

上海大学出版社

社会

S

SHEHUI HEXIE
YU FAZHAN CONGSHU

统计

CONGSHU

SHEHUI TONGJI

翁定军◎编著

社会统计发展丛书

SHANGHAI DAXUE CHUBANSHE



会
和
谐
与
发
展
丛
书

社 会 统 计

◎ 翁定军 编著

SHHX YFZCS 社会和谐与发展丛书
SHHX YFZCS 社会和谐与发展丛书
SHHX YFZCS 社会和谐与发展丛书
SHHX YFZCS 社会和谐与发展丛书
SHHX YFZCS 社会和谐与发展丛书
SHHX YFZCS 社会和谐与发展丛书

上海大学出版社
SHANGHAI DAXUE CHUBANSHE

图书在版编目 (C I P) 数据

社会统计 / 翁定军编著. — 上海: 上海大学出版社, 2006. 8

(社会和谐与发展丛书)

ISBN 7-81118-006-5

I. 社... II. 翁... III. 社会统计 IV. C91-03

中国版本图书馆CIP数据核字 (2006) 第089815号

责任编辑: 张光斌

封面设计: 王春杰

责任制作: 章 斐

社会和谐与发展丛书

社会统计

翁定军 编著

上海大学出版社出版发行

(上海市上大路 99 号 邮政编码 200444)

(<http://www.shangdapress.com> 发行热线 66135110)

出版人: 姚铁军

*

上海江杨印刷厂印刷 各地新华书店经销

开本 787 × 960 1/16 印张 18.5 字数 304 千

2006 年 8 月第 1 版 2006 年 8 月第 1 次印刷

ISBN 7-81118-006-5 / C · 061

定价: 31.00 元

内 容 提 要

随着计算机技术的发展,以“抽样—问卷—定量分析”为特征的现代统计调查得到了日益广泛的运用。本书主要介绍了社会调查研究中常用的一些统计方法。全书共分十章,前三章是描述最基本的统计方法,介绍了统计表、集中量数、差异量数和相关等内容;中间三章是推论统计的基础知识,内容主要是推论统计所涉及的概率基础知识、参数估计和假设检验;最后四章是目前调查数据分析中经常用到的几种数学模型的建立方法,内容包括多元线性回归、路径分析、Logistic 回归和对数线性模型。为便于实际应用,本书还结合目前流行的 SPSS 统计软件,介绍了相关的统计操作,并对输出结果的统计意义作了比较具体的解释。

本书可供社会学系本科生、研究生或社会学工作者研读。对于教育学、心理学或其他社会科学工作者也具有一定的参考价值。

社会和谐与发展丛书编委会



主 编：邓伟志

沈关宝

副主编：胡申生

仇立平

张钟汝

总 序

于光远

欣闻上海大学社会学系的同仁们新近编辑出版了一套“社会和谐与发展丛书”，我谨向他们表达诚挚的祝福。众所周知，世纪之交，中国社会的多元化发展，引发了各种新的研究热点，社会学的分支学科也层出不穷，社会学进入了繁荣发展的新阶段。相对于改革初期社会变迁集中在农村而言，进入 21 世纪以后，在经济全球化的背景下，新一轮的所有制改革所导致的结构性变迁，更多地是在城市的领域中演化的。城市社会的和谐发展能够成为热门研究，确实是形势所致，水到渠成。

上海大学社会学系地处中国最大的现代化都市上海，周边的长三角地区又是中国城市化水平最高、城市建设最先进的地区之一，这为他们研究城市与农村社会的和谐发展，赋予了天时地利的优势。该系的历史最早可以追溯到上世纪的 80 年代。那时，中国社会学的教学与研究刚刚得以恢复和重建，上海大学文学院的前身——复旦大学分校，便首先开设了社会学系。20 多年来，经过不断地努力和探索，该系在发展上具备了“规模效应”，“系”、“研究所”、“学术杂志”、“硕士生、博士生与博士后流动站等结合成一体”，培养了众多的社会学研究和应用型与理论型人才，逐步确立了其在学校文科建设中的重要地位。与此同时，本着“社会学的研究必须从现实生活出发”的原则，该系的教师和研究人員承担了一批国家和上海地区的重点研究课题，为国家和上海及周边地区的社会经济和谐发展出谋划策。他们的研究成果受到了国内社会学界和相关政府部门的关注。可以说，这套丛书很大程度上是这些研究成果的结晶，对于社会学分支学科的建设以及中国社会的健康而和谐地发展都具有良好的借鉴作用。

早在 2000 年，上海大学社会学系就在邓伟志、沈关宝两位教授的主持下，编著出版了“社会学与社会发展丛书”。当时这套丛书是被列为国家“十五”期间（2000—2005）重点图书出版规划的，丛书出版以后，受到社会学界和广大读

者的欢迎。应该说丛书也填补了国内某些社会学分支学科在专著和教材出版方面的空白,摒弃了其中一些脱离时代的陈旧内容,收到了良好的社会和经济效果。时隔六年,同样是由邓伟志、沈关宝两位教授主编的“社会和谐与发展丛书”又被批准列为“十一五”期间(2006—2010)国家重点图书出版规划,在他们的精心谋划、精心组织下,这套新的丛书很好地继承了原有丛书的体例,既保存了精华,又根据构建社会主义和谐社会的要求和发展目标、面临的形势和任务,予以更新和拓展;既着眼于学术探讨和交流上的价值,又适于社会学研究和教学实践之本色。相信这套丛书的问世,一定会给人以耳目一新之感。

当前,经济的高速发展与社会问题的频生如梦魇般牵绕在中国社会的急剧转型中。显然,单纯的市场化并非能够解决现代社会的所有问题,现代社会的复杂性需要更多的人文科学,特别是社会学的考量,这既是中国社会学界的责任之所在,也为社会学的创新提供了大量的机遇。我真诚地希望上海大学社会学系的同仁们能够延续他们紧密联系社会实际、严谨求新的学术风格,以敏锐的嗅觉、开阔的视野和浓厚的人文关怀不断为中国的社会学发展注入新的活力。

2006年7月21日于北京

前言

无论从哪个角度讲,社会统计都称不上是一门独立的学科。它像其他应用性统计——教育统计、心理统计、经济统计等一样,依附于数学分支学科——数理统计。任何新的统计技术、统计理论的出现和发展,全都仰仗数理统计,社会统计只不过是数理统计的理论和技術在社会研究领域的应用。社会统计不仅缺乏自身发展的能力,还缺乏自身发展的动因,它不具有自身提出研究问题的可能性,它所解决的全部问题都是由社会调查这门方法学科提出的。它是运用数理统计的原理分析社会调查研究的结果。社会统计与其说是一门学科,不如说是社会定量分析的一种手段,是处理社会调查数据的一种技术。

称其为技术,并不意味着它在社会研究中的地位无足轻重。任何事物都有质和量两个方面,社会现象也不例外。以往的社会调查研究往往只注重定性分析,忽视定量分析。然而,定性与定量相结合的研究方法已经成为目前社会调查研究中的一种趋势,定量分析在社会研究中表现出了无可否认的优越性,人们利用统计技术建立各种数学模型,揭示数据后面隐藏的各种社会现象之间的关系、规律和发展趋势。社会统计在社会研究的分析方法中已经取得了毋庸置疑的合法地位。

统计分析涉及大量的数据运算,离开了统计软件,一些复杂的运算几乎难以完成。在计算机技术出现以前,巨大的运算量妨碍了许多统计技术在社会调查研究中的应用。随着计算机技术的发展,各种统计软件的出现和普及,过去那些由于运算复杂而难以应用的统计技术在今天已经变得非常简单,人们再也不用手工运算,也不用记忆复杂的统计公式,过去需要数天甚至更长时间才能完成的运算,现在只需数分钟便能完成。一切变得迅捷、方便。统计技术的应用离不开统计软件。

然而统计软件毕竟只是一种计算工具,它不能代替人们对统计原理的理

解,不会告诉人们应该使用哪些统计方法,更不会告诉人们如何解释、分析统计结果。显然,统计软件的操作同样离不开人们对统计原理的理解。

本书将统计原理与统计软件的操作两部分内容结合在一起。重点在于介绍统计原理和统计方法,说明每种方法的主要含义和适用范围,力图以比较通俗的方式解释其中的统计原理,尽量回避过于复杂的数学演算和证明,以方便数学基础不太强的读者学习。当然,完全绕过高等数学和概率方面的知识是不可能的。在此基础上,结合具体例题介绍统计软件的相关操作。本书采用的统计软件是 SPSS for Windows 13.0 版。对于 SPSS 的操作介绍是比较简略的,只是介绍本书涉及的统计知识的软件操作步骤,解释软件的输出结果。至于本书未涉及的统计知识的软件操作,读者尚需参阅其他有关书籍。

本书原名为《社会定量研究的数据处理——原理与方法》,为与本科课程“社会统计”保持名称上的一致,本次再版改名为《社会统计》。与原书《社会定量研究的数据处理——原理与方法》相比,本次出版在内容上作了一些修改,主要是对 SPSS 统计软件的操作作了一些补充,此外,对原书中的例题也作了一些调整,删除了个别例题,增加了一些例题,而在整书的结构体系上基本未作变动。

限于作者水平,书中难免有各种不足和错误之处,真诚欢迎各位同仁指正。

翁定军

2006 年 7 月

目 录

前言	1
----	---

第一章 数据的初步整理	1
-------------	---

第一节	数据的特点与分类	1
	一、数据的特点	1
	二、数据的分类	2
第二节	统计表与统计图	4
	一、数据的统计分组和分类	4
	二、统计表	5
	三、统计图	8
第三节	SPSS 软件简介及基本操作	10
	一、数据的输入	11
	二、变量命名和数据编码	13
	三、数据的合并	14
	四、次数分布表和交互分类表的输出	15
	五、多项选择题的次数分布表和交互分类表	17
	六、数据的审核	18
	七、两个重要的命令——Recode 和 Computer	20

第二章 集中量数与差异量数

第一节	集中量数	22
	一、算术平均数	22
	二、中(位)数	25
	三、众数	28
第二节	差异量数	29
	一、全距	30
	二、异众比率	30
	三、四分位差	31
	四、平均差	33
	五、方差与标准差	34
第三节	相对差异量数	36
	一、离散系数	36
	二、相对位置量数—— z 分数	37
第四节	SPSS 软件的有关操作	38
	一、平均数与标准差的操作	38
	二、其他集中量数与差异量数的操作	40

第三章 相关与回归

第一节	相关概述	41
	一、相关的概念	41
	二、相关统计量	42
	三、消减误差比例	43
第二节	常用相关统计量的计算	44

	一、 λ (Lambda)相关	44
	二、 τ (tau-y)相关	45
	三、 G (Gamma)相关	47
	四、相关系数 r	50
第三节	一元线性回归方程	51
	一、相关与回归	51
	二、回归线、回归方程与回归系数	52
	三、最小二乘法和回归系数的计算	53
第四节	详析模式与偏相关	56
	一、虚假相关与虚假无关	56
	二、详析模式	57
	三、偏相关系数	62
第五节	SPSS 软件中的操作	66
	一、 λ 、 τ 和 G 相关的操作	66
	二、相关系数 r 的操作	67
	三、偏相关的操作	68
	四、详析模式的操作	69

第四章 概率与理论分布

第一节	概率简介	70
	一、随机现象与随机事件	70
	二、事件的相互关系	71
	三、概率的概念	71
	四、概率的加法与乘法	72
第二节	随机变量的分布	73
	一、随机变量	73
	二、随机变量的分布	74
	三、随机变量的数字特征	75

第三节	二项分布和正态分布	78
	一、二项分布	78
	二、正态分布	80
	三、正态分布中的概率计算	82
第四节	抽样分布与标准误	86
	一、抽样分布(样本分布)与标准误	86
	二、平均数分布的标准误	86
	三、比例分布的标准误	89
	四、平均数差异的标准误	90
	五、比例差异的标准误	92
第五节	大数定理与中心极限定理	92
	一、贝努里大数定理	93
	二、切贝谢夫大数定理	93
	三、中心极限定理	93

第五章 参数估计

第一节	推论统计的几个基本概念	95
	一、总体与样本	95
	二、统计量与参数	96
	三、抽样误差、随机误差与系统误差	97
第二节	参数的点估计	98
	一、衡量估计值好坏的标准	98
	二、求点估计值的方法——极大似然法	100
第三节	参数的区间估计	102
	一、总体平均数的区间估计	102
	二、总体比例的区间估计	106
	三、平均数差异的区间估计	107
	四、比例差异的区间估计	109
	五、相关系数 r 的区间估计	109

第四节	预测值的区间估计	111
	一、样本范围内的区间估计	111
	二、总体范围内 $\hat{y}$ 的区间估计	113

第六章 假设检验

第一节	假设检验的几个基本概念	117
	一、虚无假设与研究假设	118
	二、显著性水平与否定域	119
	三、双边检验和单边检验(两端检验和一端检验)	119
	四、两类错误	120
第二节	z 检验和 t 检验	121
	一、平均数单值的假设检验	121
	二、平均数差异的假设检验	124
	三、比例的单值检验	127
	四、比例差异的显著性检验	128
第三节	χ^2 检验	129
	一、 χ^2 检验的基本含义	129
	二、相关统计量的检验	131
	三、正态分布的检验	136
第四节	F 检验	137
	一、 F 检验的含义	137
	二、变异数的计算与方差分析表	140
	三、 r^2 相关	143
	四、方差一致性的检验	144
第五节	SPSS 软件的相关操作	144
	一、平均数检验与比例检验的操作	144
	二、 χ^2 检验的操作	148
	三、方差分析的操作	149

第七章 多元线性回归

第一节	多元线性回归的建立与解释	151
	一、多元线性回归方程的建立	151
	二、回归方程的解释	154
	三、标准回归方程与标准回归系数	155
	四、标准回归系数与相关系数的关系	157
第二节	回归方程的检验	158
	一、模型的整体线性拟合度检验	158
	二、自变量作用幅度的检验	164
	三、多重共线性的检验与容忍度	166
第三节	虚拟变量的运用	167
	一、虚拟变量的含义	167
	二、虚拟变量的编制	168
	三、虚拟变量在回归模型中的解释	168
第四节	SPSS 软件中的相关操作和逐步回归	169
	一、基本操作与结果解释	169
	二、有关选项的说明和操作	172
	三、含有虚拟变量模型的操作	173
	四、逐步回归及其操作	175

第八章 路径分析

第一节	路径分析概述	185
	一、回归分析与路径分析	185
	二、路径分析模型和假设条件	186
第二节	路径系数与残值路径系数的计算	188

	一、路径系数的计算	188
	二、残值项路径系数的计算	190
第三节	因果效应的分解	192
	一、直接效应与间接效应的分解	192
	二、相关系数的分解	192
第四节	路径分析的检验	203

第九章 Logistic 回归

第一节	Logistic 回归的涵义和建立	208
	一、线性回归的局限与 Logistic 回归的涵义	208
	二、回归系数的计算	211
	三、回归模型的解释	212
	四、自变量为虚拟变量的 Logistic 回归	213
第二节	模型的检验	215
	一、模型整体的检验	215
	二、回归系数的检验	216
第三节	SPSS 软件中的操作	217
	一、“Enter”(强迫进入)方式下的 Logistic 回归	217
	二、“Backward: wald”方式下的逐步回归	221

第十章 对数线性模型

第一节	对数线性模型的原理和涵义	224
	一、交叉乘积比	224
	二、对数线性模型的建立	227
	三、模型中各项参数的求解	229
	四、各项参数与交叉乘积比的关系	231

第二节	$2 \times J$ 表和三维表的对数线性模型	235
	一、从 2×2 表到 $2 \times J$ 表	235
	二、从二维表到三维表	237
	三、对模型中交互效应的解释	239
第三节	非饱和模型	241
	一、非饱和模型的涵义	241
	二、非饱和模型中的次数拟合(理论次数的估计)	242
第四节	对数线性模型的检验	243
	一、模型拟合度的检验	243
	二、分层效应和单项效应的检验	245
	三、效应参数的检验或估计	247
	四、单元格的拟合度检验	249
第五节	利用 SPSS 统计软件进行对数线性模型分析	249
	一、操作步骤简要说明	249
	二、各项输出结果的含义	250
	三、例题分析	258
附表		266
附表 1	正态分布下的概率值	266
附表 2	t 值表(双侧临界值)	268
附表 3	χ^2 数值表	270
附表 4	F_{α} 值表	272
附表 5	γ 值的 z 转换表	275
附表 6	相关系数显著性检验表	277
参考文献		278