

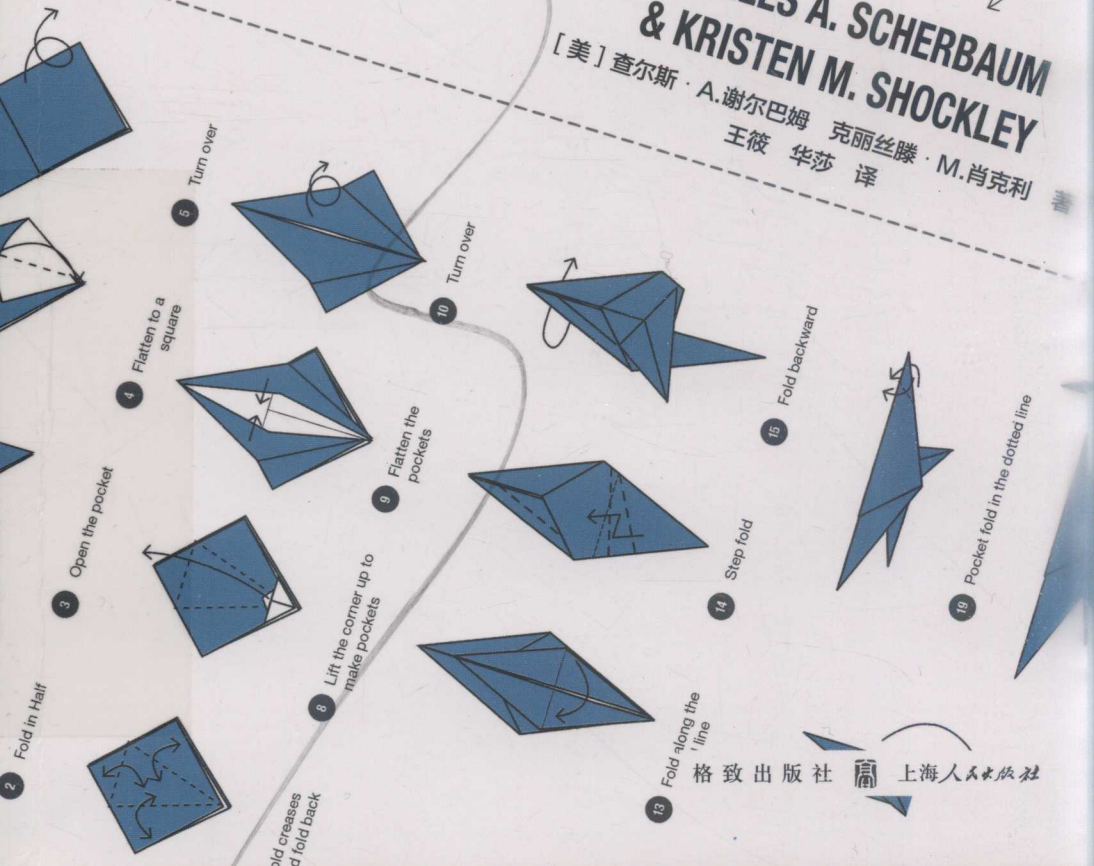
定量 数据分析

ANALYSING QUANTITATIVE DATA

for BUSINESS and MANAGEMENT STUDENTS

CHARLES A. SCHERBAUM
& KRISTEN M. SHOCKLEY

[美] 查尔斯·A.谢尔巴姆 克丽丝滕·M.肖克利
王筱 华莎 译



格致方法·商科研究方法



定量 数据分析

ANALYSING QUANTITATIVE DATA

for BUSINESS *and* MANAGEMENT STUDENTS

CHARLES A. SCHERBAUM
& KRISTEN M. SHOCKLEY
[美] 查尔斯·A.谢尔巴姆 克丽丝滕·M.肖克利
主 译
破 译
一 译
华 译
莎 译

图书馆



图书在版编目(CIP)数据

定量数据分析/(美)查尔斯·A.谢尔巴姆,(美)克丽丝滕·M.肖克利著;王筱,华莎译.—上海:格致出版社:上海人民出版社,2019.4
(格致方法·商科研究方法译丛)
ISBN 978-7-5432-2989-1

I. ①定… II. ①查… ②克… ③王… ④华… III. ①数据处理-定量分析-研究 IV. ①TP274
中国版本图书馆 CIP 数据核字(2019)第 043292 号

责任编辑 方程煜

装帧设计 路 静

格致方法·商科研究方法译丛

定量数据分析

[美]查尔斯·A.谢尔巴姆 著
克丽丝滕·M.肖克利
王筱 华莎 译

出 版 格致出版社
上海人民出版社
(200001 上海福建中路 193 号)
发 行 上海人民出版社发行中心
刷 常熟市新骅印刷有限公司
开 本 635×965 1/16
印 张 11.25
插 页 2
字 数 158,000
版 次 2019 年 4 月第 1 版
印 次 2019 年 4 月第 1 次印刷
ISBN 978-7-5432-2989-1/C·214
定 价 45.00 元



查尔斯·A.谢尔巴姆 Charles A. Scherbaum

纽约城市大学巴鲁克学院和研究生中心的心理学副教授，俄亥俄大学工业与组织心理学博士。他的研究集中在人员选拔、认知能力测试、对待被污名化员工的态度和应用心理测量学方面。其研究成果发表在《人事心理学》(*Personnel Psychology*)、《组织研究方法》(*Organizational Research Methods*)、《教育与心理测量》(*Educational and Psychological Measurement*)、《商业与心理学杂志》(*Journal of Business and Psychology*) 和《领导力季刊》(*Leadership Quarterly*) 等期刊上。

克丽丝滕·M.肖克利 Kristen M. Shockley

纽约城市大学巴鲁克学院和研究生中心的心理学系助理教授，南佛罗里达大学工业和组织心理学博士。她的主要研究领域集中在了解员工工作与家庭生活的交集。其研究成果发表在《职业行为杂志》(*Journal of Vocational Behavior*)、《管理杂志》(*Journal of Management*)、《人类行为》(*Human Performance*)、《职业健康心理学杂志》(*Journal of Occupational Health Psychology*) 和《混合方法研究和人事心理学杂志》(*Journal of Mixed Methods Research and Personnel Psychology*) 等刊物上。

本书主要包括定量分析的哲学和理论基础、数据收集方法与定量分析之间的相互影响、收集到的数据的属性对定量分析的影响形式、定量分析的实施步骤，以及该方法的实际应用案例等。全书条理清晰，简洁精炼，为试图使用定量分析方法进行学术、商业或管理研究的从业者提供了颇为实用的指导。



编辑寄语

欢迎学习商科研究方法。近年来,攻读商科硕士专业学位的学生日益增多。在攻读学位的最后阶段,研究生都要花费3—4个月的时间来撰写论文。对于大多数学生来讲,撰写论文都是在课程结束之后,这与课程是完全不同的。每个研究生都由导师来指导论文撰写或进行课题研究,研究生需要详细了解自己使用的研究方法。开始撰写论文或开始课题研究之前,研究生们通常都会接受一些研究方法的训练以完成论文或课题。如果你就是正在撰写论文的研究生,那么你不会孤军奋战,我们向你提供了一套书来帮助你。这套丛书的每本书都就某一具体的方法提供了详细的信息以帮助你的论文撰写。我们理解什么是硕士学位论文所需要的研究方法,也理解什么研究方法是硕士研究生所需要的,从而帮助你在撰写论文的时候能够出色地完成任务,这正是商科研究方法系列丛书的编写目的。

该丛书中的每一本都旨在对数据收集或数据分析方法提供足够的知识,当你进行到研究过程的每个具体阶段的时候,你都可以找到一本与其相应的方法介绍,如数据收集或数据分析。每一本都邀请了业界权威的学者来编写,他们都在研究方法的教学与写作方面具有丰富的经验,因此他们的作品清晰易读。为了让学生在学习丛书中的每一本的时候都能够迅速找到自己需要的内容,丛书使用了标准的格式,即每本书均由6章构成:

- 第1章:导论,介绍方法的目的和本书纲要;

- 第2章:研究方法的哲学假定;

- 第 3 章:研究方法的组成部分;
- 第 4 章:研究方法不同的组成部分;
- 第 5 章:提供研究中使用该种研究方法的例子;
- 第 6 章:结论,该种研究方法的优点与缺点。

我们希望阅读本书对你撰写论文有所帮助。

比尔·李、马克·N.K.桑德斯和 V.K.纳拉亚南

本书是清华大学出版社出版的一本关于定量数据分析的教材。本书共分六章，第一章介绍定量数据分析的基本概念，第二章介绍定量数据分析的组成部分，第三章介绍定量数据分析的不同组成部分，第四章介绍定量数据分析的例子，第五章介绍定量数据分析的结论，第六章介绍定量数据分析的优点与缺点。本书旨在帮助读者了解定量数据分析的基本概念和方法，并为撰写论文提供必要的参考。

本书由清华大学出版社出版，作者为比尔·李、马克·N.K.桑德斯和 V.K.纳拉亚南。本书共分六章，第一章介绍定量数据分析的基本概念，第二章介绍定量数据分析的组成部分，第三章介绍定量数据分析的不同组成部分，第四章介绍定量数据分析的例子，第五章介绍定量数据分析的结论，第六章介绍定量数据分析的优点与缺点。本书旨在帮助读者了解定量数据分析的基本概念和方法，并为撰写论文提供必要的参考。

本书由清华大学出版社出版，作者为比尔·李、马克·N.K.桑德斯和 V.K.纳拉亚南。本书共分六章，第一章介绍定量数据分析的基本概念，第二章介绍定量数据分析的组成部分，第三章介绍定量数据分析的不同组成部分，第四章介绍定量数据分析的例子，第五章介绍定量数据分析的结论，第六章介绍定量数据分析的优点与缺点。本书旨在帮助读者了解定量数据分析的基本概念和方法，并为撰写论文提供必要的参考。

丛书编辑简介

比尔·李(Bill Lee)博士,会计学教授,英国谢菲尔德大学会计与金融系负责人。他在研究方法和研究实践领域具有多年的经验,另外,他的研究方向也包括会计和会计准则。比尔的研究兴趣广泛,成果多发表在 *Accounting Forum*、*British Accounting Review*、*Critical Perspectives on Accounting*、*Management Accounting Research*、*Omega* 和 *Work, Employment & Society* 等期刊。他的关于研究方法和研究实践的科研成果发表在 *The Real Life Guide to Accounting Research* 及 *Challenges and Controversies in Management Research* 中。

马克·N.K.桑德斯(Mark N.K.Saunders)博士,英国萨里大学商科研究方法教授。他的研究兴趣是研究方法,特别是内部组织关系、人力资源管理方面的变革、组织内和组织间的信任和中小企业的研究。马克在一些学术期刊中有很多发表,如 *Journal of Small Business Management*、*Field Methods*、*Human Relations*、*Management Learning* 和 *Social Science and Medicine*。同时,他也是一些专著的合著者和合作编辑,如 *Research Methods for Business Students* (目前已经是第6版)和 *Handbook of Research Methods on Trust*。

V.K.纳拉亚南(V.K.Narayanan),美国宾夕法尼亚州费城的德雷塞尔大学教授,研究院副院长,战略和创业研究中心主任。他先后在一些顶尖专业期刊发表文章,如 *Academy of Management Journal*、*Academy of Management Review*、*Accounting Organizations and Society*、*Journal of Applied Psychology*、*Journal of Management*、

Journal of Management Studies、*Management Information Systems Quarterly*、*R&D Management* 及 *Strategic Management Journal*。纳拉亚南在印度马德拉斯的印度理工学院获得机械工程学士学位,艾哈迈达巴德印度管理学院获得管理学硕士学位,美国宾夕法尼亚的匹兹堡大学商学院获得博士学位。

1

目 录

- 1 导论 /1
- 2 理解定量数据分析 /9
- 3 定量数据分析的基本组成 /19
- 4 实施定量数据分析 /41
- 5 定量数据分析实例 /123
- 6 结论 /136
- 词汇表 /146
- 附录：Excel 公式 /160
- 参考文献 /162
- 译后记 /170

1 导 论

定量数据与分析已成为研究与商业领域不可或缺的一部分(Economist, 2010; Salsburg, 2002; Siegel, 2013)。许多学者和观察家都认为我们现在生活在“大数据”和定量分析的时代(McAfee and Brynjolfsson, 2012),研究和商业的本质正在转化为一种结果(Ayers, 2008; Davenport and Harris, 2007; Siegel, 2013)。事实上,纵观当今的学术文献或商业媒体,不提到“大数据”和定量分析是不可能的事情。有关定量数据分析如何从根本上改变某些方面的研究、商业活动或日常生活的新闻故事似乎每天都在上演。有关成功运用定量分析的书籍,如 Michael Lewis(2004)的《点球成金》(Moneyball),已经成为许多商业领域的“必读”书目。

考虑到量化数据和分析对各行各业的影响,从职业运动到医学,到互联网搜索,再到管理、运营和市场营销,就不难理解定量数据分析给我们带来的兴奋感。这种重新燃起的兴趣的好处在于新的方法、软件、书籍、网站、会议和信息都在以惊人的速度增长。对于那些对定量数据和分析感兴趣的人来说,这些都是激动人心的时刻。

在激动的同时,对于所有研究者和实践者来说,时代正在改变。我们似乎已经到达了一个临界点,即对于大多数研究领域和商业领域来说,定量数据分析的相关知识不再是可有可无的了。我们不禁要问,一个无法理解和使用定量数据分析的人能否在研究或工业领域取得成功?无数的报告和调研得出这样一个结论,我们对拥有定量数据分析知识和技能的人有巨大的需求,然而这些人才却是短缺的。有些人甚至宣称,定量分析工作将成为 21 世纪最受欢迎和最有声望的职业

之一。

即使对那些熟知且能熟练运用定量数据分析的人来说,工具和方法的不断发展和扩散也决定了我们应该做这件极具挑战性的事情。我们发现,有经验的研究人员和从业人员问的最多的问题是,“如何确定哪些分析是最适合于数据和研究方法的?”从本质上来说,适当的定量分析的选择是基于研究目的(如探索性分析和验证性分析)、研究问题(以及有待验证的相关假设)、研究设计(如实验或非实验性研究)和数据的属性(如数值测量尺度、分布的形状;Scherbaum, 2005)。然而,做出这些决定仍然是非常困难的,因为在选择定量分析时必须考虑大量的信息和选项。

这些挑战,尤其是对于学生而言,还来源于为了理解定量分析而构建框架或心智模型所带来的困难。定量分析不仅仅是一套应用于数据的技术。更准确地说,它是一种系统地思考研究问题、研究方法论和观察到的数据模式的方法。人们甚至可以把定量分析看作一种语言,这种语言使研究人员能够使用商定的词汇进行交流。正是这种定量分析的心智模型,为选择适当的分析、研究方法和研究问题,提供了宝贵的帮助。

本书旨在为学生、研究者和从业者在应对复杂的定量数据分析问题时提供辅助。我们的目标是不仅能为读者提供发展定量分析心智模型所必须的基础,也可以使读者大致了解目前所有可用的定量分析方法。为此,我们涵盖了定量分析的哲学和理论基础、数据收集方法与定量分析之间的相互影响、收集到的数据的属性如何影响定量分析,以及为准备定量分析所需数据所包括的步骤等等。无论是在学术上还是商业中,本书所包含的定量分析案例已经足以解答大部分困扰学者和从业者的常见问题。至于更特殊和更先进的定量分析方法(例如因素分析、联合分析),读者将能够在本系列的其他丛书中找到相关的信息。

我们发现,现有的有关于定量分析的书籍很少集中在思考和设计定量分析的更概念化的方面。它们也没有关注定量分析和研究设计与研究方法之间的关系。在起草这本书时,我们尝试解决这个问题。我们努力为人们需要考虑的因素、应该进行的步骤,以及在进行定量分析时必须做出的决策提供指导。我们有意避免提供一种刚性规则约束系

统,不让一种方法成为针对某种特定情形的“唯一”适用的定量分析。正如 Abelson(1995)所说,很少有一个唯一正确的选择或规则。然而,采用这种方法确实需要权衡。这种权衡是,我们更少地关注每一个定量分析和过程的各个方面的详细描述,以及在统计软件程序中运行它们的程序。考虑到定量分析得天独厚的资源具有广泛的可用性,它涵盖了研究的每个细节,以及对统计软件的使用(Cohen et al., 2003; Field, 2013; Pedhazur and PedhazurSchmelkin, 1991),我们认为这种权衡对那些刚刚开始使用定量分析的人来说是非常有益的。

1.1 什么是定量分析

定量分析是统计学,它将大量的数据简化到更易于管理的形式,以便得出关于数据模式的结论和观点。虽然定量分析有许多类型,但本书主要介绍用于描述定量数据、识别组间差异、检查变量之间的关联或关系,或进行预测的常用方法。定量分析最基本的形式是描述性定量分析。描述性定量分析可以用来将大量的数据压缩成一个较小的数字集合,用来表示数据中典型的数据和数据中的变量数量。如第 4.1 节所述,描述性定量分析包括频数、众数、中位数和平均数(平均类型)以及全距、方差和标准差。描述性定量分析可以用数字和图形两种形式表示(例如条形图和直方图)。无论是研究问题还是研究方法,描述性定量分析都应该始终是数据分析过程中的第一步。

与描述性分析相关的是定量分析,它被用于检查各组结果之间或结果集合之间的差异。例如,这些分析可能被用来研究那些接受个性化促销的人与那些接受一般性促销的人之间的客户参与度的差异。如第 4.2 节所述,这些分析通常检查组间平均值的差异,包括 t 检验和方差分析。

其他的定量分析着重于变量之间关系的方向和强度,作为这方面的延伸,其目标是基于一个变量对另一个变量进行预测。这些分析是相关性和回归方法(见第 4.3 节)分析。相关性往往是理解变量是否相

关的第一步。回归分析通过允许人们创建一个方程式来扩展相关性,该方程式可以基于一组输入来预测或预示一种结果。例如,可以开发方程式,以便从当前的销售培训投资中预测未来员工的销售业绩。

1.2 定量分析的历史基础与当前趋势

虽然目前对定量数据分析的很多报道和讨论都把它看作是刚刚出现的革命性突破,但大多数常用的定量分析在 100 多年前就已经开发出来。尽管对定量分析发展历史的详细回顾超出了本书的范畴(Lehman, 2011; Salsburg, 2002),但重要的是要考虑一些为本书中所讨论的许多分析带来最初发展,并对当今的定量分析仍然产生影响的研究挑战。

大部分商业和管理科学的研究集中在与大群体有关的问题上。例如,研究者可能会对培训销售人员的说服技巧对销售效率提高的效果感兴趣。研究者在一家公司对一小组销售专业人员进行培训,收集培训完成情况和销售业绩的数据,并分析数据。最终,研究人员对他们研究中的具体销售人员不感兴趣,他们感兴趣的是所有销售专业人员。那么问题来了,如何将他们的研究结果从(作为他们研究的一部分的)特定一组销售专业人员发展到代表所有销售专业人员的更大群体。更具体地说,问题是如何从样本(例如样本中的平均值)的统计计算得到总体中相同统计量(即总体参数)的充分估计。这个问题是早期统计学家为解决问题而开发的方法。也就是说,如果分析只是基于这个总体的样本,那么对总体的归纳就是合理的吗? 总体代表了一个特定研究问题(如英国的人口)的整个群组。样本代表总体的一个子集(例如那些在某天经过特拉法加广场的时候愿意完成调查的人)。

从总体中收集样本的过程包含一定程度的不确定性。不确定性是指研究人员不确定他们的样本是否足够代表总体。任何给定的样本都不是总体的完美反映,这并不奇怪。使用样本统计量作为总体参数的估计值时,总是存在一定程度的误差,确切来说是抽样误差。这种抽样

误差会导致同一主题的研究结果之间出现波动。研究人员对这些波动最主要的担心是,从一个不能真正反映大范围总体的样本中收集数据,然后从这个样本中得出的关于总体的推论可能是错误的。抽样误差最被人熟知的例子之一就是 1948 年的美国总统选举。在对一组无法代表整体选民的投票人进行调查后,《芝加哥论坛报》(*Chicago Tribune*)预测,Thomas Dewey 已经击败了 Harry Truman,并为此发表了一个标题。然而,Harry Truman 取得了决定性的胜利,给该报纸造成了相当程度的尴尬。定量分析是为了帮助确定从样本到总体的推论是否确实值得,尤其是当样本只占总体的很小比例时。

从样本数据中得出推论的实际问题引发了定量分析的许多发展。然而,目前定量分析使用增多的原因是不同来源的可被整合的大量数据(如“大数据”)的可用性,以及定量分析软件的发展(McAfee and Brynjolfsson, 2012; Mayer-Schönberger and Cukier, 2013)。曾经,数据一度难以获得,数量少,并且需要单独进行审查。随着技术的进步,大量的数据现在可以从多个来源获得,并且可以被整合。例如,现在可以整合来自销售点交易的数据、客户满意度调查数据、网络流量数据、完成的客户服务培训数据、员工敬业度数据,以及每笔销售交易和组织中每个员工的薪酬数据等。

例 1.1 大数据的运用

大多数大型零售商目前正在使用来自销售点、网络流量、客户调查和对营销活动响应的“大数据”,据此为客户定制优惠券和广告,从而有助于增加购买行为。一家北美零售商使用其关于所有客户购买历史的大型数据库,以及注册婴儿登记的客户数据库,以预测哪些女性顾客可能正处在怀孕的第二个妊娠期。我们可以想象,在他们向朋友和家人宣布怀孕之前,关于新生儿产品的介绍手册就已经发送至他们邮箱时,顾客会多么惊讶!尽管以这种使用数据的方式可能会引发道德问题,但正是这种整合和利用正常业务运营所创造的大量数据的能力,使定量分析的广泛运用呈指数增长。

统计软件的有效性很大程度上影响着定量分析的快速普及,特别是在商业领域。开展定量分析曾经是高技能专家的专属领域。这些软件需要专业知识才能在所有可用的分析和每个分析中存在的众多选项之间进行选择。现如今,被开发的大量新软件已经供非专业人士使用。例如,IBM 的 SPSS 建模提供了用于特定分析的选项之间的自动选择,并且使用视觉示意图进行分析,用户拖拽一个节点,表示一个特定的分析类别(例如回归),并将其与表示数据的节点进行连接。软件会完成剩下的步骤,用户不需要做任何其他的决定,所需要的仅有软件、数据和一般分析类型的想法。正如本书将要讨论的那样,不建议盲目地和不加批判地使用定量分析,这可能会导致误用定量分析,并且会对某一领域的理论进步产生负面影响。

1.3 数量分析中的关键概念

无论选择何种具体的定量分析,定量分析的使用者都需要熟悉一些关键概念。定量研究和分析中最基本的概念之一就是测量(Scherbaum and Meade, 2009)。测量可以被概念化为按照一组规则将对象、事件或人的特性或属性赋予数字(Stevens, 1968)。对象、事件或人物的这些属性是定量分析中的变量。在很多管理研究中,这些变量被用作不可观测的构念的间接指标。构念是为解释对象、事件或人物的属性的差异、共同点或模式而提取出的抽象概念。例如,人格是管理和组织科学中使用的一个众所周知的构念。人格是心理学家为了研究人们行为倾向和人与环境的交互效应之间的差异而提出的抽象概念。

构念构成了商业和管理科学许多理论的基础。这些理论反过来又是所设计的定量分析假设的来源。假设是关于总体中变量之间的预期关系或差异的可验证的陈述。这些假设将使用从相关总体中抽取的样本数据来检验。定量分析的结果被用来确定从样本中收集的数据是否

支持关于总体的推论。贯穿本书的重点是理论和假设作为定量分析的一部分的重要性。虽然定量分析可以,而且经常在没有理论或假设的前提下使用,但我们强烈不赞同这种做法。在一些领域中,如互联网搜索分析、市场调查或犯罪预防,唯一感兴趣的研究问题是“会发生什么事情”,或者“购买某一特定产品的顾客是否也购买了另一种特定产品?”对于这些描述性定量分析,理论可能没有必要。然而,在很多商业和管理研究中,研究问题或假设的兴趣都集中在为什么会出现某种效应或现象。没有理论指导下的分析,这些问题就不能得到充分的解答。

正如第2章将要详细阐述的那样,定量分析的基本概念是概率。概率可以被描述为一个特定结果的可能性。概率的范围从0.00(没有发生结果的机会)到1.00(结果肯定会发生)。概率作为所有定量分析的基础,试图从样本扩展到总体。一个假设是否是被支持的,取决于与定量分析中观察到的结果相关的概率。换句话说,概率是决定样本收集数据是否支持总体推断的主要标准之一。

1.4 理解定量分析与研究方法之间的对应关系

正如第3章所述,研究方法、研究设计和定量分析之间有着本质的联系。研究中使用的方法和设计会影响定量分析结论的属性、可能适合的分析类型、所需的数据量,以及特定定量分析假设被满足的可能性。例如,无论使用何种定量分析,非实验研究设计都不会支持关于因果关系的强有力的结论。同样,如果所收集的数据代表类别(例如,男性/女性;是/否),则只有有限的定量分析是适合的。定量分析的使用者需要了解研究方法、研究设计和定量分析之间的密切关系。也就是说,我们不主张设计研究来支持特定的定量分析的使用。理论和研究问题决定了所使用的研究设计和方法,从而决定最佳的定量分析。

研究方法和研究设计也与定量分析有着不太明显的关系。在许多科学领域中,对某些类型的研究设计已有了传统意义上的“正确的”定