研究：原理、方法与实践

著者：阿诺•巴塔查尔吉

（ANOL BHATTACHERJEE）

译者：沈校亮（武汉大学）

孙永强（武汉大学）

全球教材计划

Global Text Project

社会科学研究: 原理、方法与实践

阿诺•巴塔查尔吉（Anol Bhattacherjee）博士

南佛罗里达大学
坦帕　佛罗里达州　美国
电子信箱: abhatt@usf.edu

第二版

The Global Text Project is funded by the Jacobs Foundation, Zurich, Switzerland

社会科学研究：原理、方法与实践（第二版）
阿诺•巴塔查尔吉（著）
沈校亮、孙永强（译）
2012 年出版

ISBN-13:

目 录

前　言

本书旨在向硕博研究生介绍社会科学、商学、教育、公共卫生和其他相关学科的科学研究过程。该书根据作者在南佛罗里达大学面向博士生开设的研究方法课程教案整理而成，其受众包括硕士和博士研究生、初级研究人员，以及讲授研究方法课程的老师，当然一些高级学者也可将此书作为参考用书。

本书与市场上的其他类似教材到底有哪些区别呢？这可能是本书的潜在读者最为关心的问题。本书与其他教材区别主要体现在以下三个方面。首先，与其他教材不同的是，这本书不仅涵盖了研究方法（实证数据的收集与分析），同时它还对整个研究过程进行了介绍。研究方法只是研究过程中的一个阶段，也可能是最简单、最结构化的一部分。市面上大多数教材对研究方法进行了深入地介绍，但是却未对更具有挑战性且非结构化的重要议题进行探讨。比如应如何像研究者一样去推理和思考这样的问题，而这往往是实证研究的先决条件。根据以往的经验，很多博士生对于研究方法非常了解，却不能很好地提出有趣且实用的研究问题及构建相关的科学理论。因此，本书将用整段的章节来详细探讨此类问题，如“像研究者一样思考”以及“科学研究中的理论”等。这些部分对于初级研究人员而言是一项必备的基本技能。

第二，本书语言简洁，内容紧凑。在撰写此书的时候，我们仅讨论最重要的概念，而未罗列那些并不特别相关的材料。大部分面向博士生的研讨会都包含一些来自于某特定学科的阅读材料，本书将对这些阅读材料是个重要的补充。我们以一种简洁的方式总结了所有重要的概念，而非单纯地给学生提供大量的指定读物。

第三，本书当前和未来的所有版本都将提供免费下载。该书的亚马逊 Kindle 版本、苹果 iBook 版本以及纸质版本均能以一个象征性的价格获取。许多人会问我为什么不去售卖这本书，而是将它免费提供给社会？首先，我们不愿意看到很多学生因为教材定价过高而抱怨；再者，我们认为价格等因素不能成为限制科学知识传播的壁垒。只有当世界各地的学生和学者都能够获取科学所能提供的最好资料时，科学才能得以发展和进步。因此，这本免费教材是我们为此所做出的一点努力。当然，免费并不意味着低质量，生活中一些最美好的事物，如空气、水、阳光等也都是免费的。谷歌的资源同样是免费的，如果没有这些免费的资源，我们无法想象今天的互联网时代会是怎样。还有一些当今最先进的软件，比如 Linux 和 Apache 也是免费的，本书也是一样。

本书分为 16 个章节以适合 16 周的学期制。当然，授课老师可以根据自己的课程安排灵活地增减或扩缩相关的章节。比如，由于我们学校开设了统计学课程，我们将不会在课堂上讲授本书的第 14 章和第 15 章。在实际的授课过程中，我们一般会花两周的时间讲授第三章的理论部分，一周的时间对学术期刊进行讨论和回顾，还有一周的时间用于期末考试。考虑到有些学校并没有开设统计学课程，我们感到有必要将这两个章节包含在此书中。另外，商科博士项目所使用的教学大纲也提供在本书的附录中。

最后，我们打算基于科学研究的最新趋势不断地对本书进行更新和完善。如果您希望在未来的版本中看到任何其他内容，欢迎通过电子邮件（联系信箱：xlshen@whu.edu.cn）告知我们，并且我们将尽最大的努力对这些内容予以考虑。由于时间和能力有限，本版内容难免会有错漏之处，恳请广大读者批评斧正，多提宝贵意见，在此我们对于任何建议、批评或者指正表示衷心地感谢！

著者：阿诺•巴塔查尔吉（Anol Bhattacherjee）
南佛罗里达大学 商学院

译者：沈校亮（武汉大学经济与管理学院）
孙永强（武汉大学信息管理学院）

第一章

科学与科学研究

什么是科学？对于这个看似平淡无奇的问题，你可以从不同的人那里得到不同的答案。有些人会说他们经常研究不同的网站以找到他们想要购买的最好的商品或服务。电视新闻频道会使用观众投票的方式对一些涉及公众利益的问题（如即将进行的大选或者政府资助的项目等）进行相关课题的研究。本科生会通过研究互联网来搜寻完成课程作业或学期论文所需的信息。从事科研项目的研究生或许认为研究就是收集和分析与项目相关的数据资料。企业和咨询顾问通过研究各种可行性方案以解决组织的问题（如供应链瓶颈，或发现消费者的购买模式等。然而，除非遵循科学的研究方法，并对科学本身有所贡献，否则上述的各点都不能被认为是科学研究。本章将讨论科学究竟是什么。

科学

科学是什么？对于某些人而言，科学是指为最聪明的学生所提供的高深的物理、化学和生物等高中或大学课程。对于另一些人而言，科学是指在实验室中通过专业设备从事研究的科学家所掌握的技能。在词源上，科学这个词来自于拉丁语中的 scientia，即知识的意思。科学是指在任何需要使用科学方法进行研究的领域中所包含的有组织的系统化知识。科学大体上可以分为两大类：自然科学和社会科学。自然科学研究针对的是由自然界所产生的事物或现象，如光、物体、物质、地球、天体或人体等。自然科学可以被进一步划分为物理科学、地球科学、生命科学等。物理科学由物理学、化学和天文学等学科构成。地球科学主要由地质学等学科构成，而生命科学主要由生物学和植物学等学科构成。与之相反的是，社会科学以人或者人群（如团体、企业、社会、经济）为研究对象，以及探索与之相关的个人或集体行为。社会科学分为心理学、社会学和经济学等学科。

自然科学和社会科学在很多方面都存在着差异。自然科学非常精确，且不依赖于从事科学观测的人。以物理学中的科学实验为例，测量声音在某个特定媒介上的传输速度或者水的折射率的结果总是相同的，而与实验时间、地点及实验人员无关。如果两个学生在做同一个物理实验时得到了两组不同的数值，那么我们通常认为其中一个学生或者两个学生的实验结果都是错误的。然而，对于不那么精确的社会科学而言，我们不能得出同样的结论。比如你用同一个量表去测量某人的幸福指数，你会发现同一个人在不同的日子，有时甚至在同一天的不同时间，表现出不同的结果。某人的幸福指数可能会受到这个人在那天所获知的新闻或其

他事件的影响，更进一步而言，并没有哪个量表可以准确地测量出一个人的幸福感。因此，一个量表可能会测量出某人是幸福的，而另一个量表在同一时刻会发现这个人并不是那么幸福。换句话说，社会科学领域存在着很大程度的测量误差，并且对于社会科学而言，具有相当大的不确定性和较少的共识。比如说，自然科学领域的科学家在光速或者地球的公转速度方面并不存在什么分歧，但是社会科学领域的科学家们却在如何解决诸如打击全球恐怖主义和经济复苏等社会问题上存在这大量的争议。任何从事社会科学研究的学生都必须认识到并能灵活地处理这种高模糊性、不确定性以及伴随而来的错误，而这也反映了社会对象的高易变性。

科学也可以根据其研究目的进行分类。基础科学，也叫做纯科学，用于解释最基本的对象，它们之间的关系，以及驾驭它们的法则。这样的例子包括物理学、数学和生物学。应用科学，也叫做实践科学，主要把基础科学中的科学知识应用于具体的物理环境下。比如工程学作为一门应用科学主要是把物理和化学领域的法则用于构建实际应用，如建设更牢固的桥梁或者提高内燃机的燃油效率等。医学同样作为应用科学将生物学的法则用于医治人类的疾病。对于人类的发展而言，基础科学和应用科学都是必不可少的。但是应用科学并不能独立存在，而是需要依靠基础科学以促进其发展。鉴于应用科学的实际价值，工商业往往把重点放在应用科学上，而大学通常对基础科学和应用科学都进行研究。

科学知识

科学的目的在于创造科学知识。科学知识指通过科学方法获取的用于解释现象或行为的一系列规律和理论。规律是指所观察到的现象或行为的模式，而理论是指对于现象和行为的系统化解释。比如在物理学中，牛顿运动定律描述的是一个物体处于静止或运动状态下的状况（牛顿第一定律），移动一个静止物体或者让一个运动物体静止下来所需要的外力（牛顿第二定律），以及当两个物体发生碰撞时的状况（牛顿第三定律）。总体而言，这三个定律是经典力学的基础。同样地，光学理论解释了光的特性及它在不同媒介中的表现，电磁理论解释了电的特性及其如何产生，量子力学解释了亚原子粒子的属性，天文学解释了恒星和其他天体的特征，热力学解释了能量和机械功的特性。一本面向高中或大学本科的物理教科书会使用独立的章节来介绍这些理论。在社会科学领域，也有一些类似的理论。例如，心理学中的认知失调理论解释了当人们对一个事物的观察结果与之前对该事物的看法存在偏差时，他们将会做出何种反应；威慑理论解释了人们为何参与一些不当的甚至是犯罪的行为（如从非法网站下载音乐或者盗版软件）；计划行为理论主要解释的是人们在日常生活中如何有意识的做出各种理性选择。

科学研究的目标在于发现各种规律，并提出可用于解释自然和社会现象的理论，或者换句话说，即创造科学知识。同时，需要格外注意的是，这些知识可能并非尽善尽美，有时候甚至与真理差距甚远。在一些情况下，或许没有单一的普遍真理，而是由多个真理所构成的均衡状态。我们必须认识到科学知识所依附的这些理论仅能用于解释某个特定的现象。由于这些解释在不同程度上与现实相符，因而可能存在好的或差的解释，与之伴随的是好的或者差的理论。因此，我们要通过更精密的仪器以及更严谨的逻辑推理来获取更好的观察结果，一步步地从差的理论向好的理论发展过渡，这才是科学进步的标志。

我们通过逻辑和迹象分析找到科学的规律或理论。逻辑（理论）和迹象（观察结果）是科学知识得以发展的两大支柱。在科学研究中，理论和观察是相互依赖，不可分割的。理论为

我们所观察到的现象提供了重要的指导意义，而观察帮助我们验证和完善已有的相关理论，或构建新的理论。任何其他的知识获取方式，比如信念、信仰或者哲学都不能称之为科学。

科学研究

鉴于理论和观察是科学的两大支柱，科学研究可以从两个层面展开：理论层面和经验层面。理论层面主要考虑的是提炼自然或社会现象的抽象概念及这些概念之间的关系（即理论构建），而经验层面主要考虑的是对理论概念及它们之间的关系进行检验，发现理论概念与现实观察之间的相符程度，并最终构建更好的理论。随着时间的推移，理论不断被完善（即与现实观查更加相符），科学也逐步变得更成熟。科学研究就是在理论与观察之间的持续交替。因此，理论和观察都是科学研究的重要组成部分，比如在科学研究中完全依赖观察来进行推理而忽略理论的做法是不被认可的。

根据科研人员训练及兴趣的不同，科学探索可采取归纳和演绎两种方式。归纳研究中，研究人员根据观察到的数据来构建理论概念及其模式。演绎研究中，研究人员通过经验数据对既有理论中的概念和模型进行检验。因此，归纳研究也叫做理论构建研究，而演绎研究也叫做理论检验研究。值得注意的是，理论检验不只是单纯地对理论进行检验，同时也包括对理论的提炼、改进和扩展。图 1.1 中描述了归纳研究和演绎研究的互补特征。归纳和演绎研究是研究周期中的两个部分，并且不断地在理论和观查之间交替出现。作为一个研究者，若不熟悉研究中的理论和数据，他将无法进行归纳和演绎研究。当然，一个全面的研究人员应该具备同时开展这两种研究的能力。

理论构建（归纳研究）和理论检验（演绎研究）对于科学的进步都至关重要。如果理论不能与现实相符，即使再漂亮的理论也无济于事。同样地，如果不能用于构建新的理论，即使是堆积如山的数据也毫无意义。与其将图 1.1 中所示的两个过程看作是循环的关系，倒不如将它们看作是螺旋上升的关系，理论和数据之间的每次循环都将有助于更好地解释现象及更好地发展理论。尽管归纳和演绎这两种研究对于科学的进步都非常重要，但是在理论和现象解释均较少的领域，归纳研究更有价值。与之相反，在解释同一个现象时，存在多个相互抵触的理论，则演绎研究对于帮助研究者发现最合适的理论（即在何种情况下哪个理论可以对什么现象做出最好的解释）更富有成效。

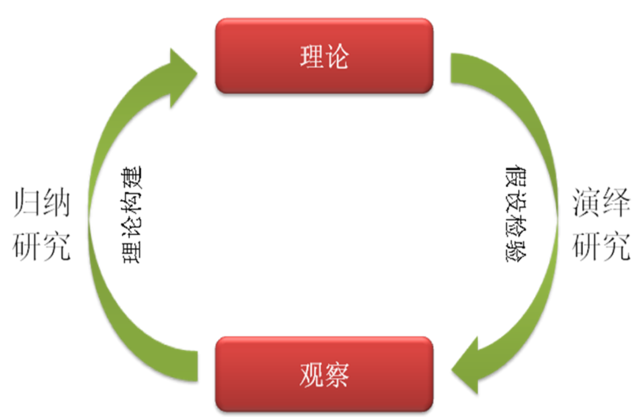


图 1.1 研究循环

由于理论概念不够精确、测量工具缺乏以及各种可能会影响社会现象的不可预知因素的存在，这些都将导致在社会科学领域理论构建和理论检验变得非常困难。 对于那些不好的理论也很难予以驳斥。比如，卡尔·马克思的共产主义理论多年来一直被认为是个有效的经济生产方式，而现在我们却认为该理论在推动经济发展和社会福利等方面并不如资本主义。以前的共产主义经济，如苏联和中国，最终走向了以利润最大化的私企为代表的资本主义经济。然而，近期发生在美国的抵押贷款和金融行业的崩溃也表明资本主义同样有它的不足，并且不能像以前所推测的那样可以有效地促进经济增长和社会福利。与自然科学的理论不同，社会科学的理论很少有完美的，这也为研究人员带来了大量的机会来发展这些理论或者构建他们自己的理论。

开展科学研究需要掌握理论和方法两种技能，这两种技能分别是理论层面和经验层面的研究所必需的。方法技能（知道如何做）对于不同的学科而言是相对标准化的，并且很容易通过博士课程的学习来获得。然而，理论技能（知道做什么）却相当难以掌握，需要基于多年的观察和思考，这种隐性的技能只能通过经验来学习而无法直接进行传授。人类历史上所有最伟大的科学家，如伽利略，牛顿，爱因斯坦，尼耳斯·玻尔，亚当·斯密，查尔斯·达尔文和赫伯特·西蒙等都是理论大师，他们因为提出了改变科学进程的理论而被人们铭记在心。方法技能对于一个普通的研究人员而言是必不可少的，但是理论技能一定是一个杰出的科学家所必备的！

科学方法

在前面的章节中，我们将科学视作通过科学方法所获取的知识。那么，什么才是科学方法呢？科学方法指一整套用于创造科学知识的标准化技术，比如如何进行有效的观测，如果对结果进行解释，以及如何概括归纳这些结果。科学方法让研究人员可以独立而公正地对已存在的理论和研究结果进行检验，并且对这些理论和研究结果进行公开地辩论、修正或改进。科学方法必须满足四个特征：

- 可复制性：其他人能够独立地复制或重复一个科学研究，并且获得一样或相似的结果。

- 精确性：对于难以测量的理论概念必须准确定义，这样其他人可以使用这些定义去测量这些概念并对理论进行检验。

- 可证伪性：理论必须以一种可被证伪的方式进行阐明。不能被检验或证伪的理论不能称作科学理论，这样的知识也不能算是科学知识。如果一个理论以不精确的术语进行描述，或者理论概念不能准确地测量，这样的理论将无法被检验，因此也不是科学。西格蒙德·弗洛伊德关于精神分析的思想正属于此类，即便精神分析在治疗某些类型的疾病时有其实用价值，但仍然无法被看作是个理论。

- 精简性：当对于某个现象有多种解释的时候，科学家总是会接受最简洁的解释。这个概念被称作精简性或者“奥卡姆剃刀”。精简性保证了科学家不至于过度追求过于复杂或奇异的理论，这样的理论有数不清的概念和关系，看上去似乎可以解释任何事物，但事实上什么也解释不了。

对于任何无法通过科学方法来测试其基本规律和理论的研究都不能称作是科学。比如，神学（宗教研究）就不是科学，因为神学的观点（如上帝的存在）无法被独立观测者用一个可复制的，精确的，可证伪的，精简的方法来检验。艺术、音乐、文学、人文学科和法律同样也不能被看作是科学，即使这些学科是具有创造性的并且也值得为之努力的。

适用于社会科学研究的科学方法包含各种研究方案、工具和技术，如定性和定量数据、统计分析、实验、问卷调查、案例研究等等。读者可通过本书来学习这些研究方法。然而，需要注意的是科学方法主要用于经验层面的研究，即如何观察、分析和解释这些结果。很少有方法与理论层面直接相关，而这正是科学研究中更具有挑战性的部分。

科学研究的类型

根据研究的目的不同，科学研究项目可以分为三大类：探索性研究、描述性研究和解释性研究。探索性研究适用于一个崭新的研究领域，它的主要目标包括：（1）发掘特定现象、问题或行为的作用范围及程度；（2）针对该现象提炼出一些初步的思路（或预测）；或者（3）检验对某个现象展开广泛研究的可行性。以国民对于政府在经济衰退期间的政策普遍感到不满意为例，探索性研究可用于测量国民的不满意程度以了解这些不满是如何显现的，如公众抗议的频繁程度，以及造成这种不满的可能原因，如政府的相关政策在处理通货膨胀，利率，失业率或高税收时的低效性。这类研究可能会包括对解决类似问题的历史实践进行调查，或者对公开报道的数据（如经济指标估算、国内生产总值（GDP）、失业率、消费价格指数等）进行检验。这些数据可以通过多种渠道获取，如第三方来源、对专家或各利益主体（如著名经济学家和政府主要官员）的访谈等。这样的研究或许并不能针对目标问题得出非常准确的见解，但是对于指出问题的性质和程度很有帮助，并且为将来更深入研究奠定了基础。

描述性研究主要是对目标现象做出仔细观察和详细记录。这些观察必须基于科学的方法（即具有可复制性、精确性等），因此，与未受过专门训练的人员所获得的观察数据相比，其可靠性更高。描述性研究的例子包括美国人口普查局的人口统计数据或者劳工部统计的就业数据等。这些部门使用相同或相似的工具，并采用多个就业问卷调查或人口普查的方式，来估算不同行业的就业情况或不同族裔的人口增长速度。若测量工具发生了变化，其变化前后的估算数据都将提供给读者以便于对人口或就业趋势进行前后的对比。其他的描述性研究包括城镇中青少年帮派活动的民俗志报告研究，选定社区中宗教、文化或种族活动的发展与演变研究，以及信息技术（如微博和即时通信工具）在中东国家民主运动浪潮中的作用研究。

解释性研究是对所观察到的现象、问题或行为做出解释。描述性研究揭示了现象是什么（what）及其何时（when）何地（where）出现，而解释性研究回答的是为什么（why）以及如何（how）这类的问题。该研究试图找出目标现象的影响因素及其结果来对既有研究进行归纳概括。这样的例子包括理解青少年犯罪或者团伙暴力背后的原因，并最终找出解决这些社会问题的对策。尽管在学术研究的初级阶段需要一些探索性和描述性研究，但是，目前大部分的学术研究仍属于解释性研究。对所观测到的事物进行解释需要强有力的理论基础和阐释技能，同时还需要直觉、洞察力以及个人以往的经验。能够做到这一点的人在他们各自的领域也都是最出色的科学家。

科学思想史

在结束本章之前，我们或许可以先回顾下历史并了解一下科学是如何随着时间不断发展的，并找出这个演变过程中的关键科学精神。尽管几个世纪以来科学进步的例子不断被记载，然而“科学”、“科学家”和“科学方法”这些术语在 19 世纪才被创造出来。在这之前，科学被认为是哲学的一部分，并且与其他的哲学分支，如逻辑学、形而上学、伦理学以及美学（尽管这些分支之间的界限并不清晰）并存。

在人类探索的初期，知识常被视作基于信仰的神学戒律。在公元前 3 世纪，希腊哲学家，如柏拉图，亚里士多德，苏格拉底，对这个观点发起了挑战。他们指出生命和世界的本质可以通过系统地逻辑推理过程（即理性主义）得以更准确地认识。尤其是亚里士多德的经典作品形而上学（字面上的意思是“超越物理的存在”）将神学（对上帝的研究）从本体论（对生命和存在的研究）和普遍科学（对逻辑所依附的第一原理的研究）中分离出来。理性主义（不要与合理性混淆）将理性视作知识和判断的源泉，并且认为真理的标准不应该是感性的，而是从一系列的第一原理或公理（如亚里士多德的“不矛盾律”）中演绎和推论而来的。

16 世纪，英国哲学家弗兰西斯·培根（1561-1626）指出知识只能从对现实世界的观察中获取，此时科学思想迎来了它的下一个重要变革。基于这个前提，培根强调知识的获取是一种经验行为（而非推理行为），并且提出了哲学中具有重要影响力的一个分支——经验主义。培根的研究成果使科学探索中的归纳方法得到了普及，推动了由系统观察、测量和实验所组成的科学方法（最初也被称作培根方法）的发展，甚至为无神论或反对神学戒律（由于其不可观测）播下了种子。

哲学家们一直在找寻最有效地获取知识的途径，而在整个中世纪，经验主义与理性主义却在这个问题上不断地发生碰撞。法国哲学家勒内·笛卡尔站在理性主义一边，而英国哲学家约翰·洛克和大卫·休谟则站在经验主义一边。其他的科学家，比如伽利略和牛顿试图将这两种研究融合为自然哲学以关注于对自然界和物质世界的认识。自然哲学因而也被认为是自然科学的先驱。伽利略（1564-1642）第一次提出自然界的规律是数学，并且创新性地将实验和数学进行结合，对天文学做出了重要的贡献。

在 18 世纪，德国哲学家伊曼努尔·康德试图在他的著作《纯粹理性批判》中解决经验主义与理性主义之间的争辩。书中指出经验是完全主观的，若在没有探究经验的主观属性的情况下，用纯粹理性对经验进行处理将导致理论上的错觉。康德的思想带来了德国唯心主义的发展，进而激发了解释性技术（如现象学、诠释学和批判社会理论）的进步。

大约在同一时期，社会学的创始人法国的哲学家奥古斯特·孔德 (1798–1857)试图用他的实证主义学说对理性主义和经验主义进行融合。他提出理论和观察之间存在着循环性依赖。尽管理论可能通过推理来构建，但是这些理论只有经过观察的验证才能说明它们是真实可靠的。是否重视验证是现代科学和哲学及形而上学的分水岭，验证推动了科学方法的发展，这是因为它是检验科学主张的主要途径。涂尔干的社会实证主义（实证主义作为社会研究的基础）和路德维希·维根斯坦的逻辑实证主义进一步发展和壮大了孔德的思想。

在 20 世纪初，德国唯心主义思想学派的解释社会学家（反实证主义者）否认实证主义具有强有力的解释力度。实证主义通常等同于定量研究方法（如实验和问卷）却没有任何明确的哲学保证，而反实证主义主要采用定性研究方法（如非结构化访谈和参与观察）。甚至实

证主义的实践者，如美国社会学家保罗·拉扎斯费尔德（首倡大规模调查研究及对调查数据进行统计分析）也承认实证调查中存在诸如观察者偏见和结构化局限等潜在问题。与此相应的是，反实证主义者强调社会行为必须通过解释性方法来研究，并要求这种研究应基于对个人行为的意义和目的的理解。这些进一步激发了一系列理论的出现，如乔治·齐美尔的象征式互动论，马克斯·韦伯的理想行政组织体系，以及埃德蒙德·胡塞尔的现象学。

在 20 世纪中后期，一些学者对实证主义和反实证主义思想进行了批判性修正。英国哲学家卡尔·波普尔爵士认为人类知识不应建立在无法挑战且固若磐石的知识基础之上，而应建立在一系列不确定的、仅能被证伪而绝对无法被完全证实的推测之上。实证经验正是证伪这些推测或理论的基础。这种元理论的立场（也叫做后实证主义或后经验主义）对实证主义进行了批判和改进。尽管这种观点保留了实证主义学家所倡导的客观真理的概念并强调了科学的研究方法，但是该观点认为证伪错误的信念是有可能的，而证实真理是不可能的。

同样地，反实证主义者也一直受到批判，因为他们仅仅试图理解社会，而没有批评和改变现有社会并使之变得更好。德国哲学家卡尔·马克思和弗里德里希·恩格斯在其所著的《资本论》中最先引入这种思想。《资本论》批评资本社会是不公且低效的，并建议通过阶级斗争和无产阶级革命推翻这种不公正。马克思主义在德国、意大利、俄罗斯和中国推动了社会革命，但是最终并未能实现它所期待的社会公正。由马克斯·霍克海默尔和尤尔根·哈贝马斯在 20 世纪所倡议的批判性研究（也叫做批判理论）保留了类似的批判和解决社会不公的观点。他们指出尽管人们有意识地去改变他们所处的社会和经济环境，但是其能力受到了各种形式的社会、文化和政治管辖的限制。批判性研究试图通过分析当今社会中存在的对立、冲突和矛盾来揭露并批判现状的限制和异化条件，并设法消除异化和控制的起因（即解放被压迫的阶级）。更多关于研究哲学和研究方法的内容将呈现在本书的后续章节中。

第二章

像研究者一样思考

做好研究首先需要重新对大脑进行训练使其能够像研究者一样思考。这就要求从实际观察中找出抽象的事物，有意识地“串起各个点”以识别出隐藏的概念和模式，并将这些模式综合形成具有普适性的理论，以用于初始观察领域之外的其他情境。研究包括在实证层面（进行观察）和理论层面（将观察抽象成普遍的规律和理论）之间的不断反复。这是一种需要多年培养才能形成的技巧，而不是在本科或研究生教程中所能学到的，也不是在行业训练中所能习得的，而这也正是目前为止大多数博士生所欠缺的部分。像研究者一样思考所需的一些在思维上的抽象过程包括分析单元、构念、假设、操作化定义、理论、模型、归纳和演绎等，这些将在本章节中予以阐述。

分析单元

科学研究的分析单元是任何社会科学研究的重要决策之一。分析单元指的是特定的个人、集体或调查所涉及的对象。典型的分析单元包括个人、团体、组织、国家、技术、物体等。例如，如果对人们的购买行为、学习成果或是对新技术的态度进行研究，那么这里分析单元便是个人。如果要研究街头帮派的特点或组织的团队精神，那么分析单元便是这个团体。如果研究的目标在于了解企业如何提高利润率或如何做出好的经营决策，，那么分析单元是企业（尽管决策是由公司内的个人所决定的，这仍被认为是代表公司的决定，而非个体的的决定）。如果研究针对民族文化的差异而实施，那么分析单元便是一个国家。　无生命的物体也可以是分析单元。例如，如果一个研究人员有兴趣了解如何使网页更加吸引用户，那么分析单元是一个网页（而不是用户）。如果想要研究两个组织之间的知识转移是如何发生的，那么分析单元是一对目标（即传递和接收知识的两个组织的集合）。

了解分析单元有时可能会相当复杂。例如，如果想要研究为什么某些社区有很高的犯罪率，那么分析单元是社区，而不是罪犯所犯下的罪行或类似的案件。这是因为询问对象是左邻右舍，而不是罪犯。然而，如果想要比较不同社区的犯罪类型，如杀人、抢劫、袭击等等，分析单元则是犯罪案件。如果要研究为什么罪犯从事违法活动，那么分析单元是个人（如，罪犯）。例如，如果要研究为什么一些创新比其他的更成功，那么分析单元是一项创新。然而，如果研究的是为什么一些组织比其他组织更具有持续创新能力，那么分析单元是组织。因此，同一研究的两个相关研究问题可能有两种截然不同的分析单元。

了解分析单元是非常重要的，这是因为它决定了研究中应该收集什么类型的数据以及从哪里进行收集。如果分析单元是一个网页，则应该从真实的网页中收集关于网页的数据，而不是调查人们如何使用网页。如果分析单元是组织，那么应该测量组织层面的变量，如组织规模、收入、等级或吸收能力等。这些数据可能会有各种各样的来源，如财务记录或是对首席执行官（CEO）的问卷调查。一些变量，如 CEO 报酬，看起来可能是个人层面的变量，但事实上，它也可以是一个组织层面的变量，因为每个组织在任何时候只有一个 CEO 报酬。有时，可能会在较低的层次收集数据，然后再将数据整合到一个较高的层次进行分析。例如，为了研究组织的团队合作，可以调查不同组织内团队的个人成员，再将个人分数进行平均来创建一个团队层次的综合得分，并进而作为团队层次的变量，比如凝聚力和冲突。我们将在下面更深入地考察"变量"的概念。

概念、构念及变量

在第一章中我们提到研究可以是探索性的、描述性的或解释性的，但科学研究倾向于解释性的，这是因为解释性研究可以为所观察到的自然或社会现象寻找可能的解释。解释性研究需要对概念或与物体、事件或人相关的普适性特征进行开发。然而，诸如人、公司或车辆这些对象不能被称作是概念，与他们相关的特定行为，如对移民的态度、公司的创新能力以及车的重量，才可以被看作是概念。

我们在日常对话中会有意识或无意识地使用不同的概念。一些概念从我们共有的语言中发展而来。有时候，我们也会借用其他学科或语言中的概念来解释我们所感兴趣的现象。例如，物理学中的引力这个概念可用于商业环境下描述为什么人们趋向于被吸引到他们更加喜爱的购物地点。同样地，距离的概念也可用来解释两个同等地位的个体之间的社会分离程度。有时，我们也创造新的概念来描述在以前的研究中未述及的独一无二的特点。例如，技术压力是一个全新的概念，指的是一个人被要求学习新技术时面临的心理压力。

概念是抽象的更高形式。一些概念，如人的体重，是精确而客观的，而其他概念，如人的个性，则可能更加抽象且难以形象化。构念是指为了解释一个给定现象所特别选定（或"创造"）的抽象概念。构念可以是一个简单的概念，如人的体重，或是一组相关概念的组合，如一个人的沟通技巧就包括几个潜在概念，如词汇、语法和拼写。前例（体重）是一个一维构念，而后例（个性）则是一个多维构念（即包含多个潜在概念）。在多维构念中，构念和概念的区别更为清晰，高阶抽象叫做构念，而低阶抽象称为概念。然而，在一维构念中，这种区别则较为模糊。

用于科学研究的构念必须有精确、清楚的定义，这样其他人才能够完全理解它意味着什么，而不意味着什么。例如，一个看似非常简单的构念，如收入，可能指的是每月或每年的收入、税前或税后的收入、个人或家庭的收入，因此既不精确也不清楚。有两种类型的定义：词典定义和操作化定义。在我们更为熟悉的词典定义中，构念通常按照同义词进行定义。例如，态度可能被定义为一种性格、一种感觉、或是一种情感，而情感反过来被定义为一种态度。这样的循环定义在科学研究中对于阐述构念的含义和内容不是特别有用。科学研究需要操作化定义，即通过如何进行实证测量来定义构念。例如温度这个构念的操作化定义必须详细说明打算用摄氏度、华氏还是开尔文标度。收入这个构念应该定义清楚是每月还是每年的收入，是税前还是税后收入，是个人还是家庭的收入。可以想象的是，学习、个性和智力这几个构念很难被操作化定义。

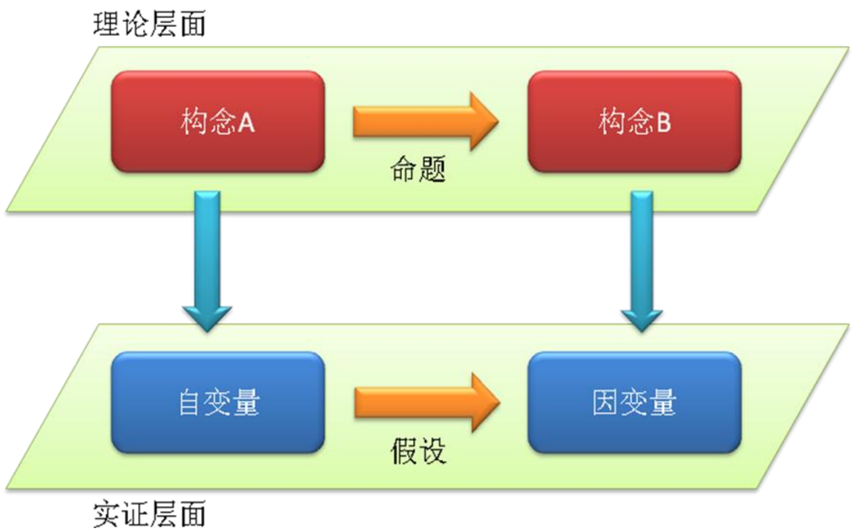


图 2.1 在理论和实证层面的研究

经常与构念联系在一起，有时还可以互换的另一个术语是变量。从词源学上看，与常数（即保持不变）相比，变量的数值可以发生变化（如，由低到高，从负到正等）。然而在科学研究中，变量是对抽象的且不可衡量的构念的一个可衡量的表示。作为抽象的实体，构念不可以被直接测量，因此，我们需要通过变量进行测量。例如，一个人的智力经常以他或她的 IQ 值（智商）衡量，这是在对人实施分析性模式匹配的检测时所产生的指数。在这种情况下，智力是一个构念， IQ 值是测量智力这个构念的变量。不管 IQ 值能否真正测量了一个人的智力是人的猜测 （虽然许多人相信 IQ 值衡量了智力水平），还是取决于其如何测量智力，IQ 值可能是智力这个构念的一个好或差的衡量方式。如图 2.1 所示，科学研究沿着两个层面进行：理论层面和实证层面。构念在理论层面被概念化，而变量在实证层面被予以测量。像研究者一样思考暗示着在这两个层面不断反复的能力。

根据预期用途的要求，变量可分为自变量、因变量、调节变量、中介变量或控制变量。解释其他变量的变量称为自变量，被其他变量解释的变量是因变量，被自变量解释同时也能解释因变量的变量是中介变量，那些影响自变量和因变量之间关系的变量是调节变量。例如，如果我们认为高智商引起学生学习的进步，那么智力就是一个自变量，而学习是一个因变量。可能存在一些其他的随机变量，并不能对某个给定的因变量进行解释，但却对因变量产生一些影响。它们是控制变量，必须在科学研究中控制这些变量。

图 2.2 所示的例子可以帮助读者了解这些不同类型变量的差异。如果我们认为智力影响（或解释）学生学业成绩，那么以衡量智力的 IQ 值作为自变量，而以衡量学业成绩的绩点作为因变量。如果我们认为智力对学习成绩的影响也取决于学生对学习的投入（也就是说，两个同样聪明的学生，更努力的学生比不努力的学生易于取得更好的成绩），那么努力则是一个调节变量。顺便需要提一下，我们也可以将努力看成是自变量而将智力看作调节变量。如果学习成绩被看作是带来高收入的一个中间过程，那么收入是自变量学习成绩的因变量，而学习成绩在智力和收入的整个关系过程中是中介变量。因此，没有任何一个变量可以固定为自变量、因变量、调节变量或中介变量。变量的类型是根据不同构念间的本质联系而确定的。一组相关构念间的所有关系网络称为法理学网络（见图 2.2）。像研究者一样思考不仅包

括对观察进行构念抽象的能力，而且还包括对联系这些抽象构念的法理学网络进行心理想象的能力。

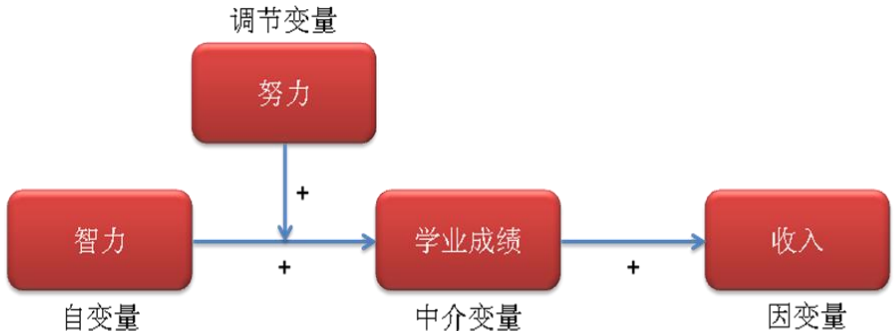


图 2.2 构念的法理学网络

命题和假设

图 2.2 展示了理论构，念如智力、努力、学业成绩和收入，在法理学网络中是如何相互联系的。每一个这样的关系被称为一个命题。在寻求对一个特定的现象或行为进行解释时，只识别出隐含在目标现象或行为中的关键概念和构念是不够的。我们也必须识别并陈述这些构念间的关系模式。这种关系模式就是命题。命题是以声明形式所阐述的暂时性、推测性的构念关系（例如，学生智力的增加引起学习成绩的增加）。这个声明式的陈述必须在实证层面可以检验（至少是间接地），并可在实际观察的基础上判断命题的真假。命题一般从逻辑（演绎）或经验观察（归纳）中得到。

跟构念一样，命题也是在理论层面陈述，而无法直接进行验证。研究人员可以通过检验变量间的关系来间接地对命题进行检验。阐述变量间关系的经验性阐述，被称作为假设（见图 2.1）。在上面的例子中，由于 IQ 值和绩点分别是智力和学习成绩的可操作化衡量指标。智力与学习成绩之间的关系是命题，而 IQ 值与绩点之间的关系是假设。假设可用于实证检验，并且若没有得到实证观察的支持，这个假设则被拒绝。当然，假设检验的目的在于推断相应命题的有效性。

假设可以或强或弱。“学生的 IQ 值与学习成绩相关”是一种典型的弱假设，因为它既没有表明假设的方向性（即，是正相关还是负相关），也没有指明它的因果关系（即，智力影响学业成绩，还是学业成绩影响智力）。“学生的 IQ 值与他们的学业成绩正相关”是较强的假设，它表明了方向性但没有表明因果关系。图 2.2 中的箭头表明了各个假设的方向性。“学生的 IQ 值对学业成绩有积极的影响”是一个更强的假设，它既表明了方向性也表明了因果关系（即智力影响学业成绩，而不是学业成绩影响智力）。

还需要注意的是，科学的假说应该清楚地说明自变量和因变量。之前的假设很清楚地说明智力是自变量（原因），学业成绩是因变量（结果）。很明显，可以进一步证明这个假说是真的（如果更高的智力导致更高的学业成绩）还是假的（如果更高的智力没有影响或导致低的学业成绩）。在这本书的后面，我们将探讨如何采用实证分析测试这样的因果关系。“学

生一般都是聪明的”或“所有学生都能取得学业成绩”这样的陈述不是科学的假设，因为自变量和因变量不明确，并且没有指明两个变量间的方向性，不能证实或证伪。

理论和模型

理论是对某个现象或行为，在一定的边界条件和假设下，提出解释和预测的一整套系统性的相关构念和命题。理论在本质上是理论命题的一个系统集合。命题最多联合了两三个构念，而理论代表了多个构念和多个命题所构成的体系。因此，理论实质上比命题或假说更复杂、更抽象、且有更大的适用范围。

我必须在这里指出，不熟悉科学研究的人经常将理论看作推论或作为事实的对立面。例如，我们常听到教师的课堂教学需要较少的理论和较多的实践知识。然而，事实和实践并不是理论的对立面，在科学意义上，它们都是测试理论有效性的必要组成部分。一个好的科学理论应当被观察到的事实予以充分地支持，也应该有实践价值，然而未被充分界定的理论往往缺乏这些维度。著名的组织研究学者库尔特·列文曾经说过：“没有实践的理论是枯燥无味的，而没有理论的实践是盲目的。”因此，理论与事实（或实践）对于科研而言都是必不可少的。

理论可以解释了社会或自然现象。正如在第一章中所强调的，可能存在好的或差的解释。换句话而言，可能存在好的或差的理论。第三章描述了可以用来评价理论好坏的标准。然而，研究者必须知道没有任何理论是神圣不可侵犯的，理论不应该仅仅是因为由某人提出就接受它，在科学进程中差的理论最终将被具有更好解释力的理论所取代。研究人员的根本挑战是建立一个比其他理论更好的、更全面的理论来解释目标现象。

通常与理论一起使用的术语是模型。模型代表了所研究的体系，并为研究该体系的部分或全部内容而建立。模型与理论的不同之处在于理论用于解释，而模型则用于表达。模型的例子包括数学模型、网络模型和路径模型。模型可以是描述性的、预测性的、或规范性的。描述性模型经常被用在复杂的体系中，用来描述体系中的众多变量和各种关系。预测性模型（即回归模型）用来预测未来事件，比如在风速、风向、温度和湿度等参数的基础上预测气候的模式，或根据参赛队伍的当前组成等来预测篮球比赛的结果。规范性模型主要用于指导我们应如何遵循公认法则来采取行动。模型可以是静态的（表示系统在任何给定时间点的状态），或动态的（表示系统的演化）。

模型的开发过程可能包括归纳和演绎推理。我们在第一章中提到演绎是一个在理论或逻辑推理（基于一系列的初始假设）基础上对现象或行为得出结论的过程。例如，如果某银行对雇员实行严格的道德准则 （假设 1），并且杰米是这个银行的员工（假设 2），那么可以认为杰米遵循了道德准则（结论）。如果假设和推理是正确的，推导得出的结论必定是正确的。

与此相比，归纳则是从一个或多个事实或观察到的证据中得出结论的过程。例如，如果一个企业在一项促销活动中花费了很多钱 （观察 1），但销售额并没有增加（观察 2），那么该项促销活动的效果很差。然而，还可能存在其他的解释，如经济衰退、竞争产品或品牌的出现、或者供应链瓶颈影响了生产。归纳得出的结论只是一个假设，并且可能被证伪。因此，演绎的结论往往比归纳的结论更具有说服力。

如图 2.3 所示，归纳和演绎推理在模型建立的过程中紧密相关。当我们看到一个事实时，并且追问“为什么发生了这件事？”，归纳便由此产生了。我们会提出一个或多个初步解释（假设）来回答这个问题。接着，我们用演绎来减少解释，直到最终基于逻辑和相关前提（即我们对于探究领域的理解）得到最可行的一个解释为止。如果想对已有的理论进行扩展或修正，或是想提出更好的理论，研究人员必须能够在归纳和演绎推理间不断反复，这是科学研究的本质。模型（从最初的理论扩展或修改而来）是这一过程的结果，并且它可以用实证来检验。因此模型是发展理论的重要手段，同时模型帮助决策者在一组给定输入的基础上做出重要决策。理论和模型在理解一个特定现象时所扮演的角色略有不同，但两者对科学研究都很有帮助。

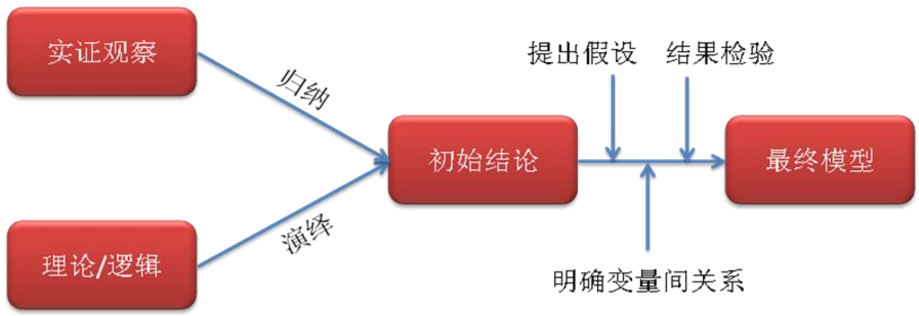


图 2.3 模型建立过程

第三章

研究过程

在第一章中，我们学习到科学研究是通过科学方法获取科学知识的过程。然而，科学研究是如何进行的呢？本章将深入探索科学研究的过程，以及研究过程的设想和结果。

社会研究范式

研究的设计和实施是基于我们用于推理和观测的构思模型或参考框架建立起来的。这些构思模型或框架（信念体系）被称作为范式。范式一词最早由托马斯•库恩 （1962）提出，在其所著的《科学革命的结构》一书中，库恩检验了自然科学的历史，相似的观点同样也适用于社会科学。特别地，不同的人会通过不同的角度理解相同的社会现象，这将束缚他们的思考及推论过程。例如，就政府在公民生活中所扮演的角色而言，保守党和自由党意见窘异，因此两党对如何解决社会问题亦各持己见。比如，保守党可能认为降低税收能增加人们的可支配收入并增加消费，从而扩大业务及就业规模，因此能更好地刺激疲软的经济。相对而言，自由党可能认为政府应更多地直接投资就业项目，如增加在社会工作和基础设施建设中雇员。同样地，西方社会重视个人权利，如隐私权、自由言论权及挟带武器的权利，而亚洲社会则倾向于平衡个人权利与家庭、组织及政府的权利，因此这些国家的政策便少了许多个人主义色彩。这种看法上的差异导致西方人士批判亚洲政府侵犯人权，而亚洲政府亦批判西方社会的自私、贪婪、高犯罪率及个人崇拜。我们的个人范式就像有色眼镜，会影响我们的世界观、我们研究世界的方法及我们构造思想及观察的方法。

范式由于是隐晦的、假设性的及被认为理所当然的，因此通常难以识别。然而，对于范式的识别是调和不同的人对相同社会现象所产生的不同看法的关键。例如，为何自由党认为提高中学教育水平的最好方法是聘请更多教师，而保守党认为教育私有化（学券制）更为有效？这是因为保守党更信任竞争市场（即学校针对教育经费的自由竞争），而自由党更信任劳动力（即更多的教师及学校）。同样地，在社会科学研究中，若想了解为何某项技术在一个组织中成功实施，却在另一个组织中一败涂地，那么透过理性视角去观察世界的研究者从技术不合适或者与使用环境不符等角度提供理性的解释；透过社会视角解决同一问题的研究者则可能试图认为社会缺失（例如使用者培训不当或缺乏管理支持）是主要原因；而透过政治视角观察的研究者则将发掘出影响技术实施的组织政策问题。因此，各自的范式将限制研

究者所测量的概念、观察以及对问题的解释。然而，鉴于社会现象的复杂性，以上范式均具一定程度的正确性，而对问题的全面理解则应采用多种范式。

目前为社会科学研究者所广泛使用的两大范式是实证主义和后实证主义。基于法国哲学家奥古斯特·孔德（1798-1857）的相关成果，实证主义是 20 世纪中叶以前的主要科学范式。实证主义认为科学或知识创造应严格限制于可观察及测量的范围，并仅仅依赖于可直接检验的理论。尽管实证主义的初衷是区分科学与宗教（宗教的概念难以客观地衡量），实证主义导致了对经验主义的盲目信仰，认为观察和测量是科学研究的核心，并拒绝任何可观测事实以外的尝试或推理。例如，由于人类思想及情绪无法直接进行测量，因此这些并不被认为是心理学中的正当议题。实证主义的弊端导致了后实证主义（或后现代主义）在二十世纪中后期的发展。后实证主义认为实际观察和逻辑推理的结合可以得出合理的推论。后实证主义认为科学并非确定的，而是具有不确定性和偶然性，并试图通过探索这些偶然性来更好地理解社会现象。后实证主义学派进一步分为通过主观思想而非客观现实观察世界的主观主义，和认为外界事实独立于个人的思想，并无法被认知的批判现实主义。

伯勒尔和摩根（1979）在《社会学范式和组织分析》一书中认为，社会科学研究者看待和研究社会现象的方法有两个基本的哲学假设：本体论及认识论。本体论指我们如何看待世界的假设（即世界是否主要由社会秩序或持续变化构成）。而认识论指我们研究世界的最好方法的假设（即我们是否应使用客观或主观的方法研究社会事实）。通过使用以上两种假设，我们能够将社会科学研究归类为以下四种（见图 3.1）。

若研究者认为世界主要由社会秩序组成（本体论），试图研究秩序化事件或行为的模式，并认为客观方法（认识论）由于可以独立于进行观察及推断的人士（例如通过实地调查等标准化数据采集工具），因而是研究世界的最好途径，那么他们所采用的范式即为功能主义范式。然而，若研究者相信研究社会秩序的最好办法是参与人员的主观解释（即通过访问不同的参与者并根据研究人员的主观看法对答复差异进行调和），那么他们所采用的范式即为诠释主义范式。若研究者相信世界由急剧变革所构成，并试图通过客观方法对这种变革进行理解，那么他们所采用的是激进结构主义范式。若研究者通过参与人员的主观解释对社会变革加以解释，那么他们所采用的则是激进人本主义范式。

图 3.1 社会科学研究中的四种范式

（来源：伯勒尔和摩根，1979）

与自然科学类似，大部分的社会科学研究采用功能主义范式。功能主义者认为研究社会秩序或模式可以通过其功能模块来予以理解，因此试图通过将问题进行分块，并使用客观方法（例如调查法及实验法）具体研究一个或多个问题来解释社会问题。然而，由于后实证主义的崛起，为数不多却不断增加的研究者试图通过使用主观方法（例如采访及种族研究）来理解社会问题。激进人本主义和激进结构主义代表了社会科学研究中极少的一部分，这是因为科学家主要关注的是理解具有推广性的行为、事件或现象，而不是与众不同的或不断变革的事件。然而，如果你打算研究社会变革，比如在中东国家为何有不断兴起的民主运动，或者为何民主运动在埃及成功了，在利比亚用了很长时间才得以成功，而在叙利亚至今尚未成功，那么激进人本主义是研究这类问题的正确途径。社会和组织现象主要由秩序和变革两个因素所构成。比如，组织成功依赖于清晰定义且正式化的商业过程、工作流程和职业责任，同时也受到了由商业环境下的竞争者、竞争产品、供应商和客户基础所引发的持续变革的限制。因此，对于社会现象（如为什么一些组织比另一些更加成功）的全面理解需要在研究中使用多种范式。

研究过程概述

那么，我们的构思范式是怎样形成社会科学研究呢？任何科学研究的核心是观察、合理化和验证之间的反复过程。在观察阶段，我们对我们所感兴趣的自然现象或社会现象、事件和行为等进行观察。在合理化阶段，我们将我们所观察到的有关谜团的不同部分有逻辑地加以联系，并试图理解或者解释现象、事件或者行为。在某些情况下，我们可能进而构造出新的理论。最后，在验证阶段，我们通过数据收集和分析的科学方法来验证我们所构建的理论，同时，我们也可能会对我们最初的理论进行修改或者扩展。然而，研究设计可能包含不同的模式，一种模式是研究人员先观察，并将所观察的现象进行合理化（归纳研究），另一种模式是研究人员从事先合理化或理论出发，并试图检验这一理论（演绎研究）。因此，观察-合理化-验证的循环过程类似于第一章中所述的归纳-演绎的研究循环过程。

大多数传统的研究往往运用演绎法且其在本质上属于功能主义范式。图 3.2 展示的是研究计划的示意图。图中描述了功能主义研究中一系列的活动，并将其分为三个阶段：探索阶段、研究设计阶段和研究执行阶段。需要注意的是，这种一般性的研究设计并不是任何研究的路径图或流程图。为了适应特定的研究计划的需要，我们应对其进行相应的修改。

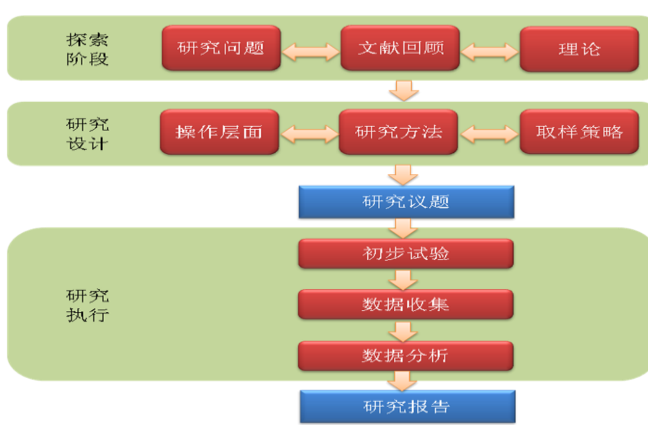


图 3.2 功能主义研究过程

研究的第一阶段是探索阶段。这一阶段包括：探索并选择研究问题，通过阅读已发表的有关领域的文献来了解领域内当前研究现状，识别出有助于回答目标问题的相关理论等。研究的第一步就是识别出一个或多个与特定行为、事件或者现象相关的研究问题。一些研究问题如下：什么因素激发了消费者在未见到或者未体验到相关商品或服务时仍然愿意购买这些商品或服务，我们怎样才能使得高中生更具有创造力，为什么有人会实施恐怖行为。研究问题可以进一步划分为是什么（What）、为什么（Why）、怎么做（How）、何时做（When）等方面。当研究课题能够吸引更广泛的观众（如“企业怎样创新”这个课题比“属于服务业的中国企业怎样创新”这个课题更能引起广泛的兴趣），能够对真实的复杂问题（而不是假设性的问题）进行阐述，并且问题的答案不那么明显，这些研究课题将会更加有趣。过于集中的问题（仅提供是否两种答案选择）往往用处不大，也缺乏趣味。为了能够捕捉到社会现象中的细微差异，广泛的研究问题往往是更好的。需要注意的是，乏味的研究问题最终将导致乏味的研究成果。

研究的下一阶段是要对感兴趣的领域进行文献回顾，文献回顾有三个目的：（1）对该领域的知识现状进行研究；（2）识别该领域中关键的作者、文章、理论和研究成果；（3）发现研究领域中的知识缺口。在网络数据库中利用与感兴趣的领域相关的关键词进行检索，我们就可以获得所应回顾的文献。接下来，研究者必须手工浏览每一篇文章（至少浏览每篇文章的摘要部分），并决定是否应该在文献回顾中包含该篇文章。文献回顾应该适当完整，并不应局限于几种期刊上的文章、几年内的文章，或者运用特定的研究方法的文章。可以以表格的形式对回顾的文献进行总结，也可以用组织框架（概念矩阵）的形式进行总结。有良好指导作用的文献回顾应该说明是否现有文献已经解决了我们最初确定的研究问题（排除了对问题进行重复研究），并会帮助我们找出新的更加有趣的研究机会，或者能够根据现有的研究成果对自己初步确定的研究问题进行修改或变更。文献回顾同时可以为我们提供与目标问题相关的直觉或潜在答案，或者能够帮助我们识别出以前被用来解决相似问题的理论。

因为演绎研究法包含理论检验，下一阶段就是要识别出与目标研究问题相关的一个或者多个理论。虽然文献回顾可能帮助我们发现各种不同的与目标现象潜在相关的概念或者构念，理论有助于识别出哪些构念可能与目标现象是逻辑相关的，并有助于理解它们是如何相关的。抛弃理论可能会导致对关系不大的、稍微相关的或者完全无关的构念进行衡量，这将降低了我们获得有意义的且非偶然性的结果的可能性。在功能主义研究中，理论可以作为实证检验中提出假设的逻辑依据。当然，并不是所有的理论都适用于研究所有的社会现象。必须根据理论与目标问题间的吻合度，以及理论假设与目标问题的假设间的吻合度来仔细挑选相关理论。我们将在下一章中详细说明理论及理论化的过程。

研究过程的下一个主要阶段就是研究设计。即设计完成研究目标和回答研究问题的蓝图。这一阶段包括选择研究方法、对感兴趣的构念进行操作化定义和设计适当的抽样方法等。操作化定义是指设计出对抽象的理论构念进行精确测量的过程。操作化定义是个很大的问题，这是因为社会科学研究中有许多构念，比如偏见、疏远和自由等构念都很难定义，更别说去精确衡量了。操作化定义的第一步是对构念的含义和所研究的变量进行详细的说明（这一过程通常也被叫做构念的概念化）。下一步，研究者可以查阅相关文献，看文献中是否有得到验证的，并可用于衡量相似构念的有效方法，是否可以对已有的衡量方法进行适当的修改，使其能够适用于当前的研究背景。如果文献中没有目标构念的衡量方法，或者已有的衡量方法不适当，或者反映的是与研究者预期所不同的概念，那么就应该为这些构念设计新的衡量工具。这就要求研究者详细准确地说明是怎样衡量预期构念的（比如包含多少事项、有哪些

事项等）。这是个长期并艰难的过程，在新设计的衡量工具被认为是“科学有效”之前，必须进行多次的预测试和修改。在第六章，我们将讨论构念的操作化。

在对构念进行操作化的同时，为了解决相应的研究问题，研究者必须决定采用何种方法来收集数据。这些方法包括如实验研究和调查研究的实证方法，或者如案例研究、行动研究或者将案例研究和行动研究相结合的定性方法。如果选择实验研究，那么怎样进行实验设计呢？如果选择调查研究，那么应选择邮件调查、电话调查、网络调查还是综合调查方式呢？对于复杂的、不确定的及多层面的社会现象而言，多种研究方法综合使用可能更加合适，因为这将有助于利用每种研究方法的独特优势，并获得通过单一研究方法所无法得到的见解。

研究者必须选择数据收集的目标总体，并确定从总体中进行样本选取的抽样方法。比如，是否应在企业内部进行个人调查、企业调查或者工作组调查？选择哪种类型的个人或者企业作为调查对象？抽样策略与研究问题中的分析单元紧密相连。当选择样本时，应该合理仔细地进行选择，以避免出现有偏样本（比如，根据研究者的便利来进行样本选择）及有偏观察。下面的章节将详细介绍抽样方法。

在这一阶段，写一份研究计划是非常好的办法。计划中应详细介绍在研究过程中所做出的决定，及每个决定背后的基本原理。在这份包含多个部分的计划书中，应该说明你想要研究的研究问题、为什么选择这个研究问题、有关这方面研究的已有知识、支持待检验假设的理论、怎样衡量构念、选择哪种研究方法、为什么选择这种方法和所选择的抽样方法等。提供经费支持的机构也会要求申请者提供研究计划，以便于其选择最好的计划进行资助。即使研究项目未成功申请到经费资助，研究计划也有助于寻求其他研究者的反馈，并有助于在数据收集前识别出研究项目中的潜在问题（比如，研究中是否遗漏了一些重要的构念）。因为在数据收集完成后再修改关键的问题往往太迟了，所以最初的反馈是非常重要的。

在决定要对谁进行研究（研究对象）、测量哪些构念（概念）和怎样收集数据（研究方法）后，研究者现在应准备进入研究执行阶段。这一阶段包括测量工具的预测试，数据收集和数据分析等过程。预测试是非常重要的，它可以发现研究设计或者研究工具的潜在问题（比如，目标样本群体是否理解他们被问到的问题），并保证研究中所运用的测量工具能够可靠有效地衡量目标构念。测试样本通常是目标总体的子集。在成功进行预测试后，研究者就应该进入通过总体抽样进行数据收集的阶段。所收集的数据根据研究方法的差异分为定性数据和定量数据。

在完成数据收集之后，要对数据进行分析和解读以得出与研究问题相关的结论。根据数据的类型（定量数据或者定性数据），数据分析可对定量数据运用统计方法进行定量分析，如回归分析法或者结构方程模型分析法等，或对定性数据进行定量分析，如对数据进行编码等，或对定性数据进行定性分析，如内容分析法等。

研究的最后阶段是准备最终研究报告，以研究论文、专题或者专著的形式对整个研究过程进行记录。报告应该描述研究过程中的所有细节（如使用的理论、选择的构念、运用的测量方法、研究方法、抽样过程等），并说明其原因和每个研究阶段的研究结果。研究过程必须详细进行描述，以便于让其他研究者能够重复该研究，检验相关研究结果，或者评估推论是否得到了科学的认可。当然，准备好研究计划将大大简化并加快撰写研究报告的过程。需要注意的是，研究者应该记录下研究的过程和结果，这些记录文档对于科学的发展是必要的。

研究中的常见错误

研究过程中充满着问题和陷进。对于一个初学者来说，其往往要投入大量的时间和精力进行研究，却在完成研究后发现研究结果并没有充分地回答研究问题，或者研究并不具备“可接受的”科学质量（如发表期刊论文）。下面将描述一些常见的错误。

动机不足的研究问题。通常，我们选择所感兴趣的问题进行研究，但是这些研究问题对整个科学界来说却没有太大的吸引力。我们在前文中提到，研究是为了获取关于自然或社会现象的新知识或新见解。因为在研究过程中研究者需要投入大量的时间和精力，所以研究者必须保证（并使他人相信）其所试图回答的研究问题是现实问题（而非假设性问题），这些研究问题会对很多人产生影响，并在先前的研究中未被充分地加以解决。

盲目跟随研究潮流。另一个常见的错误就是在有限的学术生涯中追求流行的研究主题。一个典型的例子就是研究当下普遍的技术或实务。因为研究需要几年的时间去完成并发表，所以在研究完成并提交发表时，人们对流行的研究问题的兴趣很可能已经消退。比较好的策略是选择那些多年来被人们坚持研究，但似乎没有人能得到合理答案的“永恒性的”研究问题加以研究。

不可研究的问题。单单使用基于观察的证据或利用目前公认的方法或程序可能无法充分回答一些研究问题。研究者最好避免研究这类问题。然而，一些不可研究的、定义含糊的问题可以适当地加以修改并调整为可研究的、定义明确的问题。

偏爱的研究方法。许多研究者为了使研究问题与他们所偏爱的研究方法（比如调查研究）相符，他们往往倾向于对研究问题进行改动。这是一个不幸的趋势。研究者应该选择与研究问题最吻合和研究方法，而不是颠倒过来。

盲目的数据挖掘。一些研究者倾向于先展开数据收集，再考虑如何对这些数据进行处理。需要注意的是，数据收集仅仅是漫长而精心准备的研究计划、研究设计和研究实施过程中的一个步骤。实际上，在数据收集之前，还需要完成多项步骤。如果研究者不经过精细的计划而直接跳到数据收集这一步，那么收集的数据很可能是无关的、不完美的、甚至是无用的，这样数据收集的努力就可能完全白费了。研究计划或者研究设计方面的缺陷，特别是缺乏有趣的研究问题，是不能通过丰富的数据来弥补的。

第四章

科学研究理论

正如我们从之前的章节中所了解到的，科学是科学方法衍生出的以理论集合为代表的知识体系。 在本章中，我们将研究什么是理论，为什么我们在研究工作中需要理论；什么是理论的构成要素，如何评估理论以及我们应如何在研究中运用理论。同时，我们将以社会科学研究中五个常用理论为例来阐述前面章节所提及的问题。

理论

理论是对自然或社会行为、事件或现象的解释。从更专业的角度来说，科学理论是一个由构念（概念）和命题（构念间的关系）组合而成的系统，对在某些假设和边界条件下的现象做出符合逻辑、系统且一致的解释（Bacharach 1989）[1]。

理论应当解释事件发生的原因，而不是仅仅描述或预测事件。需要注意的是，预测事件或行为只需要通过设置一系列的预测变量，而不需要解释事件发生的原因。例如，市场分析师可以利用市场公告、主要公司的财务报表以及来自美国联邦储备系统或其他机构的最新数据，并基于先前所观察到的相关性来预测股市的波动。预测仅仅需要这种相关性分析。与此相反，解释事件则建立在理解其原因或因果关系的基础上。建立因果关系需要三个条件：（1）两个构念之间存在相关关系；（2）时间优先（原因必须先于结果发生）；（3）排除备选假设（通过测试排除）。科学理论与神学，哲学或者其他对于事件的解释不同，这是因为科学理论可以通过科学的方法进行检验，甚至被证伪。

解释包括个案研究和通则研究。个案解释是指对某一情境或事件的详尽说明。例如，你考试成绩很糟糕是因为：（1）你忘记了考试日期；（2）交通堵塞导致你考试迟到；（3）你在考试中途觉得惊慌；（4）考试的前一天你工作到很晚因此没时间复习，或者甚至是因为（5）你的狗吃掉了课本。解释可以是详细、精确和有效的，但是它们可能不适用于其他相似的情境（即使是对于同一个人），因此个案解释不具有普遍性。相比而言，通则解释则用于于解释一系列的情境或事件，而不仅仅针对某一情境或事件。例如，学生考试成绩很差是因

[1]Bacharach, S. B. （1989）. "Organizational Theories： Some Criteria for Evaluation," *Academy of Management Review* （14：4）, 496-515.

为他们没有投入足够多的时间来准备考试，或者由于过度紧张、注意力下降以及其它的一些心理紊乱因素所造成。通则研究的目的是得出适用于跨情境、跨事件或跨人群的结论，因此它没那么精确、完整和详细。但是，通则研究可以用较少的变量来解释问题。由于理论也同样用于对事件、行为或者现象的模式做出普遍性解释，因此就其属性而言，理论解释通常是通则性解释。

在了解理论的同时，了解理论不是什么也是尤为重要的。理论不是数据、事实、类型学、分类学或实证结果。一堆事实不是理论，就如同它们只是一堆砖瓦，而不是房子本身。同样的，构念的特定集合（一种拓扑结构）也不是理论，因为理论不仅仅是构念，还包括了命题、解释和边界条件。理论基于逻辑推理而不是观测，在概念层面操作，而数据、事实和调研结果在实证或观测层面操作。

在研究中运用理论有很多好处。首先，理论为自然或社会现象的发生提供潜在的逻辑推理。它解释目标现象的主要驱动因素和关键结果是什么，为什么是这些因素，以及推动目标现象的潜在过程是什么。其次，理论提供了意义构建，它帮助我们在理论的框架下综合实证结果并通过不同的研究，发现影响两个构念之间关系的因素来协调矛盾的结果。再次，理论通过界定构念和关系，为今后的研究提供指导。最后，理论可以缩小理论之间的差异以及引起对现有理论的再度评估，有利于知识的积累。

然而，理论也有局限性。构念及其构念间的关系是有限的，因此，在对现实的简化说明上，理论无法对现状做出足够的说明。理论往往用于简化说明，然而现实却更加复杂。此外，理论可能限制了研究者的视野，使他们忽视了没有被理论界定的重要概念。

理论的构建模块

大卫·惠顿（1989） 认为理论的构建模块有四个：构念、命题、逻辑和边界条件/假设。念构阐述了理论是“什么”（例如，什么是解释现象的重要概念），命题阐述了理论是“怎样的”（例如，构念是怎样互相联系的），逻辑阐述了“为什么”（例如，为什么这些构念相互联系），边界条件/假设阐述了“谁，什么时候以及在哪儿”（例如，在什么情况下这些构念和关系才能实现）。 尽管我们已经在第二章讨论过构念和命题，为了结构的完整性，我们将再阐述一次。

构念是高度抽象的概括，用于明确地描述感兴趣的现象。在第二章中我们了解到构念可以是一维的（包含单一构念），例如体重、年龄，或者是多维的（包含多重构念），例如个性、文化。有些构念，例如年龄、教育和公司规模，都是容易理解的；但是有些构念，例如创造力、偏见和组织的敏捷度，更加复杂和深奥；还有些构念，例如信任、态度和学识，代表着随时间改变的趋势，不是一个稳定的状态。尽管如此，所有的构念必须有清楚明白的定义，并详细指明如何度量构念以及在在哪个层面（个人、集体，组织等）分析构念。度量抽象构念的代表就是变量。例如，智商测试的分数是变量，它可以度量智商这个抽象的构念。前面也提到过，科学研究一般沿着两个方向：理论方向和实证方向。构念是理论层面的，然而变量却是实证层面的度量和观测。除此之外，第二章中我们讨论过，变量可能是独立变量、因变量、中间变量和调节变量。图 4.1 展示的是构念（理论层面）和变量（实证层面）的差别。

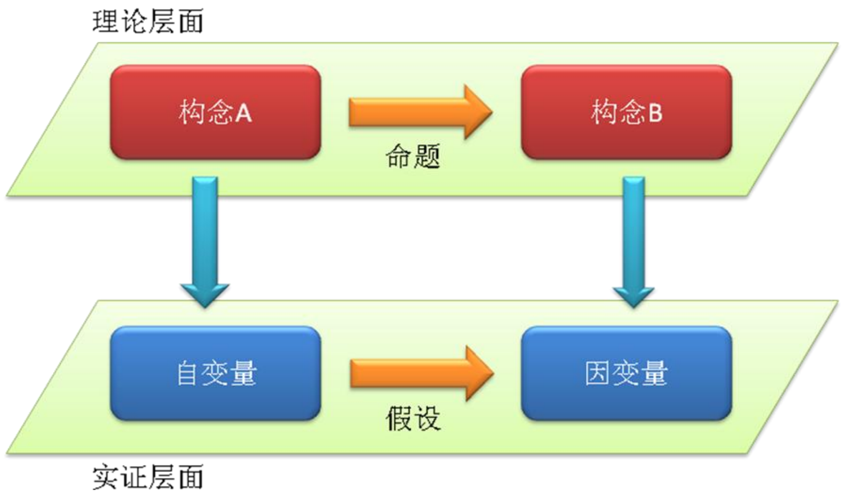


图 4.1 理论概念和实证概念的区别

命题是基于演绎逻辑对构念的相关性分析。命题是陈述的形式，应该完美地指出因果关系（例如，如果 X 发生，Y 也将随后发生）。值得注意的是，命题可能是推测出来的，但是必须经过验证，如果实证观测不支持推测，该命题就被丢弃。然而，与构念相似，命题也是理论层面的表述，只能通过考察构念的度量变量之间的相关关系来验证。假说，即命题的经验公式，可以表述成变量间的相关关系。

逻辑是理论的第三个构建模块，是证明命题的基础。逻辑像胶水一样，把理论构念粘合在一起，为构念间的关系提供意义和关联。逻辑也是存在于理论核心的解释。没有逻辑，命题将会是临时的、任意的甚至无意义的命题，无法形成紧密结合的命题体系，这是所有理论的核心部分。

最后，理论都受到价值、时间和空间等假设的限制，边界条件限定了理论适用于哪些方面，不适用于哪些方面。例如，在推测人群行为时，经济理论假设人是理性的（或者接近于理性），以及运用成本收益分析使效用最大化。然而，政治学理论假设人具有政治性，而不是理性。具体而言，人们试图在专业或者个人的领域内最大化自身的权利和对他人的控制。考虑到假设的不同，经济学和政治学的理论无法直接比较，如果研究者的目标是了解组织的权利结构和改革，那么就不应该运用经济理论。同样地，理论可能有隐含性的文化假设（例如，它们是否适用于个人主义文化或者集体主义文化）、时间假设（它们是否适用于人类行为的前期或者后期）以及空间假设（它们是否适用于某个特定的地点）。如果一个理论被合理地使用或者检验，那么我们必须充分了解所有构成理论边界的隐含性假设。然而，理论家们很少清楚地描述隐含性假设，这往往会造成研究中理论的误用。

优秀理论的特征

理论是对复杂社会现实的局部解释说明。就其本身而言，有好的解释也有坏的解释，因此有好的理论也就有坏的理论。如何衡量某个理论的好坏？不同的研究者提供了不同的标准，下面我们列出一些重要的标准：

逻辑的一致性：理论的构念、命题、边界条件和假设是否在逻辑上具有一致性？如果一个理论的构建模块不具有一致性（例如，理论假设理性，但是其构念的定义蕴含着非理性），那么这个理论是一个不好的理论。

解释力：某个给定的理论能够多大程度解释（或预测）现实？好的理论显然比其他理论更好地解释了目标现象，在回归方程中通常用方差（R^2）来度量。

可证伪性：英国哲学家卡尔波普尔在 20 世纪 40 年代曾指出，理论若是有效的，那么它必须具备可证伪性。可证伪性确保了当实证数据与理论命题不相符时（研究者可以使用实证数据来检验理论），理论是可被证明为误的。换句话说，若无法通过实证检验，这个理论就是错误的。重复的表述，例如"气温高的天是热天"无法通过实证检验，因为热天被定义（及衡量）为气温高的天，因此，像这样的表述不能被视作理论的命题。可证伪性要求存在反义的解释，这使我们可以充分度量构念。然而，理论具有可证伪性并不代表着理论应该是错的。如果一个理论基于实证证据确实是错的，那么从一开始它就是个不好的理论。

简约性：简约性检验了利用少数变量从多大程度上解释现象。这个定义出自 14 世纪英国逻辑之父奥卡姆的威廉（William of Ockham，亦被称作"Ockham's razor" 或 "Occam's razor"），他认为在众多能充分解释被观测证据的理论中，最简单的理论（运用最少变量或者最少设定的假设）是最好的。解释复杂的社会现象总是牵涉到众多的变量。然而，这却违背了建立理论的目标——简化和归纳事实。简约性与给定理论的自由度有关。简约的理论有着更高的自由度，这使得该理论更易于跨环境、跨假设和跨人群的归纳。

理论化的方法

研究者如何构建理论？斯坦菲尔德和富尔克（1990）[2]提出了四种方法。第一种，在观测到的事件或行为模式上归纳出理论。这种方法被称作"构建扎根理论"，因为理论根植于经验观测值。这种方法不仅困难，而且受限于研究者的主观偏见，有可能导致伪相关性。此外，除非研究者能够提供不同观测模式之间一致性的解释，否则观测某个事件的模式无法构建出一个理论。我们将在后面有关定性研究的章节中讨论扎根理论的方法。

第二种构建理论的方法是利用预定义框架，针对与目标现象潜在相关的不同的预测变量，构造一个自下而上的概念性分析。例如，这样的框架可能是一个简单的投入产出框架， 研究者首先寻找不同类别的输入，例如个体的、组织的或者与感兴趣的现象相关的技术因素（产出），然后描述把这些因素连接起来的过程。这也是一种归纳性方法，基于经验性观测和/或先前的知识，对研究者的归纳能力要求很高。

第三种构建理论的方法是扩展或者修改现有的理论，用来解释新的内容。例如将个人学习的理论扩展到组织学习的理论。在对现有理论进行扩展时，原有理论中特定的构念、命题、和/或边界条件可能被保留，其余的将会被修改以适用于新的内容。这种演绎的方法对以前的理论学家提出的大量社会科学理论进行了补充，是一种构建理论的有效方法。

[2]Steinfield, C.W. and Fulk, J. （1990）. "The Theory Imperative," in *Organizations and Communications Technology*, J. Fulk and C. W. Steinfield （eds.）, Newbury Park, CA： Sage Publications.

第四种构建理论的方法是总结出两种环境（旧环境和新环境）的结构相似性，将现有理论应用于新环境。这种方法依赖于类推，是利用演绎方法构建理论中最为创新的一种方法。例如，马库斯（1987）[3]通过类推分析，总结出核爆炸与不受控的网络发展或电子商务的相似性，提出网络增长的临界物质理论。就像核爆炸需要放射性材料作为临界物质来维持一样，马库斯认为网络需要用户作为临界物质来维持自身的增长，没有这样的临界物质，网络将会萎缩而最终失败。

社会科学理论的例子

本节，我们将简要概述几个来自不同社会科学领域的理论。这些理论运用一系列构念、命题、边界条件、假设和深层逻辑来解释不同类型的社会行为。本节仅对这些理论进行简单的介绍，若想详细和深刻的理解这些理论，建议读者查阅这些理论的原始来源。

代理理论。代理理论（又称委托与代理理论），是组织经济学中的经典理论，由罗斯（1973）[4]提出，解释了两个目标不一致的当事人之间的关系（例如雇员和雇主之间，组织主管和股东之间，买方和卖方之间）。代理理论的目的是约定最佳的契约和条件，最小化由于双方目标不一致造成的影响。理论假设的核心是人都是利己主义的个体，理性有限且不愿意承担风险。该理论适用于个人或者组织层面。

代理理论中的双方当事人即为委托人和代理人，委托人雇佣代理人代替他们执行某项任务。委托人会快速和有效地完成指定任务的目标，而代理人的目标是用自己的节奏完成任务，规避风险，在共同利益中寻求个人利益（例如个人薪金）。因此，两者的目标是不一致的。信息的不对称性自然而然导致委托人无法观测代理人的行为或者无法准确度量代理人的能力。这样的不对称性可能导致代理出现问题，包括代理人没有付出相应精力去完成任务（道德风险问题）或者歪曲自己的专业技能以获得代理任务，却无法像委托人期待的那样完成任务（逆向选择问题）。典型的契约是基于行为的，例如按月支付的工资，无法克服这些问题。因此，代理理论建议使用基于结果的契约，例如在任务完成后支付佣金或者基于行为与结果激励的混合契约。虽然雇员的工资是一个基于行为的契约，但是雇员的股票期权计划是基于结果契约的。代理理论也建议雇主使用工具，来提高基于行为契约的效能，例如引入监控机制的投资（如雇佣监察人），来克服由于道德风险导致的信息不对称；根据代理人的表现而定的续约契约（评定表现使契约变成部分结果导向）；或者改进指定任务的结构使其更加程序化，以便于观测。

计划行为理论，由阿耶兹（1991）[5]提出，该理论是一个广义的社会心理学范畴内的个人行为理论，用于研究范围广泛的个人行为。它假设个人行为基于理性思考和社会压力，表现出有意识的理性选择。计划行为理论假设行为是基于个人对该行为的意图，行为是一个函数，与个人对该行为的态度、行为的主观规范以及行为控制感知相关（图 4.2）。态度是指个

[3] Markus, M. L. （1987）. “Toward a ‘Critical Mass’ Theory of Interactive Media： Universal Access, Interdependence, and Diffusion,” *Communication Research* （14：5）, 491-511.

[4]Ross, S. A. （1973）. “The Economic Theory of Agency： The Principal’s Problem,” *American Economic Review* （63：2）, 134-139.

[5]Ajzen, I. （1991）. “The Theory of Planned Behavior,” *Organizational Behavior and Human Decision Processes* （50）, 179-211.

人对某项行为所持的整体正面或负面的感觉，态度可通过综合某人对某项行为所致的结果的信任来评价，态度权重由对结果的赞许度来衡量。主观规范是指个人对于是否采取某项行为所感受到的社会压力，表现为不同团体（例如朋友、同事、公司主管）预期规范的加权组合。行为控制是指为限制某种行为，个人对于内在和外在的行为控制感知。内在控制可能包括个人执行目标行为的能力（自我效能），然而外部控制是指在执行行为过程中对外部资源的获取（推进条件）。计划行为理论也提到，有时想要执行某种行为，但是缺乏需要的资源，因此，行为控制除了可以间接影响行为（以意向作为中介）之外，还可假定行为控制直接对行为造成影响。

计划行为理论是早前理性行为理论的扩展，理性行为理论认为感知的主要驱动力是态度和主观规范，而非行为控制。阿耶兹将行为控制加到计划行为理论中，用以解释人们可能无法完全控制自身行为的情况（例如在浏览网页时没有高速网络接入）。

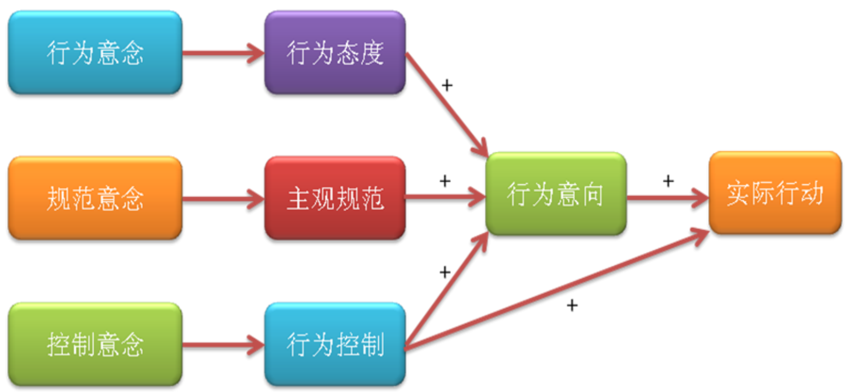


图 4.2 计划行为理论

创新扩散理论。创新扩散理论（IDT）是通信领域内具有开创性的理论，用以解释创新如何在众多潜在采纳者中扩散。埃弗雷特•罗杰斯（1962 年）基于 508 个扩散研究观测值创建了该理论，但是最早研究这个概念的人是法国社会学家塔尔德。该理论中的四个关键元素分别为：创新、通信渠道、时间和社会系统。创新，包括新技术、新实践或者新想法，采纳者可以是个人或组织。在宏观（群体）层面，创新扩散理论认为创新扩散是一个通信的过程。在社会系统中，人们通过通信渠道（例如大众传媒或者早前的采纳者）了解新的创意和该创意的潜在好处，并被说服采纳该创意。扩散是一个时间过程，在早期采纳者中扩散过程缓慢，当创新被主流人群接受时，扩散过程开始加速，当采纳人群饱和的时候，扩散过程最终减速。因此累积的采纳曲线是一个 S 形，如图 4.3 所示，采纳者的分布符合正态分布。所有采纳者都不同且根据采纳的时间，采纳者被分为创新者、早期采纳者、早期众多跟进者、后期众多跟进者以及滞后者。扩散的速度取决于社会系统的特征，例如意见领袖（其意见受到其他人重视的专家）和变革推动者（影响其他人行为的人）的出现。

在微观（采纳者）层面，罗杰斯（1995）[6]认为创新采纳过程包括五个阶段：（1）获知：当人们通过大众传媒或社交渠道来了解创新，（2）说服：当人们被早期采纳者劝服采纳创新，（3）决策：人们决定采纳或拒绝创新，（4）实施：人们初步运用创新的程度不同，

[6]Rogers, E.（1962）. *Diffusion of Innovations*. New York： The Free Press. Other editions 1983, 1996, 2005.

（5）确认：人们决定继续使用创新（见图 4.4）。该理论假定了 5 个创新特征来描述采纳者的决定：（1）相对优势：相对于早期创新，现有创新预期收益高，（2）兼容性：创新匹配采纳者的工作习惯、信念和价值的程度，（3）复杂性：人们学习和使用创新的难易程度，（4）可试用性：创新可以被试验的程度，（5）可观测性：创新结果可清晰被观测的程度。最后两个特征被许多创新研究忽视。复杂性与创新采纳呈负相关关系，然而其他四个特征与创新采纳呈现正相关。创新采纳也取决于个人因素，例如采纳者的风险倾向、受教育程度、世界大同主义和通讯的影响。早期采纳者一般偏好风险，受到良好的教育以及倾向于依赖大众传媒了解创新信息，然而后期采纳者更多的依赖人际关系（例如朋友或家人）作为他们信息的主要来源。创新扩散理论由于“创新偏见”（即认为所有创新都是好的，都会在整个群体间扩散）而受到批评，因为该理论不允许无效的创新，例如时尚，在没有被整个群体采纳或者被更好的创新所替代的情况下会快速消亡。

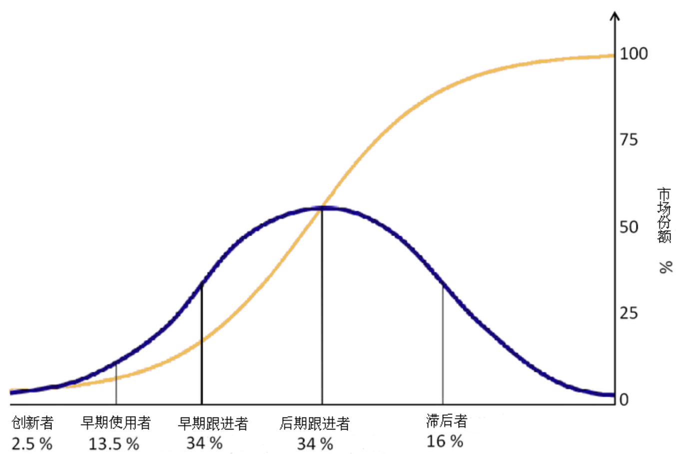


图 4.3 S 形扩散曲线

图 4.4 创新采纳过程

精细加工可能性模型。由佩蒂和卡乔鲍（1986）[7]提出，精细加工可能性模型（ELM）是心理学领域内关于态度的形成或改变的双过程理论。它解释了针对特定对象、事件或行为，如何影响个体近而改变他们的态度以及这些改变策略的相关效用。精细加工可能性模型假定个人态度可由两种影响路径来描述：中心路径和外围路径。它们的区别是大量有用信息处理不同或者精细加工不同（图 4.5）。中心路径要求个人在形成一个对目标对象的判断前，必须思考与目标事件相关的论点信息，详细地查看论点的利弊。在外围路径里，主体依赖外部“线索”，例如早前的用户数量、专家的推荐或者推荐的喜爱程度，而不是依靠论点的质量来决定他们对目标对象的态度。外围路径没那么多认知要求，我们通过运用论点质量和外围线索这两个构念，对精细加工可能性模型中的态度改变路径进行典型操作。

[7]Petty, R. E., and Cacioppo, J. T. （1986）. *Communication and Persuasion： Central and Peripheral Routes to Attitude Change*. New York： Springer-Verlag.

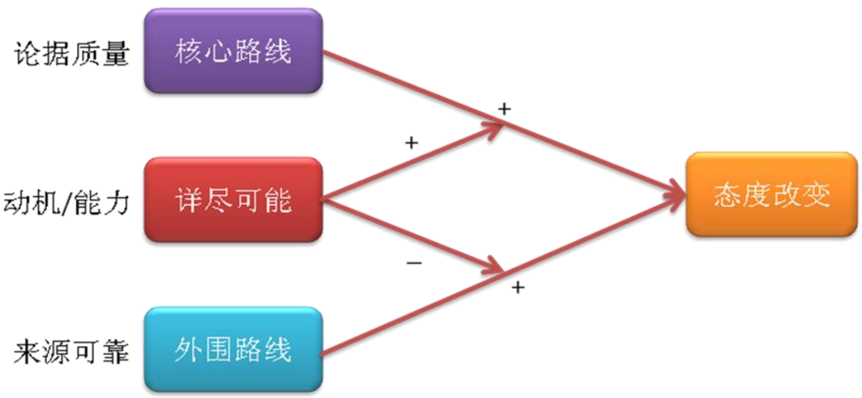


图 4.5 精细加工可能性模型

人们受中心路径影响还是外围路径影响取决于他们详尽描述论点优点的能力和动机。这种能力和动机被称作精细加工可能性。具有高精细加工可能性（高能力和高动机）的人更容易仔细地处理信息，因此受到论点质量的影响更大。然而那些具有低精细加工可能性的人更容易受到外围线索的影响。精细加工可能性是环境特征而不是个人特质。例如，医生在诊断和治疗病症时可能使用中心路径（凭借他/她的专业技能），但是在通过汽车修理工了解他的车况时使用的是外围线索。就这点而言，该理论对新产品或新想法甚至社会变革中如何改变人们态度有着广泛的启示。

一般威慑理论。两位 18 世纪的功利主义哲学家，贝卡里亚和杰里米·边沁，提出了一般威慑理论（GDT）来解释犯罪和减少犯罪的方法。一般威慑理论解释了个体做出异常、反社会或者犯罪行为的原因。该理论认为，人们都是理性的（对于附和行为和离经叛道行为），经过理性的成本效益估算后，他们大量地选择了离经叛道行为。人们自然会选择使效用最大化的行为，离经叛道的选择会增加个人收益或者快感。因此，加大对犯罪行为的惩罚力度以及提高犯罪份子的逮捕率能控制犯罪行为的发生。迅速、严厉和必要的惩罚是一般威慑理论的核心构念。

然而，经典的犯罪学实证主义研究一直在寻求产生犯罪行为的普遍原因，例如贫穷、缺乏教育、心理状况并对如何改造犯罪分子提出推荐策略，例如向犯罪分子提供职业培训和医疗服务。一般威慑理论重点研究犯罪分子的决策过程和影响这一决策过程的情景因素。因此，犯罪分子的个人情景（例如他的个人价值、富裕程度和对金钱的需求）和环境情景（例如目标被保护的程度、当地警察的效率、犯罪分子的逮捕率）在犯罪决策过程中扮演着重要角色。一般威慑理论的研究重点不是如何改造犯罪分子或避免犯罪行为，而是如何降低犯罪行为的吸引力来阻止犯罪行为的发生。为了这个目的，"目标强化"，例如安装门栓和增强自卫技能；法律威慑，例如减少某些犯罪行为的假释、"三击服刑制"（源自棒球术语三击出局，指犯罪分子犯罪满三次就强制入狱监禁，哪怕第三次犯罪很轻或者不值得监禁）、死刑；利用邻里监督等方式增加逮捕的机会；针对毒品和黑社会犯罪成立特别任务小组，增加警力巡逻和犯罪相关教育课程例如发布"犯罪分子将会被起诉"的高曝光率公告都能够有效阻止犯罪行为的发生。一般威慑理论不仅对传统犯罪有意义，而且对当代白领犯罪，例如内部交易、软件盗版和非法音乐分享也有着重大意义。

第五章

研究设计

在实证研究中，研究设计是指全面的数据收集计划。研究设计是实证研究的“蓝图”，它旨在回答特定的研究问题或者检验特定的研究假设，并且必须至少详细说明如下三个过程：（1）数据收集过程，（2）量表开发过程，和（3）采样过程。这一章主要介绍数据收集过程（一般也叫做“研究设计”）， 后续两章介绍量表开发和采样过程。在第 9-12 章，我们会对数据收集过程作更详尽的描述。

大致说来， 数据采集方法可分为两大类：实证与诠释。实证方法，例如实验研究和调查研究，的目的在于检验理论（或假设）；而诠释方法（如行动研究和人种志研究）的目的在于构建理论。实证研究采用演绎的方法，从一个理论开始并通过经验数据来检验理论假设。相反，诠释研究采用归纳的方法，通过所观察到的数据来试图获得与某个现象相关的理论。通常，这些方法被错误地等同于定量和定性研究。定量和定性的方法是指数据收集（定量数据包括数值型数据、指标型数据等，而定性数据包括访谈、观察等）和数据分析（即利用定量技术，如回归，或定性方法，如编码）的类型。实证性研究主要采用定量数据，但也可以使用定性数据。诠释性研究很大程度上依赖于定性数据，但有时也会使用定量数据。采用定性和定量数据相结合的方法可以帮助研究人员获取对某些复杂社会现象的独特见解。这种见解并不能通过单独使用任何一类数据来获知。因此，使用定性和定量数据的混合研究模式常常是非常必要的。

研究设计的关键属性

研究设计的质量取决于四个关键设计属性：内部效度，外部效度，构念效度和统计结论效度。

内部效度，也被称为因果关系，它研究因变量的变化是否确实由假定的自变量的变化所造成，而不是由与研究背景毫无关联的外生变量所引起的。因果关系的存在需要三个条件：（1）原因和结果相互关联（也就是说，如果原因发生了，那么也会产生相应的结果；如果原因不发生，那么也不会产生相应的结果）；（2）时间优先：原因必须先于结果而发生；（3）没有其他看似合理的解释（或伪相关）。某些研究设计，例如实验室实验，具有较高的内部效度。这是因为实验室实验能够在控制外生变量影响的同时，通过某些控制条件对自变

量（即原因）进行操控，并在控制施加的一段时间后观察结果（因变量）的变化。其他的研究设计，如现场调查，由于无法操控自变量（即原因），而且原因和结果都是在相同的时间点观察到的（这与原因必须优先于结果的条件相违背，并造成似乎是预期的结果影响到原因，而不是原因影响结果），因而缺乏内部效度。虽然实验研究与其他研究方法相比具有较高的内部效度，但是实验研究易受到历史记录、测试、试验设备、回归和其他因素的影响。我们将在实验设计这一章中对该内容进行详细地阐述。尽管如此，不同的研究设计在各自的内部效度水平上差异还是很大的。

外部效度或可推广性主要指所观察的相关性是否可以从样本推广到总体（总体效度），或是否适用于其他人、其他组织、其他情境，或其他时间（即生态效度）。比如，从美国金融公司样本推出的结论是否可以推广至所有金融公司（总体效度）或美国的其它公司（生态效度）？调查研究的数据来源于各种各样的个人、企业或者其他分析单位。这种研究往往比实验研究更具普遍性。这是因为实验研究会严格地人为控制外生变量，但在现实环境下这些外生变量是不可控的，因而导致实验研究的结论很难得到推广。各种研究设计的内部效度和外部效度如图 5.1 所示。

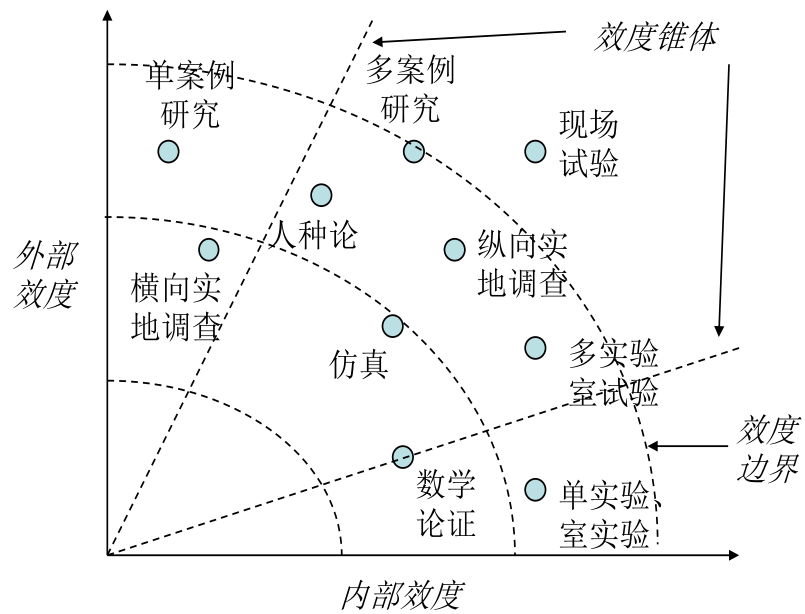


图 5.1. 内部效度和外部效度

一些研究人员认为内部效度和外部效度之间存在着权衡关系，即只有在牺牲内部效度的情况下才能获得高的外部效度，反之亦然。但是，也未必总是如此。如现场实验、纵向现场调查、以及多案例研究等研究设计就可以同时实现较高的内部效度和外部效度。就我个人而言，我更倾向于兼具合理的内部效度和外部效度的研究设计，即落于如图 5.1 所示的锥形区域内的研究设计。但这并不能表明位于锥形区域之外的研究设计的价值或有用性更低。研究者最终将根据个人的偏好和能力，以及他想获得的内部效度和外部效度水平来选择研究设计。

构念效度衡量的是给定的量表在多大程度上测量了该理论构念。社会科学研究中的许多构念，如移情、变革阻力以及组织学习等都是难以界定的，更不用说去衡量了。例如，构念效度必须保证对移情这一构念的衡量确实反映的是共同感受而不是同情怜悯。要做到这一点可

能会比较困难，这是因为这些构念在意义上比较相近。在下一章节中，我们描述了如何在实证研究中通过相关分析和因子分析来评估构念效度。

统计结论效度检验的是在多大程度上由统计方法所推导出的结论是有效的。例如，统计结论效度会分析假设检验是否使用了正确的统计方法、所使用的研究变量是否符合统计检验的假定条件(如样本大小或分布要求)等。因为诠释性研究设计并不使用统计检验，所以诠释性研究不存在统计结论效度的问题。在图 5.2 中，我们给出了不同的效度及其在理论和实证研究层面所处的位置。

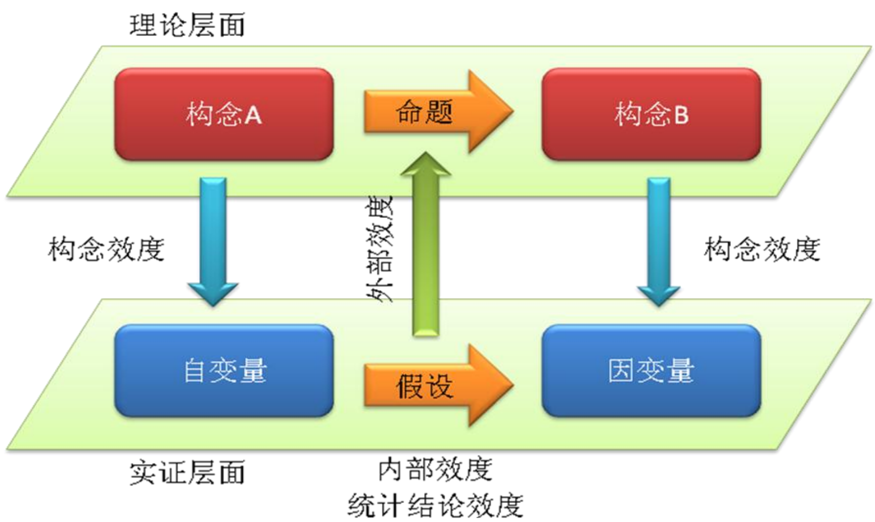


图 5.2. 科学研究中的不同效度类型

提高内外部效度

最好的研究设计应该能同时保证高水平的内部效度和外部效度。这样的设计能防止伪相关，增加假设检验的可信度，并确保小样本中所得出的结论可以推广至总体样本。研究人员可以通过以下五种控制方式来保证研究设计中的内部效度（即因果关系）：（1）操控、（2）剔除、（3）包含、（4）统计控制、（5）随机试验。

在操控方面，研究人员在一个或多个层面对自变量进行操控（称为“处理”），并与没有接受处理的对照组进行比较，以发现结果的差异。这种处理可能是医学治疗研究中对研究对象采用一种新的药物或不同剂量的药物，或是在教育领域对学生采取不同的教学风格等。这种类型的操控可以在实验研究或准实验研究中实现，但并不能在非实验研究（如调查研究）中实现。值得注意的是，如果研究对象不能充分地辨别出不同实验处理的操控的差异，则他们对于不同操控的反应可能相同，那么这样的操控是失败的。

剔除方法可通过在不同处理过程中保持外生变量的一致性来实现，比如将研究对象限定为同性别或处于相同社会经济地位的主体。包含方法可通过在研究设计中包含外生变量，并分别评估其对因变量的作用（比如包含性别因素的因子设计）来实现。这种方法得出的结论

具有普遍推广性，但同时也需要更大的研究样本。在统计控制中，外生变量在统计检验过程里被用作为协变量。

最后，若能保证外生变量的影响是随机（非系统化）产生的，则可以通过随机抽样过程来消除这种影响，我们称这种方法为随机方法。随机方法有两种类型：（1）随机选择，即在总体样本中随机选取样本；（2）随机分配，即把由非随机方式选取的研究主体随机分配到试验组和对照组中。

随机化也可以确保外部效度，并能把目标样本中所获得的结论推广到相应的总体样本中。需要注意的是，由于资源或数据获取的限制而无法进行随机选择时，需要强制进行随机分配。然而，我们很难保证研究结果在不同群体中也可以得到推广。这是因为群体之间存在多个维度上的差异，而我们只能控制其中很少的一些维度。

流行研究设计

如前文所述，基于科学研究目的，研究设计可以被分为两大类：实证研究和诠释研究。实证研究是指理论检验，而诠释研究是指理论构建。实证研究基于客观现实探索出广义的模式，而诠释研究是以参与主体的视角来寻求对社会现象的主观解释。一些流行的实证研究设计包括实验室实验、现场实验、现场调查、二手数据分析和案例研究等。而较为流行的诠释研究主要包括案例研究、现象研究和人种志研究等。需要注意的是，案例研究可用于理论构建或理论检验（虽然不可以同时进行）。并不是所有的方法都适合于各种类型的科学研究。一些研究方法，如焦点组访谈，特别适用于探索性研究，而另一些研究方法，如人种志研究，特别适用于描述性研究，还有一些其他的研究方法，如实验室实验对于解释性研究非常合适。我们将在下面的内容中对其中一些研究设计进行简单的描述。更详尽的细节会在第 9-12 章提供。

实验研究是一种在严格控制的环境下测试因果关系（即假设）的方法。实验研究可以将原因从结果中分离出来，并对接受实验处理的一组（以下简称"实验组"或"处理组"）而不是另一组（即"对照组"）施加原因因素，进而观察两组主体在原因因素影响下的差异。例如，如果我们设计一个实验来测试新药在治疗某些疾病时的功效，我们可以得到一个这类病人的随机样本，并把他们随机分配成两组（实验组和对照组），对实验组主体实施药物治疗，但另一组只服用安慰剂（比如服用没有医疗价值的糖丸）。更复杂的设计可能包括多个实验组，如根据给予药物剂量的高低进行分组，或者多种治疗方式，如将药物治疗与饮食干预治疗相结合。在真实验研究中，研究人员必须对研究主体进行随机分组。如果没有进行随机分组，那么这种研究就只能是准实验研究。实验研究可以在人工环境或在如大学这样的实验室环境下进行（实验室实验），或在现象实际发生的场所，如组织中进行（现场实验）。实验室实验允许研究者分离出感兴趣的变量并控制外生变量，而这在现场实验中是不可能实现的。因此，从实验研究中推出的结论往往具有较高的内部效度，而从现场实验中推出的结论往往具有较高的外部效度。实验数据采用定量统计方法进行分析。实验研究的主要优势在于其可以分离、控制和深入检验一些变量，因此具有较高的内部效度，而它的主要缺点在于其有限的外部可推广性，这是因为现实生活往往比人为设置下的实验环境更加复杂（例如包含更多的外生变量）。此外，如果实验研究不能事先识别和控制相关外生变量，这也有可能降低它的内部效度， 进而可能导致伪相关。

现场调查属于非实验研究，它不能控制自变量进行相关实验，而是测量相关变量并运用统计方法检验这些变量的影响。现场调查通过在现场发放问卷或偶尔通过结构化访谈来捕捉随机抽样主体的实践、信仰等相关信息。在横截面现场调查中，研究人员在同一时间点对自变量和因变量进行观测（例如通过同一份问卷进行调查），而在纵向现场调查中，因变量的测量应晚于自变量。现场调查的优势在于其具有较高的外部效度（因为数据是在现场进行收集的），以及可以捕捉并控制多个变量，并从多个角度或运用多种理论来研究某个问题。然而，由于现场调查缺少时序性的特点，研究人员很难保证它的内部效度（因果关系），并且调查主体也可能存在应答偏差（例如，调查对象会提供一个社会所期望的反应，而不是他们真实的反应），这也进一步降低了其内部效度。

二手数据分析指对之前通过其他来源所获取的数据进行分析。这些数据可能包含从政府机构获得的数据，诸如来自美国劳工局的就业统计数据或来自联合国发展计划署的国家发展统计数据，其他研究人员所收集的数据（常用于元分析研究），或者可公开获取的第三方数据，如股票市场的财务数据或来自 eBay 的实时拍卖数据等。这与大多数由研究者搜集原始数据进行的研究形成了鲜明的对比。当原始数据搜集的成本太高或不可行，并且二手数据可用于回答面向具体分析层面的研究问题时，二手数据分析不失为一个有效的研究方法。这种研究的局限性在于二手数据可能不是用系统而科学的方法所搜集的，因此可能不适用于科学研究。而且由于二手数据基于不同目的而搜集，这些数据也可能无法充分地解决研究人员所感兴趣的研究问题。同时，原因和结果的时间顺序很模糊，这也降低了研究的内部效度。

案例研究是指对一个或多个真实情境（案例现场）所进行的长期而深入的调查。数据可通过访谈、个人观察和内部或外部文档进行收集。案例研究可以是实证研究（假设检验）或诠释研究（理论构建）。这种研究方法的优势在于它能发掘出与未知现象相关的各种社会、文化和政治因素。案例研究往往是定性研究，但是它在很大程度上更加情境化和细致化。然而，对于调查结果的解释可能取决于研究人员的观测和综合能力。控制条件的缺乏也可能会导致难以确立相应的因果关系，而来自于单个案例的调查结果同样难以推广至其他案例。我们可以通过在多案例研究中对不同案例的分析结果进行比较来提高案例的可推广性。

焦点组访谈研究是指一组受访对象（通常是 6 到 10 人）聚在一起对某个特定现象讨论 1.5 到 2 个小时。在研究的过程中，一名受过专门训练的协助人员将主持和引导讨论，并为参与者确立讨论议程及提供讨论的初始问题集。主持人应确保所有参与者都表达了自己的观点和经验，并最终基于参与者的意见和经验对研究问题构建一个全面的理解。由于控制条件的缺乏，该类研究将无法保证其内部效度，同时由于较小的样本量，研究结果也不能被推广至其他情境。因此，焦点组访谈一般不会用于解释性或描述性研究，而较多地用于探索性研究。

行动研究认为对于复杂社会现象的最好理解方式是引入相关干预措施或采取相关行动，来观察这些行动对于社会现象的影响。在该方法中，研究人员通常以咨询顾问或组织成员的身份参与到某个社会情境（如组织）中，并发起一项行动（如新的组织流程或新的技术）来应对真实环境下的问题（如利润下降或运营瓶颈）。研究人员对于行动的选择必须立足于理论，并且这种理论应解释该行动为何以及如何带来预期的变化。之后，研究人员观察这种行动的结果，并在必要的时候对其进行修正，同时通过对该行动结果的观测，提出对目标问题和外部干预的理论见解。研究人员可以通过所选择的行动在何种程度上成功地解决了目标问题来对原始理论进行验证。行动研究区别于所有其它研究方法的核心特征是行动研究能够同时解决问题并产生观点。因此，行动研究可以很好地衔接研究与实践。该方法也适用于研究特定情境下的社会问题，但是这种研究方法会受到研究者偏见和主观性的影响，并且研究结果通常只适用于具体实施该研究的情境。

人种志研究是基于人类学研究的一种诠释性研究设计。人类学研究强调必须在一定的文化背景下对现象进行研究。研究人员需要在很长一段时间内（8个月至 2 年）融入于特定的文化中，并在此期间体验、观察并记录该文化下的日常生活，最终构建关于这种文化下的变革和行为的理论。研究数据主要通过观测、与其他参与者之间正式和非正式的互动，以及个人的现场记录来收集。数据分析是一种"意会"的过程。研究人员必须极其详尽地叙述自己的经历，这样可以让读者不必身临其境就可以体验到同样的文化。该方法的优点在于它的情境敏感性及由此而引发的丰富而又细致入微的理解，并可以最大程度地减少应答偏差。然而，这也是一个极其耗时和占用大量资源的方法，并且研究结果只针对特定的文化背景，而对于其他文化则具有较低的可推广性。

选择研究设计

考虑到上述众多的研究设计，研究人员在他们的研究工作中应如何选择呢？一般而言，研究人员更倾向于选择那些他们感觉良好并最有能力操作的研究设计，但是理想的选择应取决于研究现象本身。在研究问题还不是很清楚的初步阶段，研究人员主要想弄清楚特定研究问题的性质和范围，这时焦点组访谈（用于个体层面的分析）或者案例研究（用于组织层面的分析）是探索性研究中的理想选择。当我们深入到某个研究领域中，但是又没有发现好的理论可以去解释这种现象时，我们就想建立一个理论来填补这个领域的研究缺口，诠释性研究（如案例研究或者人种志研究）可能是更好的研究设计。如果存在不同的理论，并且研究人员希望检验这些不同的理论以将其整合成一个更广泛的理论时，使用实证研究如实验研究、调查研究、或二手数据分析是比较合适的。

无论选择何种研究设计，研究人员应努力通过混合研究方法，如问卷、访谈、观察、文档或二手数据来获取定性和定量数据。比如，即使是在为了搜集定量数据而设计的高度结构化问卷中，研究人员也能通过设置开放性问题来搜集定性数据，并能从中获得结构化定量数据无法得到的见解。同样地，尽管案例研究大多采用面对面的访谈来收集定性数据，但也不应忽视定量数据搜集的潜力和价值。以组织决策制定过程研究为例，案例访谈者可以记录下数值型数据，如组织耗费了几个月来制定某个决策，多少人参与了决策制定过程，以及有多少种决策可供选择等。这些数据为我们提供了大量有价值的，且无法从受访者叙述性的回答中获取的见解。无论采取哪种研究设计，研究人员应尽可能多地搜集各种数据来加深对目标现象的了解。

第六章

构念的测量

理论命题表述的是抽象的构念之间的关系。在检验构念之间的相关程度之前，我们需要运用科学的方法评估这些构念的准确性。构念的衡量需要我们对客观世界仔细进行观察，它是实证研究的精华部分。虽然一些社会科学研究中的构念，比如人的年龄，体重，或公司规模等，可以很容易进行衡量，但是其他如创造力、偏见、或疏远等构念可能很难衡量。在这一章中，我们将分析如何将构念概念化，并说明衡量这些构念的过程。

概念化

概念化是一种思维过程。在概念化过程中，我们将模糊的不准确的构念（概念）及其组成部分进行具体的准确的定义。例如，我们经常使用“偏见”这个词，当我们提到这个词时，我们的脑海中就会出现其对应的形象。不过，当我们需要确切定义这个词的意思时，我们会感到比较困难。如果有人对其他种族进行不当评论，这可以被定义为种族歧视吗？如果女性和男性从事同样的工作时，其收入却低于男性，这可以被定义为性别偏见吗？如果教徒们相信那些没有宗教信仰的人会受地狱的煎熬，这可以被定义为宗教偏见吗？“偏见”是否可以被划分为不同种类呢？如果可以，怎样进行划分呢？是否存在不同程度的"偏见"呢？如果存在，我们又应该怎样定义呢？这些问题的回答对于准确衡量"偏见"这一构念是很关键的。将“偏见”这一构念概念化其实就是理解"偏见"概念中应该包含什么或者不包含什么的过程。

由于社会科学研究中许多构念是不准确的、模糊的并容易产生歧义，因此在衡量构念时，最重要的就是将构念概念化的过程。例如，“同情心”这一构念是否等同于“同理心”或“多愁善感”这些构念呢？如果有如下一命题：人们的同情心与同理心之间存在正相关，除非你可以区分“同情心”和“同理心”的概念并利用在实证研究中准确衡量这两个非常相似的构念，否则就不能对这一命题进行检验。如果笃信宗教的教徒相信社会中的一些成员（如无宗教信仰的人，同性恋者，堕胎医生等）将因其罪恶行为而受地狱的煎熬，教徒们设法改变这些“罪人”的的行为以防止其受地狱之苦，那么教徒们的这种行为应该定义为有偏见的还是应该定义为有同情心的呢？这种构念的定义不基于任何客观的标准，而是基于我们对这些构念的共识性（“主体间的”）的主观意像（概念）。

当定义如“偏见”或“同情心”的构念时，我们必须认识到，有时候这些构念并不是真实的或者能够独立地存在，它们可能仅仅是我们脑海中虚构的事物。例如，生活在世上的群体之间可能根本不存在“偏见”甚至他们根本不知道“偏见”的具体含义。但在现实生活中，我们往往认为这个构念是真实存在的。将有关思维上的构念变为现实的过程称为“具体化”，在定义构念并找出衡量构念的变量时，“具体化”这一过程是重要环节。

在对构念进行概念化的过程中，我们必须选择是单维度地还是多维度地对构念进行衡量。单维构念只预期包含一个基本维度。这些构念能够用单个变量进行衡量。一些简单的构念，如人的体重，风速等都属于单维构念。甚至一些复杂的构念，比如"自尊心"（虽然认为自尊可从单一维度进行解释是一个不切实际的假设），也可以属于单维构念。多维构念包含两个或两个以上的基本维度，比如，我们可以用数学能力和语言能力两个维度来定义某人拥有的学术资质。那么学术资质就是多维构念。在这种情况下，每个基本的维度都应该分别进行衡量。比如说，分别用不同的方法衡量数学能力和语言能力，或者以加权的方式综合这两方面的衡量得分结果，这样就可以从大体上对学术资质进行衡量。

构念操作化

在我们对理论构念进行定义后，我们应该怎样衡量这些构念呢？操作化指的是找出衡量这些构念的指标。比如，如果一个不可观测的理论构念如社会经济状况可以被定义为家庭收入水平，则该构念可操作化为这样一个指标，即让受访者回答这样一个问题：你的家庭的年收入是多少？考虑到社会科学研究中的构念有较强的主观性且其本质定义不是很严谨（除了一些人口统计学的构念，如年龄，性别，教育和收入等）我们倾向于采用多个指标来衡量这些构念。这样我们就可以通过检查这些指标的紧密程度来评估衡量结果的准确性（可靠性）。

相较于作为理论层面的概念的构念而言，指标是处于实证层面的。在实证层面上，用于代表构念的指标组合被称之为变量。如前一章所述，根据变量在研究中的不同使用情况，可将其分为自变量，因变量，中介变量和调节变量。每个指标有几个属性层级（或者水平）且每个属性层级代表一个值。比如，性别这一变量有两个属性层级，即男性和女性。同样地，代表顾客满意度这一构念的量表中可以包含五个属性层级：“强烈不满”、“稍有不满”、“中立”、“稍微满意”和“非常满意”。属性值可以为定量的（数值型）或者定性的（非数值型）。定量数据可以通过定量分析方法来分析，比如通过回归或者结构方程模型等方法。但是，定性数据需要通过定性分析方法进行分析，如对数据进行编码等方法。需要注意的是，在社会科学研究中，即使很多变量用定量方法进行表示，他们仍然属于定性数据。比如，我们可以用五个属性层级来测量顾客满意度：“强烈不满”、“稍有不满”、“中立”、“稍微满意”和“非常满意”，并对这五种属性从 1 到 5 分别赋值，通过这种操作，我们就可以运用高级统计工具进行定量分析。然而，我们应该认识到，这些数据仅仅代表受访者个人评估的自身满意度。即使我们对基本变量（满意度）用定量方法进行表示，但是其仍然是定性数据。

指标可以是反映型指标或是构成型指标。反映型指标是“反映”某个潜在构念的一个测量。比如，虔诚度是衡量信奉宗教程度的构念，参加宗教仪式可以作为虔诚度的反映指标。构成型指标是“构成”某个潜在构念的一个测量。构成型指标可以代表其所感兴趣的构念的多种维度。比如，如果宗教信仰包括信仰，祷告和宗教仪式这几个维度，选择衡量这些不同维度的指标即为构成型指标。单维度构念通过反映型指标来衡量（即使会通过多个反映型指标来衡

量深奥难懂的构念，比如自尊等）。多维度构念由多维度的构成型指标来衡量，即使每个基本维度可能由一个或多个反映型指标来衡量。

测量尺度

在对构念进行操作化定义时首先考虑的应是其测量尺度。测量尺度也叫做评点量表，它指的是指标的取值（而不是指标本身）。比如，男性和女性（或 M 和 F，或 1 和 2）是"性别"的两个取值水平。在心理学家斯坦利·史密斯史蒂文斯（1946）的一篇发表于《科学》的题为"测量尺度理论"的开创性文章中，作者定义了科学测量中的四种广义的评点量表：即名义、顺序、等距和等比量表。这些量表的统计特征如表 6.1 所示。

量表	集中度	统计量	转换
名义	众数	卡方检验	一对一的（平等）
顺序	中位数	百分位数，非参数统计	单调递增（顺序）
等距	算术平均值、范围、标准差	相关分析，回归分析，方差分析	正向线性相关 （仿射）
等比	几何平均值、调和平均值	变异系数	正相似性 （乘数， 对数）
注： 所有高等级量表均可用于低等级量表			

表 6.1 等级量表的统计特征

名义量表又称类别量表衡量的是类别型数据。这种量表常用来衡量具有互相排斥属性的变量或指标。比如性别（两种取值：男性和女性），行业类别（制造业，金融业和农业等）和宗教信仰（基督教徒、穆斯林和犹太信徒等）。即使我们对不同的属性赋不同的值，比如衡量性别时，若为男性则赋值为 1，女性则赋值为 2，但是，这些数值并不真的意味着什么。（比如，1 并不代表其比 2 小，或是 2 的一半），这些属性可以通过非数值型数据进行替代，比如用 M 代表男性，用 F 代表女性。名义量表仅仅提供不同属性值的名称或者标识。衡量名义量表的集中度的最佳统计量是众数，而不是平均数或者中位数。可用的统计量包括卡方和频率分布，并且只允许一对一的（等价）转换（比如 1=男性，2=女性）。

顺序量表衡量的是排序型数据，比如基于学生的平均成绩所得出的第一，第二和第三等的排序。然而，这类数据属性的真实值，相对值或者属性值之间的差异却不能体现出来。比如，学生在班级的排名不能反映学生真实的考试成绩或者绩点，也不能反映学生之间的差距。在自然科学中涉及到顺序量表的典型例子是摩氏硬度。摩氏硬度利用一种矿物相对另一种矿物的刮伤力来衡量矿物之间的硬度。比如，钻石可以在地球上的任何其他天然矿物上留下刮痕，所以，钻石是"最硬的"矿物。然而，这种测量不能反映矿物的真实硬度也不能够反映各类矿物硬度之间的差距。顺序量表也进行属性（锚点）的定义，比如定义"严重"、"中等"、"良好"，或者"强烈不满"、"稍微不满意"、"中立"，或者"稍微满意"、"非常满意"。在对满意程度进行衡量时，我们可以认为回答"稍微满意"的被调查者比回答"非常满意"的被调查者有较低的满意度，但是我们不可以量化他们的满意程度。可以用中位数或者众数来衡量顺序量表

数据的集中度。因此，可以对顺序数据进行百分位数统计分析和非参数分析，但是不可以进行如相关分析，回归分析，方差分析等更高级的分析。我们也可以对数据进行单调增的变换（即保留原有排序）。

等距量表的数据不仅是有序的并且彼此相邻之间是等距的，比如温度标（华氏度或者摄氏度）中 30 华氏度与 40 华氏度的差距和 80 华氏度与 90 华氏度的差距是一样的。同样地，当用如 0 到 10000 美元，10000 美元到 20000 美元，20000 美元到 30000 美元等属性（范围）来问受访者的年度收入时，由于每个范围的中间值（即 5000 美元，15000 美元和 25000 美元等）之间是等距的，因此得到的数据就属于等距量表。智商的衡量也是运用定距量表，因为这样才使得智商得分 100 与 110 的差距和智商得分 110 与 120 的差距是相等的（虽然我们不知道真正的情况是否是这样）。运用等距量表，我们可以衡量变量之间的差距，但是运用名义量表或者顺序量表却不可以做到这一点。我们可以用平均数，中位数或者众数来衡量等距数据的集中度，用范围和标准差来衡量数据的离散程度。我们除了可以利用名义和顺序数据中的统计分析方法对等距数据进行分析外，还可以利用相关分析，回归分析和方差分析等方法。我们同样也可以对量表数据进行正线性的变换。需要注意的是，上文讨论的衡量顾客满意度的量表并不严格是等距量表。因为我们不能获知“非常满意”与“稍微满意”的差距和“中立”与“稍微满意”的差距或者“稍微不满意”与“非常不满意”之间的差距是否相等。但是在社会科学的研究中我们常“假称”（不正确地）这些差距是相等的，所以可以运用统计方法对这些数据进行分析。

等比量表除了具有名义、顺序和等距量表的特征外，还有“绝对零点”这一特征（零值意味着缺乏或不存在这一基本构念）。很多自然科学和工程学中变量的测量，如质量，平面倾斜度，电荷量等都用到等比量表。同样的，社会科学研究中的变量，如年龄，在单位的任期和公司规模（用员工数或者总收入来衡量）等也运用等比量表。比如公司规模为零时表示公司不雇佣员工或者公司没有收入。绝对温度不同于华氏或者摄氏温度，因为绝对温度中的零值（即-273.15 摄氏度）不是任意的值而是代表在这一温度下，物质粒子没有动能，所以绝对温度属于等比量表。由于等比量表中数据之间的比值是有意义且是可解释的，因此这些数据就属于等比量表。比如公司的雇员数为 10 人时，表示其规模是雇员数为 5 人的公司规模的两倍。同样地，公司的雇员数为 10000 时表示其规模是雇员数为 5000 的公司的两倍。所有可以衡量集中度的统计量都适用于等比数据，比如几何平均值和调和平均值等。所有衡量离散程度的统计量也适用于等比数据，比如范围或者变异系数等。所有统计方法都适用于等比数据的分析。正相似（如乘数或者对数）的复杂变换也可以在等比数据中使用。

基于上述讨论的四种基本类型的量表，我们还可以开发出特别的量表进行社会科学研究。常见的量表包括二元量表，李克特量表，语义差分量表或者格特曼量表。其他一些不常见的量表在此不做讨论。

二元量表。二元量表对衡量变量的语句，只提供是还是不是，或正确还是错误等两种可供选择的答案。比如，“政治活动”构念中典型的二元量表可能包含如表 6.2 所示的六个语句。量表中的每个语句都只有两种选择。被调查者选择“是”的次数（0 到 6）可以全面衡量其“政治活动”。如果要理解是怎样确定这六个语句的，可以参考这章后面的"量表开发"部分的说明。二元量表中也可以有其他值，比如性别构念中的男性或者女性，雇佣状态构念中的全职或者兼职等。如果对雇佣状态构念进行修改，使其包括两个以上的选择时，它就不再属于二元量表，但是仍然是常见量表的语句。

你曾经给政府工作人员写过信吗？	是	否
你曾经在政治请愿书上签过名吗？	是	否
你曾经捐钱支持过政府事业吗？	是	否
你曾经捐钱给候选人帮助其竞选公职吗？	是	否
你曾经写过政治信并将其寄给报纸或者杂志社的编辑吗？	是	否
你曾经试图说服他人改变其投票计划吗？	是	否

表 6.2 衡量政治活动的六语句的二元量表

李克特量表。这量表是由伦西斯 李克特设计的，它是社会科学研究中普遍的衡量有序数据的评估量表。李克特量表中对包含的语句进行简单说明，使得被调查者可以用 0-5 分或者 0-7 分之间的分值反映自己从“强烈反对”到“强烈赞同”的不同程度的态度。衡量“雇员自尊”构念的包含六个语句的典型李克特量表如表 6.3 所示。李克特量表对各个语句中被调查者所做的选择进行加总，最终得到量表各个语句所反映的受访者态度的加总值。

	非常不赞同	稍微不赞同	中立	稍微赞同	非常赞同
我对自己的工作感到很满意	1	2	3	4	5
在工作中，我能与他人和睦相处	1	2	3	4	5
在工作中，我为我和上司之间的关系感到自豪	1	2	3	4	5
我能肯定工作中的其他人很乐意和我相处	1	2	3	4	5
我能肯定我的同事是尊重我的	1	2	3	4	5
我觉得我在工作中做了有用的贡献	1	2	3	4	5

表 6.3 衡量“雇员自尊”的李克特量表

李克特量表比二元量表更能够反映更多不同程度的回答（即得到更加细致的回答），包括受访者是否对这一论点持中立态度的回答。可以用 0-3 或者 0-9 的值来代表受访者的不同态度，但是，考虑到受访者可能保持中立（既不同意也不反对）的态度，我们应该取奇数值进行衡量。在有些研究中，研究人员去掉了中立态度的选项并用偶数值来衡量受访者的不同态度，强制受访者进行“同意”或者“反对”的选择。因为有的受访者确实对某一论点持中立态度，如此"强制选择法"不能反映受访者中立的立场。虽然李克特量表中各个语句涉及不同的内容，但是对每个语句均有相同的选择范围（从“非常不赞同”到“非常赞同”）。即使有时我们将其视为等距量表，但李克特量表中各个集合锚点之间并不是等距的，基于此，李克特量表还是顺序量表。

你怎样评定你对国民健康保险的意见等级？						
	非常	稍微	中立	稍微	非常	
好的	□	□	□	□	□	坏的
有用	□	□	□	□	□	无用
关注	□	□	□	□	□	不关注

引起兴趣的	□	□	□	□	□	令人厌烦的

表 6.4 衡量国民健康保险的语义区分量表

语义区分量表是复合（多项）量表。在量表中，受访者对一组完全对立的形容词之间的不同程度进行选择，这一组形容词仅仅是对一个语句的描述。例如"对国民健康保险的态度"这一构念可以通过表 6.4 进行衡量。在李克特量表中，总体得分是每个语句所反映的受访者态度的得分的加总，需要注意的是，在李克特量表中，可能有不同的语句描述，但是形容词的不同程度的描述是一样的，但是会有不同的对立形容词组。在衡量人们对客观物体，事件或者行为的看法和感受时，我们认为运用语义区分量表是很好的方法。

格特曼量表是由路易斯 格特曼设计的。这个量表由程度从最轻到最重的语句组成。比如，"对移民者的态度"这一构念的衡量可用表 6.5 中的五个语句。在格特曼量表中，根据语句程度轻重的不同对语句赋予权重，对这些语句进行选择后的得分进行加权组合就得到量表的综合衡量得分结果。

对下述有关移民者的陈述，你有怎样的观点？		
你愿意移民者成为你国家的公民吗？	是	否
你愿意移民者生活在你社区里吗？	是	否
你愿意移民者居住在你隔壁吗？	是	否
你愿意接受移民者作为你的密友吗？	是	否
你愿意你家庭成员中有人与移民者结婚吗？	是	否

表 6.5 衡量对移民者态度的格特曼量表

量表开发

前面主要讨论了如何衡量受访者对基本构念中已设计的语句或指标的回答。但是我们怎样设计出这些指标呢？设计这些指标的过程就叫做量表开发。更正式地来说，量表开发也是测量的一部分。对构念的测量需要将与不可观测的构念有关的定性判断与定量的测量单位结合。史蒂文斯（1946）就曾指出，量表开发就是根据一定的规则对客观实体赋值。在实证科学研究中，将抽象概念具体化仍然是最难解决的。

量表开发最终是为得到包含给定构念的衡量语句或者指标的量表。这部分讨论的量表开发和前面部分讨论的量表评价是不同的。量表评价主要是收集受访者对既定量表语句的回答。比如，定类量表收集受访者"是"或者"否"的回答，等距量表收集受访者从"非常不赞同"到"非常赞同"的不同程度的回答。对某一陈述回答进行量表评价并不是量表开发过程。量表开发指的是在量表评价之前开发出量表语句。

量表可能是单维的也可能是多维的，这取决于基本构念是单维的（比如体重，风速，公司规模）还是多维的（比如学术才能，智商等）。单维量表根据由高到低的尺度来衡量构念。需要注意的是，单维量表也可能包含多个语句，但是这些语句都只是衡量同一基本维度。单

维量表在社会科学构念衡量中比较普遍。比如"自尊"可通过从高到低的单一维度来衡量。多维量表，运用不同的语句来分别衡量构念的各个维度，或者从不同的维度对构念进行测量，然后加总每个维度的得分得到多维构念的综合衡量得分。比如，学术才能的衡量可通过对学生的数学能力和语言表达能力分别进行衡量来实现，然后对这些得分进行加总得到学术才能的综合得分。由于社会科学研究中大多运用单维量表，接下来，我们将介绍开发单维量表的三种方法。

单维量表开发方法发展于二十世纪上半年代，各种方法根据其发明者进行命名。三种最普遍的单维量表开发方法如下：（1）瑟斯顿的等距量表法、（2）李克特的总加表法和（3）格特曼的累积量表法。三种量表开发方法在很多方面是相似的，只是评判者对量表语句的评级和选择最终量表语句时所运用的统计方法有所区别。接下来我们对每种方法进行讨论。

瑟斯顿等距量表法。路易斯·瑟斯顿是最早的最著名的量表开发理论家，他在 1925 年发明了等距的量表开发方法。这种方法以构念的概念定义为基础，基于概念产生相关量表语句。语句由对构念有所了解的专家进行确定。最初的待选语句用相似的方式进行表述，比如，将其表述成受访者是同意或是反对某一论点（而不是要求受访者对问题进行具体陈述）。接下来，要求相关评判者从待选语句中选出最能反映构念的语句。评判者可以选择在量表构造上受过培训的专业学者，也可以随意选择对该构念有兴趣的调查对象（比如熟悉该现象的调查对象）。在语句选择过程中，评判者根据自身的观点用 1 到 11 的值来评判每个语句能够反映该构念的程度（1 表示该语句非常不适合代表该构念，11 表示该语句很适合代表该构念）。计算出每个语句的中位值和四分位差值（75%分位值和 25%分位值的差，可以用来衡量离散程度）描绘成如图 6.1 所示的柱状图。最终选择的语句应有相等的四分位差值。我们可以用每个语句得到的中位值代表语句得分，并选出具有最小的四分位差值的语句。但是，除了完全依靠统计分析方法进行语句选择外，对待选语句的每个层面进行检查并选出最清晰最有意义的表述也不失为一个更好的策略。每个量表语句的中位值代表最后加总所感兴趣的构念的各个语句得分时所运用的权重。这种量表好比一个标尺，每个语句或者陈述根据 1 到 11 这一尺度进行评分（同样根据这一尺度进行加权）。因为每个语句得到 1 到 11 的评分的概率相同，这种方法就叫做等距法。

瑟斯顿也发明了单维量表开发的其他两种方法：连续区间法和配对比较法。除了评判者评定数据方法的不一样外，这两种方法和等距法的方法基本类似。比如，配对比较方法要求每个评判者对每对陈述进行评价（而不是根据 1 到 11 的尺度评价单个陈述），因此取名为配对比较法。由于需要很多配对陈述，这种方法是相当耗时的，而且与等距法相比，其适用度低。

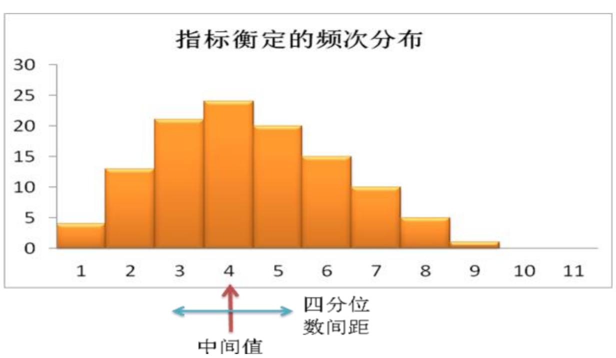


图 6.1 瑟斯顿量表语句的柱状图

李克特加总量表开发方法。这个单维量表开发方法是由墨菲和李克特（1938）发明的，它很可能是这章介绍的三种量表开发方法中最普遍的方法。和瑟斯顿的量表开发方法一样，李克特的方法是基于感兴趣的构念的定义，利用专家的工作，产生 80 到 100 个待选的量表语句。评判者根据 1 到 5 的评价尺度对这些语句进行评价（1 表示非常不赞同，2 表示稍微不赞同，3 表示中立态度，4 表示稍微赞同，5 表示非常赞同）。在这种评价尺度下，最终选出代表构念的语句。这里介绍了几种选择方法：（1）计算评判者所评判的每个语句与总语句（对所有单个语句进行加总）的二元相关系数，剔除相关系数低（如 0.6 以下）的单个语句。（2）截取最高分数端的 25%为高分组，最低分数端的 25%为低分组。求出这两个组的平均值，对高低两组的平均值做 t 检验，选出有较高 t 值的语句。（即选出高分位组合与低分位组区分度大的语句）。最后，研究人员选出相对有限的语句集，这些语句与总语句的相关系数高或者其语句本身有较大的区分度（即 t 值较大）。李克特量表开发方法假定所有语句具有相同的权重，因此可以对受访者对于每个语句的回答直接加总评分。因此这种方法叫做加总量表开发。需要注意的是，在加总得分之前，与构念原意义相反的语句应该进行相反方向的赋值（即 1 代表非常赞同，2 代表稍微赞同等等）。

格特曼的累积量表法。这个方法是格特曼根据埃默里·博加德斯的社会距离方法进行设计的。这个方法假设人们和他人一起参加社会活动有不同的意愿，用从“最不强烈的”到“最强烈的”一系列的语句衡量这种意愿程度。这个方法的思想是只要受访者同意其中的一条语句那么代表其也同意先前所有的语句。但在实践中，我们很少能够找到与这种渐增模式相吻合的语句组合。语句组合与渐增内涵的相关程度可以运用量表图分析法检验。

和前面介绍的量表开发方法一样，格特曼量表开发方法也是基于感兴趣的构念的定义并利用专家的工作，产生很多待选的量表语句。评判者对这些语句进行评价，如果这些语句与构念相符，评判者就选择“是”，如果其认为语句与构念不符，他们就选择“否”。接下来，利用矩阵或者表格来表示评判者对待选语句的回答。将表的资料首先依据各被访者回答"是"的个数由上到下进行降序排列，然后再按答案“是”的数目由多至少将各语句从左至右顺序排列。得到重新整理的表格如表 6.6 所示。需要注意的是，量表从左至右（即横穿各个语句）阅读时是渐增的。但是，如表 6.6 所示，存在一些异常情况，因此表格并不完全是渐增的。可以运用量表图分析方法选出最能实现量表渐增特性的语句（当语句数量较少时，可直接通过观察挑选语句）。这种统计方法评估每个语句的得分，并利用各个语句的得分加总计算得出全部语句的分值。

回答者	**语句 12**	**语句 5**	**语句 3**	**语句 22**	**语句 8**	**语句 7**	**…**
29	Y	Y	Y	Y	Y	Y	
7	Y	Y	Y	-	**Y**	-	
15	Y	Y	Y	Y	-	-	
3	Y	Y	Y	Y	-	-	
32	Y	Y	Y	-	-	-	
4	Y	Y	-	**Y**	-	-	
5	Y	Y	-	-	-	-	

23	Y	Y	-	-	-	-
11	Y	-	-	**Y**	-	-
Y 表示妨碍矩阵累积性的异常情况						

表 6.6 格特曼量表中已排序整理的矩阵

指数

指数是指根据一定的规则和公式对多个构念（即组成部分）进行衡量并加总计算其得分。它不同于量表那样需要集合衡量，但是指数衡量的是构念的不同维度或同一维度。比如，消费者价格指数（CPI）就是众所周知的指数。美国劳工部劳动统计局每个月都要计算这个指数。CPI 指数衡量的是消费者通常为商品和劳务支付多少，这个指数分为五类（食品和饮料、房地产、服装、交通、医疗、娱乐、教育、通讯、和其他商品和服务领域），每类可进一步细分为 200 多个物品。政府人员每个月对全国各地 80000 多种物品的当前价格进行统计。考虑到每种物品购买的地点和购买的可能性，计算指数时需要使用复杂的权重计算方法。分析师运用一定的规则和公式对这些价格进行加总得到综合指标值。

另一个指数的例子是社会经济状况指数，也叫做邓肯社会经济指数。这一指数综合了收入，教育和职业三种构念。收入用美元来衡量，教育用受教育年限或者所获学历来衡量，职业根据不同的情况划分为不同的类别和层次水平。结合这些不同的衡量角度，对“职业教育”（某个职业中受过一年或者一年以上大学教育的人数比例）和“职业收入”（某个职业中收入高于特定的年收入的人数比例）进行加权组合，得到综合社会经济状况指标值。然而，研究人员对社会经济状况指数的衡量有很多的争议和不同的看法。

构造指数的过程与量表开发的过程类似。首先对指数及其组成部分进行概念化（定义）。这看似简单，但是对于指数中到底应该包含哪些部分（构念），人们对此有着很多不同的看法。比如，在社会经济状况指数中，收入难道不和教育和职业相关吗？如果相关，我们应该仅考虑一个组成部分还是应该仍然考虑三个组成部分呢？回顾这方面的文献，运用这方面的理论，或者向专家及关键的利益相关者进行咨询等可能有助于我们解决这一问题。其次，对各组成部分进行操作化定义和衡量。比如，你怎样对职业进行分类，特别地，一些职业是随着时间变化的（比如，在互联网出现前是没有网站开发这一职业的）。最后，确定计算指数值的规则和公式。同样地，这个过程中有很多主观影响因素。最后，还要根据现有的或者新的数据来验证所得的指数值。

虽然指数和量表最终都是通过数值得分来代表所感兴趣的构念，但是它们在很多方面是不同的。首先，指数通常由互不相同的部分组成（比如，社会经济状况指数中的收入，教育，职业）而且用不同的方法进行衡量。但是量表包含的语句是相似的，这些语句用的是相同的评级尺度（如五等级的李克特量表）。第二，指数通常综合的是客观的可衡量的值，比如价格或者收入等。但是量表的设计是为了评估主观的或者判断性的构念，如态度，偏见或者自尊等。一些人认为量表研究方法的精细度高于指数研究方法。但是，其他人认为指数研究方法同量表研究方法有相同的精细度。虽然如此，指数和量表都是社会科学研究中的基本工具。

类型

量表和指数对单维构念进行等距测量。但是，研究人员有时希望总结两个或者两个以上的构念进行类别创造，这就叫做类型学。与指数和量表不同，类型是多维的但是仅仅包含名义变量。比如，可以根据报纸对本国及外国政策的不同倾向将报纸分为不同的政治类别，如图 6.2 所示。这种分类可以将报纸归为四个"理想类型"中的一类（A 类到 D 类）。可以观察报纸在这些类型中的分布情况，甚至可以根据报纸的其他属性特征创造新的模式将报纸分为四种理想类型，并将每种报纸归为其中的一类。

图 6.2 报纸的多维分类

总结

最后，我们认为社会科学研究中的量表（或者指数）构造是一个复杂过程。在这个过程中，我们应该做如下考虑：

- 应该使用量表、指数和类型中的哪一种进行构念的衡量？
- 计划怎样分析数据？
- 测量的水平或评点量表是什么（类别、顺序、等距或者等比）？
- 应该使用哪些量表属性值。（比如 1 到 10，1 到 7，-3 到+3）？
- 应该使用奇数还是偶数的属性值？（即是否预期获得有关态度中立的值？）
- 最后，利用哪种方法开发量表语句（比如瑟斯顿、李克特或者格特曼方法）或者挖掘出指数的组成部分。

这章讨论了量表开发的过程和结果。下章将介绍怎样评估用上述方法开发出的量表的可靠度和效度。

第七章

量表的信度和效度

前一章讨论了社会科学研究中构念测量的一些困难。例如，我们如何知道我们所测量的是“同情心”而不是“同理心”，因为这两个构念在含义上比较接近？或者同情心和同理心是一样的吗？更复杂的是，有时候这些构念是想象中的概念（也就是说，它们在现实中并不存在），并且是多维的（在这种情况下，我们又多了一个难题，即确定它们的构成维度）。因此，仅用任何我们所喜好的量表来测量社会科学的构念是不够的，我们必须测试这些量表以保证：（1）这些量表确实测量了我们想要测量的潜变量（也就是说，这些量表是“有效的”）；（2）它们一致而精确地测量了这些构念（即量表是“可靠的”）。信度和效度（合在一起被称为测量量表的“心理计量特性”）是科学研究中测量程序的充分性和精确性的评价尺度。

一个量表可以是可靠但无效的，例如当一个量表非常一致地测量某个错误的构念。同样地， 一个量表也可以是有效但不可靠的，例如一个量表测量的是正确的构念，但结果却并不总是一致。以射击做类比（如图 7.1 所示），一个构念的可靠而有效的多指标量表由聚集在靠近靶心的较窄范围内的射点所构成。有效但不可靠的量表由那些在靶心附近，但并没聚集在较窄的范围内，而是在靶心周围零散分布的射点所构成。最后，可靠但无效的量表由聚集在较窄范围内却偏离靶心的射点所构成。因此，为确保对构念进行充分地测量，信度和效度这两个属性都是必不可少的。

图 7.1 信度和效度的比较

信度

信度是指在多大程度上对某个构念的测量是一致的或可靠的。换句话说，如果我们用这个量表多次测量相同的构念，在潜在现象没有发生变化的情况下，我们是否每次都能得出相同的结果？不可靠测量的一个例子是让别人猜你的体重。人们的猜测很可能是不同的，不同的测量将会不一致，因此，通过“猜测”来进行测量是不可靠的。一个更可靠的测量是使用体重表，这样除非你的体重在测量间隔期发生了实际的变化，否则你每次踏上体重计都有可能得到相同的数值。

注意，信度是指一致性而不是精确性。前例中的体重计，如果体重计校准错了（比如，为了使你感觉较好，从你真实的体重中减去 10 磅），这将无法衡量你的真实体重，因此这不是一个有效的测量。然而，校准错误的体重计在每次称量时都会给出相同的重量（比真实重量少 10 磅），因此该量表是可靠的。

社会科学计量中不可靠观测的来源是什么？观测者（或研究者）的主观性是主要来源之一。如果公司的员工士气用观察员工是否互相微笑、是否开玩笑等等来衡量，若观测者在非常繁忙的工作日（没时间开玩笑、交谈）或是在非常轻松的工作日（有更多交谈和玩笑的时间）观察员工，那么不同的观测者对士气水平有不同的推测。即使在同一天，两个观测者对士气水平形成不同的推测，这取决于他们认为什么是笑话、什么不是笑话。“观测”是一个定性的测量手段。有时，运用定量测量可以提高可靠性。例如，通过计算一个月内提出不满的次数来衡量（逆）士气。当然，无论不满是否是一个有效测量，它更少受制于人类的主观性，因此更可靠。第二个不可靠观测的来源是提出不准确、含糊不清的问题。例如，如果询问人们的工资是多少，不同的人对这个问题有不同的理解：月薪、年薪、或每小时的工资，因此，导致结果可能会大不相同并且不可靠。不可靠的第三个来源是提出的问题是受访者不熟悉或不关心的问题，如问一个美国大学毕业生对斯洛文尼亚和加拿大的关系是否满意，或者要求公司的首席执行官评价公司技术策略的有效性——而他很可能将技术策略委派给技术主管。

那么如何才能创造可靠的测量呢？正如许多社会科学研究一样， 如果测量需要向其他人征集信息，那么就可以从替换数据收集方式开始，即用主观性低的方法（如问卷）替代主观性高的方法（如观察）；只询问受访者知道答案的问题或他们关心的事件；在测量时避免模糊不清的指标（即清楚地界定收集的信息是否是年薪）；简化指标的措辞确保受访者不会错误理解（即避免使用受访者不理解的复杂文字）。这些措施即使不会使测量完全可靠，也会提高我们测量的可靠性。尽管如此还是需要对测量指标的可靠性进行检测。有很多种方法可以评估信度，下面将讨论这个问题。

评判间信度。评判间信度，也称为观测者间信度，是对两个或两个以上独立评判（观测者）关于同一构念一致性的测量。通常在试验研究中对此进行评估，并且根据构念的测量类型可采用两种方式来实现。如果测量是分类的，而且对所有类别都进行了界定，评判核查每个观察落在哪一类，评判间一致性的比例可作为评判间信度的估值。例如，如果两个评判将 100 个观测分到三种可能类别中的一类中，他们分类的重合度为 75%，那么评判间信度是 0.75。如果这种方法是测量区间或是比例 （例如，两个评判每隔 5 分钟用 1-7 级的量表测量课堂活动），两个评价者衡量之间简单的相关也可用来评估评判间信度。

重测信度。重测信度衡量了不同时点对相同样本的同一构念进行两次测试的一致性。如果观测在两次测试间没有发生实质的变化，这种测量是可靠的。两次测验得到的观察值的相关系数是重测信度的估计。这里指出两次测试的时间间隔是至关重要的。一般来说，时间间隔越长，这段时间内（由于随机误差）两个观察值发生变化的可能性越大。

分半信度。分半信度是构念测量的两个一半之间一致性的衡量。例如，如果对一个构念有一个 10 个指标的测量，首先，将这 10 个项目随机等分为两组 （如果整个数字是奇数可以不等分），并且在相同的受访者中实施测试。然后，计算出每个受访者在每组测试中的总分，每组之间的相关系数作为分半信度的衡量。指标越多，每组测量的相似性越大（因为随着更多指标的加入误差降低了），因此，该技术倾向于系统性地低估长量表的信度。

内部一致信度。内部一致信度是对同一构念不同测量指标一致性的衡量。如果对受访者采用一个多指标量表，那么受访者在多大程度上采用相似的方式来评价这些指标反映了内部一致性。该信度可以用指标间的相关系数均值、单个指标与总体指标间的相关系数均值、或更普遍地用 Cronbach’s α 来估计。例如，如果有一个包含六个指标的量表，那么，将有十五种不同的指标配对方式，即这六个指标之间的十五个相关系数。指标间相关系数均值是这十五个相关系数的平均数。为计算单个指标与总体指标的相关系数均值，首先应对六个项目赋值创建一个“总体”指标，然后计算每个指标与总体指标的相关系数，最后再计算这六个相关系数的均值。上述两种测量都没有考虑测量指标的个数（在这个例子中是六个指标）。李·克隆巴赫在 1951 年提出了信度测量——Cronbach’s α，该指标在信度估计时注重量表规模，用以下的公式计算：

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum_{i=1}^{K}\sigma_{Y_i}^2}{\sigma_X^2}\right)$$

K 是测量指标的数量，σ_X^2是观测到的总评分的方差（标准差的平方），$\sigma_{Y_i}^2$是观察指标 i 的方差。

标准化的 Cronbach’s α 可以用一个简单公式计算：

$$\alpha_{\text{standardized}} = \frac{K\bar{r}}{(1+(K-1)\bar{r})}$$

K 是指标的的数量，$\bar{r}$ 是指标间相关系数的均值，即矩阵的上三角（或下三角） K（K-1）/ 2 个系数的均值。

效度

效度，常被称为构念效度，是指测量在多大程度上反映了想要研究的构念。举例来说，同情心的测量是否真的反映了同情心，而不是衡量另一个构念同理心。效度可以用理论或实证的方法来评估，并且最好结合二者来评估。理论评估主要关注于操作性的测量在多大程度上转化了或表示了理论构念。这种类型的效度称为转化效度（或表示效度），包括两大类：表面效度和内容效度。转化效度通常用专家组的判断以及称作 Q-sort 的定性技术来评估，专家组对每个测量项 （指标）进行打分以评价其与构念的概念化定义的匹配程度。

效度的实证评估基于实际观察检验既定的测量与一个或多个外部标准相关的程度。这种类型效度被称为效标关联效度，包括四种类型：聚合效度、区效度分、同时效度以及预测效度。转化效度检验了测量是否是潜在构念的有效反映，效标关联效度则是根据构念的理论来检验给定的测量是否以应该的方式进行。这种评估是基于观测数据的定量分析，使用了诸如

相关分析、因子分析等统计技术。图 7.2 表明了效度的理论与实证评估的区别。然而，为充分保证社会科学研究测量的效度，这两种方法都是必要的。

注意这里讨论的不同类型的效度是指测量过程中的效度，这不同于假设检验过程的效度，如内部效度 （因果关系），外部效度（普遍性），或统计结论的效度。假设检验过程的效度将在后续章节中讨论。

表面效度。表面效度指一个指标是否在“表面上”对潜在构念的合理测量。例如，无需做过多解释就可以理解一个人参加宗教祈祷仪式的频率可以作为一个人虔诚的指标。因此这一指标具有表面效度。然而，如果我们想要把员工从单位图书馆借书量作为测量员工士气的指标，那么这种测量很可能缺乏表面效度，因为这讲不通。有趣的是，一些在组织研究中广泛应用的测量似乎也缺乏表面效度。例如，一个组织的吸收能力（为改进组织流程而能吸收的新知识的数量）经常用研发能力（即，研发费用除以总收入）来测量。如果研究包含非常抽象的构念或难以区分的构念（如同情心和同理心），则值得考虑邀请一组专家来评估构念测量的表面效度。

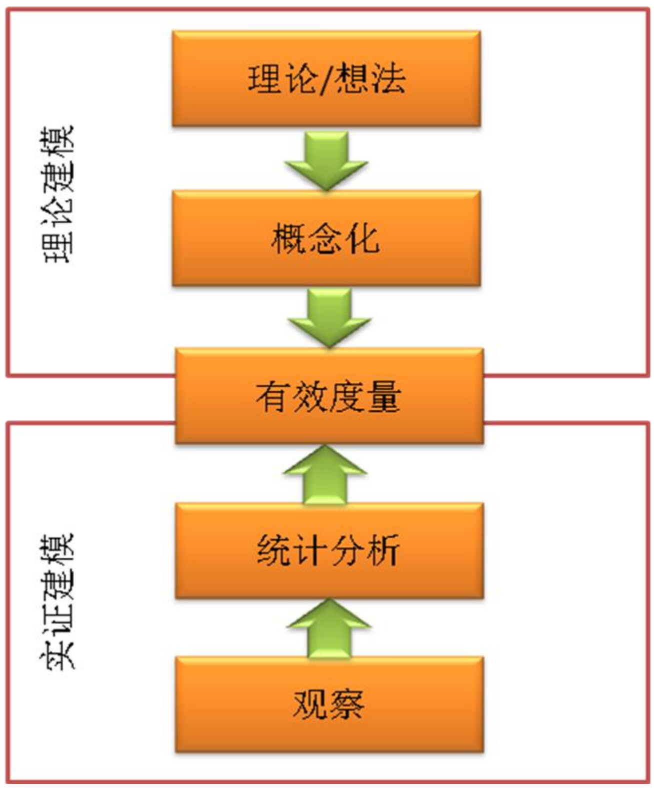


图 7.2 效度评估的两种方法

内容效度。内容效度是对一系列测量项与构念欲测的内容范围的匹配程度的评价。例如，若要对“饭店服务满意度”进行测量，可将其饭店服务的内容范围限定为食物的质量、侍者的态度、以及饭店的整体环境（是否嘈杂、有烟味等）；为了达到足够的内容效度，需要建立一系列的指标来衡量饭店内顾客对食物质量、侍者态度以及等待时间长短、饭店整体环境等的满意度。当然，这种方法需要对一个构念的内容范围进行描述，这对诸如自尊、智力等较复

杂的构念而言是非常困难的，所以对内容效度不太可能进行充分的评价。像表面效度一样，可以请一组专家来对构念的内容效度进行评价。

聚合效度是指测量与其意图测量的构念之间的接近程度，区分效度是指测量与未意图测量的构念之间的分歧程度，通常，聚合效度与区分效度被运用于一组相联系的构念。例如，如果你期望组织知识与组织组织绩效是相关的，那么该如何确定对组织知识的测量确实测量了组织知识（聚合效度）而不是组织绩效（区分效度）呢？聚合效度可以通过比较同一构念的各不同指标观测值之间的相似度（或高相关性）来衡量。区分效度则可以通过比较不同构念的指标差异来衡量，即具有较低的相关性。在上例中，如果组织知识和组织绩效各有三个项目，根据样本数据，我们可以通过软件对组织知识和组织绩效的每对项目做二元相关分析，如果组内的相似度很高，而组间的相似度很低，那么我们就同时实现了聚合效度和区分效度（如表 7.1 所示）。

	KL1	KL2	KL3	PF1	PF2	PF3
KL1	1.00	0.83	0.79	0.23	0.21	0.19
KL2		1.00	0.75	0.11	0.20	0.03
KL3			1.00	0.03	-0.11	0.17
PF1				1.00	0.84	0.91
PF2					1.00	0.77
PF3						1.00

同一构念的各指标间
具有较高的相关性
（聚合效度）

不同构念的各指标间
具有较低的相关性
（区分效度）

表 7.1 聚合效度与区分效度二元相关分析

另一个替代的方法也是更普遍使用的方法，即采用探索性因子分析来评价聚合效度和区分效度。作为一种数据精简方法，该分析基于上文中所提到二元相关系数结构，并通过主成分分析方法将一组给定指标聚合成更少的因子集。这些因子应完美地对应于我们所试图测量的理论构念。因子提取的一般标准是每个被提取的因子其特征值应高于 1。根据潜在构念是否相关，可以对被提取的因子进行斜交旋转或正交旋转，以获得可用于将各构念的指标聚合为复合测量的因子权重。

正如图 7.2 中旋转负荷矩阵的例子，为了达到足够高的聚合效度，同一构念的各指标应在单个因子上具有 0.6 或更高的因子负荷（也叫做同因子负荷）。而为了达到更高的区分效度，这些指标应在所有其他因子上具有 0.3 或更低的因子负荷（也叫做跨因子负荷）。

多质多法（MTMM）方法是评价聚合效度和区分效度的更为精细的方法，它要求对每个构念运用两种以上的方法（如调查、个体观察，或者对教师和父母两组不同的被试进行调查）。这是一种比较繁琐并且实际运用不多的方法，所以在这里不再赘述。

效标关联效度通过变量是否能很好与当下或将来的标准相关联来衡量，包括同时效度和预测效度。预测效度是指测量在多大程度上预测了理论的预期结果，例如：成绩（SAT 成绩）是否能用来预测学生将来的学术成就(GPA)？ 衡量预测效度需要设计反映构念理论联系的关系网。同时效度衡量的是某个指标在多大程度上与同时发生的其他标准相关联，例如：学生的微积分成绩与线性代数成绩之间是否有关联，这两门课程都属于数学课程，因此其成绩应同时相关。与聚合效度和区分效度不同的是，同时效度与预测效度在实证研究中通常忽略不计。

	Factor1	Factor2
KL1	0.88	0.13
KL2	0.93	0.11
KL3	0.87	0.03
PF1	0.17	0.93
PF2	-0.03	0.85
PF3	0.07	0.78

同因子负荷高（聚合效度）

跨因子负荷低（区分效度）

表 7.2 聚合效度与区分效度的探索性因子分析

测量理论

了解各种效度与信度之后，让我来用数学表示的方法（经典检验理论或真值理论）对信度与效度的相关理解进行综合。这是一个用来研究测量是如何工作的，其测量了什么以及没有测量什么的一个心理测量理论。该理论假设每次观测有一个真实值 T，这个真实值可以在没有测量误差的情况下精确观测得到，然而，观测过程通常会存在误差 E，这些误差的存在使得观测值 X 与真实值 T 之间存在偏离。

X = T + E

观测值 真实值 误差

根据观测结果，观测值与真实值方差的关系可以用相似的公式来表达：

var（X）= var（T）+ var（E）

心理测量分析的目标是估计并尽可能地减小 var（E），这样观测值 X 可作为真实值 T 的一个好的测量。

测量误差分为两种类型：随机误差和系统误差。随机误差是指由不可预知和不可控制的偶发外部因素所引起的误差。比如，在测量过程中，一些被试的情绪状态可能比其他被试好一些，这将会影响被试对测量项目的反应，情绪好的被试对自尊、满意度以及幸福指数等构念项目的反应可能比情绪状态差的人积极，然而，预测或者控制被试的情绪状态是不可能的；

同理，在进行组织层面的测量时，如测量企业绩效，制度和环境变化可能会影响观测样本中的一些企业，而其他的企业则不受影响，因此，随机误差被视为测量中的“噪音”，通常忽略不计。

系统误差是指由系统整体因素所引起的，影响样本中所有个体观测值所形成的普遍误差。接我们前面的例子，相比于制造业或者服务业企业，金融危机更多的影响了金融企业的绩效，假设样本只包含金融企业，我们预期金融危机给样本带来了普遍性的绩效下降。随机误差可能是正向、负向或者是零，然而系统误差则会表现为样本中所有个体观测值的同向变化（正向或者负向），因此，系统误差通常被认为是偏差，且需要得到修正。

由于观测值可能包含随机误差和系统误差，可将上述的真实值等式修正如下：

X = T + Er + Es

Er 、Es 分别代表随机误差和系统误差，随机误差会对观测值分布的差异性指标（如方差）产生影响，不会影响居中趋势；而系统误差则会影响居中趋势（如均值），不会影响差异性指标（如图 7.3 所示）：

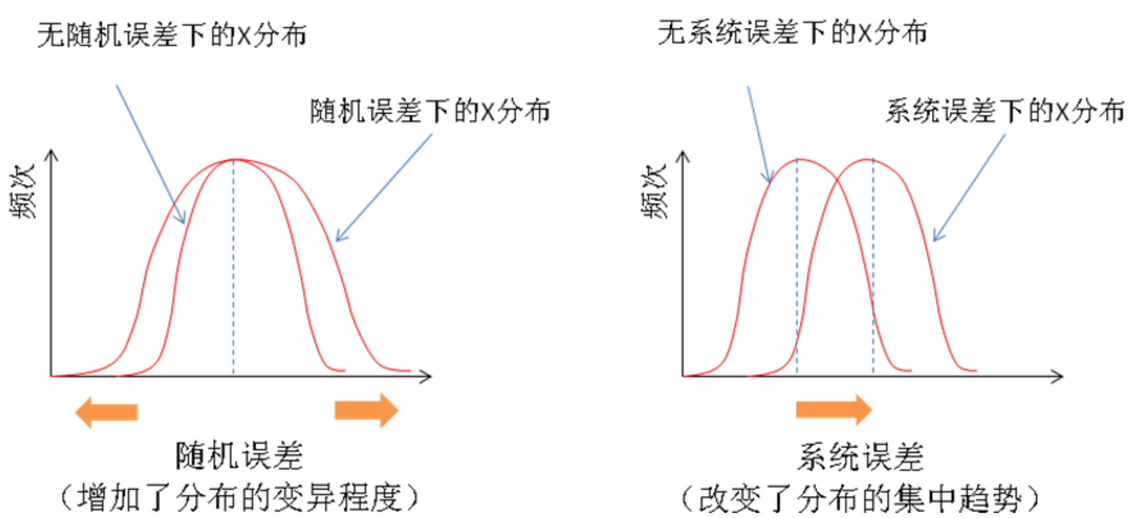


图 7.3 随机误差与系统误差对分布的影响

随机误差和系统误差对测量过程意味着什么呢？随机误差通过增加观测值的变异程度使得测量的信度降低，相反地，系统误差通过影响居中趋势降低测量的效度。

在测量中，效度通常比信度更重要，这是由于低效度的测量意味着可能测量的并不是我们试图测量的构念，因此，效度对基于数据分析的研究特别关键。

由于信度是真实值与观测值之间的比例关系，因此，信度可以由下式表示：

var（T） / var（X） = var（T） / [var（T） + var（E）]

如果真实值与观测值的方差相等，那么信度为 1。

测量效度的整合方法

对效度进行充分的测量需要同时使用理论方法和实证方法。图 7.4 是一个评价量表信度与效度的流程图。

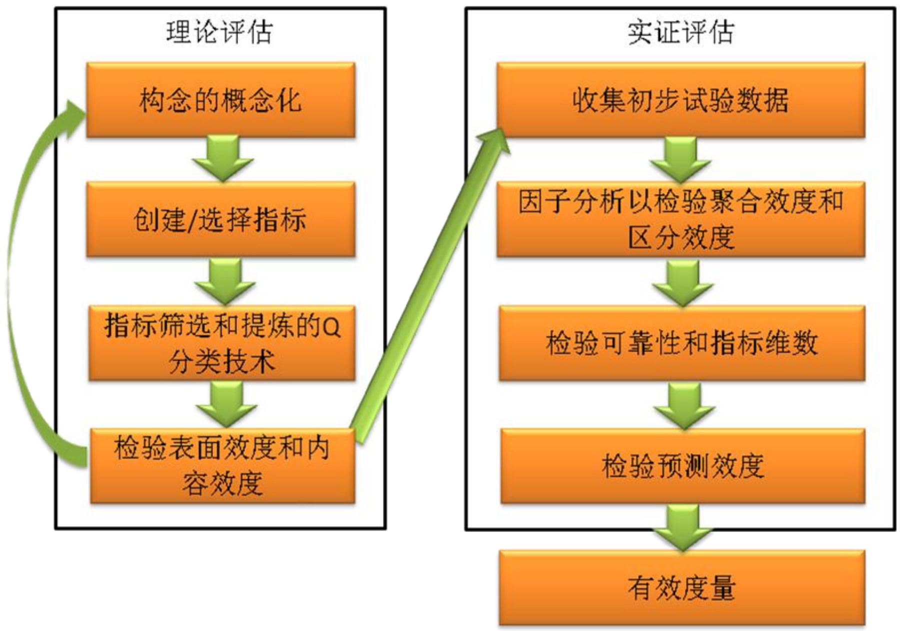


图 7.4 测量有效性的整合方法

整合测量始于理论方法，第一个步骤是将构念概念化，这包括给构念下定义并限定构念范围（维度）；接下来需要根据概念化的过程来确定指标或者项目（如第 5 章的量表过程所述）。在确定指标的过程中，文献述评是非常有益的，应以简单易懂的文字对每个项目进行一致描述。

接下来的一步是评判小组进行评测（研究方法领域的学者或具有典型性的一组被试），这个测评小组将对每个指标进行检验并进行 Q 分类分析。评判会得到写有构念及其概念化定义和操作指标的卡片（每张卡片上写有一个指标），然后他们将独立地阅读每张指标卡，检验其明晰度、可读性以及指标语义，并根据所得到的构念定义，将每张指标卡归类到最合适的构念之中。

评判间信度用以测量评判对分类的一致性。多个评判未领会的具有歧义的项目可能需要重新检验、重述或者剔除。评判间信度较高的项目（一般 10 到 15 个）可以保留以备进一步的分析。保留的项目需经过评判对表面效度和内容效度进行进一步检验，如果在这一阶段达标的项目不够，则需根据概念化的定义引进新的项目，一般来讲，要得到最能代表构念且评判间信度高的一系列项目需要两到三轮的 Q 分类。

随后进行的是实证领域的有效性考察，量表需包含所有精确的项目，并从目标人群中选取具有代表性的被试组成试验性被试组进行试验，由此收集到的数据应以表格列示并运用软件

程序（如 SAS 或者 SPSS）进行相关分析或者探索性因子分析来评价聚合效度和区分效度。不满足因子载荷标准的项目（同因子载荷高于 0.6，交互因子负荷低于 0.3），剩下的项目需利用内在一致性系数（如克隆巴赫系数）进行信度测量。

量表维度也需要校验，这取决于目标构念被概念化为单维度还是多维度。然后，运用回归分析和结构方程模型对特定理论关系网中构念的预测能力进行评价，如果构念测量满足了本章中所述的大部分信度和效度，那么基本可以确信量表是精确和合理的。

对量表有效性的综合测量需要花费研究者大量的时间和精力，不过，为了确保量表符合研究规范，上示测量程序是不可或缺的，由于有缺陷的量表会导致无意义的结果，量表的有效性测量仍然是实证研究中的重要领域。

第八章

抽样方法

抽样是在感兴趣的总体中选择子集（样本）的统计过程，以达到对该总体样本进行观察和统计推断的目的。社会科学研究一般是关于推断特定人群的行为模式。由于可行性和成本约束我们不能研究整个群体，因此，我们必须从总体中选择具有代表性的样本来进行观察和分析。选择真正代表总体的样本是极其重要的，这样从样本得出的推论才能推广回到总体。民意调查和投票后民意调查报告显示不适当和有偏的样本是经常性分歧和错误推论的主要原因，这是由 CNN/Gallup Poll、ABC、以及 CBS 等不同的民意调查团体在每次美国总统选举前实施的。

抽样过程

图 8.1 抽样过程

抽样过程包括几个步骤。第一步定义目标总体。总体可以被定义为所有的有所要研究特点的人或物体（分析单元）。分析单元可以是一个人、群体、组织、国家、物体、或是想要从中得到科学推断的其他任何实体。有时总体是显而易见的。例如，如果一个制造商想要确定生产线上制造的产成品是否符合一定的质量要求或是必须废弃或返工，那么该总体由该生产线制造的全部完工产品构成。其他时候，推断目标总体可能会有点困难。如果想要识别高中生学术学习的主要驱动因素，那么目标总体是什么呢？是高中生、老师、学校领导还是家长？这种情况下的正确答案是学生，因为你感兴趣的是学生的成绩，而不是老师、家长或学校的成绩。同样地，如果你想分析赌轮的性能来识别有问题的轮子，总体就不是单个赌轮的不同观察，而是不同的赌轮（即无数个轮子的性能）。

抽样过程的第二个步骤是选择抽样框。这是从目标总体（通常是联系方式的清单）中选取一个易接近的部分。如果目标总体是工作的在职人员，因为不能接近世界上所有的在职人员，一两个愿意参与研究的本地企业的雇员名单是一个更为实际的抽样结构。如果目标总体是组织，那么世界财富 500 强的企业清单或在纽约证券交易所注册的标准普尔企业清单就是可以被采用的抽样框。

注意，抽样框可能没有完全代表当前的总体，如果是这样，从这样的样本中得出的推论可能不能推广回到总体。例如，如果你的目标总体是组织当前的雇员（即，你想要研究该总体中雇员的自尊），而抽样框是美国中西部汽车企业的雇员，从这些群体中得到的结果甚至可能不能推广到美国当前的劳动力，更不用说全球的工作环境了。这是因为美国汽车行业近 50 年来一直面临着严峻的竞争压力，已经经历过很多次重组和缩小规模，这可能导致低的员工士气及自尊。此外，大多数的美国劳动力从事服务行业或是小型的商业，而不是汽车行业。因此，美国汽车行业的员工样本不能用来代表美国的劳动力。同样地，世界财富 500 强名单包含了美国最大的 500 家企业，而这并不能代表美国的一般企业，美国的大部分企业是中小企业而不是大企业，因此这是一个有偏的抽样框。相反，标准普尔企业清单将允许选择大型、中型和/或小型的企业，这取决采用的是标准普尔的大型、中型还是小型企业清单，但是它只包括上市企业（而非私人企业），因此仍有偏差。还应指出，从总体中抽取的样本未必和总体一样与我们真正想要的信息相关。例如，如果一个研究者想要获得新型“戒烟”项目的成功率，那么目标总体是已经利用该项目的所有吸烟者，这可能是一个未知的总体。因此研究者可能以到当地戒烟机构戒烟的病人为样本，他们中的一些人可能并没有接触到这种特殊的“戒烟”项目，在这种情况下，抽样框与总体并不相符。

最后一步就是用明确定义的抽样技术从抽样范围中选取样本。抽样技术可以划分为两大类：概率（随机）抽样和非概率抽样。如果结果的普遍性对研究很重要，概率抽样是最理想的，但是可能存在特定的情况，非概率抽样也是合乎情理的。这些技术将在下面的两部分中讨论。

概率抽样

概率抽样技术是指总体中的每个单元在抽样时都有被抽到的机会（非零的概率），并且这个机会是可以准确确定的。样本统计量就这样产生了，比如样本均值或标准差，只要根据样本单元的选取概率对它们赋予权重，样本统计量就是总体参数的无偏估计。所有概率抽样有两种共同的属性：（1）总体的每个单元的中选概率是已知的、非零的；（2）抽样程序包含随机选择采样点。概率抽样的不同类型包含：

简单随机抽样。在该技术中，总体（更准确地说应是抽样范围）的所有可能的子集被选中的概率都是均等的。从总数为 N 的抽样范围中选择任何 n 个样本的组合的概率是 C_N^n。因此，不采用任何加权，样本统计量是总体参数的无偏估计。简单随机抽样包括从样本范围内随机选择受访者，但对于大型的抽样范围通常会用到随机数表或计算机随机号码发生器。例如，如果想从有 1000 家企业的清单中选择 200 企业来调查，若这份清单在一个电子表格软件如 Excel 中，就可以 Excel 的 RAND 函数来为清单上 1000 个客户生成随机编号。接着，将它们按相应的随机数以升序的方式排序，然后在排过序的清单中选择前 200 名客户。这是所有概率抽样技术中最简单一种；然而，简单也正是这个技术的优点。因为抽样范围并没有进行细分或分割，样本是无偏的，并且该推论在所有概率抽样技术中最具普遍性。

系统抽样。在该技术中，抽样范围是根据某个标准有序排列的，并且样本是按固定的间隔从有序的列表中选取的。系统抽样涉及随机的起点，然后从起点向后每隔 K 个样本进行选取，这里 K= N/n ，称作采样率，是抽样范围规模 N 以及理想样本规模 n 的比值。清单上的第一个样本不能自动成为起点，而要从清单上前 K 个样本中随机选择一个作为样本起点。前面从 1000 家企业中选择 200 家企业的例子中，可以按照规模（员工数或年营业收入）将企业按升序（或降序）排列，在有序列表的前五家企业中随机选择一家企业作为起点，然后选择列表上的每个第五家企业。这个过程将确保样本中没有过多选取大型或小型企业的样本，所有规模的企业都被均等选取，正如抽样范围一样。换句话说，样本代表了总体，至少在分类的基础上是这样的。

分层抽样。在分层抽样中，抽样范围被分成同类的和不交叉的子集（称为"层"），再按照简单随机抽样方法在每个子集中抽取样本。前面从 1000 家企业中选择 200 家企业的例子中，可以先按照规模将企业分成大型（超过 500 名员工）、中型（50 ~ 500 名员工）和小型（少于 50 名员工）三个子集。然后在每个子集中随机选取 67 家企业来组成 200 家企业的样本。然而，由于在样本范围内，与大型企业相比存在更多的小型企业，选取相同数目的小型、中型和大型企业将会使样本具有较低的代表性（即，考虑到目标总体中大型企业更少，这将是有偏的）。这就是所谓的不成比例的分层抽样，因为每个子集的样本比例没有反映其在抽样范围（或总体）中的比例，较小的子集（大型企业）被过多取样了。一个替代技术是根据它们在总体中的比例在子集中选取样本。例如，如果有 100 家大型企业、300 家中型企业、600 家小型企业，可以从大型组中选取 20 个企业，从中型组中选取 60 个企业，从小型组中选取 120 个企业。在这种情况下，总体中企业的比例分布在样本中得到保持，因此，该技术称作比例分层抽样。注意，非比例分层抽样方法在代表小的子集如大型企业时非常有效，只要根据子集在总体中的比例对非比例方式的结果赋予权重，与比例分层抽样相比，非比例分层抽样方法未必不能代表总体。

整群抽样。如果总体分散在一个大的地理区域，对整体实施简单随机抽样是不可行的。在这种情况下，将整体分"群"（通常按地理界线）是合理的，随机抽取几个群，然后测量群中的每个样本。例如，如果你想以纽约州的城市政府为样本，而不是周游整个州去采访那些关键城市的官员（如在简单随机抽样中必须要做的），可以基于郡县将这些政府分群，随机选取三个郡县，然后采访这些郡县的每一个政府官员。然而，由于群体间的差异，群体样本中样本估计的变异性通常会比简单随机样本的变异性高，因此与从简单随机样本的到的结论相比，该技术的结果不具有普遍性。

配对抽样。有时，研究者基于特定的标准想要比较总体内两个子集。例如，为什么一些企业始终比其他企业更有利可图？实施这样的研究需要根据毛利率、每股收益或其他盈利衡量指标将抽样范围内的企业分成两类："高盈利"企业和"低盈利"企业。接着，在一个子集中随机

选择样本，再根据规模、行业分部、和/或其他匹配标准将这个组中的每个企业与第二子集中的某个企业配对。这样，就有两个配对的高利润和低利润企业样本供更详细的研究。这种配对抽样技术通常是了解特定总体中不同子集间差异的理想方式。

多级抽样。先前讨论的概率抽样技术都是单级抽样技术的例子。根据样本需求，可以将这些单级抽样技术结合起来实施多级抽样。例如，可以基于企业规模将清单上的企业分层，然后在每层实施系统抽样。这是将分层抽样与系统抽样结合起来的两级抽样。同样地，可以先将纽约州的学区分群，在每个群中，用简单随机抽样选取学校；在每个学校中，用简单随机抽样选取年级；对每个年级，用简单随机抽样选取学生来进行抽样。在这种情况下，就有一个包括整群抽样和简单随机抽样的四级抽样过程。

非概率抽样

非概率抽样是指一些分析单元被抽取的概率是零或是选取的概率不能准确确定的抽样技术。通常，分析单元以某种非随机的标准选取，如配额或便利性。由于选择是非随机的，非概率抽样不能进行抽样误差的估计，并可能受限于样本偏差。因此，样本的信息不能被推广回到总体。非概率抽样技术包括以下几种类型：

便利抽样。也称作偶遇抽样或机会抽样，这种技术是指从近手的、唾手可得的或便利的部分总体中抽取样本。例如，如果在一个购物中心的外面向人们分发问卷，或是在他们进入购物中心时采访他们，这样获得的受访者样本就是便利样本。这是非概率抽样，因为这在系统上排除了到其他购物中心购物的人。从选择的样本中获得的观点可能反映了这个购物中心的独特特点如商店的性质（即，高端商店将吸引更富裕的阶层）、顾客的社会阶层或位置（靠近大学的购物中心主要吸引有特定购物习惯的大学生），因此它不代表当前消费者总体的观点。因此，这些观察的科学性将是有限的。其他便利抽样的例子如在某个课程注册的学生样本或在某个医疗诊所治病的病人样本。这种类型的抽样对引导测试是最有用的，其目标是指标测试或测量检验而不是获得普遍性的推论。

配额抽样。在这种技术中，总体被划分为相互排斥的子集（正如分层抽样），然后在每个子集中选取一组非随机的观察来满足事先确定的配额。在比例配额抽样中，每个子集的受访者比例应当与总体匹配。例如，如果美国人口由 70%的白人、15%的西班牙裔美国人、13%非洲裔美国人组成，若想了解样本数为 98 的美国人的投票偏好，可以在购物中心外面询问人们的投票偏好。但当你有了 15 个西班牙裔（或 13 个非洲裔）的回应后你必须停止对这个子集的询问，即便仍在抽取其他种族的样本，这样样本的种族构成与一般的美国人口构成相匹配。非比例配额抽样的限制更少些，这是因为没有必要获得等比例样本，但是或许会要符合每个子集最小数目的要求。在这种情况下，你可能会决定分别从三个种族子集（白人、西班牙裔美国人、非洲裔美国人）中选取 50 个受访者，并在每组配额达到时停止取样。两种配额抽样的类型都没有代表美国人口总体，这是因为，根据你在纽约还是在堪萨斯州实施研究，结果可能会完全不同。非比例配额抽样技术甚至更不具有代表性，但可能会更有用，因为它通过过多取样能够获取小的、被低估了的群体的观点。

判断抽样。在这种技术中，受访者是专家根据他们对所研究现象的经验而选取的，是非随机的。例如，为了解新政府政策如萨班斯-奥克斯利法案的影响，可以抽取一组熟悉该法案的

公司会计人员。这种方法的优点在于专家比非专业人士更了解该事项，来自专家样本的意见比包含专家和非专业人员的样本更可信，尽管结果仍然不能推广到总体。

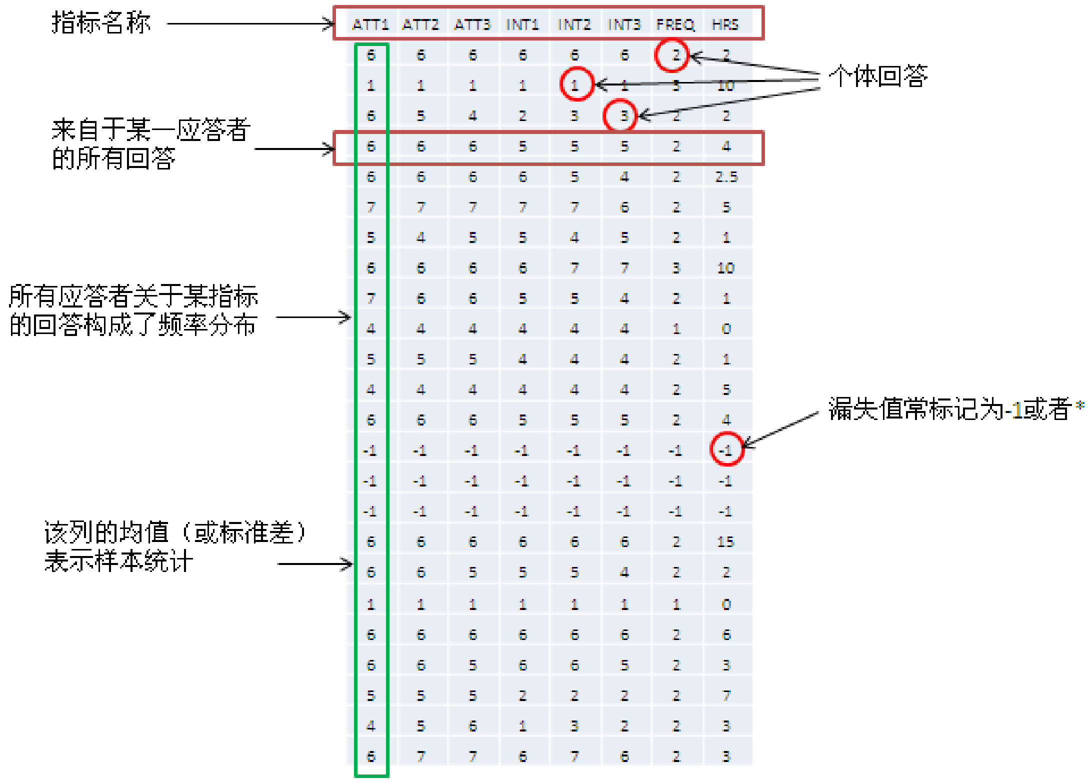
滚雪球抽样。在滚雪球抽样中，先找出符合研究标准的受访者，然后要求他们推荐他们知道的符合选择标准的其他人。尽管这种方法难以实现样本的代表性，它有时是接近难以接近人群、或是没有抽样框时的唯一方法。

抽样统计

在前面的部分中，我们介绍了诸如总体参数、样本统计量、样本偏差等术语。在本部分，我们将试图了解这些术语的意思以及它们如何相互关联。

当测量一个给定的分析单元的某个观察时，比如一个人对李克特式量表项目的回应，那个观察就叫做回应（见图 8.2）。换句话说，回应是一个单元样本提供的测量值。每个受访者对一张表上的不同的项目有不同的回应。可以根据不同受访者对相同项目或观测的回应发生的频率绘制频率分布图。对于样本中的大多数回应，这个频率分布趋向于一个钟形曲线，称作正态分布，它可以用来估计全部样本的特性，如样本均值（样本中所有观测的平均数）或标准误（样本中所有观察的变异性或差幅）。这些样本估计称作样本统计量（"统计量"是一个值，这个值就是观测数据的估计）。如果能够对全部总体取样，也将获得总体的均值和标准误。然而，总体不能全部取样，总体特征总是未知的，它们称作总体参数（而不是"统计量"，因为它们不能通过数据的统计分析得到估计）。如果样本不能完美地代表总体，样本统计量可能与总体参数不同。两者间的差异称作抽样误差。从理论上讲，如果我们可以逐渐增加样本容量，样本越来越接近总体数量，那么抽样误差将会降低，样本统计量将与相应的总体参数越来越近似。

如果一个样本真正代表了总体，那么估计的样本统计量应该与相应的理论总体参数完全相同。我们怎么才能知道样本统计量是否至少合理地接近总体参数？ 在这里我们需要了解抽样分布的概念。假设从给定的总体中选取三个不同的随机样本，如图 8.3 所示，并推断出每个样本的样本均值和标准误等样本统计量。如果每个样本都真正代表了总体，那么三个随机样本的三个样本均值将是相等的（并且与总体参数相等），样本均值的标准差将是零。但是这是极不可能的，因为每个随机样本可能由总体的不同子集构成，因此，它们的均值可能会有微小的不同。然而，可以取这三个均值绘制样本均值的频率直方图。如果这些样本的数量从 10 增加到 100，频率直方图成为抽样分布。因此，一个抽样分布是一系列样本的样本统计量的频率分布，而作为参考的一般频率分布是单个样本的回应（观测）的分布。就像一个频率分布，抽样分布也往往有更多聚集在均值（当然是总体参数的估计）周围的样本统计量 。无限大的样本的分布将接近正态分布。抽样分布的样本统计量的变异性或差幅（即，抽样统计量的标准差）称作标准误。相比之下，标准差这个术语用于单个样本观测值的变异性。



ATT1	ATT2	ATT3	INT1	INT2	INT3	FREQ	HRS
6	6	6	6	6	6	2	2
1	1	1	1	1	1	3	10
6	5	4	2	3	3	2	2
6	6	6	5	5	5	2	4
6	6	6	6	5	4	2	2.5
7	7	7	7	7	6	2	5
5	4	5	5	4	5	2	1
6	6	6	6	7	7	3	10
7	6	6	5	5	4	2	1
4	4	4	4	4	4	1	0
5	5	5	4	4	4	2	1
4	4	4	4	4	4	2	5
6	6	6	5	5	5	2	4
-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1
6	6	6	6	6	6	2	15
6	6	5	5	5	4	2	2
1	1	1	1	1	1	1	0
6	6	6	6	6	6	2	6
6	6	5	6	6	5	2	3
5	5	5	2	2	2	2	7
4	5	6	1	3	2	2	3
6	7	7	6	7	6	2	3

图 8.2 样本统计量

所在抽样分布中抽样统计量的均值被认为是未知总体参数的估计。基于抽样分布的差幅（即，基于标准误），也可以估计总体参数预测的置信区间。置信区间是总体参数落于某个特定样本统计值区间的概率。有的正态分布倾向于遵循 68-95-99 百分比法则（见图 8.4），它表示在这个分布中超过 68%的样本分布在均值的一个标准偏差范围之内（μ+ 1σ），超过 95%的样本分布在均值的两个标准偏差范围之内（$\mu + 2\sigma$），以及超过 99%的样本分布在均值的三个标准偏差范围之内（$\mu + 3\sigma$）。

因为有无限样本数的抽样分布接近正态分布，68-95-99 法则也同样适用于抽样分布，可以这样表述：

（样本统计+标准误）代表了总体参数 68%的置信区间。

（抽样统计+两个标准误）代表了总体参数 95%的置信区间。

（抽样统计+三种标准误）代表了总体参数 99%的置信区间。

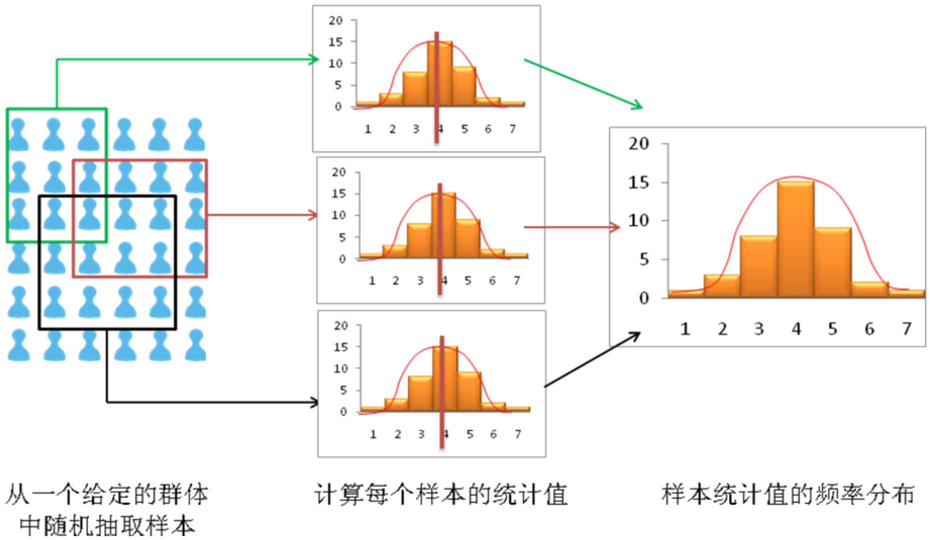


图 8.3 抽样分布

如果抽样分布无法估计或违反了 68-95-99 百分比法则，样本就是有偏的（不能代表总体）。另外值得注意的是，在大多数用 $p<0.05$ 检验系数显著性的回归分析中，我们正试图看看抽样统计（回归系数）是否在 95%的置信区间预测了相应总体参数（真实数值）。有趣的是，“六个 δ”标准尝试识别 99%的置信区间以外的或是六个标准差（用希腊字母 δ 表示），代表 $p<0.01$ 时的显著性检测。

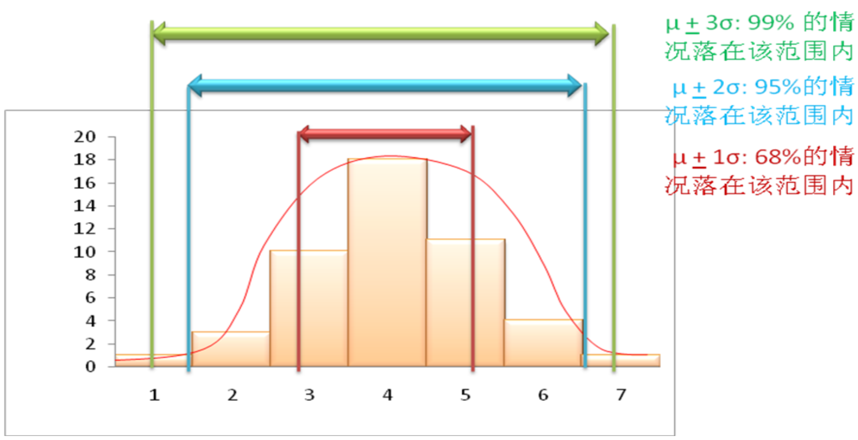


图 8.4 置信区间的 68-95-99 百分比法则

第九章

调查研究

调查法指通过使用标准化问卷或访谈的形式对人们的偏好、想法和行为进行系统化数据收集的一种研究方法。尽管调查法早在古埃及已经被采用，但其作为一种正式的研究方法则应追溯到 1930-40 年代的社会学家鲍尔-拉扎斯菲尔德，他利用该方法分析了无线电设备对美国政见形成的作用。从那时起，该方法得以在社会科学的定量研究中广泛流行。

调查法可以用于描述性、探索性或解释性的研究。该方法最适用于那些以个体作为分析单元的研究。尽管基于其他分析单元，如群组层面、组织层面或者双向关系层面（成对的组织，例如买方与卖方）的研究也可以使用调查法，这些研究经常将分析单元中的某个特定人物作为分析单元的"关键信息源"或者"代理人"。由于所选择的调研对象对于所调研的现象未必有足够的知识或者有偏差性的观点，这样的调研容易产生应答偏差。例如，首席执行官可能并不太了解其员工的观点或者公司的团队合作情况，因此选取他们作为调研对象来研究组织动态性或员工自负感将不太合适。

调查法相对于其他研究方法具有一些固有的优势。第一，调查法是测量各类不可观测数据的极好工具，诸如人们的偏好（如政治导向），个性（如自尊），态度（如对于移民的态度），信念（如对于新法律的看法），行为（如吸烟与酗酒行为）或事实信息（如收入）。第二，调查法也非常适合对一个庞大而不易直接观测人群进行远程数据收集。通过信件、电子邮件、或者电话调查等方式可以对一个大的区域如整个国家的群体进行覆盖，此时应注意采用认真的采样设计以保证该样本能够被小样本所代表。第三，由于问卷调查具有非冒犯性以及允许被试根据其自身便利性来回答问题的能力，被试较偏好于问卷调查法。第四，对于某些特定人群如无家可归者或者非法移民者，访谈调查法也许是唯一可行的方法，因为不存在固定的抽样框。第五，大样本调查使得在进行多元变量分析时即使很小的效果亦能被检测出来，此外根据调查设计，大样本调查还有利于不同子群体之间进行比较分析，例如组内比较与组间比较。第六，调查法相较于其他研究方法如实验法与案例法而言，其在研究者的时间、努力与成本上都较为经济实用。同时，调查法也有其缺点，主要是该方法可能导致一系列偏差的产生，如无应答偏差，抽样偏差，社会称许性偏差，回忆偏倚等，这些我们将在后文进行详细讨论。

根据数据的具体收集方式，调查法可以进一步分为两大类：问卷调查法如邮件调查法、集体调查法与在线调查法等，以及访谈调查法如个人、电话或焦点组访谈法等。问卷是指让被调查者根据量表回答相关问题，而访谈是指访谈者与被调查者进行语言沟通所获得的回答。正如下文所将讨论的，每种方法根据其成本、目标群体覆盖率以及研究者在提问时的灵活性等方面都有其优势与劣势。

问卷调查法

问卷调查法最早为弗朗西斯•高尔顿所创立，问卷指的是由一系列标准化地捕捉被调查者的反应的问题所组成的研究性量表。问题可以是非结构化的或结构化的。非结构化的问题让被试用他们自己的语言来回答，而结构化的问题则要求被试从给定的选项集中选择一个答案。被试对一个结构化的问卷上的问题的回答可能将加总为一个综合指标以进行统计分析。在设计问卷时，问题应该能让被试阅读和理解以便他们可以有效地回答问题，因此，问卷调查法可能对某些人群例如儿童与文盲并不适用。

大部分问卷倾向于采用自填式邮件调查法。采用该法，研究者将同样内容的问卷邮寄给大量潜在被试并邀请他们在方便时给予填写，被试将填好的问卷放入已支付邮资的信封中寄回研究者。邮件调查具有非冒犯性的优势，并且该法的管理成本很低因为大宗邮资在大多数国家是很便宜的。尽管如此，邮件调查的回收率可能会非常低因为大部分人会忽视调查请求。此外，被试填写并返回问卷的延迟也可能会比较久（几个月），或者他们将这些问卷遗失。因此，研究者必须不断地监控已经回收的问卷，并多次对未回收的问卷进行追踪和发送提醒，如发送两到三次提醒信，每次间隔一个半月。问卷调查对于那些需要对被试进行详细阐明以及需要被试提供详尽回答的研究情景并不适合。纵贯性设计可以用于那些针对同一被试群体在不同时间点的调查，但是从一个时间点的调查到下个时间点的调查，问卷的回收率将陡然下降。

第二类调查类型是群填式问卷。被试们在同一时间被聚集在同一个地方，然后对调查问卷进行填写。每个被试在不与他人沟通的情况下各自独立填写自己的问卷。这种方式对研究者来说比较方便，因为回收率比较高。如果被试对某些特定问题不理解时，他们可以要求研究者进行阐明。在许多组织中，将员工们安排在一个会议室或午餐室进行调查是相对容易的，特别是在该项调查得到公司高层的批准时。

最近的一种问卷调查方式是采用网络在线调查。这些调查可以利用互联网上的互动式表格来进行管理。研究者可以给被试发送电子邮件邀请其参与调查，在邮件中附上在线调查的问卷网站的链接。或者，研究者可以将问卷包含在电子邮件之中，然后让被试在完成问卷后用电子邮件回答。这些调查方式管理成本非常低，调查结果可以马上记录在在线数据库中，所以调查问卷可以在需要时方便地进行修改。尽管如此，如果调查网站没有采用密码保护或多次提交限制，回收问卷的质量将大打折扣。此外，抽样偏差将是该法的重大问题，因为该调查并不能送达到那些没有电脑或互联网接入权的人群手中，例如那些贫穷、年老和少数派群

体，从而使得回应样本偏向于那些具有大量时间上网并且有能力完成这些调查的年轻人。如果调查是贴在一个讨论组或公告板上而非直接发电子邮件给目标被试时，问卷回收率的计算也将是一个问题。基于这些原因，许多研究者偏好于采用双媒体调查法（比如邮件调查和在线调查），这样可以使得被试选择他们偏好的回应方式。

构建一个调研问卷是一项艺术。针对问题的内容、措辞、格式和顺序等需要做出很多决策，这些都将对调查问卷的回答具有重要影响。

回答格式。调查问题可以是结构化的或者非结构化的。对于结构化的问题可以采用下面的回答格式：

- 二分式回答。被试可以从两个可能选项中选择一项，例如真/假，是/否，或同意/不同意。这类问题的一个例子是：你认为死刑在某些情景下是合理的吗（二选一）？是/否。
- 名义型回答。被试需要从两个以上的非排序型的选项中进行选择，例如：你所在公司的产业类型是：制造业/消费者服务业/零售业/教育业/医疗业/旅游与酒店业/其他。
- 序数型回答。被试可以从两个以上的有序的选项中进行选择，例如：你的最高学历是：高中/大学/研究生。
- 连续型回答。被试可以填写一个基于一个有意义的零点的连续值或比例刻度，例如年龄或工作年限。这些回答倾向于采用填空的形式。

问题内容与措辞。调研所获得的回答很容易受到问题形式的影响。粗制滥造或者表意不清的问题将很可能导致毫无价值的回答。迪尔曼（1978）提出了构建好的调研问题的几项规则。调查中的每个问题都需要认真审核如下几项议题：

- 问题是否表意清晰，容易理解：调研问题需要用非常简单的语言来表述，最好采用主动语态，并且避免采用难以理解的词语或术语。问卷中的所有问题需要用相似的方式进行措词以方便被试阅读与理解。唯一的例外情况是，当你的目标被试是那些在他们的日常生活中采用术语的特定人群如医生、律师和研究者。
- 问题是否采用否定句式：采用否定句式的问题（如地方政府是否应该不增税）将使大部分被试感到困惑并引致不精确的回答。这些问题应该予以避免，特别是不要采用双重否定句式。
- 问题是否模糊不清：调查问题的设置应避免不同的被试对于某些措辞有不同的理解（如“有些”或“仅仅”）。例如，如果你问一个被试“你的年收入如何”，被试将不清楚你指的是基本工资或者红利、租金或其他收入，也不清楚你指的是个人收入，家庭收入（包括配偶工资），个人和业务收入？不不同被试的不同理解将导致得到的回答不具可比性。

- 问题是否具有偏向性或累赘的措辞：偏向性指的是问题中任何引导被试按照某一特定方式回答的属性。肯尼思和拉辛基（1989）对几项关于人们对政府支出的态度的研究进行了分析，发现被试对“援助贫困计划”比对“福利计划”表示出更强烈的支持，尽管这两个措词具有同样的意思。在该项研究中，同样发现被试对“阻止犯罪率上升”的支持胜于“加强法制”，对“援助大城市”的支持胜于“解决大城市的问题”，对“毒品康复治疗”的支持胜于“解决毒品上瘾问题”。具有偏向性的语言或语气将扭曲回答。提前预测到具有偏向性的措辞通常比较困难，但是在调查之前需要对调研问题进行细致的检查以尽最大可能防止偏向性的用语。

- 问题是否双重意义：具有双重意义的问题可能导致多种答案。例如，你是否满意为你工作所提供的软件和硬件？在这种情景下，如果被试对硬件满意而对软件不满意，那么其应该如何作答？一般而言建议将具有双重意义的问题分成几个独立的问题：（1）你是否满意为你工作所提供的硬件，和（2）你是否满意为你工作所提供的软件。另外一个例子：你的家庭是否喜欢公共频道电视？一些人可能自己喜欢公共频道电视，但是也可能因为子女的缘故而喜欢有线电视节目例如芝麻街。

- 问题是否太笼统：有时候问题太笼统以至于不能精确地传达被试的想法。如果你问一个人他有多喜欢一本书并要求其根据从“一点也不”到“非常喜欢”这样的度量来作答，当这个人选择“非常喜欢”时，他表达的意思到底是什么？相反地，问一些与具体行为相关的问题，诸如你是否愿意将这本书推荐给他人，或者你是否计划阅读该作者的其他作品？同理，不要问“你们的公司有多大”（不同的人对这个问题的表述不同），而是问“你们公司有多少人”或者“你们公司的年收入是多少”，这些都可以测量公司的规模。

- 问题是否太细：避免问过于细节的但是并不必要的问题，这样做并不能为研究目的服务。例如，你是否需要一个家庭里每个孩子的年龄，或者一个家庭的孩子数量是否可以接受？尽管如此，如果不确定，宁可错在细节也不要错在过于宽泛。

- 问题是否自以为是：如果你问“你认为减税的好处是什么”，你其实在假定被试认为减税是有利的。但是很多人可能并不认为减税是有利的，因为减税将导致对公共学校的财政支持减少，更大的班级规模，更少的公共服务例如警察、急救和消防。避免在问题中存在前提假定。

- 问题是否是虚构的：在电视游戏中的一个极其常见的问题是“如果你在该节目中赢了100 万美元，你计划怎么花掉它？”大多数被试从来没遇到过这么一大笔钱也从未想过如何处理它（大部分人甚至不知道税后他们只能拿到 64 万美元，或者在美国，大多数情况下这笔资金是被分配到 20 年的期限中的，这样按照净现值来算将更少），所以他们的回答将非常随机，例如环游世界，买一个酒店或酒吧，花在教育上，存起来做退休金，帮助父母或子女，或者举行一场奢华的婚礼。虚构性问题导致虚构性的回答，而这些回答并不能用作科学参考。

- 被试是否具有正确回答问题所需要的信息：通常情况下，我们假定被试具有回答问题所需要的信息，但是在现实中，他们并不具有这些信息。即使我们已经得到了他们的答案，鉴于其本身知识的匮乏，这些答案也是不准确的。例如，我们不应该向一个公司的 CEO 询问日常的运作细节，因为他们对这些也许并不知情，也不应该向老师们询问他们的学生学到多少，也不应该问一个高中生"你认为美国政府在猪湾事件中的作为是否合适？"

问题顺序。一般来说，问题应该在逻辑上自然地从一个问题跳到另一个问题。为了获得最好的回收率，问卷需要从最不敏感的问题到最敏感的问题，从事实性或行为性的问题到态度性的问题，从宽泛性的问题到具体化的问题。关于问题顺序的几个常用规则包括：

- 从容易回答、不具威胁性的问题开始，这些问题容易被回想起。例如对于个体层面的调查可以先问人口统计学问题（年龄，性别，学历等），对于企业层面的调查可以先问与企业特征相关的问题（员工数，年收入，产业类型等）。

- 不要从开放式问题开始。

- 如果问题牵涉到多个事件的先后顺序，按照事件的发生时间从早到晚设置问题。

- 每次只问一个话题。当转换话题时，需要加入过渡性说明，如"下面的部分将调查您对…的看法"。

- 需要时选用过滤性或关联性问题，例如："如果你对问题 5 回答'是'，请跳转至第二部分；如果你回答'否'，请跳转至第 3 部分"。

- 其他黄金准则。从被试者的角度出发进行问题设置。尊重与关注被试的时间、注意力、信任与个人信息的保密性。在所有调查研究中采用下述策略：

- 人们的时间非常宝贵。尊重别人的时间。尽量保证你的问卷在满足需求的情况下尽可能的短。被试不愿意在任何一项调查中花费 10-15 分钟以上的时间，不管它有多么重要。长的问卷将显著地降低回收率。

- 保证被试答卷的保密性，告知他们你将如何使用数据（如用作学术研究）以及数据将如何汇报（通常来说只汇报总体结果）。

- 对于组织层面的调查，保证你将调查的最终结果的复印件发送给他们，以信守承诺。

- 感谢被试参与你的调查。

- 最后，在将问卷发放到目标被试之前，一定要对你的问卷进行前测（Pretest），最起码要利用方便性样本进行前测，以发现问卷中的表意不清，阐述不足，或者措辞偏向性等问题，并尽力在正式调研前消除这些偏差。

访谈调查

相较于问卷而言，访谈是一种更为个性化的数据收集方法，其一般是由经过特定训练的访问者采用与问卷类似的研究手段（如标准化的问题）来进行的。不过，与问卷不同，访谈的脚本包含一些被访问者看不到的特殊的指引，并且留有一些供访问者记录个人观察与评论的空白区域。此外，与邮件调查不同，访谈者可以对被调查者提出的问题与困惑进行阐明。访谈者需要具备特定的访谈技巧。作为测量量表的一个组成部分，访谈者需尽力保证其对观察到的答案进行无偏差的记录。

最常见的访谈形式是个人访谈或面对面访谈，访谈者直接向被调查者提出问题并记录他们的回答。个人面谈可以在被调查者的家中或工作场所进行。有些被调查者喜欢这种方法，但另外一些却可能因为陌生人进入他们的家庭而感到不适。不管如何，具有高超技能的访谈者可以说服被调查者进行合作并显著地提升回收率。

个体面试的一个变化形式是群面或者焦点组。采用这种方法时，通常需要将一群被调查者（一般 6-10 人）聚集到同一个地点进行面试。访谈者的角色是促使小组成员进行讨论并保证每个人都有发言的机会。与其他形式的调查方法相比，焦点组有助于深度挖掘复杂问题，因为当一个人听到其他人的发言时，会引起他思考一些本来没有考虑过的方面。不过，焦点组也可能被某个具有统治力的个体所控制，其他人可能并不愿意在同行或上司面前说出自身的看法，特别是针对那些敏感事件时，如员工表现不佳或办公室政治等。由于焦点组的样本量较小，其经常被用作探索性而非描述性或阐释性的研究。

第三种面试调查法是电话调查。采用这种方法时，访谈者基于从一个电话薄上随机选择的号码来联系被调查者，向其询问一系列调查问题。最新的一种基于技术改进的方法是计算机辅助电话访谈。这种方法正越来越多地在学术、政府以及商业领域得以应用。访谈者作为一个电话操作员，根据计算机程序所生成的一系列指示与问题对被调查者进行询问。被调查者也是根据电子拨号技术从众多电话号码中随机选择出来的，其回答可以通过语音俘获技术来进行记录。一旦被调查的电话接通，调查的回收率将比较高。不过这种方法在电话不普及的乡村地区不太理想，而且这种方式也不适用于获取非音频信息如图形或产品展示等。

访谈者的角色。在访谈过程中，访谈者具有复杂和多面的角色，包括下述任务：

- 访谈准备。因为访谈者处于数据收集的最前端，数据收集质量很大程度上依赖于访谈者是否具备良好的训练。访谈者必须在访谈流程与调查方法方面受过良好的训练，同时他们也需要对研究目的、如何对被试的回答进行存储与利用以及可能导致访谈偏差的原因非常熟悉。在正式调研之前，他/她还需要进行预演并控制时间。

- 定位受访者并寻求其合作。特别对于那些在家庭进行的个体面试，访谈者必须能找受访者的具体地址，避开那些受访者不愿接受访谈的时间如周末。他们还需要向销售人员一样，想法设法让受访者参与到调查中来。

- 激励受访者。受访者往往因为访谈者的激励而参与调查。如果访谈者本身并不感兴趣或不专心，受访者也将不会受激励而提供有用或具有信息量的回答。访谈者必须展现出对于研究的热情，与受访者沟通该项研究的重要性并在访谈过程中认真倾听受访者的回答。

- 阐明困惑与担忧。访谈者必须从受访者的角度考虑问题，消除受访者的顾虑与担心以使其满意。此外，访谈者应该在必要的时候向受访者询问一些试探性的问题，尽管这些问题并不在访谈脚本上。

- 观察回答质量。访谈者处于判定收集的信息质量的最佳位置，他们可以将自身观测到的表情或肢体语言适当地记录在案。

访谈执行。在访谈之前，访谈者需要准备一系列用品以带至访谈现场，包括从研究负责人或赞助方获得的附函，足够数量的调研量表复印件，照片认证以及受访者的电话号码以致电确认访谈者的身份真实性。访谈者还需要提前致电受访者以预约。开始访谈时，访谈者应采用自信与祈使语句，例如"我想占用您几分钟时间来采访您以完成这样一项重要研究"而非"我可以进来做个访问吗？"访谈者需要首先进行自我介绍，提供证明材料并用一两句话阐明研究目的并保证受访者回答的保密性与自愿参与。所有这些陈述都不能超过 1 分钟。不要使用专业术语，也不要提供过多细节（除非受访者想要了解这些）。如果访谈者想要录制访谈过程，其应争求受访者的同意。当然，即使录制访谈过程，访谈者还是需要对关键议题、探讨进行逐字记录。

在访谈过程中，访谈者需要严格遵循问卷脚本进行提问，不要改变措辞以使问题变得更友好。他们也不应改变问题的顺序或者跳过任何问题，即使这些问题在前面已经有所回答。关于问题的任何事由应该在正式访谈前的预演阶段予以解决。访谈者也不应该帮受访者完成句子。如果受访者给出一个非常简单粗略的回答，访问者应该试探性地引导受访者给出更为彻底全面的回答。一些有用的试探性技术包括：

- 沉默试探。仅仅是暂停与等待（不要跳转到下一个问题），这会向受访者暗示访谈者正在等待更多的细节阐释。

- 公开鼓励。偶然地采用"啊哈"或"好的"等来鼓励受访者提供更多细节。不过，访谈者千万不要对于受访者的回答给予赞成或不赞成的表述。

- 寻求详述。例如"您能否再对此进行详细阐述？"或"刚才您提到您在高中的一次经历，您是否能再多谈一些这次经历？"

- 重述。访谈者可以采用精神诊疗师的方法，重复受访者的回答。例如，“我听您说到那次经历非常痛苦”，然后暂停并等待受访者详细阐述。

当访谈结束时，访谈者应向受访者表示感谢并告知其何时可以获得相关结果。此外不应匆忙离开。在离开之后，访谈者应立即记录下访谈过程中的观察以帮助更好地描述受访者的回答。

调查研究中的偏差

尽管具有很多优点与长处，调查研究经常也容易沾染一些系统性偏差，使得从调研中得出的启示失效。这些偏差可以分作五类：无回应偏差、抽样偏差、社会称许性偏差、回忆偏倚及共同方法偏误。

无回应偏差。调查研究经常因其低回收率而为人所诟病。即使经过两到三轮的提醒，邮件调查的回收率水平还是停留在 15-20%左右。如果大部分目标被试都没有对调查给予回复，针对未回应样本的一个合理猜测是“这是不是由于系统原因导致的”，这样的话，问卷与研究结果的有效性将备受质疑。例如，不满意的客户相对于满意的客户而言更倾向于表达自己的看法。因此，样本中不满意客户的比例要高于总体中满意客户的比例。这样的话，研究结果将不具有广义性，所观察到的结果很可能是基于这种偏差性样本的一种假象。为提高调查的回收率，研究人员可采用如下策略：

- 提前告知。在调研开始之前，提前向目标被试发一封短信以邀请他们参与调查以使得他们有所准备，从而提升他们回复的意愿。信件内容应包括研究的目的与重要性，数据收集的方式（比如通过电话或邮件调查等），并对他们的合作表示感激。该技术的一种变化形式是让潜在被试回复一张已提前支付邮资的明信片以表明其是否愿意参与。

- 内容相关性。如果一项调研对于被试而言是紧密相关而意义重大时，被试将非常可能给予回复，而对于那些与其无关的调研则不会回复。

- 对被试友好型的问卷形式。短问卷将比长问卷具有更高的回收率。此外，问题明确、非侵犯性以及容易回答的问卷将具有更高的回收率。

- 获取支持。对于组织调研，获取组织高层对研究的重要性肯定的相关证明对调研很有帮助。这种证明材料可以以一种推荐信的形式出具，在被试看来这样可以提高研究者的可信度。

- 后续请求。多次向未回应者提出后续调研邀请可以促使未回应者填写问卷，尽管这些回应会比较迟。

- 访谈者培训。访谈调研的回收率可以通过对访谈者进行一系列培训而提高，培训内容包括如何提出访谈请求，如何利用计算机化的拨号技术来发掘潜在被试，以及如何对未回应者再次安排访谈。

- 激励。采用现金或礼物卡、钢笔或压力球等赠品、抽奖、优惠券或承诺慈善捐赠等激励措施可提高回收率，至少对于某些特定人群是有效的。

- 非物质激励。相对于物质激励而言，商家更倾向于获得非物质激励。举例而言，研究者可以根据整体调查数据而非个体调查数据为商家提供标杆报告作为非物质激励。

- 保密与隐私。承诺被试隐私数据将不会落入第三方手中，这样也可以提高回收率。

抽样偏差。随机从公共电话本中抽取样本进行电话调研，然后系统性地从样本中剔除那些不在电话本中的电话、手机以及那些无法接听电话的人（如正在工作中）。这样，样本中使用固定电话以及大部分时间待在家里的被试（如失业、残疾或老人）的比例较大。与此类似，在线调研使得样本中经常在线的学生与年轻人比例升高，而将那些不太容易上网或接触计算机的人群（如穷人和老人）从样本中剔除出来。同样，问卷调研将那些不会阅读、理解或者对问卷给予有效回复的人群如儿童与文盲从样本中剔除出来。另外一种抽样偏差是在错误的人群中进行抽样，例如，向老师（或父母）询问其学生（或子女）的学习情况，或者向 CEO 询问公司的运作细节等。这些偏差使得样本并不能代表目标总体从而使得根据偏差性样本得出的结论不具广义性。

社会称许性偏差。很多被试会试图避免他人对其自身、雇主、家庭或友人的负面观点与尴尬性的评价。对于负面性的问题如"你是否觉得你们的项目团队功能失调"，"你的工作场所是否有很多办公室政治的存在"，或"你是否曾经从互联网上非法下载过音乐"，研究者并不一定能得到真实的回答。被试的这种"歪曲事实"从社会称许的角度来描述自己的倾向被称为"社会称许性偏差"。其将破坏所获得的回复的有效性。在实践上并没有方法消除问卷调研中社会称许性偏差，在访谈调研中，一个机敏的访谈者可以根据不一致的回答来进行试探，并利用个人观察来补充受访者的回答。

回忆偏倚。对于调研问题的回答往往依赖于被试的动机、记忆以及回答的能力。特别是对于那些发生在过去很久以前的事情，被试不一定能够回想起他们的动机、行为或者说他们关于这些事情的记忆会随着时间而变化以至于无法获取。例如，如果一个被试被问及其一年前的计算机使用情况或者他们儿时的回忆（如生日）等，他们的回答可能不会那么精确，因为这些事情很难回想起。消除回忆偏倚的一个可行方法是让被试回答其记忆中的具体事件的描述，而不是问他们对于那些事件的看法或动机。

共同方法偏误。共同方法偏误指的是在同一时间点测量的自变量与因变量之间的协方差虚高，例如在横向调研中，采用同样的量表（如问卷）时容易发生这种偏误。在这种情况下，需要研究的现象不能与测量数据的假象分离。有一些标准统计检验可用于检验共同方法偏误，如哈蒙的单因子测试（波德萨科夫等 2003），林德尔和惠特尼（2001）的标记变量法

等。当因变量与自变量是在不同时间点进行测量时，或采用纵向调研设计时，或者这些变量是利用不同的方法进行测量时（如根据计算机记录来测量因变量，而用自评问卷来测量自变量），这种偏误可以在一定程度上消除。

第十章

实验研究

被视作研究设计的“黄金标准”的实验研究是最为严谨的研究设计方法之一。在这种设计中，一个或多个自变量被研究者操控为不同实验处理，被试被随机分配到不同的实验处理中，这样可以观测到实验处理对于产出（因变量）的影响。实验研究的独特优势在于其内部效度或因果联系，因为其可以在控制那些随机变量的虚假影响的同时，通过实验操控来发掘原因与效果之间的联系。

实验研究非常适用于解释性研究而非描述性或探索性研究，此类研究的目的是检验因果关系。该法同样适用于那些包含相对有限且明确界定的自变量的研究，这些自变量可以通过操控或控制的方法来处理。实验研究可以在实验室进行也可以在实际现场进行。在实验室进行的实验室实验具有高的内部效度，但却是以低的外部效度（普遍性）为代价的，因为人为或实验室情景可能不能反映真实世界。在实际现场（如真实组织）进行的现场实验同时具有高的内部效度与外部效度。不过这样的实验为数不多，因为要在现场情景做到既操控实验处理又控制随机变量非常困难。

实验研究可以分为两大类：真实验设计与准实验设计。这两种设计方法都需要实验操控，不同之处在于，真实验设计需要随机分布样本而准实验设计不需要。有时我们也用非实验研究来指代所有不采用实验操控或随机分布的研究设计，如调查研究，观测研究及相关性研究。

基本概念

实验（处理）组与控制（对照）组。在实验研究中，一些被试将接受一个或多个称之为“处理”的实验刺激（实验组），而另外一些被试并不接受这些刺激（控制组）。当实验组中的被试对产出变量的评价优于控制组被试对该变量的评价时，则该实验处理被视作为成功的。实验刺激可能具有多个水平，因此会有多个实验组。例如，为了检验一项用于治疗某种病情

（如痴呆）的新药的作用时，如果由痴呆病人组成的样本被随机分成三个组，第一个组接受高剂量的药，第二组接受低剂量的药，第三组接受安慰剂如糖片（控制组），那么前两组是实验组而第三组是控制组。经过一段时间的处理，如果实验组的被试比控制组的被试的病情获得显著改善，我们可以声称新药是有效的。我们还可以通过比较高剂量与低剂量两个实验组的被试来判断采用高剂量是否比采用低剂量更有效。

实验处理的操控。实验处理是实验研究区别于其他研究方法的独特属性。处理的操控有助于对因果关系中“原因”进行控制。显然，实验研究的有效性取决于是否很好地对实验处理进行了操控。处理的操控必须通过正式实验前的前测或初步试验来验证。所有在处理执行之前的测量都称之为前测试测量，而那些在处理执行后的测量则被称为后测试测量。

随机选取与分派。随机选取是指从总体或者抽样框中随机抽取样本的过程。这种方法也在调查研究中所采用，其可以保证总体中的任何一个个体都有可能被选入样本。随机分派是指随机地将被试分派到实验组或控制组中的过程。这是在真实验研究中的一种标准方法，用以确保在处理执行之前实验组与控制组是相似或等价的。随机选取是与抽样相关的，因而与研究的外部效度或普遍性是相关的。然而，随机分派是与设计相关的，因而与内部效度紧密相关。一个设计良好的实验研究可以同时满足随机选取与随机分派，但准实验研究则即不能做到随机选取也不能做到随机分派。

内部效度的威胁。虽然实验设计因其在内部效度方面的优越表现（即其有能力通过处理的操控来对原因进行控制）而被认为比其他研究方法更为严谨，其并非一点也不存在内部效度的威胁。在下文中，结合一项用于研究数学补习班对提升高中生数学能力的作用的研究，我们将列出内部效度的几种威胁。

- 历史威胁是指观测到的结果（因变量）是由随机变量或者历史事件而非实验处理所引致的概率。例如，学生在参与补习班之后的成绩提升可能是因为其好好准备数学考试而非数学补习班而导致的。

- 成熟威胁是指观测到的结果是由被试的自然成熟度（例如理解复杂概念的智力的一般性提升）而非实验处理而导致的概率。

- 测试威胁是指由于前测-后测设计构成的一种威胁，因为被试的后测结果是建立在前测结果的基础之上的。例如，如果学生记住了前测阶段中的答案，他们很可能再后测阶段重复该答案。不进行前测有助于避免此类威胁。

- 量表威胁指的是在前测-后测设计中，前后测之间分数的差距并非因为数学补习班的作用而是因为在前后测中采用不同的测试（如在后测的考试难度低于前测）而引起的。

- 实验对象流失威胁指的是由于系统性原因（如退出实验的大多数是在前测中分数低的学生），从实验组和控制组中退出的被试的比例不同的。如果低成绩的学生中途退出实验，后测结果将被人为地放大了因为留下来的成绩好的学生居多。

- 回归威胁又称均值回归，指的是一个组在后测阶段在某些测量上的整体表现有向该测量的均值而非向预期方向回归的一种统计倾向。例如，当被试在前测中成绩优异时，其后测成绩有降低倾向（靠近均值）因为他们在前测中的高成绩（偏离均值）可能是一种统计偏差。这种问题在非随机样本以及两次测量并不完全相关的情况下更普遍。

双组实验设计

最简单的真实验设计是包括一个实验组和一个控制组的双组实验设计，该设计非常适于检测单个自变量的影响，该自变量可以操控为一种实验处理。双组实验的两个基本设计是前测-后测控制组设计与只有后测控制组的设计，其变化形式可能包括协方差设计。这些设计往往被表示为一种标准化的设计标注，即用 R 来代表随机分派被试，X 表示将实验处理执行到实验组，O 表示因变量在前测与后测中的观测值（可以用不同的下标来区分实验组与控制组在前测与后测中的观测值）。

前测-后测控制组设计。在这种设计中，被试被随机分派到实验组与控制组，在对关注的因变量在前测阶段进行初始测量之后，实验组被施以实验处理（代表感兴趣的自变量），然后再重新测量因变量（后测）。该设计可用图 10.1 来标注。

R	O_1	X	O_2	(试验组)
R	O_3		O_4	(对照组)

图 10.1 前测-后测控制组设计

在前测-后测设计中，实验处理的效果 E 可以用实验组与控制组的前测与后测分值的差来表示：

$$E = (O2 - O1) - (O4 - O3)$$

此类设计的统计分析包括对比实验组与控制组的简单的 ANOVA 分析。前测-后测设计解决了内部效度的几项威胁，如成熟威胁，测试威胁和回归威胁，因为这些威胁可以被认为对实验组和控制组具有相似（随机）的影响机制。样本选取威胁可以通过随机分派进行控制。尽管如此，该设计仍受到一些其他内部效度威胁。例如，如果在两个组中的样本流失率不同，那么实验对象流失将会是一个问题。此外，前测可能会对后测造成偏差，特别是当前测引入一些不寻常的话题与内容时。

仅有后测控制组的设计。该设计是前测-后测设计的简化版，省略了前测。该设计的标注可以用图 10.2 来表示。

R	*X*	O_1	(试验组)
R		O_2	(对照组)

图 10.2 后测控制组设计

实验处理的效应可以简单地通过比较两个组后测的分值差异来实现。

$$E = (O1 – O2)$$

两组 ANOVA 分析也是针对该设计的一种合适的统计分析方法。该设计的简明性使得它比前测-后测设计更具吸引力。在内部效度方面，该设计控制了成熟，测量，回归，选样，以及前测后测交互等威胁，不过被试流失威胁仍然存在。

协方差设计。有时因变量的测量可能会受到随机变量（即协变量）的影响。协变量是指那些并非实验研究的主要研究兴趣但需要在实验设计时进行控制的变量，这样做可以消除其对因变量的可能影响并且更精确地测量感兴趣的自变量的作用。前面提到的实验设计并没有控制这些协变量。协方差设计（又称伴随变量设计）是一种特殊形式的前测-后测控制组设计，不同之处在于前测阶段对因变量的测量被对协变量的测量所代替。该设计可以用图 10.3 来标注，其中 C 表示协变量。

R	*C*	*X*	O_1	(试验组)
R	*C*		O_2	(对照组)

图 10.3 协方差设计

因为前测测量是对协变量而非因变量的测量，实验处理的作用可以用实验组与控制组在后测中的分值差异来表示。

$$E = (O1 – O2)$$

由于协变量的存在，正确的统计分析方法是采用双组 ANCOVA 分析。这种设计具有仅后测设计的所有优势，但因为控制了协变量而更具内部效度。协方差设计也可以被扩展至前测-后测控制组设计。

因子设计

当一项研究需要对两个或两个以上的自变量（处理）进行操控时，双组实验设计就不适用了。在这种情况下，需要进行 4 组或更多组的设计。下面将要谈到的在实验研究中非常普遍的一种实验设计方法，一般被称作因子设计。在该设计中，每个自变量可以被称为一个因

素，每个因素的不同的等级又被称作水平。因子设计使得研究者不仅能够检测每项实验处理对因变量的独立作用（称为主效应），而且可以检测他们的联合作用（称为交互作用）。

最基本的因子设计是 2 x 2 因子设计，该设计包含两个实验处理，每个实验处理包括两个水平（例如高/低，有/无）。例如，如果你想要比较两种教学方式（课堂教学与在线教学）对学习效果的影响，你还想要检验这些影响是否随着授课时间（每周 1.5 或 3 小时）而变化。在这种情况下，有两个因素：授课方式与授课时间。每个因素具有两个水平（课堂教学与在线教学的授课方式，以及每周 1.5 和 3 小时的授课时间）。该设计可以用图 10.4 表示。如果你想要为授课时间增加第三个水平（如每周 6 小时），那么第二个因素将由三个水平组成，那么需要进行 2 x 3 的因子设计。另外，如果你想要引入第三个因素比如团队工作（有/无），你需要进行 2x2x2 的因子设计。在这种标注中，每个数字表示一个因素，它的值表示每个因素的水平。

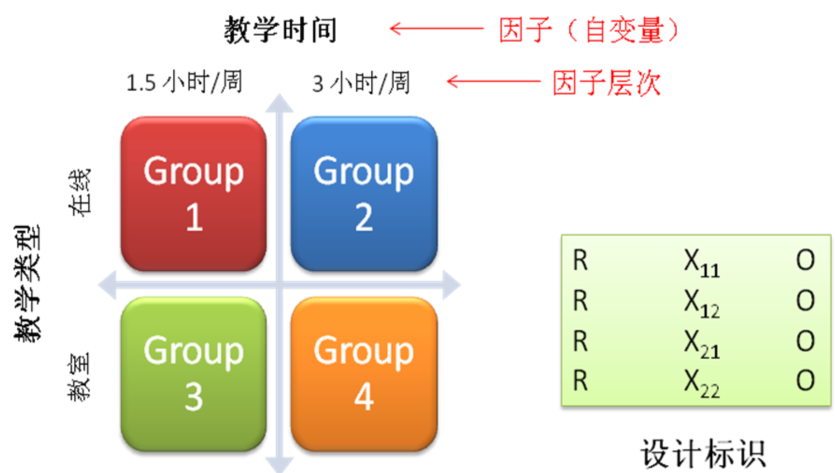


图 10.4 2 x 2 因子设计

因子设计也可以通过设计标注来描绘，如图 10.4 右侧所示。R 代表将被试随机分派到实验组，X 代表实验组（X 的下表代表了每个因子的水平），O 代表因变量的观测值。注意到 2 x 2 的因子设计将形成四个实验组（相对于每个因子的两个水平的组合结果）。相应地，2 x 3 设计将形成 6 个实验组，2 x 2 x 2 设计将形成 8 个实验组。根据经验法则，因子设计的每个单元需要最少 20 个被试（这是参照科恩的效力计算公式，按照中等影响规模得来的估计值）。那么一个 2 x 2 x 2 因子设计需要至少 160 个被试，每个单元至少 20 个被试。你会发现，在因子设计中，数据收集的成本将随着因子水平的增加而显著地增加。出于资源限制的考虑，因子设计中的有些单元并不接受任何实验处理，即非完全因子设计。这种非完全设计削弱了从不完全因子中获得结论的能力。

在因子设计中，当因变量的值在一个因子的不同水平下具有显著差异时，并且该结论在其他因子的任何水平都成立时，说明该因子的主效应存在。主效应的判定是以假设因变量不随因子的水平变化而变化作为零假设或基准线的。在上面的例子中，你可以发现授课方式与授

课时间对学习效果的主效应。当一个因素对因变量带来的变化依赖于第二个因素时，可以认为交互效应存在。在我们的例子中，如果授课方式对学习效果的影响在每周 3 小时的授课时间的情境下高于在每周 1.5 小时的授课时间的情景下的情况时，我们可以说授课类型与授课时间对学习效果的具有交互效应。注意，交互效应的存在可能会使主效应变得并不重要，交互作用显著可能使对主效应的描述失去意义。

复合实验设计

复合设计是通过组合几种成熟的设计方式而实现的。主要有三种复合设计方法：随机区组设计，所罗门四组设计与切换重复设计。

随机区组设计。这是仅有后测控制组设计或前测-后测控制组设计的一个变化形式。被试总体被分作相对同质的子小组（即区组），在这些区组中进行重复使用。例如，你想要在大学生与全职工作者（两个同质区组）中重复同一个仅后测控制组的设计，在两个区组中，被试都是被随机地分作实验组（接受同样的实验处理）或控制组（见图 10.5）。这种设计的目的是减少因为不同区组中的被试的差异带来的“噪声”，以更准确地检测出所关心的变量的实际作用。

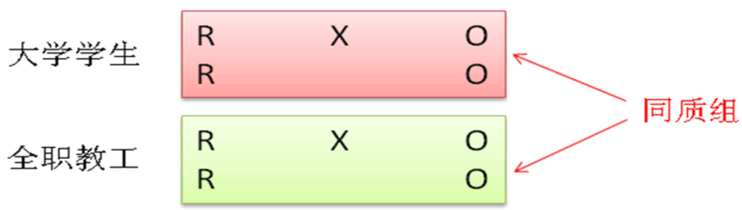


图 10.5 随机区组设计

所罗门四组设计。在这种设计中，样本被分为两个实验组和两个控制组。一个实验组和一个控制组接受前测而另两个不进行前测。这种设计体现了仅后测控制组设计与前测-后测控制组设计的结合，用于检验前测测量对后测测量带来的偏差性影响，这种偏差性影响存在于前测-后测设计而不存在于仅后测设计。设计标注如图 10.6 所示。

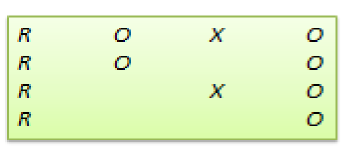


图 10.6 所罗门四组设计

切换重复设计。这是将双组设计应用于两个阶段之中，进行三轮测量。第一阶段的实验组作为第二阶段的控制组而第一阶段的控制组作为第二阶段的实验组，如图 10.7 所示。换言

之，在不同的时间对原来的设计进行重复并切换了实验组与控制组的角色。在研究结束时，所有被试都接受了实验控制（在第一阶段或者第二阶段）。这种设计在组织研究情景下更可行，因为组织项目（如员工培训）是分阶段实施的或者按照某种固定时间间隔重复进行的。

R	O	X	O		O
R	O		O	X	O

图 10.7 切换重复设计

准实验设计

准实验设计几乎与真实验设计完全相同，除了一个方面不同：随机分派。例如，整个班级或组织被当做实验组而另一个班级或同一个行业内的另一个组织被当做控制组。由于缺乏随机分派，这些组并不等价，例如一个组比另一个组在某些内容上更优越（比如说在上学期有更好的老师），这将有可能导致选择偏差。准实验设计在内部效度上不如真实验设计，因为其存在各种各样的选择相关威胁如选择成熟度威胁（实验组与控制组具有不同的成熟水平），选择历史威胁（实验组与控制组受随机因素或历史事件的影响不同），选择回归威胁（实验组与控制组对前测后测均值的回归比例不同），选择测量威胁（实验组与控制组对测量的反应不同），选择测试威胁（实验组与控制组对前测的反应不同），选择流失威胁（实验组与控制组的样本流失率不同）。鉴于这么多的选择威胁，一般尽可能地避免采用准实验设计。

许多真实验设计可以在忽略随机分派时被转化为准实验设计。例如，前测-后测控制组设计的准实验版本被称作不等对照组设计，如图 10.8 所示，随机分派 R 被替换为不等（非随机）分派 N。同样，切换重复设计的准实验版本被称作不等切换重复设计（见图 10.9）。

N	O	X	O
N	O		O

图 10.8 不等对照组设计

N	O	X	O		O
N	O		O	X	O

图 10.9 不等切换重复设计

此外，很少有那种真实验设计没有其对应的准实验版本。下面将讨论其他一些有用的设计。

界断点回归设计。这是一种不等前测-后测设计，根据事前某一测量的界点值将被试分派到实验组与控制组中。例如，重症病人被分到实验组以检测某种新药或处理方式的有效性，那些病情轻微的病人被分派到控制组中。另一个例子，那些落后于标准考试成绩的学生被选入课程补习班以提高他们的成绩，而不将那些考试成绩优异的学生选入补习班。这种设计可以用下图来标注，其中 C 代表界点值。

C	*O*	*X*	*O*
C	*O*		*O*

图 10.10 界断点回归设计

因为采用了界点值，观测结果很可能是由于界点值本身影响的而非实验处理影响的，这将威胁到内部效度。尽管如此，采用界点值也促使将有限的资源分配给那些最需要这些资源的人而非随机进行分配，同时这也可以作为一种准实验的实验处理。在界断点回归设计中，考虑两组之间的系统性的不等价性，控制组的成绩并不作为与实验组进行比较的标杆。如果在控制组中的前测与后测并不存在不连续性，而这种不连续性出现在了实验组，那么这种不连续性可视作实验处理的影响的证据。

代理前测设计。如图 10.11 所示，这种设计与标准的不等对照组设计（前测-后测）非常相像，区别在于前测值是在执行实验处理之后收集的。这种设计的一种典型应用情景是：研究者想要检验某个项目（如一个教育项目）的有效性，然后该项目已经实施使得前测数据无法获取。在这种情况下，研究者的最好选择是利用一个提前记录的一个测量，例如学生在项目开始时的平均积分点作为前测的代理数据。代理前测设计的一种变化形式是在被试进行后测时对前测数据进行重新收集，不过这种方法具有回忆偏倚的风险，但不管如何，这也是提供了一种对因变量的感知收益或变化的测量。

N	O_1	*X*	O_2
N	O_1		O_2

图 10.11 代理前测设计

个别样本的前后测设计。当基于同一种原因从同一个样本收集前测与后测数据不太可能时，这种设计方法非常适用。如图 10.12 所示，该设计有四个组，其中两个组来自同一个不对等组，而另两个组来自不同的不对等组。例如，你要检验客户对于某种在一个城市予以实施但在另一城市没有实施的新的在线服务的满意度，那么在第一个城市的客户可以作为实验组，而另一个城市的客户作为控制组。如果不可能从同一批客户中获得前测与后测的数据，

你可以在某个时间点测量客户满意度，实施新服务项目，然后测量项目实施后的客户满意度（此时的客户与前测时的客户不同）。在同一时间点对控制组的客户满意度进行测量，不过该组并没有实施新项目。这种设计并不非常严谨，因为你不能测量特定客户在新项目实施前后满意度分值的变化，你只能检验客户满意度的平均值。尽管该设计内部效度较低，其仍然是在前后测数据无法从同一被试获取的情况下的一种非常有用的准实验数据收集方法。

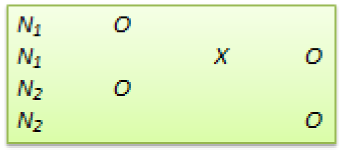


图 10.12 个别样本的前后测设计

非对等因变量设计。这是针对同一组群的前后测准实验设计，具有两个结果测量，其中一个测量从理论上应该受到实验处理的影响而另一个测量并不受此影响。例如，如果你要为高中生设计一个新的微积分课程，该课程将很可能影响微积分成绩而不会影响代数成绩。尽管如此，在后测时代数成绩也可能会因为随机因素（如历史或成熟性因素）而有所变化。因此，前后测的代数成绩可以用作一种控制测量，而前后测的微积分成绩则被视作实验测量。该设计标注可以用图 10.13 来表示，N 表示单组，O1 和 O2 表示同一组学生在前后测中的微积分与代数学成绩。该设计在内部效度上存在不足，但其优势在于不需要一个独立的控制组。

非对等因变量设计的一种变化形式被称作模式匹配的非对等因变量设计。这种设计采用多个结果变量并采用一个理论来解释这些变量将如何受到实验处理的影响，然后研究者可以检测理论预测值是否与实际观测值所匹配。这种基于理论模式与观测模式的对应程度的模式匹配技术可以有效地消除原始的非对等因变量设计的内部效度问题。

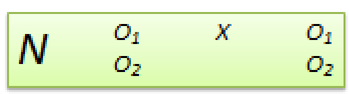


图 10.13 非对等因变量设计

实验研究的危险

实验设计是最难的研究设计之一，因此不能草率为之。这类研究也经常出现大量的方法论问题。第一，尽管实验研究需要理论作为假设检验的基础，现在大量的实验研究不是基于理论的。如果没有理论，假设检验将会是临时构建的，不合逻辑的，并且没有意义的。第二，很多实验研究中用到的测量量表并没有进行信度与效度检验，而且这些量表在不同实验中是不兼容的，因而基于这些量表的结果也将是不兼容的。第三，许多实验研究采用不合适的研

究设计如不相关的因变量，没有交互效应，没有实验控制或者在不同实验组采用不对等的实验刺激等。基于这些研究的发现也是缺乏内部效度而备受质疑的。第四，实验研究中的实验处理（任务）在不同研究中是有差异的，不兼容的和不一致的，有时这些实验处理对被试而言是并不合适的。例如本科生经常被要求扮演营销经理来制定负责的预算，然后其并没有这方面的经验。采用这样的不合适的任务也会给内部效度带来威胁（如被试的绩效很可能是由任务情景的内容或难度形成的人为产物），这样得来的研究发现是难以表述和没有意义的，将这些发现与其他研究进行整合就更不可能了。

合适的实验控制设计是实验设计的最重要的任务，因为实验控制是实验方法的根本所在，决不能草率或忽视。为设计好一个令人满意的合适的任务，研究者需要事先验证任务的有效性，进行实验处理的操控检测以证实任务的充分性（在被试完成任务后简要向其询问情况），进行预测试（需要时可重复进行），如有疑问最好采用对被试来说简单熟悉而非复杂的不熟悉的任务。

总之，本章介绍了实验研究的主要概念以及各种各样的真实验与准实验的设计方法。尽管这些设计在内部效度上千差万别，内部效度不佳的设计也不能轻视，因为根据某些特殊情境和研究需要，他们可能会非常适用。

第十一章

案例研究

案例研究是一种在自然情景下的一个或几个站点对某个现象进行长期的集中研究的方法。多种数据收集方法如访谈，观测，预录文件以及二手数据可以用于数据收集，针对所关注的现象所得出的结论也将是丰富、详尽和情景化的。案例研究可以采用实证的方式以对理论进行检验，也可以采用描述的方式来进行理论构建。相对于其他社会科学领域，这种方法在商业研究中更为普遍。

案例研究相对于其他研究方法如实验和调查研究具有一些独特的优势。第一，案例研究既可用作理论构建也可用作理论检验，而实证研究方法只能用于理论检验。在描述性案例研究中，感兴趣的变量不必提前预知，而是可以随研究的开展从数据中浮现出来。第二，在研究过程中，当发现原来的问题设置并不相关时，研究者可以在研究过程中对问题进行修改。这是在其他任何一种实证研究方法中所不允许的。第三，相比与其他研究方法，案例研究有利于对感兴趣的现象提出更丰富、更情境化与更真实可信的描述，因为其可以对一系列情境数据进行把握。第四，该方法可以从多个参与者的角度，采用多层面（如个体层面与组织层面）的分析方法对感兴趣的现象进行研究。

同时，案例研究也有其本质缺陷。首先，因为其并不进行实验控制，研究结论的内部效度较弱。当然，这是除了实验研究外其他所有研究方法的通病。尽管如此，正如下文将讨论的，案例研究也可以通过“自然控制”来解决这一问题。其次，从案例研究中生成的结论的质量很大程度上依赖于研究者的整合能力。一个经验丰富的研究者能发掘出案例数据中的概念与模式，而经验不足的研究者则可能错过这些发现。因此，基于案例研究的发现经常被批判为过于主观。最后，因为研究结论是非常情境化的，将研究结论推广至其他研究情境或组织时则比较困难。

值得指出的是，案例研究与案例描述如在商业课上讨论的哈佛案例不是一个概念。案例描述是对一个组织的问题进行详细的描述以引发课题讨论，启发学生判断思考或者分析一个组织如何处理具体问题，而案例研究是一种针对组织现象提出解释的正式研究技术。

案例研究是一种较难的研究方法，其要求研究者具有高超的研究技能，因此这类研究也经常容易出错。本巴萨等人（1987）[8]阐述了案例研究中经常遇到的五个问题。第一，很多案例研究在研究之初并没有具体的研究问题，从而在研究结束之后也不能提供针对性的答案或者有洞察力的结论。第二，案例的站点选择经常以易接触性和方便性为标准而不是以与研究问题的匹配性为标准的，因而不能充分地解决所感兴趣的研究问题。第三，研究者往往不对数据收集方法进行确认或者采用多种方法来进行数据收集，这样可能会因为受访者的偏差而导致偏差性的描述。第四，很多研究仅仅提供很少的数据收集细节（如采用了哪些访谈问题，分析了哪些文件，以及受访者在组织中的职位等）或分析细节，这使得结论的可信度饱受质疑。最后，尽管该方法具有纵向研究的优势，很多案例研究却并没有按照纵向研究的方式对现象进行研究，而是仅仅呈现出横断面的分析，对于那些与时序相关的组织过程或现象认识不足。

案例研究中的关键决定

当一个研究者考虑案例研究方法时，其需要考虑如下几项重要决定。第一，这种方法是否是解决研究问题的正确的方法？案例研究方法适用于那些用于在理论形成阶段发现相关变量的研究，适用于那些参与者经验与行为的情景至关重要的研究，适用于那些旨在理解复杂的时序过程（如为什么现象会发生或现象如何发生）而非因素或原因（关于什么）的研究。这种方法对于涉及多个参与者与多个连续事件的复杂的组织过程（如组织变革或大规模的技术实施项目）的研究更为适用。

第二，案例研究的合适的分析单元是什么？因为案例研究可以同时考查多个分析单元，所以研究者必须决定其研究对象是个体层面、群组层面、组织层面还是多层面的现象。例如，一个关于群决策或群协作的研究可能会整合个体层面的变量（如个体参与组活动的行为）与组层面的变量（如小组凝聚性和领导风格）从而获得比单层面分析更丰富的理解。

第三，研究者应该采用单个案例还是多个案例的设计？单个案例设计更适合于理论创造型的研究，特别是当此个案是独特或极端情况时，当其具有启示性时（即该情况在以前的科学研究中不太容易出现时），或当其是一种批判性的或与成熟理论相违背的案例时。多案例设计对于理论检验更为适用，因为其可以增加结论的广义性，并且可以对现象进行更为丰富与详尽的表述。殷（1984）[9]依据重复性的逻辑建议采用多案例站点，将每个案例站点视作一个实验研究，遵循实证研究中的科学严谨性原则。

第四，应该选择什么案例研究站点？由于案例研究结论的情境化的本质，站点选择非常重要，因为选择错误的站点将导致错误的结论。如果研究目的是对理论进行检验或者检验结论的广义性，那么应该选择不相像的站点以增加观测的方差。例如，如果研究目的是理解企业

[8]Benbasat, I., Goldstein, D. K., and Mead, M. (1987). "The Case Research Strategy in Studies of Information Systems," *MIS Quarterly* (11:3), 369-386.
[9] Yin, R. K. (2002), *Case Study Research: Design and Methods*. Thousand Oaks, CA: Sage Publications.

的技术实施过程，应选择混合大型、中型与小型企业的站点以检验技术实施过程是否因企业规模而不同。站点选择不应抱乐观态度或基于便利性原则，而应该与研究问题所匹配，这个过程称之为“理论化抽样”。

第五，案例研究应该采用什么数据收集技术？尽管访谈（开放化/非结构化或封闭式/结构化）是至今案例研究最为流行的数据收集技术，访谈数据可以与其他技术相结合，例如直接观察（如参与执行层会，传达指示会，和规划会）、文件整理（如内部报告，演讲，备忘录，以及外部新闻报道等）、档案记录（如组织图表，财务记录等）和物理用品（如设备，输出，工具等）。此外，研究者可以通过比较来自不同受访者的回答来验证观测数据。

实施案例研究

大部分案例研究在本质上是阐释性的。阐释性的案例研究是一种归纳性的技术，其通过对从一个或多个案例站点收集的证据进行系统性的分析与整合来发现相关概念与模式以构建新理论或扩展已有理论。艾森哈特（1989）[10]提出了从案例研究中构建理论的“路标”，我们将其修订颁布阐述如下。对于那些实证性的案例研究，需要对下述的阶段进行重新安排与修改，但其抽样、数据收集和数据分析技术是基本一致的。

定义研究问题。像其他科学研究一样，案例研究必须从定义那些在理论与实践中有趣的研究问题开始，并且需要对这些问题的可能答案或者初步的变量进行直觉预测以指导初步的案例研究。在实证性案例研究中，初步的变量是基于理论的，然后在阐述性研究中这些理论或假设并非提前存在的。这些研究问题或变量可能随着阐述性研究的开展而发生改变，但这在实证研究中是不会发生的。

选择案例站点。研究者应该采用“理论化抽样”（而非随机抽样）的过程来确定案例站点。按这种方法，案例站点的选择应给予理论性而非统计性考虑，例如重复先前的案例，扩展初步理论或填充理论类别。应注意确保站点选择与研究问题的本质相匹配，减少由企业规模、行业等造成的随机方差或噪声并最大化所感兴趣的因变量的方差。例如，如果研究目的是检验某些企业如何比其他企业创新性强，研究者应该选取在同一行业内规模相似的企业以减少行业或规模的影响，并选择一些创新性强的企业和一些创新性弱的企业以增加企业创新性的方差。不要不经预约就跟潜在的站点致电或写信，而应与公司管理执行层里具有批准项目实施的权力的人或者能够确认这些人的人进行联系。在初始交谈中，研究者应该对项目的本质与目的、为案例站点带来的可能收益、数据如何收集与利用、数据收集涉及到的人（其他研究者、研究助理等）、想要访问的人、赞助公司所需付出的时间、精力和花费一一进行介绍。研究者还应对公司以及个体受访者的保密性、隐私和匿名性进行承诺。

[10]Eisenhardt, K. M. (1989). “Building Theories from Case Research,” *Academy of Management Review* (14:4), 532-550.

生成量表和草案。因为案例研究的主要数据收集方式是访谈，所以需要设计一份访谈草案以指导访谈过程。一般包括一系列将要提问的问题。这些问题可以是开放式的（非结构化的）、封闭式的（结构化的）或者二者的结合。在访谈过程中，访谈者必须严格遵循访谈草案，不能擅自改变问题顺序或调过任何问题，尽管访谈者可以稍微偏离草案以试探性地对受访者表意模糊或者其他感兴趣的问题进行扩展。访谈者必须保持中立的观点，不要对受访者的回答做出同意或不同意的表态以将受访者引致某个特定方向。关于访谈的其他技术细节已经在关于调查的章节中进行了讨论。此外，其他数据源诸如内部文件和备忘录、年报、财务声明、新闻文章和直接观测也应作为访谈数据的补充材料。

选择受访者。从组织的不同层级、部门和职位对受访者进行选择以获得针对感兴趣的现象的不同观点。最好对受访者进行随机抽样，不过滚雪球式的抽样也是可以接受的，只要样本能体现出多样性的观点。受访者必须根据其对与研究现象的相关水平以及他们认真充分地回答研究者问题的能力与意愿进行选择，而不要仅仅根据便宜性原则。

开始数据收集。最好将访谈进行电子化录制以备后续查询。当然，这种录制必须征得受访者的允许。即使对访谈进行录音，访谈者还是需要将关键性的观察、受访者的行为反应（如肢体语言）以及研究者对于受访者及其回答的个人印象记录在案。每个访谈结束之后，整个访谈需要被整理成一个文本文件以进行分析。

个案内数据分析。数据分析应依照数据收集或与其有交叉。使数据收集与分析存在重叠的好处在于，可以根据数据分析中出现的主题来调整数据的收集过程以进一步探析这些主题。数据分析可分两阶段进行。在第一阶段（个案内分析），研究者需要检验每个独立个案中的概念以及这些概念间的模式，从而生成关于研究问题的初步理论。研究者可以结合自身的个人观测与体会主观地对访谈数据进行分析以使研究问题“讲得通”。或者，采用一种编码策略如格拉瑟和施特劳斯（1967）的扎根理论方法等，利用开放编码、主轴编码、选择编码等技术也可以挖掘出一系列的证据与结论。此外也可以采用一些自主开发的技术如图形化数据表示（如网络图）或顺序分析（适用于纵向数据）。因为对各种各样的案例数据并不存在某种特定方法，所以可以根据研究项目的特性来对数据分析技术进行适当修改。

跨案例分析。多站点案例研究需要进行跨案例分析作为第二阶段的数据分析。在这种分析中，研究者需要在不同案例中寻找相似的概念或模式而忽略那些因为情境差异而导致的不一致结论。这些模式可用于验证初始理论或者进行理论修订（通过添加或去除一些概念或关系），从而开发出一个更具包容性的可推广的理论。这种分析有几种方式。例如，研究者可用选择某些类别（如企业规模、行业等）来寻求组内相似性与组间差异性（如高或低的绩效，创新者或落后者）。或者，研究者可以采用成对比较的方式来比较企业间的相同点与不同点。

建立与检验假设。可以基于那些在不同案例站点间具有广义性的概念或主题来构建尝试性的假设。这些假设应不断地与观测到的现象进行比较以判断其是否与观测数据相匹配，如果不匹配，则因对变量或关系进行重新修订。研究者还应将出现的变量与假设与先前文献中的

报告做比较，以保证其内部效度与广义性。不要去拒绝冲突性的发现，而应通过创造性的思考以将更深刻的见解融入到理论中。当理论与数据间的迭代不再形成新的见解或变化时，那么就达到了“理论成熟”，理论构建过程也告完成。

撰写案例研究报告。在撰写报告时，研究者应对抽样、数据收集、数据分析以及假设开发等细节过程进行清晰的描述，这样读者可以独立地对研究的合理性、优势以及所报告的结论的一致性进行评估。研究者还需要对研究模型进行非常清晰的阐述以保证这些发现并不是因为研究者的先入为主造成的。

阐释性案例研究示例

学习阐释性案例研究的最好方式是研读示例。这里的一个例子是艾森哈特（1989）[11]用于研究经理人在动荡性环境下的决策行为的研究。建议读者在看这部分介绍前先阅读其刊登在 Academy of Management Journal 上的原文。在这个研究中，艾森哈特探析了一些动荡性环境下，一些企业的经理团队可以做出快速决策，而另外一些则不能做到，并进一步分析了快速决策在这种环境中将提升还是降低企业绩效。动荡性环境指的是需求、竞争与技术迅速而不连续地变化从而使得获有的信息经常是不精确、不可获得或者过时的。该研究的潜在的前提假设是：（1）如果没有足够信息很难做出快速决策；（2）快速决策未必更有效，其可能反而导致差的企业绩效。

在对先前关于经理人决策文献进行阅读后，艾森哈特发现了几种模式，尽管这些模式并不是针对动荡性决策环境的。文献指出作为权宜之计，那些做出更快决策的企业从少数信息源获取信息，考虑较少的备选项，做出有限的分析，限制用户参与决策，采用中心化的决策模式以及具有有限的内部冲突。尽管如此，艾森哈特认为这些观点并不足以解释决策者是如何在动荡性环境中进行决策的，因为不仅需要利用不完全信息进行快速决策，而且需要保持高的决策质量。

为研究这一现象，艾森哈特实施了一项基于个人电脑行业的八家企业的归纳性研究。随着 UNIX 操作系统，RISC 架构，以及上世纪 80 年代 64K 随机存取存储器等技术的引入，个人电脑行业正发生着巨变，随着 IBM 进入个人电脑领域，市场竞争加剧，顾客需求呈两位数增长，这与动荡性发展环境是相匹配的。该研究采用基于重复性逻辑的多案例设计方法，期望每个个案与其他个案所形成的结论一致或不一致。案例站点的选择是基于研究者的就近与可触及的原则的，所有这些企业都在加州硅谷区域经营市场动荡的个人电脑行业。这些企业都分布在同一区域的同一行业，这就排除了行业类型或地理位置差异对因变量（决策速度或绩效）的“噪声”影响。

[11]Eisenhardt, K. M. (1989). “Making Fast Strategic Decisions in High-Velocity Environments,” *Academy of Management Journal* (32:3), 543-576.

该研究采用了多层分析的设计方法：决策（比较每个企业内的多个战略决策）、经理团队（比较参与战略决策的不同团队）、企业（整体企业绩效）。数据从五个数据源进行收集：

- 与首席执行官（CEO）进行初始访谈。向 CEO 们询问其公司的竞争战略、独特能力、主要竞争者、绩效以及最近或正在进行中的主要战略决策。基于访谈可以从每个企业中选取一些战略决策以备后续研究。在选择这些决策时应参考四个标准：（1）决策应包括企业的战略定位；（2）决策意义重大；（3）决策包括多项功能；（4）决策应代表公司的战略决策过程。

- 与部门领导进行访谈。向部门领导询问十六个开放性问题，包括企业的竞争战略、功能战略、高管层成员、团队讨论的频率与属性、典型决策过程、先前提及的那些决策是如何制定的以及这些决策耗费了多长时间。访谈持续了 1.5 到 2 个小时，甚至有时延长至 4 个小时。为了关注于事实或真实事件而不是受访者的主观看法与表述，访问采用“审判式”的提问方式，例如何时发生以及你做过什么等。两个人负责进行访谈，访谈数据需经访谈者和记录者对事实与印象进行复核。尽管对所有访谈进行了录音，对每次访谈还是进行了笔录，并且记有访谈者的整体印象。根据“24 小时制”，所有现场详细记录需要在访谈结束 24 小时内完成，这样相关数据或印象才不会被遗忘。

- 问卷。每个企业的经理团队成员还被要求完成一份调研问卷以对企业的冲突与权力分布进行把握。

- 二手数据。收集行业报告与内部文件如参与战略决策的经理团队的统计学特征，企业财务绩效等等。

- 个人观测。最后，研究者还参加了两家企业的为期一天的战略规划会议以及经理层周会。

数据分析包括定性分析与定量分析。关于冲突与权力的定量数据用以分析不同企业或决策的模式。定性访谈数据用于把握决策文化剖面，这可以通过多于一个经理人所提出的剖面特征（如不耐心）来刻画。对于个案分析，需要将经理人的关键决策事件按时间轴梳理成每个战略决策的决策故事。对于跨个案分析，可以成对比较企业的相似点与不同点，从而按照感兴趣的变量如决策速度和企业绩效进行归类。在这些分析的基础之上，可以从每个决策故事中归纳出一些常识性的变量或假设，然后对决策个案进行回访以确认所假设的关系。推理出来的假设可以与已有文献进行比较以解决所发现的差异进而从案例发现中形成新的理解。最终，将经过证实的假设整理为一个关于在动荡性环境下的企业战略决策的归纳性理论。

从多案例研究中所生成的推论与已有文献所预期的一些决策模式是冲突的。第一，在动荡性环境中，快速的决策者会采用更多的信息而非如先前文献说暗示的较少信息。不过，这些决策者利用更实时的信息（先前研究中并未涉及的观点）以帮助他们发现与回应问题、机遇以及快速变化的情境。第二，快速决策者考察更多（而非更少）的备选方案。不过，他们可以用并行的方式考虑这些备选方案，而那些缓慢决策者则以串行方式考虑较少的备选方案。

第三，与文献所启示的不同，快速决策者并非采用中心化的决策模式或限制他人的参与。相反，这些企业采用双层决策过程，即先让一些经验丰富的顾问完成第一阶段的决策，然后在第二阶段对这些决策进行快速的比较与筛选。第四，快速决策者并不像文献所预测的那样具有很少的冲突，而是具有更好的冲突处理技巧以减少冲突和提供决策速度。最后，快速决策者显现出更好的企业绩效，得力于其内在的认知、情感及政治能力使其可以迅速地接近主要决策。

实证案例研究示例

案例研究也可以用作一种实证手段来检验理论与假设。这样的研究很少，而马库斯（1982）[12]关于金三角公司（GTC）（假名）的技术实施的研究是一个范例。该研究的目的是理解为什么一个旨在提升 GTC 公司的会计师们的效率与绩效的金融服务系统（FIS）会受到 GTC 总部会计师们的支持却遭到 GTC 分公司的会计师们的抵制。由于研究现象的特殊性，该研究采用个案研究的方法。

为了探索 FIS 的用户抵制行为的原因，马库斯提出了三种可能解释：（1）系统决定性理论，即用户抵制是因为系统缺陷的原因造成的，如技术机能不足，人体工程学设计差，或者用户友好性欠缺；（2）用户决定性理论，即用户抵制是由用户的内在因素，如会计师的认知方式或个性与使用该系统不兼容；（3）交互式理论，即用户抵制并不仅仅是因为系统或用户的内在因素所决定的，而是因为这两类元素之间的交互而造成的。具体而言，交互式理论提出 FIS 造成了组织内部权力的重新分配，那些失去组织地位、相关性或权力的会计师们会抵制 FIS 而那些得到这些的会计师们则偏好 FIS。

为了检验这三种理论，马库斯根据每个理论解释预测了可能的结果，然后分析了这些预测在多大程度上与其在 GTC 的观测相一致。例如，系统决定性理论暗示由于用户抵制是有系统不足所引起的，那么处理好系统的技术问题将消除这种抵制。随后对运行 FIS 系统的计算机进行了一系列升级，如采用更强大的操作系统，在线处理（本来采用的是批处理，这样会延迟会计信息的及时处理）以及一个用于创立新账号的简化软件。一年之后，抵制系统的用户仍然在抵制系统，觉得应该将其替换掉。因而，系统决定性理论被拒绝了。

用户决定性理论预言在去除用户抵制因素或将抵制用户与非抵制用户指派在一起时可以减弱他们对 FIS 系统的抵制。随后，GTC 采用了一种工作轮换与流动性政策，将会计师们移进或移出抵制性的部门，但是抵制不仅仍然存在，反而变本加厉。一个具体的例子是，一个参与系统设计的会计师，其在公司财会部门工作时对系统表示支持，而将其移至部门控制办公室时却开始抵制系统。由于并未发现用户决定性理论所预言的结果，所以这个理论也被拒绝了。

---

[12] Markus, M. L. (1983). "Power, Politics, and MIS Implementation," *Communications of the ACM* (26:6), 430-444.

最后，交互式理论预言无论是改变系统或者人员（如用户教育或轮岗政策），只要技术实施没有造成权力不均衡和再分配都将削弱抵制行为。在 FIS 实施之前，GTC 每个部门的会计师们觉得他们拥有与其部门的运作相关的会计数据。他们将这些数据保留在厚的手工账本里，使得他人无法接触到这些数据，从而可以在发布那些报告之前解决一些非常规会计事件。公司层级的会计师需要依靠分部门的会计师才能接触到这些分部数据，进行公司报告与合并。由于 FIS 系统可以自动地从各个源泉进行数据收集并将它们合并为单个公司数据库，这削弱了部门会计师的存在必要，使它们失去了对其部门的财会数据的控制权与自主性，使得它们的工作变得无足轻重。公司层级的会计师可以直接从数据库查询和获取部门数据而不需要经过部门会计师，分析和比较每个部门的绩效，向经理委员会汇报不正常的情况与行为，从而使得部门的权力受到侵蚀。尽管马库斯并没有实证检验这个理论，她对组织权力的重新分配的观测以及对另两个备选理论的拒绝使得交互式理论得以证实。

与传统研究进行比较

用于检验假设的实证案例研究经常被自然科学研究者批判说缺乏控制性的观测，控制性的演绎，可复制性以及发现的广义性，而这些都是实证研究的传统准则。尽管如此，这些批判可以通过合适的案例研究设计来克服。例如，关于控制性的观测的问题指的是在案例研究中难以获得实验或者统计控制。然而，案例研究者可以采用“自然控制”的方式来弥补这种控制上的不足。在马库斯的研究中，自然控制指的是作为系统最初的拥护者，公司层的会计师在被移至其所控制的部门时便开始抵制这个系统了。在这种情况下，其行为变化可以归咎为其新的部门职位。不过这种自然控制并不能提前预知，除非他们非常积极地去寻求这种控制否则他们会忽略这种控制。顺便提一下，自然控制也应用于自然科学领域，如天文学、地质学、人类生物学，例如等待彗星足够接近地球时再对彗星的组成成分进行推断。

关于控制性的演绎的问题指的是由于大部分案例研究数据是定性的，所以缺乏充足的定量证据以支持推论。尽管缺乏定量数据以进行假设检验（如 t 检验），控制性的演绎仍然可以通过根据理论考虑进行行为预测进而长期地去检验这些预测来实现。马库斯在其研究中采用这种策略形成了三种关于用户抵制的可能性理论假说，当前两个理论的预测与实际观测不匹配时，拒绝了这两个理论。在这种情况下，假设是通过逻辑观点而非数学检验而检验的，其与统计结论同样有效，因为数学也是逻辑的一个子集。

第三，复制性的问题指的是由于案例站点的特殊性与异质性而难以观测到同样的现象。不过，以马库斯的三个理论为例，不同的研究者可以在不同的案例站点检验同一个理论，那么基于三种异质的新站点会形成三个不同预测，然后可以对这三个预测进行相应的检验。换言之，案例研究的结论是可能复制的，尽管案例研究站点或情境并不能复制。

第四，案例研究倾向于研究独特性的、不可复制的现象，使得其不能推广到其他情境。自然科学中的广义性是通过累加的研究建立的。相似地，在不同的情境中进行的累加性案例会形成不同的预测，如果这些发现在不同的研究中是一致的，则其广义性便证实了。

最后，英国的哲学家卡尔·波普尔刻画了科学理论的四个条件：（1）理论应该是可证伪的；（2）理论应该是逻辑一致的；（3）理论应具有充分的预测能力；（4）理论应比其他竞争性理论提供更好的解释。在案例研究中，前三个条件可以通过增加所观测到的研究发现的自由度而实现，例如增加案例站点数量、增加其他的备选预测以及增加分析层面的数量。这些已经在马库斯的研究中通过检验多组行为（部门会计师与公司会计师）和提供多个（三个）竞争性解释得以实现。在该研究中，波普尔的第四个条件也在发现一个假设比另外两个竞争性假设与观测事实更加匹配时得以实现。

第十二章

诠释研究

最后一章介绍诠释性研究，或者更具体地来讲，诠释性案例研究。该章将探索其他几种诠释性研究。那些实证性或演绎性的方法如实验室实验或者调研法都是具体地用于理论（或假设）检验的，而诠释性或归纳性的方法例如行动研究法和人种志研究法则用于理论构建。与那些从理论出发并采用实证数据进行理论检验的实证研究方法不同，在诠释性研究中，研究者从数据出发并尝试从观测到的数据中抽取出与感兴趣的现象相关的理论。

“诠释性研究”一词经常被不严谨地用作“定性研究”的同义语，尽管这两个概念非常不同。作为一种研究范式（见第三章），诠释性研究基本前提假设是：社会现实并不是单一或客观的，而是由人的经验与社会情境所塑造的（本体论），因而相关研究应该在其所属的社会历史情境中进行以调和不同参与者的主观表述（认识论）。因为诠释性研究者将社会现实看作嵌入于社会情境中并不太可能从社会情境中所抽象出来的，他们通过“意会”的过程对其进行“诠释”而不是进行假设检验。这是与那些实证主义者或机能主义者不同，那些人假定社会现实是可以独立于情境而存在的并可以从这些情境中进行抽象的，因而可以通过客观技术如标准化测量进行分解式地研究的。一个研究者是应该追求诠释性还是实证性的研究取决于其针对现象本质以及解决现象的最优方式的范式方面的考虑。

然而，定性与定量研究指的是关于数据收集类型与分析方法方面的实证性或数据导向性的考虑。定性研究主要基于非数字数据，如访谈与观测，而定量研究则采用数字型数据如分值与度量。因此，定性研究不适用于统计过程如回归分析等，而是采用类似内容分析的技术进行编码的。有时，编码定性数据可以用表定量地显示编码的出现频率，但这些数据并不能进行统计分析。很多纯诠释性研究者反对采用编码的方法，认为这种对本质上具有主观性的社会现象寻求一致性或客观化的努力是没有意义的。

虽然诠释性研究很大程度上依赖于定性数据，定量数据可以比定性数据提供针对现象的更为精确与清晰的理解。例如，艾森哈特（1989）在其关于动荡环境中的决策问题的诠释性研究中（在上一章案例研究中进行过讨论），收集了一些数字型数据，如每个企业做出某项战略决策所花费的时间（介于 1.5 个月到 18 个月之间），针对每项决策有多少个备选方案，以

及受访者对于组织冲突的看法等。这样的数字型数据有助于清晰地区分快速决策企业与慢速决策企业而不需要依赖于受访者的主观判断，也有助于对确定快速与慢速决策企业所考虑的备选方案数量以及冲突的程度。诠释性研究应该尽量根据研究现象对定性与定量数据进行收集，实证性研究也是一样。联合运用定性与定量数据（经常被称之为"混合模式设计"）可能会引发独特的理解，应该在学术界予以赞赏。

诠释性研究源于人类学、社会学、心理学、语言学与符号学，在 19 世纪早期就已经出现，远早于实证研究方法。由于其定性数据收集与诠释过程的主观性的本质，很多实证研究者认为诠释性研究是有误的和有偏的。尽管如此，由于很多实证研究方法并不能产生有趣的观点与新的知识，导致诠释研究从上世纪 70 年代开始复苏，尽管对诠释性研究结论的信度与效度提出了很多苛刻的标准。

与实证研究的区别

除了前面讨论过的在本体论与认识论的前提假设上的根本性的范式差异，诠释性研究与实证性研究还存在其他一些差别。第一，诠释性研究采用"理论性抽样"的策略，研究的站点、被试或案例都是基于理论考虑而进行选择的，如其是否与研究现象相匹配（如可持续实践只能在那些实施了可持续实践的组织进行研究），他们是否具有与研究相符合的某些特质（如一项研究企业创新驱动力的研究应该包括一些高创新性的企业和一些低创新性的企业，从而可以对这些企业进行比较），案例是否从总体中进行随机选择（为了研究广义性的目的）。因此，只要与研究的本质相符合，方便性样本和小样本对诠释性研究而言是可行的，这在实证研究中是行不通的。

第二，研究者在诠释性研究中的作用被予以高度关注。在某些方法中，如人种志法、行动研究、和参与者观测法等，研究者被当作社会现象的一部分，其在研究过程中的具体角色与参与需要在数据分析时予以清晰描述。在另外一些方法中，如案例研究中，研究则必须在数据收集与分析过程中保持中立或者不偏不倚的姿态，以确保其个人的偏差与先入之见不要玷污了诠释性研究的主观性结论的本质。在实证研究中，研究者被认为是脱离与研究情境而独立存在的，其不应使得数据收集与分析过程有偏差。

第三，诠释性分析是整体性的和情境性的而非简化或孤立的。解释性的描述倾向于从社会现象的参与者的角度关注语言、符号与意义，而不像在实证研究中那样大量使用统计分析技术。诠释性研究的严谨性可以从数据收集与分析的系统性与透明性方法来判别，而不像实证研究中那样依据变量效度或显著性检验的统计标准。

最后，在诠释性研究中，数据收集与分析可以同时和重复进行。例如，一个研究者可能进行了一次访谈然后在进行下一次访谈之前对其进行编码。同时性分析可以帮助研究者修正访谈草案中的潜在缺陷或对其进行调整以更好地把握研究现象。研究者甚至可以在其意识到原

来的问题不能产生新的或有用的观点时对原来的问题进行修改。这是诠释性研究的一种非常有价值但经常被低估的好处，而在实证研究中，一旦数据收集开始则不能再进行任何更改。

诠释性研究的好处与挑战

诠释性研究具有一些独特的优势。第一，其非常适于探索那些复杂、相关或多面的社会过程背后隐藏的原因，如企业间的关系或办公室政治，这些通过定量方法获得的证据将是有偏的、不精确的或者很难获取的。第二，在那些先前理论不存在或不充分的研究领域，这种方法有助于理论构建。第三，这种方法非常适于研究情境具体化的、独特的、异质的事件或过程。第四，诠释性研究有助于发现有趣而相关的研究问题或议题，以备后续研究。

同时，诠释性研究也面临着自身的一系列挑战。第一，这类研究需要比实证研究在数据收集与分析过程花费更多的时间与资源。太少的数据可能会导致虚假或不成熟的猜想，而太多的数据对研究者来说又不能有效地进行操作。第二，诠释性研究需要受过良好训练的研究者，其应该能够从参与者的角度来审视和描述复杂的社会现象并调和这些参与者的各种观点，而不应将自己的偏见或成见灌输于结论中。第三，并不是所有的参与者或数据源都是同样可信的、无偏差的或者对研究现象了解深刻的，其可能是具有未公开的政治目的的，这些都将导致误导性的或虚假的印象。参与者与研究者之间的信任不足可能会阻碍参与者给予诚实全面的自我表述，而这样的信任建立则需要时间。这就要求诠释性研究者能够“看透迷雾”（隐藏或偏差性的目的），理解问题的真正本质。第四，由于诠释性研究所得出的结论在本质上是非常情境化的，这使得这些结论不具可复制性或广义性。最后，诠释性研究可能有时并不能解决感兴趣的研究问题或预测未来行为。

诠释性研究的特征

所有诠释性研究必须遵循如下所述的一系列共同的准则。

自然性的调查。社会现象必须在其所属的自然情景下进行研究。因为诠释性研究假定社会现象是嵌入于社会情境中并不能从这个情景中孤立出来的，对于这类现象的描述必须建立于其社会历史情境之上。这表示应该对情境性变量进行观测并在对现象进行解释时予以考虑，尽管情境敏感性可能会限制结论的广义性。

研究者作为工具。研究者经常是嵌入于其研究的社会情境之中的，其应该被看做是数据收集工具的一部分，所以他们必须利用他们的观察技巧，他们与参与者的信任以及他们提取正确信息的能力。此外，研究者对社会情境的个人洞察力、知识与经验对于精确地描述研究现象至关重要。同时，研究者也需要充分意识到自身的偏见与成见，使得这些偏差不要影响其公正精确地描述现象的能力。

诠释性分析。需要按置于社会情境中的参与者的眼光来对观测进行诠释。诠释可分为两个层次。第一个层次涉及到从社会参与者的主观角度来看待与感受现象。第二个层次是对参与者的感受的意义进行理解，从而对研究现象进行“深度阐述”或构建丰富的叙述故事，以理解参与者为什么那样做。

采用有表现力的语言。对参与者的文字和非文字的语言进行记录以及对这些语言进行分析是诠释性分析的组成部分。研究必须确保所提出的故事是从人的视角而非机器的视角来观测的，必须描述出那个人的表情与感受，这样读者才能够理解这个人并把这个人跟现象联系起来。采用想象、比喻、讽刺或者其他修辞手法在诠释性分析中是很常见的。

时序性本质。诠释性研究往往并不是为了寻求某些具体答案，而是为了随着社会过程的展开对动态的社会过程进行理解或“意会”。因此，这类研究需要研究者切身处地地在研究站点待一段时间以把握研究现象的整体演化过程。

解释性循环。诠释性表述是一个在个体观测（文本）与整个社会现象（情境）之间来回游走的迭代过程，从而协调纷争以建立一个与各种主观观点以及置于其中的参与者感受相一致的理论。这些在对现象的理解或意义与观察之间的迭代过程将一直持续到达到“理论成熟”阶段，这时其他新的迭代并不能带来对现象更多的洞察。

诠释性数据收集

在诠释性研究中可采用多种技术进行数据收集，其中最为常用的方法是访谈（包括面对面、电话或焦点组）。访谈类型与策略已经在关于调查研究的章节中进行了详述。第二个方法是观察。观察性的技术包括直接观察（研究者保持中立，作为一个被动的局外人并不亲自涉入研究现象，如案例研究）和参与者观察（研究者作为现象的一个积极的参与者，其投入或仅仅是出现将影响被研究的现象，如行动研究）。第三种技术是文件整理，研究者利用外部与内部文档如备忘录、电子邮件、年报、财务声明、新闻报纸、网站等形成对研究现象的进一步洞察或者与其他证据相互印证。

诠释性研究设计

案例研究。正如在上一章所讨论的，案例研究是针对一个或多个研究站点的现象进行集中纵向研究以期获得详尽的情境化的结论或对研究现象的动态过程进行理解的方法。案例研究是一种独特的研究设计，因为其既可以用作阐述性研究以构建理论也可以用作实证研究以验证理论。上一章已经对案例研究的技术进行了深入探讨并给出了典型示例。案例研究者应该是社会情境中的中立观测者（直接观察）而不是积极参与者（参与者观察）。像其他诠释性研究方法一样，能否从案例研究中获取有意义的结论很大程度上依赖于研究者的观察技巧与整合能力。

行动研究。行动研究是一种定性研究方法但其采用实证研究设计旨在理论检验而非理论构建（之所以在本章讨论是因为缺乏合适的地方）。这是一种交互式设计，其假定复杂的社会现象最好可以通过在这些现象中引入改变、干涉、或"行动"并观察这些行动对现象所产生的影响来进行研究。通过这种方法，研究者经常作为一个咨询顾问或者组织成员嵌入到这种社会情境中（例如一个组织），发起应对一个社会问题的行动，然后查看其行动是如何影响现象的，同时可以从行动与现象之间的关系来学习并生成深刻的见解。行动的例子包括组织变革项目，如将新的组织过程、流程、人或技术引入组织，或者替代那些旧的，以期提升一个组织在其商业环境中国的绩效或利润。研究者的行动选择应基于理论，这些理论能够解释这些行动为什么或如何带来预期的社会改变。理论的验证可以通过查看所选择的行动在多大程度上成功地修复了目标问题而实现。同时进行的问题解决与见解生成是行动研究区别与其他研究方法（其可能不涉及问题解决）和咨询（其可能不涉及见解生成）的关键属性。因此，行动研究是连接研究与实践的很好的方法。

行动研究有几种变化形式，而这些方法中最有名的是萨斯曼和埃弗雷德（1978）[13]设计的参与型行动研究。这种方法遵循一个由五个阶段组成的行动研究循环：（1）诊断，（2）行动计划，（3）行动开展，（4）评估，（5）学习（见图 12.1）。诊断包括在社会情境中确认与定义问题。行动计划包括确认与评价针对问题的备选方案并基于理论原理决定未来将采取的一系列行动。行动开展是执行计划的一系列行动。评估阶段分析发起的行动在多大程度上成功地解决了原来的问题，即根据理论得出的效果是否在实践中真实发生。在学习阶段，利用从行动评价中得来的体会与反馈以形成对问题的领悟并对未来的行动的修改或改进提出建议。基于行动评估与学习，应对行动进行修改或调整以更好地解决问题，然后根据修改的行动重复这个研究循环。一般建议将整个行动研究循环至少遍历两遍，这样从第一次循环中学到的知识可以被用于第二次循环的实施过程。数据收集主要采用参与观察的方法，其他的技术如访谈或文档证据也可用于佐证研究者的观察。

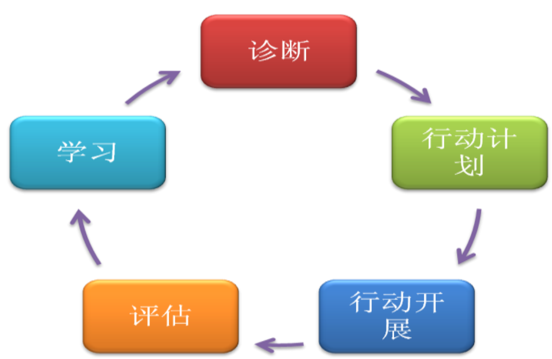


图 12.1 行动研究循环

人种志法。人种志研究法源于人种志学，其强调在特定文化背景下对现象进行研究。研究者必须在其社会文化环境中浸入很长一段时间（一般 8 个月到 2 年），并应参与、观察和记录

[13]Susman, G.I. and Evered, R.D. (1978). "An Assessment of the Scientific Merits of Action Research," *Administrative Science Quarterly*, (23), 582-603.

所研究的文化及其社会参与者在该自然情境中的日常生活。主要的数据收集方法是参与型观察，数据分析包括"意会"的方法。此外，研究者必须做大量现场笔记，用细节描述来记录其体会，这样读者才能像研究者一样感受到同一个文化。在这种方法中，研究者起两个作用：利用其特殊的知识与参与来形成深刻理解（理论）以及说服学术界知晓所研究的现象的跨情境本质。

人种志学研究的一个经典的例子是珍妮·古道尔关于灵长目行为的研究。在该研究中，她与黑猩猩一起生活在其位于冈贝国家公园的栖息地，观察他们的行为，与他们进行互动，共享他们的生活。在此过程中，他掌握并记录下黑猩猩是如何寻找食物与庇护，他们是如何与同类进行交往，他们的沟通方式以及交配行为等。更近的一个例子是迈拉·兰格（1996）[14]关于有子女受到有生命威胁的疾病的家庭的决策行为的研究，以及影响这种决策的物理、心理、环境、道德、法律与文化议题。研究者跟踪了大约 80 个患有不可治愈疾病的儿童及其家庭 2 年以上的时间。数据收集包括参与型观察以及与儿童及其父母与亲属以及医疗提供者之间的交谈，从而整理他们的生存体验。

现象学。现象学是一种强调研究有意识的体验并将此作为理解我们周围的现实的一种方式。它基于 20 世纪早期德国哲学家埃德蒙德·胡塞尔的想法，认为人的体验是一切知识的源头。现象学对与有意识的体验（如人的判断、看法与行动）相关的现象进行系统性的思考与分析，从而实现（1）从相关参与人的各种主观看法来体会与描述社会现实，以及（2）理解这些主观感受背后的符号意义（"深度结构"）。现象学调查需要研究者摒弃其任何先见与个人偏向，置身于参与者的情境，设身处地感受这种情景，从而对影响参与者的思考、感受与行为的深度结构进行充分理解。

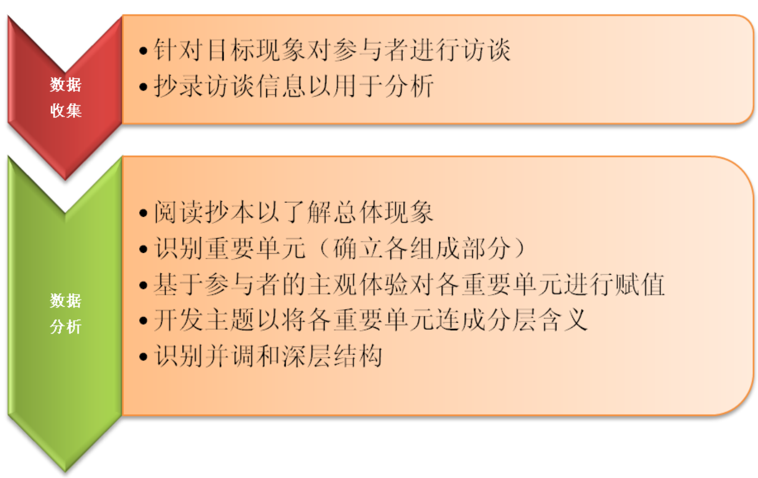


图 12.2 现象学研究方法

[14]Bluebond-Langer, M. (1996).*In the Shadow of Illness: Parents and Siblings of the Chronically Ill Child*. Princeton, NJ: Princeton University Press.

有些研究者将现象学看作一种哲学而不是一种研究方法。针对这个批判，乔治和乔治（2003）[15]开发出一种存在现象学研究方法以指导该领域的研究。如在图 12.2 中所示，这种方法可以分作数据收集与数据分析两个阶段。在数据收集阶段，首先对置身于某个社会现象中的参与者进行访谈以获取其关于调研现象的主观感受与观点。示例性的问题可以是“你能否描述下你平常的一天”或“你能否对那个特别事件给出更细节的描述”。对这些访谈进行录制并抄录，以备进一步分析。在数据分析阶段，研究者阅读记录以（1）获得一个整体印象，并（2）列出那些能够代表参与者主观感受的“显著性单元”。显著性单元的例子是某种概念如“感受到的空间”和“感受到的时间”，这些都可用于对参与者的心理感受进行归档。例如，参与者是否在体验一种现象（“感受到的空间”）时感到安全、轻松、深陷其中或快乐？参与者是否感受到压力、慢速或不连续（“感受到的时间”）？现象学分析应该考虑参与者的临时景象（即他们对过去，现在和将来的感觉），研究者必须通过想象将自身换位于参与者的情景（即临时生活在参与者的生活中）。将参与者的生活感受采用陈述方式或或主题刻画方式进行描述，并在这种脆弱与模糊的生活感受中确定参与者所表达的意思的多个层面。

诠释性研究的严谨性

实证研究采用“演绎”的方法将社会现实简化为简约的理论和法则，阐释性研究则试图通过涉入现实所处的情境之中的参与者的主观视角来诠释社会现实。这些诠释是非常情境化的，自然难以推广到其他研究情境。因为诠释性分析是主观性的且对嵌入其中的研究者的经验与洞察力非常敏感，这种方法经常被实证主义（功能主义）研究者认为是不严谨的。不过，由于诠释性研究与实证主义研究是基于不同的本体论与认识论假设的，实证主义者所认为的严谨如信度，内部效度与广义性在此并不适用。尽管如此，林肯和古帕（1985）[16]提出了一系列用于判断诠释性研究的严谨性的标准。

可靠性。如果两个研究者采用同一套证据独立地对同一个现象进行评价可以得出同样的结论，或者同一个研究者在不同的时间对同一个或相似的现象得出相似的结论，则可以认为该诠释性研究是可靠的。这个概念与实证研究中的信度是相似的，两个独立研究者的一致性与评估者间信度是相似的，同一研究者对同一现象的两次观测的一致性与重测信度是相似的。为确保可靠性，诠释性研究者必须对他们的研究现象与研究现象所在的社会情境给出充足的细节，这样可以使读者独立地鉴定他们的诠释性结论。

可信性。当读者认为其结论可信时，诠释性研究可看做是可信的。这个概念与实证研究中的内部效度很相像。诠释性研究的可信性可以通过多种方式来提高，如提供研究者在现场的参与的证据，展示数据是如何从不同参与者的聚合或收集的，保持一丝不苟的数据管理与分析流程，例如访谈的逐字记录，精确地录制访谈，清晰地记录理论上和方法上的决策，必要的时候还可以对数据收集与分析进行独立的审计。

[15]Giorgi, A and Giorgi, B (2003) Phenomenology. In J A Smith (ed.) *Qualitative Psychology: A Practical Guide to Research Methods*. London: Sage Publications.
[16]Lincoln, Y. S., and Guba, E. G. (1985). *Naturalistic Inquiry*. Beverly Hills, CA: Sage Publications.

可确证性。可确证性指的是在诠释性研究中报告的发现在多大程度上能够独立地被他人所确证（例如参与者）。这与实证研究中的客观性的概念相似。因为诠释性研究否认客观现实的概念，可确证性可以根据“间主观性”来表征，即研究的参与者认同研究者得出的结论。例如，如果一个研究的参与者总体上认同研究者从研究现象中得出的结论（基于对研究文章或报告的阅读），那么可以认为这些发现是可确证的。

可转移性。诠释性研究中的可转移性指的是研究发现在多大程度上能推广到其他情境。这与实证研究中的广义性很相似。研究者必须提供关于研究情境的丰富的细节的描述（“深厚描述”），贯穿地对数据所揭示的结果、假设与过程进行描述，这样读者可以独立地判断所报告的发现是否或多大程度上可转移到其他情境。

第十三章

定性分析

定性分析是对定性数据如从访谈脚本中提取的文本数据的分析。与定量分析是统计导向的以及很大程度上独立于研究者的不同，定性分析很大程度上依赖于研究者的分析与整合技术以及研究者对数据收集的社会情境的个人知识。定性分析中的重点是“意会”或对现象的理解而不是预测或解释。基于道德性启迪，置参与者于情境中的态度以及一系列分析策略，定性分析还需要创新性的与审查性的思维模式。该章将对一些定性分析的策略进行简要的概述，感兴趣的读者可以参考其他更权威和详细的文献如迈尔斯和休伯曼（1984）[17]关于该话题的奠基性的书籍。

扎根理论

你应该如何分析大量通过参与者观察，深度访谈，焦点组，音频（视频）录音（录相）的叙述或者二手文档等途径得来的定性数据呢？用于分析文本数据的一项技术是扎根理论研究法，它是一种基于社会现象的描述性记录数据进行归纳分析并构建相关理论的技术。这项技术是由格拉泽和施特劳斯（1967）[18]开发的，其提出了扎根理论研究的不断比较分析方法，施特劳斯和科尔宾（1990）[19]随后对其进行了改进并提出了具体的编码技术—一个将文本数据的片段归类为一系列代码（概念），类别（变量）和关系的过程。由于诠释是“扎根于”（或基于）观测到的实证数据的，因此得名“扎根理论研究法”。为了确保理论仅仅是基于观测到的证据的，扎根理论法需要研究者在数据分析前禁止一切提前存在的理论预期或偏差，让数据告知你理论是如何形成的。

---

[17]Miles M. B., Huberman A. M. (1984). *Qualitative Data Analysis: A Sourcebook of New Methods*. Newbury Park, CA: Sage Publications.
[18]Glaser, B. and Strauss, A. (1967).*The Discovery of Grounded Theory: Strategies for Qualitative Research*, Chicago: Aldine.
[19]Strauss, A. and Corbin, J. (1990).*Basics of Qualitative Research: Grounded Theory Procedures and Techniques*, Beverly Hills, CA: Sage Publications.

施特劳斯和科尔宾（1998）介绍了分析文本数据的三种编码技术：开发式的，主轴式的和选择式的。开发式编码是一个旨在确定隐藏于文本数据中可能与研究现象相关的概念或关键想法的过程。研究者对原始文本数据一行一行进行分析以确定离散事件、突发事件、想法、行动、观点以及相关性的交互，这些被编码为概念（因此称为“有机编码”）。每个概念是与以备后续验证的具体比例的文本（编码单元）相联系的。某些概念可能是简单、清晰与无异议的，而另外一些则可能是复杂、模糊或者仁者见仁的。编码的单元会随着提取的概念而变化。简单的概念如“组织规模”可能只包括文本中的几个词，而复杂的概念如“组织使命”则可能横跨几页。概念可以用研究者自己的命名习惯或者从研究文献中得来的标准化标签来命名。一旦确立好基准概念，这些概念就可以用于对其余数据进行编码，而且需要同时寻找新概念与更新旧概念。编码时，确定每个概念可识别的特征（如大小、颜色、水平高或低）非常重要，这样相似的概念可以被归类到一起。这种编码方法称之为“开放式”编码因为研究者对与研究现象相关的新概念持开放态度并对其积极搜寻。

然后，相似的概念被归类到更高的层次类别。虽然概念可能是情景化的，类别倾向于广泛与可推广的，其最终将演化为扎根理论中的变量。需要用类别来降低研究者需要处理的概念的数量，从而勾画出用于理解社会现象的主要议题的“全局图”。归类是分步完成的，先将概念组合为子类，然后再将子类组合为更高层级的类。已有文献中的变量可用于对类别进行命名，特别是当研究的目的是扩展现有的理论时。不过，在利用已有变量进行命名时需要非常谨慎，因为这些变量可能会带来惯有的信念与偏差。需要对每个类别的特征（或属性）和每个特征的维度进行确认。维度表示一个特征的连续性的值。例如，一个名为“沟通媒体”的类别可能有一个特征称之为“速度”，其可以分为快、中和慢。这样的归类有助于区分不同类别的沟通媒体并使研究者可以确定数据的模式，例如对于哪类任务采用了哪类媒体。

扎根理论法的第二个阶段是主轴式编码。此时，类别或子类被放置于因果关系或假设之中，可以常识性地对研究现象进行解释。尽管与开放式编码不同，主轴式编码可与开放式编码同时进行。类别之间的关系可能可以通过数据清晰的判别或者非常微弱或不可见。在后面的情况下，研究者可以采用编码主题（经常称为“编码范式”，但其与第三章所提到的范式概念不同）来理解哪种类别代表条件（现象所处的情况），行动或交互（个体对事件在某种条件下的反应），和结果（行动或交互的产物）。当条件、行动/交互和结果确定后，理论假说就开始浮现了，研究者可以开始解释为什么一个现象会发生，在什么条件下发生，以及其结果是什么。

扎根理论法的第三个也是最后一个阶段是选择性编码，涉及到确定核心类别或核心变量并系统性地合乎逻辑地将核心类别与其他类别相联系。核心类别可以从已存在的类别进行演化，也可以是包括已编码类别的更高层级的类别。对新的数据进行选择性抽样以验证核心类别以及其与其他类别的关系（即尝试性理论）。选择性编码限制了分析的范围，使分析更快。同时，编码者必须注意可能从与研究现象相关的新数据中出现的其他类别（开放式编码），这将促使初始理论的进一步优化。这样，开放式、主轴式和选择式编码可能同时进行。对新数据进行编码并对理论进行精炼直到到达理论成熟，即当新数据并不会给核心类别或关系带来任何边际改变。

“持续比较”过程暗示了基于增加的理解深度对类别、关系和表述不断的重新安排、综合与精炼，实现下述四阶段活动的迭代：（1）比较分配到各个类别中的事件或文本以验证该类别的有效性，（2）整合类别及其属性，（3）界定理论，关注于那些核心概念而忽略那些不太相关的概念，以及（4）用下面将要讨论的诸如书写备忘录、叙事和图解的方式来撰写理论。形成一个中心类并不意味着所有其他类别可以很好地与其整合在一起。为了确定核心类的条件、行动/交互和结果等关键类别，施特劳斯和科尔宾（1990）推荐了叙事、回忆录或概念映射等几种整合技术。采用叙事方式，可以用类别与关系来对关于所观测的现象的故事进行精炼。备忘录是指对那些随扎根理论分析的开展而记录下的关于概念及经过理论化编码的概念间关系的观点，它是在分析过程中对观点进行追踪与精炼的重要工具。书写备忘录就是利用这些备忘录，以二乘二的表格、图表、图示或其他展示方式来发现类别之间的模式或关系的过程。概念映射是对概念及其关系的图形化表示（如方框和箭头）。将主要概念在一页或多页纸、黑板或图形软件程序上进行展示，通过箭头连接不同概念，对其进行调整以适应观察数据。

在形成一个扎根理论之后，必须对其进行精炼以符合内部一致性与逻辑。研究者必须确保中心变量具有所陈述的特征和维度，如果没有，那么需要重新进行数据分析。研究者必须保证这些特征与维度在不同类别间具有方差。例如，如果行为频率是这样一个类，那么数据应该包括针对该行为的高频行为者与低频行为者。最后，理论必须与原始数据相比较以验证理论的有效性。如果理论与观测证据相冲突，可能需要重复编码过程以消除这些冲突或未解释的方差。

内容分析

内容分析是以定量或定性的方式对文本内容（如谁对谁说了什么，为什么这么说，有什么效果，效果的程度如何等）进行的系统性分析。内容分析一般按如下步骤开展。首先，当要分析的文本很多时（如新闻故事、财务报告、博客、在线评论等），研究者需要从总体样本的文本中抽取一部分文本进行分析。这个过程并不是随机的，那些含有丰富内容的文本被选择性地选取。其次，研究者应确定并应用某种规则将文本分割为可作为独立分析单元的部分或“块”。这个过程称之为单元化。例如，每个单元可能由文本中的假设、效果、推动因素与阻碍因素而构成。再次，研究者构建并应用在每个单元化的文本部分中的一个或多个概念的过程被称作编码。为了满足编码的目的，在对文本进行归类时，研究者需要采用一种基于其正在寻求或揭示的主题的编码方案。最后，对编码数据进行定量和定性分析以决定哪些主题出现频率最多，在何种情境出现以及这些主题间是如何相互关联的。

一种简单的内容分析类型是情感分析，它是一种用于刻画人对一个对象、人或现象的观点或态度的技术。这种分析的一个例子是，阅读一个在线论坛上关于某位政治候选人信息，并将每条信息分作正向的、负向的或中立的。在这个例子中，每条信息代表了一个分析单元。这种分析有助于确认样本整体是否对候选人持正向、负向或中立的态度。用同样的方式对在

线评论的内容进行分析是另外一个例子。尽管这种分析可以手动完成，当数据集非常庞大时（上百万的文本记录），基于自然语言处理和文本分析的软件程序可用于自动编码过程，并可以保持人们随时间的情感波动记录。

对于内容分析的常见的一种批评是其缺乏让其他研究者复制该分析的一系列系统性的程序。针对这一批评，席林（2006）[20]提出了综合不同内容分析步骤的螺旋模型。该模型由文本描述的五个水平或阶段组成：（1）将录音转换成可供内容分析的原始文本数据或脚本，（2）将原始数据转换成压缩协议，（3）将压缩协议转换成初步的类别系统，（4）利用初步类别系统形成编码协议，（5）分析编码协议，形成对研究形象的表述。

内容分析有几点缺陷。第一，编码过程受制于文本形式的信息。例如，如果研究者有兴趣研究人们对死刑的看法，但不存在与之相关的文档，那么该分析无法进行。第二，采样是必须非常认真以防止采样偏差。例如，如果你采用的样本总体是关于某一话题的发表的研究文献，那么你已经系统性地忽略了那些未发表的研究或者那些最近的即将发表的文献。

解释学分析

解释学分析是一种特殊类型的内容分析，研究者可以采用该方法尝试对在特定社会历史情境中的某个文本的主观意义进行"诠释"。与那些在编码过程中忽视文本文件的情境与意义的扎根理论或内容分析不同，解释学分析是一种真正的分析定性数据的诠释性技术。这种方法假定所写的文本描绘了作者在某个特定社会历史情境中的感受，应该结合特定情境进行诠释。因此，研究者不断地在文本的孤立表示（部分）与对情境的全盘理解（整体）之间进行迭代，以形成对嵌入情境中的现象的更深刻理解。德国哲学家马丁 海德格尔(将其称为解释性循环)。"解释性"一词指的是诠释的一种独特的方法或分支。

一般而言，解释学是关于诠释的研究以及关于诠释的理论与实践。传统解释学源于宗教研究（如圣经注解学）或语言学，指的是对书写的文本特别是在文献学、宗教与法律（例如圣经）领域的诠释。在 20 世纪，海德格尔提出理解社会现实的更为直接、无中介和真实的方式是对现实进行体验而不仅仅是观测，因而提出了哲学释义学。海德格尔指出文本不仅仅是读者阅读作者的体验的工具，更是在想象中再次经历作者的体验的工具。由海德格尔的学生如伽达默尔发展的当代或近代解释学进一步分析了用于交流社会经验的书写文本的局限性，并进一步提出了整合所有沟通形式（包括书写的、口头的、非口头的）的诠释性过程的框架，探索了限制书写文本的沟通能力的原因，如预设、语言结构（如语法与句法等）和符号（关于书写标志如象征主义、隐喻、类比和反讽等的研究）。解释学一词有时也与注释学不太准确地进行互用，注释学仅仅是针对书写文本特别是经书的解释或批注。

[20]Schilling, J. (2006). "On the Pragmatics of Qualitative Assessment: Designing the Process for Content Analysis," *European Journal of Psychological Assessment* (22:1), 28-37.

结论

最后，标准的软件程序如 ATLAS.ti.5，Nvivo 和 QDA Miner 可用于定性研究方法中的自动编码过程。这些程序可以根据用户定义的规则快速有效地组织、寻找、归类与处理大宗的文本数据。为了给自动分析提供指引，研究者需要制定一个编码规则，在对样本文本数据进行初始的收到分析后明确在对文本进行搜索时的关键词或代码。编码规则可以按层级将代码分为高阶代码或变量。需要利用一个不同的文本样本对编码规则验证其精确性与充分性。如果编码规则出现偏误或不准确，那么对于整个总体文本的分析将存在严重缺陷，而难以作出解释。不过，软件程序并不能破译某些字词、短语或这些字词或短语的应用情境（如在反讽或隐喻中）的意义，这将导致大规模定性分析的显著的错误诠释。

第十四章

定量分析：描述性统计

研究项目中搜集的数字数据可以运用两种不同的统计工具进行定量分析。描述性分析是指统计性地描述总体及展示兴趣的构念或构念间的联系。推断性分析是指假设的统计检验（理论检验）。在这章中，我们将检验用于描述性分析的统计技术，在下章中将检验用于推断分析的统计技术。当前大多数定量数据分析是用软件程序处理的，比如 SPSS 或 SAS。为了理解本章中描述的概念，建议读者熟悉一种这样的程序。

数据准备

在研究项目中，数据可能从各种各样的来源被收集：邮件调查、访谈、事前测试或者事后测试的实验数据，观察数据等等。这个数据必须被转成机器可读的、数字格式，比如电子表格或文本文件，以使它们可以被电脑程序分析，如 SPSS 或 SAS。数据准备通常需要遵循以下的步骤。

数据编码。编码是将数据转化成数字格式的过程。需要创建一个编码簿来指导编码过程。编码簿是一个综合的文件，其包含研究中每个变量的详细描述、那些变量的条目或计量、每个条目的格式（数字、文本等等）、每个条目的反应尺度（例如，它是按名义变量、顺序量表、等距量表还是等比变量来计量；这样的量表是五点式、七点式还是一些其他类型的），以及怎样将每个值编码成数字表格。例如，如果我们有一个用七点式李克特量表计量条目，锚点从“强烈不同意”到“强烈同意”变化，那么我们可以这样编码这些条目， 1 表示“强烈不同意”，4 表示“中性”，7 表示“强烈同意”，中间的锚点介于其间。名义变量如行业类型，可以按照这样一套编码计划的数以字形式编码：1 表示制造业，2 表示零售业，3 表示金融业，4 表示医疗业，以此类推（当然，定类数据不能被统计分析）。定比数据，如年龄、收入，或测试得分可以按照应答者进入那样编码。有时，数据可能需要被加总成不同于数据搜集格式一样的格式。比方说，为了计量如“电脑的好处”这样的构念，如果一项调查向调查对象提供一个好处的清单，他们从中进行选择（即，他们可以选择他们想要的尽可能多的好处），那么受查条目的总数可以被用作好处的一个总体计量。注意许多其他形式的数据，如访谈手稿不能被转换为用于统计分析的数字格式。编码对于大量复杂研究尤为重要。这些研究中包括许多变量和计量条目，编码过程被不同的人实施，为了帮助编码团队按照一致的方式编码，而且为了帮助其他人解释编码数据。

数据输入。编码数据可以被录入电子表格、数据库、文本文件或直接进入统计程序如SPSS。大多数统计程序提供录入数据的数据编辑器。然而，这些程序以它们自己的原本格式存储数据（比如，SPSS 以.sav 文件存储数据），这将使得在其他统计程序中共享数据变得困难。因此，一般来说将数据录入电子表格或数据库中比较好，这样数据可以根据需要被重整，在程序中共享，而且数据的子集可以为分析而被提取。更小的数据集，一般少于 65000 个观测值和 256 个条目，可以存储在如 Microsoft Excel 的电子表格中。然而拥有几百万观测值的较大数据集就需要一个数据库。每个观测值被录入进电子表格中的每行，每个计量条目列在每列中。输入的数据需要在输入过程中或结束后被经常检查其准确性，可以通过对一组条目和观测值的不定期抽查。而且，当录入数据时，编码者应该注意明显较差的数据，如调查对象选择“强烈同意”回答与内容无关的所有条目，包括反向编码的条目。如果这样，这些数据可以被录入但应该从随后的分析中被剔除。

缺失值。缺失数据是任何实证数据集不可避免的一部分。如果它们用词模糊或太敏感的话，调查对象可能不回答特定的问题。此类问题应该在事前测试中尽早地被发现，而且在主数据收集过程开始前被改正。在数据录入时，一些统计程序自动地将空白记录视为缺失值，而其他程序则要求录入具体的数字值如-1 或 999 来代表缺失值。在数据分析中，大多数软件程序中处理缺失值的默认模型是仅仅删掉整个观测值，包括即使一个单一的缺失值，该技术叫成列删除法。此类删除会显著减少样本规模，使得发现细微影响极其困难。因此，一些软件程序允许选择将缺失值替代成估计值，通过一个叫插补法的程序。例如，如果缺失值是一个多要素尺度的条目，缺失值可能是调查对象对剩余条目在那个尺度上反应的平均值。如果缺失值属于单要素尺度，许多研究者用其他调查对象对那个要素反应的均值作为插补值。如果缺失值是系统而不是随机的，此类插补可能会有偏误。两种可以产生相对无偏估计的插补方法是最大似然程序和多重插补方法，它们都被流行的软件程序如 SPSS 和 SAS 所支持。

数据转换。有时有必要在数据可以被有意义地解释前转换数据值。例如，反向编码要素表达了与它们潜在的构念相反的意思，应该在它们与没有反向编码的要素比较或结合前被反转（例如，在 1-7 的间隔尺度中，8 减去观测值将反转原值）。其他类型的转换可能包括通过增加单个尺度要素创造尺度计量，从一组观测的测度中创造加权指数，以及将多个值拆叠为更少的分类（例如，将收入值拆叠为收入的变化）。

一元分析

一元分析，或单变量分析，是指可以描述一个变量总体属性的一套统计方法。单变量统计包括：（1）频数分布，（2）集中趋势，和（3）离差。一个变量的频数分布是那个变量的单个值或是范围值的频数（或百分比）。例如，我们可以度量一个样本的调查对象参加了多少次宗教服务（表示他们的“宗教性”的一个变量），通过用一个分类尺度：从来不，每年一次，每年几次，大约每月一次，每月几次，一周几次，和一个对于“不回答”的选择性分类。如果我们计数每个分类（除了“不回答”，实际表示一个缺失值而不是一个分类）下的观测值数量（或百分比），并以表格的形式列示，如图 14.1。这就是频数分布。这个分布也可以以直方图的形式描绘出，即图 14.1 右侧，横轴表示那个变量的每个分类，纵轴表示每类中观测值的频数或百分比。

	频次	百分比
从不	85	29.9%
每年一次	23	8.1%
每年多次	36	12.7%
每月一次	44	15.5%
每月多次	56	19.7%
每周多次	21	7.4%
无回答	19	6.7%
总计	284	100.0%

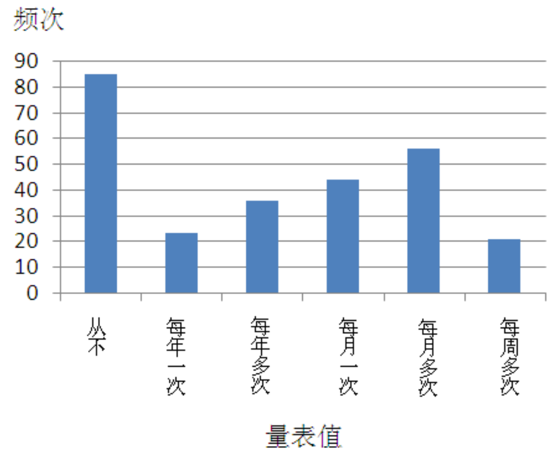


图 14.1. 宗教狂热的频数分布

由于许多大样本的观测值是独立和随机的，频数分布倾向于逼近一个图，看起来像钟形曲线（频数分布的一个光滑的直方图），类似于图 14.2 所示，大多数观测值向中心聚集，越来越少的观测值趋向极端。这样的曲线叫正态分布。

集中趋势是分布的中心值的估计。有三种集中趋势的主要估计：均值、中位数和众数。算术平均值（经常仅被叫做“均值”）是一个给定分布的所有值的简单平均。考虑一组 8 测试得分：15, 22, 21, 18, 36, 15, 25, 15.这些值的算术平均数是(15 + 20 + 21 + 20 + 36 + 15 + 25 + 15)/8 = 20.875.其他类型的均值包括几何平均数（一个 n 个数值分布的 n 次方根）和协和平均数（一个分布中每个值的倒数的算术平均数的倒数），但是这些均值在社会研究数据的统计分析中不是很流行。

第二种计量集中趋势的方法——中位数，是一个分布的值域的中间值。这通过用增序排列所有的值并挑出中间值计算得到。在有两个中间值的情况下（如果分布有偶数个值），这个两个中间值的平均数代表中位数。在上述例子中，排序后的值是：15, 15, 15, 18, 22, 21, 25, 36. 这两个中间值是 18 和 22，因此中位数是(18 + 22)/2 = 20.

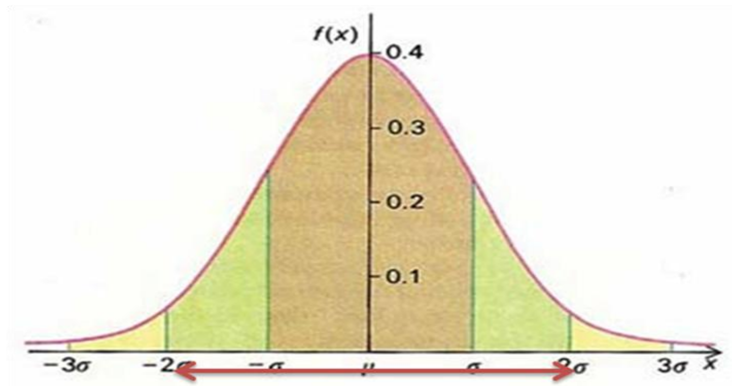
最后，众数是值的分布中出现最频繁的值。在前述例子中，最频繁出现的值是 15，这就是上述测试得分的众数。注意从样本中估计出的任何值，比如均值、中位数、众数，或者任何后来的估计都叫统计。

离差是指值分布于集中趋势周围的方式，例如，值是怎样紧密或怎样广泛聚集在均值周围的。两种普遍的离差的测度是极差和标准差。极差是最高值和最低值之间的差异。我们前述例子中极差是 36-15 = 21.

极差对极端值的出现特别敏感。例如，如果上述分布的最高值是 85，其他值保持不变，极差就是 85-15 = 70。标准差，第二种离差的测度，通过使用一个公式修正了这种极端值。这个公式考虑了每个变量与分布的均值的距离：

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n - 1}}$$

σ 是标准差，x_i 是第 i 个观测（或值），μ 是算术平均数，n 是观测值的数量，Σ 表示所有观测值的总和。标准差的平方叫做分布的方差。在正态的频数分布中，68%的观测值位于均值的一倍标准差之间(μ + 1 σ)，95%的观测值在两倍标准差之间(μ + 2 σ)，99.7%的观测值位于三倍标准差内(μ + 3 σ)，如图 14.2 所示。



在正态分布中，95%的观察落于均值±两个标准差的范围内 (μ ± 2σ)

图 14.2 正态分布

双变量分析

双变量分析检验两个变量是如何相互关联的。最普通的双变量统计是双变量相关系数（经常被简称为"相关系数"），一个介于-1 与 1 之间的值，表示了两个变量之间的相关强度。假设我们希望研究年龄如何跟自尊相关，以 20 个调查对象为样本，比如随着年龄的增长，自尊是会增加，减少，还是不变。如果自尊增加，那么我们得到这两个变量间的正相关关系；如果自尊减少，我们得到负相关关系；如果保持不变，我们得到零相关关系。为了计算相关系数的值，考虑表 14.1 的假设数据集。

观察	年龄(x)	自尊(y)	xy	x^2	y^2
1	39	4.1	159.9	1521	16.81
2	45	4.6	207	2025	21.16
3	29	3.8	110.2	841	14.44
4	42	4.4	184.8	1764	19.36
5	19	3.2	60.8	361	10.24
6	22	3.1	68.2	484	9.61
7	39	3.8	148.2	1521	14.44
8	30	4.1	123	900	16.81
9	33	4.3	141.9	1089	18.49
10	23	3.7	85.1	529	13.69
11	20	3.5	70	400	12.25
12	18	3.2	57.6	324	10.24
13	24	3.7	88.8	576	13.69
14	22	3.3	72.6	484	10.89
15	29	3.4	98.6	841	11.56
16	35	4.0	140	1225	16.00
17	36	4.1	147.6	1296	16.81
18	37	3.8	140.6	1369	14.44
19	35	3.4	119	1225	11.56
20	32	3.6	115.2	1024	12.96
总计(Σ)	**609**	**75.1**	**2339.1**	**19799.0**	**285.45**

$r_{xy} = 0.79$

表 14.1 年龄和自尊的假设数据

这个数据集中的两个变量是年龄（x）和自尊（y）。年龄是一个定比变量，而自尊是从多要素自尊量表计算的平均得分而来的，运用七点式李克特量表，从“强烈不同意”到“强烈同意”变动。每个变量的直方图显示在图 14.3 左列。计算双变量相关系数的方程是：

$$r_{xy} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x s_y} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\ \sqrt{n\sum y_i^2 - (\sum y_i)^2}}.$$

r_{xy} 是相关系数，x 和 y 是变量 x 和 y 的样本均值， s_x 和 s_y 是变量 x 和 y 的标准差。如表 14.1 所示，手工计算年龄和自尊之间的相关系数是 0.79。这个数字表明了年龄和自尊之间强烈正相关。即自尊倾向于随着年龄的增长而增强，随着年龄的减小而减弱。这样的模式也可以从比较图 14.3 中的年龄和自尊的直方图看得出，两个直方图的顶点大致相互紧随。图 14.3 的纵轴表示真实的观测值而不是观测值的频数（如图 14.1 所示）。因此，这不是频数分布而是直方图。图 14.3 的右侧是双变量的散点图，纵轴表示自尊，横轴表示年龄。这个图粗略地聚集成了一个上倾的斜线（即正斜率）。如果两个变量是负相关，散点图应该是向下倾斜的（负斜率），表明年龄的增长与自尊的减弱相关，反之亦然。如果两个变量不相关，散点图将接近于一条水平线（零斜率），表明年龄的增长与自尊非系统相关。

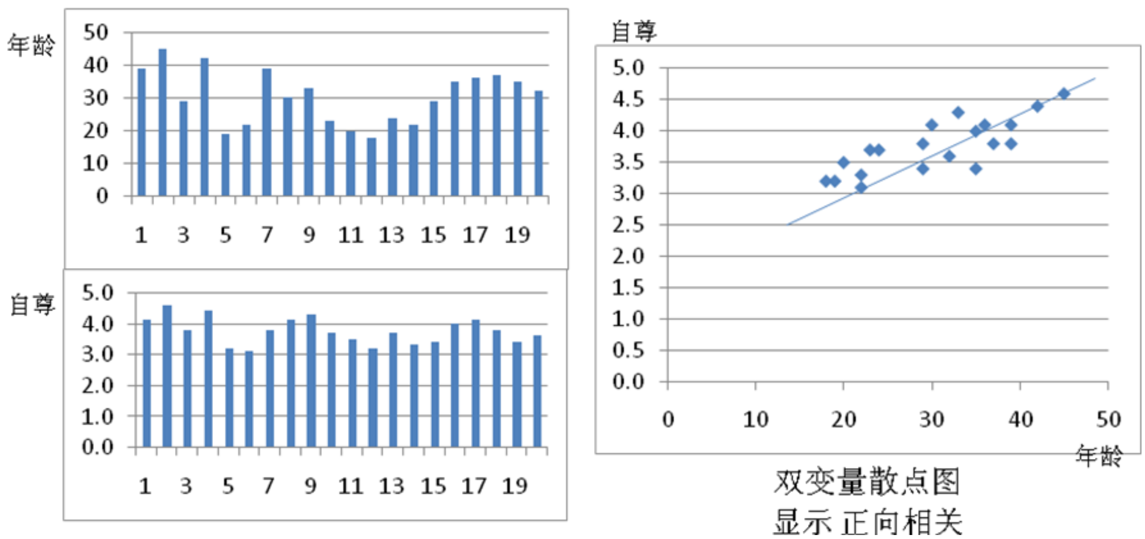


图 14.3 年龄和自尊的直方图和相关系数图

在计算了双变量相关性后，研究者通常比较有兴趣知道相关性是否显著（即一个真正的）或是仅仅偶然引起的。回答这样一个问题需要检验如下假设：

H_0: r = 0

H_1: r ≠ 0

H_0 被称为原假设，H_1 被称为备择假设（有时也表示为 H_a）或我们真正想要检验的假设（即相关系数是否异于 0）。尽管他们可能看起来像两个假设，H_0 和 H_1 联合代表了一个单一的假设，因为他们彼此相反。也要注意到 H_1 是非方向性的假设，因为它没有具体说 r 是大于还是小于 0。方向性假设具体为：H_0: $r \le 0$; H_1: $r > 0$（如果我们检验正相关）。方向性假设的显著性检验用单尾 t 检验做，而非方向性检验用双尾 t 检验做。

在统计检验中，备择假设不能被直接或者推定证明。然而，它可以通过在特定概率水平下显著拒绝原假设而被间接证明。统计检验经常是概率性的，因为我们从来不确定基于样本数据的推断是否适用于总体，因为我们的样本从来不等于总体。纯属巧合的统计推断的概率被称为 p 值。p 值是跟显著性水平（α）相对的。α 表示我们愿意接受推断是不正确的最大风险水平。对于大多数统计分析，α 被设定为 0.05。p 值小于 α=0.05 表明我们有足够的统计证据拒绝原假设，因此，间接接受了备择假设。如果 p>0.05，那么我们没有充足的证据拒绝原假设或接受备择假设。

检验上述假设最简单的方法是在统计表格中查阅 r 的临界值。统计表格在任何标准的统计教科书或互联网上（大多数软件程序也执行显著性测试）都可以得到。r 的临界值依赖于我们所要求的显著性水平（α = 0.05），自由度（df）以及要求的检验是单尾检验还是双尾检验。自由度是指在统计的任何计算中可以自由变动的值的数量。在相关系数中，自由度等于 n – 2。如在表 14.1 中，自由度是 20 – 2 = 18。对于单尾检验和双尾检验有两种不同的统计表格。在双尾检验中，对于 α = 0.05、df = 18，r 的临界值是 0.44。要是我们计算的 0.79 的相关性显著，r 必须大于临界值 0.44 或小于-0.44。由于我们计算的 0.79 大于 0.44，可以推论出在我们的数据集中，年龄和自尊有显著的相关性。也就是说相关性是偶然性的概率小于 5%。因此，我们可以拒绝 $r \leq 0$ 的原假设，间接表明了 $r > 0$ 的备择假设是可能正确的。

大多数研究涉及多于两个变量。如果有 n 个变量，那么在这 n 个变量中，我们将共有 n*(n-1)/2 个相关系数。此类相关系数通过如 SPSS 样的软件程序可以轻易地计算出，而不是相关系数公式（如表 14.1）手工计算，并且以相关系数矩阵的形式列示，如表 14.2 所示。一个相关系数矩阵是这样一个矩阵：在第一行和第一列列示变量名称，并在矩阵合适的单元格中描述一组变量间的双变量相系数。这个矩阵的主对角线（从左上角到右下角）的值一般为 1，因为任何变量都跟它自身完全相关。而且，由于相关性是非方向性的，变量 V1 和 V2 间的相关系数与 V2 和 V1 间的相关系数是一样的。因此，下三角矩阵（主对角线以下的值）是上三角矩阵（主对角线以上的值）的一面镜子。如此，为了简洁，我们便常常只列示出下三角矩阵。如果相关性包括以等距尺度计量的变量，那么相关系数的具体类型被称为皮尔逊积差相关系数。

	V1	V2	V3	V4	V5	V6	V7	V8
V1	1.000							
V2	0.274	1.000						
V3	-0.134	-0.269	1.000					
V4	0.201	-0.153	0.075	1.000				
V5	-0.095	-0.166	0.278	-0.011	1.000			
V6	-0.129	0.280	-0.348	-0.378	-0.009	1.000		
V7	0.171	-0.122	0.296	0.086	0.193	0.233	1.000	
V8	0.518	0.238	0.238	-0.227	-0.551	0.082	-0.102	1.000

表 14.2 八个变量的假设相关系数矩阵

列示双变量数据的另一种有用的方式是交叉表格（常简称为交叉表，有时也被更正式地称为列联表）。交叉表是一个描述两种及两种以上的定类或分类变量的所有组合频数（或百分比）的表格。例如，我们假设有 20 个学生样本的关于性别和分数的观测值，如图 14.3。姓名是一个定类变量（女/男），分数是一个有三个等级（A/B/C）的分类变量。在一个 2x3 的矩

阵中，数据的简单列联表列示了性别和分数间的联合分布（即每个等级每个性别的学生有多少个，以行频数计数或百分比表示）。这个矩阵将帮助我们看清 A、B、C 三个等级的分数是否均匀分布在男生和女生中。表 14.3 中的交叉表数据表明 A 级分数分布偏向于女生：样本有 10 个男生 10 个女生， 5 个女生拿到了 A，而仅一个男生拿到 A。相反，C 级分布偏向于男生：3 个男生拿到 C，而仅一个女生拿到 C。然而，B 级的分布是有点统一的，6 个男生和 5 个女生拿到 B。表格中的最后一行和最后一列被称为边际总数，因为他们表示每类的总数并列示在表格的边上。

观察	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
性别	F	M	F	M	F	M	M	M	F	F	M	M	M	F	F	M	F	F	F	M
分数	A	B	B	B	C	A	C	B	B	A	B	C	C	B	B	B	A	A	B	B

假设数据集

			分数			总计
			A	B	C	
性别	男	数目	1	6	3	10
		预期数目	2.5	5.5	2.0	
	女	数目	4	5	1	10
		预期数目	2.5	5.5	2.0	
总计			5	11	4	20

边际总和

性别和年龄的交叉列表

表 14.3 交叉表分析的例子

尽管我们可以看到表 14.3 中男生和女生的分数分布有明显不同的模式，这个模式是否是真正的或“统计显著”呢？换句话说，以上的频数计数与纯属巧合的预期是否不同呢？为了回答这个问题，我们应该在 2 x 3 的交叉表矩阵的每个单元格中计算观测者的预期数。这个可以通过乘以边际列总数和边际行总数再除以总观测值的数量计算得到。例如，男生/A 级的单元格，预期计数=5 * 10 / 20 = 2.5。换而言之，我们预期有 2.5 个男生得到 A，但是实际上只有一个男生拿到 A。预期值和真实值之间的差异是否显著可以通过卡方检验来检验。卡方统计量是计算所有单元格的观测值和预期值均值的差异。然后我们可以将这个数与临界值相比较，临界值是通过期望概率水平（$p < 0.05$）和自由度确定。自由度即(m-1)*(n-1)，m 和 n 分别表示行和列的数量。任何教科书中标准卡方表格中，对于 p=0.05 和 df=2，临界的卡方的值为 5.99。基于我们的观测数据计算的卡方值是 1.00，小于临界值。由此，我们一定可以推论出观察到的分数模式与纯属巧合的预期不是统计上显著区别的。

.

第十五章

定量分析：推断性统计

推断统计是一种研究变量之间关系的统计程序。他们与描述性统计的差别在于他们是专门为检验假设而设计的。许多的统计程序都是属于推断统计，他们其中的大部分都可以通过现代统计软件实现，比如 SPSS 和 SAS。这一章仅提供一个简短的关于最为基本和常用的程序的初级读本，建议读者查阅关于统计的正式文献或者选修一门关于统计的课程来了解更为高级的程序。

基本概念

英国的哲学家卡尔 波普尔曾经说过理论永远不能被证实，只能被证伪。比如说，我们如何能够证明明天太阳会升起来？波普尔认为仅仅因为太阳在我们所记得的每一天都升起来了不足以说明它在明天也会升起来，因为归纳得出的理论仅仅是猜想，对未来的现象可能有预测作用，也可能没有。然而，他认为我们也许可以假设太阳每天都会升起而不需要证明它，当太阳有一天没有升起的时候，这个理论就被推翻了。同样地，我们能在相反的事实上拒绝一个假设而永远不能完全地接受一个假设。因为支持性证据的出现并不意味着我们在后面不会观察到反对性证据。因为我们永远不能真正接受一个我们所关心的假设（被择假设），因而我们拟定一个与被择假设相反的原假设，然后通过用实证证据来拒绝原假设的方式来间接地、概率性地支持我们的被择假设。

在进行社会科学的研究时，第二个与检验假设的关系有关的问题在于因变量可能被无数个无关变量所影响，然而我们不可能测量和控制所有的这些外来影响，因此，即使在我们的观测样本中，两个变量似乎是有关系的，但是他们在总体中可能并不是真正相关的，因而推断统计永远不是确定的，只能是概率性的。

我们怎样才能确定我们从样本中观察到的两个变量之间的关系是显著的（有意义的），而不是一种巧合呢？历史上最为杰出的统计学家之一，罗纳德 费舍尔制定了显著性检验的基本

准则。他说当一种统计结果被拒绝的概率等于或者小于 5%的时候，这个结果就是显著的。在推断统计中，这个概率被称为 P 值，而 5%被称为显著性水平（α），P 值和 α 之间的理想关系是 $p \leq 0.05$。这个显著性水平是我们使用从样本来推断总体这一本来就存在不确定性的方法时愿意承担的最大风险水平。如果 P 值小于或者等于 5%，意味着我们拒绝原假设有 5%的可能性是错误的，这种错误也被称为 I 类错误。如果 P 值大于 0.05，我们就没有充分的理由去拒绝原假设或者接受被择假设。

同时，我们必须理解这三个有关的统计概念：抽样分布、标准误差和置信区间。抽样分布是从你所感兴趣的总体中抽取无限数量的样本时，这些样本的理论分布，然而，由于一个样本永远不可能跟总体完全一样，每一个样本都会有一些固有水平的错误，这被称为标准误差。如果这个标准误差很小，那么从样本中得出的统计估计量（比如样本均值），就是总体的合理估计量。我们的样本估计量的精确性被定义为置信区间。一个 95%的置信区间被定义为在均值估计量的基础上加上或者减去两个标准差的一段区域，这段区域内的数值可以被看作来自于一个抽样分布中的不同样本。因此，当我们说我们的观察样本的置信区间为 95%的时候，我们的意思是我们可以肯定，在 95%的置信区间下，总体参数的值会落在我们观测样本的估计量加减两个标准差的区间内。P 值和置信区间共同给了我们一个好方法来确定我们的结果是正确的概率，以及它与相应的总体参数的接近程度。

一般线性模型

社会科学研究中的大部分推断统计程序都来源于一组统计模型，叫做一般线性模型（general linear model，GLM）。一个模型是指一个估计的数学方程，它可以用来代表一组数据，而线性是指一条直线。因此，一个 GLM 是指可以用来代表观察数据中的一种线性关系的一个方程组。

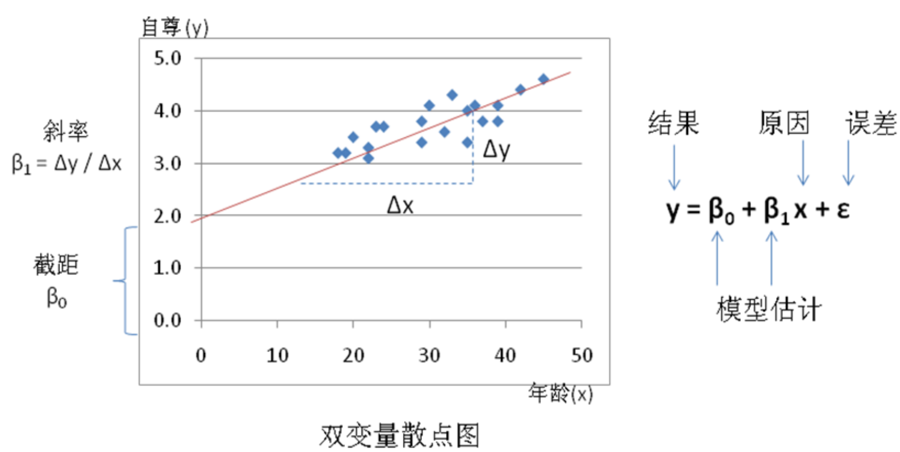


图 15.1 双变量线性模型

最简单的 GLM 是研究自变量（原因或者预测因素）与因变量（影响或者结果）之间关系的双变量线性模型。我们假设这两个变量分别是年龄和自尊，二者关系的二元散点图见图 15.1，水平线或者 X 轴表示年龄（预测因素），垂直线或者 Y 轴表示自尊（结果）。散点图显示，代表同时被标注出年龄和自尊的个体观测者的点在图中沿着一条虚构的向上倾斜的直线分布。我们可以通过 GLM 估计这条线的参数，比如它的斜率和截距。从高中的代数中我们知道，这条线可以用数学方程 $y = mx + c$ 来代表，其中 m 是这条直线的斜率（x 每变动一个单位，y 变动的值），c 是截距项（当 x 等于 0 的时候，y 的值）。在 GLM 中，这个方程正式地表示为：

$$y = \beta_0 + \beta_1 x + \varepsilon$$

其中 β_0 是斜率，β_1 是截距项，ε 是误差项。由于大部分观测值都很接近那条线但并不是刚刚落到那条线上（即 GLM 不是完美的），所以存在 ε 来表示实际观察值与它们的估计值之间的偏差。注意一个线性模型可以有两个以上的预测因素。为了形象地表示一个有两个预测因素的线性模型。我们一个想象一个三维的立方体，结果（y）用垂直轴表示，两个预测因素（比如，x_1 和 x_2）用立方体的两条水平轴表示。一条描述两个或者更多变量之间关系的线叫做回归线，β_0 和 β_1（以及其他的贝塔值）都被称为回归系数，而估计回归系数的过程被叫做回归分析。有 n 个预测因素的回归分析 GLM 见下：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_n x_n + \varepsilon$$

在上面的方程中，预测变量 x_i 可以代表自变量或者协变量（控制变量）。协变量是我们在理论上不感兴趣的变量，但是它们可能会对因变量 y 有影响，所以应该被控制，这样我们感兴趣的自变量的残余影响能够被更加精确地检测出来。协变量抓住了一个回归方程中的系统误差，而误差项（ε）则抓住了随机误差。虽然 GLM 中的大部分变量往往是等距或者等比变量，但也不必要总是这样。一些预测变量甚至可以是名义变量（例如，性别：男性或女性），他们被编码为虚拟变量。这些变量假定只能从两个可能的值当中取其中之一，这两个可能的值为 0 或 1（在性别的例子中，“男性”被指定为 0，“女性”被指定为 1；反之亦然）。n 个定类变量用 n-1 个虚拟变量来表示。比如，包括农业、制造业和服务业的产业类别，可以用两个联合的虚拟变量（x_1, x_2）来表示，（0,0）代表农业，（0,1）代表制造业，（1,1）代表服务业。我们不用关心赋予不用的类别多大的值，是 0 还是 1，因为 0 和 1 被认为是两个截然不同的组，而不是数字量（就像在一个实验设计中的实验组和对照组一样），而每一个组的统计参数都是分别估计的。

GLM 是一个非常强有力的统计工具，因为它不是一个单一的统计方法，而是一组可以被用来分析不同种类和大量预测变量与结果变量之间复杂关系的方法。如果我们有一个虚拟预测变量，同时我们在比较这个变量的两个取值（0 和 1）对结果变量的影响，那么我们在进行一个方差分析。如果我们在控制了一个或者多个协变量之后进行方差分析，那么我们在进行一个协方差分析。我们也可以有多个结果变量（比如，$y_1, y_1, \ldots y_n$），可以用一个方程组来表示，对于方程组中的每个结果变量都对应着一个不同的方程（每个方程都有自己的回归系数）。如果多个结果变量都被同一组预测变量来解释，这种分析被称为多变量回归。如果我

们使用多个结果变量来进行方差分析或者协方差分析，这种分析被称为多变量方差分析或者是多变量协方差分析。如果我们在一个相关联的方程组中使用一个回归方程的结果作为另外一个回归方程的预测因素，那么我们就得到一个非常复杂的类型的分析，被称为结构方程模型。在 GLM 中最为重要的问题就是模型设定，即怎么确定一个回归方程（或一个方程组）来最好地描述我们所关心的现象。模型设定应该基于对所研究的现象的理论思考，而不是什么模型能够最好地拟合观测数据。数据是用来验证模型的，而不是用来设定模型的。

双组比较

最简单的推理分析之一就是比较事后测试的实验组和对照组的结果，比如进入一项特别数学项目的学生是否比那些传统数学课程的学生表现得更好。在这个例子中，预测变量是一个虚拟变量（1=实验组，0=对照组），同时结果变量，即表现，是一个定比变量（比如，在特别项目之后的一个数学测试的分数）。这个简单设计的分析技术是单因素方差分析（单因素是因为它只包含了一个预测变量），统计检验被称为学生（Student）t 检验（或者简略地成为 t 检验）。

t 检验于 1908 年被一个在爱尔兰都柏林的吉尼斯啤酒厂工作的化学家威廉·西利·戈塞特用于控制烈性黑啤酒的质量，这种黑啤酒在 19 世纪伦敦的守门人中很受欢迎。因为他的雇主不想暴露他们使用了统计方法来控制质量的事实，戈塞特在《生物统计学》上使用自己的笔名"学生"发表了这个检验方法（他曾经是罗纳德·费舍尔爵士的学生），同时这个检验需要计算 t 值，t 这个字母经常被费舍尔用来表示两组之间的差异。因此，这个检验被称为学生 t 检验，虽然这个学生的身份对于统计学家而言都是熟知的。

t 检验研究两组的均值是否在统计上有差别（非定向或双尾检验），或者一组的均值在统计上比另外一组要大（或小）（定向或者单尾检验）。在我们的例子中，如果我们想检验在特殊数学课程中学习的学生是否比那些在传统课程中学习的学生表现更好，我们需要一个单尾检验。这个假设可以这样写：

$H_0: \mu_1 \leq \mu_2$ （原假设）

$H_1: \mu_1 > \mu_2$ （被择假设）

μ_1 表示参加了特别课程的学生的表现的均值（实验组），μ_2 表示参加了传统课程的学生的表现的均值（对照组）。注意，原假设永远是有等于号的那一个，所有的统计显著性检验都是为了拒绝原假设。

我们怎样能够用从不同总体中抽取的样本来推测出总体均值的差别呢？从图 15.2 中的假设的实验组和对照组分数的频数分布中我们可以看到，对照组似乎有一个平均分数为 45（在 0-100 的区间上）的钟形（正态）分布，而实验组看起来平均分数为 65。这两个均值看起来是有差异的，但是他们实际上是样本的均值（$\bar{X}$），由于抽样误差的存在，这可能与它们所对应的总体均值（μ）有差异。样本均值是在一定置信区间对总体参数的概率估计（95%的置信区

间是样本均值+两个标准误差，其中标准误是从总体中抽取出的无限个样本的均值的分布的标准差）。因此，总体均值的统计显著性不仅仅依赖于样本均值的数值，也依赖于标准误或者说样本均值频数分布的发散度。如果发散度很大（比如，这两条钟形曲线有很多重叠），那么这两个均值的 95%的置信区间也可能会重叠，这样我们就不能得出高可能性的（$p<0.05$）他们相对应的总体均值显著不同的结论。然而，如果这两条曲线的发散度较低（即，他们重叠的部分很少），那么每个均值的置信区间就可能不会重叠，这样我们就可以拒绝原假设，认为这两个总体均值在 $p<0.05$ 的水平下显著不同。

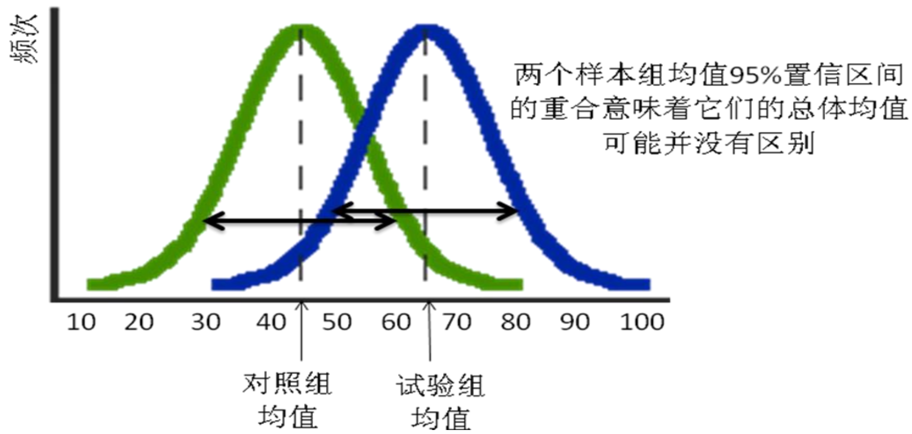


图 15.2 学生 t检验

为了进行 t 检验，我们必须首先计算一个 t 统计量，表示这两组样本均值的差别。这个统计量是两个样本均值的差与它们的分数之差的变异程度（标准误）的比值：

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_{\overline{X}_1 - \overline{X}_2}}$$

其中分子是实验组（组 1）和对照组（组 2）的样本均值之间的差异，分母是这两组之间差异的标准误，分母也可以这样来估计：

$$s_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

s^2 是方差，n 是每一组的样本规模。如果实验组的均值比对照组的均值大，t 统计量会为正。为了确定 t 统计量是否足够大而不是偶然地为正，我们必须从标准统计课本中的统计表中或者因特网上查阅与我们计算出的 t 统计量相关的概率或 p 值，或者使用统计软件程序比如 SAS 和 SPSS 来计算。这个值是一个关于 t统计量，t 检验是单尾还是双尾，和自由度（df）或者在计算这个统计量时能够自由变动的值的数量的函数（通常是一个样本容量和所进行的检验的类型的函数）。t 检验的自由度的计算方法如下：

$$\text{d.f.} = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1-1) + (s_2^2/n_2)^2/(n_2-1)}.$$

通常接近于（n_1+n_2-2）。如果 p 值比理想的显著性水平（例如 $\alpha=0.05$），或者得出有实验处理效应而实质上没有（I 类错误）的结论时我们愿意承担的最大风险（可能性）要小，我们就可以拒绝原假设。

在研究了实验组是否显著地比对照组有更高的均值之后，通常来说，下一个问题是效应规模或者是相对于对照组的试验处理效应的大小是多少？我们可以通过进行一个代表表现分数的结果变量（y）与一个代表是否是实验组的虚拟预测变量（x）的双变量一般线性回归来进行回归分析。实验变量的回归系数（β_1）同时也是回归线的斜率（$\beta_1 = \Delta y/\Delta x$），就是对效应规模的预测值。在上面的例子中，由于 x 是一个只有两个值（0 和 1）的虚拟变量，因此效应规模或者 β1 就是实验组和对照组均值的差异（$\Delta y = y_1 - y_2$）。

因子设计

沿用上面的例子，我们假设特别课程（实验）相对于传统课程（对照）的效应依赖于教学时间（每周 3 小时或者 6 小时）。现在，我们有了一个 2 x 2 的因子设计，其中两个因子分别是课程类型（特别还是传统）和教学时间（每周 3 小时还是 6 小时）。这样一个设计不仅仅帮助我们估计每一个因子的独立效应，称为主效应，也估计出两个因子的联合效应，称为交互效应。为这种双因素因子设计而设计的一般线性模型如下：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

其中 y 代表学生实验之后的表现分数，x_1 是课程类型（特别还是传统课程），x_2 是教学时间（每周 3 小时还是 6 小时）。注意 x_1 和 x_2 都是虚拟变量，虽然 x_2 看起来像是等比变量（3 或 6），它实际上代表了在因子设计中的两个不同的组。回归系数 β_1 和 β_2 提供了主效应的效应规模估计值，β_3 则提供了交互效应的效应规模估计值。此外，也可以使用一个双因素方差分析来进行同样的因子模型分析。有多个预测变量的回归分析有时候被称为多元回归，这与有多个结果变量的多变量回归是不同的。

以下是关于解读交互效应的注意事项。如果 β_3 是显著的，那么意味着课程类型对学生表现的影响有赖于教学时间。在这个例子中，我们就不能显著地解读出课程类型（β_1）或者教学时间（β_2）的独立效应，因为这两个效应不能被区分开来。只有当交互效应不显著的时候，主效应才能被解读出来。

协变量也可以连同新的回归系数（比如，β_4）一起被当作新的变量加入到因子设计中。协变量可以是定距或定性变量，即使我们关心的预测变量都是虚拟变量。协变量的解读与其他的预测变量的解读遵循同样的准则。

其他定量分析

有很多其他基于一般线性模型变形的有用的推断统计技巧，下面将简单地介绍一些。有兴趣的读者可以去更为高级的教科书或者是统计学课程中查阅更多关于这些技巧的信息。

- 因子分析是一种基于大量的观测到的观测值（项目）与较小的不可观测的（隐性的）被称为因子的变量组之间潜在的双变量相关模式，将前者转化为后者的数据缩减技术。这种技术在社会科学研究的多项目测量尺度中被广泛的用于评价收敛和判别的有效性。

- 判别分析是基于预测变量的线性组合将给定的观测值分类到一些名义的类别当中去的一种分类技术。除了因变量是分类变量，这种技术与多元回归很相似，在市场营销中的应用很广泛，比如说根据从大规模的问卷中识别出的突出属性将客户或者产品进行分类。

- Logistic 回归（或 logit 模型）是一种结果变量为二进制的（0 或 1）并且符合 logistic 分布的一种一般线性模型，这种回归分析的目的是通过将数据与一个 logistic 曲线进行拟合来预测某种结果发生的可能性。比如说基于预测因素比如年龄、身体质量指数、养生法等等来预测在一段时期内心脏病发作的可能性。Logistic 回归在医学研究中运用非常广泛。效应规模估计是基于 “胜算比”，这表示一件事情在一个组中相对于其他事情发生的胜算。

- Probit 回归（或 probit 模型）是一种结果变量在 0 和 1 之间变动（或者可以假定为 0 或 1），并且符合标准正态分布的一般线性模型。回归的目的在于预测各个结果出现的可能性。这种技术是在精算学、金融服务、保险和其他行业中非常流行的预测分析技术。被用于比如说根据一个人贷款申请上的信用等级、薪水、负债和其他信息来给他进行信用评分。Probit 和 logit 回归在比较应用（二进制结果）中倾向于得出相似的回归系数；然而 logit 模型更加容易计算和解释。

- 路径分析是一种用来分析一组变量的定向关系的一种多变量一般线性回归技术。它可以进行复杂的规律模型的研究，在这种模型中，一个方程中的因变量是另外一个方程的自变量。它被广泛地应用于当代的社会科学研究中。

- 时间序列分析是一种研究时间序列数据或者会随着时间逐渐改变的变量的技术。应用的例子包括预测股票市场的波动和城市犯罪率。这种技术在计量经济学、数理金融学和信号处理中很流行。特别的技术被用来修正自相关性，或者同一变量在时间趋势上的相关性。

第十六章

研究道德

在韦伯斯特字典中，道德定义为遵守相关职业或组织的行为标准。这些标准通过各个学科的一系列职业准则加以定义，并由机构审查委员会强制执行。即使没有明确的规定，科学家仍应了解并遵守学术界所达成的共同协议，诸如在专业科学研究领域哪些行为是允许的，哪些是禁止的。比如，科学家不应对数据的收集、分析和解释程序进行操纵，倘若这种操纵是与科学准则或科学方法相矛盾的，或者为了提前实现了研究者的个人计划。

为什么研究道德是重要的呢？这是因为一些个人或者组织为了提前实现其计划往往以不道德的方式对科学进行操纵，而这些行为与学术行为准则是相违背的。违反研究道德的典型案例是制药巨头默克公司关于万络（抗关节炎药）的药物试验。公司对科学界隐藏了这种药物致命的副作用，这导致了服用该药物的人中有 3468 人因心跳骤停而死亡。2010 年，公司同意赔偿 48.5 亿美元，与此同时，公司成立了两个独立的委员会并任命一名首席医疗官来监督医药开发过程中的安全性。默克公司的行为是不道德的，并且违背了与数据收集、分析和解释相关的科学准则。

道德要求从道德层面分清是非，但不道德的行为不一定是不合法的。如果科研人员的行为属于道德和法律之间的灰色地带，那么她在法律上可能并不会受到处罚，但是这种行为在其专业领域中将会被排斥，使其专业名誉受损，甚至可能由于渎职而失去工作。不同的社会群体可能会有不同的道德规范，我们在此介绍西方国家的科学研究中的所通行的道德标准。

科学研究中的道德准则

科学界所广泛接受的道德行为准则如下：

自愿参加和无恶意原则。研究项目中的被调查者应该认识到他们是自愿参与该项研究的，他们可以在任何时候退出该研究，并且对其自身不会带来任何不利的结果。被调查者不

会因为是否参加这个研究项目而受到任何伤害。最为恶名昭著的违反自愿原则的例子或许是纳粹科学家在第二次世界大战期间对囚犯强制实施的医学试验，这个事件记载于战后的纽伦堡审判中（“反人类罪”这一专业术语就来源于这些实验）。不太为众人所知的还有由美国公共卫生署在 1932-1972 年间实施的塔斯基吉梅毒实验。在这个实验中，400 多名贫困的非洲裔美国人虽身患梅毒，但是却没有得到青霉素的治疗，而是被迫接受脊椎穿刺的错误治疗方式。即使被试验者没有受到致命的威胁，他们也不应因为参与试验而遭受任何痛苦。1971 年，心理学家菲利普 • 赞巴尔度发明了斯坦福监狱实验，实验为斯坦福的学生随机分配囚犯或警卫的角色，并要求其扮演各自到的角色。当发现扮演囚犯的学生因这种模拟监禁过程而受到精神伤害，而扮演警卫的学生也产生与自我形象相反的虐待倾向时，这个实验就被终止了。

如今，若教员要求学生填写问卷，教员应告诉学生可自愿决定是否参与，并且学生不用担心他们由于没有参加实验而影响到其在班级中的成绩。比如，对参与实验的学生提供奖励分而对不参与实验的学生不提供奖励分，这种做法是不道德的，因为这使得不参与实验的学生处于明显的劣势。为了避免这种情况的发生，教员可为不参与实验的学生提供可替代的任务，这样学生就可弥补其因为没有参与这项研究所损失的奖励分。或者，教员也可以为每个学生都提供奖励分，而不管其是否参与研究实验。此外，所有的参与者都必须接受并签订知情同意书，并明确规定参与者在回答被记录之前，有权决定不参与研究或退出此项研究。在医学研究中，知情同意书必须详细说明参与主体可能承担的所有风险。对于 18 岁以下的参与主体，知情同意书应由其父母或其法定监护人签署。为了遵照学科或工作环境下的科学行为规范，研究人员应在数据收集程序结束后的一段时间内（通常是三年）保存知情同意书。

匿名原则和保密性原则。在科学研究中，为了保护被调查者的利益，研究人员必须保护他们的身份。这个可通过匿名和保密双重原则来实现。匿名原则是指研究人员或研究报告和论文的读者不能识别出某个特定被调查者的回答。科学研究中运用匿名原则的例子是在邮件问卷中不使用可找出特定被调查主体的识别号。在调查离经叛道或非正常的行为时，比如吸食毒品和非法下载音乐等，若不能承诺匿名调查，被调查者将可能不会提供真实的回答。更进一步而言，匿名原则必须对被调查者与侦查并追踪这些非法行为主体的法律监管部门或其他机构进行隔绝。

在一些研究设计中，如面对面访谈，匿名原则是无法保证的。在其他研究设计中，如纵向维度的实地调查研究，由于需要对同一主体在不同时间点的回答进行匹配并进行纵向分析，因此匿名原则也是不可行的。在这些情形下，研究人员必须对被调查者的信息进行保密，即研究人员虽然可识别被调查主体对应的回答，但是应保证不在任何报告、论文或公共论坛中泄露被调查者的身份。相对于匿名原则而言，对被调查主体的信息进行保密是较弱的一种保护方式，这是因为社会研究数据并不能如与牧师或律师之间的交流那样在美国法院中享有特许交流权。比如，在埃克森•瓦尔迪兹油轮在阿拉斯加的瓦尔迪兹港口附近泄漏了一千万桶原油的两年后，经济和环境受到侵害的社区委任圣地亚哥的一家研究公司调查被波及的家庭中所存在的心理问题，及与之相关的各种私人且尴尬的细节。因为美国原住民的文化规范会让公开披露这类信息的居民变得痛苦且困难的，所以被调查者之前被承诺他们的回答将严格保

密。当证据被提交至法庭时，埃克森要求法庭呈示原始的调查问卷（及身份识别信息）以盘问被调查者之前的回答，而这些回答是在保密的承诺下提供给调查研究公司的，这个要求最终获得法庭的许可。幸运的是，埃克森•瓦尔迪兹油轮泄漏事件在受害者被强制要求进行公开作证之前得到了圆满的解决，但是，类似的与保密原则相违背的行为仍然存在。

更极端的例子是华盛顿大学研究生里克・斯卡尔斯的故事。斯卡尔斯进行了动物权利激进分子的参与式观察研究，并将观察结果记录于 1990 年发表的一本名为《Ecowarriors：Understanding the Radical Environmental Movement》的书上。1993 年，斯卡尔斯被陪审团传唤，并要求指出他所研究的动物权利保护组织中的活跃分子。这名研究人员拒绝回答陪审团的问题，并坚守他作为美国社会学协会成员所应该遵守的道德责任。最终，这名学者在斯波坎县监狱被强制拘留 159 天。为了使自身免于遭受类似于里克・斯卡尔斯的痛苦经历，研究人员应在不需要相关文件和数据资料时立即删除其中的身份识别信息。2002 年，美国卫生与公共服务部颁布了保密证明书来保护研究项目的参与者免于受到警方或其他监管部门的调查。并不是所有的研究项目都能得到这种保护，然而在许多情形下，这为保护参与者隐私提供了重要的支持。

披露原则。通常，研究人员有义务向被调查主体提供有关该研究的信息来帮助其决定是否参与这项研究。比如，这项研究由谁来实施，实施目的和预期研究结果是什么，谁将受益于这一研究等方面的信息。然而，在一些情形下，披露这类信息会带来被调查者的回答偏差。例如，如果研究目的在于检验被调查者在多大程度上会放弃个人观点，而遵照集体的想法，那么在此实验中，被调查者在发表个人见解之前，会听取其他人的观点。若在实验之前公布了研究目的，这将使得被调查主体对实验处理过程变得敏感。在这种情形下，即使在实验前并没有披露研究的目的，研究人员也应该在数据收集结束之后立即对情况进行简要的说明，并列出参与实验的潜在风险或伤害。

分析和报告。对学术界而言，研究人员也有汇报数据是如何进行分析和报告的道德责任。未预期的或负面的调查结果也应该予以充分地披露，即使这会导致人们质疑研究设计或调查结果。类似地，在研究完成之后由于偶然的机会或者通过数据挖掘等方法也会发现许多有趣的关系。而将调查结果作为故意设计的产物是不道德的，也就是说，在实证研究中不能基于数据分析的结果提出假设，因为数据是用来检验假设的，而不是用来建立假设的。把数据分割成不同的部分来对相关假设进行证实或证伪，或者宣称使用不同的数据集来撰写多篇论文都是不道德的。基于部分的、不完整的或不合适的数据分析，将可疑的观点错误地解释为真实有效的做法同样是不诚实的。只有做到公开诚实，科学才能进步。研究人员只有充分地披露研究中的问题才能更好地为服务于科学，也只有这样，他们才能帮助其他研究人员解决类似的问题。

机构审查委员会

在美国，联邦法律会对以人作为参与主体的研究中的道德问题进行管制。任何代理机构，比如高校或者医院，只要其想申请联邦政府研究基金来支持研究项目，就必须确定这一研究过程不能违背联邦法律中对人这一参与主体的权利保护的相关规定。研究过程应由机构审查委员会（IRB）的专家小组进行监督。机构审查委员会对所有涉及到以人作为参与主体的研究申请文件进行审查，保证不违背自愿参加，无恶意性，匿名性和保密性等原则，并将其对人的威胁降到最低程度。虽然联邦法律适用于联邦政府基金支持的科研项目的审核，但是这些标准和审核程序同样也适用于非联邦政府基金支持的科研项目，甚至于一些学生项目。

在机构审查委员会的审批流程中，研究人员需要提交结构化的申请，并提供关于研究项目和项目主要成员的完整信息，以及对如何保护被调查者权利的细节进行描述。还有一些附加性文件，如知情同意书、研究问卷或采访协议等也可能需要提供。研究人员可以提供他们参加研究道德课程培训的证明来表示其非常熟悉研究中所涉及的道德问题。在研究项目被机构审查委员会批准后，研究人员才可以开始收集数据。

专业道德准则

许多专业协会确立并发布了正式的道德行为准则，以明确其会员的哪些行为是可以接受的，而哪些行为是不可接受的。比如，作为全球信息系统学科的专业协会，国际信息系统协会（AIS）的行为准则摘要如表 16.1 所示（完整的行为准则可具体参见以下链接：http://home.aisnet.org/displaycommon.cfm?an=1&subarticlenbr=15），在其他学科中也有类似的行为准则。

信息系统协会的行为准则将违反研究道德的行为划分为两大类。第一类包括严重的违反研究道德的行为，比如剽窃、数据伪造、研究过程或数据分析中的造假等，这些行为可能会导致其被协会开除，被雇佣单位解除职务，甚或被提起法律诉讼，并将对其专业名誉造成巨大的损害。第二类包括程度相对较轻的违反研究道德的行为，比如不尊重被调查者的权利、虚假陈述研究项目的创新点、或在未获得他人许可的情况下使用其他人已发表的数据。这些行为将导致专业名誉受损，以及受到学术期刊的惩戒等。这些准则同样为好的研究提供了指导，指出在发现违反研究道德的行为时应该怎么做（对于违反者和受害者而言），以及信息系统协会处理违反道德准则行为的程序等。尽管这些道德准则并不能完全消除不道德的行为，但是其有助于明确学术道德行为的界限，并减少了违背学术道德的行为发生。

类别 1： 必须永远遵守这类行为准则，对该准则的忽视将构成严重违背学术道德的行为，并会导致你被学术协会开除，被雇主解除职务，被提起法律诉讼等，并将对你的学术声誉带来严重的损害。

1．不准剽窃。 2．不准对数据、研究过程或者数据分析进行造假。
类别 2： 这类行为准则是我们所推荐的学术道德行为。公然漠视这些准则或者其他的专业规范，即使是较低程度的，也会导致你的学术声誉受损，不仅会受到期刊编辑的制裁，更将给职业生涯带来难堪，并会引起法律诉讼和同行的厌恶。 3．尊重受调查主体的权利，特别是保密权和知情权，即被调查者有权要求对其私人信息进行保密，并有权知晓研究的性质和即将参与的研究类型。 4．递交给杂志编辑和会议项目主席的论文，不允许对论文创新点进行虚假描述。 5．不滥用自身作为编辑、审稿人或主管所拥有的权力，并保证私人关系对你的评判不会造成干扰。 6．在评阅提交的论文、项目申请、软件或承担外部工作时，对任何影响自身客观公正的利益冲突必须做出说明。 7．不准在使用已公开发表的数据时不标明出处，不准在使用未公开发表的数据时未获他人允许和不标明出处。 8．无论参与者是同行或学生，都应该根据其在研究中所做出的实质性贡献做出感谢。 9．在未获得原作者许可的情况下，不准利用其未发表的作品、信息、观点、概念或数据等。 10．根据来源档案的规则适当地使用档案资料。
建议：以下是关于怎样使你免受原创作者的质疑，避免采取错误的步骤，避免犯错甚至免于引起法律诉讼的建议： 1．保存能证明学术著作权的文档和数据。 2．不要重复发表你的旧观点以形成自己做出新贡献的假象。 3．在数据整理之前解决数据所有权的问题。 4．如果仍有疑问，可咨询合适的同行人士。

表 16.1. 国际信息系统协会道德准则

道德争论

罗伯特·艾伦·汉弗莱斯是美国的社会学家，他因 1970 年发表的题为《茶室交易》的博士论文而闻名。这本书是关于公园公厕（在美国俚语中也叫作茶室）里男同性恋的人种志研究。汉弗莱斯发现大多数男同性恋者表面上都是正常的异性恋者，他们和周围的群体一样过着传统的正常生活，这一现象引起了他极大的兴趣。然而，在进行同性恋调查时，保持被调查者的匿名性是非常重要的。

典型的茶室邂逅包括三个人——两个进行性行为的男同性恋者和一个看客。看客的任务是当有警察或其他人接近他们的时候，通知这两个同性恋者，但是看客会从这种偷窥行为中感到愉悦。因为没有其他途径可以接近这些同性恋者，所以汉弗莱斯出现在公厕里，并伪装成看客来进行研究。作为对这些参与对象的观察者，和政治抗议和其他社会现象的研究一样，汉弗莱斯能够为他的论文实施现场观察。

汉弗莱斯需要更多的关于这些同性恋者的信息。但是由于这些同性恋者不愿意被采访或泄露其个人身份，汉弗莱斯记下了他们的车牌号码，并且无论在何处，侦查他们的姓名，进而从公共数据库中提取相应信息。接下来，他对这些人进行上门拜访，精心伪装自己以防被识别出来，并宣称自己正在实施一项研究，进而搜集到其他方式无法获得的私人信息。

汉弗莱斯的研究在学术界引起了很大的争论。许多批评者认为他不应该以科学研究的名义侵犯他人的隐私权，还有一些人对他这种误导同性恋性行为的参与者，并使他们相信自己只是看客的欺骗行为表示不满，因为他明显是别有用心的。另一些学者认为对茶室交易的观察是可以接受的（因为这些同性恋者是在公共设施中进行性行为），但是随后以不正当的方式获取同性恋者的家庭地址，且在未获得受访者同意的情况下以虚假的借口对同性恋者进行上门调查是不道德的。少数的研究人员认为汉弗莱斯的方法是合乎情理的，因为同性恋是非常值得研究的社会现象，但是却没有获得这方面数据的其他方法，并且调查过程中的欺骗行为并没有任何恶意，因为汉弗莱斯没有对任何人泄露被调查者的身份。这个争议一直悬而未决，而且一直在关于研究道德的课堂和论坛上被予以激烈地讨论。

附录

博士课程大纲

QMB 7565：研究方法介绍
南佛罗里达大学
商学院
2011 秋季学期

目标：

该课程的目的在于向博士研究生介绍从事学术研究的程序与方法。我们将学习如何在科研项目的概化、设计、执行和评估等方面像研究者一样进行思考。该课程的部分内容需要你在导师的指导下设计和撰写独立研究报告。除此之外，期末考试有助于你准备博士资格考试中的研究方法部分。

结构：

该课程以研讨会的方式展开。博士研讨会的核心在于对指定阅读材料的讨论和分析。为了实现这一目标，你必须在课程之前阅读所有的指定论文，在研讨的过程中认真思考相关的议题，并与你的同学一起讨论和总结这些议题。请注意：如果你不能在上课之前做好充分地准备，你将无法从该课程中学到任何东西。

作为一个博士研讨会，该课程将包含以下内容：（1）比你迄今为止所上过的任何硕士课程有着更大的工作量；（2）大量的令人厌烦的阅读材料；（3）大量的批判性（甚或令人沮丧）的思考。这不是一门简单的课程，并且你将不能轻松地获得高分。如果你认为该课程学起来比较吃力，那么你应该退出该课程以及你所在的博士项目。

书籍和资料：

Bhattacherjee, A.*Social Science Research: Principles, Methods, and Practice*, Ver 2.0, 2011, Free download from Blackboard (my.usf.edu).

Kuhn, T. J. *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago, 1996.

分数：

分数组成:	期末考试	35 分
	研究报告	35 分
	论文评审	10 分
	课堂参与	20 分

分数区间：A+: 97-100; A: 92-97; A-: 90-92; B+: 88-90; B: 82-88; B-: 80-82; C+: 78-80; C: 70-78

期末考试：

这是一个持续 3 个小时的考试，类似于你的博士资格考试，期末考试将要求你对 3-4 个方面的论文类问题进行作答。你可以从教学平台上下载到考试样题。考试中你可以手写答案或将答案输入电脑。你的回答应阐明：（a）对课堂中所覆盖的材料进行归纳总结；（b）你对这些材料的批判性分析，而不仅仅是重复论文或重复教授的评论。你的分数完全取决于回答的质量，而不是取决于你写了多少页。如果你的回答是不相关的，没有包含足够的细节，或者仅仅表达了对课堂讨论材料的肤浅理解，那么你将会因此而失分。考试采用开卷的方式，但是需要告知的是，如果你并没有提前准备这次考试，那么即使把所有的材料放在你的面前对你而言用处也并不大。考虑到课堂上所提及的大量话题和材料，你在考试前一周才开始准备的话，时间将是来不及的。

研究报告：

在学期结束之前，你需要提交一份研究报告。该报告应有 10-12 页（单倍行距，不包含参考文献和附录）。你可选择任何研究问题，但是最好采用实证研究的方式。你的研究必须具有原创性（不能是你为另一门课或为其他教授所准备的材料），必须检验一个真实的问题（不能是假设问题或游戏问题），并且必须至少达到会议论文的要求。你的报告应包含五个部分：（1）研究问题及其重要性；（2）文献回顾；（3）理论和假设；（4）研究方法；和（5）研究计划。数据收集和分析并非必须的。但如果你打算完成一个解释性的研究项目，那么数据收集和分析则是必要的。当我们学完相应的阶段性课程，你可以在学期中的任一时刻提交论文的章节，以便于我及时地给予反馈意见和修正建议。需要注意的是，学期中所提交的论文章节并不记分，只有你最后的研究报告才被纳入总成绩中。你应在课程的最后一周以 15-20 分钟正式演讲的方式（以及另外的 5 分钟问答环节）介绍你的研究报告，正如同你在学术会议上所做的那样。

论文评审：

学术生涯的一个关键组成部分是批判性地评审他人的研究。在该学期的后半阶段，你将要对提交到顶级商业期刊上的论文进行评审。为了帮助你准备这件事，我将在课堂上讲解如何撰写评审，并为你提供一个论文评审的框架，以及让你在课堂上完成一次实际的评审活动。在你写完评审之后，我将把来自匿名审稿人和期刊编辑的意见发给你，以帮助你评价你和专业评审人之间的差距，并找出所欠缺的地方。论文评审作为课后作业，应在期末考试前一周提交。在撰写论文评审报告之前，你需要做一些背景研究，或阅读一些其他的文章，但是你不可以和其他任何人进行讨论。

课堂参与：

课堂上所讨论的每篇论文都会指定一名引导人，他的职责包括：（1）以结构化的方式准备一页的论文概要，（2）引导课堂讨论。如果你可以按时上交所有的的论文概要，并在课堂讨论中适度地参与进来，你将获得课堂参与的满分。请查阅论文概要的样本以知晓如何撰写结构化概要。如果你非常细致地完成了这些工作，那么概要对于考试而言也是一个非常有用的辅助资料。我将在课堂上对你的概要进行即时点评。这样你可以在下次撰写概要的时候有所提高。如果你想写其他文章的概要，你可以与你的同学交换各自分配到的论文。然而若在沟通上有误差，并且在课堂讨论时没有任何摘要，那么最初分配到这篇文章的学生将在他的课堂参与成绩上被扣分。除了概要撰写的任务之外，每个人都应参与所有的课程，阅读所有的论文，并为所有的课堂讨论做出贡献。

课堂纪律：

参与：我将会保留一份关于出勤情况的记录。如果你认为你将缺席超过一周的课程，你应该退出这门课。

学术诚实：任何形式的剽窃都是被严格禁止的，并将直接导致一个最低分。请参照你们的学生手册上所列出的南佛罗里达大学学术诚实政策。

残疾：需要残疾服务设备的学生应在最初的两次课中由学生残疾服务办公室发邮件通知我。

手机：在上课时间，手机必须关机，但是在课间休息的时候你可以使用它们。

课程安排：

第一周：研究介绍

课程大纲和课程介绍
第一章
研究和发表流程

第二周：像研究者一样思考

第二章.
两篇论文的故事:

- Fichman, R.G. and Kemerer, C.F., "The Illusory Diffusion of Innovation: An Examination of Assimilation Gaps," Information Systems Research (10:3), September 1999, pp. 255-275.
- Williams, L.; Kessler, R.R.; Cunningham, W.; and Jeffries, R., "Strengthening the Case for Pair Programming," IEEE Software, July/August 2000, pp. 19-25.

批判性思考:

- Dialog on Leadership, "Awareness is the First Critical Thing," A Conversation with Wanda Orlikowski, 1999.

研究道德:

- AIS Code of Conduct: http://home.aisnet.org/displaycommon.cfm?an=1&subarticlenbr=13
- IRB Process: http://www.research.usf.edu/cs/irb_forms.htm

第三周：文献回顾和元分析

文献回顾:

- Fichman, R.G., "Information Technology Diffusion: A Review of Empirical Research," Proceedings of the Thirteenth International Conference on Information Systems, Dallas, 1992, 195-206.
- Alavi, M. and Leidner, D.E., "Knowledge Management and Knowledge Management Systems," MIS Quarterly (25:2), March 2001, pp. 107-136.

元分析:

- King, W.R, and He, J., "Understanding the Role and Methods of Meta-Analysis in IS Research," Communications of the AIS (16), 2005, pp. 665.686.
- Henard, D.H. and Szymanski, D.M., "Why Some New Products are More Successful Than Others," Journal of Marketing Research (38), August 2001, pp. 362-375.

第四周：科学哲学

第三章
科学研究中的范式:

- Kuhn, T., The Structure of Scientific Revolutions, University of Chicago Press, 1996 (entire book). [See Dr. Pajares' notes ONLY if you completely lost]

社会科学范式:

- Krugman, P., "How Did Economists Get It So Wrong," New York Times, Sept 6, 2009.
- Gioia, D.A. and Pitre, E., "Multiparadigm Perspectives on Theory Building," Academy of Management Review (15:4), 1990, pp. 584-602.

第五周：组织研究中的理论

第四章
为什么需要理论:

- Steinfield, C.W. and Fulk, J., "The Theory Imperative," in Organizations and Communications Technology, Janet Fulk and Charles W. Steinfield (eds.), Sage Publications, Newbury Park, CA, 1990.

评估组织理论：

- Bacharach, S.B., "Organizational Theories: Some Criteria for Evaluation," Academy of Management Review (14:4), 1989, pp. 496-515.

代理理论：

- Eisenhardt, K.M., "Agency Theory: An Assessment and Review," Academy of Management Review (14:1), 1989, pp. 57-74.

交易费用理论：

- Williamson, O.E., "The Economics of Organization: The Transaction Cost Approach," American Journal of Sociology (87:3), 1981, pp. 548-577.

第六周：组织研究中的理论（续）

交易费用理论：

- Rindfleisch, A. and Heide, J.B., "Transaction Cost Analysis: Past, Present, and Future Applications," Journal of Marketing (61:4), 1997, pp. 30-54.

资源基础理论：

- Barney, J.B., "Firm Resources and Sustained Competitive Advantage," Journal of Management (17:1), 1991, pp. 99-120.
- Priem, R.L. and Butler, J.E., "Is the Resource-Based 'View' a Useful Perspective for Strategic Management Research?", Academy of Management Review (26:1), 2001, pp. 22-40.

动态能力理论：

- Teece, D.J.; Pisano, G.; and Shuen, A., "Dynamic Capabilities and Strategic Management," Strategic Management Journal (18:7), 1997, 509-533.

第七周：测量和效度

第六、七章

量表效度和单维性：

- Straub, D.W., "Validating Instruments in MIS Research," MIS Quarterly (13:2), June 1989, pp. 146-169.
- MacKenzie, S. B., Podsakoff, P. M., and Podsakoff, N. P., "Construct Measurement and Validation Procedures in MIS and Behavioral Research: Integrating New and Existing Techniques," MIS Quarterly (35:2), 2011, pp. 293-334.

第八周：调查研究

第五、八、九章

实地调查范例：

- Tsai, W., "Knowledge Transfer in Intraorganizational Networks: Effects of Network Position and Absorptive Capacity on Business Unit Innovation and Performance," Academy of Management Review, 2001.

调查研究偏差:

- Malhotra, N. K, Kim, S. S., and Patil, A., "Common Method Variance in IS Research: A Comparison of Alternative Approaches and a Reanalysis of Past Research," Management Science (52:12), 2006, pp. 1865-1883.

第九周：实验和准实验研究

第十章

实地实验范例:

- Hunton, J.E. and McEwen, R.A., "An Assessment of the Relation Between Analysts' Earnings Forecast Accuracy, Motivational Incentives, and Cognitive Information Search Strategy," The Accounting Review (72:4), October 1997, pp. 497-515.

实验研究的问题:

- Jarvenpaa, S. L.; Dickson, G. W.; and DeSanctis, G., "Methodological Issues in Experimental IS Research: Experiences and Recommendations," MIS Quarterly, June 1985, pp. 141-156.

第十周：研究评审

撰写论文评审:

- Lee, A.S., "Reviewing a Manuscript for Publication," Journal of Operations Management (13:1), July 1995, pp.87-92.
- Agarwal, R; Echambadi, R; Franco, A.M.; and Sarkar, M.B., "Reap Rewards: Maximizing Benefits from Reviewer Comments," Academy of Management Journal (49:2), 2006, pp. 191-196.

论文被拒的原因:

- Daft, R.L., "Why I Recommended that Your Manuscript be Rejected and What You Can Do About It," in L.L. Cummings & P.J. Frost (eds.), Publishing in the Organizational Sciences, 2nd ed., 1995, pp. 164-182.

正式评审:

- 根据上面的文章中所提供的指南，为以下论文撰写一篇正式的评审意见。你可以使用模板来结构化你的评审。来自匿名审稿人、副主编和主编的真实评审意见将会在你完成论文评审后发给你。

论文评审:

- 你所要评审的论文应在第十三周提交，这是你课程成绩的一部分，因此请为论文评审计划好足够的时间和精力。

第十一周：案例研究

第十一章

案例研究:

- Benbasat, I.; Goldstein, D.K.; and Mead, M., "The Case Research Strategy in Studies of Information Systems," MIS Quarterly, September 1987, pp. 369-386.

案例研究范例:

- Beaudry, A. and Pinsonneault, A., "Understanding User Responses to Information Technology: A Coping Model of User Adaptation," MIS Quarterly (29:3), September 2005, pp. 493-524.
- Eisenhardt, K.M., "Making Fast Strategic Decisions In High-Velocity Environments," Academy of Management Journal (32:3), 1989, pp. 543-577.

实证主义与解释主义分析：

- Trauth, E.M. and Jessup, L.M., "Understanding Computer-Mediated Discussions: Positivist and Interpretive Analyses of Group Support System Use," MIS Quarterly (24:1), March 2000, pp. 43-79.

第十二周：解释性研究

第十二、十三章

演示：运用扎根理论进行内容分析

定性研究：

- Shah, S.K. and Corley, K.G., "Building Better Theory by Bridging the Quantitative-Qualitative Divide," Journal of Management Studies (48:3), December 2006, pp. 1821-1835.

行动研究：

- Kohli, R. and Kettinger, W., "Informating the Clan: Controlling Physician Costs and Outcomes," MIS Quarterly (28:3), September 2004, pp. 1-32.

人种论：

- Barley, S.R., "Technicians in the Workplace: Ethnographic Evidence for Bringing Work into Organization Studies," Administrative Science Quarterly (41), 1996, pp. 404-411.

第十三周：混合研究方法

第十六章

演示：用 SPSS 进行统计分析

二手数据分析：

- Chaney, P.K. and Philipich, K.L., "Shredded Reputation: The Cost of Audit Failure," Journal of Accounting Research (40:4), September 2002, pp. 1221-1245.

分析模型：

- Bayus, B.L., Jain, S., and Rao, A.G, "Truth or Consequences: An Analysis of Vaporware and New Product Announcements," Journal of Marketing Research (38), February 2001, pp. 3-13.

第十四周：期末考试

3 小时开卷考试。

第十五周：学生讲演

20 分钟的讲演和 10 分钟的问答环节

关于此书

该书用于向硕博研究生介绍在社会科学、商学、教育学、公共卫生及其他相关领域从事科学研究的流程。本书为行为研究中的基本概念提供了一站式、全面而简洁的资料。此书也可以作为任何博士研讨会和研究方法课程的独立教材或辅助阅读材料。

关于作者

阿诺•巴塔查尔吉是美国南佛罗里达大学信息系统系终身教授。其论文发表在国际顶级期刊，如 MIS Quarterly 和 Information Systems Research 上，他也是信息系统领域世界范围内的十名顶尖学者之一（2000-2009 年的十年总排名列第七）。在他十五年的研究生涯中，巴塔查尔吉博士发表了 2 本书和 50 余篇论文，其论文在谷歌学术上的被引总数超过 3000 余次。他目前是 MIS Quarterly 的编委，并经常受邀在全球各地的大学和会议上进行讲演。巴塔查尔吉博士从美国休斯顿大学获得博士学位和 MBA 学位，并从印度理工学院获得学士学位。

关于全球教材计划

全球教材计划是一项非盈利性的创举，并以在全球范围内发布免费的高质量电子教材为宗旨。自 2004 年 1 月成立以来，该项目总部一直位于美国乔治亚大学，并受到瑞士雅各布斯基金会的资助。