

吴柏林 曹立人 著

Xiandai

XIANDAI TONGJIXUE JIQI YINGYONG

现代统计学及其应用



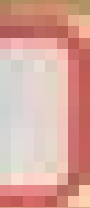
浙江教育出版社

第1版 第1次印刷

WILEY-INTERSCIENCE

JOHNS WILEY & SONS, INC.

现代统计学及其应用



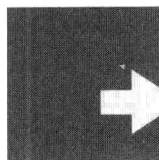
WILEY-INTERSCIENCE

吴柏林 曹立人 著

Xiandai

XIANDAI TONGJIXUE JIQI YINGYONG

现代统计学及其应用



浙江教育出版社

图书在版编目(CIP)数据

现代统计学及其应用/吴柏林、曹立人著. —杭州:浙江教育出版社,2007.12

ISBN 978-7-5338-7232-8

I. 现... II. ①吴... ②曹... III. 统计学 IV. C8

中国版本图书馆 CIP 数据核字(2007)第 161085 号

责任编辑 周俊

责任校对 雷坚

封面设计 曾国兴

责任印务 温劲风

现代统计学及其应用

吴柏林 曹立人 著

出版发行 浙江教育出版社

(杭州市天目山路40号 邮编:310013)

印刷 杭州富春印务有限公司

开本 880×1230 1/32

印张 11.5 插页 1

字数 309 000

印数 0 001-3 000

版次 2007 年 12 月第 1 版

印次 2007 年 12 月第 1 次

书号 ISBN 978-7-5338-7232-8

定价 22.00 元

联系电话: 0571-85170300-80928

e-mail: zjjy@zjcb.com 网址: www.zjeph.com

序

统计是用来分析、处理自然科学与社会科学信息的工具。统计学则是探讨有关数量关系研究方法的科学,内容包括资料的收集与分析、解释相关变量、总体特征的估计与检验、模式的建构与预测以及决策判定等。

近年来由于科技日新月异,研究方法不断更新,统计学已成为各门学科都不可缺少的研究工具。如何以有限的时间或成本获得可靠的信息?如何根据统计资料分析社会经济现象?如何从复杂的实验数据中攫取出规则或定理?如何建构模式进行决策应变或预测未来?这些都是我们极其关心且感兴趣的问题。

统计学历经 20 世纪以来的发展,理论与方法不断改进和创新,这使初学者面对这么多繁杂的理论与方法架构感到茫然,甚至望而生畏。为此作者依据多年的教学、研究经验,从现代社会发展的角度编写了这本书,以适合理、工、医、农、商学与教育等学科大学生和对统计学有兴趣的读者学习统计学的需求。

本书借生动活泼的笔法、简洁的编排,配合相关图表以吸引读者的兴趣,摆脱传统统计学给人的艰涩难懂的印象。在内容方面,除了传统统计理论与方法外,适度加入新的统计观念与方法,如统计软件绘图、人工智能、无参数统计检验、指数、时间序列分析与预测方法等,以适应 21 世纪统计科学发展的大趋势。

在撰写本书过程中,特别注意到以下几点:

(1) 由学习动机引进理论:为了让读者容易了解、掌握统计理论与方法,每章都先由介绍学习动机与目的入手,使读者能简要、迅速地了解该章内容。在结构安排上,配合该章内容附上相关的图表等直观

材料,编排生动活泼,加深读者的印象。

(2) **可读性与亲和力**:考虑到过于繁杂的数理证明或公式,常导致初学者视统计学为天书,因此本书对统计理论与方法的介绍力求简洁明了,文字说明清楚明白。该说的说,该提示的地方一定交代清楚,可有可无的内容就一并删除。文字阐述方面,注意避免模糊的解说或语意不清的文句。

(3) **理论与生活实例配合**:每讲一个观念或定理,必定紧跟着生活实例,使读者能充分了解统计理论在实际生活中的应用。因为借着生活实例的比较,读者就能自然地理解理论或公式的来龙去脉,而不必死背公式。

(4) **注重系统性概念与连贯性**:采用以学习者为中心的编写方式,由浅入深进行论述。前后章节注意逻辑上的连贯性,将初学者一步一步带进统计学的领域。本书中,凡是重要的定理和公式一定交代得清清楚楚,不必要的定理、公式或证明绝对不多列。每章最后都有本章摘要、重点提示,以利于读者复习本章内容,也可以给读者作为准备口试问答的参考。

(5) **合适的习题安排**:习题的演练是让读者熟练掌握内容的最佳方法。每章均列有作者精心设计的习题,完全比对课本内容逐一呈现。一道习题一个观念,不重复也不漏列。读者只要熟悉每章内容,都可以轻松地回答出来。而对于内容还不甚清楚的读者,亦可借由习题的思考与演练,进一步理解课本内容。书末附有习题解答,以便读者练习参考。

(6) **应用计算机学习**:鉴于目前计算机与统计软件的普及,书中各章节的编排中采用的符号、公式都与常用统计软件如 SAS、MINITAB、SPSS 等保持一致。如此既可以减少工作量,又能立即对统计计算结果做比较,增加对统计概念的理解与应用统计软件的熟练度。

本书的撰写历时三年才完成,是海峡两岸科学工作者成功合作的一项成果。在本书的撰写过程中,我们两位来自海峡两岸的学子通过

e-mail、电话等多种方式,就书稿中的各种问题进行过无数次的讨论,充分交换各自的学术意见。统计公式的校验、术语的表征习惯以及文字的推敲、标点符号的规范等,无不经过反复斟酌才敲定。最后,虽然本书已经过多次排版、校对,错误和疏漏的地方恐怕仍在所难免,敬请读者指正。

吴柏林 曹立人

2007年8月

我不知道我在这世界上的地位;但是在神游科学的世界里,
我常把自己比喻成一位充满好奇心的小孩,
在缤纷的海滩边,捡拾一些更美丽的贝壳。
而浩瀚的知识大海则不断地拍打在我的脚丫上,
激起无数的美丽浪花。

目 录

序

第 1 章 导论 / 1

- 1.1 认识统计 / 2
- 1.2 统计学简史 / 4
- 1.3 统计精神就是科学研究的精神 / 7
- 1.4 统计在现代社会所扮演的角色 / 8
- 1.5 现代统计学的发展 / 9
- 摘要 / 15
- 习题 / 15

第 2 章 统计资料的整理与描述 / 17

- 2.1 统计资料的收集 / 18
- 2.2 数据处理与展示——统计图表 / 20
- 2.3 统计数据的特征数 / 31
- 摘要 / 37
- 习题 / 37

第 3 章 概率导论 / 41

- 3.1 概率的概念 / 42
- 3.2 统计独立与条件概率 / 46
- 3.3 概率分布函数及其特征值 / 51
- 3.4 联合概率密度与边际概率密度 / 55
- 3.5 概率不等式 / 60
- 摘要 / 62
- 习题 / 63

第 4 章 常用的概率与统计分布 / 66

- 4.1 离散型概率分布 / 67

4.2 连续型概率分布 / 75

4.3 常用的统计分布 / 80

4.4 中央极限定理 / 84

摘要 / 88

习题 / 88

第 5 章 抽样方法与抽样分布 / 91

5.1 抽样方法 / 92

5.2 样本比率 $\hat{p}$ 的抽样分布 / 95

5.3 样本平均数 $\bar{x}$ 的抽样分布 / 97

5.4 样本变异数 S^2 的抽样分布 / 100

摘要 / 102

习题 / 103

第 6 章 统计估计 / 106

6.1 点估计 / 107

6.2 如何评估点估计的优良性 / 110

6.3 区间估计 / 113

6.4 常态总体变异数 σ^2 的区间估计 / 119

6.5 总体比例 P 的估计 / 122

6.6 决定样本数 / 123

摘要 / 126

习题 / 127

第 7 章 统计检验 / 132

7.1 统计检验的概念 / 133

7.2 总体平均数 μ 的检验 / 138

7.3 总体变异数 σ^2 的假设检验 / 144

7.4 总体比例 p 的假设检验 / 147

摘要 / 150

习题 / 151

第 8 章 变异数分析 / 155

8.1 由实验设计谈起 / 156

8.2 一因子变异数分析 / 159

8.3 集区随机化设计 / 164

8.4 二因子变异数分析 / 168

摘要 / 173

习题 / 174

第 9 章 相关分析与回归模型 / 179

9.1 资料散布图与相关程度 / 180

9.2 单变量回归模型 / 185

9.3 多变量回归模型 / 193

9.4 回归模型的诊断 / 200

摘要 / 205

习题 / 205

第 10 章 非参数统计检验 / 210

10.1 非参数统计的特性 / 211

10.2 符号检验 / 212

10.3 威克生符号等级检验 / 216

10.4 威克生等级和检验 / 219

10.5 科莫果夫—史迈诺夫检验 / 223

10.6 克洛斯科—瓦力士检验 / 227

10.7 随机性检验 / 230

10.8 斯皮尔曼等级相关检验 / 233

摘要 / 236

习题 / 237

第 11 章 指数 / 242

11.1 指数的意义与性质 / 243

11.2 物价指数 / 244

11.3 指数编制要点与应用 / 249

摘要 / 254

习题 / 254

第 12 章 时间数列分析与预测 / 256

12.1 预测概论 / 257

12.2 时间数列的古典分析方法 / 261

12.3 应用平滑法做预测 / 263

12.4 ARIMA 模型 / 266

12.5 模型的认定与检验 / 275

12.6 时间数列模型的预测 / 277

摘要 / 281

习题 / 282

第 13 章 类别数据分析—列联表与卡方检验 / 285

13.1 类别数据分析 / 286

13.2 卡方适合度检验 / 288

13.3 卡方齐一性检验 / 292

13.4 卡方独立性检验 / 296

摘要 / 300

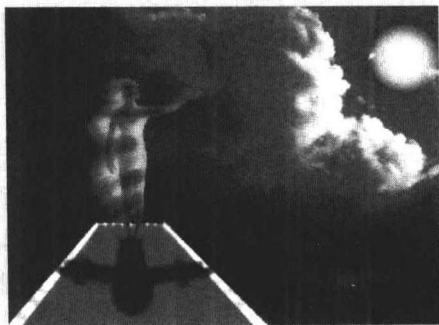
习题 / 300

参考书目 / 303

索引 / 308

附录 A 统计表 / 318

附录 B 习题参考解答 / 346



第1章 导论

统计学是一门探讨如何收集数据与分析数据的研究方法的学科,帮助人们在复杂的自然或社会现象中,借由样本数据所提供的信息,经归纳分析、推论检验、决策、预测等过程,使我们对现实状况更了解,从而客观有效地处理现实世界中所面临的问题。

1.1 认识统计

自古以来,真理的追寻一直是人类从事科学活动的主要目标,也是社会文明进步的源动力。但是在寻求真理的过程中,人类也付出了巨大的代价,包括时间的花费、财力的投入、资源的分配、信仰的转变等,甚至是经历残酷的战争。走过历史,可以发现通往真理的路上充满了混沌与挫折。从假设、观察、实验到结论的每一过程,真理并不是那么明显地摆在那里的。即使是同样一个实验、同一种信息或者是同一组样本,不同的科学家也可能会推演出不同的结论。

随着人类社会的演变,科学技术不断进步,信息来源也越来越多。每天我们不断地从报纸、杂志、广播、电视、网络等媒体接触到各种各样的信息,这些资料中通常还附有大量的数据和图表。如何从大量的信息当中去芜存菁、理清真相,是现代人面临的一大考验。现代统计学的主要目标就是以科学方法来处理、分析、研判并应用这些信息。因此在众多探寻真相的实践过程中,统计学就成为一门科学研究中必不可少的方法科学。

传统概念中“统计”一词通常解释为数据的收集与数据的解说。其实现代的统计学已被积极推广到如何收集数据、如何分析数据、如何进行模型分析、如何做统计决策以及动态数据如时间序列趋势分析与预测等多个领域。现在不仅政府部门需要用高效率的统计方法来调查人口、财经、交通、教育、国防等状况,民间各大行业、工商业界的经营者,也常借统计数据来评估公司是否处在预定的目标状态、营运状况是否有效率等问题,并借此来比较其他同类型企业的营运状况,用以评估或改进企业自身的发展策略,寻求最合适的经营模式。

例如：

年金政策推行后，是否提高了人民对政府的满意度？

口蹄疫流行期间，有多少头猪被感染？感染后有多少头猪被灭杀？

央行降低利率1%，对商业景气有多大影响？对证券市场有多大影响？

近几年犯罪率是否显著降低？人民对社会治安是否较有信心？

幼儿奶粉添加特种维他命对婴儿成长是否有帮助？补脑丸是否能增强记忆力？

针灸对病患的改善效果多大？鸡尾酒疗法对艾滋病患者的病情能控制多少？

北京市区到底要再增加多少停车位才够用？

中小学实施小班化教学以后，未来五年全国将需要增加多少教师？

现代统计学的目的主要是针对诸如此类的问题，拟定一套衡量测度的方法，以期得到客观的结论。其过程是：①设定合适的理论或模型；②收集样本数据，实验设计、抽样或仿真；③数据分析与研判；④参数估计与检验；⑤决策或预测。

在我们的日常生活中，常常碰到类似统计却非统计的叙述论调，尤其在广告方面的误导更是严重，值得我们注意。例如，经常有人说，一般人一生中仅用到其脑力的10%；老师感叹现在的学生自理能力越来越差；广告传单上宣称学琴的孩子不会变坏；医院墙上贴着海报，强调吸烟得癌的比率较未吸烟者高出20倍，等等，其目的主要是为了让人们赞同他们的看法或接受他们的论点。但是这些看法或论点，有的是尚未经过科学的统计检验程序而下的结论，有的甚至是假借统计之名，对事实做出某种程度的夸大或扭曲的报道。这样不仅异化了统计学的功能，也违反了实事求是的科

学基本精神。如：①一般人的脑力如何定义？②什么是社会成本的定义？因素有哪些？③病患到哪一种程度才算治愈？能维持多久？④社会治安好坏的指标是什么？⑤不学琴的孩子比学琴的孩子变坏的比例高吗？是否显著？孩子变到什么程度才叫做变坏？等等。我们在叙述真相、做出归纳性结论之前，不仅要先说明定义，而且要清楚命题的假设与内容，最好能有一些可靠的数据来支持，否则很容易变为玩弄统计手段与文字的魔术，得出一些危言耸听的谬论。统计分析的另一功用是可以帮助我们思考命题的意义及可信度，理清混淆的叙述，以避免无谓的争论。

1.2 统计学简史

统计学是有悠久历史的科学研究方法，内容包括抽样调查、实验设计、变异数分析、相关与回归分析、非参数统计、指数与时间序列、类别资料分析、统计计算等。

有关统计活动的记载，最早可追溯到公元前 3000 年巴比伦人 (Babylonians) 与罗马人使用小泥板记录农业生产与商品买卖资料。古埃及人在建金字塔之前，曾运用统计方法来记录人口与物资数据。类似的数据在中国公元前 2000 年的夏商朝代亦有记载。约公元前 594 年，希腊人曾进行人口普查以确定征税标准。罗马帝国最早广泛地收集人口、土地、控制区财富数据。

到了中世纪欧洲，统计活动更频繁。如英王威廉一世于 1086 年进行人口普查，结果记录于 *Domesday Book*。16 世纪初期英国开始有死亡、出生登记。1662 年，第一本著名的人口统计书 *Observations on the London Bills of Mortality* 问世。

而从近代统计学发展的历史来看，大致可分为四个阶段。

古典统计学时期(从古代到 19 世纪初)

(a) 描述统计学:偏重数据与图表的显示、事物性质的解释。统计学的名词正式被提出,见艾岑瓦(Achenwall,1690)的《欧洲各国国势学概论》。

(b) 政治统计学:偏重事物观察、分类、对比的量的计算。其代表作有格朗特(Graunt,1662)的《关于死亡表的自然和政治的观察》、派蒂(Petty,1690)的《政治算术》。

(c) 概率论:有系统地将数学分析应用于概率统计理论。代表作有拉普拉斯(Laplace,1812)的《概率论》,以及谢比雪夫(Chebychev)与马尔可夫(Markov)1890年提出的概率不等式。

(d) 重要事件:比利时天文与统计学家奎特莱特(Quetelet)于1853年创办第1届国际统计会议。高斯(Gauss)与拉普拉斯分别研究误差分布而发现常态分布(1930)。

近代统计学时期(从 19 世纪初到 20 世纪初)

(a) 实验设计与统计检验:费雪(Fisher)于1935年创立实验设计,研究自然与工程现象所用的随机化模型,并提出了变异数分析、小样本技术等统计检验方法。

(b) 社会统计学:强调社会科学中的实证性科学,强调在统计研究中必须以事物的性质为前提。代表人物有尼斯(Knies,1821~1898)和安吉尔(Engel,1821~1889)。

(c) 生物统计学:强调将统计应用于生物科学的研究,使生物统计成为应用统计学的重要一环。代表人物有高尔顿(Galton,1880),他应用统计方法研究遗传因果关系,发现了祖先遗传律与子女趋中律。另一代表人物是皮尔逊(Pearson,1901),他运用统计推断方法,以数学公式来描述未知总体。

(d) 重要事件:学生(Student,1908)(本名为 W. S. Gosset)出

版《精密样本论》，这是第一本研究样本分布的著作。吉伯士(Gibbs, 1870)将概率统计方法应用于化学研究创立了相位律(Phase rule)。波尔兹曼(Boltzmann, 1877)应用统计学从事分子能量研究，提出分子动能的常态分布理论。

现代统计学时期(20 世纪初至计算机革命)

(a) 数理统计学: 强调自然科学中的方法论科学, 偏重统计公式的数学基础与理论的推演归纳。代表人物有尼曼(Neyman)与皮尔逊(Pearson, K. Pearson 之子), 他们在 1933 年合作研究, 发展了统计假设检验的系统理论。劳(Rao)、布莱克威尔(Blackwell)、黎曼(Lemann)、斯凯夫(Scheffe)于 20 世纪 60 年代提出了估算理论, 为后来的统计估计论奠定了基础。

(b) 非参数统计学: 20 世纪 30 年代末, 统计学家开始用不同的方法来进行统计推论, 特点在于尽量放宽对分布或模型假设的限制。其代表人物是魏克生(Wilcoxon, 1945), 他提出配对样本的符号等级检验方法。哈杰克(Hajek)与西达克(Sidak)在 1967 年合作提出等级检验理论。

(c) 重要事件: 科莫果夫(Kolmogorov, 1933)出版《概率论基础》, 以集合论与测度论开拓现代概率论领域。考克兰(Cochran, 1963)出版《抽样技术》, 创立各种类型的抽样设计。

后现代统计学时期(计算机革命时期迄今)

这一时期, 社会统计学逐渐由实质性科学向方法论科学转变, 一个重要的特点是吸收自然科学研究中的方法, 应用于社会科学的研究。动态统计学如随机过程、时间序列分析, 生物统计学如遗传基因统计分析、神经计算技术等都得以如火如荼地发展。数理统计学也从方法论科学走向实质性科学, 并随着计算机技术的快速发展, 衍生出更多统计学的新分支与相关学科, 以适合社会科学