

# 第二讲 实验设计基础

---

孟天广 副教授

清华大学政治学系

2021年3月1日

# 内容提要

---

- 实验设计与因果推论
- 实验设计的要素与程序
- 实验设计的效度：表现与克服
- 常用实验设计的类型

# 实验设计与因果推论

---

- 实验是从假设（猜想）出发，人为创造或控制一些因素，使因变量随自变量的变化而变化，从而揭示变量或现象间的因果关系。
- 实验设计的原理
- 实验设计的基本要素
- 实验的程序
- 伦理要求

# 因果关系： $X \longrightarrow Y$

---

- ◆ 因果关系四个标准（Lazarsfeld, 1959）
- ✓ 两个变量间的因果关系必须有时间顺序，原因先于结果；
- ✓ 两个变量间有可观察的相关性；
- ✓ 最重要的是，可观察的相关性不能被第三个变量解释
- ✓ 可以被理论（逻辑）所辩护/解释

# 因果关系： $X \longrightarrow Y$

---

- ◆ 相关关系VS因果关系
- ◆ 当X与Y存在相关关系时，至少存在三种可能性：
  - ✓ 虚假相关：X和Y同时受第三个变量C影响，当控制C后，X和Y的相关关系不存在；
  - ✓ X引起Y：当控制其他变量后，X和Y仍然有相关关系；
  - ✓ Y引起X：当控制其他变量后，X和Y仍然有相关关系；
- ◆ 情况（2）和（3）的因果关系方向不明确
- ◆ *Neyman-Rubin counterfactual framework of causality*

# 实验方法的原理（优势）

---

- 内在效度 (Internal Validity)
- 可控条件 (Controlled environment)
- 随机化分配 (Randomization): 可比性 (Comparable) + 比较 (Comparison)
- 可复制性 (Replication)
- ✓ 因果关系 (机制)
- ✓ 研究假设明晰而简单
- ✓ 伦理限制

# 实验设计：案例

---

- 如何设计实验检验最低生活保障政策对幸福感的影响？
- 集体行动与政府回应的关系？
- 民主制度对治理绩效的影响？
- 政体类型与跨国冲突的关系？
- 基因决定政治偏好？

# 实验设计的基本要素(1): 变量

---

- 自变量（干预）：由实验人员选择、操纵的变量
- 因变量（结果）：衡量自变量对实验单位的影响的变量。
- 外生变量（控制）：自变量以外的影响被试反应的所有其他变量。干扰效应



# 实验设计的基本要素(2): 操作

---

- 实验单位（被试）：是指对自变量或处理的反应被作为实验对象的个人、组织或其他实体
- 前测与后测
- 实验组与控制组
- 可复制

# 实验设计的程序

---

- 提出研究假设
- 确定实验组（干预）和控制组（未干预）
- 前测（干预前测量）
- 随机分配及引入干预（自变量）
- 后测（测量因变量）
- 检验假设： within subject design vs between subject design

# 自变量+因变量

---

- 自变量（干预）：研究者可以操作、引入或改变的干预条件
- ✓ 信息指导、情景设计、物理环境或社会环境；
- 因变量（结果）：研究者关心的社会后果或效应
- ✓ 行为、态度、偏好、决策、交易等
  
- ✓ 凡是在实验中能够干扰自变量在结果反应上的变化的所有因素都属于干扰变量，譬如社会特征、实验环境或社会演变

# 比较+复制思维

---

- 最好的实验是比较（Good experiments are comparative）；
- 比较实验组与（同时发生）控制组（而不是历史控制组）
- 可复制性(Replication): 降低不可控因素的效应（增加精度precision）；量化不确定性程度（uncertainty）

# 随机化 (Randomization)

---

- 建立干预组与控制组的可比性(Comparable)
- ✓ 可观察confounders vs 不可观察confounders
- 避免（可能的）偏差(Bias)
- 控制偶然性(role of chance)

# 实验设计的类型

---

- 单因素实验设计
- 多因素实验设计

# 单因素实验设计

- 单个自变量（干预）
- 单一对照法
- 单一对照法包括后测法与前后测法。
- ❑ 后测法：只对引入自变量（干预）后的结果变量进行测量

组别	处理	后测值
1	√	$y_1$
2	×	$y_2$

# 单因素实验设计

- **前一后测法**。在引入自变量（干预）之前对被试进行测量，并比较前后测的结果变量。前一后测法又可分为自身前一后测与两组前一后测。
- **自身前一后测**是指被试引入单一自变量前与后均进行测量，并比较所测的结果。

前测	处理	后测
$\sqrt{y_1}$	$\sqrt{\quad}$	$\sqrt{y_2}$



# 单因素实验设计

- 两组前—后测法是较为普遍的实验法，具体方法是使用随机分派或配对法创设一个实验组、一个控制组，实验组引入自变量（干预），并进行前后测比较。
- 用随机法或配对法产生实验组和控制组
- 两组均进行前后测。实验组引入自变量（干预），控制组不引入自变量。

组别	前测	处理	后测
实验组	$\sqrt{y_1}$	$\sqrt{\phantom{x}}$	$\sqrt{y_2}$
控制组	$\sqrt{y_1'}$	$\times$	$\sqrt{y_2'}$

# 多因素实验设计

- 在一个实验中引入两个或两个以上的自变量（干预）的实验设计。
  - 根据自变量的数目和每个自变量的水平（干预类型），分为两因素和三因素实验设计等。
- 两因素实验设计
  - 一个实验中引入两个自变量的实验设计。
  - 常用的两因素实验设计包括随机 $2 \times 2$ 因素实验设计、重复测量 $2 \times 2$ 因素实验设计、随机区组 $2 \times 2$ 因素实验设计、 $2 \times 2$ 因素混合实验设计等。
  - 随机 $2 \times 2$ 因素实验设计最常用。

- 随机 $2 \times 2$ 因素实验设计中的“随机”指的是根据随机原则把被试分派到各个处理组中
- “ $2 \times 2$ ”分别指两个自变量呈现两个水平变化
- 具体自变量组合参见下表：

		自变量 B	
		$b_1$	$b_2$
自变量A	$a_1$	$a_2b_1$	$a_1b_2$
	$a_2$	$a_1b_1$	$a_2b_2$

# 析因设计(Factorial design)

---

- 多个自变量对因变量的同时效应
- 多个干预（或干预水平）的结合
- 主效应(Main effects): 单个自变量的效应
- 交互效应(Interaction effects): 两个或三个自变量的结合之效应

# 实验举例

- 政治资源与集体行动对政府回应的影响
- 两个自变量（干预）+两个干预水平
- 具体自变量组合参见下表：

		B（集体行动）	
		b <sub>1</sub> (有)	b <sub>2</sub> (无)
A (政治资源)	a <sub>1</sub> (有)	a <sub>1</sub> b <sub>1</sub>	a <sub>1</sub> b <sub>2</sub>
	a <sub>2</sub> (无)	a <sub>2</sub> b <sub>1</sub>	a <sub>2</sub> b <sub>2</sub>

# 实验操作

---

- 研究人员分析自变量分为a1b1、a2b1、a1b2、a2b2时政府回应的程度。
- **被试** 政府官员
- **程序** 随机把被试分为4个小组，引入4个自变量，观察、记录因变量的结果。
- **假设** 依据理论假设推论出工作假设，按照作用大小预测干预效果（结果变量变化）的顺序： $a1b1 > a1b2 > a2b1 > a2b2$ 。
- **结果** ? ? ?

## □ 三因素实验设计

- 一个实验中引入三个自变量的实验设计。
- 按每个自变量的水平数目不同，分为随机三因素、重复测量三因素等实验设计。
- 随机 $2 \times 2 \times 2$ 因素实验设计

(自变量) $a_1$				(自变量) $a_2$			
(水平) $b_1$		(水平) $b_2$		(水平) $b_1$		(水平) $b_2$	
(水平) $c_1$	(水平) $c_2$	(水平) $c_1$	(水平) $c_2$	(水平) $c_1$	(水平) $c_2$	(水平) $c_1$	(水平) $c_2$

# 实验效度：表现及克服

---

- 实验效度是指根据实验设计进行因果关系研究所得出结论的有效程度。
- 内部效度*internal validity*
  - 对自变量或干预的操作是否确实导致了因变量的变化。
  - 自变量是否观测到之因变量变化的唯一原因。
  - 它是实验研究的最低要求
  - 解决内部效度的必要条件：控制外生变量



# 实验的效度

---

- 外部效度 external validity
  - 是指实验中所发现的因果关系是否具有普遍意义
  - 实验结果适用于真实世界的程度
  - 产生问题的原因：实验环境的处理
- 实验需要在内部效度与外部效度之间权衡
- 增加效度会增加成本，在效度和成本之间权衡

# 实验效度的威胁

---

- ❑ 历史误差，发生在实验开始和结束之间，不受控制的、可影响因变量的特殊外部事件。
- ❑ 成熟：生理条件误差，受试者在心理或生理方面变得成熟，从而会影响到实验结果
- ❑ 工具化：工具误差，指实验期间由于测量工具或测量程序的改变而对实验结果产生的影响。
- ❑ 选择误差：由于处理组和控制组在实验前就存在差异而导致的实验结果失真。
- ❑ 损耗，也称流失误差，在实验过程中实验单位的流失对实验结果的影响。
- ❑ 回归：指具有极端行为的被试在实验过程中向着行为的均值发展的趋势。

# 实验效度的威胁

---

- 测试效应：由实验过程引起。前期测量误差，指前面的观察对后来观察的影响；相互作用误差，是一次先前的测量影响到实验单位对自变量的反应。
- 反应误差：因实验环境或实验组织者的行为对因变量产生影响而产生。
- 测量时间：由于测量结果的时间不当引起的实验误差。
- 替代情境：由于实验情景与实际情景不同而引起的。

# 应对效度威胁

---

- 保证实验效度的关键是控制外生变量
- 实验效度的提高需要控制较多的外生变量，因而需要较大成本
- 实验不可能消除全部误差，只能在给定条件下接受一个可能的误差程度。

# 控制外生变量的方法

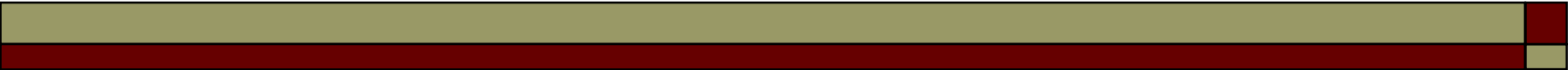
---

- 随机化：通过使用随机数字将实验单位分派到不同实验组。处理条件也被随机分派到不同实验组。
- 匹配：将实验单位分派到各处理条件之前，就一系列关键背景变量对它们进行比较。匹配方法的缺点：只能就几个特性进行匹配
- 统计控制：测量外生变量并通过统计分析来对其影响加以调整。
- 设计控制：使用特殊的实验方法控制特定的外生变量。  
双胞胎实验

# 实验设计： 随机抽样vs随机分配

---

- 随机化分配（Random assignment）： 保证被试者以无偏化方式被随机地分配到不同干预组
- 随机抽样（Random sampling）： 从大总体 (population) 中随机抽取小样本(sample)， 以反映总体特征
- 理想设计： 随机抽样+随机分配； 随机分配
- 随机化分配： 计算机辅助； 硬币/卡片；



# 实验设计：干预分配

---

- 完全随机设计
- 随机区组设计
- 组内设计
- 组间设计

# 实验类型

---

- 经典实验设计(True experiment)
- Pre-experiment: 缺乏经典实验设计部分要素的实验设计
- Quasi-experiment: 接近经典实验设计, 但缺乏部分要素
  - 单组后测设计(One-Shot Case Study Design)
  - 单组前后测设计(One-Group Pretest-Posttest Design)
  - 静态组间比较(Static Group Comparison)
  - 中断时间序列(Interrupted Time Series)



# 实验研究的伦理要求

---

## □ IRB

- 禁止伤害原则
- 欺骗：dishonesty (NO!); Deception在特定条件下可以接受；
- Debrief
- Blinding

# 常用实验设计(1): Lab Exp

---

- 在实验室内进行，干预变量与环境受到严格控制，实验结果可以清楚观测的实验。
- 米尔格莱姆经典的权威一服从实验，就是社会科学领域经典的实验室实验。
- ✓ 优势：更好控制；更精准测量；更丰富的实验“世界”
- ✓ 局限：人造环境、社会动物、情景异质性
- ✓ 应用：心理学、行为经济学、博弈论等

# 常用实验设计(2): Field Exp

---

- 现场实验是在现实生活中进行的实验研究。
- 现场实验的被试行为往往是不知情自己是被研究对象情况下的自然反应，因而较能真切反映人在真实环境中的行为与心理变化。
- 如营商环境的实验
  - ✓ 优势：现实场景；真实行为；外部效度
  - ✓ 局限：伦理挑战；实验控制；现实效应？
  - ✓ 应用：竞选与投票、社会交易、政府行为、制度变革等
  - ✓ 干预：操作信息接收；信息干预的伦理困境更少

# 常用实验设计(3): Survey Exp

---

- 调查实验是在抽样调查中嵌入实验设计，跨对象系统地改变调查的一个或多个要素（改变），对于一个或多个结果变量的作用的实验设计。
- 优势：Field experiment + Lab experiment；外部效度；
- 局限：缺乏控制；外溢（污染）效应
- 应用：行为方式、价值偏好、社会互动、制度变革、组织等；

# 常用实验设计(4): Natural Exp

---

- 自然实验并非人为设计、计划的实验，而是事后控制组比较
- 是现场实验的一种，具备实验的部分特征
- 干预通常为偶发事件(Event)，利用实验逻辑进行比较分析
- 优势：Field experiment；自然干预
- 局限：Field experiment；缺乏控制
- 应用：自然灾害、政策实验等

# 小结：什么是好的实验设计？

---

- 无偏(Unbiased)
  - Randomization
  - Blinding
- 精度(High precision)
  - Uniform measure
  - Replication
  - Blocking
- 简化(Simple)
  - avoid mistakes
- 适用性(Wide Applicability)
  - Deliberate variation
  - Factorial designs
- 不确定性(Estimate uncertainty)
  - Replication
  - Randomization
  - Statistical power

# Q&A

---