

高级计量经济学

理论经济学博士课程 2023-2024

Lecture 0: Introduction

黄嘉平

工学博士 经济学博士
深圳大学中国经济特区研究中心 讲师

办公室 粤海校区汇文楼1510
E-mail huangjp@szu.edu.cn
Website <https://huangjp.com>

课程网站

- 直达链接：

<https://huangjp.com/teaching/deconometrics.html>

- 首页链接及路径（个人网站）：

<https://huangjp.com/>

首页 > Teaching > Graduate > 高级计量经济学（理论经济学博士）

课程简介

课程目标

- 课程目标：深入学习计量经济学**核心理论**和**前沿方法**
 - 必要的背景知识储备：
微积分、线性代数、统计学
尤其是**矩阵运算**和**条件概率**非常重要
 - 需要了解的初级计量经济学内容：
线性回归模型、最小二乘估计、假设检验、非线性回归函数、
虚拟变量的用法、内生性等内容
- 需要自主学习的内容：
 - 数据的处理、计量经济软件的使用（任何软件都可以，推荐使用 R）
 - 文献阅读

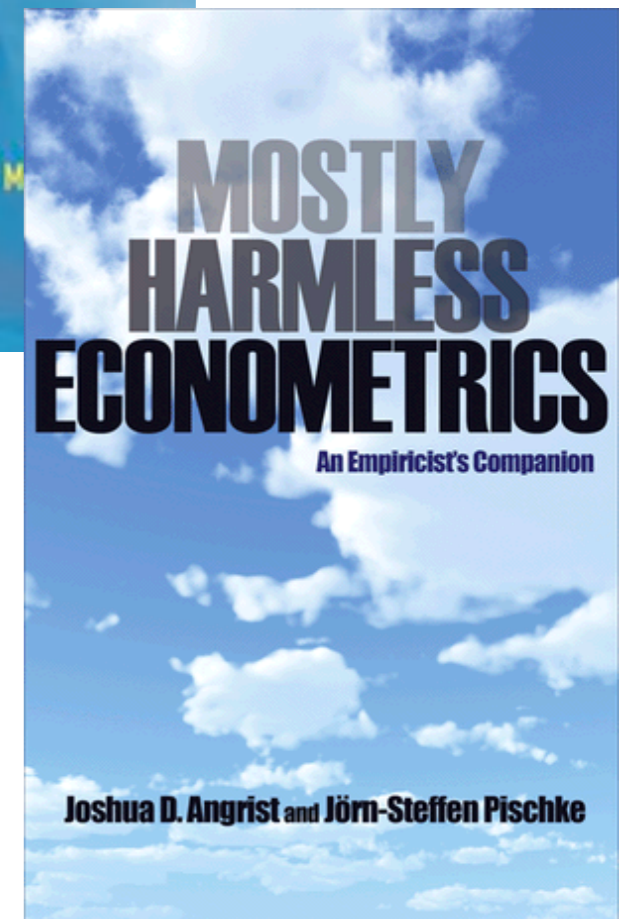
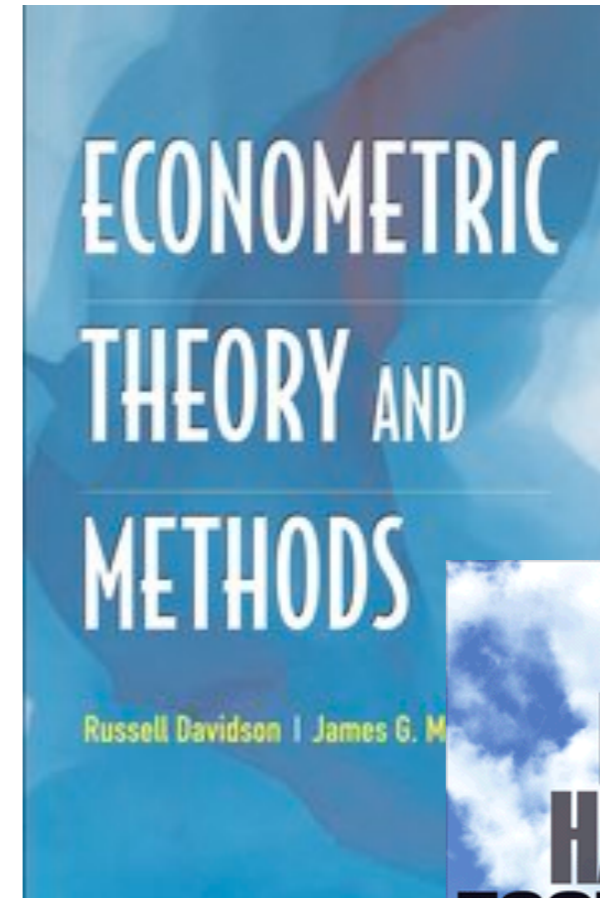
课程内容

- 第一学期：线性回归模型及其扩展
 - 计量经济学基础：概率、统计、线性代数
 - 线性回归模型：最小二乘法、FWL定理、Gauss-Markov定理、假设检验
 - 扩展：异方差性、重采样方法、非线性回归、IV、回归方程组
- 第二学期：其他回归模型和方法、因果推断
 - 其他回归模型和方法：面板数据模型、最大似然估计、离散因变量模型、非参数估计
 - 因果推断：Neyman-Rubin潜在结果模型、匹配、双重差分、合成控制、IV、断点回归设计

课程教材

主要参考以下书籍、同时借鉴其他书籍和论文

- Davidson & MacKinnon (2021), *Econometric Theory and Methods*, Oxford University Press.
可以从作者网站下载
<http://qed.econ.queensu.ca/ETM/>
- Angrist & Pischke (2008), *Mostly Harmless Econometrics*, Princeton University Press.



成绩评价

- 评价方法：课堂表现 20% + 期中考核 40% + 期末考核 40%
 - 课堂表现：包括出席率、课堂参与度、文献汇报等
 - 期中考核：第一学期末进行，拟采取随堂考试的形式
 - 期末考核：第二学期末进行，拟采取课程论文的形式
- 学校关于夸学期课程的处理方法
 - 本学期结束后，研究生院会将课程注销，期中考核成绩由任课教师保存
 - 第二学期，研究生院会重新开设课程（课程代码相同），学生需重新选课，期末考核后录入两学期的综合成绩

深圳大学研究生课程教学与
成绩考核管理办法（修订）
— 深大〔2017〕196号
第二十八条 研究生无故缺
课超总学时 1/3 的，取消考
试资格，该门课程成绩按零
分记；无故缺课超总学时
1/6 的，由授课教师视情节
和本人态度决定是否准予参
加考试。

如何更好地学习计量经济学

- 计量经济学要求学生同时具备**数学推导能力**和**直观理解能力**
 - 经过以往的学习和研究，多数同学都掌握了直观理解能力
 - 本门课程的重点是培养数学推导能力
- 那有的同学可能就要问了，我不是主攻计量经济理论，在学习中跳过数学推导不行吗？
 - 你可能会会有很多论文读不懂
 - 你将无法确认自己的直观理解是否正确
 - 你会无法分辨自己对计量方法的运用是否正确
- 怎么掌握数学推导能力？ → 参与到课堂中来
- 如果课程内容我已经学过了，能不能不听课做其他的事情？ → 强烈不建议

课程内容综述

计量经济学 = 计量方法 × 经济学问题

- 计量方法：运用**数学模型**和**数据**探索经济变量间的关系
 - 数学模型：对经济现象的抽象总结
 - 数据：随机变量的样本，因此需要探究隐藏在数据面纱下的真实关系
- 计量经济学的两个侧面：解释、预测
 - 试图**解释**变量间的因果关系（causal inference）：自变量的变化如何引起（cause）因变量的变化
 - 在给定自变量的取值时**预测**因变量的值

从经济学模型到计量经济模型

以Gary Becker的犯罪模型为例

- 犯罪的经济学模型

- 犯罪行为作为个人选择的结果，其中包含理性决策和非理性行为。传统经济学的核心是对理性选择的研究，因此可以被用来分析犯罪行为。
- 诺奖得主 Gary Becker 将针对犯罪行为对社会带来的损失进行度量，并探讨了使损失最小化应该付出的社会成本，将经济学方法运用到犯罪的研究中。
- 我们可以把犯罪活动的强度（次数、时间等）总结为包含一系列变量的函数

$$y = f(x_1, x_2, \dots, x_n)$$

此式中

在经济学中，我们通常不去假设 f 的具体形式，而是假设它的某些特征，如增减、凹凸等

y = 犯罪活动的强度

x_1 = 犯罪活动带来的回报（如非法所得的钱财、内心需求的满足等）

x_2 = 犯罪活动的机会成本（如同样时间内合法工作的回报等）

...

其他变量包括逮捕率、量刑等犯罪成本，教育程度、收入等个人特征

从经济学模型到计量经济模型

以Gary Becker的犯罪模型为例

- 犯罪的计量经济模型

$$\text{crime} = \beta_0 + \beta_1 \text{wage}_m + \beta_2 \text{othinc} + \beta_3 \text{freqarr} + \beta_4 \text{freqconv} \\ + \beta_5 \text{avgsen} + \beta_6 \text{age} + u$$

此式中

计量经济学需要假设具体的函数形式，这里以线性函数为例

crime = 参与犯罪活动的频率

wage_m = 从事合法工作带来的报酬

othinc = 其他收入（投资、财产继承等）

freqarr = 被逮捕的频率（作为被逮捕概率的近似）

freqconv = 被判刑的频率

avgsen = 平均服刑年数

回归分析

- 核心问题：从样本数据推测总体中变量间的关系
 - 在线性回归模型中，我们通过推测因变量的系数探讨变量间的关系
- 参数模型的拟合：根据样本观测值计算最优的参数组合
 - 什么是最优？使目标函数最小，例如最小二乘法就是使残差平方和最小
- 参数的估计：根据样本观测值猜测参数的真实值（不可观测）
 - Gauss-Markov 定理指出：线性回归模型的最小二乘估计量是 BLUE 的
 - 估计的准确性（是否非偏）和精确性（误差的大小）往往不可兼得
 - 根据研究目的确定估计量的选择标准：目的为解释时更注重非偏性、目的为预测时更注重精确性。

最小二乘估计量仅在线性非偏估计量中是最优的

线性回归模型

Linear regression model

我们假设观测数据服从回归模型，因此此式也被称作
data generating process

- 多变量线性回归模型

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i$$

若假设 $E[u_i | x_{i1}, \dots, x_{ik}] = 0$ ，则

$$E[y_i | x_{i1}, \dots, x_{ik}] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$$

- 线性指的是上述回归函数关于系数 $\beta_0, \beta_1, \dots, \beta_k$ 是线性的。
- 回归分析的主要目的：
 - 正确选择回归函数（基于经济学理论或对问题的理解）
 - 正确估计系数 $\beta_0, \beta_1, \dots, \beta_k$ 的值（基于样本数据）
 - 对总体中系数的值进行推断（基于模型假设和统计学理论）

最小二乘估计

Ordinary Least Square Estimation

- 假设单变量回归模型 $y_i = \beta_0 + \beta_1 x_i + u_i$
- 定义残差平方和

$$SSR = \sum_i (y_i - b_0 - b_1 x_i)^2$$

- 最小二乘估计量是使 SSR 取值最小的 (b_0, b_1) 的值
- 解法：求一阶条件，然后解联立方程

$$b_1 = \sum_i (x_i - \bar{x})(y_i - \bar{y}) / \sum_i (x_i - \bar{x})^2$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Gauss-Markov 定理

- 最小二乘估计量 (OLS) 是众多估计量中的一种，因为它满足很多好的性质，因此是最常用的估计方法。
- 估计量是观测值的函数，因此也是随机变量。若 $\hat{\beta}$ 是 β 的估计量，而 β 的真实值是 β_0 ，我们说
 - 当 $E[\hat{\beta}] = \beta_0$ 时， $\hat{\beta}$ 是非偏的 (unbiased)。
- OLS估计量是非偏的。
- OLS估计量是线性的 (它是 y_i 的线性函数)。
- 在适当的假设下，OLS估计量在所有线性非偏估计量中是方差最小的，因此被称作 the best linear unbiased estimator (BLUE)。

内生性

Endogeneity

- 内生性和外生性（exogeneity）源自联立方程模型。
 - 当一个变量可以被系统内其他变量解释时，该变量被称作内生变量，一般出现在方程左侧。当该变量被用来解释其他变量时，也可以出现在方程右侧。
 - 当一个变量无法被系统内其他变量解释时，该变量被称作外生变量，一般只出现在方程右侧。
- 若变量是内生的，则有可能和误差项相关，因此现在对内生性的定义，主要指解释变量和误差项之间存在相关性。
- OLS的主要假设是**不存在内生性**，表达为 $E[u | x] = 0$ 。

异方差、自相关、误差项的分布

- 异方差性 (heteroskedasticity) 是指不同观测值的误差项的方差可能不同。OLS一般假设同方差，即

$$\text{Var}[u_i | x] = \sigma^2$$

- 自相关 (autocorrelation) 是指不同观测值的误差项之间存在相关。OLS一般假设不存在自相关，即

$$\text{Cov}[u_i, u_j | x] = 0$$

- 有时我们也假设误差项服从多变量正态分布。

理论在实证研究中的重要性

估计边际生产率

- 让我们考虑一下如何估计生产要素的边际生产率的问题
- 以劳动 (labor) 为例, 令总产出为 Y , 劳动投入为 L , 则劳动的边际生产率定义为 $\frac{\partial Y}{\partial L}$ 。

- 如果假设 Cobb-Douglas 生产函数 $Y = AL^\alpha K^\beta$, 则

$$\frac{\partial Y}{\partial L} = \alpha AL^{\alpha-1} K^\beta = \alpha \frac{Y}{L}$$

这里假设了经济学模型

- 因此, 在实证研究中通常通过估计 α 和 β 来估计边际生产率, 即考虑下面的回归模型

$$\log Y_{it} = \log A + \alpha \log L_{it} + \beta \log K_{it} + u_{it}$$

这里将经济学模型转化为计量模型

此处, i 可以代表地区、企业等生产单位。

理论在实证研究中的重要性

Wang, H. & Lall, S. (2002). Valuing water for Chinese industries: a marginal productivity analysis. *Applied Economics*, 34:759-765.

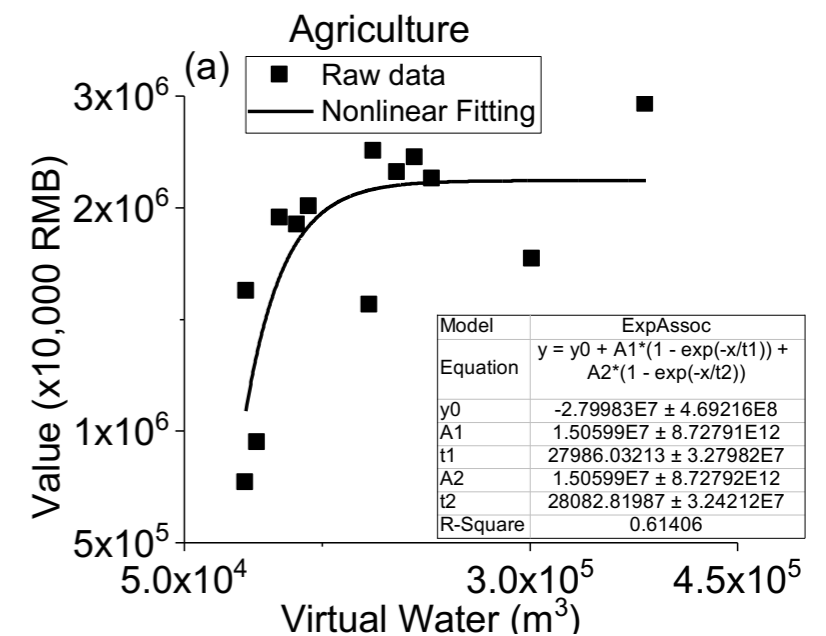
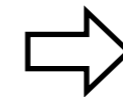
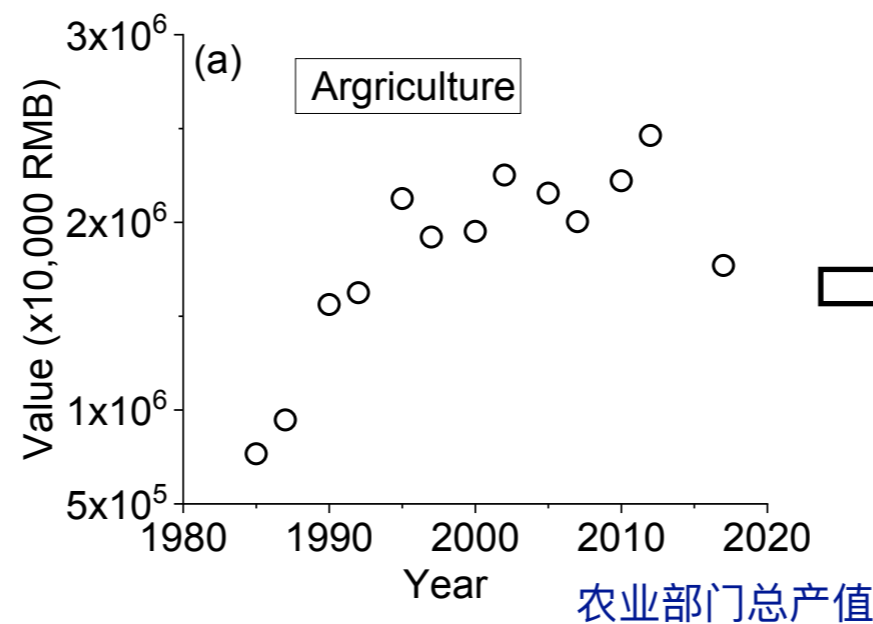
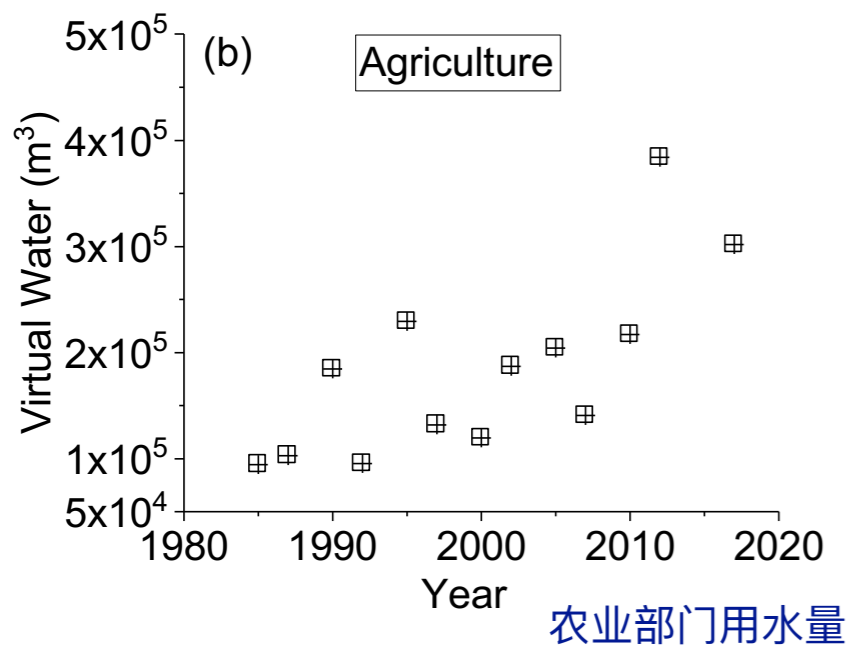
- 在实证中估计生产函数并不容易，最大的障碍往往是无法获得有效数据！
- Wang & Lall (2002) 用企业级数据估计了中国各产业部门的水资源边际生产率。
 - 模型：translog production function
$$\log Y_i = \beta_0 + \beta_1 \log K_i + \beta_2 \log L_i + \beta_3 \log W_i + \beta_4 \log E_i + \beta_5 \log M_i + \text{二次项}_i + u_i$$

水 能源 原材料
 - 数据：作者在世界银行工作，通过和中国环境保护部（现生态环境部前身）的合作研究项目，获得了1993年国有大中型企业中约2000家工厂的环境数据，变量包括 Y, K, L, W 以及其他特征变量
 - 文章内容比较简单，但是当你想要参考它做类似研究时，你会发现很难获得数据

理论在实证研究中的重要性

一个失败的硕士论文

- 在关于水与能源的边际生产率及相关研究中，企业级数据极难获得。公开可获得的数据多为地区级投入产出表（基于普查和估算，时间间隔2、3或5年不等）。
- 文章作者为了估计北京地区不同生产部门的水和能源的边际生产率，利用1985-2017年间13个年份的投入产出表计算了水资源投入量、能源投入量、以及总产值。以水为例，作者对用水量 and 总产值间的关系进行了非线性拟合，继而通过对拟合出的函数进行求导，得出“**边际生产率**”的函数：



- 那么这位作者错在哪儿呢？

因果效应

Causal effects

- 因果关系 (causality) : 特定的行为导致特定的、可度量的结果
 - 增加教育投入是否会引起犯罪率的减少?
因: 教育投入 果: 犯罪率减少
(个人层面: 增加受教育程度是否会减小犯罪的可能性)
- 计量经济学学习和应用的重点之一就是鉴别分析结果是否能够解释为因果效应。而不同的数据获得方式需要不同的方法。
 - 实验: 理想实验获得的结果本身可以解释为因果效应
 - 准实验: 理想实验在社会科学中非常困难, 因此需要在研究设计层面创造近似理想实验的状态
 - 观测数据: 对于如何利用观测数据正确地总结出因果效应, 学术界还没有统一的认识。一些教科书将满足 *ceteris paribus* 解释为因果效应, 另一些则会在反事实推论的框架内讨论。

其他条件不变

Ceteris Paribus

- Ceteris Paribus — 指其他条件保持不变。
- 当我们讨论一个变量的变化（原因）是否会引起另一个变量变化（结果）时，我们一般假设没有其他变量一起发生变化。如果有，则很难说清是哪个变量导致了结果的产生。
- 经济学的问题多数都假设其他条件不变
 - 例如：缩小班级规模会提高小学教育质量
这里假设教师能力、学生能力、学生家庭环境、学校所处城市的社会环境、学校设备等其他所有可能影响教育质量的变量都不变，而只缩小班级规模。
- 现实中很难实现
 - 班级规模不同的小学，其他条件往往也不同（例如学校所处的社区人口密度、学生家庭收入等）
- 在回归分析中，ceteris paribus 等同于控制其他变量。但往往我们不知道该控制多少变量，且有些变量无法观测，因此也就无法控制。

The Neyman-Rubin Causal Model

- “No causation without manipulation.”

Rubin (1975), Bayesian Inference for Causality: The Importance of Randomization, *Proceedings of the Social Statistics Section of the American Statistical Association*.

- Treatment (处理): 对每个个体进行的干预行为。
- Unit (个体): 在某一特定时间点上的物体 (physical object)、人、公司、或团体。不同时间点上的同一物体被视为不同的个体。
- Potential outcome (潜在结果): 在进行处理前 (ex anti) 我们知道处理结果可能是两种处理状态 (treatment and control) 分别对应的变量取值 $Y_i(\text{treated})$, $Y_i(\text{control})$ 。而进行处理后 (ex post) 我们只能观测到其中一种 $Y_i(\text{observed})$, 无法观测的则被称为反事实 (counterfact) 。
- Causal effect 的**定义**: $Y_i(\text{treated}) - Y_i(\text{control})$
因果效应是对同一个体在同一时间点 (处理后) 上的潜在结果间的比较。
- Causal effect 的**估计**: 基于可观测变量的实际观测值之间的比较。可能是同一个物体在不同时间点上的比较, 也可能是多个物体在同一时间点上的比较。

随机对照实验

Randomized controlled experiment

- 实验：针对被试验者进行特定的处理，然后观察被试验者的状态
- 保持其他条件一致
 - 理学实验：条件可控，可以获得个体因果效应
 - 生命科学或社会科学实验：条件很难控制
 - 例外：针对双胞胎进行实验可以得到近似个体因果效应的结果
- 随机对照实验（randomized controlled experiment）：个体层面无法控制时，在均值层面控制
 - 随机从总体中选择被试验者（random sampling），并随机分配处理组和对照组（random assignment）
 - 对比处理组和对照组结果的均值： $\bar{Y}(\text{treated}) - \bar{Y}(\text{control})$
- 多数时候，经济学研究无法进行随机对照实验
 - 例如，我们不能以实验为名随机决定大学入学资格

分配机制与选择偏差

Assignment mechanism and selection bias

Unit	Potential Outcomes		Causal Effect
	$Y_i(\text{drug})$	$Y_i(\text{surgery})$	$Y_i(\text{surgery}) - Y_i(\text{drug})$
Patient 1	1	<u>7</u>	6
Patient 2	<u>6</u>	5	-1
Patient 3	1	<u>5</u>	4
Patient 4	<u>8</u>	7	-1
Average	4	6	2

Y 为治疗后的生存年数
 假设我们可以观测到所有潜在结果，则真实的平均因果效应为 2
 即 surgery 比 drug 使患者平均多活 2 年

Unit	Treatment	Observed Outcome
	$W_i = \text{surgery}$	Y_i^{obs}
Patient 1	1	7
Patient 2	0	6
Patient 3	1	5
Patient 4	0	8

假设医生有办法给每个病人进行最有效的治疗，则 surgery 组的平均生存年数为 6，小于 drug 组的平均生存年数 7
 单纯比较观测值会受到分配机制的影响而无法准确估计因果效应

- 因果效应推断的核心思想：估计平均处理效应 (average treatment effect, or ATE)

$$\text{处理组结果 (6)} - \text{对照组结果 (7)} = [\text{处理组结果 (6)} - \text{处理组未处理时的结果 (1)}] + [\text{处理组未处理时的结果 (1)} - \text{对照组结果 (7)}]$$

$$\text{可观测的结果之差 (-1)} = \text{处理组的平均处理效应 (5)} + \text{选择偏差 (-6)}$$

因果效应推断

如何消除选择偏差

- 处理变量可操控时：随机对照实验
 - 随机性消除了选择偏差
 - 什么问题适合做实验？例：性别、种族可以作为处理变量吗？
 - 实验的成本有多大？
- 处理变量不可操控时：准实验（自然实验）
 - 准实验的分配通常不是随机产生的，因此为了消除选择偏差我们需要满足其他假设或使用特殊方法，以达到和随机分配近似的效果
 - 常用方法：工具变量（IV）、双重差分（DID）、断点回归（RD）等

因果效应估计的重要应用方向：政策评估

- 政策的事前评估与事后评估
 - 事前评估：基于理论预测与模拟
 - 事后评估：基于可获得信息的定性与定量分析
- 政策评估的特征：难以进行理想试验
 - 基于反事实推断的计量经济方法
- 例：对创新型企业的优待政策是否促进了创新？
 - 样本符合随机分配吗？
 - 是否存在潜在的选择偏差？
 - 如何进行反事实推断？