



高等学校经济与管理类核心课程教材

统计学

STATISTICS

夏南新 著

高等教育出版社

高等学校经济与管理类核心课程教材

STATISTICS

统 计 学

TONG JI XUE

夏南新 著

高等教育出版社·北京

图书在版编目(CIP)数据

统计学 / 夏南新著. —北京: 高等教育出版社,
2014. 2
ISBN 978-7-04-038986-9

I. ①统… II. ①夏… III. ①统计学—高等
学校—教材 IV. ①C8

中国版本图书馆 CIP 数据核字 (2014) 第 015265 号

策划编辑 刘自挥 责任编辑 刘自挥 王驰宇 封面设计 吴昊 责任印制 蔡敏燕

出版发行	高等教育出版社	咨询电话	400-810-0598
社 址	北京市西城区德外大街 4 号	网 址	http://www.hep.edu.cn
邮政编码	100120		http://www.hep.com.cn
印 刷	上海华教印务有限公司		http://www.hepsh.com
开 本	787mm×1092mm 1/16	网上订购	http://www.landaco.com
印 张	23		http://www.landaco.com.cn
字 数	565 千字	版 次	2014 年 2 月第 1 版
购书热线	010-58581118	印 次	2014 年 2 月第 1 次印刷
	021-56717287	定 价	39.00 元

本书如有缺页、倒页、脱页等质量问题, 请到所购图书销售部门联系调换

版权所有 侵权必究

物 料 号 38986-00

作者简介



2013年7月12日,作者与妻子姜穗琦医生在克里姆林宫留影

夏南新,江西省南昌市人。1994年7月开始任教于中山大学岭南学院,教授职称,系书记。1997年9月—2000年7月,厦门大学统计学专业博士研究生,师承统计学大师钱伯海教授,获博士学位。2000年至今担任六届中国数量经济学会理事。2002—2003年在美国春田大学进修。在《统计研究》上独立发表论文7篇,在《数量经济与技术经济研究》上独立发表论文3篇,共独立发表论文30篇。在中国财经出版社出版专著《新概念统计学》、《地下经济估测》。独立完成的研究报告获2011年广东省哲学社科成果三等奖。

2005—2007年主持的国家社科基金一般项目“地下经济对国家经济安全影响的实证研究”,以题为《关于防范和遏制地下经济对国家经济安全影响的对策建议》的研究报告形式,被中宣部全国哲学社科规划办编入2009年第12期国家社科基金项目《成果要报》,上报给中央政治局委员,中央书记处书记,国务院总理、副总理,国务委员等领导以及中央办公厅、国务院办公厅等机构参阅。国家社科基金项目《成果要报》是呈送中央政治局常委、委员等党和国家领导人的内部参阅件,专门摘报国家社科基金项目研究成果中有重要现实意义和应用研究、对策研究成果,作为党和国家重要决策的参考(引自中山大学新闻网,2009年5月18日)。2009年12月中宣部全国哲学社科规划办向广东省委宣传部、中山大学党委发来《关于中山大学夏南新同志研究成果受到有关领导和部门重视的通报》。该《通报》表示,《建议》中提出的观点和建议已受到国家有关部门负责同志的重视。《通报》指出,作为哲学社科研究工作者,夏南新教授带领课题组坚持正确导向,自觉关注现实问题,深入开展调查研究,努力推出高质量的理论研究成果,体现了较高的责任感和使命感,为国家社科基金更好地服务党和国家工作大局作出了贡献(引自中山大学新闻网,2009年12月21日)。在人大经济论坛上,被推举为统计学科唯一学术带头人;并在其人物专栏排行中,名列第8位。夏南新格言:我讲的知识你们都能听懂,甚至能超过我,不过,有一点你们恐怕永远超不过我——在学术上坚决不搭便车!2014年1月8日,在北京召开的中国统计学会第九次全国会员代表大会上,31个省市自治区统计局长、国家统计局各省调查总队长,以及院士等当选常务理事,夏南新当选常务理事。

前言

1980年,纽约未来学巨擘阿尔文·托夫勒(Alvin Toffler, 1928—)在《第三次浪潮》中指出:“如果说IBM的主机拉开了信息化革命的大幕,那么大数据才是第三次浪潮的华彩乐章。”

所谓大数据(big data),它是指大量以文字、图片、视频等形式存在的非结构化和半结构化,下载到关系型数据库中无法透过传统架构的信息系统工具挖掘的资讯数据。大数据特点:①数据体量从TB级别跃升到PB级别——Volume;②物联网、云计算、手机、PC、GPS系统、QR码、传感器、社交网络等承载图片、视频、地理位置、网络日志等海量数据流——Variety;③价值密度低,商业价值高。例如:对视频连续监控,由于大量水军差评、好评和刷量等噪声数据干扰,所以提纯出蕴含价值的可能只有几秒钟的视频数据——Value;④信息数据瞬息万变,遵循“1秒定律”,优化流程,提升挖掘技术——Velocity。

大数据分析是商业智能的演进。数据挖掘前端化(the front-end of data mining),让用户个性化体验,开展直觉感知推荐服务,数据的价值在于将正确的信息在正确的时间交到正确的人手中。数据越大,机遇也越大。

以青少年为主力的大众群体社交堆砌的大数据,主观随意性很大,毫无价值的“负面清单”式的信息比比皆是,“捕风捉影”式地从海量数据中挖掘出潜在价值的点滴信息,真可谓“积小流成江海”,可见,样本是难当其大任。实际上,抽样还不如降维(dimension reduction),即减少变量个数,大数据分层或聚类可能比抽样更有效。统计分析是以洞察总体或群体特征为目标,而大数据分析是以每个个体特征分析并以其特征作为分析应用支撑为目标,所以,大数据分析对象显然不能基于抽样,其实无须抽样。大数据量大和维度高,维度低(即变量少)的数据分析根本不需要大数据,抽样完全可以解决。再大的局部还不如随机样本更有代表性。样本应当与其所对应总体的分布规律一致,而大数据几乎不可能掌握总体分布特征,也就无所谓抽样代表性了。从成本和效率的视角考虑,抽样是不可或缺的。分布式计算(map/reduce)和实时流计算(内存计算/流式计算)成为大规模数据处理平台。数据统治世界,大统计式思维,将改变人们利用数据的方式。过去,受限于处理数据的软、硬工具,抽样便成了折中的统计技术。如果大数据本身就是总体而非大样本,那么基于样本推断总体的假设检验和估计技术就毫无价值了。没有了抽样,统计学将被改写,“大数据统计学”正应运而生!

为了依据翔实可靠的统计数据进行科学决策,经修订的《中华人民共和国统计法》于

2010年1月1日起施行。该项法律确保了我国统计数据的真实性、准确性、完整性和及时性。统计法第一章总则中的第二条规定：统计的基本任务是对经济社会发展情况进行统计调查、统计分析，提供统计资料和统计咨询意见，实行统计监督；第六章法律责任中的第四十五条指出：违反本法规定，利用虚假统计资料骗取荣誉称号、物质利益或者职务晋升的，除对其编造虚假统计资料或者要求他人编造虚假统计资料的行为依法追究法律责任外，由作出有关决定的单位或者其上级单位、监察机关取消其荣誉称号，追缴获得的物质利益，撤销晋升的职务。

无论是生产数据（注：不是编造数据），还是使用数据（注：不能篡改数据），都必须遵守统计法。论著中的模型拟合以及显著性检验也是数据生产方式，一旦作者蓄意篡改结果，将会误导读者，在一定程度上也有悖于统计法。“真实”应当成为做学问的底线。

面对爆发式增长的天量数据，肩负处理大数据使命的统计学，借助描述统计方法和推断统计技术，有效发掘事物的信息价值，从量化分析上认识事物的质，进而探索数据的内在规律性。

美国统计学会前会长伯克斯(G. E. P. Box)在他的学术报告(JASA, 1979, Vol. 74)中曾指出：统计学界呈现出偏重研究统计理论的统计学者群和偏重研究统计应用的统计学者群的双峰分布群体，期待出现统计理论与统计应用研究日臻完美结合的单峰分布学者群体。统计学者群体呈现单峰或双峰分布，孰优孰劣，这是仁者见仁、智者见智的问题。不过，本书旨在力求打造单峰分布统计学者群体。

2011年5月，国务院学位办《关于按〈学位授予和人才培养学科目录〉进行学位授权点对应调整的通知》，各学位授予单位需将按原目录批准的部分学位授权点，对应调整到新目录相应的一级学科。在此次调整中，统计学升格为一级学科，下设数理统计、经济统计、生物与医学统计等4个方向。在贯彻落实《国家中长期教育改革和发展规划纲要（2010—2020年）》下，中国大统计正迎来繁荣的发展契机。值此中国圆了大统计梦之际，本书付梓，大数据、大统计构筑了本书双引擎。

本书布局力求凸显人性化的体例风格。“不求所有，只盼拥有”（“Not for ownership, only for holding”）。凡涉及到的知识点，力求透彻。也许会有悖于“尽量避免烦琐的数学推导”的思维定式。考虑到篇幅，没有以事例切入每章开篇的手法。为了追根溯源，适当穿插了统计名人小传。

选择剑走偏锋的探究路径，书中出现的瑕疵，恳请同仁赐教！

夏南新 中国·广州·祈乐苑

2014年1月

目 录

第一章 统计学褒贬、诠释和史论	001
第一节 统计学的褒与贬	001
第二节 统计学的诠释	004
第三节 统计学的史论	005
研讨问题一	010
第二章 统计学的范畴、方法论、量化尺度与试验设计	011
第一节 统计学的基本范畴	011
第二节 统计方法论	012
第三节 统计数据的量化尺度	014
第四节 试验设计	015
研讨问题二	018
第三章 统计资料的搜集、整理和描述	020
第一节 统计资料的搜集	020
第二节 统计资料的整理	024
第三节 频数分布	025
第四节 统计表的制作	035
研讨问题三	038
第四章 综合指标	040
第一节 绝对指标	040
第二节 相对指标的概念与划分	042
研讨问题四	049
第五章 集中与离散趋势测定指标	050
第一节 集中趋势测定指标	050
第二节 离散趋势测定指标	074
第三节 是非标志	081
第四节 离散系数	083
研讨问题五	085
第六章 时间序列分析指标	090
第一节 时间序列的概述	090
第二节 水平指标	092
第三节 速度指标	101
研讨问题六	110
第七章 动态趋势与波动的测定	112
第一节 动态趋势与波动的概念	112

第二节	动态趋势与波动的经典模式	112
第三节	动态序列的结构模型	114
第四节	长期趋势的测定	115
第五节	季节波动的测定	133
第六节	循环波动的测定	145
第七节	不规则波动的测定	150
	研讨问题七	151
第八章	指数	154
第一节	指数的概念、发展史、作用和分类	154
第二节	个体指数	156
第三节	总指数编制之一——简单指数	157
第四节	总指数编制之二——加权综合指数	160
第五节	总指数编制之三——加权平均数指数	175
第六节	总指数编制之四——平均指标指数	181
第七节	价格总指数的修正公式及数理方法	184
第八节	股票价格指数	187
	研讨问题八	208
第九章	抽样技术	211
第一节	抽样技术的概述	211
第二节	抽样推断中的基本概念	212
第三节	抽样推断的方法论基础	215
第四节	样本的概率分布	217
第五节	抽样估计的优良标准	219
第六节	抽样误差	221
第七节	必要样本容量的确定与全及总体指标的推断	230
第八节	随机抽样的组织形式	238
	研讨问题九	266
第十章	相关分析与回归分析	268
第一节	相关分析	268
第二节	回归分析	290
	研讨问题十	311
附录一	部分章节部分研讨问题的参考答案	313
附录二	“商业与经济统计”课程期末试题及其答案	323
附录三	统计分布表	336
参考文献		358
教学资源索取单		

第一章

统计学褒贬、诠释和史论

第一节 统计学的褒与贬

一、统计学的褒扬

印裔美籍著名统计学家 C. R. 拉奥(Calyampudi Radhakrishna Rao, 1920—)指出:在终极的分析中,一切知识都是历史;在抽象的意义下,一切科学都是数学;在理性的基础上,所有的判断都是统计学。人们可以强调科学定律不是靠权威的原则推进,不是由信仰或中世纪哲学所证明,统计是诉诸新知识的唯一的法庭^①。

统计活动是认识现象总体数量方面的活动,它经历了四五千年漫长的艰苦探索和发展过程。早在原始公社时期,就有结绳记事和计量的方法,《周易正义》中写到原始社会:“事大,大结其绳;事小,小结其绳;结之多少,随物众寡。”^②这可谓统计萌芽。据晋皇甫谧(公元 215—282)《帝王世纪丛书》记载,早在 4 000 多年前的夏禹王朝,为了治国治水的需要,曾进行过初步的国情统计,查明当时全国人口为 13 553 923 人,土地为 24 308 024 顷。并依照山川土质、人口物产及贡赋多寡,将全国分为九州。公元前 3050 年,埃及建造金字塔和兴修大型灌溉系统时,为了征集所需的财物和劳力,对全国人口、财产进行过调查。公元前 600 年古希腊就进行过人口普查。公元前 400 年古罗马就建立了出生、死亡登记制度。尤其是我国西周朝代建立的统计报告制度——日报称日成,月报称月要,年报称岁会,被西方经济学家推崇为统计学最早的萌芽。远古时期的统计活动搜集的数字资料,主要是为了满足当时奴隶主和封建王朝课税、徭役等统治需要。然而,统计在学术上作为一门学科的历史只是始于 17 世纪中叶的欧洲,至今不过 300 多年。尤其是近代将概率论引

① [印度] C. R. 拉奥,《统计是寻求真理的科学》,《统计译文》,1985 年第 5—6 期。

人物介绍

C. R. 拉奥(Calyampudi Radhakrishna Rao, 1920—),美国宾夕法尼亚州立大学统计学荣誉教授(professor emeritus)。1920 年出生于印度 Karnataka 省的一个贵族家庭。1940 年获印度安德拉(Andra)大学数学硕士学位,1943 年获加尔各答(Calcutta)大学统计学硕士学位。经印度著名统计学家马哈拉诺比斯(Mahalanobis)推荐,进入剑桥大学师从数理统计学奠基人 R. A. 费歇(Fisher)教授,1948 年获统计学博士学位。在多元分析、渐进推断、估计理论、概率分布、矩阵代数、组合分析、统计学中的微分几何方法等研究领域作出了重要贡献。克拉默-拉奥不等式是寻求一致最小方差无偏估计重要工具,由拉奥和瑞典统计学家 H. 克拉默在 1945 年和 1946 年先后独立证明。被 14 个国家授予 25 个荣誉学位。当选为美国科学院、第三世界科学院、英国皇家统计学会等 31 个国际著名研究机构的院士、名誉院士或理事。2002 年获得美国总统科学奖。

② 参阅《周易正义》,郑玄注。

入统计学中,使统计学有了惊人的发展。由于人类社会现在正处在信息化时代,电脑在瞬息之间可以准确地处理大量的信息,通过处理后的大量信息,经过筛选分析,对于社会现象来说,可以从中发掘规律性,为制定方针政策提供依据;而对于自然现象来说,可以发现控制它们的手段,为揭示科技奥秘提供依据。这就是所谓的大量统计信息当中存在着反映社会现象和自然现象的“规律”或“法则”。

赞誉一个人的学识用“才高八斗,学富五车”,这也是足以证明此人大大脑中摄入的知识信息量大,在某种程度上可以说,此人已经掌握了社会想象或自然现象某领域中的“规律”或“法则”。正如西方学者 H. G. 威尔斯对统计的比喻:“统计思维,如同读写能力一样,总有一天会成为讲求效率的公民所必需的本领。”^①

我国著名的统计学家陈希孺院士^②(1934—2005,当时中国统计学界唯一的院士)指出:“我们现处在一个所谓‘信息爆炸’时代,信息的一种最常见的形式就是数据。现代人在工作和生活中,不时地要从这大量的、杂然纷呈的数据中发掘出事物的规律,做出正确的判断,以决定适合的行动方针。可以说,这方面的能力如何,实在是衡量一个人聪明与否的一个极重要的外现指标。‘仙人持玉尺,度君多少才’。要准确无误地测出一个人有多聪明,恐非仙人莫办,可惜红尘十丈,何处去找那手持玉尺的仙人?但处事是否得宜,是有客观检验的。古人说与其载之空言,不如见之于行事之深切著明。而统计学,正是这样一种通过分析数据而达致尽可能正确的结论的技术。掌握了它,就可以使你在工作和生活中少犯错误,赢得主动。也就是说,使你变得更聪明。”^③

主张以技术角度研判历史的睿智史学家黄仁宇,在其《万历十五年》^④中指出:仅靠儒家的思想管理一个国家是不够的,还得有点别的;中国必须从一个靠尊卑有序来管理的国家,过渡到一个靠数目字来管理的国家。传统中国之所以未能成功转型,无关道德和个人因素,而是在技术上不能实现“数目字管理”。不过,在现实中的判断或决策,都应当尽量避开数字的误导。

二、统计学的贬责

由于有些人对统计学知识懂得不多,对统计方法更是一知半解,因此,长期以来对统

① 达瑞尔·哈夫等著,《怎样利用统计撒谎》,中国统计出版社,1989年,第4页。

② 人物介绍

陈希孺,1934年1月生于湖南望城县,1956年武汉大学数学系毕业,同年入中国科学院数学所任研究实习员,1961年调往中国科技大学数学系任教,1986年调入中国科学院研究生院。1978年任副教授,1980年晋升教授,1981年获准为首批博士生导师。新中国首批18位博士学位获得者中3位是他指导的,在人民大会堂举行的学位授予仪式上,这些博士受到总理接见。1997年被选为中国科学院数学学部院士。曾任中国统计学会副会长、中国概率统计学会理事长、中国现场统计学会理事长。2002年当选为IMS的Fellow。他是中国线性回归大样本理论的开拓者。在参数和非参数统计领域做出了具有国际影响的工作。他解决了在一般同变损失下位置-刻度参数的序贯Minimax同变估计形式和存在问题;在序贯、固定和两阶段抽样机制下,作为分布泛函的一般参数存在精确区间估计的条件,推翻了国外学者有关猜测;其U统计量逼近正态分布的非一致收敛速度的研究成果被美国“Encyclopedia of Statistical Sciences”引用,苏联学者在“Theory of U-Statistics”著作中专门详尽地论述了他的这一工作;在广义线性模型和自变量带误差线性回归模型研究中取得若干重要成果。曾与美国P. R. Krishnaiah教授和C. R. Rao教授开展过合作研究。

③ 汤其林、柯惠新编著,《使人聪明的技术》,人民交通出版社,1996年,第3~4页。

④ [美]黄仁宇,《万历十五年》,中华书局,2007年。

计方法论产生了种种误解。比如,《怎能利用统计撒谎》一书中写道:“一份装帧精致的统计资料,胜过希特勒的‘弥天大谎’,它使人误入歧途,却不负任何责任。”这是由于没有诚实的态度来深刻理解统计语言的作者和虚心领会统计语言含义的读者,结果造成以上表述语义上的荒诞不经,最终导致人们对统计的种种错觉。

反对统计方法论的人说:“统计分析经常意味着对有分歧的数字进行篡改,用意义含糊不清的办法去解决不明确的问题。”^①还有误用这样的比喻来批评人运用统计:“它使用统计就如同一个醉汉使用街灯一样,是用来支撑他,而不是用来照明。”^②

19世纪70年代美国著名作家马克·吐温(Mark Twain, 1835—1910),常引用英国保守党领袖、曾两度出任英国首相(1868、1874—1880)本杰明·迪斯雷利(Benjamin Disraeli, 1804—1881)的名言:“世界上有三种谎话:谎言、糟透的谎言、统计数据(There are three kinds of lies: lies, damned lies, and statistics)”。显然,这是一种情绪逐渐加重的语句,蕴涵的意思为:比谎言更糟糕的是糟透的谎言,而比糟透的谎言更糟糕的是统计数据。其本意是嘲讽官僚为了政绩而操纵统计数据,导致统计数据虚假、失真,从而无形地剥夺了公众的知情权。其实,统计数据本身并无罪,随着统计法的实施和完善,统计数据造假现象正在根除,但是,由于一些官员或学者动辄引用统计数据支撑自己的观点或立场,在各种场合几乎见不到没有数目字的决策或主张。人们普遍认为,只有得到数字支持的判断才是可靠的,只有数字化的决策才是科学的。社会各阶层对数据的痴迷和滥用,从而出现了像本杰明·迪斯雷利和马克·吐温一样的对统计数据愤世嫉俗地诅咒。本杰明·迪斯雷利也是一位小说家,作为本杰明·迪斯雷利、马克·吐温的读者,对上面这句话就应当像对待他们创作小说中的情节一样,都可以认为是虚构的,不能当真!

统计以定量方式反映现实,人们必须恰当地使用均值、百分率等指标数据描述社会经济现象。统计数据就如同望远镜,它既可以把小东西放大,又可以把大东西放小。德国学者瓦尔特·克莱默在《统计数据的真相》书中表述道:“增长率的增长率是统计学家的‘毒药柜’,他们既可以把一个平淡无奇的数据吹嘘夸张成为让人为之激动不已的指标,又能够把正面的、良好的数据全部淹没在一大堆看起来丝毫没有任何价值的指标中,通常来说,统计学家并不使用真正的、真实的基础数据。”^③

美国著名统计学家达莱尔·哈夫在《统计陷阱》^④论著中写到,从大类上分,操纵统计数据方法有:样本选择偏差、平均数误导、遗漏部分数据(对比、显著性等)、平面图形(偷换维度)等方法。

三、统计学的正名

由于统计学与不确定性有关,所以,常被人滥用来作为“谋生”的手段,由此统计学就

①② [英] H. T. 海斯莱特著,《简明统计学》,黑龙江人民出版社,1981年,第2~3页。

③ [德] 瓦尔特·克莱默著,隋学礼译,《统计数据的真相》,机械工业出版社,2008年12月1日,第四章,第36~45页。

人物介绍:瓦尔特·克莱默,1948年生,德国多特蒙德大学经济和社会统计学教授。

④ [美] 达莱尔·哈夫著,廖颖林译,《统计陷阱》,上海财经大学出版社,2002年6月。

人物介绍:达莱尔·哈夫,1913年出生于美国爱荷华州,获爱荷华州立大学(the State University of Iowa)学士和硕士学位。在《时尚先生》、《纽约时报》、《星期六邮报》和《哈波斯》等美国顶尖纸媒上发表文章。1963年被授予国家学院钟奖(National School Bell)。

无故地背上了“臭名昭著”的恶名。

尽管统计学不像以上说的那样,但是在一定程度上至少反映了这样两种情况:一是统计数据被滥用;二是统计知识普及程度不够。也许有些人提出他们明知是不能证明的假设,或者,也许有些人事先已经做出了结论,却再去挑选样本或篡改数据反过来证明他们的结论,这些都是故意滥用统计方法的做法。

我们似乎都有一种数字情结或崇拜症,特别迷信数字。毫无疑问,人类已经昂首阔步地跨进了数字化时代,我们每天都生活在数字(包括数字通讯)的汪洋大海中。

“爱”与“恨”就如同“形”与“影”,它们总是相互伴随、形影不离。凡是有人喜爱的,就会有人憎恨,统计数据的处境和遭遇也不例外。

相关性与因果性混淆使用,导致诉诸权威的逻辑性谬误;依据小样本衍生出规律性结论,犯了以偏概全的错误。毋庸置疑,只有学会了统计,才能“驾驭”数据,就能够知道怎样说理!

第二节 统计学的诠释

英语统计一词 Statistics 的词源出自中世纪拉丁语 Status 和 Statista。Status 译为各种现象的状态和情况;Statista 译为通晓政治和熟知各国国情者。以这个词根 Stat-构成的意大利语 Stato,则表示国家以及有关各国的国家结构和国情知识的总称。

1749 年,德国哥廷根大学的高特弗里德·阿亨瓦尔(Göttfried Achenwall, 1719—1792)在他的《近代欧洲各国国势学论》中率先使用“统计学”这个名词,他把“国势学”(“Staatenkunde”)命名为 Statistik,即统计学。直到 18 世纪末,德语统计一词 Statistik 才由齐默尔曼(E. A. W. Zimmerman)译成英语 Statistics,同时赋予了它用数字表述事实这一新内涵。在我国古代,统计与合计、总计意思相近。现代意义上的统计,大约是在 20 世纪初期才传入我国。

统计一词的英文 Statistic 大致有以下三种含义:

- (1) 它表示事实、数字、图形和表格。在这个意义上,这个词为复数。
- (2) 它表示从样本资料推算出来的数量群。在这个意义上,这个词也为复数。
- (3) 它表示拥有它自己的术语、方法论和知识体系的学科。在这个意义上,这个词为单数。

而在我国国内普遍把统计一词释为这样三种涵义,即统计工作、统计资料和统计学。统计工作是指搜集、整理、分析和提供数字资料的实践活动;统计资料是指统计实践活动过程所取得的各项数字资料;统计学是关于大量数据资料的收集、整理、描述和分析,以认识客观现象总体数量特征和数量关系的科学。它们三者关系是:统计工作与统计资料是统计活动过程与成果的关系,而统计工作与统计学是统计实践与理论的关系。

统计学(statistics)根据它的研究对象和应用范围,可以分为描述统计(descriptive statistics)和统计推断(inference statistics)。描述统计是以卡尔·皮尔逊(K. Pearson, 1857—1936)为代表,它是对搜集到的大量数据资料进行加工、整理、列表、图示,以及计算综合指标,用以探索数据内在的数量规律。推断统计是以费歇尔(R. A. Fisher, 1880—1962)为创始人,它是借助抽样技术,从样本推断总体,以达到认识总体数量特征,对不确

定的事物做出决策。推断统计有估计和检验两种形式。统计估计是以从总体中抽取有限样本为依据,从而对总体的某一数值做出估计;统计推断是事先作出某种假设,根据总体分布特征,以实验所得的数据或以样本数据计算结果,来检验事先的假设是否可以接受。应用统计就是描述统计和推断统计的方法的具体应用。

由此可知,统计学具有研究现象数量方面的特点。

因而,统计学研究对象就是指人们所要认识的独立存在于人的主观意识之外的事物客体,即现象总体的数量方面——现象总体的数量特征和数量关系。

统计学研究对象客体具有数量性、总体性和变异性的特点。

一是数量性。一切客观事物都有质与量的两个方面。统计对现象的定量认识是建立在定性认识的基础上的,统计研究是密切联系现象的质来研究现象的量的,仅这一点与数学研究抽象的数量关系是迥然不同的。

二是总体性。由于个别现象通常有其特殊性和偶然性,所以统计研究应从个别入手,通过对个别单位的具体事实的调查登记,以达到认识现象总体特征的目的;而又由于总体现象具有相对普遍性和稳定性,所以总体现象表现出的特征带有一般的规律性,因而,统计的数量研究应当是对现象总体中各单位普遍存在的事实进行大量观察和综合分析,以得出反映现象总体的数量特征,而不能仅从对个别或少量不具有代表性的单位的观察和分析,匆忙得出一般性结论。

三是变异性。统计上把总体各单位由于随机因素引起的标志表现的差异称为变异。统计研究同类现象总体的数量特征,它的前提条件是总体各单位至少有一种特征表现存在着差异,并且该差异不是由某种固定的原因事先给定的。若总体各单位不存在任何差异,如同用同一细胞“克隆”的动物无差异一样,那么也就用不着统计;若总体各单位之间的差异是按照已知条件事先可以推论的,那么也无需统计方法。

第三节 统计学的史论^①

未渗入系统统计理论的统计实践活动经历了漫长的艰苦探索过程,时至封建社会末期,统计活动发展到了一定程度,积累了一定的经验、材料和方法,客观上要求统计实践上升到理论。真正意义上的统计学理论的研究开始于古希腊亚里士多德时代,自这时起,统计学历经了三大发展时期。

一、古典统计学时期

在公元前3世纪至公元17世纪期间,出现了具有浓厚的哲学色彩的用文字记述政情的统计。最早始于古希腊的亚里士多德(Aristotle,公元前384—公元前322),他用文字对希腊城邦政情进行比较叙述,撰写了许多有关城邦政情的论文。为此,被后人称为“城邦纪要形态的统计”。^②到17世纪和18世纪,在欧洲特别是在德国,城邦纪要形态的统计受到空前的重视,并被广泛地应用,逐渐形成了一个学派——记述学派或称国势学派,这是统计学史上第一个学派。

① 夏南新,《统计学史论》,《统计与预测》,1999年第2期。

② 胡孝绳著,《统计学》,小木屋社,1976年,第2页。

(一) 记述学派

这是一个有统计学之名,而无统计学之实的学派。这一学派的代表人物是德国的海耳曼·康令(H. Conring, 1606—1681)和高特弗雷德·阿亨瓦尔(G. Achenwall, 1719—1772)。康令将国家土地、生产及其他的数字资料作为“国情纪要”课程的教材。阿亨瓦尔所编《近代欧洲各国国势学论》,对欧洲各国的物产、人口、土地、兵力等状况也作了详细的叙述和比较。国势学原意是记述国家的显著事项的学说,它是研究各国基本制度的学问,当然,它也是当时用来歌颂普鲁士君主政体的。尽管阿亨瓦尔于1749年把国势学定名为统计学,但是,它实际上只是徒有其名。准确地说,它是一种历史科学。因而,这也就是后来出现“统计是静态的历史,历史是动态的统计”之说的原因之一。

然而,到此间后期,继承记述学派衣钵的人逐渐分化,出现了主张专用文字记述的文字记述学派和主张用统计表、图形记述的图表学派。即便后期有了学派内部分化和论争,但是他们仍然还没有摆脱“城邦纪要”的窠臼,最终依然表现为纯粹哲学形态的统计分析。理所当然,国势学是属于实质性的社会科学。

不过,真正以总体数量比较的方法研究分析社会经济问题却是起源于英国伦敦,其主要理论和方法是出自《政治算术》一书中。

(二) 政治算术学派

这是一个无统计学之名,而有统计学之实的学派。这一学派的代表人物是英国的威廉·配第(W. Petty, 1623—1687)和约翰·格朗特(J. Graunt, 1620—1674)。威廉·配第的代表作是《政治算术》和《爱尔兰的政治解剖》。《政治算术》^①成书于1671—1676年,遗憾的是却是在配第去世之后的1690年才在伦敦出版发行。应当指出的是,配第的著作带有鲜明的政治倾向,它是为当时新兴资产阶级服务的。

英国是最早完成资产阶级革命的国家。随着资本主义经济的兴起,英国、法国和荷兰不断向国外扩张,然而,英国统治者在这场争霸过程中总是感到严重不安。为此,配第在他的这两本著作中,以数字资料为基础,用计算和对比的方法,分析英、法、荷三国的政治、经济、军事等方面的实力,论证出英国可以超过法国和荷兰而称雄世界,以实证分析的结果批驳了英国的悲观论调。配第的政治算术原本应从属于他的政治经济学的一种研究方法,他给政治算术下的定义是“对于人口、土地、资本、产业的真实情况的认识方法”^②。他主张用数字、重量和尺度来表达自己的问题。只进行诉诸人们感觉的议论,而不采用比较级或最高级的词语进行形而上学的思辨式议论。他不仅利用数字说话,而且还开始利用朴素的图表形式概括和显现数字资料。

不过,发表第一篇关于政治算术文章的人却不是配第,而是配第的好友格朗特。格朗特是一名商人。17世纪中叶,欧洲各国发生了周期性的瘟疫,人口大量死亡,于是,格朗特在1662年发表了一篇《对死亡表的自然与政治观察》的论文。该论文以人口问题为研究中心,利用大量数据对社会人口变动进行统计分析,从而发现人口的出生与死亡率经久不变,据此创立了“大量恒静定律”(Law of Inertia of Large Numbers)。“大量恒静定律”是指某一事物的某一性质或状态在反复观察或反复试验中保持不变。它成为当时统计学的基本原理。格朗特根据这一定律按照不同年龄的死亡率与人口寿命编制出了最早的生

① 威廉·配第,《政治算术》,陈冬野译,商务印书馆,1960年。

② 杨坚白、莫日达等著,《统计学原理》,上海人民出版社,1987年,第8页。

命表。

18 世纪的德国牧师苏斯米尔希(J. P. Süssmilch, 1707—1767)被视为系统地论述政治算术的人。他写了一本《神定秩序》一书。他是以“神定秩序”来说明社会现象的。其实,他论证的并非神的秩序,而是通过大量观察,证明了大数法则的存在。直到 1850 年,德国经济学家克尼斯(Kniex, 1821—1896)在《独立科学之统计学》中,主张将政治算术更名为统计学,至此统计学才成为有政治算术真正统计内容的有名有实的统计学。

综观上述可知,政治算术学派是以数量分析为特征,研究客观现象数量关系。就其内容和方法来看,政治算术应当是统计学的正统起源。不过,从其本质上来讲,他仍然未能在根本上从政治经济学中分化出来。这一派所探讨的规律还都是一些用数字表示的实质性的社会经济规律,因此,政治算术还是属于实质性的社会科学。

二、近代统计学时期

近代统计学起源于 19 世纪概率论(又称或然率的数学理论)的思想方法被引进统计学。

早在 16 世纪初就开始了或然率的研究,当时在意大利的目的是试图解决机遇游戏及其赌博输赢的得点问题。17、18 世纪许多数学家探讨了这个问题,如瑞士的贝努里家族。佳科·贝努里(Jokob Bernouili, 1645—1705)系统论证了概率的客观存在。他在 1733 年出版的《概率原理》一书中率先提出大数法则——贝努里定理。1786 年法国数学家和统计学家拉普拉斯(P. S. Laplace, 1749—1827)撰写了一篇关于巴黎人口的出生、婚姻、死亡的文章,文中提出根据法国特定地方的出生率来推算全部人口的问题。当时,法国的资产阶级革命推翻了统治法国 200 多年的波旁王朝,处死国王,并宣布共和。拉普拉斯的建议受到新政权的赞同,于是抽取 1799—1802 年 30 个县市人口资料来推算全国人口数,并测出了误差区间。就它的统计方法论而言,他的确是一个巨大创举,从此正式开始实现了统计学与概率论的联系。拉普拉斯认为应以研究自然科学的方法来研究社会现象。比利时的阿道夫·凯特勒(L. A. J. Quetelet, 1796—1874)由此深受启发,他为了学习概率论曾专门就学于拉普拉斯。凯特勒开辟了运用概率论认识随机现象数量规律性的理论和方法的新领域。1867 年德国数学家威特斯坦(T. Wittstein)发表了《数理统计学及其在经济学和保险学中的应用》的论文,并将凯特勒创建的新领域命名为数理统计学。随着数理统计学的产生和进一步发展,关于统计学究竟是研究客观现象规律性的实质性科学还是方法性科学这一问题有不同回答,后来逐渐形成了两大学派——近代数理统计学派和近代社会统计学派。

(一) 近代数理统计学派

这一学派的代表人物是凯特勒。他博学多才,既是数学家、统计学家、物理学家、天文学家和人类学家,又是社会活动家。他担任比利时统计委员会主席时,在英国支持下发起召开国际统计会议。他著有《论人类》(1835 年)、《概率论书简》(1846 年)和《社会物理学》(1869 年)等 60 余种书。他在统计理论上的重要贡献是率先把概率引入统计学。他认为正态分布适用于各学科,而正态规律只有借助概率论才能够确切地解释。他对社会领域的规律性的探索是他一大主要成就。他在《社会物理学》著作中利用大数定律论证了社会生活中的随机偶然现象贯穿着必然的规律性。他认为任何现象都可能出现偏差,只有通过大量观察和计算平均数等方法,才能够排除现象的偶然性,从而反映出现象有规律的必

然性。为此,他运用概率论原理提出了“平均人”的概念。所谓“平均人”就是运用统计方法计算的人类自身各性质标志的综合平均值。通过计算“平均人”就可以用来探索社会规律。若社会所有的认同平均人的差异愈小,社会矛盾就愈易缓和。凯特勒为了解决统计计量上的准确性,创造性地运用概率论手段研究人口、人体测量、婚礼、自杀和犯罪等诸多问题。鉴于他所作的贡献,他被当时资产阶级学者奉为“近代统计学之父”,他所处的19世纪上半叶被称为统计时代。凯特勒把统计方法发展为既可应用于社会现象研究,又可应用于自然现象研究的通用方法,从此,统计学就不再是单纯的社会科学了。他在《论社会制度》一书中将力学规律与社会现象相类比,提出社会现象与自然现象一样也有客观规律可循的正确观点。然而,应当指出的是,凯特勒及其一批追随他的近代统计学者所探讨出的规律性,不过是被研究对象通过统计方法在数量变化上表现出来的统计规律性。在这个意义上讲,近代统计学实际上已经不再是实质性的社会科学,而是可以通用于揭示社会现象和自然现象的统计规律的方法论科学。

数理统计学在凯特勒创立的通用的统计学的母体中孕育成长。至19世纪末期开始,由于英美数理统计学派葛尔登(F. Galton, 1822—1921)等学者把数理统计学应用于生物学等领域,加之当时进化论的确立,细胞学的发展,以及能量守恒定律的肯定,致使数理统计从统计学中分立出来自成体系。这样,凯特勒当之无愧地被公认为统计学史上第三个学派——数理统计学派的奠基人。

(二) 近代社会统计学派

正值英美数理统计学派异军突起之际,以德国为中心的社会统计学派便悄然地登场。社会统计学派的主要代表人物是恩格尔(C. L. E. Engel, 1821—1896)和梅尔(G. V. Mayr, 1841—1925)。尽管该学派不同意国势学派的所谓统计学,他们认为统计学不是方法论的科学,而是采用大量观察法研究社会集团的实质性的科学。为此,欧洲、美国和日本的统计学界都受其影响,但是,从学术渊源上看,他们实际上融会了比较记述和政治算术的观点,又继承和发扬了凯特勒同时兼顾研究社会现象的传统,并且把政府统计与社会调查结合起来,进而形成社会统计学。

随着社会科学学科发展及其定量化发展的趋势,把更多统计和数学方法运用于统计实务已经成为必需。从恩格尔起,就逐渐开始融合规律论和方法论方面的分歧。尔后,他们的追随者逐步从实质论向方法论转化。比如,第一次世界大战后,社会统计学的代表人物日本的蜷川虎山,主张吸收以概率论为基础的数理统计方法,当时在日本称其为“蜷川理论”。尽管出现了向方法论过度的倾向,但是,他们仍然强调以事物的质为前提,认为社会现象的核心即质的规律性是不可能一概转化为以量来表示的。

三、现代统计学时期

从20世纪20年代开始,统计学在其内容有了一定丰富的基础上进行了相当大的改进和深入的发展,而其中又数数理统计学发展迅猛。数理统计学派的观点逐步占据了优势地位,从而使得数理统计学在欧美西方国家的应用日益广泛,以致现在在欧美等地区,一提到统计学,便毫无疑问地就是指只计量不计质的数理统计学。当然,此间社会统计学依然存在,不过已转为以苏联为中心了,并且在一些社会主义国家中得到传播和发展。

(一) 现代数理统计学派

这一学派的主要代表人物是英国数学家戈塞特(W. S. Gosset, 1876—1937)、英国统

计学家费歇尔(R. A. Fisher, 1890—1962)和波兰统计学家奈曼(J. Neyman, 1894—?)、爱古·皮尔逊(E. Pearson)(K. 皮尔逊之子)。这一学派着重研究如何从样本提供的信息推论总体,在数理统计中的随机抽样基础上建立起推断统计学。戈塞特创立了以小样本代替大样本的方法——T分布理论,为样本资料的统计分析开创了新纪元。他的理论方法被费歇尔充实和发展,他与奈曼、皮尔逊共同领导了关于“统计实验结果显著性的测定理论和估计理论”的研究活动。尔后,美国统计学家瓦尔德(A. Wasld, 1902—1950)将统计估计和假设理论进行了归纳,建立了决策理论。美国的科克伦(W. G. Cochran, 1909—1980)等学者在1957年提出实验设计的理论和方法,从而进一步拓展了数理统计学的应用新天地。自20世纪60年代以后,逐步创立出抽样技术、非参数估计、多变量分析和时间系列分析等数理统计学的新分支;建立起计量经济学、生物统计学、天文统计学等以数理统计为基础的边缘学科。

(二) 现代社会统计学

这一学派的代表人物主要是莫斯科大学教授丘普洛夫(1842—1908)。他的《统计学教程》就是传播梅尔的统计思想的。他主张统计学是一门社会科学,并认为统计学的特点是利用大量观察法探讨社会现象,以发现其规律。不过,从学术渊源上看,苏联社会经济统计学是深受德国的社会统计学派的影响的,它是恩格尔、梅尔等德国学者的理论的发展。

社会经济统计的根本性质和作用认识社会。从20世纪30年代起,苏联曾一度批判统计消亡论、数理统计学派以及通用论与方法论。在1954年召开的有760人参加的苏联统计科学会议上,以会议决议的形式给统计学定义为:“统计学是一门独立的社会科学,它在与质的方面的密切联系中研究大量社会现象的数量方面,研究社会发展规律在具体时间和地点条件下的数量表现。”决议还强调马克思主义哲学和政治经济学是统计学的理论基础。决议认为数理统计学是数学分支,但是统计学可以利用数理统计学的方法。至20世纪70年代,有学者开始反对这一决议,持反对意见者声称社会经济统计学是方法论学科,且数理统计学是统计学大家族中的一员。

至于我国统计界,早期统计学观点的形成和发展,受到过1903年从日本翻译过来的社会统计学派的学者横山亚男的《统计学》著作中的统计思想的影响。新中国成立以后,我国统计主要是学习苏联的,而且一学就是30年,当然还接受了苏联1954年统计学科学会议决议中的观点。在这30年,我国统计机构和管理模式都是在苏联成套经验基础上建立起来的,指导我国统计活动的理论也是全盘照搬苏联的以马列主义政治经济学为理论基础,以社会经济现象的数量方面为对象的社会经济统计学。当时,我国统计界普遍认为统计学只是指社会经济统计学,而数理统计是数学的一个分支。

随着苏联解体,实质性统计学而今又面临西化(方法论的数理统计学)的挑战。为此,我国统计是随波逐流,还是我独自以不变而应万变呢?应辩证地回答此问题。我国是一个拥有13亿人口的大国,照情理上讲怎么能随着别人一哄而起呢?即便是科学,尤其是应用型科学,也还是讲究实用。而社会经济统计学在中国大有用武之地,那就应当让它繁华叶茂。绝不能刻意迷信“洋货”(指西方国家唯一指的统计学——数理统计学)。全国有些高校不得已将统计系更名换牌,这是当时我国社会主义初级阶段的权宜之计,是社会主义市场经济不发达的表现。纵观统计学历史,我们了解到统计学有着广阔的发展和应用空间,我们决不能对其短视。