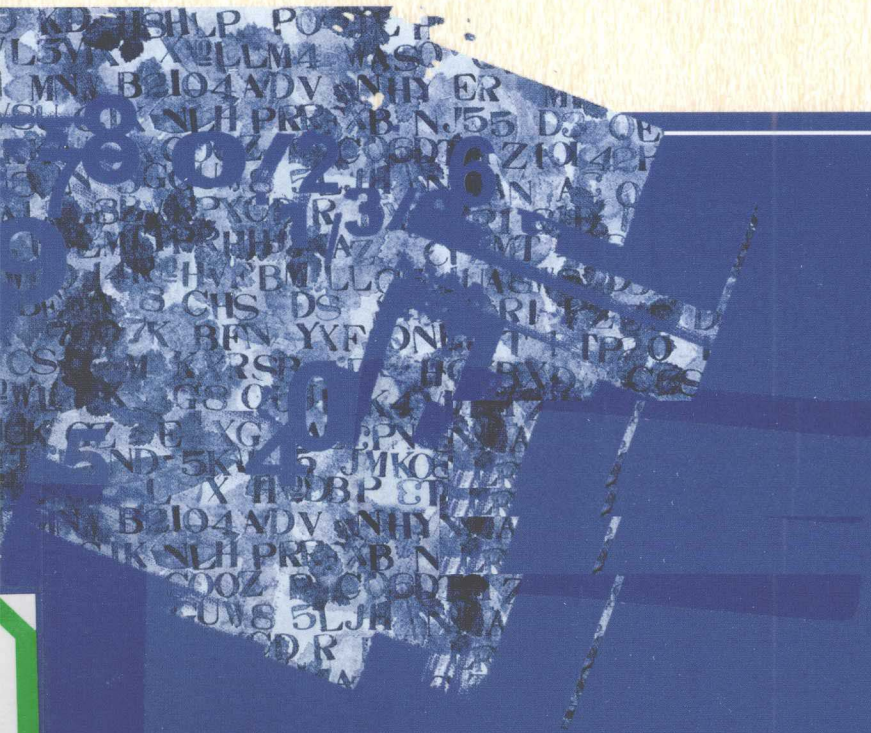
社会统计学

对问卷调查数据的统计分析

Social Statistics An Introduction to Statistical Analysis of Survey Data



游正林 著

社会科学文献出版社
SOCIAL SCIENCES ACADEMIC PRESS (CHINA)

社会统计学

对问卷调查数据的统计分析

游正林 著

社会科学文献出版社
SOCIAL SCIENCES ACADEMIC PRESS (CHINA)

图书在版编目（CIP）数据

社会统计学：对问卷调查数据的统计分析/游正林著．—北京：社会科学文献出版社，2010．9
ISBN 978－7－5097－1721－9

Ⅰ．①社… Ⅱ．①游… Ⅲ．①社会统计 Ⅳ．①C91－03

中国版本图书馆 CIP 数据核字（2010）第 148515 号

社会统计学

——对问卷调查数据的统计分析

著　　者／游正林

出 版 人／谢寿光
总 编 辑／邹东涛
出 版 者／社会科学文献出版社
地　　址／北京市西城区北三环中路甲 29 号院 3 号楼华龙大厦
邮政编码／100029
网　　址／http：//www. ssap. com. cn
网站支持／（010）59367077
责任部门／社会科学图书事业部（010）59367156
电子信箱／shekebu@ ssap. cn
项目经理／童根兴
责任编辑／杨桂凤
责任校对／姜夕芬
责任印制／郭　妍　岳　阳　吴　波

总 经 销／社会科学文献出版社发行部
（010）59367080　59367097
经　　销／各地书店
读者服务／读者服务中心（010）59367028
排　　版／北京中文天地文化艺术有限公司
印　　刷／北京季蜂印刷有限公司

开　　本／787mm×1092mm　1/16
印　　张／13. 75
字　　数／218 千字
版　　次／2010 年 9 月第 1 版
印　　次／2010 年 9 月第 1 次印刷

书　　号／ISBN 978－7－5097－1721－9
定　　价／29. 00 元

作者前言

社会统计学是“统计学”与“社会研究”的结合，旨在探讨如何将统计学的方法应用于社会研究当中。在社会研究当中，最常用的结构化的经验资料是问卷调查数据，本书主要关注问卷调查数据的统计分析。

本书主要是写给社会学和社会工作专业的大学本科生的，当然，对问卷调查研究感兴趣的其他专业的同学也可以阅读本书。我从多年的教学经验中了解到，在社会学、社会工作等专业就读的大学本科生中，有不少同学都不喜欢甚至害怕学习社会统计学或者担心学不好社会统计学。我之所以编写这本教材，主要目的之一就是想从教材编写的角度“矫治”这种“统计厌学症”或“统计焦虑症”。我所用的“矫治”手法主要有以下四种。

首先是简化全书的内容结构。问卷调查研究的基本目的可以概括为两个：一是描述被调查者的有关特征，也即描述有关变量的特征，主要是把有关变量的变异描述清楚；二是进一步解释为什么会产生这种变异。本书的内容始终围绕如何描述变异、如何解释变异以及如何将这种描述和解释由样本推论至总体而展开。解释变异属于相关分析的范畴，在讲解这部分内容时，重点放在以 PRE（减少误差比例）为基础的相关测量上；在讲解推论统计时，则把重点放在样本统计值的抽样分布的性质上。这样，全书主线清晰，重点突出，结构简明，便于同学们从整体上把握社会统计学的内容。

其次是简化对具体内容的讲解，其具体做法有二：一是淡化统计方法背后的数理基础，侧重于介绍统计分析的逻辑与原理；二是注意与统计分析软件（SPSS）衔接，所使用的数学符号、统计术语等与 SPSS 保持一致，所有复杂的计算过程都交给 SPSS 去做，不在手工计算上耗费时间。

再次是简化语言表述，力求做到深入浅出。在这方面，我给自己定的基本标准是：即使没有老师亲自讲授，用心的同学也能看得懂这本书。不过，统计学毕竟拥有一套比较抽象的、独特的语言，即统计语言，我们不可能把它们都“翻译”为日常口语，因此，我并不希望同学们抱着浏览小说、散文那样的心态来阅读这本书。

最后是在每章的最后都设置一些形式比较新颖的思考、练习题，希望通过这种方式进一步引导同学们深入思考有关的问题，帮助他们巩固所学的知识，提高他们学习本课程的兴趣。

在本书的编写过程中，我参考了比较多的相关文献。中文文献我主要参考了卢淑华教授的《社会统计学》（第三版）（北京大学出版社，2005）和李沛良教授的《社会研究的统计应用》（社会科学文献出版社，2001）；所参考的英文文献则更多，详见书后的“主要参考书目”。每一章每一节的编写，我都从这些文献中汲取了营养，尤其是从英文文献中汲取了如何深入浅出地讲解有关统计内容的营养。我尽量以页下注的形式注明了这些文献的出处，有兴趣的同学可以进一步阅读这些文献。

本书的蓝本是我给本院 2006、2007 和 2008 级社会学专业的同学讲授“社会统计学”课程的讲稿。这三个年级的同学的勤学好问，不但使我真正体会到了教学相长的乐趣，而且激励着我努力把讲稿“升级”为书稿。在书稿即将出版之际，我首先要感谢这些同学。本书的出版，得到了社会科学文献出版社范广伟先生与童根兴先生的热情帮助以及杨桂凤女士的悉心编辑，还得到了中国政法大学社会学院“国家级特色专业经费”的资助，在此一并致谢。

游正林

2010 年 6 月于中国政法大学社会学院

目录
CONTENTS

第一章
导　论

社会统计学（social statistics）是“统计学”与“社会研究”的结合，旨在探讨如何将统计学的方法应用于社会研究。这里所说的“统计学”是指一系列整理、计算、概括和推断数字信息的方法，即统计学的方法；社会研究是指经验性的社会研究，即指研究者通过自己的感觉（如听觉、视觉等）器官来收集资料并寻求有关社会世界的问题的答案的研究方式。以这种方式收集的资料称为经验资料（empirical data）。经验资料分为结构化资料（structured data）和非结构化资料。如果是结构化资料，我们就可以通过编码把它们转换为数字形式并运用统计学的方法进行统计分析。在社会研究中，最常用的收集结构化资料的工具是调查问卷。因此，就像本书的副标题所显示的那样，本书主要关注问卷调查数据的统计分析。

本章主要介绍与问卷调查研究有关的内容，它由6个部分组成：问卷调查研究的基本特征、因果分析的基本逻辑、问卷调查研究的方法论背景、问卷调查研究的基本过程、测量的四个层次、统计分析的类型与本书的内容编排。

第一节　问卷调查研究的基本特征

为了更直观地说清楚问卷调查研究的基本特征，我们不妨先来思考一

个问题：假如我们想调查了解本校（大学）一年级和二年级本科同学对学习英语是否感兴趣、感兴趣的程度如何等问题，怎么来了解？方法之一就是根据所要调查研究的问题事先设计一份如下所示的调查问卷，然后，请全校所有一年级和二年级的同学或随机地从中抽取一部分同学来回答这份问卷中的问题。回答的方式有两种：一是由被调查者自己填答问卷，二是由调查员拿着问卷当面或通过电话询问被调查者，[①] 并在问卷上如实记录被调查者的回答。统计分析所有被调查者对这些问题的回答，即可以达到调查了解的目的。这种做法就是社会研究中最常用的一种研究策略——问卷调查研究。

1. 首先，请问你的性别是什么？[②]

①男；　　②女。

2. 你现在读大学几年级？

①一年级；　　②二年级。

3. 本学期以来，你对学习英语是否感兴趣？感兴趣的程度如何？

①很不感兴趣；　　②不太感兴趣；　　③一般；

④比较感兴趣；　　⑤很感兴趣。

4. 最近一次英语课（指本校统一安排的英语课），你去上了吗？

①去了；　　②没有去。

5. 你对本学期教你们班的英语老师的教学水平是否满意？满意的程度如何？

①很不满意；　　②不太满意；　　③一般；

④比较满意；　　⑤很满意。

6. 进入本校学习以来，除了上本校统一安排的英语课以外，你参加过其他收费的英语培训班吗？

①参加过；　　②没有参加过。

……

① 在询问过程中，调查员必须严格遵守事先规定好的提问顺序、提问方式和其他具体要求。这种做法也叫结构式访问。

② 如果是访问，则不要询问这个问题，只需记下被访者的性别即可。

目前，这种做法的应用范围很广，一些非专业人士也常常使用这种策略来了解情况、研究问题。不过，本书所说的问卷调查研究方法主要指专业（如社会学专业）意义上的一种研究社会现象的方法，它具有以下几个基本特征。

第一，它的调查对象和分析单位都是一个一个的人，且至少调查两个人，对每个人所询问的问题也至少在两个以上。之所以要做这样的要求，主要是出于探讨事物之间的关系尤其是因果关系的需要。一般来讲，每次问卷调查的被调查者人数都在50人以上，每份问卷所询问的问题也远远多于两个。这样，每次问卷调查往往能收集到大量的调查资料，因而需要利用专门的统计软件（如SPSS）来统计分析这些资料。

第二，它通常只收集一个时间点上的资料，即要求所有被调查者在同一个时间或者大体上在同一个时间回答问卷中的问题，以便描述被调查者在这个时间点上的有关特征并探讨这些特征之间的关系。被调查者的特征可以分为三类：一是稳定性很高的基本特征，如上述问卷中第1题和第2题所问的内容，此外，通常还包括年龄、职业、文化水平等；二是行为特征，如上述问卷中第4题和第6题所问的内容；三是态度（包括价值观念）特征，如上述问卷中第3题和第5题所问的内容。由于这些特征，特别是行为特征和态度特征往往会随着时间的推移而不断发生变化，而我们每次调查只不过了解了其变化过程中的某个“横截面”而已，因此，它又被称为横剖研究（cross-sectional research）。在这一点上，它不同于自然科学研究中常用的实验研究（详见下节）。

第三，它收集资料的工具是结构化的调查问卷。所谓“结构化”，是指问卷中的问题及其备选答案都是事先设计好了的，这意味着对所有的被调查者都是以同样的方式询问同样的问题，所有的被调查者都只能从所给的备选答案中选择符合自己情况的答案。这种“结构化”的好处主要有两个：一是便于被调查者回答；二是能够将所收集的问卷调查资料转换为可以进行直接对比的数字形式，从而使对这些调查资料进行统计分析成为可能。在如上所示的问卷调查中，如果用备选答案前的序号来代替该备选答案，那么，就可能将所有被调查者的回答转换为如表1－1那样的数据表格。在这张数据表格中，所有被调查者对同一问题的回答的信息都放在

同一列里，每个被调查者对所有问题的回答的信息都放在同一行里，这意味着有多少位被调查者，这张表格就有多少行。这样，本次调查的所有原始数据就都呈现在这张表格里。这也是统计软件 SPSS 呈现问卷调查数据的形式。有了这样的数据表格，我们就可以直接比较不同的被调查者对有关问题的回答，从而发现不同的被调查者对某个问题的回答是否不一样即是否存在变化。谈及这种“不一样”或“变化”的现象，有必要在此先介绍一下“变量”（variable）概念，我们以后会经常和它打交道。

简单地讲，一个变量就是被调查者所具有的可以变化的某个特征。这里所说的“变化”，是指在被调查的人群中，不同的被调查者在某个特征上并不都是一样的，而是存在差别、有所变化的。比如，在被调查者中，如果既有男的，又有女的，那么，被调查者的“性别”特征就是一个变量。反过来，如果被调查者都是男的或都是女的，也即在性别特征上没有差别、没有变化，那么，他们（或她们）的“性别”特征就是一个常量，而不是变量。在社会研究中，几乎碰不到这样的常量。即使偶尔能碰到常量，社会研究者也对它们不感兴趣，而只对变量感兴趣，即关注它们变化了多少、是怎么变化的、为什么会变化。其实，表 1－1 中所列的几个问题（问题 1 至问题 6）所对应的都是变量，它们分别是“性别”、“年级”、“学习英语的兴趣程度”、“是否去上了最近一次英语课”、“对英语老师的教学水平的满意程度”和“是否参加过其他收费的英语培训班”。

表 1－1　被调查者的回答资料被转换为数据表格

	问题 1	问题 2	问题 3	问题 4	问题 5	问题 6	…
被调查者 1	1	1	1	1	1	1	…
被调查者 2	1	2	2	1	2	2	…
被调查者 3	2	1	4	2	4	2	…
被调查者 4	2	1	2	1	3	1	…
被调查者 5	1	2	1	2	2	1	…
被调查者 6	1	1	2	2	1	2	…
被调查者 7	2	2	3	1	5	2	…
…	…	…	…	…	…	…	…

第四，它的基本目的可以概括为以下两个：一是描述被调查者的有关特征，也即描述有关变量的特征，主要是要把有关变量的“变异或变化”（variation）[①] 描述清楚；二是进一步解释为什么会产生这种变异。要进行这种因果解释（causal explanation），首先必须把一个变量的变化与另一个（或几个）变量的变化进行比较，看看二者之间是否存在系统性的关联，然后，进一步推断这个变量的变化是不是由于另一个（或几个）变量的变化所引起的。比如，在上述问卷调查中，当发现不同的被调查者学习英语的兴趣程度“不一样”也即存在“变化”时，我们就需要进一步解释为什么会产生这种变化。这种解释性研究，其实是探讨两个或两个以上变量之间的关系，属于相关分析并进一步进行因果分析的范畴。

第五，它通常采取随机抽样的方法从调查对象的总体中抽取一部分人来填答调查问卷或进行结构式访谈。被随机抽取出来的这部分人叫做样本。一般来讲，研究者希望得出关于总体的结论，他们之所以观察、描述、概括样本的特征，是希望通过样本的特征来猜测或推论总体的特征。

第二节　因果分析的基本逻辑

如上所述，问卷调查研究的基本目的：一是描述变量的变化，二是进一步解释为什么会产生这种变化。这种对变化的解释常常涉及变量之间的因果关系。因此，有必要先介绍一下因果分析的基本逻辑。

简单地讲，如果 X 是 Y 的一个原因，则意味着 X 的变化会引起 Y 的变化。[②] 因果分析的第一步是拿一个变量的变化与另一个变量的变化进行比较，看看二者之间是否存在共变关系。在因果分析中，需要把变量分为自变量和因变量两种基本类型。自变量（independent variable）是引起其他变量变化的变量，或者说是用来解释其他变量变化的变量，因此，有些

① 在本书中，“变异”与“变化”同义，有时交替使用。

② Blalock, H. M. 1961. *Causal Inferences in Nonexperimental Research*, p. 9. Chapel Hill: The University of North Carolina Press.

学者称之为解释变量（explanatory variable）。因变量（dependent variable）则是被其他变量引起变化的变量，或者说是被解释的变量，有些学者亦称之为回应变量（response variable）。

如果用 X 表示自变量，用 Y 表示因变量，则可以把这种因果关系表示为 $X \rightarrow Y$，箭头“→”代表影响的方向。

探讨社会现象或不同变量之间的因果关系之所以重要，主要是因为通过研究因果关系不但可以建立起系统化的、有效的关于社会世界的知识（即认识社会世界），而且可以为制定政策干预社会发展过程提供科学依据，以便预测、控制人类的某些社会行为。所以，老一辈社会学家李景汉认为，“社会调查的最大使命，是发现社会现象因果的关系”①。

那么，怎样探讨社会现象之间（不同变量之间）的因果关系呢？通过收集结构化的经验资料来检验因果假设的方法有两种：一种是实验的方法，另一种是问卷调查研究的方法。

实验法通常用在自然科学研究中。在最简单的实验研究中，研究者首先将研究对象分为各方面的特征都相同的两个小组（一般采用随机分配的做法来分组），即实验组和控制组。然后，研究者给予实验组以实验刺激，再观察和测量它产生的结果，并与没有接受实验刺激的控制组的结果进行比较；如果两者之间出现了差别，则认为这种结果上的差别是由实验刺激上的差别（接受与不接受刺激）引起的。②

可以举一个简单的例子来说明这种因果推断的逻辑。假定在两块相同的耕地（彼此之间应该相隔一定的距离，以免互相影响）里分别栽种相同的玉米甲和玉米乙，在其他耕作条件也都相同并且都不存在病虫害的情况下，给玉米甲施加某种肥料 X，却不给玉米乙施加这种肥料。过了一段时间（比如两周）之后，我们再观察和测量这两株玉米的生长情况。如果二者之间的生长情况出现了差别，实验者就有信心认为这种生长情况上的差别是由施肥上的差别（施加与不施加肥料 X）引起的——因为实验

① 李景汉：《实地社会调查方法》，北平星云堂书店印行，1933，第12页。

② 对实验设计的详细讨论，可参考 de Vaus，D. A. 2001. *Research Design in Social Research*，chapter 4 – 6. London：Sage Publications Ltd.。