中国社会学函授大学教材之八

# 社會研究中的統計分析

（内部教材，不得翻印）

中国社会学函授大学编印

# 目　录

（上　册）

# 第一章　社会调查方法简介

## 第一节　社会调查与科学研究

一、社会调查，指的是根据实际工作或社会科学研究的需要，有目的的对社会上的特定现象作系统的全面的考察和了解，并从中得到有意义的结果和推论，以便对我们所从事的工作有所裨益。

社会调查也是一门学问。特别是在当代，随着社会信息量的大量增长，对信息在社会组织、管理、抉择、决策等方面价值认识的提高，研究如何科学而及时地从社会提取信息，进行正确的分析和处理，已不仅是社会学工作者关心的问题，就是对于广大的非社会学工作者，诸如形势分析，政治思想动态，社会评论，经济预测等部门的同志，也是极为重要的。在当前的改革中，更要及时了解人们在物质与精神两方面的心理与愿望。

社会调查方法的内容是很多的。它既包括具体方法的研究，例如个案、问卷、观察等等，也包括宏观方法论方面的研究，因此是多层次的。

要作好一项科学的社会调查，必须做好以下三方面工作：一是正确的课题选择；二是确当的科学研究方法；三是成果的科学总结和表叙。以下分述之。

（一）正确的课题选择，必须具有研究价值。社会学除了少部分研究社会学理论外，大部分社会学研究的课题，都是来源于当前的社会现实和要解决的实际问题。因此，社会学的研究就是直接为四个现代化服务的。例如，为配合我国当前人口政策的贯彻执行，社会学工作者进行了大量生育意愿的调查与分析。为配合

价格体系改革的顺利进行，了解群众对于价格改革的社会心理反应，中国经济体制改革研究所与北京大学社会学系对全国十一个市、十六个县、郊区开展了大型社会问卷调查。获取了大量有价值的信息，为有关领导的决策提供了可靠的依据。

（二）常见的研究方法共有四种：

1．实验法：

它的研究方法与在实验室进行科学实验是相同的。所不同的是社会现象中的因素控制不及自然科学易于实现。当研究A变量的影响，为了使A变量之外的因素得以控制。一般要建立条件相同的两个组：实验组和控制组。实验组改变A变量，制控组不改变A变量。然后比较A变量变化前后两组的差别。实验法虽然逻辑严谨，但社会学中较少运用。它常见于教育、心理方面的研究。例如教学法的研究等等。

2．参与研究法：

为了对某一地区、某一民族进行实地调查，与所在地群众同吃、同住、同劳动，打成一片。这样可以收集到很多深入的调查材料。这方面我国累积了丰富的调查研究做群众工作的好经验。

3．内容分析法：

它是通过某一时期报章，杂志、报告对某种问题讨论的频次和态度来分析当前倾向性的看法。

4．调查研究法：

由于社会研究中，很难像实验室里，人为地改变某些因素，因此一般不采用实验法而采用调查研究法。社会调查法的特点是不改变社会现状，就地取材，互相比较从中研究因素之间的关系。它是社会学中最常用的研究方法。社会调查又分全体调查和抽样调查。全体调查是调查全部个案，而抽样调查是从全体中抽出一部分调查，然后根据抽样调查结果推论到全体。抽样调查是较大范围进行社会研究最常用的方法。本书将主要介绍抽样调查。

（三）成果的总结与表叙

为了有效的用文字表达研究成果，目前研究报告或论文的书写已渐趋规范。它在撰写方法上已有了一套比较定形的格式。以后我们将在有关章节中介绍。

二、社会调查研究步骤

现以抽样调查为例，说明社会调查大致有如下步骤：

（一）根据需要、确定研究课题

例如为什么当前会出现大龄女青年找对象难的问题？为什么独生子女政策城市比农村易于推行？当前体制改革中，人们对物价、工资、人材流动、退休一系列问题的社会心理反应如何？等等都是研究的课题。

（二）了解情况

通过查阅文献和向有经验、有知识的人，了解本课题已有的进展。同时，更重要的还要向社会进行初步探索，开展个别访问、小组座谈，了解人们现实的想法与动态，取得第一手资料。

（三）建立假设

在前两步的基础上，明确研究的范围。并在初步探索的基础上，提出一定的想法和建立假设。

例如，我们确定研究的课题是当前有关独生子女问题。目前由于家庭子女数显著减少，出现了家庭十分偏爱子女的情况。这种“子女偏重”的现象。根据初步分析，它与“父母的生活价值观”；“父母一切为了孩子的感情强度”；“生育意愿是否满足”；“父母对子女的期望”和“溺爱子女的程度”有关。

在此基础上建立的模型是：

父母生活价值中孩子的地位越高，一切为了孩子的感情越强，生育意愿越得不到满足，对子女期望越高和溺爱子女的程度越高则“子女偏重”的现象越严重。

模型建立了现象与现象之间的关系，但须要强调的是这种关

系只是一种假设或想法。最终能否确立，必须经过实际的检验。

模型中的关系有两种表达方式：

1．函数式：A高则B高；（正比）

A高则B低；（反比）

2．差异式：A不同则B不同

例如：择偶标准男、女有别。

（四）确定概念的定义和测量方法

例如，前例所涉及的概念有“生活价值”；“感性强度”；“生育意愿”；“对子女的期望”；“溺爱子女”和“子女偏重”。这些概念首先必须给出抽象的明确定义，以免发生误解。举例说，所谓“子女偏重”，这里指的是经济上过分偏重子女、感性上过分依恋子女和生活上过分优先子女。但仅有以上抽象的定义还不够。为了进一步定量化处理的需要，还必须给出这些概念的具体变量方法。例如，我们通过父母闲暇化在子女身上的时间来度量生活上优先子女的程度。这称作概念的操作化定义。所谓操作化定义（operational definetion）或许是翻译不够传神，它的意思是通过抽象概念的操作化，使之定性研究和定量化之间建立起了桥梁。通过概念的操作化，实际是非定量的概念得以进行运算（operation）了，可以测量了。可见，概念的操作化是定量研究的飞跃和艺术所在。下节我们还将讨论它。

（五）在概念操作化的基础上，设计问卷。

问卷是指一组与研究目标有关的问题。这些问题则是根据概念操作化所提出的。问卷包括的内容一般有

1．事实：被访人的年龄、性别、职业、文化程度等等。这些在问卷中属于基本资料，在分析资料时，往往当作基本的自变量来考虑。

2．态度与看法：例如对某种行为、政策是否赞成，对某种职业的评价。

3．行为趋向：这种问题具有假设性。要了解某一情况下，

被访者会有什么样的行为。

4．理由：了解被访人采取某种态度和行为趋向的原因。

问卷的回答有两种方式：固定答题式和自由答题式。固定答题式一般是多种答案选择。这种问卷在大规模调查中经常使用。固定答题中答案的设计，取决于研究人员对问题实际情况了解的深度。为此，在探索性研究阶段，不妨采用自由答题式，以便收集到更多的活思想、新情况。

（六）试填问卷

把问卷发给调查中的少数人试填，以便问卷设计不周或遗漏之处，尽量在试填阶段予以纠正。否则，当大规模调查一经开始，纠正起来将相当困难，甚至成为不可能。这点凡具有实际经验的人，都知道它在调查研究中的实际价值和不可缺少性。

（七）抽样调查

根据抽样原则，科学、合理的选择抽样对象。它是进一步分析的依据。鉴于它在抽样调查中的主要性，今后还有专门章节讨论。

（八）培训访问员，发放问卷

调查访问成功与否，取决于被访者能否有效的合作，提供可靠的资料。而访问员的表现往往会影响被访者的合作。因此对访问员要进行挑选和培训。同时研究人员在问卷调查中，自己也要参加一部分实地调查。以便及时发现问题，指导访问员和对访问员的调查进行质量检查。

（九）对回收问卷进行校核、登录或将资料存入计算机

（十）统计分析

回卷回收之后，数据、资料成千上万。这时如果不能有效的整理、分析资料，提取隐藏在数据内部的规律。往往使人有淹没在数据之感而同时又感到信息之饥饿。统计技巧可以对数据进行可靠的分析与检验。这方面内容，下面章节还要介绍。

（十一）最后根据实际资料的统计分析，检验最初在第三步

骤中所建立的假设是否成立；对研究的社会课题提出有益的建议；并在现有研究的基础上，提出进一步研究的方案。从而一个新的研究循环又开始了。

三、社会调查与科学研究的基本方法

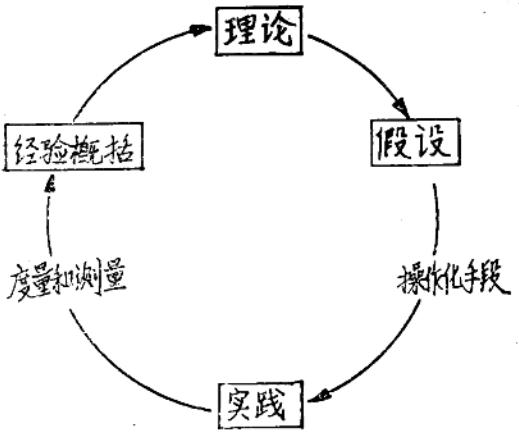
通过以上社会调查步骤的介绍，可以看出社会调查是遵循科学研究的基本方法的：即来经于实践而又回到实践中去、理论联系实际的方法。纵观上述全部步骤，它包括了科学研究中不可缺少的两个层次：抽象层次和经验层次：

科学研究方法

| | | | | |
|---|---|---|---|---|
| 抽象层 | | 概念 | 命题 | 理论 |
| 经验层 | | | | |
| 研究设计 | 原则： | 观察<br>量度 | 测定<br>分析 | 测定<br>分析 |
| | 数据： | 搜集 | 使用计算机对数据记录、储存、管理与分析。 | |

抽象层次包括前面所列举的1，2，3步骤，主要目的在于确定课题、概念以及概念与概念之间假设的关系（命题）或一组命题。但作为科学研究，仅此还不够，还必须得到经验层次支持与证实。为此，必须搜集数据。而概念与搜集之间，必须通过观察与量度才能使研究得以量化。这就是研究设计的概念操作化。作为研究，一般都要分析社会现象与现象之间的关系，因此，在对概念进行操作化，搜集数据的同时，还要对所假设的命题或理论进行测定、分析和检验。这些统计工作涉及的数据量都比较大，一般都由计算机完成。

下面我们用一个简化的循环图来说明社会学研究称之为科学研究的进程




可见，社会学理论和假设是指导我们应该收集那些资料，它是研究的基础，是定性的研究。而实践和经验概括则要解决资料如何收集，如何整理，如何分析和如何推论。因此，社会统计学的研究对象是数据的收集、整理、分析和推论。而社会调查方法论研究对象则是理论，假设和操作化手段。本书主要介绍社会统计学，但为了了解社会统计学在研究中的地位，在本节介绍了社会调查方法全过程。同时为了使没有专门学习过调查方法的读者，也能对调查方法略知一、二和能完整的尝试社会学的研究。下节将就调查方法的核心环节“概念和概念的量变”进行介绍。

## 第二节　概念与概念的量变

### 一、概念

概念是科学研究的基石或基本组成单位，它犹如大厦建筑中的砖瓦。因此，在社会研究中，首先必须发现和选用适当的概念来代表所研究的现象。例如权力、社会地位、态度、社会化、角色等等。概念是从个体中抽象出来的共同属性。一般没有时间和

空间的限制。概念一般都是抽象的。抽象的层次越高，所能概括的现象越高，代表的信息也就越多，时同也就越不具体。

概念不仅包括一般人所习惯了的术语，例如性别、职业、出生地等等，同时还包括研究人员根据研究的需要所设计出来的抽象概念，因此在使用之前，必须先给出明确定义，例如社会学中常用“角色”这样的概念。但对它的理解可以是三方面的：客观上对行为者的要求；行为者主观上对自己的要求以及行为者的实际行动。因此在使用之前，必须说明你所使用的“角色”指的是那一种定义，以免混淆。

二、变量

概念或属性在数量或质量上往往有所变化。或者说，概念的表现形式是取多种可能。这种多种可能性称作随机性。因此，概念是随机变量，或简称变量。例如：性别是变量，那么，男、女则是该变量两种可能的取值。它表示一个人的性别只可能是男或是女。

| 变　　量 | 变量的可能取值 |
|---|---|
| 性别 | 男；女 |
| 家庭子女数 | 1个；2个；3个…… |
| 重　要　性 | 非常重要；较重要；<br>重要；不太重要；不重要。 |
| 文化程度 | 大学；中学；小学；不识字 |

变量取值的选择要注意以下两点

（一）完备性：即变量的取值要一一无遗的列举出来。例如性别，它的取值只有男和女，因此是完备的。

（二）互斥性：即变量的各数取值是互斥的。不能产生使观察即可属A类同时亦可属B类的情况。例如，如果把收入分为三挡，它表示三种取值：0；0—100元；100元以上。那么，收入恰好是100元，就有两挡可以归入，因此这样的取值就不互

斥。

## 三、概念的量变——操作化定义

### （一）概念的抽象定义和操作化定义

概念是抽象层次的，在抽象层次必经给出明确的定义。下面举例来说明概念及其抽象定义。

| 概　念 | 抽象定义 |
| --- | --- |
| 智　力 | 抽象思维的能力或适应环境的能力 |
| 都市化 | 现代都市的生活形象 |
| 个人现代化 | 一个人由于经济、工业等现代工业因素影响所产生的内部变化 |

这些抽象化定义，只能帮助我们理解所要研究的概念指的是什么，但却无法进行具体变量，因此也就无法进行定量化处理。为了使概念过渡到可以变量，必须通过具体的指标来摸拟它。它诸如自然科学中的间接测量。所谓指标，是一串可以观察和可以测量到的具体数字或对某些具体问题赞成的态度。它属于经验层次。例如：人数多少？资金多少？是否赞成“养儿防老”这样的看法……等等都可称作指标。

概念或变量通过指标得以量化和度量，而指标或一组指标的综合则代表了所要测量的概念的操作化定义。下面是概念及其操作化定义。

| 概　念 | 操作化定义 |
| --- | --- |
| 工厂规模 | 工人数；资金；设备；面积 |
| 社会经济地位 | 职业；收入；教育 |
| 都市化 | 妇女就业人数；子女数；家庭住宿形式 |
| 现代化 | 对时间；效率；家庭；亲属；消费；自信等具体问题的看法。 |

（二）概念与操作化定义之间的关系

概念的操作化定义类同于自然科学中的间接测量。例如人们通过测量水银柱高度的变化去测量温度的变化。但间接测量手段是不唯一的。除了水银还可以用酒精。同样，操作化定义对于同一概念也不是唯一的。两者只是相似关系而不是恒等关系。因为无法排除间接测量所带来的误差。但一个好的操作化定义应尽量摸拟和包含抽象定义的内容，例如，为了解被访者是否喜爱看电影，如果问：

“每星期你平均看几次电影？”

就比问：

“昨天你看电影了吗？”

反映概念“喜爱看电影”准确得多。

操作化定义并不是很好下的，如果说统计分析还可以借助于统计工作者协作的话，那么，操作定义的设计则必须是由课题研究人员自己完成的。而操作定义设计的好坏则取决于研究人员对本课题理解的深度，情况的掌握以及研究的素质和艺术。下面介绍几个国外操作化定义很成功的实例。

四、操作化定义实例

态度是行为的主要因素。以前一直不能数量化。二十年代开始，很多人开始探索态度的变量方法。态度不是单一问题、单一指标所能代表的，它往往需要多方面因素的综合。例如前面所提到的现代化观念，它涉及到人们对时间、效率、家庭、亲戚、消费、自信等多种因素的综合。下面介绍 Likert 量变法，又称总和尺度法。

（一）收集大量能反映概念的有关语句。例如为了度量孩子的价值，可有这样问题：

1．孩子可以增进夫妻间的感情

2．养大一个孩子是件很辛苦的工作

3．有孩子可以“养儿防老”

4．生育孩子是对社会应尽的责任。

5．没有孩子的家庭是很寂寞的

6．生孩之前，夫妻应慎重考虑孩子会给大人工作、学习各方面带来的不便。

根据例句赞成的程度，分为五等，最高等“很赞成”给5分；最低等“很不赞成”给1分。

（二）小范围内进行试填。剔除那些总分差别不大的，筛选出看法差别大的问题。作为问卷中要询问的问题。

（三）设选中的问题共几个。那么几项问题得分之和：

$$A = a_1 a_2 + \cdots a_n$$

就是被访者所测概念的量化值。

设以下是某问卷的回答：

| 问题号 | 很赞成（5分） | 赞成（4分） | 不肯定（3分） | 不赞成（2分） | 很不赞成（1分） |
|---|---|---|---|---|---|
| 1 | | | | | |
| 2 | | | | | |
| 3 | | | | | |
| 4 | | | | | |
| 5 | | | | | |

某人得分 = 4 + 5 + 3 + 4 + 5 = 21（分）

五、小结

本节介绍了概念、变量和操作化定义。一个科学研究往往涉及一个以上的概念。而一个概念的量化，往往须要通过一个以上具体指标的摸拟和对应。这就称作概念的操作化定义。操作化定义的设计必须要包含抽象定义中的内容。一个好的研究，必须是概念、操作化定义和统计技巧诸方面的结合。

最后须要提醒的是必须注意分析单位。如果分析的单位是个

人，称作个人变量。如果分析的单位是群体，则称作群体变量。两者在分析和推论中不能混淆。也就是说，以个体单位研究的结论不能推论到群体，反之亦然。试想，穷队生育率高，并不表示穷队中穷人生育率比其富人高。很可能它的富人生育率也高。

## 第三节　社会调查资料的特点和数理统计学的运用

以上介绍了社会调查研究的全过程，以及非量化的抽象概念如何转化为量化的操作化定义，从而实现了从抽象层次向经验层次的过渡。下面将研究根据这些量化的操作化定义所收集到的资料，具有那些特点以及在进行分析时需要采用数理统计学的原因。

（一）社会调查资料的特点

1．随机性

客观现象可以分作确定性现象和非确定性现象。例如，物体在重力作用下的降落是确定性的。我们只要知道物体开始降落时刻的高度和速度，就可以完全肯定的预言在随后任一时刻的运动情况。同样，水在常压下，加热到100°C必然沸腾，这也是确定性现象。对于确定性现象，其因果关系可归纳为

“若A，则必有B”。

A与B之间，存在着确定性的函数关系

$$B = f(A)$$

和确定的函数图形（图1—1）

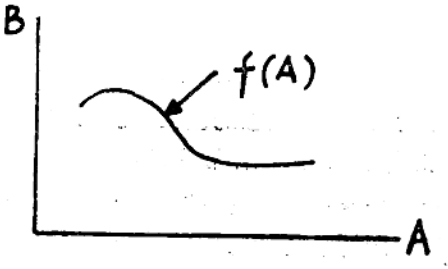



图1—1

但对非确性现象，其因果关系则可归纳为：

“若A，则可能有B，

但也可能：C

D

E。”

A与B之间，表现为非确定性关系。A和B之间虽然没有确定的函数关系和确定的函数图形，但A和B之间，仍然存在某种连系，其图形为图1—2


图1—2

通过散点图（图1—2），我们仍能看出A与B之间的连系。例如，身高与体重之间之关系就表现在为如上的散点图。

现在回到我们的社会现象中来，实际上，大部分社会现象都属于非确定性现象。现象与现象之间连系的命题也往往是非确定性的。我们不能像水到100°C必然沸腾那样来预言人到了某一年龄必然结婚。同样，也不能用抽查一滴水而知所有水成份那样，抽查一部分人的情况就知道全体人的情况。下面举例说明：

例一：下面列举了某厂全部女工的结婚年龄。假设总数N=100

| 人名代号 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 结婚年龄 | 25, | 25, | 24, | 27, | 25, | 26, | 24, | 28, | 27, | 26, | 25, | 25, | 26, | 22, | 21, |

| 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 25, | 25, | 27, | 22, | 24, | 26, | 27, | 28, | 24, | 26, | 27, | 27, | 25, | 26, | 27, | 28, | 27, | 24, | 27 |

| 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 26, | 29, | 27, | 22, | 22, | 19, | 24, | 27, | 26, | 24, | 20, | 30, | 26, | 25, | 24, | 28, | 32, | 25, | 26 |

| 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 24, | 25, | 24, | 19, | 25, | 25, | 27, | 23, | 30, | 21, | 25, | 28, | 19, | 24, | 26, | 27, | 25, | 25, | 26 |

| 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 24, | 22, | 26, | 28, | 25, | 25, | 26, | 26, | 26, | 25, | 24, | 25, | 25, | 27, | 26, | 24, | 22, | 23, | 26 |

| 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100。 |
|---|---|---|---|---|---|---|---|---|
| 28, | 26, | 24, | 28, | 26, | 25, | 25, | 27, | 24。 |

全厂女工的平均结婚年龄（总体平均值）：

$$\overline{n} = 25.26\text{岁}$$

现在如果进行的不是全体统计，而是抽查。例如从中任意的抽查10名，并计算抽查的平均结婚年龄。并假设这样的抽查共进行了四次。于是有：

$$\overline{n_1} = 25.9\text{岁}$$
$$\overline{n_2} = 25.7\text{岁}$$
$$\overline{n_3} = 25.5\text{岁}$$
$$\overline{n_4} = 26.1\text{岁}$$

可见，四次抽样结果相互都不相等，且都不等于总体的平均值：

$$\overline{n_1} \neq \overline{n_2} \neq \overline{n_3} \neq \overline{n_4} \neq \overline{n}$$

读者如果有兴趣，不妨自己也试想一下：把人名代号作成100阄，充分搅乱，从中摸10个，计算它的平均结婚年龄。

从上面四次抽样结果可以看出，对于社会调查资料，不存在

局部平均值等于总体平均值的公式。这是和确定性现象，例如，化验一滴水的成份就知道所有水的成份，不同的。

例二，以下列举某厂职工对独生子女的看法。其中括号内的人名代号表示不赞成独生子女的。假设男、女总数都是100名

男：

1，2，3，（4），5，6，7，8，（9），10，（11），12，13，14，（15），（16），17，18，（19），20，21，22，23，（24），25，26，27，（28），29，（30），31，32，33，34，35，（36），（37），（38），39，40，（41），42，43，44，45，（46），47，48，49，（50），51，52，53，（54），55，（56），57，58，（59），60，（61），（62），63，64，65，（66），67，68，69，70，（71），72，73，74，（75），76，77，78，（79），80，81，82，83，（84），85，86，87，88，（89），90，91，（92），93，94，（95），96，（97），98，99，（100）。

女：

1，（2），3，（4），5，6，7，（8），9，10，（11），12，13，14，15，（16），（17），18，19，20，21，22，23，（24），25，26，（27），28，29，（30），31，32，（33），34，35，36，（37），（38），39，40，（41），42，（43），44（45），46，47，48，（49），50，51，52，53，（54），55，（56），57，58，59，60，（61），62，63，64，65，66，（67），68，69，70，（71），72，73，74，（75），76，77，（78），（79），80，81，82，（83），84，85，（86），87，88，（89），90，91，（92），93，94，（95），（96），97，98，99，100。

于是有：