testing

- AB testing was inspired by RCT used in pharmaceutical industry; it is commonly used in the Internet age.

- It is commonly used for Website and APPs.

    - Which website design is better?
    - Which APP design is better?
    - Which image is more attractive?


VISITORS RANDOMLY DISTRIBUTED
VERSION
A
VERSION
B
VERSION
25
VERSION B IS BETTER THAN VERSION A
VERSION
75

Experiment

Unfortunately, we are not always able to run our experiments or AB testing. For example, experiments can be illegal, too costly, too time consuming, or simply impossible.

For example, suppose that you want to understand how smoking affect one’s health; however, you cannot run an AB test forcing some people to smoke/not to smoke. So, how to obtain the results?

Two Common Solutions

Experiments

Instrumental Variable

Course Overview

2021 Nobel Prize in Economics


Joshua D. Angrist and Guido W. Imbens

"for their methodological contributions to the analysis of causal relationships"

Instrumental Variable

When running an experiment is impossible, we may also consider the instrumental variable approach.

Idea: Find a new variable that affects your X but does not affect your Y directly.

Theory (Optional)

Suppose that you want to estimate how X affects the value of Y. Mathematically, suppose that when X increases by 1, Y will increase by b. We want to find out the value of b.

You find a variable Z that affects X but does not affect Y directly.

People have proved that

$$b = \frac{Cov(Y,Z)}{Cov(X,Z)}$$

where cov stands for covariance.

A FACT

Someone may be telling you that he/she is an economist. However, what he/she does everyday is simply looking for instrumental variables.

Scientists that are famous for finding instruments

Daron Acemoglu

Professor, MIT

Joshua Angrist

Professor, MIT

Steve Levitt

Professor, UChicago

James Snyder

Professor, Harvard

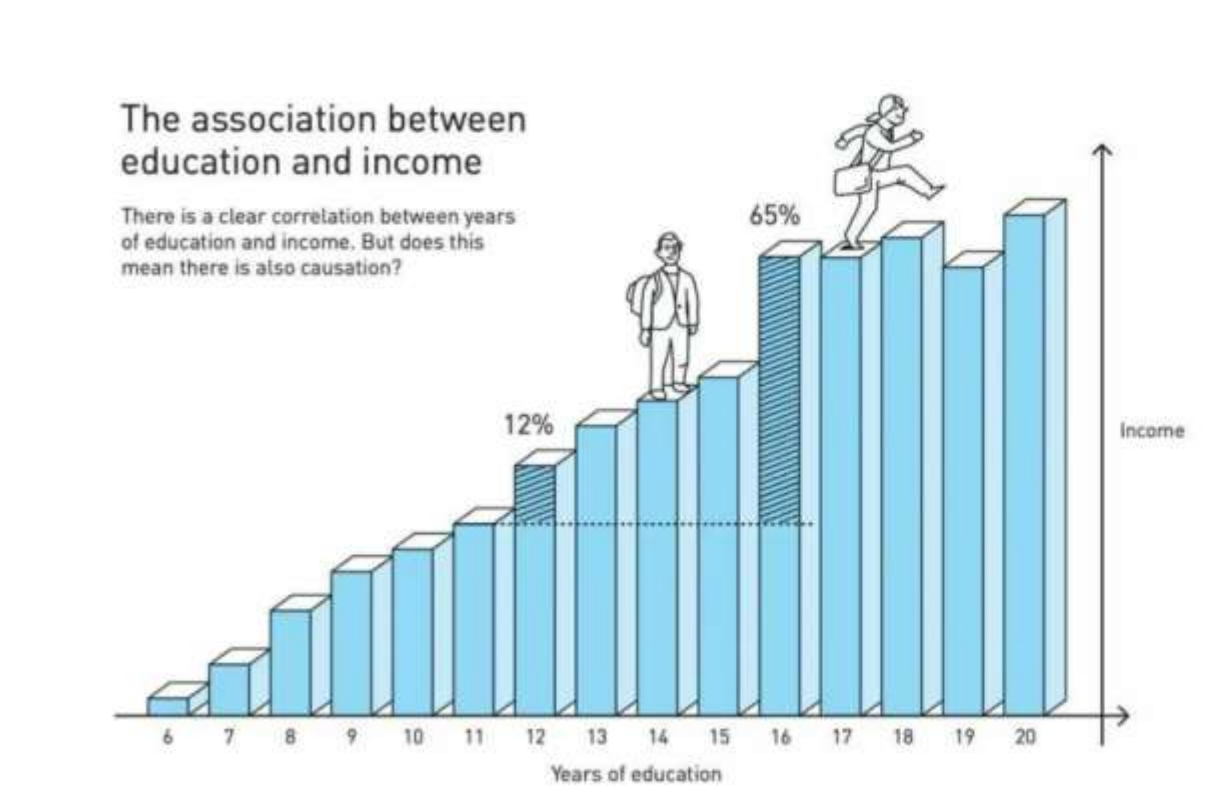
One of the most classic research!

DOES COMPULSORY SCHOOL ATTENDANCE AFFECT SCHOOLING AND EARNINGS?*

JOSHUA D. ANGRIST AND ALAN B. KRUEGER

We establish that season of birth is related to educational attainment because of school start age policy and compulsory school attendance laws. Individuals born in the beginning of the year start school at an older age, and can therefore drop out after completing less schooling than individuals born near the end of the year. Roughly 25 percent of potential dropouts remain in school because of compulsory schooling laws. We estimate the impact of compulsory schooling on earnings by using quarter of birth as an instrument for education. The instrumental variables estimate of the return to education is close to the ordinary least squares estimate, suggesting that there is little bias in conventional estimates.

One of the most classic research!



Any Other Instrumental Variable?

Econometrica, Vol. 82, No. 1 (January, 2014), 137–196

THE ELITE ILLUSION: ACHIEVEMENT EFFECTS AT BOSTON AND NEW YORK EXAM SCHOOLS

BY ATILA ABDULKADIROĞLU, JOSHUA ANGRIST, AND PARAG PATHAK[1]

Parents gauge school quality in part by the level of student achievement and a school's racial and socioeconomic mix. The importance of school characteristics in the housing market can be seen in the jump in house prices at school district boundaries where peer characteristics change. The question of whether schools with more attractive peers are really better in a value-added sense remains open, however. This paper uses a fuzzy regression-discontinuity design to evaluate the causal effects of peer characteristics. Our design exploits admissions cutoffs at Boston and New York City's heavily over-subscribed exam schools. Successful applicants near admissions cutoffs for the least selective of these schools move from schools with scores near the bottom of the state SAT score distribution to schools with scores near the median. Successful applicants near admissions cutoffs for the most selective of these schools move from above-average schools to schools with students whose scores fall in the extreme upper tail. Exam school students can also expect to study with fewer nonwhite classmates than unsuccessful applicants. Our estimates suggest that the marked changes in peer characteristics at exam school admissions cutoffs have little causal effect on test scores or college quality.

Comment on Machine Learning

Nowadays, machine learning is very powerful in making predictions. That is, given X, machine learning algorithms predict the value of Y.

Many algorithms such as neural networks, naïve Bayesian, matrix factorization etc.

Basically, machine learning algorithms generate a function f such that $Y \approx f(X)$.

Comment on Machine Learning

There is a frequent challenge on machine learning:

“Machine Learning-based projects focus on predicting outcomes rather than understanding causality.”

For this reason, machine learning algorithms are often called “black box” --- we know it works, but we don’t know how it works (只知其然，不知其所以然).

Why should we care about causality?

In an e-commerce context, we could determine which specific factor impacts the decision to purchase a product. With this information, we could better allocate resources to improve a specific KPI. We could also rank the impact of different factors on the purchasing decision. We could determine if a given customer would have purchased a specific product if he/she had not bought other products for the last two years.

Why should we care about causality?

In the agricultural field, we often try to predict if a farmer's crop yield will be lower this year. However, using casual inference, it will become to better understand what steps should we take to increase the harvest.