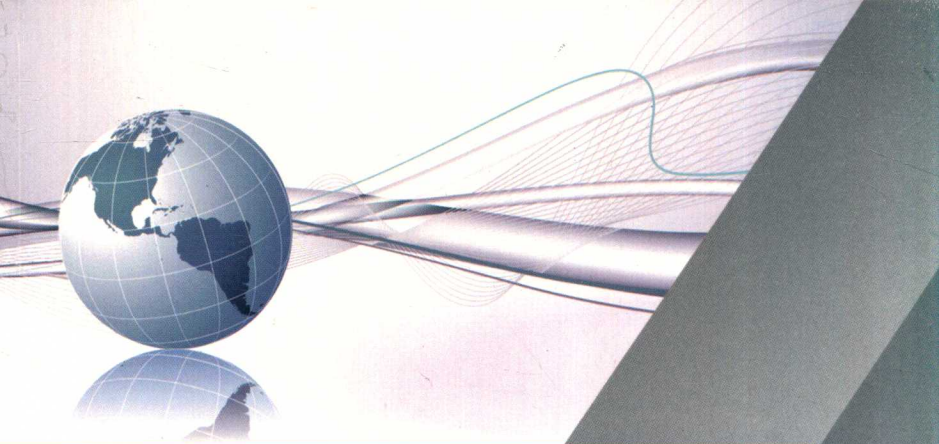




# 社会科学定量研究

## 原理、方法与应用

Quantitative Research in the Social Sciences:  
Principles, Methods, and Practices

陈 健·著

大学出版社  
ST UNIVERSITY PRESS

# 社会科学定量研究

## 原理、方法与应用

陈 健 著



东南大学出版社

SOUTHEAST UNIVERSITY PRESS

· 南京 ·

## 图书在版编目(CIP)数据

社会科学定量研究:原理、方法与应用 / 陈健著.

—南京:东南大学出版社, 2018. 8

ISBN 978-7-5641-7596-2

I. ①社… II. ①陈… III. ①社会科学—定量方法  
IV. ①C3

中国版本图书馆 CIP 数据核字(2017)第 323901 号

社会科学定量研究:原理、方法与应用

SheHui KeXue DingLiang YanJiu: YuanLi, FangFa Yu YingYong

著 者 陈 健

出版发行 东南大学出版社

社 址 南京市四牌楼 2 号 邮编:210096

出 版 人 江建中

网 址 <http://www.seupress.com>

电子邮箱 [press@seupress.com](mailto:press@seupress.com)

经 销 全国各地新华书店

印 刷 南京玉河印刷厂

版 次 2018 年 8 月第 1 版

印 次 2018 年 8 月第 1 次印刷

开 本 850 mm×1 168 mm 1/32

印 张 7.875

字 数 227 千字

书 号 ISBN 978-7-5641-7596-2

定 价 38.00 元

本社图书若有印装质量问题,请直接与营销部联系。电话(传真):025-83791830

## 前 言

贝利(1986)曾这样评述韦伯(Max Weber, 1864—1920)的方法论:社会科学“使用自然科学所遵循的科学方法……是合理的,但仅此是不够的”;社会科学家与他研究的现象之间的关系,为直接理解或投入理解(verstehen)开辟了可能性,这“在自然科学中是不可能的”。韦伯的这些主张无疑是深刻而富于启发意义的,换句话说,社会科学要有(可以有)比自然科学“更多的东西”。如果我们考察社会科学不得不面对的个体“异质性”(heterogeneity),这种“更多的东西”立刻就是必不可少的。谢宇(2006)曾以两个简单而深刻的例子说明了个体“同质性”(homogeneity)假定在自然科学中是允许的(事实上是必要的),但在社会科学中则未必成立:研究温度对灯泡寿命的影响,只要研究两个不同温度下的灯泡即可,因为假定灯泡都是一样的没有问题,但研究读大学对收入的影响,假定所有人(个体)都是没有差异的(差异仅在读不读大学),毫无疑问是不成立的。——社会科学中差异是本质,这就是“变异性原理”(variability principle)。因此对于社会科学来说,要得到一般结论(例如即使是对于一个特定的总体),仅研究个体无济于事。“相同的”的灯泡任取两个,在控制其他条件不变的情况下观察处于不同温度的两个灯泡的寿命,并根据观察结果检验温度如何影响灯泡寿命的有关“理论”(命题或假设),是受控实验的一个最为简单的例子。查尔默斯(Chalmers, 1999)在一本关于科学性质的入门书中曾仔细论证了实验的性质以及实验在科学中的作用。他

指出:“如果有构成科学基础的事实,那么那些事实表现为实验结果”(查尔默斯,2002)<sup>51</sup>——但是对社会科学来说就不是这么回事了,因为受控实验对于社会科学来说远不是那么现实。

定量社会科学也遵从某种“科学程序”(人们把它称为定量范式<sup>①</sup>),以便根据观察结果来检验理论(假设)——这些“观察”主要基于调查(survey)。社会科学家广泛使用调查数据(也被称为观察数据或观测数据)来检验理论,这在本质上与自然科学(指古典科学)不同,这种“检验”是概率意义上的,它借助于统计分析技术。(定量)社会科学在什么意义上是科学?本书不能讨论此类社会科学的科学性质问题(这属于另一门学问),但从这点出发,我们关注调查数据的“客观”性质和概率性质,关注分析调查数据的统计逻辑。和自然科学中实验需要“实践上的努力”(practical struggle)(查尔默斯,2002)<sup>53</sup>一样,社会科学中作为观察的调查(survey),也是一种实践上的努力,本书尝试描述并讨论社会科学家的这种努力(它们遵循定量范式):如何设计(定义)变量的测量以得到测量工具(问卷),如何设计概率抽样以得到随机样本,如何设计数据收集模式以测量样本单位——有了样本数据之后——如何进行统计分析以检验研究假设。本书上篇中,“变量测量”解决“用什么测量”的问题,这部分的讨论基于人类交流本质对测量“客观”性质的影响。“概率抽样”要解决的是“测量谁”的问题,基本概率抽样方法的讨论主要基于样本概率性质的保持。“数据收集”解决“怎么去测量”的问题,数据收集模式的比较讨论,基于“测量过程”对调查数据的影响。本书下篇的目的是讨论经典统计分析技术中最为

① 在库恩(Kuhn, 1970)“范式”(paradigms)概念的意义下,定量范式这个说法描述了定量社会科学在方法上的一个侧面:“基于一定主张的一套社会科学家共同遵循的做法”,它们为社会科学家群体的研究活动提供规则和指导。范式还意味着,需要在其中学习、从事研究活动,方得门径。

基础的回归分析,依本书作者看来,回归分析最直截了当地表明了“概率意义上的检验”。回归分析“其他条件不变”的解释,还有利于将统计控制与受控实验的实验控制相对照。讨论回归分析,“绕不开”数学推导,但本书不假定读者已经具备概率论、数理统计学、线性代数等知识,而只需有些初等集合论和微积分的基本知识。本书下篇中“概率基础”“统计推断”“简单回归”和“多元回归”是循序渐进的,这样安排与本书阐述定量研究方法原理的“最小系统”这一目的相一致。这样做还带来一个额外好处,就是它对没有这些数学基础的读者是有益的。出于这个目的,本书也不基于线性代数阐述回归分析方法,这个做法得益于J. M. 伍德里奇(伍德里奇,2010)。

有关“数学推导”,对没有数学基础的读者(例如初学者、感兴趣于或希望转到定量研究的文科背景的研究者等)还要说上几句:必要的数学推导是不应避免的,因为在本书作者看来,没有什么语言比数学语言更严格、更清晰明了的,例如不用数学,“概率分布”就说不清;不推导抽样方差公式,要搞清“同方差假定”就十分困难。“数学推导”经常给人造成这些数学很难的“错觉”,因而“望而生畏”,但实情或许是:所需要的数学“不是很难而是较多”。如果确实如此,我希望本书下篇的处理(解决了“数学太多”)有助于解决这个问题。

我要感谢多年来我教过的学生们,当我和他们一起把教学要求推进到“做一项定量研究”时,他们的“定量概念”很快让我明白传统课程教学的弊端:上了不少定量研究相关课程,但知识的“零碎”积累对规范地做一项研究帮助不大。这是促使我写这本书的最初起因。在写作本书的最后一年,我越来越明白我应该对社会科学定量研究方法原理的“最小系统”加以阐明,这就是现在呈现在读者面前的这本书。我希望它至少能服务于以下三个目的:1)作为从没学过(或学过一些)定量研究相关课程的本科生或研究生

的教材或自学用的参考书。2)为想要了解、学习定量研究方法的读者,尤其是文科背景的研究者提供一本高效率的入门书。3)为想要掌握更多定量研究技术(如跟随最新调查方法的进展,复杂抽样设计,更高深统计分析技术,甚至哲学上的方法论背景,等等)的读者提供一个有用的较易接近的起点。

如把本书用作教材,则它尤适用于那些以学生自主学习和综合运用为主导的课程安排,如小组讨论课(seminar)、案例讨论课等。如以讲授为主导,则本书也适用于开设一门定量研究相关课程(如“社会科学定量研究”“社会科学研究方法”“调查研究方法”等),上、下篇的内容均可增删,取决于学生基础和教学计划的要求。

出于篇幅和结构上的考虑,本书省去了若干需要很多篇幅的那些推导,但都指明了参阅书目。本书还省去了按惯例作为附录的随机数字表和统计分布表,它们容易从互联网上下载到,统计分布表还可以用统计软件来生成。

最后,感谢东南大学出版社责任编辑张丽萍专业而辛勤的工作,也感谢负责审校的所有编辑,书稿完成后尽管我仔细通读校对了许多遍,但严格的审校流程还是帮助我减少了不少我自己没能发现的疏漏。

# 目 录

## 上篇 收集数据

|                    |    |
|--------------------|----|
| 第 1 章 研究设计 .....   | 2  |
| 1.1 研究问题 .....     | 3  |
| 1.2 变量测量 .....     | 6  |
| 1.3 抽样设计 .....     | 10 |
| 1.4 数据收集 .....     | 12 |
| 1.5 统计方法 .....     | 15 |
| 1.6 数据性质 .....     | 17 |
| 第 2 章 变量测量 .....   | 23 |
| 2.1 概念是什么 .....    | 23 |
| 2.2 概念的测量 .....    | 26 |
| 2.3 测量的性质 .....    | 27 |
| 2.4 复合测量 .....     | 29 |
| 2.5 测量尺度 .....     | 33 |
| 2.6 测量工具: 问卷 ..... | 35 |
| 第 3 章 概率抽样 .....   | 43 |
| 3.1 基本术语 .....     | 43 |

|              |                   |           |
|--------------|-------------------|-----------|
| 3.2          | 非概率抽样及其问题 .....   | 46        |
| 3.3          | 简单随机抽样 .....      | 49        |
| 3.4          | 等距抽样 .....        | 54        |
| 3.5          | 分层抽样 .....        | 58        |
| 3.6          | 整群抽样 .....        | 64        |
| 3.7          | 多段抽样 .....        | 68        |
| 3.8          | PPS 抽样 .....      | 71        |
| <b>第 4 章</b> | <b>数据收集 .....</b> | <b>78</b> |
| 4.1          | 无回答误差和测量误差 .....  | 79        |
| 4.2          | 数据收集方法及其比较 .....  | 84        |

## 下篇 分析数据

|              |                   |            |
|--------------|-------------------|------------|
| <b>第 5 章</b> | <b>概率基础 .....</b> | <b>94</b>  |
| 5.1          | 概率是什么 .....       | 94         |
| 5.2          | 概率公理化 .....       | 98         |
| 5.3          | 随机变量 .....        | 106        |
| 5.4          | 数学期望 .....        | 118        |
| <b>第 6 章</b> | <b>统计推断 .....</b> | <b>130</b> |
| 6.1          | 总体与样本 .....       | 130        |
| 6.2          | 统计量与抽样分布 .....    | 135        |
| 6.3          | 点估计 .....         | 143        |
| 6.4          | 点估计的优良性准则 .....   | 150        |
| 6.5          | 区间估计 .....        | 156        |
| 6.6          | 假设检验 .....        | 160        |

---

|                       |     |
|-----------------------|-----|
| 第 7 章 简单回归 .....      | 170 |
| 7.1 回归分析是什么 .....     | 170 |
| 7.2 简单回归模型 .....      | 176 |
| 7.3 简单回归模型的参数估计 ..... | 178 |
| 7.4 最小二乘估计的代数性质 ..... | 184 |
| 7.5 简单回归模型的基本假定 ..... | 187 |
| 7.6 最小二乘估计的统计性质 ..... | 193 |
| <br>第 8 章 多元回归 .....  | 204 |
| 8.1 多元回归模型 .....      | 204 |
| 8.2 多元回归模型的基本假定 ..... | 209 |
| 8.3 多元回归模型的参数估计 ..... | 216 |
| 8.4 最小二乘估计的统计性质 ..... | 224 |
| 8.5 多元回归模型的假设检验 ..... | 229 |
| <br>参考文献 .....        | 239 |

上篇

## · 收集数据 ·

## 第 1 章 研究设计

简要地说,研究就是提出问题,解决问题。因此,研究设计要解决的两大问题,其一是研究什么(以及为什么),其二是如何研究。“研究什么”,是对研究问题的清晰阐述。“为什么研究”,则是论证一项研究为什么是值得的。在一个研究领域提出某个值得做的研究问题,需要长期的学术训练,更需要问题意识。学术研究是学术共同体的共同事业——在这个意义上讲,研究问题的提出,不是任意的,不是研究者“个人的”(或者“随意的”),而是“公共的”。“如何研究”,是考虑一项研究的全盘计划。计划的一个要义在于:它是依此可执行的东西,是可以“照此办理”的文本。研究计划的设计是结构性的,也就是说,一项具体的研究的各个部分是有机结构在一起的。一项具体的研究的各个部分一般包括研究问题的提出及其意义的论证、变量测量、问卷设计、概率抽样设计、数据收集模式和统计分析方法的选择运用,还包括研究经费预算以及研究进度安排。这里结构一词的含义意味着各个部分是相互影响,甚至相互决定的,任何一个部分在研究中的失败,也将导致全部研究的失败。因此,研究设计是一项需要极其小心谨慎,并且尽力在各个部分之间找寻适当平衡的工作。通常研究所需经费与时间(合称为研究成本),是各个部分选择时的一个硬约束。

本章大体从各项结构要素的日常见解开始,对每个结构部分,以概要的方式加以阐明,同时从“一项具体的研究”这个角度,并在强调“计划是可执行的”这个意义上,尝试论述各部分的设计要

点。在本章最后的部分,在整体的意义上讨论“研究中的数据”,意在说明要做什么样的研究,就需要什么样的数据,或者倒过来,有什么样的数据,才能做什么样的研究。有关各部分的更加深入、更为全面的讨论,可详见之后的相应章节。

## 1.1 研究问题

前面已经提到研究问题,也提到问题意识。问题意识可能最易理解但也最难以捉摸。简单来说,能够提出有意义的、有价值的研究问题的能力或意识,可称之为问题意识,这看起来相对无章可循,而且我们需要首先搞清楚什么叫研究问题。

研究问题,不是泛泛的“关于什么”,而是明确且具体的设问,是提出有待数据加以检验的一个或一组命题的适当概括。例如,我们经常可以见到一些“消费情况调查研究”。这些研究是“关于消费”,并没有提出研究关于消费的什么问题,或者这类研究中,可能涉及诸多问题方面,如消费水平、消费倾向、消费心理、消费行为等等。把这些若干问题方面概括成“消费情况”,不是一种“适当概括”,同行或读者并不能知晓研究者要研究关于消费的“什么情况”。“情况”,通常是一种无意义的概括(过大的概括),因为这一概括并没有提供任何信息。此外,还应强调,即使是这些若干问题方面,在某种程度上,也还是某种“关于什么”。可以注意到,上述所列四个问题方面,在信息量的提供上大致是递减的,换句话说,在“清晰性”或“具体性”(clarity or specificity) (Frankfort-Nachmias et al., 2000)上是递减的。“消费水平”最接近于一个“具体性”问题,但比如“消费水平研究”,基本上还只是一项研究的题目(标题,而且不够好),研究者需要进一步明确提出研究“消费水平”的“什么(问题)”,比如可以是消费水平的差异,或者消费水平的影响因素,或者收入与消费水平的关系。这三者,我们称其为

定量研究中的研究问题,从这三者,很容易得到一组研究假设。例如,如果研究的分析单位是高校大学生,则从“消费水平的差异”可以提出以下假设:消费水平男女有别,或者不同年级的大学生消费水平不同,等等。这就是典型的“组间差异”研究,值得一提的是,如果研究“消费水平的性别差异”,这里并非只要有“消费水平”和“性别”这两个变量就足够了,而是需要考虑其他的一组控制变量,比如年级、专业、地区以及家庭收入等等,这样才可以在其他条件不变的情况下,检验消费水平的性别差异。事实上,这就可以建立一个多元线性回归模型。可见,在研究问题提出时,就有必要同时考虑所用的统计方法,称之为选题阶段的“统计先行”。

其次,研究问题并不是随意提出的,而是要符合“学理”。这就意味着,作为一项科学研究工作,自当符合科学事业的范式。一项具体的研究工作,应当是整个科学事业的一部分,尽管它可能极其微不足道。在这个前提下,研究者自当清晰阐述并严格论证,一项研究的研究问题是什么,以及为什么它是值得研究的。这种论证过程,往往同时也是研究问题的厘清过程。例如,在一个问题领域中,哪些问题是本项研究要研究的,为什么研究这个,而不是另外一个,已有的研究研究了哪些,用了哪些方法来研究,有什么样的研究结果,等等。由此,还有什么问题尚未得到关注,或者尚未有更好的方法加以研究,或者已有的研究结果存在怎样可质疑或可商榷的部分,等等。这一过程涉及大量已有的文献,这就是为什么学术上常常有这样的说法——做研究即是掌握文献。因为这一过程保证了一个研究问题的最终提出,同时也给出了研究问题本身及其价值的一个有力的论证,即现在研究者提出的这个研究问题(要做的这项研究),或多或少是重要的或有价值的——这就是通常所说的“选题及其意义的论证”。这里,还值得另外一提的是,不时可以看到一些研究(尤其研究新手的研究)在论证时的两种倾向:其一是“自称必要”,其二是“论证不相干”。“自称必要”是指,

无论是否经过一番怎样的讨论(论证),最后必明确声称若干结论性的断言,比如“本项研究有其理论意义和现实意义”,或者“本项研究有一定的理论意义和现实意义”,如此等等。尽管第二种说法意识到了“谦逊”之必要,但还是或多或少地属于“越俎代庖”。一个研究问题的提出,一项研究者打算着手的研究——有无价值——是由同行,由读者来评价,而不必由研究者“自称”。研究者的工作是提供足够有用的证据,以及在证据基础上的逻辑论证。“论证不相干”是指,不能充分论证或不能论证“这项研究为什么有价值”(注意着重号)。这经常表现为只谈“研究背景”,不及其他。例如在上述“大学生消费研究”中,只谈“消费”(消费这个事情,或消费这个问题)是如何重要——“消费”或“消费问题”是如此重要,因而这项研究是重要的,这经常被看成是不成问题的推理(论证),但这里混淆了“事情”“问题”与“研究问题”(清晰的、具体的),也混淆了“研究问题本身重要”与“研究这个研究问题(研究者做这项研究)重要”。还可以注意到,两种情形中前者对于后者来说,都是必要的,但不是充分的,因此在逻辑上推不出:“研究问题”或“研究者要做的这项研究”是重要的。可以证明的不过是:“事情”或“问题”不重要,则研究问题不重要;研究问题不重要,则做这项研究不重要(无论用多好的方法)。这个例子中“不能论证”还可包括论据(前提)不成立,例如只谈“大学生如何是一个特殊群体”“这个特殊群体有怎样的重要性”,等等。然而,什么叫做“特殊群体”?这个词是在什么意义上使用的?一个群体相比其他群体的“特殊性”因而“重要性”本身,怎么论证?!一系列的问题都还不甚了了。以上是两个简单的例子,总之,需要十分清楚的是,“选题及其意义的论证”,那么论题是一项“研究者要做的研究”的意义或价值——很明显这需要首先明确提出研究问题。因此,只谈研究背景(不及其他),首先意味着尚未明确提出研究问题,其次意味着在论证上(逻辑上)有误:用错推理规则,论据不足,或者偷换论题——论题被偷

换成了与这项研究有关的问题领域(甚至“事情”),或者与这项研究有关的分析单位(分析单位的讨论详见1.6节)。

综上所述,什么叫研究问题,取决于所陈述的命题(假设)的精确性,或者它所能提供信息的程度。我们在这里主张的一个要点是,在定量研究中,研究问题应当可以通过明确定义的变量来陈述。

## 1.2 变量测量

一项研究的研究问题的厘清并提出的过程,也是研究者搞清楚这项研究所涉及的全部变量,尤其是主要或关键变量如何测量的过程。如前所述,所谓研究问题的厘清,不过就是恰当地提出一个或一组有待检验的命题(即假设)。这里,命题的含义是,由确定的陈述来表达的概念间的关系。现在这些概念被我们看作变量,相应地,有待检验的命题(假设)被看作由确定陈述来表达的变量关系。“概念即变量”,这是主张概念应当并且可以被测量。显然,这一测量在相当程度上,应保有其客观基础。从逻辑上来看,这一主张,是(定量)社会科学研究得以展开,并且能够作为科学事业的某种范式的一个方法论前提。事实上,在定量社会科学研究中,概念与变量这两个术语,常常作为同义语而交替使用。变量测量的结果——数据——用来检验所提出的假设(对于抽样数据来说,即是统计推断)。毫无疑问,测量是定量社会科学研究的基础和关键,也常常是一项研究最为困难的部分。相当多的研究之所以有价值,常常在于对于研究的主要变量的测量有所学术贡献。

变量这一术语的含义,除了上述“概念测量主张”,或也可称为“概念可测性”主张,还在于在某种意义上,概念实际上多少是可以测量的,这样就可获得用于计算(统计)的数据。然而,与自然科学中的测量不同,例如使用称重的秤这样的测量工具——社会科学

研究中,概念的测量,其所使用的测量工具主要地是“问卷”,即通过询问被访者,以便通过被访者的回答,得到有关变量的信息(数据)。

例如,教育程度是一个变量,那么通过询问“您的最高学历是什么”,可以按小学及以下、中学、大学本科、硕士研究生以及博士研究生,编码为1到5,代表由低到高不同的教育程度。再如,收入变量的测量,是通过询问类似“您上个月的薪水”这样的问题,得到收入数据。这只是两个简单的例子(尽管实际上并不简单),测量这两个变量,可能是由于研究者打算进行某项“教育回报”的研究,当然,值得一提的是,还有其他很多的因素影响收入,例如行业或职业、工作经历,以及能力等,都是这项研究需要测量的变量,以便可以在这些“其他条件不变”的情况下,检验教育对收入的影响。

如果研究者打算做一项有关同情心的研究,例如同情心是否男女有别,或者研究人们的事业成功究竟是否取决于毕业于一所好学校,那么,“同情心”“事业成功”是这些研究中的关键变量。这样的变量,其测量如何可能——如果,尤其当我们尚不十分清楚这两个概念究竟是什么,简单地询问被访者类似“您的同情心可以打几分?”以及“您觉得您的事业成功吗?”——可能是相当误导的。因为,首先这样的询问所测量的,其实都是被访者对于同情心或者事业成功的自我评价,而这是与所测变量不同的另外两个变量,换句话说,这样的询问并没有真正测量人们的同情心或者事业成功这两个概念。其次,即使作为同情心或事业成功的自我评价的测量,那么上述询问事实上也并不能算作一个好的测量。对于前者来说,存在由所谓的“社会期许”(social desirability bias)(Bradburn et al., 2004)而导致的答案修饰,即倾向于高报同情心分值,因为谁都不太愿意被认为没有同情心或者不太有同情心。对于后者来说,不同的人对于“事业成功”如果理解不同,那么他们