

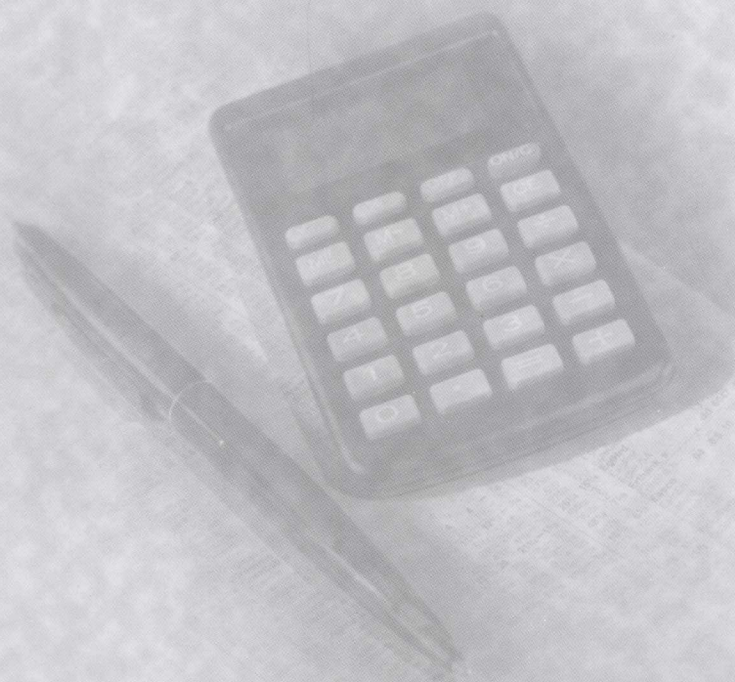


21世纪高等院校教材

应用统计学

(第三版)

耿修林 谢兆茹 编著



科学出版社

013047920

C8
74-3

21 世纪高等院校教材

应用统计学

(第三版)

耿修林 谢兆茹 编著



科学出版社

北京



北航

C1655681

内 容 简 介

本书主要介绍统计学的基本原理和常用的统计方法,内容包括绪论、数据来源与数据质量、定量数据的统计描述、定性数据的统计描述、随机变量与概率分布、统计量抽样分布、参数估计、参数假设检验、非参数检验、方差分析与正交实验设计、相关与回归分析、时序分析与动态预测和风险决策分析。

本书可作为财经类、工商管理类本科生的教材使用,也可作为工商管理硕士及统计工作者的学习参考资料。

图书在版编目(CIP)数据

应用统计学 / 耿修林, 谢兆茹编著. —3 版. —北京: 科学出版社, 2013

21 世纪高等院校教材

ISBN 978-7-03-037436-3

I. ①应… II. ①耿…②谢… III. ①应用统计学—高等学校—教材 IV. ①C8

中国版本图书馆 CIP 数据核字 (2013) 第 094228 号

责任编辑: 兰 鹏 / 责任校对: 王艳利
责任印制: 徐晓晨 / 封面设计: 蓝正设计

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>

化学工业出版社印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2002 年 8 月第 一 版 开本: 787×1092 1/16
2008 年 7 月第 二 版 印张: 19
2013 年 6 月第 三 版 字数: 448 000
2013 年 6 月第九次印刷

定价: 39.00 元

(如有印装质量问题, 我社负责调换)

第三版序

本书的第三次修订，基本上沿用了第二版的架构，主要进行的改进有以下几个方面：

(1)将原来的第三章与第四章顺序做了对换处理，做这样改动的原因是：定量数据毕竟是统计分析的主体，虽然定性资料也是统计应用的重要领域，并且其自身也发展了许多独特的统计分析方法，但适应于定量分析的很多统计方法可以直接运用到定性资料的分析中，可以看成是一般统计方法在定性资料分析中的专门化运用。所以，先讲清一般的统计特征数字计算和分析原理，然后再针对定性资料介绍相关的特征数字计算，从逻辑上方便于初学者理解和掌握。

(2)充实了第二章第一节统计数据来源的内容，特别是有关统计数据直接来源的内容，除了继续介绍统计数据直接来源的两大重要渠道(即统计调查和统计实验)外，还介绍了内容调查、参与观察等方法，此外对样本调查也区分了随机样本调查和非随机样本调查的情况。有关间接资料的来源，补充介绍了国内外重要的数据网站。

(3)在定性资料频数分布的编制中，增加了两变量频数分布表编制的知识，在定性资料特征数字的计算中，补充介绍了定性变量间相关关系计算的内容。

(4)对非参数检验的内容进行了压缩处理，删除了“柯尔莫哥洛夫与斯米尔诺夫检验”这一节，考虑到卡方检验的实际应用价值，加强了对卡方检验相关应用的介绍。

(5)删除了第二版中的第十三章抽样调查的知识，这主要是考虑到开设经济管理专业的学校多半都有专门的社会调查、市场调查、调查技术之类的课程，同时也是适应课时精简的教学组织的需要。

(6)修改和更换了书中的部分例题，增加了电子表格软件统计资料处理方法的介绍，以增强学生统计方法实现的动手能力。

从接到科学出版社责任编辑的修订通知至今已近一年时间，在这期间我们参考了国内外很多优秀的著作和教科书，吸收了对我们从事教学富有启发的素材，并陆陆续续地将其符合本书风格的有价值的知识点充实进书稿中。虽然我们很努力，但能力有限，书中存在的缺点、不足难以避免，在此，我们诚恳地接受一切批评指正。最后，特别感谢本书参考文献中的那些作者，从这些长期从事统计科学教学及科研的专家们的著作中，我们不仅了解到了最新的统计知识传播理念，而且也进一步提高了统计学专业修养。

作者

2013年1月于南京大学

第二版序

本书自出版以来，几乎每年都要重印，根据实际使用反馈的意见和科学出版社的要求，从 2007 年开始我们着手进行了修订，前后大约持续了一年多的时间。

本书在保留第一版通俗易懂、条理清晰、内容系统等特色基础上，本次修订我们精炼了内容，使体系更合理，内容更实用，增加了教学课件与案例分析支持，同时内容方面主要做了以下修订：

第一，增加了定性资料的统计描述，包括：定性数据的图示，定性数据频数分布的编制，定性数据统计特征数字的计算方法。

第二，对定量数据的统计描述做了压缩，把原来分两章介绍的内容缩写成一章。

第三，考虑到高等数学课程中，对概率论介绍比较多，在第二版的编写中我们删除了第一版的“第五章概率原理”。

第四，考虑到一学期教学时间的安排，把非参数检验的内容进行了精简，主要是删除了多样本非参数检验和秩相关检验的内容。

第五，对时间序列分析的内容只集中介绍古典分析的基本原理，并给出了具体的分析过程示例。

希望修订后的本版，能给读者学习带来方便。不知能否达到目的，还需求教于大家。

作者

2008 年 6 月于南京大学

第一版序

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,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的技巧的感性认识。③对一些重要概念的解释,考虑到经济管理类学生的数学背景,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,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

第三版序

第二版序

第一版序

第一章 绪论	1
第一节 统计学的渊源及其体系	1
第二节 统计学的概念	6
第三节 统计学的性质与特点	8
第四节 统计学的应用	9
第二章 数据来源与数据质量	11
第一节 统计数据的来源	11
第二节 统计数据的测量尺度	17
第三节 统计数据质量问题	19
第三章 定量数据的统计描述	24
第一节 定量数据的频数分布	24
第二节 定量数据的图形描述	26
第三节 定量数据的特征数字	31
第四章 定性数据的统计描述	45
第一节 定性数据的频数分布	45
第二节 定性数据的图形表示	49
第三节 定性数据的特征数字	52
第五章 随机变量与概率分布	61
第一节 基本概念	61
第二节 随机变量的期望、方差及矩	64
第三节 重要的离散型概率分布	68
第四节 重要的连续型概率分布	81
第六章 统计量抽样分布	91
第一节 参数、统计量与抽样分布	91
第二节 由正态分布导出的分布	95
第三节 常用的抽样分布	101
第七章 参数估计	111
第一节 综述	111

第二节 点估计·····	111
第三节 区间估计·····	121
第八章 参数假设检验·····	129
第一节 基本问题·····	129
第二节 总体均值检验·····	133
第三节 总体比例检验·····	137
第四节 总体方差检验·····	139
第九章 非参数检验·····	145
第一节 非参数方法评述·····	145
第二节 卡方检验·····	146
第三节 符号检验·····	153
第四节 秩和检验·····	158
第五节 游程检验·····	160
第十章 方差分析与正交实验设计·····	163
第一节 方差分析原理·····	163
第二节 单因素方差分析·····	166
第三节 两因素方差分析·····	173
第四节 实验设计基本问题·····	185
第五节 正交实验分析·····	190
第十一章 相关与回归分析·····	196
第一节 相关分析·····	196
第二节 一元线性回归分析·····	202
第三节 多元线性回归分析·····	214
第四节 非线性回归分析·····	220
第十二章 时序分析与动态预测·····	224
第一节 时间序列的含义及其作用·····	224
第二节 时间序列的分解与假定模型·····	227
第三节 长期趋势的测定与预测·····	231
第四节 季节变动测定与预测·····	243
第五节 循环变动的测定·····	252
第十三章 风险决策分析·····	254
第一节 概述·····	254
第二节 不确定型决策·····	257
第三节 风险型决策·····	261
第四节 决策树·····	272
第五节 贝叶斯决策问题·····	275
参考文献·····	278
附录 常用统计表·····	279

第一章

绪 论

作为一门科学的统计学是怎样产生和发展起来的？现代统计学的知识体系构成如何？统计思想与统计方法有哪些基本特征？从事统计应用需要注意什么问题？在这一章里，我们将对诸如此类的问题进行介绍和说明。

第一节 统计学的渊源及其体系

一、统计学的起源与发展

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 年)一书的问世，标志着统计学的肇端。配第在这本书以及他的其他有关著作中，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，试图把想要说明的事情建立在可靠的事实根据上。正如他在《政治算术》“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辨式的议论，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问题，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。格朗特(J. Graunt)是“政治算术”的另一位具有重要影响的人物，他的代表性著作《关于死亡表的自然观察与政治观察》(1662 年)，通过对人口变动数据的分析揭示了人口变化的规律，尤其是他在这本书里运用了不少独特的资料整理和估算方法，给后来统计学的发展带来了有益的启示。

与“政治算术”产生的时期差不多，在德国也出现了一门统计学——“国势学”(staatenkunde)。“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

—德文词转译成英文“statistic”后,逐渐得到了人们的接受并沿用至今。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做了具体深入的阐述。经过一百多年的发展,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年)与康令的《国势学讲义》(1730年)被奉为“国势学”的经典文献。“国势学”的特点是较少使用数字工具,主要诉诸文字记述和逻辑比较认识。

历史上,人们曾就“国势学”与“政治算术”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)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。

统计学的另一个重要起源是概率论,14~15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”问题产生了兴趣。不过,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”(problem of points)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. Leibnix)、贝努里(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,建立了较为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,高斯还导出了正态分布曲线。“政治算术”研究的是简单的、确定的数量关系,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,统计学进入了全新的发展时期。统计学的三个来源及其影响如图1.1所示。



图 1.1 统计学的三个来源及其影响

凯特勒(A. Quetelet)是统计学发展史上承前启后的重要人物,一生写了大量的有关统计学方面的著作,代表性的有《社会物理学》(1869年)、《概率论书简》(1846年)、《犯罪学》(1831年)、《比利时统计研究》(1829年)等。凯特勒在统计学上的突出贡献是他把概率论全面引进到“政治算术”、“国势学”以及其他社会问题的研究中,从而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 世纪后半叶起, 统计学几乎在各个方面都取得了创新性的成果。其中, 尤为引人注目的是兴起于生物遗传学、农业田间实验等领域的所谓统计数学方法的发展。高尔登(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, 认为正态法则适用于许多情况, 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)是高尔登的学生, 他系统地发展了高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”(moment)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”的体系。

进入 20 世纪后, 统计学经历了新一轮飞跃, 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。此外, 统计方法不仅作为科学研究和社会管理活动的工具, 而且直接融合在科研、管理和生产过程之中, 发挥着巨大的社会效益。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工作体会, 戈塞特洞察到大样本统计方法并不适合一切场合, 有时候只能根据少量观察做出结论, 按他的话来说, “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、实验设计等统计分支, 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 年)、《试验设计》(1955 年)、《统计方法与统计推断》(1956 年)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 年)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。随着计算机的普及, 由威夏特(J. Wishart)、赫特林(H. Hotelling)等发展起来的多变量统计重新活跃起来。自 20 世纪五六十年代以来, 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, 都取得了重要的进展。

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、社会实践需要的推动, 统计学也在不断地丰富和完善。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; 从确定型问题的研究, 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认识的大样本理论, 到控制实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壮大起来的, 现在仍然充满着勃勃生机。

二、统计学的学科体系

统计学的产生虽可追溯到 17 世纪,但其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,统计学的进步更是快马加鞭。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,把后者看做随机变量一同纳入模型。仅就这一点,统计学的结果,比单纯只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。统计学对此做了简单的处理,把那些不便控制的因素当成随机干扰,这可能会降低分析的逼真性,但更有利于求解。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,统计学已经发展成了由若干个分支组成的庞大的学科体系,如图 1.2 所示。

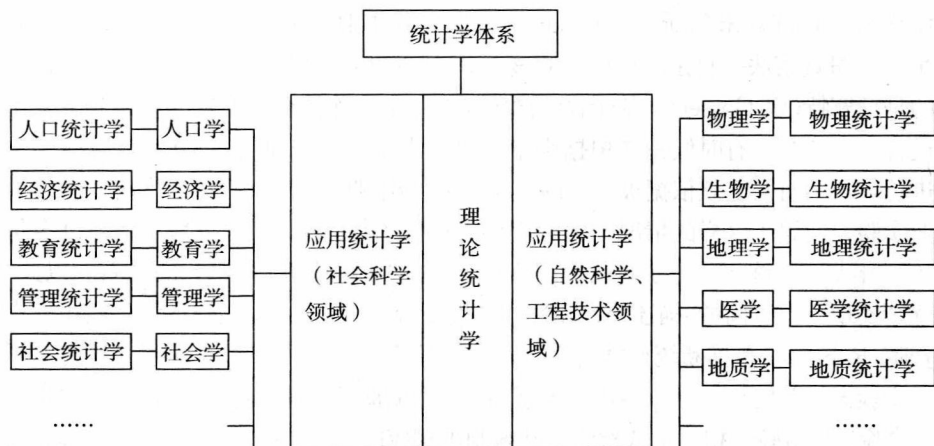


图 1.2 统计学的整体结构

从整个科学体系的大坐标着眼,统计学属于应用学科。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。假设检验是统计学的另一重要内容,它是指在总体分布已知的情况下,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构造、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)实验设计。实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)、田口玄一等都做过一定的贡献。实验设计主要研究如何安排实验方案、如何分析实验结果等问题。

(5)非参数统计。非参数统计与参数统计相对应,主要研究总体分布未知或不依赖于总体分布的各种统计问题。近年来,非参数统计发展较快。

(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”。其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,视每次观察的结果确定是否需要增加新的样本单位。序贯分析可以节省时间、人力、物力,减少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。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、路径分析、多层统计分析等。

(10)统计诊断。这是一个最近二十多年来发展起来的统计学的新领域,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,后经过鲍克斯(G. Box)、塔基(J. Tukey)、瑚珀(P. Huber)、汉姆配尔(F. Hampel)等的努力,其影响越来越大。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,然后在此基础之上建立分析模型。探索性数据分析与塔基的贡献分不开,它们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,可被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,狭义的应用统计学,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,以至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,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的结合,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。

第二节 统计学的概念

一、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,这或多或少影响到了人们对统计学范畴的表述。因此,在解释统计学概念之前,我们先对统计学认识上的一些分歧作下简单评介。

统计学的学科归属。对于统计学到底属于数学学科还是属于社会科学,学者们存在着不同的意见。一种看法是,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学者当中比较流行。统计史上,高尔登、K. 皮尔逊、戈塞特、费喧、内曼等著名人物就是这种看法的代表。而由“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、行政统计,一般坚持认为统计学属于社会科学。除配第、康令、阿肯瓦尔外,后来的苏歇米尔斯、凯特勒、克尼斯、梅约(G. von Mayr)、恩格尔(C. Engel)等也曾作过这样明确的陈述。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,没有完整兼顾统计学广泛的实践;最后,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。但如果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。现在我们国家的《学位授予和人才培养学科目录》(2011年)明确规定,统计学是一级学科,下设数理统计、社会经济统计、生物卫生统计、金融统计与风险管理和精算、应用统计五个二级学科。

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,人们一直争论不休。主张统计学是方法论科学的人认为,统计学主要研究数据资料的处理方法,目的是提供数量认识的方法和工具。在这一派人当中,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,认为统计学主要研究社会经济现象总体数量方法论,强调统计方法的特定用途。“通用方法论”则认为,统计方法既适用于社会科学,又适用于自然科学,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,把统计学当做实质性学科的人比较多,他们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“通用”的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”,这种观点同统计学的“通用方法论”观点一脉相承。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,首先从事物质的分析入手,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重要理由。可是另一部分人则认为,统计学研究客观事物确定的数量,其主要活动内容是对客观事物的变化进行记录,并对记录的结果做出分析。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而言,存在着描述统计学和推断统计学之说。描述统计学与推断统计学之间没有明确的界限,只是大致上所作的划分。根据内曼的看法,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做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。

二、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会相应地加以改变。所以,描述性定义应主动适应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,所以描述性定义具有非唯一性、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。公理化定义是一种语言或符号,可以脱离具体事物的背景,一经给出就必须遵守,如做出改动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。公理化定义具有确定性、精确性、主观性等特点,评判公理化定义只能依据它使用的语言或符号。

有人曾作过考察,关于统计学概念的解释,有不下几十种。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,我们侧重于统计应用,把统计学解释为关于资料搜集、整理、分析和解释的科学。

统计学研究客观现象的数量方面,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,就是计数和获取量值的过程。统计资料的质量,直接关系到能否得出客观的正确的结论。因此,应搞好统计观察和调查活动,设法提高统计调查的水平。

统计整理在统计活动过程中处于承上启下的地位,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建立数据库、资料审核、分组分类、汇总、编制统计表等。

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,目的是形成概括的、精确的数量描述。

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,它往往涉及如何依据样本观察对同类大量现象做出估计、检验、预测等问题。

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,但又存在着密切的联系,只有同时做好各项工作,才能圆满完成统计活动的任务。

第三节 统计学的性质与特点

一、统计学的性质

统计学是一门什么性质的科学?统计方法有哪些特点?了解这两个方面,有助于我们正确掌握统计学的基本原理、提高对统计问题的认识能力、搞清统计学的学科地位以及它与其他相关学科的关系。

现代统计学具有如下几个性质:

第一,数量性。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,只要是出现大量数据的地方,皆用得着统计也需要统计。根据辩证唯物主义的认识论,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。统计研究客观事物的数量,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,总体性。数量有个体数量与总体数量之别,统计学主要研究后者,它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,不确定性。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、随机因素的作用,客观事物的实际数量表现存在一定程度的“不可确知性”。例如,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,这时就不存在统计问题。如果有测量误差,但误差的大小可以完全掌握,这时也无需用到统计。唯有存在误差,而误差又不可预测的时候才需要统计。在现代统计学中,处理不确定性问题是统计学的主要课题和任务。

第四,归纳推断性。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;二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达到目的,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,而是基于观察到的样本情况,对总体的可能情况做出判断。

二、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、运算法则同样适用,必须严格遵守。但是也要注意,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,要正确使用统计方法,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力。其次,统计学主要研究不确定性问题。最后,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就必然存在着联系。这种联系体现在,统计学能为各个专门学科中带有普遍性的数据搜集、整理、分析和解释活动提供方法和理论指导,帮助它们更精确更深入地进行认识。但是统计方法只是定量分析的工具而已,它不会涉及各专门学科中的具体问题。举个例子,经过长期的观察人们发现,吸烟人群肺癌的发病率比较高,吸烟与肺癌之间是否存在关系,这可以通过统计进行论证,但为什么吸烟会导致肺癌却是医学和病理学的课题,统计学无法做出回答。以上所述只是一方面,对此不能作歪曲的理解,片面地认为统计专业的人不需要了解各专门学科的知识。恰恰相反,学习统计的人如果具有专门学科的知识,就能与该学科的专家有更多的共同语言,能够更好地运用统计方法解决问题。

第四节 统计学的应用

今天,人们无论从事生产活动、科研活动还是社会活动,大多离不开数据资料的搜集、整理、分析和解释。在工农业生产和商业活动方面,在社会学和政治学方面,在史学和考古方面,在物理、化学和生物方面,在天文、地理方面,在交通运输和能源供应方