

# 《实验经济学讲义》笔记

Hischun Ma

June 2022

# 目录

<b>1 实地实验</b>	<b>3</b>
1.1 实地实验的简史 . . . . .	3
1.2 实地实验的分类 . . . . .	3
1.3 实地实验的研究方法 . . . . .	4
1.4 有效性与可复制性 . . . . .	4
1.5 审计实验法及通讯审计实验法 . . . . .	5
1.6 实地实验与 RCT 之比较 . . . . .	6

# 1 实地实验

本书中的“实地实验”指：在真实的自然环境中，选择自然人群作为被试，通过实验方法将其随机分配到处理组和对照组，赋予处理组以外生的冲击，观察两组被试行为决策的差异，以获取变量间的因果关系。

实验方法有许多优点。第一，它可以有效地处理“外部有效性”的争论，即实验室中得到的实验结果是否（或者说，在多大程度上）可以被推广至实验室外。第二，实地实验能够回应经济学研究者对“现实的报酬激励”有关问题的质疑（275）。第三，实地实验所体现的“试点”功能受到了政策制定者的青睐。

但是，实地实验的实施也需要考虑诸多因素（275）。常见的问题有：如何控制真实情景中的要素、如何确定参与者并且与其协商，以及如何简化操作程序等。

## 1.1 实地实验的简史

实地实验的发展大概经历了三个阶段。在第一个阶段中，随机化的思想被引入；第二个阶段为“大规模的社会实验”，它有四个特征：随机分配、政策干预、随访/后续数据收集与评估。

但是，大规模的社会实验存在两个问题：随机偏误（randomization bias）与损耗偏误（attrition bias）。随机偏误指因随机分配进入实验的被试和与那些实际政策利益相关的群体的差异而引起干预效果的有偏估计。比如，一些学者（Heckman, Smith and Manski）认为，小规模实验的参与者并不具有代表性，选取特定群体作为被试可能导致系统性差异。而损耗偏误则指的是“处理组与对照组之间参与者的不同损失”，比如，有些参与者很可能在实验过程中退出实验，也有一些参与者会受到“霍桑效应”的影响。

实地实验发展的第三个阶段为“实验室外的受控实验”，在这个阶段中，研究对象往往没有意识到它们已成为实验的一部分，其规模也小于社会实验。<sup>1</sup>

## 1.2 实地实验的分类

《讲义》中介绍了两种分类方法，其分别基于两篇论文，我们记为 Harrison-List 分类法（Harrison and List, 2004）与 Charness 分类法（Charness, 2013）。

在 Harrison 与 List 的分类中，实地实验被分为三种：人为实地实验（artefactual field experiment）、框架实地实验（framed field experiment）与自然实地实验（natural field experiment）。人为实地实验与实验室实验的唯一区别在于它使用的是“非标准”的被试池，即被试不再是学生。除此之外，它有两个重要的应用：发展经济学与解释、预测非实验结果。前者可见 Henrich et al. (2004)；后者可见 Barr and Serneels (2004)，以及 Carpenter and

---

<sup>1</sup>关于实地实验的详细发展状况，请见 List 的网站：[www.fielddexperiments.com](http://www.fielddexperiments.com)

Seki(2005)。

框架实地实验直接使用了真实的社会环境，其以一种确保被试理解他们正在参与实验的方式进行的，被试的行为会被记录和详细审查（被试知道自己处于实验中）。而自然实地实验则不然，被试并不知道自已处于实验中。但是，在 Charness 的分类中，实验室也可以作为实地实验的场景。

### 1.3 实地实验的研究方法

实地性的程度可以用四个维度来测量，它们分别是：干预真实性、参与者真实性、背景真实性以及结果真实性（Gerber and Green, 2012）。实地实验的基本方法是在真实的社会环境（real setting）中，随机抽选实验对象，然后将其**随机**分配到实验组和对照组，控制其他因素不变，对对照组实施真实的干预，并对随后两组被试的行为决策差异进行比较，进而得出**因果效应**。

实验的基本过程一般分为五个步骤（List, 2011）。第一，确定实验参与对象。第二，对实验参与者随机分组，一般根据实验问题分为一个对照组以及一个或者多个实验组。第三，实施干预。第四，收集数据。第五，分析数据。

### 1.4 有效性与可复制性

有效性是一个多维度的概念，其中两个最为关键的维度是内部有效性和外部有效性。内部有效性是指所观察到的相关性确实是因果关系，外部有效性是指研究中所发现的关系推广至其他群体。实验室实验与实地实验均面临着有效性之间的权衡，但是，这种权衡不应该被解释为两种方法之间的竞技。相反，诸多实验方法应该是互补的，它们都是不同的工具，适用于不同的目的。（Charness, 2015）

外部有效性是实地实验的优势之一。因为实地实验在真实的情景中选取真实的被试，采用真实的干预措施，所得出的实验结果对于所运用的环境情景更为普遍。但是，实地实验的内部有效性要弱一些。因为实地实验研究人员所能施加的控制更少，这主要体现在以下几个方面。第一，研究对象的控制。一般来说，实地实验的被试群体文化水平较低且被试的生活经验较为丰富，这将增加被试理解实验说明的难度且容易对被试的决策造成干扰。第二，实验环境的控制。由于实地实验在自然的环境中进行，可控制的变量相对较少。第三，实施过程中的控制。研究人员或许会面临实施方面的困难，例如寻找可信赖的合作伙伴（非营利组织、公司或学校的从业者）以及被试中途退出问题，Gueron(2016) 描述了在—项合作中如何建立和发展合作关系，可以作为参考。<sup>2</sup>

一般来说，可复制性分为三个维度：获取并重新分析原始数据、重新实施类似的实验以

---

<sup>2</sup>研究者必须提出一个另各方都接受的实验设计。实地实验需要研究者、实验干预者以及结果测量者之间不断协调达成一致决定，而研究合作者和资助方可能反对采用随机分配的思想。相反，他们有时更倾向于全部干预或精心挑选被试。

及涉及新实验。对于实地实验来说，第二个维度较为困难，因为很多实地实验都是机会性的，并且需要与第三方合作，这些外部的客观现实条件都会使复制变得困难（Levitt and List, 2009）。<sup>3</sup>

## 1.5 审计实验法及通讯审计实验法

在现实中开展实地实验是很困难的，但是两种相对容易操作的实验方法——审计实验法（audit study）与通讯审计实验法（correspondence studies）——逐渐被主流经济学接受。

所谓审计实验法，是采用**角色扮演**的方法建构一个真实的实地实验环境，通过考察所扮演的角色与现实中行为主体的互动来研究所要探讨的社会经济问题。角色扮演是指两个只在**某一方面**存在差异的个体分别于真实市场中的另一方行为人进行互动，研究者通过对比这两个个体多受到的**差别对待**来分析另一方的市场行为（陆方文，2014，2020）。

审计实验法大概有以下三个步骤：首先，设计两位不同的审计员（tester），审计员之间只有某一方面的差异，其他方面必须保持**几乎完全相同**。其次，当两个审计员分别与真实存在的另一方进行互动时，另一方在整个互动过程中**完全不知情**。最后，收集双方互动信息。

虽然审计实验法操作相对简单，但是其本身也有局限性，主要为以下几点。第一，有效匹配问题（Heckman and Siegelman, 1993; Heckman, 1998）。它意味着，即便在实验前对审计员进行培训，也很难消除审计员之间存在的差异。第二，实验设计的非双盲性。在审计实验法中，审计员知道实验的目的，他们对实验的认识可能会影响他们的期望或行为，进而影响实验结果，一个代表性的例子参见 Pager(2003)。

为了弥补审计实验法的不足，出现了通讯审计实验法。从本质上来说，它可以被视为审计实验法的书信形式（陆方文，2014）。在通讯审计实验法中，审计员作为虚构的申请者与真实存在的市场主体交流。该方法具有明显的优势。第一，实验员所收到的信息在不同组之间有严格的可比性，进而保证了任何观察到的差异都是由研究者操纵的少数特征引起的。第二，该方法一般情况下不会受到需求效应的影响。第三，成本低廉。

但是，通讯审计实验法也有局限性。第一，可以研究的实验结果变量比较单一和粗糙。在大多数情况下，将回复率作为唯一结果不能考察出申请者在诸多方面所受到的区别对待。比如，Riach and Rich(2002) 认为，如果出现多数和少数候选人都被拒绝的情况，并不能构成平等对待的证据，只有使用更多连续的变量才可以进一步判断，但是，这些变量通常在通讯审计实验法中很难获得。第二，伦理道德问题。虚假的申请者带有“欺骗”的性质，这会浪费雇主的时间。但是，List(2009) 认为，当研究者能使参与者生活得更好，对社会有益，并赋予所有研究对象匿名性和公正对待时，缺乏知情同意是可以理解的。

---

<sup>3</sup>不同的社会背景对实地实验的实施也有影响，在不同地区或国家的群体中复制已有的实地实验研究发现能进一步地提高实地实验的推广性，进而提高外部有效性。

## 1.6 实地实验与 RCT 之比较

本小节将比较实地实验与随机对照实验，主要按以下顺序展开：首先说明实地实验与随机对照实验的发展；其次说明二者的相同之处与不同之处；最后比较二者的有效性。

实地实验的发展可以分为两个路径。其中一条路径是从农业领域转向大规模的社会实验，并在社会实验的基础上做方法论的改进；另一条路径起源于学者对实验室方法的质疑和反思，因此将实验室实验扩展至实地实验（罗俊，2015）。而随机对照实验发展则较晚，其主要应用与发展经济学中，尤其是政策评估领域。<sup>4</sup>

从核心思想的角度来看，实地实验与随机对照实验几乎是一致的：利用随机化的思想，将具有真实决策行为的参与者随机分为是实验组和对照组，通过对比两组（多组）参与者的行为差异来检验研究者所关心的某种干预、政策或项目的处理效果（Duflo, 2006）。除此之外，二者的共同点还在于对“反事实”问题的回答，这也是因果推断的本质所在（Stock and Waston, 2015; Duflo et al., 2007）。

但是，二者也有差异。第一，从狭义的范围而言，用于政策评估的实验方法一般是指随机对照实验，实验依据发展经济学理论设计，数据的收集也是基于微观个体的行为。第二，随机对照实验用于政策检验更偏向于由果溯因的过程：观察并评估干预效果，根据该结果回溯影响微观个体行为的原因；而实地实验更像是一个由因到果的过程：预设产生某种效果的机制或措施，并在此基础上检验该机制或干预是否产生了所预测的影响。<sup>5</sup>

接下来从内部有效性和外部有效性的角度比较实地实验和随机对照实验。第一，从内部有效性的角度来看，随机对照实验的内部有效性较强。因为在实地实验中，研究人员所能控制的变量较少且在实验实施的过程中容易受到其他未知因素的影响，故不利于因果关系的识别。但是，随机对照实验可以通过控制单一因素或某几个因素的组合实现对因果关系的准确识别，以实现内部有效性。第二，从外部有效性的角度来看，实地实验的外部有效性较强。因为实地实验是在真实的自然环境背景下，堆积悬泉真实的实验对象施加真实的感具。而随机试验面临着以下三个方面的考验：实验样本的非代表性、实验的实施过程的复杂性，以及实验环境的差异（Duflo et al., 2006）。

综上，随机对照实验可以被视为广义的实地实验，其主要目标在于政策的检验与评估，病因局限于特定地区而削弱了外部有效性。但是，整体而言，二者都是进行因果推断的强有力的工具，在研究方法与内容上有交集（不互相包含）。更重要的是，二者都要面临内部有效性和外部有效性的权衡，因此并不谁更优之说，研究者应根据研究问题来选择最适合的方法。

---

<sup>4</sup>当经济学者评论“实地实验”时，其通常指 Levitt and List(2009) 中的自然实地实验，而随机控制实验是指在所有自然实地实验中，直接用于项目评估目的，或者以解决实际问题，而非检验经济学理论为导向的自然实地实验研究（包特，2020）

<sup>5</sup>有一种实验叫做 lab-in-the-lab（实地中进行的实验室实验），可以被用来测量被试的偏好，可以为实地实验丰富维度。详见《讲义》299 页