

应用统计学系列教材 Texts in Applied Statistics

# 统计学(第二版)

Statistics (Second Edition)

贾俊平 编著

Jia Junping

3

清华大学出版社

C8  
72D=2

应用统计学系列教材 Texts in Applied Statistics

# 统计学(第二版)

Statistics (Second Edition)

贾俊平 编著

Jia Junping

清华大学出版社  
北京

## 内 容 简 介

统计学是收集、分析、表述和解释数据的科学。作为数据分析的一种有效工具,统计方法已广泛应用于社会科学和自然科学的各个领域,是各学科领域研究者和实际工作者的必备知识。《统计学》(第二版)一书结合了作者多年的教学实践经验和国外优秀统计学教材的成果,在内容上包括描述统计方法、推断统计方法以及工商管理中常用的一些统计方法;在写法上与计算机紧密结合,大部分统计方法都给出了 Excel 的计算过程和结果,并在书后配有教学和学习辅助光盘,方便教师授课和学生自学。

本书可作为高等院校经济管理类专业本科生统计学课程的教材,也可作为 MBA 的教材或参考书,对广大实际工作者也极具参考价值。

版权所有,翻印必究。举报电话:010-62782989 13501256678 13801310933

### 图书在版编目(CIP)数据

统计学/贾俊平编著. —2版. —北京:清华大学出版社,2006.7

(应用统计学系列教材)

ISBN 7-302-12928-2

I. 统… II. 贾… III. 统计学—高等学校—教材 IV. C8

中国版本图书馆 CIP 数据核字(2006)第 043830 号

出 版 者: 清华大学出版社

<http://www.tup.com.cn>

社 总 机: 010-62770175

地 址: 北京清华大学学研大厦

邮 编: 100084

客 户 服 务: 010-62776969

责任编辑: 王海燕

印 刷 者: 北京市清华园胶印厂

装 订 者: 北京鑫海金澳胶印有限公司

发 行 者: 新华书店总店北京发行所

开 本: 170×230 印 张: 36 字 数: 646 千 字

版 次: 2006 年 7 月第 2 版 2006 年 7 月第 1 次印刷

书 号: ISBN 7-302-12928-2/F·1523

印 数: 1~5000

定 价: 47.00 元(含光盘)

# 序

# PREFACE

随着社会经济的飞速发展,统计学课程设置的不断调整,统计学教材已经有了很大的变化。为了适应这些变化,我们从2000年开始编写面向21世纪统计学系列教材,经过近4年的实践,该系列教材取得了较好的效果,基本实现了预定的目标。然而目前学科的发展和社会的进步速度相当快,其中的一些教材已经需要进一步修订,也有部分内容成熟、适合教学需要的教材没有列入编写计划。

为满足应用统计科学和我国高等教育迅速发展的需求,清华大学出版社和施普林格出版社(Springer-Verlag)合作,倡议出版这一套“应用统计学系列教材”,作为对现有统计学教材的全面补充和修订。这套教材具有以下特点:

1. 此套丛书属于开放式的,一旦有好的选题,即可列入出版计划。

2. 在教材选择上,拓宽了范围。有些教材主要面向经济类统计学专业,包括金融统计、风险管理与精算方面的教材。部分教材面向人文社科专业,而另外一些教材则面向自然科学领域,包括生物统计、医学统计、公共卫生统计等。

3. 本套教材的编写者都是活跃在教学、科研第一线的教师,他们能够积极地、广泛地吸收国内外最新的优秀成果。能够在教学中反复对教材进行补充修订和完善。

4. 强调与计算机应用的结合,在教材编写中,注重计算机软件的应用,特别是可编程软件的应用。对于那些仅限于应用方法的教材,充分考虑读者的需求,尽量介绍简单易学的“傻瓜”

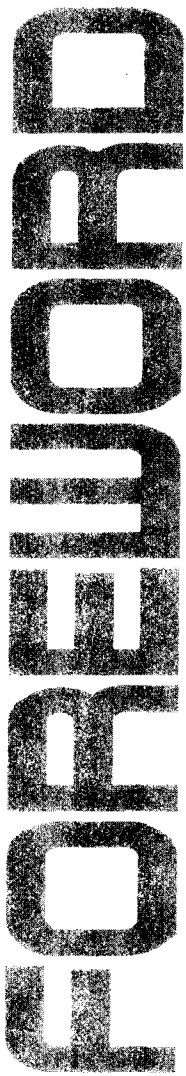
软件。

5. 本套教材包括部分优秀国外教材译著,对于目前急需,而国内尚属空白的教材,选择部分国外具有广泛影响的教材,进行翻译出版。

我们希望这套系列教材的出版能够对我国应用统计科学的教育和我国统计事业的健康发展起到积极作用。感谢参与教材编写的中国人民大学统计学院和兄弟院校的教师以及进行审阅的同行专家。让我们大家共同努力,创造我国应用统计学科新的辉煌。

易丹辉

2004年1月



# 第一版前言

多年来,我一直从事统计教学工作,也编写过不同版本的统计学教材,但在教学中始终感到教材不满意。目前,公开出版的统计学教材版本繁多,但内容和写法大同小异,基本雷同。同国外的一些优秀统计学教材相比,我们的教材水平相差甚远。主要问题有四个方面:一是内容肤浅,许多方法介绍得不够深入;二是废话太多,抓不到统计方法的实质;三是写法陈旧,与计算机结合得不够;四是缺乏案例,习题量太少。多年教学实践中所感受到的这些问题,希望在本书中部分得到解决。

随着计算机的应用,传统的教学方式发生了根本性的转变,很难想像,脱离计算机的统计学教学效果会怎么样。特别是对非统计学专业的学生来说,学习统计学的目的主要是应用,不会用计算机,就很难将统计方法应用到实际问题之中。对他们而言,做过多的统计公式的推导显然意义不大,学会统计计算不等于能使用统计方法。将计算机应用到统计教学之中,不仅可使学生从复杂的统计计算中解脱出来,教师也可以在同样多的时间里讲授更多的内容,特别是有更多的时间讲授统计方法的思想、特点和应用场合。只有理解了统计方法的思想,才能真正地使用统计方法,而统计计算显然不是应用的障碍,因为这些工作可以交给计算机来完成。

Gudmund R. Iversen 与 Mary Gergen 合著的《统计学——基本概念和方法》一书,在其前言对统计学教学的讨论中作了很精辟的论述:“正如大部分统计学教师所敏锐地意识到的那样,统计的教学方式已经发生了戏剧性的变化。计算机与教学环境的结合,尤其是界面越来越方便友好的统计软件的使用,已经使旧

的学习方法——特别是记忆并运用统计公式,已不再适用于大部分学生……我们的经验是,统计公式就像一门外语,如果一个人理解了这种语言,那么公式会大大增加对统计学的理解;否则,这些公式就像密码一样难以破译。我们已经看到,很多同学在学习统计时,公式反倒成了一种障碍。我们坚信,不用公式,也照样有可能获得对统计思想的深刻理解。”这段话,对于非统计专业的学生来说尤为合适,自然也适合本书的初衷。

计算机的应用不仅需要彻底变革传统的统计学教学方式,对统计学教材的写作也提出了更高的要求。本书在写作上,力求淡化统计方法的推导和计算,而重点突出统计方法的思想 and 计算机的应用。对大多数计算问题,都给出了计算机输出的结果。考虑到非统计学专业的教学内容和统计软件的普及程度,本书中仍然使用 Excel 作为计算工具。在计算时,我们直接给出了由 Excel 输出的结果。由于大多数人都熟悉 Excel,我们只给出 Excel 的简单操作步骤。本书不主张,确切地说是反对学生用手工进行统计计算,要求学生必须能使用计算机这一工具,退一步说,至少能看懂 Excel 输出的统计结果。我们也不主张学生在学习时死记公式,而应把重点放在对统计思想的理解上,只有真正理解了,才能在实际中应用。

本书的出版得到各方面的帮助。感谢中国人民大学统计学院领导的支持。感谢曾讲授过该门课程的王燕、王星、张云、姚嘉秋、赵以立、陈龙、刘文卿、李静萍、黄向阳等各位老师,他们为作者提供宝贵的意见。感谢谭英平同学,她翻译整理了部分章后给出的人物传记。有幸作为应用统计学系列教材之一出版,还要感谢清华大学出版社的王海燕同志,她对书稿的认真编辑使本书增色不少。

尽管几年来对统计学教材的编写一直进行着各种努力和尝试,很想奉献给读者一本满意的教材,但仍然有可能达不到各方面的要求。书中的不当或疏漏之处在所难免,恳请同行和读者提出宝贵意见,以便共同努力把统计学教材建设好。

贾俊平

2004年1月于中国人民大学

# 第二版前言

我多年从事统计教学的体会是：学习一遍，很难学会统计；记住公式，不等于理解统计；学会计算，不等于会用统计。统计的真谛在于它所体现的思想，在于它所提供的思维方式。

作为一门受众面很广的课程，怎样在有限的时间内把统计讲清楚是一件很难的事情。一本好的教材对于最初接触统计的人来说很重要，多年来我一直在朝着不断完善统计学教材的目标努力，但一直力不从心。尽管《统计学》受到广大读者的厚爱，在一年多的时间里先后 6 次印刷，但不足之处仍然不胜枚举。近两年来，经过在讲授统计学课程过程中的不断学习，又有一些新的想法，很想把它写进教材。这就是《统计学》（第二版）的初衷。

与第一版相比，第二版仍然维持原有的框架和风格，与第一版的不同之处体现在以下几个方面：

第一，增加了一些新的内容，特别是与 Excel 应用有关的一些内容，比如 Excel 统计函数的应用、数据透视表的应用、概率分布表的生成，同时结合了其他一些统计软件，如 STATISTICA 软件的输出结果，包括正态概率图、箱线图、三维相关图等。

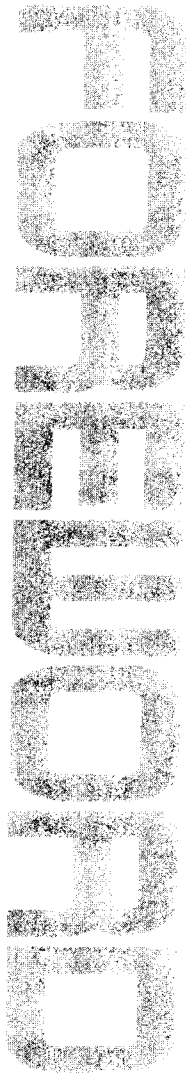
第二，在每章的开头增加了“统计应用”的一个实例。

第三，增加了每章的习题量，习题的类型也趋于多样化。

第二版的出版仍然要感谢清华大学出版社的支持。感谢编辑王海燕同志，她的辛勤劳动使本书的错误降到最低。感谢书后所列参考文献的作者，本书的部分素材选自所列参考文献。

贾俊平

2006 年 1 月于中国人民大学





---

# 目 录

<b>第 1 章 导论</b> .....	1
统计应用:质量管理中的 $6\sigma$ .....	1
1.1 统计及其应用领域 .....	2
1.1.1 什么是统计学 .....	2
1.1.2 统计的应用领域 .....	3
1.1.3 历史上著名的统计学家 .....	6
1.2 统计数据的类型 .....	7
1.2.1 分类数据、顺序数据、数值型数据 .....	7
1.2.2 观测数据和实验数据 .....	8
1.2.3 截面数据和时间序列数据 .....	8
1.3 统计中的几个基本概念 .....	9
1.3.1 总体和样本 .....	9
1.3.2 参数和统计量 .....	11
1.3.3 变量 .....	12
思考与练习 .....	13
人物传记——Adolphe Quetelet .....	15
<b>第 2 章 数据收集</b> .....	17
统计应用:北京市 2005 年 1% 人口抽样调查 .....	17
2.1 数据来源 .....	19
2.1.1 数据的间接来源 .....	19
2.1.2 数据的直接来源 .....	19
2.2 调查设计 .....	25
2.2.1 调查方案的结构 .....	25
2.2.2 调查问卷设计 .....	26

2.3 数据质量 .....	36
2.3.1 数据的误差 .....	36
2.3.2 数据的质量要求 .....	36
思考与练习 .....	37
人物传记——William G. Cochran .....	37
<b>第3章 数据的图表展示 .....</b>	<b>39</b>
统计应用:把数据画图之后要用用脑袋! .....	39
3.1 数据的预处理 .....	40
3.1.1 数据审核 .....	40
3.1.2 数据筛选 .....	40
3.1.3 数据排序 .....	43
3.1.4 数据透视表 .....	44
3.2 品质数据的整理与展示 .....	48
3.2.1 分类数据的整理与图示 .....	48
3.2.2 顺序数据的整理与图示 .....	55
3.3 数值型数据的整理与展示 .....	57
3.3.1 数据分组 .....	57
3.3.2 数值型数据的图示 .....	62
3.4 合理使用图表 .....	75
3.4.1 鉴别图形优劣的准则 .....	76
3.4.2 统计表的设计 .....	76
思考与练习 .....	79
人物传记——John W. Tukey .....	86
<b>第4章 数据的概括性度量 .....</b>	<b>87</b>
统计应用:一种测量的平均数比单个的测量更可靠 .....	87
4.1 集中趋势的度量 .....	88
4.1.1 分类数据:众数 .....	88
4.1.2 顺序数据:中位数和分位数 .....	89
4.1.3 数值型数据:平均数 .....	93
4.1.4 众数、中位数和平均数的比较 .....	98
4.2 离散程度的度量 .....	99

4.2.1	分类数据:异众比率 .....	100
4.2.2	顺序数据:四分位差 .....	100
4.2.3	数值型数据:方差和标准差 .....	101
4.2.4	相对离散程度:离散系数 .....	107
4.3	偏态与峰态的度量 .....	109
4.3.1	偏态及其测度 .....	109
4.3.2	峰态及其测度 .....	110
	思考与练习 .....	113
	人物传记——Pafnuty Lvovich Chebyshev .....	118
<b>第5章</b>	<b>概率与概率分布 .....</b>	<b>120</b>
	统计应用:买彩不是“押宝” .....	120
5.1	事件及其概率 .....	122
5.1.1	试验、事件和样本空间 .....	122
5.1.2	事件的概率 .....	124
5.1.3	概率的性质和运算法则 .....	125
5.1.4	条件概率与事件的独立性 .....	130
5.1.5	全概率公式与逆概率公式 .....	134
5.2	离散型概率分布 .....	137
5.2.1	随机变量 .....	137
5.2.2	离散型随机变量的概率分布 .....	138
5.2.3	离散型随机变量的数学期望和方差 .....	140
5.2.4	几种常用的离散型概率分布 .....	141
5.3	连续型概率分布 .....	151
5.3.1	概率密度函数 .....	151
5.3.2	正态分布 .....	152
5.3.3	其他连续型概率分布 .....	166
	思考与练习 .....	170
	人物传记——James Bernoulli Carl Friedrich Gauss .....	173
<b>第6章</b>	<b>抽样与抽样分布 .....</b>	<b>176</b>
	统计应用:“抓阄”征兵计划 .....	176
6.1	概率抽样方法 .....	177

6.1.1	简单随机抽样	177
6.1.2	分层抽样	179
6.1.3	系统抽样	180
6.1.4	整群抽样	180
6.2	3种不同性质的分布	181
6.2.1	总体分布	181
6.2.2	样本分布	182
6.2.3	抽样分布	182
6.3	一个总体参数推断时样本统计量的抽样分布	183
6.3.1	样本均值的抽样分布	183
6.3.2	样本比例的抽样分布	189
6.3.3	样本方差的抽样分布	190
6.4	两个总体参数推断时样本统计量的抽样分布	194
6.4.1	两个样本均值之差的抽样分布	194
6.4.2	两个样本比例之差的抽样分布	195
6.4.3	两个样本方差比的抽样分布	195
	思考与练习	198
	人物传记——William Sealy Gosset	201
<b>第7章</b>	<b>参数估计</b>	<b>203</b>
	统计应用:一次失败的民意调查	203
7.1	参数估计的一般问题	204
7.1.1	估计量与估计值	204
7.1.2	点估计与区间估计	205
7.1.3	评价估计量的标准	209
7.2	一个总体参数的区间估计	211
7.2.1	总体均值的区间估计	211
7.2.2	总体比例的区间估计	217
7.2.3	总体方差的区间估计	219
7.2.4	正态总体未来观测值的预测区间估计	220
7.3	两个总体参数的区间估计	222
7.3.1	两个总体均值之差的区间估计	222
7.3.2	两个总体比例之差的区间估计	228

7.3.3	两个总方差比的区间估计 .....	229
7.4	样本容量的确定 .....	234
7.4.1	估计总体均值时样本容量的确定 .....	234
7.4.2	估计总体比例时样本容量的确定 .....	235
7.4.3	估计两个总体均值之差时样本容量的确定 .....	236
7.4.4	估计两个总体比例之差时样本容量的确定 .....	237
	思考与练习 .....	237
	人物传记——Jerzy Neyman .....	244
<b>第8章</b>	<b>假设检验 .....</b>	<b>246</b>
	统计应用:药物筛选中的假设检验 .....	246
8.1	假设检验的基本问题 .....	248
8.1.1	假设的陈述 .....	248
8.1.2	两类错误与显著性水平 .....	252
8.1.3	检验统计量与拒绝域 .....	255
8.1.4	利用 $P$ 值进行决策 .....	257
8.1.5	统计显著性与实际显著性 .....	261
8.2	一个总体参数的检验 .....	264
8.2.1	总体均值的检验 .....	264
8.2.2	总体比例的检验 .....	272
8.2.3	总体方差的检验 .....	275
8.3	两个总体参数的检验 .....	278
8.3.1	两个总体均值之差的检验 .....	278
8.3.2	两个总体比例之差的检验 .....	290
8.3.3	两个总体方差比的检验 .....	293
	思考与练习 .....	297
	人物传记——Egon Sharpe Pearson .....	304
<b>第9章</b>	<b>方差分析与试验设计 .....</b>	<b>306</b>
	统计应用:SARS病毒灭活疫苗临床试验 .....	306
9.1	方差分析引论 .....	308
9.1.1	方差分析及其有关术语 .....	308
9.1.2	方差分析的基本思想和原理 .....	310

9.1.3	方差分析中的基本假定 .....	313
9.1.4	问题的一般提法 .....	315
9.2	单因素方差分析 .....	316
9.2.1	数据结构 .....	316
9.2.2	分析步骤 .....	317
9.2.3	关系强度的测量 .....	324
9.2.4	方差分析中的多重比较 .....	325
9.3	双因素方差分析 .....	327
9.3.1	双因素方差分析及其类型 .....	327
9.3.2	无交互作用的双因素方差分析 .....	328
9.3.3	有交互作用的双因素方差分析 .....	334
9.4	试验设计初步 .....	338
9.4.1	完全随机化设计 .....	338
9.4.2	随机化区组设计 .....	339
9.4.3	因子设计 .....	341
	思考与练习 .....	342
	人物传记——Ronald Aylmer Fisher .....	348
<b>第 10 章</b>	<b>一元线性回归</b> .....	<b>351</b>
	统计应用:回归分析在投资风险中的应用 .....	351
10.1	变量间关系的度量 .....	353
10.1.1	变量间的关系 .....	353
10.1.2	相关关系的描述与测度 .....	354
10.1.3	相关关系的显著性检验 .....	359
10.2	一元线性回归 .....	361
10.2.1	一元线性回归模型 .....	362
10.2.2	参数的最小二乘估计 .....	365
10.2.3	回归直线的拟合优度 .....	370
10.2.4	显著性检验 .....	374
10.2.5	回归分析结果的评价 .....	378
10.3	利用回归方程进行估计和预测 .....	379
10.3.1	点估计 .....	379
10.3.2	区间估计 .....	380

10.4 残差分析 .....	384
10.4.1 用残差证实模型的假定 .....	384
10.4.2 用残差检测异常值和有影响的观测值 .....	388
思考与练习 .....	390
人物传记——Francis Galton .....	397
<b>第 11 章 多元线性回归</b> .....	<b>400</b>
统计应用:预测大学足球比赛的获胜得分差额 .....	400
11.1 多元线性回归模型 .....	401
11.1.1 多元回归模型与回归方程 .....	402
11.1.2 估计的多元回归方程 .....	403
11.1.3 参数的最小二乘估计 .....	403
11.2 回归方程的拟合优度 .....	406
11.2.1 多重判定系数 .....	406
11.2.2 估计标准误差 .....	407
11.3 显著性检验 .....	408
11.3.1 线性关系检验 .....	408
11.3.2 回归系数检验和推断 .....	409
11.4 多重共线性 .....	411
11.4.1 多重共线性及其所产生的问题 .....	411
11.4.2 多重共线性的判别 .....	412
11.4.3 多重共线性问题的处理 .....	413
11.5 利用回归方程进行估计和预测 .....	415
11.6 变量选择与逐步回归 .....	416
11.6.1 变量选择过程 .....	416
11.6.2 向前选择 .....	417
11.6.3 向后剔除 .....	418
11.6.4 逐步回归 .....	418
11.7 虚拟自变量的回归 .....	420
11.7.1 含有一个虚拟自变量的回归 .....	420
11.7.2 用虚拟自变量回归解决方差分析问题 .....	426
11.8 非线性回归 .....	429
11.8.1 双曲线 .....	430

11.8.2 幂函数曲线 .....	430
11.8.3 对数曲线 .....	430
思考与练习 .....	433
人物传记——George Waddell Snedecor .....	440
<b>第 12 章 时间序列分析和预测 .....</b>	<b>441</b>
统计应用:平均增长率的计算争议 .....	441
12.1 时间序列及其分解 .....	443
12.2 时间序列的描述性分析 .....	446
12.2.1 图形描述 .....	446
12.2.2 增长率分析 .....	447
12.3 时间序列预测的程序 .....	451
12.3.1 确定时间序列的成分 .....	451
12.3.2 选择预测方法 .....	454
12.3.3 预测方法的评估 .....	455
12.4 平稳序列的预测 .....	457
12.4.1 简单平均法 .....	457
12.4.2 移动平均法 .....	458
12.4.3 指数平滑法 .....	460
12.5 趋势型序列的预测 .....	463
12.5.1 线性趋势预测 .....	463
12.5.2 非线性趋势预测 .....	465
12.6 季节型序列的预测 .....	475
12.7 复合型序列的分解预测 .....	479
12.7.1 确定并分离季节成分 .....	479
12.7.2 建立预测模型并进行预测 .....	483
12.7.3 计算最后的预测值 .....	484
12.8 周期性分析 .....	485
思考与练习 .....	487
人物传记——Abraham Wald .....	494
<b>第 13 章 指数 .....</b>	<b>496</b>
统计应用:报道价格指数 .....	496



13.1	引言 .....	497
13.2	加权指数 .....	498
13.2.1	加权综合指数 .....	498
13.2.2	加权平均指数 .....	500
13.2.3	价值指数与指数体系 .....	502
13.3	几种常用的价格指数 .....	503
13.3.1	零售价格指数 .....	503
13.3.2	消费者价格指数 .....	504
13.3.3	生产价格指数 .....	506
13.3.4	股票价格指数 .....	507
13.4	多指标综合评价指数 .....	508
13.4.1	多指标综合评价指数的构建 .....	508
13.4.2	几种常用的综合评价指数 .....	510
	思考与练习 .....	513
	人物传记——Karl Pearson .....	515
<b>附录 1 各章练习题答案 .....</b>		<b>518</b>
<b>附录 2 常用统计表 .....</b>		<b>543</b>
表 1	标准正态曲线下的面积 .....	543
表 2	$t$ 统计量的临界值 .....	545
表 3	$\chi^2$ 统计量的临界值 .....	546
表 4	$F$ 统计量的临界值 .....	548
<b>参考文献 .....</b>		<b>556</b>