

高等学校文科教材

社会统计学

附社会科学统计学统计软件应用

丘海雄

编著

中山大学出版社

社会统计学

——附社会科学统计软件应用

丘海雄 编著

中山大学出版社

· 广州 ·

(粤)新登字 11 号

版权所有,翻印必究

图书在版编目(CIP)数据

社会统计学/丘海雄编著. —广州:中山大学出版社,1993. 12

附社会科学软件应用

ISBN7—306—00761—0

- I 书名
- II 丘海雄
- III ①社会科学 ②数理统计
- IV C8

责任编辑:邱琼英 责任校对:钟水源

中山大学出版社出版发行

(广州市新港西路 135 号)

韶关新华印刷厂印刷 广东省新华书店经销

*

850×1168 毫米 32 开本 13 印张 35 万字

1993 年 12 月第 1 版 1993 年 12 月第 1 次印刷

印数:1—2000 册 定价:5.85 元

内 容 摘 要

本书清晰、直观和系统地介绍了单变项和两变项的描述统计与推论统计,以及扼要地介绍了多变项统计的基础知识,并在上述的基础上,介绍了社会科学计算机软件 SPSS/PC+ 的应用。本书偏重介绍统计方法的选择原则和统计结果的意义解释,结合实例对统计原理作尽可能通俗的介绍,并且融汇了对社会学研究的规范、意念和思路的阐述。本书是高等学校社会学专业的社会统计学课程教材,同时也可供社会科学研究者以及对社会、经济现象作数量分析的人员作参考。

前 言

1989年3月国家教委在广州中山大学召开“全国高等院校社会学系(专业)系主任(专业负责人)联席会议”。议题之一是组织编写社会学本科教材。参加“社会研究法”和“社会统计学”组讨论的同志在评议各院校呈交的《社会统计学教学大纲》后,认为有必要编写本科的社会统计学教材,并提议由中山大学社会学系负责整理、完善教学大纲和组织编写教材。笔者承担了这项任务,在综合各兄弟院校教学大纲之长的基础上整理出一份新的教学大纲,咨询了一些兄弟院校担任这门课程教学的教师的意见,最后根据教学大纲编写出这本教材。

本书介绍的统计知识适用于各领域的社会科学研究的统计分析。社会科学将统计作为应用工具。本书与一般的统计学原理教科书相比,侧重介绍处理社会科学研究过程中如何恰当地选择统计手段和如何准确地解释统计结果,通过实例对统计原理作尽可能直观、通俗的分析,而统计公式的数理推导则从略。目前社会科学对多数的社会现象只能作类别或等级的测量,本书注重介绍处理类别和等级变量的统计手段。此外,本书在介绍各种统计手段的过程中,还贯穿了对社会科学研究的规范、意念和思路的阐述。

本书选用了一套名为《企业工人工作情况调查》的问卷资料作为贯穿全书的例子和练习素材,以便读者能对社会科学研究的过程有整体的了解和提高处理原始数据的能力。

本书的第一篇介绍了统计发展的历史、社会统计的基本概念、过程及选择统计手段的准则,目的是让读者对于社会统计有概括性的认识。

第二篇是描述统计,介绍单变项简化的基本技术、集中趋势和高散趋势测量法以及两变项的相关性质和相关测量。

第三篇是推论统计,介绍抽样调查与统计推论的基本知识、参数估计以及单变项、两变项的假设检定。

第四篇是多变项分析,介绍详析分析和统计控制,简略地介绍复相关、多元回归、因径分析以及因子分析等主要的多变项分析技术。

第五篇是社会科学统计软件应用,介绍了 SPSS/PC+ 的基本使用方法、常用的操作、定义、运算和程序指令。这套软件的全称是“社会科学统计软件包”(statistical package for social sciences),“SPSS”是每个英文单词第一个字母的缩写。“PC”是指微机(personal computer),即指示该软件适合于在微机上运行。“+”是指增强的版本。在国际三大统计软件(其余两套是 SAS、BMDP)中,SPSS 最适合于社会学、政治学、人类学等目前量化程度较低的社会科学的统计分析。目前国内社会学本科、研究生社会统计课程涉及的统计手段,SPSS 都基本涵括了。本书简要地介绍了使用该套软件的基本方法和一些最常用的指令,所举的例子都是本书前面各章引用过的《企业工人工作情况调查》里的例子。读者可同时了解对同一问题作手工计算与计算机计算的方法和过程。同时可用本书所附的上述调查资料的前 50 个个案的原始数据(见附三、四)输入计算机自行作练习。此外,还介绍了笔者及其学生刘晓东在使用这套软件过程中摸索出来的一种十分简易的使输出结果部分汉化的方法。

本书的多数章节都附有练习题,并且选择目前国内比较好的由李沛良、布莱洛克和卢淑华著的三本社会统计教科书作为参考书目,列出每章的参考页码。

这本书得以完成,首先要感谢我读研究院时教我方法进阶课程的老师李沛良教授。他不但教我学习社会统计学的知识,还教我

如何观察社会,如何做一个称职的教师。他的《社会研究的统计分析》一书一直是国内社会学界最受欢迎的社会统计学教材。本书的体系设置、陈述方式很大程度上参照了他的这本著作。同时要感谢刘创楚博士、张德胜博士在我求学期间在学业上和生活上给予的热情关怀,感谢中山大学社会学系的何肇发教授、丘士杰副教授多年来对我的专业训练和教学科研给予的支持和帮助。此外要感谢曾经教过或指导过我学习社会统计学的林南教授、邝振权博士、陈膺强博士、梁作繁博士,吴伯骏博士,及在计算机应用上给予指点的余雪明小姐。在本教材的写作过程中,中山大学社会学系的刘晓东和丛彤同学帮我整理资料和校对文稿,罗仰明、蔡瑞琼小姐帮我打印文稿,在此一并致谢。

作者 1991年冬于康乐园

目 录

前 言	1
第一篇 导 论	1
第一章 社会学研究与社会统计	1
第一节 什么是社会统计	1
第二节 统计手段的选择	9
第二篇 描述统计	15
第二章 单变项统计分析	15
第一节 基本技术	16
一、定类变项:次数分布、比例、百分比、对比值、圆形比例图、长条图	16
二、定序变项:累加次数、累加百分比	23
三、定距变项:直方图、多边形图	27
第二节 集中趋势测量法	35
一、定类变项:众值	36
二、定序变项:中位值	37
三、定距变项:平均值	39
第三节 离散趋势测量法	45
一、定类变项:离异比率	45
二、定序变项:四分位差	46
三、定距变项:标准差	50
第四节 正态分布与标准值	54
第三章 两变项的统计分析	62
第一节 交互分类	62
第二节 相关的基本性质	68

一、相关的程度	68
二、相关的正负	69
三、相关的对称性	70
四、消减误差比例	70
第三节 相关测量法	73
一、两定类变项： $\lambda, \lambda_{yx}, \text{Tau}-y$	73
二、两个定序变项： $G, d_{yx}, d_{xy}, \text{Tau}-a, \text{Tau}-b, \text{Tau}-c, \rho$	78
三、两定距变项：简单直线回归， r	93
四、定类变项与定距变项： E	105
第三篇 推论统计	112
第四章 抽样调查与统计推论	112
第一节 抽样调查	112
一、抽样的意义和问题	112
二、抽样的过程	113
三、非随机与随机抽样	115
第二节 机率与抽样分布	123
一、二项抽样分布	124
二、平均值抽样分布	126
第五章 参数估计	130
第一节 平均值、百分比间距估计	131
一、平均值的间距估计	131
二、百分比的间距估计	134
第二节 积矩相关系数的间距估计	135
第三节 决定样本大小	137
第六章 假设检定	141
第一节 假设检定的基本概念	141
一、研究假设与虚无假设	141
二、否定域与显著度	143
三、一端检定与两端检定	146

四、甲种误差与乙种误差	148
五、参数检定与非参数检定	149
六、假设检定的基本过程	150
第二节 平均值和百分比假设检验	150
一、单平均值假设检定	150
二、两平均值差异假设检定	159
三、单百分比假设检定	163
四、两百分比差异假设检定	164
第三节 两变项相关假设检定	165
一、两定类变项:卡方(X^2)检定	165
二、两定序变项:G 检定、S 检定	173
三、定类/定距变项:F 检定、单因方差分析	176
四、两定距变项:F 检定	181
第四篇 多变项统计分析	186
第七章 详析分析和统计控制	186
第一节 详析分析的基本概念	186
第二节 因果分析	188
第三节 阐明分析	196
第四节 条件分析	202
第五节 净相关	208
一、定类变项: λ_p , $Tau - y_p$	208
二、两定序变项: G_p , $dyxp$	210
三、定距变项: Y_p , $Y_x(y \cdot 1)$, $Y_y(x \cdot 1)$	212
第八章 多因、多项相互关系分析简介	217
第一节 复相关	217
第二节 多因回归	222
第三节 因径分析	236
一、因径分析的变项和效应测量	237
二、因果网的类型	238

三、完整单向模式	240
第四节 因子分析	252
一、计算相关矩阵	253
二、抽取共同因子	254
三、旋转因子	259
第五篇 社会科学统计软件应用	268
第九章 SPSS/PC+概述	268
第一节 SPSS/PC+简介	269
一、什么是 SPSS/PC+	269
二、SPSS/PC+要求的系统环境	269
三、SPSS/PC+的安装与运行	270
第二节 SPSS/PC+命令概要	273
一、操作命令	273
二、数据定义与运算命令	274
三、程序命令	274
第三节 SPSS/PC+应用流程	275
一、建立数据文件	276
二、建立命令文件	276
三、清理数据文件	276
四、编制资料手册	277
五、转换和控制资料	278
六、分析资料	278
七、输出结果编辑、打印	279
第十章 SPSS/PC+命令	281
第一节 操作命令	281
一、INCLUDE	281
二、SET	282
三、HELP	283
四、FINISH	283

第二节 定义、运算命令	284
一、BEGIN DATA END DATA	284
二、DATA LIST;固定格式	286
三、DTTA LIST;自由格式	287
四、RECODE;自由格式	288
五、COMPUTE	289
六、IF	290
七、MISSING VALUE	291
八、PROCESS IF	292
九、SELECT IF	292
十、VALUE LABELS	293
十一、VARIABLE LABELS	294
第三节 程序命令	294
一、FREQUENCIES	294
二、NPAR TESTS	300
三、T-TEST	303
四、CROSSTABS	307
五、REGRESSION	313
六、FACTOR	322
七、ANOVA	327
第十一章 SPSS/PC+应用的其它问题	333
第一节 关于 SPSS/PC+的新版本	333
第二节 SPSS/PC+输出结果部分汉化的方法	337
附一 社会研究的过程:流程图式的概念整理	340
附二 数表	356
表一 随机数表	356
表二 正态曲线下的面积	361
表三 r 值与 Z 值的转化	365
表四 t 分布表	368
表五 F 分布表	369

表六	X^2 分布表	372
附三	《企业工人工作情况调查》登录册	374
附四	《企业工人工作情况调查》部分原始数据	381
附五	习题参考答案	385
附六	英汉统计词汇索引	390

第一篇 导 论

第一章 社会研究与社会统计

第一节 什么是社会统计

一、统计的基本概念

统计一词的德文是 Statistik, 英文是 Statistics, 法文是 statistique, 意大利文是 statistica, 它们均以拉丁文 status(现状、状态)为词根。该词最早出现于德国学者的论著, 意指国家重要事项的比较和记述, 也可称为“国势学”。近百年来伴随着统计学的发展和统计学的广泛应用, “统计”一词包含了多方面的涵义:

第一是指统计工作, 即收集、整理、分析、推论量化资料的活动过程。

第二是指统计资料, 即获得的量化资料, 包括原始数据, 经简化和整理的数据等。

第三是指统计学, 统计学(statistics)是收集、整理、分析和推论量化资料的科学方法。

本节讨论的主要是统计的第三种涵义, 将简单地介绍统计学的发展历史、统计学的内容、统计学的作用和局限性等。

二、统计学的发展

统计学的发展大体可划分为三个阶段。^①

第一阶段可称为古典统计学阶段,约自16世纪中至19世纪初,这个阶段占主要地位的学派包括国势学派、机率学派和政治算术学派。

国势学派以德国学者为主,创始人是赫谟斯大学教授康令(H. Corning, 1606—1682)。他继承古希腊亚里士多德记述希腊城邦政情的传统,在大学中讲授记述欧洲各国政治、经济、军事、地理等现状的课程。1723年,耶拿(Jena)大学的施梅尔兹(M. Schmertzel, 1679—1747)开设了统计学讲座。法国的阿西昂(G. Achenwall, 1719—1772)在格廷根(Göttingen)大学正式统计学课程,首次将国势记述的学问命名为统计学,认为统计学是关于各种重要事项的学问。同一时期,丹麦的安查森(J. D. Ancherson, 1700—1765)首创以表格分栏排列国家的土地、人口、宗教、军事、货币和度量衡等数字,称为“表式统计学”,也属国势学派。国势学派的特征可以说主要是以国家的各种重要事项的统计数据记述国家的状态。

机率学派以法国学者为主,创始人是帕斯卡(B. Pascal, 1623—1662)和费尔马(P. de Fermat, 1601—1665)。他们以通信的方式讨论赌博时的机率问题,发表了《骰子赌博理论》一书。穆瓦韦尔(A. de Moivre, 1667—1754)发现了常态方程式。同一时期瑞士的伯努利(J. Bernoulli, 1654—1705)提出了二项分布理论。1814年,法国的拉普拉斯(P. S. Laplace, 1749—1827)发表了《机率分析论》一书。该书奠定了古典机率理论基础并将机率理论应用于自然和社会的

^① 有关统计学发展的资料主要转引自林义男,《统计学导论——人文与社会科学以及工商与行政管理应用》,台湾巨流图书公司,1981出版,9~11。

研究。此后法国的泊松(S. D. Poisson, 1781—1840)提出了泊松分布。德国的高斯(C. F. Gauss, 1777—1855)提出了最小平方法。这些有关的机率理论对现代的统计学仍然有影响。

政治算术学派以英国人为主,创始人是格兰特(J. Graunt, 1620—1674)和配第(W. Petty, 1623—1687)。格兰特1662年发表了《关于死亡表的政治观察》一书,在书中提出了人口发展变化是有一定规律性的观点。配第注重社会经济领域的数量计算和分析,企图以数量作为支配社会经济现象法则的客观依据。政治算术学派超越了国势学派对客观事实的记述,以数量关系的计算来说明、推论有关现象。

第二阶段可称为近代统计学阶段,约自19世纪中至20世纪初。

这一阶段以比利时的统计学家凯特勒(A. Quetelet, 1796—1874)为起点。他集国势学派、机率学派和政治算术学派于一身,发表了《社会物理学》一书。认为人类的社会活动也有一定的规律,可以通过大量的观察去认识这些规律。他以18世纪发展起来的机率论为基础,计算和分析观察数据的误差,强调正态分布、平均数的应用可以扩展到对社会现象的研究。

英国的生物学家高尔顿(F. Galton, 1722~1911)在研究生物遗传学和优生学中创立了回归分析和相关的重要概念。他的学生皮尔森(K. Pearson, 1857—1936)探讨了各种不同的次数曲线理论,创造相关系数、卡方检定等统计方法,同时注意到了从大样本推论总体的误差问题。因此有人称皮尔森是近代统计学的创始人。

第三阶段可称为现代统计学阶段,自20世纪初至现在。

这一阶段始于英国的哥色特(W. S. Gosset, 1876—1937),他以笔名“Student”发表了有关 t 分布的著作,提供了小样本抽样理论的基础。英国的费歇(R. A. Fisher, 1890—1962)提出了 Z 分布、显著度、假设检定、自由度、实验设计、方差分析等重要的统计学概念

与方法。

这一阶段美国的学者也对统计学的发展作出了杰出的贡献。

塞科(G. W. Snedcor)将 Z 分布转换成 F 分布。沃尔德(A. Wald, 1902—1950)提出了决策理论和逐次抽样法。诺伊曼(J. von Neumann)和摩根斯顿(O. Morgenstern)提出了对局理论,使决策理论更系统化。内曼(Neyman)和邓明(Deming)提出了抽样调查法。维纳(N. Wiener)的控制论和仙依(C. E. Shannon)的信息论使推论统计学的理论更加完善。

由于现代统计学对于现代化的大生产管理有重要的应用价值,因而在战后蓬勃发展。电子计算机的出现使人们可以迅速地对大规模的自然、社会现象数据作精确的统计分析,使当今的统计学更上一层楼。

三、统计学的内容与分类

按流行的划分方法,统计学的内容分为两大部分:描述统计和推论统计。

描述统计(Descriptive Statistics)是运用统计技术对数据资料进行简化,并表达其分布特征的统计方式。缺乏统计学知识的人在社会研究的过程中往往陷入面对大量的数据而束手无策的困境,原因是资料包含的信息量很大。例如若有500份问卷,每份问卷有50个问题,即使每个问题可用一位数字来代表,总共也有25000个数据,对于这些数据,是难以靠直觉来掌握其内容和特征的。必须设法将这些数据加以浓缩提炼,即必须对这些资料进行简化、描述。例如利用百分比、平均值、标准差及相关系数等统计手段,用简单的数量指标反映资料的分布特征。当然,用少数的数量指标代表大量的数据,难免损失信息,假如选择的指标不合适或解释不当,还会产生误导的作用。因此,描述统计的运用必须十分审慎。

推论统计(Inferential Statistics)是利用随机样本推断总体情况