因果推断

Peng Ding's Homepage

Home Research Papers Teaching Resources

丁鹏

I am an Associate Professor in the Department of Statistics, UC Berkeley. I obtained my Ph.D. from the Department of Statistics, Harvard University in May 2015, and worked as a postdoctoral researcher in the Department of Epidemiology, Harvard T. H. Chan School of Public Health until December 2015. Previously, I received my B.S. (Mathematics), B.A. (Economics), and M.S. (Statistics) from Peking University.

Research

My research focuses on causal inference including:

1. Design and analysis of randomized experiments: randomization tests, covariate adjustment, rerandomization, clustered experiments
2. Observational studies: sensitivity analysis for unmeasured confounding, overlap of covariates, integrating multiple data sources
3. Natural experiments: instrumental variable, difference-in-difference, regression discontinuity
4. Causal mechanisms: treatment effect variation, principal stratification, mediation, interference

I am also fascinated by the following directions:

1. Survey sampling
2. Missing data: identification and estimation with data missing not at random
3. Measurement error
4. Intersection of Frequentist and Bayesian statistics: Professor Carl Morris called it FB (Morris' football)
5. Applied statistics in social sciences and biometrical studies

Peng Ding

A First Course in Causal Inference

CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY

首页 | 关于 | 论坛 | R会 | 投稿 | 搜索

← 返回首页 2021-06-27 编辑本页 →

推荐文章

因果推断——现代统计的思想飞跃

丁鹏

关键词：因果推断; 统计

编辑：雷博文

转载自《数学文化》2021 / 第 12 卷第 2 期

causality

因果推断何以成为可能？

David Hume

An Enquiry Concerning Human Understanding.

David Hume.

Sect. IV. Sceptical Doubts concerning the Operations of the Understanding

PART I.

math, maybe also logic

20. All the objects of human reason or enquiry may naturally be divided into two kinds, to wit, Relations of Ideas, and Matters of Fact. Of the first kind are the sciences of Geometry, Algebra, and Arithmetic; and in short, every affirmation which is either intuitively or demonstratively certain. That the square of the hypothenuse is equal to the square of the two sides, is a proposition which expresses a relation between these figures. That

22. All reasonings concerning matter of fact seem to be founded on the realtion of Cause and Effect. By means of that relation alone we can go beyond the evidence of our memory and senses. If you were to ask a man, why he believes any matter of fact, which is absent; for instance, that his friend is in the country, or in France; he would give you a reason; and this reason would be some other fact; as a letter received from him, or the knowledge of his former resolutions and promises. A man finding a watch or any other machine in a desert

I shall venture to affirm, as a general proposition, which admits of no exception, that the knowledge of this relation is not, in any instance, attained by reasonings a priori; but arises entirely from experience, when we find that any particular objects are constantly conjoined with each other. Let an object be presented to a man of ever so strong natural reason and abilities; if that object be entirely new to him, he will not be able, by the most accurate examination of its sensible qualities, to discover any of its causes or effects. Adam, though his rational faculties be supposed, at the very first, entirely perfect, could not have inferred from the fluidity and transparency of water that it would suffocate him, or from the light and warmth of fire that it would consume him. No object ever discovers, by the qualities which appear to the senses, either the causes which produced it, or the effects which will arise from it; nor can our reason, unassisted by experience, ever draw any inference concerning real existence and matter of fact.

I shall content myself, in this section, with an easy task, and shall pretend only to give a negative answer to the question here proposed. I say then, that, even after we have experience of the operations of cause and effect, our conclusions from that experience are not founded on reasoning, or any process of the understanding. This answer we must endeavour both to explain and to defend.

R: lm(y ~ x)

$\uparrow$ $\hat{\beta}$?

lm(x ~ y)

$\uparrow$ $\hat{\gamma}$?

相关关系,

Pearson : $\rho_{XY} = \frac{cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$

lm(y ~ x)

lm(y ~ x + w)

(Yule-)Simpson Paradox

$$\begin{cases} X \overset{+}{\sim} Y \\ X \overset{-}{\sim} Y / Z \end{cases}$$

$$\rho_{XY} > 0$$

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} \cdot \\ \cdot \\ \cdot \end{pmatrix}, \begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} \right)$$

$$\Downarrow$$

$$\rho_{XY|Z}$$

$$\rho_{XY|Z} : \begin{pmatrix} X \\ Y \end{pmatrix} \Big| Z \sim \text{Normal}\left(\begin{pmatrix} \cdot \\ \cdot \end{pmatrix}, \begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix} \right)$$

例子: $\rho_{XY} > 0$, $\rho_{XY|Z} < 0$

potential outcomes

潜在结果

hypothetical experiment

假想实验

$i = 1, \ldots, n$

treatment 处理   treatment 1 / 0 control

intervention 干预

outcome 结果 Y_i

potential outcomes: $Y_i(1)$, $Y_i(0)$

定义因果

Neyman (1923) 论文

Neyman (1935 JRSS)

Rubin (1974)

名字: Neyman model
Rubin Causal Model
Neyman-Rubin model

$$\tau_i = Y_i(1) - Y_i(0)$$

无法同时观测

$$\tau = \frac{1}{n}\sum_{i=1}^{n} \tau_i$$

平均因果作用

$$= \frac{1}{n}\sum_{i=1}^{n} Y_i(1) - \frac{1}{n}\sum_{i=1}^{n} Y_i(0)$$

① 个体无干扰 no interference

$Y_i(z) \Rightarrow Y_i(z_1\ z_2 \cdots z_n)$

treatment of unit i

SUTVA
Stable Unit Treatment
Value Assumption

violated ⇒ spillover effect

② 一致性 consistency

$Y_i(1),\ Y_i(0)$

只有一个版本

例 1 Drug / No

例 2 Gym / No

例 3 College / No

例 4 Smoking → Cancer

例 5 BMI → health

Judea Pearl

BMI ← exercise

BMI ← 抽脂

BMI ← diet

Experimentalist's view

$$\left(Y_i(1),\quad Y_i(0) \right)$$

Z 连续 : $Y_i(Z)$

$$\left(Y_i(1),\ Y_i(0)\right)_{i=1}^{n}$$

	$Y_i(1)$	$Y_i(0)$
$i=1$	✓	?
2	✓	
	✓	
⋮	✓	
	✓	✓
		✓
n	?	✓
		✓

缺失数据

$$Y_i = \begin{cases} Y_i(1) & \text{if } Z_i = 1 \\ Y_i(0) & \text{if } Z_i = 0 \end{cases}$$

$$Y_i^{mis} = \begin{cases} Y_i(0) & \text{if } Z_i = 1 \\ Y_i(1) & \text{if } Z_i = 0 \end{cases}$$

the fundamental problem of causal inference

Fisher 随机化

- 随机 $Z_i = 1$
- 随机 $Z_i = 0$

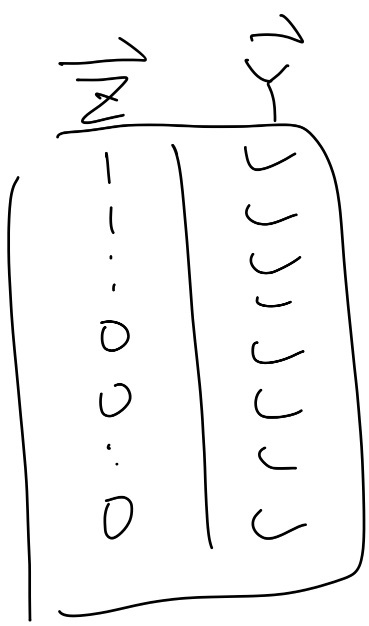
— 中：党项 人参

— 英：Lind 柠檬汁

Fisher: randomization as a "reasoned" basis

— comparable treatment & control

— 保证：随机化检验：in Fisher Randomization Test (FRT)
合理性



实验：$\vec{Z}$ $\begin{cases} n_1 & 1 \\ n_0 & 0 \end{cases}$

random permutation

随机置换

分析：$t(\vec{Z}, \vec{Y})$

$\vec{Z}$：随机　$\vec{Y}$：固定

$$p_{FRT} = \frac{1}{M} \sum_{r=1}^{M} \mathbb{1}\left(t(\vec{Z}^r, \vec{Y}) \geq t(\vec{Z}, \vec{Y}) \right)$$

$$M = \binom{n}{n_1}$$

Fisher:

Design of Experiments (1935)

Fisherian Principles

- randomization
- blocking
- replication