



21世纪高等院校教材

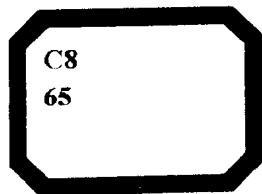
应用统计学

耿修林 谢兆茹 编著



科学出版社

www.sciencep.com



21 世纪 高 等 院 校 教 材

应 用 统 计 学

耿修林 谢兆茹 编著

科 学 出 版 社

2 0 0 2

内 容 简 介

本书系统地介绍了现代统计学的基本理论和常用的方法,全书共分为15章,内容包括:统计资料的来源及其质量,数据资料的描述方法,概率论初步,统计分布及抽样分布,统计估计与假设检验,非参数方法,时间序列分析,相关与回归分析,方差分析与试验设计,统计决策。本书从统计应用的角度出发,力求运用浅显的语言叙述和介绍统计学的思想和方法。

本书可作为财经类、工商管理类本科生的教学教辅材料,也可供工商管理硕士研究生及统计工作者学习参考。

图书在版编目(CIP)数据

应用统计学/耿修林,谢兆茹编著. —北京:科学出版社,2002

(21世纪高等院校教材)

ISBN 7-03-010284-3

I. 应… II. ①耿… ②谢… III. 应用统计学 IV. C8

中国版本图书馆 CIP 数据核字(2002)第 016454 号

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

新蕾印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2002年7月第 一 版 开本:B5 (720×1000)

2002年7月第一次印刷 印张:30 1/2

印数:1—4 000 字数:542 000

定价: 36.00 元

(如有印装质量问题,我社负责调换〈北燕〉)

前 言

从数量角度研究社会经济问题的意义已为大多数人所认识,通过数据资料对客观现象数量方面进行分析,不仅能将认识推向深化和精确化,而且还能对理论及其猜想进行论证和说明。在这一过程中,统计学的理论方法提供了重要的工具。统计学是一门应用性很强的学科,在社会经济管理和科学研究活动中能够发挥重要的作用,相信学了统计学之后,读者能感受到这一点。

本书是在参阅国内外优秀统计学著作和教科书的基础上,并结合近年来我们承担统计学课程教学的体会编写出来的,先后持续了相当长的一段时间。在本书的写作过程中,我们特别考虑了以下几方面的要求:①注意用通俗易懂、深入浅出的语言阐述统计学的思想,以便初学者和经济管理类学生容易学会统计学研究问题的思维方式。②对一些比较复杂的统计方法,我们从方法产生的背景、原理出发,循序渐进地进行介绍和讲解,最后还给出精选的例子,目的是使学生在理论学习方法的同时,还能产生对统计学的兴趣,加深对统计学解决问题技巧的感性认识。③对一些重要概念的解释,考虑到经济管理学学生的数学背景,尽量使用浅显的文字进行叙述,但又力求不失数学上的严格性,以使那些想进一步学习统计学的学生能打下一个良好的基础。④我们长期从事统计学的教学和科研工作,比较了解学生在专业课程学习过程中对统计方法的需要,因此,本书所介绍的统计方法大多数是经济管理理论研究和实践工作中经常用到的工具。⑤在本书编写的过程中,注意吸收了当今统计学发展的一些新的、颇具实用价值的成果。总之,这本教科书内容系统,介绍完整,条理清晰,通俗易懂。

在本书的编写过程中,我们参考了大量的统计学著作和有关文献资料,在此,我们要向这些作者表示深深的谢意。尽管投入了许多时间和精力,但书中难免还会存在一些问题甚至是错误,为此,我们欢迎一切善意的批评和指教。

作 者

2002年初于南京大学

目 录

前言

第1章 绪 论	1
1.1 统计学的渊源及其体系	1
1.2 统计学的概念	7
1.3 统计学的性质与特点	9
1.4 统计学的应用	11
第2章 统计数据的来源与数据质量	14
2.1 统计数据的来源	14
2.2 统计数据的测量尺度	18
2.3 统计数据质量问题	22
第3章 数据分布与图示法	28
3.1 几个概念	28
3.2 统计频数分布	30
3.3 统计资料图示与频数分布类型	39
3.4 探索性数据分析	48
第4章 频数分布的特征数字	53
4.1 引言	53
4.2 集中趋势	53
4.3 离散趋势	74
4.4 偏度与峰度	82
第5章 概率原理	87
5.1 基本概念	87
5.2 概率	93
5.3 概率运算法则	98
5.4 全概公式与贝叶斯定理	104
第6章 随机变量与概率分布	107
6.1 随机变量	107
6.2 概率分布及其种类	108
6.3 随机变量的期望、方差及矩	112
6.4 重要的离散型概率分布	116

6.5 重要的连续型概率分布	128
第7章 抽样分布	138
7.1 参数、统计量与抽样分布	138
7.2 大数法则与中心极限定理	141
7.3 由正态分布导出的分布	144
7.4 常用的抽样分布	148
第8章 参数估计	160
8.1 综述	160
8.2 点估计	161
8.3 区间估计	175
第9章 参数假设检验	190
9.1 参数假设检验的基本问题	190
9.2 总体均值检验	195
9.3 总体成数检验	202
9.4 总体方差检验	205
第10章 非参数检验	210
10.1 非参数统计方法的评介	210
10.2 χ^2 检验	211
10.3 柯尔莫哥洛夫与斯米尔诺夫检验	218
10.4 符号检验	222
10.5 秩和检验	226
10.6 游程检验	229
10.7 多个样本的非参数检验	232
10.8 秩相关检验	236
第11章 相关与回归分析	240
11.1 相关分析	240
11.2 回归分析简述	253
11.3 简单线性回归	256
11.4 多元线性回归	277
11.5 非线性回归模型	286
第12章 方差分析与正交试验设计	290
12.1 方差分析原理	290
12.2 单因素方差分析	293
12.3 双因素方差分析	304
12.4 试验设计基本问题	321

12.5	正交试验分析方法——极差分析·····	328
12.6	正交试验分析方法——方差分析·····	334
第 13 章	时间序列分析 ·····	342
13.1	概述·····	342
13.2	时间序列的对比分析·····	346
13.3	时间序列的古典分析·····	353
13.4	时间序列的现代分析·····	375
第 14 章	抽样调查 ·····	388
14.1	抽样调查的基本问题·····	388
14.2	简单随机抽样·····	394
14.3	分层随机抽样·····	407
14.4	整群抽样·····	417
14.5	系统抽样·····	426
14.6	多阶段抽样·····	433
第 15 章	统计决策 ·····	443
15.1	概述·····	443
15.2	不确定型决策·····	446
15.3	风险型决策·····	450
15.4	决策树·····	462
15.5	贝叶斯决策问题·····	466
附录	常用统计表 ·····	470

第1章 绪 论

1.1 统计学的渊源及其体系

1.1.1 统计学的产生与发展

统计活动源远流长,可以说,自从人类社会有了数的概念,有了计数活动,统计就开始了。但作为一门科学的统计学,它的出现却晚得多。统计学究竟产生于什么年代,迄今为止人们的看法还不尽一致,不过多数人认为,统计学大概兴起于17世纪。

17世纪中叶,在英国首次出现了有意识地用数字语言说明问题,注重从数量角度探索客观事物变化规律的研究活动。配第(W. Petty)《政治算术》(1676)一书的问世,标志着统计学的肇端。《政治算术》揭示了一些经济学的科学原理,研究了许多经济范畴和经济关系,所以它是一部重要的经济学著作,但与此同时,它又是一部很有价值的统计学文献。配第在这本书以及其他有关著作中,采用了不同于前人的研究方法,明确地用大量的数据资料分析问题,试图把结论建立在可靠的事实根据上。正如他在《政治算术》的“序言”中所写到的,“我进行这种工作所采用的方法,在目前还不是常见的,因为我不使用比较级或最高级的词语只进行思辩式的议论,相反却采用了这样的方法,即用数字、重量和尺度来表达自己的想说的内容,进行诉诸于人们感觉的议论,借以考察在自然中有可见根据的原因”。在统计史的研究中,一般把以配第为代表的关于社会经济现象“算术”式的研究,称为“政治算术”统计学。“政治算术”还有一位重要的人物,就是英国人格朗特(J. Graunt),他的代表性著作《关于死亡表的自然观察与政治观察》(1662),通过对人口变动数据的分析,揭示了一系列的人口变化规律,尤其是他在这本书里运用了不少新颖独特的资料整理方法和估算方法,给后来统计学的发展留下了有益的启示。

与“政治算术”产生的时期差不多,在欧洲的大陆国家德国,也出现了一门统计学——“国势学”。“国势学”主要研究“国家的有关显著事项”,即有关国家兴衰强弱的重大问题与治理国家必备的知识。它之所以也被称为“统计学”,其中一个重要的原因是,“国势学”的代表人物阿肯瓦尔(G. Achenwall)给“国势学”起了一个统计学(statistik)的新名称。这一德文词汇转译成英文“statistic”后,逐渐得到人们的接受,从而影响日深。康令(H. Conring)是“国势学”的奠基人,他对“国势学”的研究目的、研究对象、研究方法等基本问题作

了深入细致的阐述。经过 100 多年的发展,到 18 世纪中期,“国势学”达到了顶峰。阿肯瓦尔在继承康令开创的研究体系和研究方法的基础上,全面发展了“国势学”,他的《欧洲主要国家国势学纲要》(1749)与康令的《国势学讲义》(1730),被奉为“国势学”的经典文献。“国势学”最大的特点是它较少使用数字工具,主要诉诸于文字记述和逻辑比较。由于没有方法论上的特点,容易造成与地理学、历史学、政治学等学科的混淆。这一点,在“国势学”后来的内部分化中,有明显的反映。

历史上,“国势学”与“政治算术”曾就哪一个才是统计学的真正起源问题发生过重大争论,持续的时间长达一个多世纪。1850 年,德国人克尼斯(K. G. Knies)审时度势,根据当时统计学发展的实践,概括了大多数人的意见,认为“国势学”尽管有统计学之名,但没有统计学之实,应该仍叫“国势学”,而“政治算术”才是真正的统计学。克尼斯的《独立科学之统计学》一文的发表,标志着“国势学”与“政治算术”的争论告一段落。此后,“政治算术”与“国势学”的影响在一定范围内依然存在,并且局部上还有新的进展,但总体上皆处于衰微的地步。

统计学另一个重要的起源是概率论。14 世纪,在工商业比较繁荣的意大利以及地中海沿岸其他地区,由于赌博游戏盛行和保险活动的萌起,人们已经对“机会”问题发生了兴趣。但真正意义上的概率论,是从 17 世纪开始的。帕斯卡(B. Pascal)和费马(P. Fermat)关于“得点问题”的讨论,奠定了概率论的基础。在早期概率论的研究中,作过重要贡献的数学家有:莱布尼兹(G. Leibniz)、伯努利(J. Bernoulli)、德·莫佛尔(A. de Moivre)、贝叶斯(T. Bayes)、拉普拉斯(F. Laplace)、高斯(C. Gauss)、勒让德(A. Legendre)、贝塞尔(F. Bessel)、德·康多塞(M. de Condorcet)、辛普森(T. Simpson)、布丰(C. de Buffon)、泊松(S. Poisson)等。其中,拉普拉斯是古典概率研究的集大成者,他给出了概率的“古典”解释,并把数学分析方法系统地引进概率论,建立了严密的概率数学体系。高斯和勒让德在误差研究过程中,提出了最小二乘法,高斯还导出了著名的正态分布曲线。“政治算术”研究的是简单的确定性数量关系,而概率统计则研究复杂的、随机性现象,这极大地充实和深化了数量问题研究的内容。以概率论为基础,统计学进入了一个新的发展时期。

凯特勒(A. Quetelet)是统计学发展史上承前启后的重要人物,他一生写了大量的有关统计学方面的著作,代表性的就有《社会物理学》(1869)、《概率论书简》(1846)、《犯罪学》(1831)、《比利时统计研究》(1829)等。凯特勒在统计学上的突出贡献,是他把概率论全面引进“政治算术”、“国势学”以及其他社会问题的研究,从而大大推动了概率论和数学方法在社会科学领域的应用,促进了数量研究由“算术”水平向“数理”阶段的迅速转化。凯特勒还是一位出色

的社会活动家,他是国际统计学会的创始人之一。

从19世纪后半期起,统计学几乎在各个方面都取得了创新性成果。其中,尤为引人注目的是,兴起于生物遗传学、农业田间试验等领域的所谓统计数学方法的发展。戈尔登(F. Galton)是生物统计学的主创者,受凯特勒的影响,他也利用正态法则研究优生学、遗传学问题,认为正态法则适应于许多情况,先后提出了“百分位数”、“中位数”、“四分位数差”、相关与回归等概念和计算方法。K. 皮尔逊(K. Pearson)是戈尔登的学生,他系统发展了戈尔登的相关与回归理论,研究了复相关和偏相关,把物理学上“矩”的概念移植到统计学中,给出了极大似然估计方法,导出了重要的 χ^2 分布。以皮尔逊为代表,以大量观察和正态分布为基础的关于总体分布曲线的研究,确立了“大样本”统计理论,奠定了“描述统计学”的框架体系。

进入20世纪,统计学经历了又一轮的飞跃。一方面新的统计思想和统计方法大量涌现,另一方面带有归纳性质的统计推断逐渐占据了主流地位。自苏歇米尔斯(J. Süssmilch)首倡大数法则到19世纪末20世纪初的大约一个半世纪里,大量观察法一直是统计思想的核心,并被理所当然地接受下来。最先对这一理论发起挑战的是戈塞特(W. Gosset),基于在酿酒公司多年的实验观察,戈塞特洞察到大样本统计方法并不适合一切场合,有时候还只能根据少量观察就必须作出结论,按他的话说,“有些实验不能多次重复进行,这个时候,必须依据极少量的实验结果作出判断,像有些化学实验、很多的生物实验和农业实验,便属于这样的情况,但它们也应该成为统计学的研究对象”。经过多年的潜心研究,1908年戈塞特终于导出了重要的 t 分布,这是“小样本”统计研究的基石。以此为标志,统计学逐渐实现了由以前的“描述统计”阶段向“推断统计”阶段、由大样本统计向小样本理论的转变。费喧(R. Fisher)是推断统计学的建立者,他对统计学进行了深入独到的研究,开辟了方差分析、试验设计等统计分支,给出了戈塞特 t 分布的简洁证明,论证了相关系数的抽样分布,提出了 t 检验、 F 检验、相关系数检验,并编制了相应的检验概率表,费喧的代表著作包括《供研究人员使用的统计方法》(1925)、《试验设计》(1955)、《统计方法与统计推断》(1956)等。费喧在统计学发展史上有着辉煌崇高的地位,美国统计学家约翰逊(P. Johnson)在《现代统计方法:描述和推断》(1959)一书中指出:“从1920年一直到今天的这段时期,称为统计学的费喧时代是恰当的,他的名字和他的著作,受到一切不带偏见的人的尊敬和传播”。

内曼(J. Neyman)和E. 皮尔逊(E. Pearson)是继费喧之后杰出的统计学家。他们二人经过多年的合作,共同完善了现代统计学的核心内容——区间估计和假设检验的理论。20世纪50年代,瓦尔德(A. Wald)提出了“统计决策理论”和质量检验的“序贯分析”,进一步推动了统计学研究和应用的范围。

随着计算机在统计中的应用,由威夏特(J. Wishart)、赫特林(H. Hotelling)等人发展起来的多变量统计又重新活跃起来。50、60年代以来,稳健统计、时间序列、抽样理论、统计诊断、探索性分析、贝叶斯统计等皆取得了重要的进展。

通过对统计学发展历史的简单回溯,我们可以看出,随着人们认识的不断深化和社会实践需要的推动,统计学始终在不断地丰富和完善着。它经历了从意义和概念不甚明确的阶段到作为一门独立学科的转化,从数量研究的“算术”水平到需要较高数学知识的“数理”阶段的转化,从确定性问题的研究到随机现象研究的转化,从大量观察消除误差干扰以达到对客观现象规律性认识的大样本理论,到控制试验次数提高数据质量的小样本推断的转化。统计学就是这样一步步发展起来的,现在它仍然充满着活力。

1.1.2 统计学的学科体系

统计学的产生虽然可以追溯到17世纪,但它的快速发展却是自凯特勒之后出现的。伴随着应用数学在20世纪的崛起,统计学的进步更是迅速有加。统计学在处理问题、拟合分析模型时,区分主要的必然性因素和次要的偶然性因素,把前者作为基本变量,后者看作随机变量一同纳入模型。仅就这一点,统计学的结果比单纯考虑基本变量的确定性分析似乎要准确一些。有果必有因,现象的结果是由多种原因引起的,如果在结果和所有原因变量之间建立关系,将会使模型复杂化。统计学对此作简单处理,把那些不便控制的因素当成随机干扰,这可能会降低模型的逼真性,但却更有利于求解。所以,统计学能够把复杂的问题转化成简单的问题,这正是统计学能够取得快速发展的重要原因之一。目前,统计学已经形成了由若干个分支组成的庞大的学科体系,已经长成了一棵枝繁叶茂的参天大树(参见图1.1)。

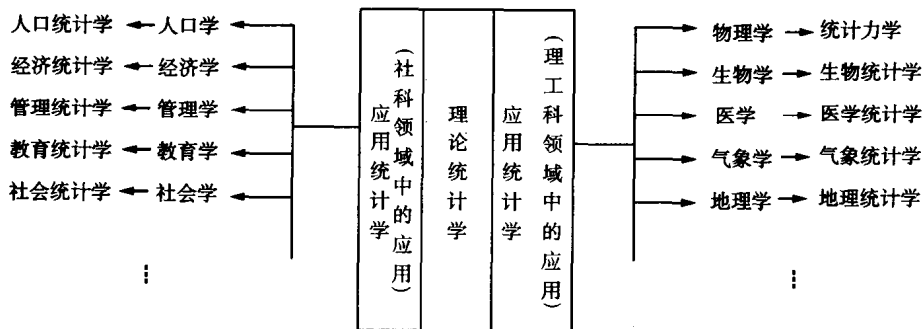


图 1.1 统计学的整体结构

这里所说的理论统计学是指,侧重于从数学学科中吸取营养,研究统计学的数学方法和基础原理,以解决统计学自身发展中的重大问题为目标的统计学研究分支。理论统计学主要包括:

(1)统计估计。统计估计是统计学的核心内容之一。它有两个方面的内容,一是指总体分布已知时,对总体未知参数或参数组合的函数进行估计,另一是指总体分布未知时,对有关分布的特征数字(如均值、方差或标准差等)以及分布密度进行估计。统计估计的主要研究课题有:估计量的确定,估计量的评价等。

(2)假设检验。参数假设检验是统计学另一重要内容。它是指在总体分布已知的前提下,根据样本资料,对总体参数的某种假设命题进行检验和判断。假设检验主要研究:检验统计量的确定,假设检验原理,检验效率等问题。

(3)抽样调查。抽样调查是搜集统计资料的基本手段之一。最初是由挪威人凯耳(A. Kiaer)倡导的,后经鲍莱(A. Bowley)、詹森(A. Jensen)、费喧、E. 皮尔逊、内曼、马哈拉诺比斯(P. Mahalanobis)、科克伦(W. Cochran)、汉森(M. Hansn)、耶茨(F. Yates)、赫维茨(W. Hurwitz)等人的努力,逐渐成了统计学的一个重要分支。抽样调查主要研究:抽样方案设计,样本抽取方法,抽样分布,抽样效果等问题。

(4)试验设计。试验设计是费喧首先明确提出来的,耶茨、哈歇巴格(B. Harshbarger)、尤顿(W. Youden)、田口玄一等人作过一定的贡献。试验设计主要研究如何安排试验方案,如何分析试验结果等问题。

(5)非参数统计。非参数统计与参数统计相对应,它主要研究总体分布未知时或不依赖于总体分布的各种统计问题。

(6)时间序列。按时间顺序排列的一组数据称为时间序列。时间序列分析大体可分为两个阶段,前期主要研究时间序列的基本结构、自回归过程等,20世纪50年代后,时间序列进入一个新的发展阶段,这一阶段的主要特点是与随机过程和数理统计结合更加紧密。时间序列的内容包括:时域描述分析、频域推断分析、时概模型、非线性系统模型、自回归估计、空间序列分析等。

(7)统计决策。统计决策理论认为,统计所研究的一切问题都可以归结为是在作“决策”。起初的意图是想用一个统一的模型把参数估计和假设检验联结起来,后来引进了博弈论、规划论等方面的知识,逐渐演化成统计学的一个理论分支。统计决策的基本概念包括:风险函数、损失函数、决策标准、决策函数等。

(8)序贯分析。序贯分析是在产品质量统计检查的基础上产生出来的,其特点是,在检查产品质量时,不事先定下抽检单位数,要视每次观察的结果决定是否增加新的样本单位。序贯分析可以节省时间、人力、物力,减小抽

样检查的破坏性损失。它的内容有:抽样方案,序贯检验统计量,判别风险等。

(9)多元统计。从一元统计到多元统计,是统计理论发展的必然趋势,也是统计研究追求“逼真”的结果。多元统计,又称多元统计分析,指的是多维随机变量的统计处理办法。主要包括:多元分布,判别分析,典型相关分析,主成分分析,因子分析,聚类分析等。

(10)统计诊断。这是近 20 年发展起来的统计学的一个新领域,统计诊断主要研究观察数据、统计模型、统计推断方法的合理性问题,并针对诊断出来的缺陷进行相应的治理和改进等。

(11)稳健统计。由于稳健统计与非参数方法、统计诊断、探索性分析有一定的联系,所以起初它的概念和学科性质并不明确。经过鲍克斯(G. Box)、塔基(J. Tukey)、瑚珀(P. Huber)、汉姆配尔(F. Hampel)等人的努力,影响越来越大。稳健统计主要研究当理论分布与实际分布不一致时,如何确定“不敏感”的统计方法。

(12)探索性分析。简单地说,探索性数据分析就是通过对观察数据进行详细的考察,力求发掘出数据本身的结构和特征,然后在此基础上建立分析模型。探索性数据分析与塔基的贡献分不开,现在对统计学的发展产生了重要的影响。

应用统计学就是运用统计思想和方法,处理实际中属于统计方面的问题。应用统计学的构成比较复杂,大体上可以归结为四类。第一类,统计计算方法,它把统计方法、数学计算方法和计算机应用结合起来,重点解决数据处理过程中所碰到的各类计算问题。统计计算方法,可看成是应用统计学的方法基础。第二类,应用统计学理论基础,同理论统计学相比,应用统计学理论基础减少了繁杂的数学符号、理论推导和公式体系,一般带有较强的应用背景,力求用简明扼要的语言,深入浅出地阐述统计学原理和方法,但它同纯粹的统计方法应用相比,又比较倾向于理论统计学。所以,应用统计学理论基础是站在理论统计学角度的应用统计学。第三类,应用统计学,这是完全意义上的应用统计,它基本上从实际问题的背景出发,着重介绍如何使用统计方法,与具体学科结合比较紧密,以致于常常被归结到所应用学科的学科体系中。应用统计学按应用的学科性质不同,可区分为应用于理工科的应用统计学和应用于社会科学的应用统计学。前者的数学味可能浓一些,而后者对数学的介绍可能更少。第四类,统计学与其他应用数学学科结合,形成了新的应用数学方法基础学科,如统计学与运筹学的结合,有博弈论、多目标决策、排队论、随机规划等。

在统计学的学科体系中,理论统计学是核心。只有理论统计研究的扩大和深入,才能更好地促进统计学的应用。当然,广泛的统计实践必将为统计理

论准备更多的素材,从而推动理论统计学的进一步发展。

1.2 统计学的概念

1.2.1 统计学认识上的分歧

什么是统计学?不同的人可能会给出不同的回答。尽管统计学作为一门科学的地位已经得到了确立,但在一些重要问题的认识上仍然存在着分歧,这或多或少影响到人们对统计学范畴的表述。因此,在解释统计学概念之前,我们先来对统计学认识上的一些分歧作一评介。

统计学的学科归属。统计学到底属于数学学科,还是属于社会科学,存在着不同的意见。一种看法认为,统计学是应用数学,属于数学的一个分支。这种观点,在从事理论统计学研究的人当中比较流行。由“政治算术”和“国势学”繁衍下来的人口统计、经济统计、社会统计、政府统计,一般坚持认为统计学属于社会科学。把统计学完全当成数学似乎有一些不足,首先统计学发展的史实没有切实得到尊重,其次也没有完整兼顾统计学广泛的实践,再就是容易导致数学化倾向,需知统计学的高度数学化和抽象化,已经阻碍了统计学的发展,使当初统计学与实际问题的结合紧密的特点几乎丧失殆尽。但仍然固守统计学只是社会科学,显然不利于统计数量研究水平的提高。针对目前的状况,有的人提出,在学科的分类上,统计学应单列出来自成一类。

统计学的性质。长期以来,关于统计学是方法论性质的学科,还是实质性学科,一直争论不休。主张统计学是方法论科学的人认为,统计学主要是研究数据资料的处理方法,目的是要提供数量认识的方法和工具。在这一派人中,由于立足点不同,又有“特定方法论”和“通用方法论”之分。“特定方法论”立足于统计学是社会科学,认为统计学主要是研究社会经济现象总体数量分析方法的,强调统计方法的特定用途。而“通用方法论”则认为,统计方法既适应于社会科学,同时也适应于自然科学,像数学一样是一切数量问题研究的公共工具。在统计学发展的早期阶段,把统计学当作是实质性学科的人比较多,要求统计学要研究客观现象的因果关系和变化规律。从现在来看,统计学是一门方法论性质的学科已为多数人所认可。

数量属性。统计学研究客观现象的数量方面,但至于研究什么样的数量也存在着不同的意见。一种意见认为,统计学研究的是抽象的数量即“计量不计质”。这种观点,同统计学是“通用方法论”的观点是一脉相承的。另一种意见认为,统计学研究的是客观现象具体的数量,它总是联系客观事物的质来研究数量,首先从事物的分析入手,通过对数量方面的认识,最后上升到质的认识,也就是遵从由质到量,再由量到质的认识过程。

在统计学研究对象上,多数人把“随机性”或“不确定性”问题当成是统计学的当然研究对象,并认为这是统计学作为一门学科存在和发展的坚强理由。可是另一部分人则认为,统计学研究客观事物确定的数量,其主要活动内容就是对客观事物的变化进行记录,并对记录的结果作出分析。事实上,统计学既研究随机现象,也研究确定性数量,在过去以确定性数量研究为主,现在则主要研究随机性问题。

就统计方法的特征和功能,存在着描述统计学和推断统计学之说。描述统计与推断统计之间没有明确的界限,只是大致上所作的划分。费暄之前的统计学,可称为描述统计学,费暄之后的统计学是推断统计学。描述统计学主要研究资料的系统搜集、整理、表述和计算。推断统计学一般研究如何根据部分观察资料,对总体情况作出具有一定可靠性的推断。在二者的关系中,描述统计学是推断统计学的基础。自推断统计学上升到主流地位后,认为统计学的实质是归纳推断的观点颇为流行。

在统计学三个多世纪的发展历程中,先后发生过若干次大大小小的争论。正常的学术争论,有利于科学的繁荣和发展,只要不囿于偏见皆值得提倡。以上所述的关于统计学基本问题认识上的差异,无疑是有趣的课题。有些问题已接近于解决,另外一些问题尚需等待学科的进一步发展才能下结论。

1.2.2 统计学的概念

概念是客观事物在人的意识中的反映。依据定义概念的手法不同,通常有描述性定义和公理化定义之分。描述性定义,顾名思义,就是用语言文字对科学概念或范畴予以简练、准确的说明。公理化定义,则是人为给定的标准或准则。

描述性定义,一般要以概念所反映的事物的存在为条件。事物发生了变化,描述性的定义解释也要相应地加以改变。所以,描述性定义应主动适应和附和客观事物而不是相反。人们对事物的观察、体会和总结不同,给出的描述性定义也随之有所差别。描述性定义,具有非惟一性、非确定性的特征。判断描述性定义的好坏或正确与否,要看它能否真实地描绘客观事物存在的方式和运动的状态。公理化定义仅是一种语言或符号,它可以完全脱离具体事物的背景,一经给出就必须遵守,要想改动它,就会引起在既定公理化定义下的一整套理论和方法体系的改变。评判公理化定义只能是它使用的语言或符号,公理化定义具有确定性、精确性、主观性等特点。

有人曾作过考察,关于统计学概念的解释不下几十种之多。为什么会出现这样的情况,原因就是采用了描述性的定义方法,加之统计学自身一直在不断地发展着,并且在认识上还存在不少的分歧。

在此,我们侧重于统计学的应用,把统计学解释为:统计学是关于资料搜集、整理、分析和解释的科学。

统计学是研究客观现象数量方面的,要达到这个认识目的,首先要搜集能够反映或说明客观现象的数字资料,这是统计活动首要的、基本的环节。搜集资料,就是计数和获取量数的过程。统计资料的质量如何,关系到能否得出客观正确的结论。因此,要搞好统计调查活动,设法提高统计调查水平。

统计整理在统计活动过程中处于承上启下的位置。一方面它是统计搜集资料工作的继续,另一方面又是统计分析的前提,能够为统计分析准备系统的资料形式。统计整理,就是运用各种合适的形式展示和表述统计资料。统计整理的内容包括:资料的审核和甄别、分组分类、编制频数分布、绘制资料分布图等。

统计分析是从已有的数据资料出发,提取有关综合信息的过程,其目的是要使一个概括的、全面的数量描述能够形成。

在统计分析的基础上,导出一般性结论的过程就是统计解释,它往往涉及到如何依据样本观察,对同类大量现象作出估计、检验、预测等问题。统计解释不是别的,其实质就是进行统计推断活动。

完整的统计活动过程,就是由资料搜集、整理、分析和解释这四个部分组成的。它们虽然各有各的任务和作用,但又存在密切的联系,只有同时做好各项工作,才能圆满完成统计活动的任务。

1.3 统计学的性质与特点

1.3.1 统计学的性质和特点

统计学是一门什么性质的科学?统计方法有哪些特点?了解这些东西,有助于正确掌握统计学的基本原理,提高统计问题的认识能力,搞清统计学的学科地位以及与其他相关学科的关系。

统计学的性质,主要表现在以下几个方面:

第一,统计学是研究数量问题的学问。统计的语言是数字,没有数字,不是从数量方面入手进行认识,就谈不上统计。无论是社会科学还是自然科学,只要出现大量数据的地方,皆用得着统计也需要统计。根据辩证唯物主义认识论,任何事物都是由数量和质量两方面组成的,是二者的统一体。统计研究客观事物数量,如果数字资料准确,统计方法运用得当,同样可以达到正确认识和反映的目的。统计研究客观事物的数量,主要包括:数量状态、数量关系和数量变化规律。统计学的数量性质,能够把它与那些非数量性质的学科,如哲学、经济学、政治学、历史学等区分开来。

第二,统计学研究的是客观现象总体的数量。数量有个体数量与总体数量之别,统计学主要研究后者,它要对大量同类现象的数量方面进行综合反映。单个数字不能称为统计,也不可能指望从它身上发现什么有价值的东西,只有对大量的现象,或对某一现象进行多次重复的观察,才有可能找到统计关系和统计规律。统计虽然研究总体数量,但必须从个体数量的调查入手,遵循由个体数量到总体数量的认识逻辑。

第三,统计学主要研究不确定性现象。所谓不确定性是指,由于受到偶然的、随机因素的作用,使得客观事物的实际数量表现存在一定程度的“不可确知性”。比如测量某一物体的长度,如果没有误差,测量一次就能确切知道结果,那就不存在统计问题了。如果有测量误差,但误差的大小完全可以掌握,这时也无须用到统计。惟有存在误差,且误差又不可预测,这时候就需要统计。现代统计学中,处理不确定性问题是其主要课题和任务。

第四,统计方法带有归纳推断的特点。统计对总体的认识有两条途径,一是对构成总体的全部事物逐一进行调查,取得全面资料,另一是从总体中抽取部分事物组成样本,然后依据样本观察结果,对总体进行推断。至于前者,运用算术方法和统计描述手段就可得到目的,而后者相对比较复杂,需要运用概率论知识和统计数学方法。实际中,全面调查与非全面的抽样调查都会用到,但由于全面调查受到诸多因素的约束,从经济性、时效性、实用性和可行性方面考虑,利用样本资料进行推断的优势比较明显。统计方法的归纳推断性质,主要是相对推断统计而言的,同逻辑学意义上的归纳推断有着明显的区别,统计推断不是从假设、命题出发,按严格的逻辑推理程序进行推断,只是基于观察到的样本资料,对总体的可能情况作出判断。

1.3.2 统计学与其他相关学科的关系

统计学与数学的关系。统计学研究客观现象的数量,不可能不用到数学,非但如此,而且还需要用到很多的数学知识。学好统计学,尤其是理论统计学,必须要有坚实的数学基础。在统计学中,数学的一些定理、运算法则同样适应,必须要严格遵守。但是也需要注意到,统计学与数学其他分科相比又有其特殊性。首先,统计学有较强的应用背景,统计方法的正确使用,不仅要有数学基础,而且要懂得相关学科的知识,具备一定的实际经验和良好的判断能力。其次,统计学主要研究不确定性问题。再就是,现代统计学的本质是归纳推断,与数学演绎方式有较大的差别。

统计学与各专门学科的关系。统计学的应用相当广泛,它在各门学科、各个领域、社会生活的各个方面,几乎都有重要的应用。仅就这一点,统计学与各个专门学科必然存在着联系。这种联系体现在,统计学能为各个专门学科