

SHE HUI TONG JI FEN XI



王宗烘 编著

社会统计分析

厦门大学出版社

厦门大学社会工作系列丛书

社会统计分析

王宗烘 编著

厦门大学出版社

(闽)新登字 09 号

厦门大学社会工作系列丛书

社会统计分析

王宗烘 编著

*

厦门大学出版社出版发行

福建省第二新华印刷厂印刷

*

开本 850×1168 1/32 7.5 印张 185 千字

1993 年 12 月第 1 版 1993 年 12 月第 1 次印刷

印数:1—2000

ISBN 7-5615-0828-X/O·49

定价:5.20 元

目 录

第一章 导论	(1)
第一节 社会统计分析的性质.....	(1)
第二节 社会统计分析的主要内容和作用.....	(3)
第二章 单变量描述性统计分析	(6)
第一节 简化资料的基本技术.....	(6)
第二节 集中趋势测量法	(14)
第三节 离散趋势测量法	(19)
第三章 双变量描述性统计分析	(27)
第一节 相关关系和相关测量法	(27)
第二节 两定类变量相关的测量	(31)
第三节 两定序变量相关的测量	(35)
第四节 两定距变量相关的测量	(42)
第五节 定类变量与定距变量相关的测量	(48)
第四章 概率和概率分布	(55)
第一节 随机事件及其概率	(55)
第二节 概率的基本性质	(58)
第三节 随机变量及其分布	(67)
第四节 随机变量的数字特征	(72)
第五章 几种重要的概率分布	(80)
第一节 二项分布及其它离散型分布	(80)
第二节 正态分布及其它连续型分布	(86)
第三节 大数定律和中心极限定理	(94)
第六章 参数估计	(99)

第一节	推断性统计分析	(99)
第二节	参数的点估计	(101)
第三节	抽样分布	(105)
第四节	区间估计	(109)
第七章	均值、成数和方差的假设检验	(123)
第一节	假设检验的基本概念及主要步骤	(123)
第二节	总体均值的假设检验	(128)
第三节	二总体均值差的假设检验	(132)
第四节	成数及成数差异的假设检验	(136)
第五节	方差及方差比的假设检验	(139)
第八章	两变量相关性的假设检验	(144)
第一节	两定类变量相互独立性的假设检验	(144)
第二节	两定序变量相关性的假设检验	(148)
第三节	两定距变量相关性的假设检验	(152)
第四节	单因方差分析	(157)
第九章	多变量分析	(163)
第一节	详析的几种模式	(163)
第二节	净相关系数	(169)
第三节	复相关系数	(173)
第四节	多因回归和逐步回归	(176)
第五节	虚构变量和典型相关	(181)
第十章	社会趋势分析	(185)
第一节	长期趋势分析	(185)
第二节	季节变动分析	(191)
第三节	复查分析	(196)
参考书目		(201)
附表		
附表一	二项分布表	(202)

附表二	泊松分布表	(206)
附表三	标准正态分布表	(210)
附表四	X^2 分布表	(212)
附表五	t 分布表	(216)
附表六	F 分布表	(218)

第一章 导 论

第一节 社会统计分析的性质

一、社会统计分析的性质

社会统计分析是一门关于如何运用统计学基本原理和方法对社会调查资料进行定量分析的学问。它是由社会学和统计学相互交叉而形成的边缘学科。

社会统计分析属于社会学，是社会学中的方法性学科。今天的社会学已不是一门简单的学科，而是一个门类众多的学科体系。这个体系包括社会学一般理论（社会学概论）、具体社会学（应用社会学）和社会学研究方法等三大部分。社会学一般理论的派别很多，我国的社会学一般理论是在历史唯物主义一般原理指导下建立起来的关于社会生活各个方面的基本理论。具体社会学是用社会学的观点和方法研究某一类社会现象和社会问题而形成的学科，如经济社会学、政治社会学、科学社会学、教育社会学、法律社会学、人口社会学、老年社会学、劳动社会学、城乡社会学、社会工作、社会管理、社会保障等。社会学研究方法也分为许多不同学科，如社会研究法、社会调查、个案研究、社会统计学、计量社会学等。社会统计分析属于社会学研究方法，是社会学学科体系中的一个方法性学科。社会学研究方法又可分为定性方法和定量方法两大类。社会统计分析属于定量方法，它通过对大量社会调查资料的统计分析来揭示各种社会关系的本质和变化规律。

社会统计分析不仅可以应用在社会学研究中，而且可以运用到其它社会科学的研究中，它已成为人们定量研究各种社会现象的一种重要工具。

社会统计分析也属于统计学，它的方法和技术是建立在数理统计基础之上的。统计学产生于 17 世纪，创始人是英国的经济学家威廉·配第和约翰·葛兰特。到了 19 世纪，由于概率论的引入，统计学发展成数理统计，并且广泛运用于社会科学和自然科学的经验层次研究。今天的统计学也是一个门类众多的学科体系，除了统计理论、数理统计，还有许多应用统计学（也称统计技术），如经济统计学、人口统计学、教育统计学、心理统计学、统计物理、化学统计学、生物统计学等。社会统计分析也属于应用统计学，是统计学学科体系中的一个分支学科。它的理论基础是数理统计，它的方法和技术是建立在以概率论为出发点的数理统计基础之上的。

二、社会统计分析与社会统计学的关系

社会统计学是一门应用统计学。就一般应用统计学的内容来说，一般包括统计调查的设计，调查资料的收集和整理，调查资料的分析和推论等。因此，社会统计学应包括统计性社会调查的设计，社会调查资料的收集和整理，社会调查资料的分析和推论等。社会统计分析则专门介绍对统计性社会调查资料的分析、推论的方法和技术。它是社会统计学的重要组成部分，是对社会统计学中分析和推论方法的发展。然而，由于社会调查理论已成为一个专门的学科，一些标以“社会统计学”的书，其内容就是社会统计分析，或以社会统计分析作为最主要的部分。

这里要强调指出，社会统计分析与社会调查的关系极为密切。社会调查的种类很多，有全面调查（普查）、抽样调查、典型调查、重点调查、个案调查、追踪调查等等。统计性的社会调查主要指

全面调查和抽样调查，它是进行统计分析的前提条件。只有做好统计调查，才能为统计分析提供较可靠的资料，为获得正确的结论奠定良好的基础。特别是推断性统计分析，它的资料来自随机抽样调查，只有实施满足随机性和代表性要求的抽样调查，才能使本质上属于或然性推论的统计推论获得有价值的结果。

第二节 社会统计分析的内容和作用

一、社会统计分析的主要内容

社会统计分析的主要内容可分为描述性统计分析和推断性统计分析两大部分。描述性统计分析包括统计图表的制作、集中趋势和离散趋势的测量、统计相关的测量等，它适用于对全面调查资料和抽样调查资料的分析。推断性统计分析（统计推论）是根据样本资料对总体的特征（集中、离散、相关等）进行推断，即估计总体中的参数，或者检验研究中所提出的关于总体情况的假设。它具体包括参数估计、均值和方差的假设检验、列联表的检验、等级相关的检验、方差分析、虚构变量分析等。它仅适用于对抽样调查资料的深入分析。

统计分析的内容一般是按单变量、双变量、多变量的顺序展开的。对于单变量，主要研究变量分布的表示以及变量分布的集中、离散特征等。对于双变量，主要研究它们之间是否存在关系以及关系的密切程度（强度）。对于多变量，除了研究两两之间可能存在的联系，还可以研究一因多果、一果多因以及网状、链状等形式的联系。本书主要介绍单变量和双变量的统计分析方法。

推断性统计分析要运用概率论的重要结论。本书将简要介绍概率论的基本知识，包括概率、概率分布、重要的概率分布以及中心极限定理等。

二、变量的层次及统计分析方法的选择

社会统计分析把各种各样的变量划分成定类、定序、定距、定比等四个不同的层次。

定类变量是最低层次的变量。它的取值只有类别属性之分，无大小、程度之分。如性别就是常见的定类变量，它的取值是男或女。从数学运算特性来看，它只具有等于或不等于（=、 $\neq$ ）的性质。

定序变量是较高一层次的变量。它的取值除了有类别属性之分外，还有等级、次序的差别。如教育程度就是常见的定序变量，它的取值是文盲、小学、中学或大学。就其数学运算特性来说，不仅有等于或不等于的性质，还有大于或小于（ $>$ 、 $<$ ）的性质。

定距变量是更高层次的变量。它的取值除了有类别、次序之分外，取值之间的距离还可采用标准化的距离去量度它。如智商就是典型的定距变量，智商为 100 的人和智商为 110 的人相比，后者的智商比前者多了 10 个单位。就其数学运算特性来说，除了等于、不等于和大于、小于之分外，还可以进行加或减（+、-）的运算。

定比变量是最高层次的变量。这种变量除了具有上述三种属性外，其取值还可以构成一个有意义的比率。如年龄就是常见的定比变量，30 岁的人和 10 岁的人相比，前者的年龄是后者的三倍。就其数学运算特性来说，除了可以进行上述各种运算，还可以进行乘或除（ $\times$ 、 $\div$ ）的运算。

由于只满足定距而不能同时满足定比要求的变量并不多，所以人们在实际应用中不再区分定距与定比，而是当成一类，并称为定距变量。另外，当定类变量只有两个取值时，常称为二分变量，并给予专门的研究。

统计分析的方法和技术多种多样，选择时既要看资料的性质

和变量的个数，还要看变量的层次以及一些其它的条件。例如，测量某个变量的集中趋势时，如果资料来自全面调查，变量属定距变量，那么可以计算其算术平均值；测量两个变量的相关关系时，如果资料来自抽样调查，两变量都属定序变量，那么可以先计算G系数（或d系数），然后用Z检验法进行假设检验。这里要特别指出，统计分析方法和变量的层次不是一一对应的，适用于较低层次变量的方法一般也适用于较高层次的变量。例如，求众值的方法是测量定类变量集中趋势的基本方法，也是测量定序变量、定距变量集中趋势的方法。为了尽量保留资料中所提供的信息，对较高层次变量的分析，一般不采用既适用于较低层次变量又适用于较高层次变量的方法。例如，测量定距变量的集中趋势时，通常是计算算术平均值而不是计算众值或其它。

三、统计分析在社会学研究中的作用

社会学研究的整个过程，大致可以分为筹划、执行、总结三个阶段。利用社会统计分析所提供的方法和技术对调查资料进行有效的处理，是总结阶段的一项极其重要的工作。同时，从三个阶段的密切关系看，社会统计分析的思想实质上贯穿于社会学研究工作的始终。

统计分析方法在社会学研究中的运用，可以把庞大的资料加以简化，使人易于理解；可以在已知样本结果的基础上，推论出总体的情况，从而使研究结论具有普遍性和概括性；可以对研究中所提出的理论、假设作出解释说明，对某种社会现象的发展趋势作出预测。此外，它还有助于人们提高对社会关系性质的认识，并促进应用社会学的发展。当然，我们不能夸大社会统计分析的作用。这一方面在于它不是社会学定量研究方法的全部，更不是社会学研究方法的全部；另一方面在于统计推论本质上属于或然性推论，其结论带有不确定性。

第二章 单变量描述性统计分析

第一节 简化资料的基本技术

对于统计调查所得到的大量数据资料,必须进行整理简化,使之成为集中、系统的资料,成为能够说明整个总体的资料。这一节先介绍简化一个变量的资料的基本技术。

一、定类变量:频次分布、比例及图示

1. 频次分布

人们在整理数据资料时,总要按照某种标志把总体(或样本)中的基本单位加以分配,以显示这种基本单位在各种类别中所出现的次数,即显示各种类别占有基本单位的数量。所谓频次分布,就是指总体(或样本)的基本单位分布在各类中的情况。各类中的基本单位的数量称为频次或频数,并用 f (或 n)表示。如果我们用表格来表示频次分布,就形成频次分布表。表2—1就是某工厂人员类型的频次分布表(表中的频次等同于人数)。

表 2—1 人员类型的频次分布

人员类型 (X)	频 次 (f)
工 人	320
科 技 人 员	80
行 政 人 员	20
总 数 (N)	420

2. 比例

比例指各类成分在总体中所占的比重。这里的比例，就是各类的频次在总频次中所占的比重，也就是相对频次。比例一般用符号 P 表示，因此我们有：

$$P = \frac{f}{N} \quad (2-1)$$

通过计算同一名称的类在不同总体中的不同比例，可以进行资料间的比较。例如，已知工厂甲有职工 420 人，其中科技人员 80 人；而工厂乙有职工 360 人，其中科技人员 76 人。根据式 2-1 可得：

$$P_{\text{甲}} = \frac{80}{420} = 0.190$$

$$P_{\text{乙}} = \frac{76}{360} = 0.211$$

显然，工厂甲科技人员在职工中所占的比例小于工厂乙的同项比例。

比例值也叫成数，在参数估计和假设检验中所说的成数都指比例值。

比例 P 可以看成是以 1 为基数的数值，即 $P/1$ 。如果我们把基数放大成 100、1000、10000 等，便得百分率、千分率、万分率等，可分别写成 $100P\%$ 、 $1000P\%$ 、 $10000P\%$ 等。以上各种比率都有自己的用场，其中百分率在社会学研究中较常使用。计算比率时，应该根据精确度的要求保留小数点后面一定位数的数字。在社会学研究中，通常是保留小数点后面一位或两位数字。

如果我们把频次分布表中的频次换成百分比（百分率），便得到百分比分布表。表 2-2 就是和表 2-1 相对应的百分比分布表。

表 2—2 人员类型的百分比分布

人员类型(X)	百分比(%)
工 人	76.2
科技人员	19.0
行政人员	4.8
总和百分数	100.0
(统计总数)	(420)

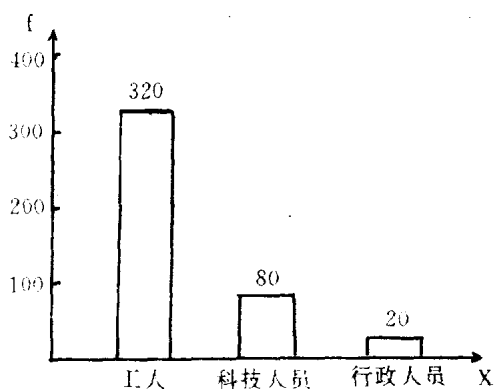


图 2—1

在进行资料间的比较时，还可以使用对比值，如

$$P_{\text{甲}}:P_{\text{乙}}=0.190:0.211=1:1.11$$

3. 图示法

频次分布和百分比分布还可以用图形表示。定类变量常用的图形有条形图和圆瓣图。

(1) 条形图

条形图用相同宽度的长方形的长短表示频次或百分比的大小。为清楚起见，每个长方形之间有一定的间隔。和表 2—1、表

2—2 相对应的条形图可以作成图 2—1、图 2—2。

(2) 圆瓣图

圆瓣图用圆平面中扇形的大小表示各个类在总体中所占比重的大小。和表 2—2 相对应的圆瓣图可作成图 2—3。

$$\alpha_{\text{工人}} = 360 \times 72.2\% = 274.3(\text{度})$$

$$\alpha_{\text{科技}} = 360 \times 19.0\% = 68.4(\text{度})$$

$$\alpha_{\text{行政}} = 360 \times 4.8\% = 17.3(\text{度})$$

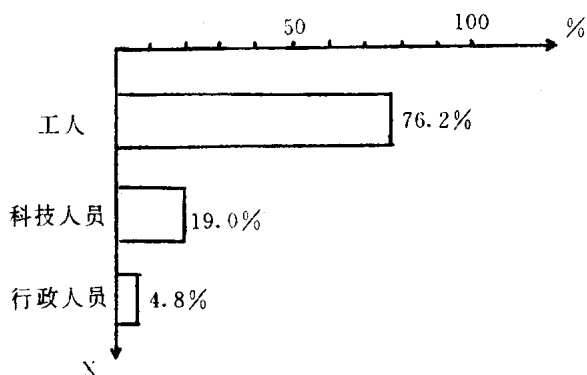


图 2—2

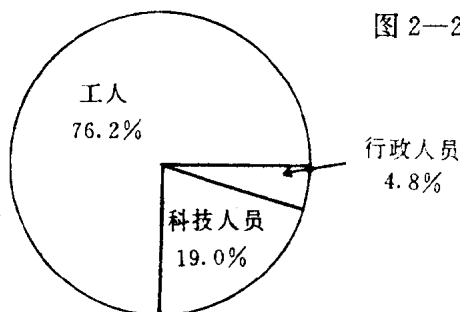


图 2—3

二、定序变量：累计频次和累计百分比

简化定序变量资料的基本技术，除了上面所介绍的，还有累计频次和累计百分比。

1. 累计频次

累计频次就是将各类的频次逐级相加起来，可以由高级到低级累加 ($cf \downarrow$)，也可以由低级到高级累加 ($cf \uparrow$)，其作用是使人们容易知道某值以上（或以下）的频次的总和。表 2--3 所表达的是某工厂 500 名职工教育程度的频次分布和累计频次分布。

表 2--3 频次分布和累计频次分布

教育程度 (X)	f	cf $\uparrow$	cf $\downarrow$
1. 本科	10	500	10
2. 专科	40	490	50
3. 高中	150	450	200
4. 初中	250	300	450
5. 小学	50	50	500
总数 (N)	500		

2. 累计百分比

累计百分比就是将各类的百分比逐级相加起来，可以由高级到低级累加 ($C\% \downarrow$)，也可以由低级到高级累加 ($C\% \uparrow$)。表 2--4 是和表 2--3 相对应的累计百分比分布表。

表 2—4

百分比分布和累计百分比分布

教育程度 (X)	%	c% ↑	c% ↓
1. 本科	2	100	2
2. 专科	8	98	10
3. 高中	30	90	40
4. 初中	50	60	90
5. 小学	10	10	100
总和百分数	100		
(统计总数)	(500)		

三、定距变量：频次分布与图示

简化定距变量资料的基本技术：除了上面所介绍的，还有一些特别适用的技术。

1. 频次分布

定距变量可以分成离散型和连续型两种。离散型定距变量（如家庭子女数、工厂科技人员数等），只能取得一些离散的数值。它的频次分布表（或百分比分布表）的制作和定序变量相同，变量的取值必须按次序排列。

连续型定距变量（如年龄、身高、体重、收入等）的取值连续地充满某个区间。制作频次分布表（或百分比分布表）时，首先要把变量的取值范围分成若干小区间（即若干组），然后计算每个组内的频次（或百分比）。分组时应该注意如下几个问题：

（1）关于组数的确定

组数不宜太多，也不宜太少。太多了会突出偶然因素的作用，太少了会掩盖变量变动时频次的变化情况，都不利于反映总体