



Statistics

21世纪统计学系列教材

Applied Statistics with SPREADSHEET

应用统计学

——基于SPREADSHEET工具

耿修林 编著

 中国人民大学出版社



Statistics 21 世纪统计学系列教材

Applied Statistics with SPREADSHEET

应用统计学

——基于SPREADSHEET工具

耿修林 编著

中国人民大学出版社
· 北京 ·

图书在版编目 (CIP) 数据

应用统计学：基于 SPREADSHEET 工具/耿修林编著. —北京：中国人民大学出版社，2014. 6
21 世纪统计学系列教材
ISBN 978-7-300-19368-7

I. ①应… II. ①耿… III. ①统计分析-软件包-高等学校-教材 IV. ①C819

中国版本图书馆 CIP 数据核字 (2014) 第 122691 号

21 世纪统计学系列教材
应用统计学
——基于 SPREADSHEET 工具
耿修林 编著
Yingyong Tongjixue

出版发行	中国人民大学出版社		
社 址	北京中关村大街 31 号	邮政编码	100080
电 话	010-62511242 (总编室)		010-62511770 (质管部)
	010-82501766 (邮购部)		010-62514148 (门市部)
	010-62515195 (发行公司)		010-62515275 (盗版举报)
网 址	http://www.crup.com.cn		
	http://www.ttrnet.com (人大教研网)		
经 销	新华书店		
印 刷	北京密兴印刷有限公司		
规 格	185 mm×260 mm 16 开本	版 次	2014 年 7 月第 1 版
印 张	17.75 插页 1	印 次	2014 年 7 月第 1 次印刷
字 数	390 000	定 价	35.00 元

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,借助数据资料对客观现象进行分析,不仅能将认识深化、精确化,还能对理论及其猜想进行论证说明。在林林总总的数量分析应用学科中,统计学因其独特的研究逻辑和研究方法,居于不可或缺的地位。

本书以经济管理类统计学初学者为对象编写,着重介绍常用统计方法在专业课程学习和实际问题中的应用。考虑到“高等数学”课程中一般都会讲授概率原理和基础数理统计方法,所以本书强调统计方法的实现过程。学习统计,掌握一款数据处理软件很有必要。现在电子表格已得到广泛普及,20世纪90年代以来电子表格建模大行其道,对非专业人员和实际工作者来说,了解并掌握用电子表格处理统计数据,不失为有效策略。

本书共有13章:第1章介绍了统计学发展过程及其形成的学科体系,阐述了统计学的性质特征,说明了统计方法应用的注意事项;第2章较为系统地介绍了统计数据的采集渠道,比较了统计数据的测量水平以及统计数据质量的检查方法;第3章介绍了描述统计的图表方法和电子表格的处理过程;第4章、第5章分别针对数量资料和属性资料,介绍了统计特征数字的计算方法和应用问题,通过例子讲解了电子表格的“数据分析”功能;第6章介绍了常用的概率分布及其应用;第7章对正态分布和卡方分布、 t 分布、 F 分布的性质特点做了说明,讨论了常用统计量的抽样分布;第8章介绍了统计参数估计和假设检验电子表格的处理过程;第9章介绍了非参数检验原理;第10章介绍了方差分析的电子表格实现方式;第11章介绍了正交实验设计的原理和实验数据的基本分析方法;第12章介绍了回归分析的原理和应用;第13章重点说明了统计多因关系分析,包括统计控制原理、调节效应和中介效应的应用。

编写过程中,笔者参考了国内外很多优秀的著作和教材,发现了很多对教学富有启发的素材,并将其中符合本书风格的有价值的知识点吸收进来。尽管本人付出了很大的努力,无奈能力有限,书中仍可能存在缺点、不足乃至错误,在此诚恳接受一切批评指正。最后,特别感谢本书参考文献的作者,从这些长期从事统计科学教学科研专家的著作中,笔者了解到了最新的统计知识传播的理念,也提高了统计学专业修养。

教师教学服务说明

中国人民大学出版社工商管理分社以出版经典、高品质的工商管理、财务会计、统计、市场营销、人力资源管理、运营管理、物流管理、旅游管理等领域的各层次教材为宗旨。

为了更好地为一线教师服务，近年来工商管理分社着力建设了一批数字化、立体化的网络教学资源。教师可以通过以下方式获得免费下载教学资源的权限：

在“人大经管图书在线”（www.rdjg.com.cn）注册，下载“教师服务登记表”，或直接填写下面的“教师服务登记表”，加盖院系公章，然后邮寄或传真给我们。我们收到表格后将在一个工作日内为您开通相关资源的下载权限。

如您需要帮助，请随时与我们联系：

中国人民大学出版社工商管理分社

联系电话：010-62515735，62515749，82501704

传 真：010-62515732，62514775

电子邮箱：rdcsjg@crup.com.cn

通讯地址：北京市海淀区中关村大街甲 59 号文化大厦 1501 室（100872）

教师服务登记表

姓 名	☐ 先生 ☐ 女士		职 称		
座机/手机			电子邮箱		
通讯地址			邮 编		
任教学校			所在院系		
所授课程	课程名称	现用教材名称	出版社	对象（本科生/研究生/MBA/其他）	学生人数
需要哪本教材的配套资源					
人大经管图书在线用户名					
院/系领导（签字）： 院/系办公室盖章					

目 录

第 1 章 绪 论	1
1.1 统计学渊源及其体系	1
1.2 什么是统计学	5
1.3 统计学的性质与特点	7
第 2 章 数据来源与数据质量	10
2.1 统计数据的来源	10
2.2 数据测量尺度	16
2.3 统计数据质量问题	19
第 3 章 统计资料的图表显示	23
3.1 统计表的种类与应用	23
3.2 频数分布编制	27
3.3 统计资料的图像描述	36
第 4 章 数量资料的特征数字	45
4.1 集中趋势分析	45
4.2 离散趋势分析	55
4.3 对称性测度	60
4.4 相关性测量	64
4.5 Excel 中的“数据分析”	66
第 5 章 属性资料的特征数字	68
5.1 属性资料的集中趋势	68
5.2 属性资料离散性分析	75
5.3 属性资料的相关性测量	77



第 6 章 随机变量与概率分布	86
6.1 基本概念	86
6.2 随机变量的特征数字	89
6.3 常用的离散型概率分布	93
6.4 常用的连续型概率分布	101
第 7 章 统计量抽样分布	107
7.1 参数、统计量与抽样分布	107
7.2 由正态分布导出的分布	110
7.3 统计量的抽样分布	119
第 8 章 参数统计推断	129
8.1 引 言	129
8.2 参数估计应用	130
8.3 假设检验应用	147
第 9 章 非参数检验	157
9.1 非参数方法评述	157
9.2 卡方检验	158
9.3 符号检验	170
9.4 秩和检验	174
9.5 游程检验	177
第 10 章 方差分析及其应用	180
10.1 方差分析原理	180
10.2 单因素方差分析	183
10.3 双因素方差分析	191
第 11 章 正交实验应用	204
11.1 正交实验原理	204
11.2 正交表的应用	205
11.3 正交实验数据分析	214
第 12 章 回归分析及其应用	224
12.1 线性回归分析	224
12.2 非线性回归分析	239
12.3 线性回归分析的推广	246



第 13 章 多因关系分析	258
13.1 相关关系与因果关系	258
13.2 统计控制原理	260
13.3 调节效应分析	265
13.4 中介效应分析	268
参考文献	274

统计学是怎样产生发展起来的？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？本章将对诸如此类的问题进行讨论。

1.1 统计学渊源及其体系

一、统计学的起源与发展

人类社会有了数的概念便产生了计数活动，统计实践也就开始了，但作为一门科学的统计学却出现得较晚。统计学究竟产生于什么年代，迄今为止人们的看法不尽一致，不过多数人认为统计学兴起于 17 世纪。

17 世纪中叶，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的研究活动。配第（W. Petty）《政治算术》（1676）的问世，标志着统计学的肇端。在这本书以及其他有关著作中，配第采用了不同于前人的研究方法，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》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”学派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），通过对人口变动数据的分析，揭示了人口变化的规律。在这本书里，他运用了不少独特的资料整理和估算方法，为后来统



计学的发展带来了有益的启示。

几乎在“政治算术”产生的时期，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。之所以称为“统计学”，是因为“国势学”的代表人物阿亨瓦尔（G. Achenwall）给“国势学”起了一个统计学（Statistik）的新名称。这一德文词转译成英文“Statistic”后，逐渐得到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作了具体深入的阐述。经过 100 多年的发展，到 18 世纪中期，“国势学”达到顶峰。阿亨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”的特点是诉诸文字记述和逻辑比较认识，较少使用数字语言。

历史上，曾就“国势学”与“政治算术”哪个是统计学真正起源的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”有统计学之名却没有统计学之实，“政治算术”才是“真正的统计学”。克尼斯《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和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，做过重要贡献的人物有：莱布尼兹（G. Leibniz）、伯努利（J. Bernoulli）、棣莫佛（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）、康多塞（M. de Condorcet）、辛普森（T. Simpson）、布丰（C. de Buffon）、泊松（S. Poisson）等。拉普拉斯是古典概率研究的集大成者，给出了概率的“古典”解释，把数学分析方法系统引进概率论，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”研究的是简单确定的数量关系，概率统计研究的是复杂的随机现象，这极大地充实并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。

凯特勒（A. Quetelet）是统计学发展史上一位承前启后的重要人物，一生写了大量有关统计学方面的著作，具有代表性的就有《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中，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”阶段转化。

19 世纪后叶，统计学几乎在各个方面都取得了重要进展，尤其是兴起于生物遗

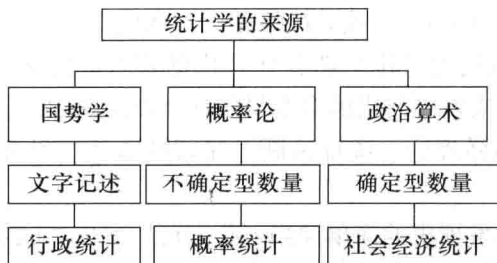


图 1—1 统计学的三个来源及其影响

传学、农业田间实验等领域的所谓统计数学方法。高尔顿 (F. Galton) 是生物统计学的主创者，受凯特勒的影响，他利用正态法则研究遗传优生问题，认为正态法则适用于许多情况，还提出了百分位数、中位数、四分位数差、相关与回归等概念。高尔顿的学生 K. 皮尔逊 (K. Pearson) 系统发展了高尔顿的相关与回归理论，研究了复相关和偏相关，把物理学上“矩” (moment) 的概念引入到统计学中，给出了极大似然估计方法，并导出了重要的 χ^2 分布。以 K. 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，构建了“描述统计学”体系。

进入 20 世纪，统计学经历了质的飞跃。一方面，新的统计思想和统计方法大量涌现；另一方面，带有“归纳逻辑性质”的统计推断逐渐成为主流。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、社会管理和生产过程之中，带来了巨大的社会效益。自苏斯密尔西 (J. Sussmilch) 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。最先发起挑战的是戈塞特 (W. Gosset)。基于在酿酒公司工作多年的体会，戈塞特洞察到大样本并非适合一切场合，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 年戈塞特导出了 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、实验设计等统计理论分支学科，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)、《实验设计》(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) 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) 是继费喧之后的杰出统计学家，共同完善了现代统计学的核心内容——区间估计和假设检验理论。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”，进一步推广了统计学研究和应用的范围。随着计算机的普及，由威沙特 (J. Wishart)、霍特林 (H. Hotelling) 等人发展起来的多变量统计重新活跃起来。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，都取得了重要进展。

通过对统计学发展史的简单回溯，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，统计学始终在不断完善，经历了从意义和概念不甚明确，到作为一门独立学科的转化；从数量研究的“算术”水平，到需要较多数学知识的“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。统计学正是这样一步步发展壮大起来的，现在依然充满勃勃生机。

二、统计学的学科体系

统计学的产生虽可追溯到 17 世纪，但其快速发展却是在凯特勒之后出现的。伴随着应用数学在 20 世纪的崛起，统计学的发展更是快马加鞭。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，并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，统计学的结果就比单纯只考虑基本变量的确定性分析要准确一些。有果必有因，现象结果由多种原因引起，在结果和所有原因因素之间建立关系会使模型复杂化。统计学对此作简单化处理，把那些不便控制的因素当作随机干扰，尽管可能会降低分析的逼真性，但却更有利于求解。因此，统计学能够把复杂的问题转化成相对简单的问题，这正是统计学能够取得快速发展的重要原因之一。目前，统计学已经形成了有若干分支的庞大学科体系（见图 1—2）。

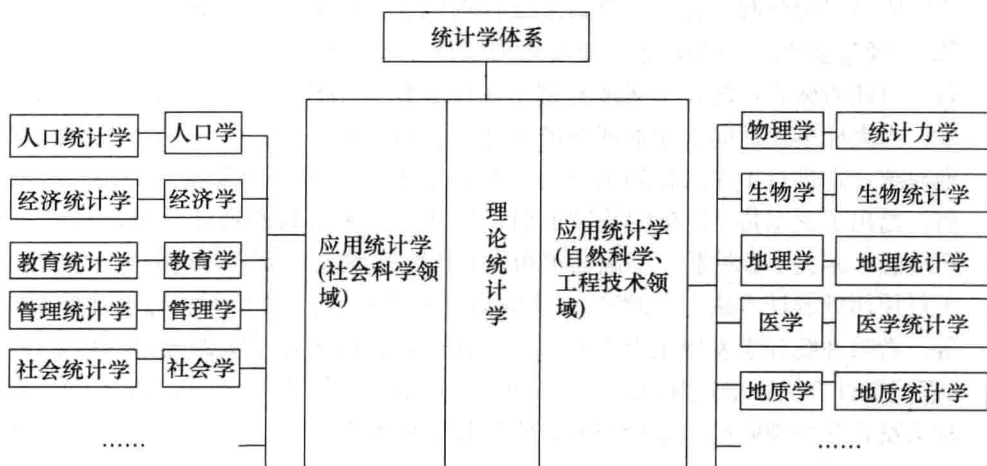


图 1—2 统计学的整体结构

从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重大问题为目标的统计学研究分支。理论统计学由一系列统计分支学科组成,主要有抽样调查、统计实验设计、概率模型与统计分布、参数估计和假设检验、非参数方法、统计决策、多变量统计、时间序列分析、探索性分析、贝叶斯统计、层次统计分析、成分统计、空间统计等。

应用统计学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,主要把统计方法、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碰到的各类计算问题。统计计算方法可看做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,其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,狭义上的应用统计学,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。前者的数学味可能更浓一些,后者对数学的介绍可能较少。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 什么是统计学

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答案。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,这或多或少会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,先来对统计学认识上的分歧作一评介。

统计学的学科归属。统计学属于数学学科还是属于社会科学,人们存在着不同的意见。一种观点认为,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。统计学史上,高尔顿、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。从事由“政治算术”和“国势学”延续下来的人口统计、经济统计、社会统计、行政统计研究的人们,一般坚持认为统计学属于社会科学。除配第、康令、阿亨瓦尔不论,后来的苏斯密尔西、凯特勒、克尼斯、迈



尔 (G. Von Mayr)、恩格尔 (C. Engel) 等人, 也曾作过这样明确的陈述。把统计学纳入数学, 似乎有些不妥。首先, 统计学发展的史实没有得到切实尊重; 其次, 没有完整兼顾统计学广泛的实践; 再次, 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结合紧密的特点几乎丧失殆尽。但一味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。现在我国的《学位授予和人才培养学科目录》(2011) 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、应用统计等 5 个二级学科。

统计学的性质。统计学是方法论性质的学科还是实质性学科, 长期以来一直存在争议。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”认为, 统计方法既适用于社会科学也适用于自然科学, 像数学一样, 是一切数量问题研究的公共工具。统计学发展的初期阶段, 把统计学当作实质性学科的人比较多, 要求统计学研究客观现象的因果关系和变化规律。现在看来, 统计学是一门“通用方法论”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, 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”。这种观点与统计学是“通用方法论”的观点一脉相承。另一种意见认为, 统计学研究的是客观现象具体的数量, 总是联系客观事物的质来研究数量, 首先从对事物的分析入手, 通过对数量的认识上升到对质的认识, 也就是遵从由质到量、再由量到质的认识过程。

在统计学研究上, 多数人把“随机性”或“不确定性”问题, 当作统计学的必然研究对象, 并认为这是统计学作为一门学科存在和发展的重要原因。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, 并对记录的结果进行分析。事实上, 统计学既研究随机现象, 也研究确定性数量, 过去以研究确定性数量为主, 现在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。描述统计学与推断统计学之间没有明确的界限, 只是大致上有所划分。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收集、整理、表述和对比计算分析。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。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以后, 认为统计学的实质是归纳推断的观点颇为流行。

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辩。正常的学术争辩, 有利于科学的繁荣和发展, 只要不囿于偏见就值得提倡。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, 另外一些问题则尚需等待学科的进一步发展才能解决。

二、统计学的概念

概念是客观事物在人的意识中的反映。依据定义概念的方法不同,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发生相应的改变,因此描述性定义应主动适应客观事物。人们对事物的观察、体会和总结不同,给出的描述性定义也有所差别,所以描述性定义具有非唯一性、非确定性的特点。判断描述性定义好坏或正确与否的标准是其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,如果做出改动便会引起在既定公理化定义下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,公理化定义的好坏只能针对其使用的语言或符号进行批判。

有人曾作过考察,关于统计学概念的解释不下几十种。出现这种情况的原因就是采用了描述性定义,加之统计学自身一直在不断发展,并且在认识上还存在不少分歧。这里侧重于统计应用,把统计学定义为:关于资料收集、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,首先要收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、必需的环节。收集资料,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结论。因此,要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位置,一方面是收集资料工作的继续;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恰当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、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,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,由资料收集、整理、分析和解释这四个部分组成。它们各有各的任务和作用,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。

1.3 统计学的性质与特点

一、统计学的特点

统计学是一门什么性质的科学,统计方法有哪些特点,了解这些内容,有助于掌



握统计学的基本原理,提高对统计问题的认识,搞清楚统计学的学科地位。

现代统计学的性质和特点如下:

第一,数量性。统计学是研究数量问题的科学,统计的语言是数字。无论是社会科学还是自然科学,只要出现大量数据的地方,都用得上统计学,也需要统计学。根据辩证唯物主义的认识论,事物是由数量和质量两方面组成的,是二者的有机统一体。统计研究客观事物的数量,如果依据的数字资料准确,方法运用得当,可以达到正确认识和正确反映的目的。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,总体性。数量有个体数量与总体数量之分,统计学主要研究后者,旨在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,只有对大量的同质现象或某一现象多次重复的观察,才有可能找到统计关系和统计规律。

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,客观事物的数量表现存在一定程度的“不可确知性”。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。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,一条是对构成总体的全部事物逐一进行调查,取得全面资料;另一条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以实现,而后者则需要运用概率论知识和统计数学方法。全面调查受诸多因素的约束,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,主要是针对推断统计而言的,它与逻辑学意义上的归纳推断有明显区别。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,只是基于观察到的代表性样本情况,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

1. 统计学与数学的关系

统计学研究客观现象的数量,不可能不用到数学知识,非但如此,往往还需要用到很多数学知识。要学好统计学,尤其是理论统计学,需要有坚实的数学基础。在统计学中,数学的一些定理、运算法则同样适用,且必须严格遵守。但也要注意,统计学与数学相比又有其特殊性。首先,统计学有较强的应用背景,统计方法的正确使用,不仅要有良好的数学基础,而且要懂得相关学科的知识,还需要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。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差别。

2. 统计学与各专门学科的关系

统计学的应用相当广泛,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。因此,统计学与各专门学科必然存在联系。这种联系体现在,统计学能为各门学科中带有普遍性的数据收集和分析提供方法和理论指导。但统计方法只是



定量分析的工具而已，不会涉及各专门学科中的具体问题。举个例子，经过长期观察发现，吸烟人群中肺癌的发病率比较高，吸烟与患肺癌之间是否存在关系，这可以通过统计学进行论证，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，对此不能作错误的理解，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，从而能够更好地解释运用统计方法得出的结论。