



SHEHUI YANJIU
DE TONGJI FENXI

社会研究的统计分析

李沛良 著

湖北人民出版社

E, TONGJI FENXI

社会研究的统计分析

李沛良 著

社会科学出版社

社会研究的统计分析

李沛良 著

*

湖北人民出版社出版、发行 新华书店湖北发行所经销

湖北省新华印刷厂印刷

850×1168 毫米 32 开本 11.375 印张 2 插页 283,000 字

1987 年 1 月第 1 版 1987 年 1 月第 1 次印刷

印数 1—1,800

统一书号：3106·750 定价：2.45 元

ISBN 7—216—00000—5 /C·1

目 录

第一篇 导 论

第一章 社会研究历程	8
第一节 筹划	6
第二节 执行	21
第三节 总结	26

第二篇 统计叙述:

单变项与两变项

第二章 简化一个变项之分布	33
第一节 基本技术	33
第二节 集中趋势测量法	42
第三节 离散测量法	48
第四节 正态分布与标准值	57
第三章 简化两个变项之分布	62
第一节 统计相关的性质	62
第二节 交互分类与百分表	65
第三节 简化相关与消减误差	70
第四章 相关测量法与测量层次	74
第一节 两个定类变项: Λ , τ_{yx}	74
第二节 两个定序变项: γ , D_{yx}	80
第三节 两个定距变项: 简单直线回归与积矩相关	91

第四节	定类变项与定距变项：相关比率与非直线相关	99
第五节	定类变项与定序变项：Lambda, Tau-y	103
第六节	定序与定距变项：相关比率	105
第七节	综合	107

第三篇 统计推论：

单变项与两变项

第五章	抽样与统计推论	113
第一节	抽样的意义与问题	113
第二节	抽样的历程	114
第三节	随机与非随机抽样法	117
第四节	机率与抽样分布	123
第六章	参数值的估计	130
第一节	点值估计与间距估计	130
第二节	间距估计：均值、百分率、积矩相关	131
第三节	决定样本的大小	136
第七章	假设的检定：均值与百分率	140
第一节	基本知识	140
第二节	单均值与均值差异	150
第三节	单百分率与百分率差异	161
第八章	假设的检定：两个变项之相关	165
第一节	X^2 检定及其相关测量法	166
第二节	Gamma 及其他级序相关的 Z 检定	173
第三节	单因方差分析与 F 检定	177
第四节	积矩相关与回归系数的 F 检定	182
第五节	综合	185

第四篇 多变项分析

第九章 详析模式与统计控制	191
第一节 因果分析	193
第二节 阐明分析	202
第三节 条件分析	205
第四节 净相关系数	208
第五节 其他的净相关测量法	214
第十章 多因分析	219
第一节 复相关	220
第二节 多因回归	223
第三节 逐步回归	236
第四节 交互分类与虚构变项	239
第五节 典型相关	245
第六节 判别分析	248
第七节 多因方差分析	253
第十一章 因径分析	257
第一节 基本原则	257
第二节 完整单向模式	260
第三节 部分含蓄单向模式	268
第四节 缺径单向模式	270
第五节 结语	273
第十二章 多项相互关系	275
第一节 聚类分析	275
第二节 因子分析	277
第三节 最小空间分析	288
第四节 Q技术与R技术	291

第五篇 社会变迁分析

第十三章	趋势分析	295
第一节	横剖研究与纵贯研究	295
第二节	两个时期的趋势	296
第三节	时间系列分析	297
第四节	系列相关	301
第十四章	同组分析	304
第一节	转动分析	304
第二节	条件性转动	306
第三节	共同转动	307
第四节	同组分析之问题	309
参考书目		311
附 录		
(一)	随机数表	319
(二)	正态曲线的各部分面积	325
(三)	Z 检定: 常用的显著度(P)与否定域($ Z \geq$)	326
(四)	t 值化为 Z' 值	327
(五)	t 的分布	330
(六)	X^2 的分布	331
(七)	F 的分布	332
习 题		338
习题解答		344
英汉索引		347

第一篇 导论

1000

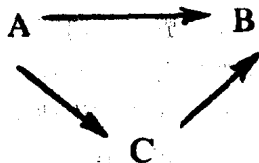
1000

第一章

社会研究历程

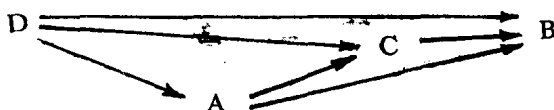
统计分析是社会学研究的一个环节。如要了解统计分析的作用，就要先掌握整个社会学研究的历程，包括怎样规划一项研究，怎样落实研究计划，和怎样总结研究的成果。本章的目的，是摘要地介绍社会学研究历程的来龙去脉，让大家有一个概括的了解以后才转入正题，介绍各种常用于社会学研究的统计方法和技术。

所谓社会学研究，就是运用科学的方法来搜集和分析社会事实，以理解社会现象之间的关系。比如有两个社会现象 A 与 B，社会学研究就要了解 A 与 B 是否相关，尤其是因果关系。以家庭关系不和 (A) 对青少年犯罪 (B) 的影响来说，可表示为：家庭不和 $\rightarrow$ 青少年犯罪，或 $A \rightarrow B$ 。当然，发生关系的社会现象，不一定是这么简单。家庭对青少年的影响，可能是通过另一个因素，例如结交不良分子 (C)，因而形成一个更复杂的因果模型 (causal model)：



这个模型表示家庭不和睦，除了会直接导致青少年犯罪以外，也会使青少年易于结交坏朋友，因而犯罪。进一步来说，这个模

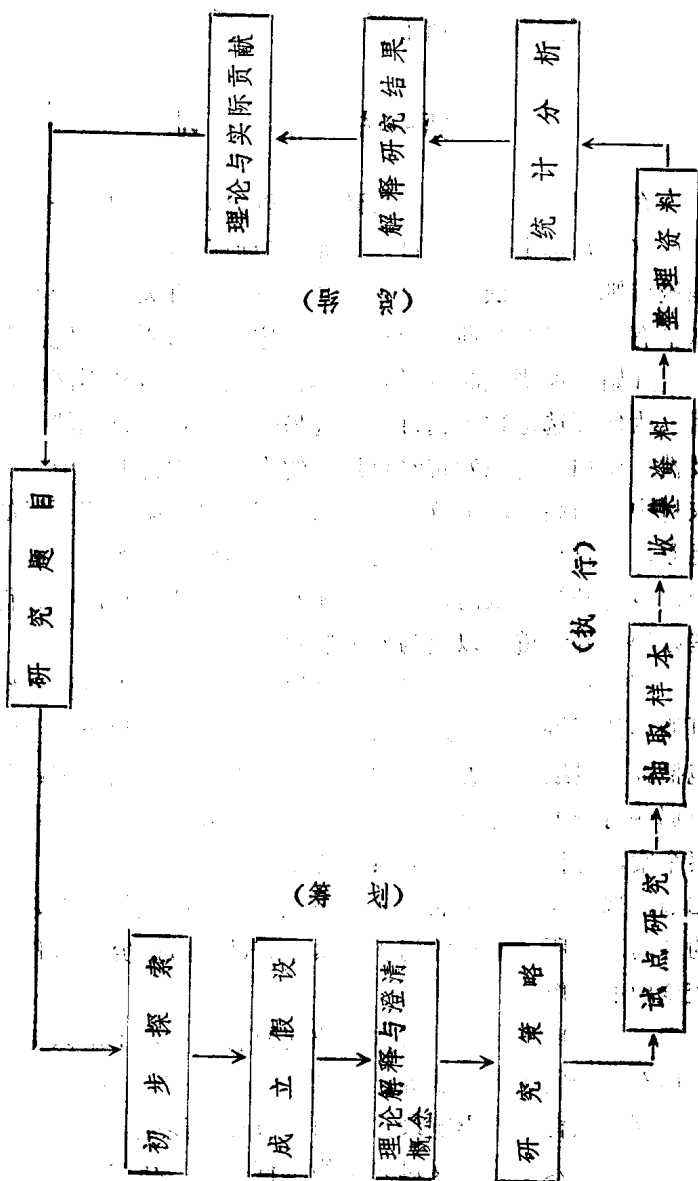
型所包括的社会现象及因果关系，也可能城、乡有别。如果考虑城乡情况(D)，则因果模型会变得更为复杂：



从上可见，社会学研究除了叙述社会现象间(A与B)的关系以外，也要力求解释相关的原因(C)，同时要注意到社会现象间的关系在不同社会情况下(D)的变化。问题是：社会学研究采用什么方法来叙述和解释社会现象间的关系？这就要运用科学方法了。所谓科学研究，就是运用客观的、逻辑的和系统的方法来搜集事实及分析事实。换言之，科学方法的一个基本特征是实事求是，不能单凭主观的或随意的构思。百多年来，世界各地的社会学界创立了很多理论，这些理论可能是源于个人的社会经验和感受，但必须经过科学的验证，获得事实的支持，才可得到他人的信赖。

采用科学方法的社会学研究，如何搜集事实？如何进行分析？社会学研究的整个历程，大致上可分为三个阶段：(1) 筹划，(2) 执行，(3) 总结。换言之，开展一项研究时，必须首先规划每一步的研究工作，然后才付诸实践；执行各步研究计划以后所得的资料，又必须简明地总结起来，才算完成该项研究。具体讲，每个阶段又分为若干的小步骤。从附图1.1所列的各项步骤，可见社会学研究在开始时是要面对一个问题，随之而来的各项步骤就是要解决这个问题。但是，在研究终结时却又会出现新的问题，让将来的研究再通过各项步骤来追求解答。社会学的知识，便随着这个既循环而又有累积性的科学研究历程，不断地壮大起来，使我们对人类社会的了解愈来愈广泛和深入。下面各节，撮要地

图 1.1 社会学研究之阶段与步骤



介绍每一阶段中的各项步骤。

第一节 筹 划

在规划一项研究时，大致上可以分为下面的五个步骤：

（一）选定研究题目。比如，为何有些城市犯罪率高，而有些城市则犯罪率低？这是否与其人口情况有关，或与其一般教育水平有关，或与其刑罚轻重有关？这就成为所要研究的题目。

开始一项研究时，虽然一定要选个题目，但又不能随便定个题目。为何要选定这个题目而不选另一个题目？首先就要说明其重要性。比如，现在发现农村地区的许多中学都办高中。如何办好农村中学的高中，十分重要，这可使农民子弟受到更好的教育。但是，农村高中有办得好的，有办得不好的；例如，有升学率高的，也有升学率低的，为什么呢？“为何有些农村高中办得好，有些办得不好？”就可以定为研究题目。

（二）初步探索。选定题目以后，我们对于所要研究的事情可能不甚了解，就要进行初步的探讨。首先是收集有关的文献，包括翻阅报纸、杂志、档案资料、前人的研究成果等；其次是咨询那些对研究的题目有经验、有知识的人，进行了解；其三是观察个案，即挑选一些较为典型的个案，亲自观察所要研究的现象。比如，选定四、五所农村中学去观察，其中包括办得好的和不好的各两、三所。上述的初步探索，有助于对研究题目作较深入和较全面的了解。

（三）成立假设。对选定的问题有了初步了解以后，就要把自己对这个问题的想法提出假设 (hypothesis)。例如，对农村地区的高中教育素质影响最大的两个条件可能是：民办教师和公办教师的比率及学生的父母之平均教育水平。在这个构思的基础上可以成立如下的假设：

(1) 如果公办教师的比率愈大，则学校的办学成绩愈好。

(2) 如果父母平均教育水平愈高，则学校的办学成绩愈好。

所谓假设，就是根据我们对问题的了解，假定现象与现象之间的关系。每一种现象均有变化，例如年龄有老、中、青之分，民族背景有汉、蒙、回、藏等之别。故此，社会学研究的社会现象，一般称为变项 (variables)，而假设就是假定某一种现象的变化（即 A 变项）与另一种现象的变化（即 B 变项）具有某种关系。假设的方式，可以书写成函数式或差异式。函数式是：

如果 A 愈高，则 B 也愈高；或
如果 A 愈低，则 B 愈高

函数式要求变项之数值有高低之分，如：学校的办学成绩、父母的教育水平等。但有的变项不存在高低的性质，如民族、性别等，只能进行分类研究，这时就要使用差异式写法。差异式是：

A 不同则 B 不同

在书写时，应该尽量指出如何不同。例如性别与成绩的关系，若写成为：“成绩高低，男女有别。”这个假设就不够好，没有说清楚二者之关系。若写成：“男（或女）学生之成绩高于女（或男）学生的成绩。”这就清楚了。

以事实来验证所成立的假设，是科学研究的目标。在成立假设以后的各项研究步骤，就是为求知道假设是对还是错。因此，在成立假设时，必须尽可能说清楚变项与变项之间的关系会是怎样的。如果研究的目标（即假设）含糊不清，我们便难以规划和组织整项研究工作。当然，所成立的假设，可以是一个，也可以是很多个，视乎研究的需要。同时，研究的假设，一般是说明两个变项的关系，但也可以只假定一个变项的情况或假定多个变项的相关情况。

不是所有的社会学研究都成立假设的。国内、外都有些调查

研究，没有成立假设便进行搜集资料，然后汇报“事实”。这些研究有其贡献，尤其是当我们对实际情况缺乏了解，难以作出有意义的假设时。但是，由于事前没有充分的思考，根据资料而作出的结论基本上是属于“放马后炮”的做法，其正确性难以判断。如果在整个研究历程中有任何差错，所得的资料就会使我们妄下结论。科学的研究是要求事前想清楚，即先成立假设，然后力求推翻这个假设。不能推翻的话，便说这个假设“可能”是正确的；当然，其正确性尚待将来的研究再度检定。当前的社会学研究，为求符合科学的要求，大部分是争取先成立假设，才搜集事实来加以验证。

（四）理论解释与澄清概念。在成立假设以后，便要予以解释。比如假设为： $A \rightarrow B$ 。但为什么A会影响B呢？这就需要讲出理由来。在上面的例子中，我们可以说：学生的成绩不是天生的，也不是上帝赐给的，而是学习环境决定的，父母的教育水平及公办教师的比率是决定学习环境的重要条件，因而影响了学生的成绩。一般来说，公办教师的素质比民办教师好，能够给予学生较佳的指导。父母的教育水平高，就会使学生在家庭中得到较为优良的管教，因而在学业上的成绩会较好。当然，假设之解释是好还是坏，与我们的人生经验和学术水平有很大关系。

另外，对假设中的变项A、B等所包含的概念，要清楚界定它们的意义，以免发生误解，因为同一个词汇，往往有多种理解。我们对于每个变项内的概念所下的定义，要清楚和明确。例如父母的教育水平，我们可以界定为在学校读书的年期（即学校以外所得的教育不算）。此外我们也要说明什么是民办教师？公办教师？高中的办学成绩是指些什么？

（五）研究策略。怎样证明假设呢？要证明假设，必须首先设计研究方式，然后确定每个变项的测量方法及其测量层次，最后还要计划怎样抽取样本和用哪些统计方法来分析资料。

(a) 研究设计 (Research design): 较为常用的研究方式有两种, 一是实验法 (experiment), 一是社会调查法 (social survey), 皆可验证假设。

实验法的基本逻辑, 是有意地改变A变项, 然后看看B变项是否随着变化; 如果B变项显然是随着A变项的变化而变化, 就说明A变项对B变项有影响。这种方法的逻辑很严密, 操作程序如下: 首先将研究对象分为两组, 一组为实验组 (experimental group), 另一组为控制组 (control group)。为求证明假设: $A \rightarrow B$, 我们要改变实验组的A变项, 但不改变控制组的A变项, 然后比较两组的B变项在实验前、后的变化。如果变化的程度不同, 表示A对B有影响; 如果相同, 则表示无影响。具体来说, 假定实验前的实验组的B变项的数值为 b_1 , 控制组为 b_3 ; 实验后的实验组数值为 b_2 , 控制组为 b_4 。实验模型便如下图 (图中的a表示A变项的数值变动):

	实验前		实验后
实验组	b_1	a	b_2
控制组	b_3		b_4

依据上图, 实验组和控制组在实验前后的变化, 分别是 $b_2 - b_1$ 和 $b_4 - b_3$ 。那么, A变项的效果 = $(b_2 - b_1) - (b_4 - b_3)$ 。如果数值是0, 即无效果; 如果不是0, 则是有效果, 即表示A对B有影响。

将研究对象分为实验组和控制组时, 要特别注意的是两组的成员在实验前的情况 (例如性别与年龄的分布等) 要基本相同, 尤其是那些会影响实验效果的情况。否则, 即使我们发现两组的B变项的变化不同, 也不一定是A变项的效果, 而可能是由于两组原来的情况不同。那么怎样使两组的情况相同呢? 有两种方法,

一是随机法(randomization)、一是配对法(matching)。所谓随机法,是按照随机的原则,将研究对象分派两组中。换言之,何者属何组,完全听命于机遇,不受人为影响。实行时,可以用类似幸运抽奖的办法,也可以用随机数表(在本书第五章详谈)。理论上,用随机法分派可以使两组研究对象的各方面情况相同,但实际上不一定完全相同。为求保证两组在某些方面是相同,尤其是使那些甚可能影响实验结果的情况相同,我们可以采用配对法。所谓配对法,就是在分派时有意识地使两组在某些方面相同,例如年龄组合、男女比例等。这种方法适用于匹配的条件少,容易配合起来的研究对象。如果条件太多,难以配对,则要采用随机法。当然,配对法与随机法可以结合运用,以加强实验的效能;例如将研究对象分为男、女二组,然后将每组人随机分派到实验组和控制组。

现在举一个例子,说明实验研究。为求证明改善师资会提高农村高中学校的办学成绩,我们在某地区选择十所高中学校,用随机法分为实验组与控制组,各有五所学校。1980年至1982年间,实验组学校的师资获得改善,而控制组的师资无改变。从搜集的资料,发现在1980年实验组学校毕业生的平均升读大学比率是10%,控制组是12%;在1982年实验组是25%,控制组是11%。因此,实验效果 = $(25\% - 10\%) - (11\% - 12\%) = 16\%$,证明在实验过程中所用的改善师资办法,可以提高农村高中学校的水平。

上述的实验设计,也称为典型(classical)或理想(ideal)实验法,能够有效地验证假设。但有些研究受到限制,不一定可以采用这么典型的方式,便要改用较为简单的实验设计,例如不作实验前的衡量或不用控制组。

如果有实验组和控制组,但不作实验前的衡量,实验设计便如下图: