

应用统计学系列教材 Texts in Applied Statistics

# 统计学 Statistics

贾俊平 编著

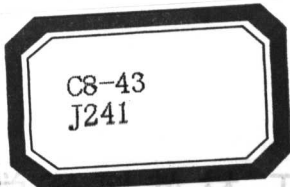
Jia Junping



清华大学出版社



Springer



郑州大学 \*04010080967Z\*

应用统计学系列教材 Texts in Applied Statistics

统计学

Statistics

贾俊平 编著

Jia Junping



00055/66

C8-43

J241



清华大学出版社  
北京



Springer

## 内 容 简 介

本书结合作者多年的教学实践经验和国外优秀统计学教材的成果编写而成。内容包括描述统计方法、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、管理类各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。

### 图书在版编目(CIP)数据

统计学/贾俊平编著. —北京:清华大学出版社,2004

(应用统计学系列教材)

ISBN 7-302-08013-5

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(2004)第 004944 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

封面设计: 常雪影

版式设计: 刘祎森

印 刷 者: 北京四季青印刷厂

装 订 者: 三河市金元装订厂

发 行 者: 新华书店总店北京发行所

开 本: 170×230 印 张: 29.5 字 数: 513 千字

版 次: 2004 年 2 月第 1 版 2004 年 2 月第 1 次印刷

书 号: ISBN 7-302-08013-5/F·736

印 数: 1~5000

定 价: 34.00 元(含光盘)

---

本书如存在文字不清、漏印以及缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话:(010)62770175-3103 或(010)62795704

# 应用统计学系列教材

Texts in Applied Statistics

## 编审委员会

主任：吴喜之

委员：（按姓氏拼音字母排序）

杜子芳 冯士雍 耿直 何书元 贾俊平

金勇进 易丹辉 袁卫 张波 赵彦云

# 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

# 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,许多方法介绍的不够深入;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,做过多地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

FORWARD

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,他们多年的教学经验为本书写作积累了宝贵财富。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。

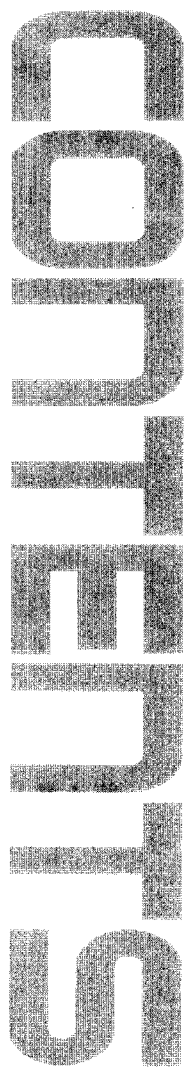
贾俊平

2004年1月于中国人民大学



# 目 录

<b>第 1 章 导论</b> .....	<b>1</b>
1.1 统计及其应用领域 .....	1
1.1.1 什么是统计学 .....	1
1.1.2 统计的应用领域 .....	2
1.1.3 历史上著名的统计学家 .....	5
1.2 统计数据的类型 .....	5
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 .....	7
1.2.3 截面数据和时间序列数据 .....	7
1.3 统计中的几个基本概念 .....	8
1.3.1 总体和样本 .....	8
1.3.2 参数和统计量 .....	10
1.3.3 变量 .....	10
思考与练习 .....	12
人物传记 .....	13
<b>第 2 章 数据收集</b> .....	<b>15</b>
2.1 数据来源 .....	15
2.1.1 数据的间接来源 .....	15
2.1.2 数据的直接来源 .....	16
2.2 调查设计 .....	21
2.2.1 调查方案的结构 .....	21
2.2.2 调查问卷设计 .....	22
2.3 数据质量 .....	32
2.3.1 数据的误差 .....	32



2.3.2 数据的质量要求 .....	32
思考与练习 .....	33
<b>第3章 数据整理与展示 .....</b>	<b>34</b>
3.1 数据的预处理 .....	34
3.1.1 数据审核 .....	34
3.1.2 数据筛选 .....	35
3.1.3 数据排序 .....	38
3.2 分类和顺序数据的整理与显示 .....	39
3.2.1 分类数据的整理与图示 .....	39
3.2.2 顺序数据的整理与图示 .....	44
3.3 数值型数据的整理与显示 .....	46
3.3.1 数据分组 .....	47
3.3.2 数值型数据的图示 .....	51
3.4 统计表 .....	61
3.4.1 统计表的构成 .....	61
3.4.2 统计表的设计 .....	62
思考与练习 .....	63
<b>第4章 数据的概括性度量 .....</b>	<b>68</b>
4.1 集中趋势的度量 .....	68
4.1.1 分类数据:众数 .....	68
4.1.2 顺序数据:中位数和分位数 .....	69
4.1.3 数值型数据:平均数 .....	72
4.1.4 众数、中位数和平均数的比较 .....	77
4.2 离散程度的度量 .....	79
4.2.1 分类数据:异众比率 .....	79
4.2.2 顺序数据:四分位差 .....	80
4.2.3 数值型数据:方差和标准差 .....	80
4.2.4 相对位置的度量:标准分数 .....	84
4.2.5 相对离散程度:离散系数 .....	86
4.3 偏态与峰态的度量 .....	87
4.3.1 偏态及其测度 .....	88

4.3.2 峰态及其测度 .....	89
思考与练习 .....	91
人物传记 .....	95
<b>第 5 章 概率与概率分布 .....</b>	<b>97</b>
5.1 事件及其概率 .....	97
5.1.1 试验、事件和样本空间 .....	97
5.1.2 事件的概率 .....	99
5.1.3 概率的性质和运算法则 .....	100
5.1.4 条件概率与事件的独立性 .....	105
5.1.5 全概率公式与逆概率公式 .....	109
5.2 离散型概率分布 .....	112
5.2.1 随机变量 .....	112
5.2.2 离散型随机变量的概率分布 .....	113
5.2.3 离散型随机变量的数学期望和方差 .....	115
5.2.4 几种常用的离散型概率分布 .....	116
5.3 连续型概率分布 .....	125
5.3.1 概率密度函数 .....	125
5.3.2 正态分布 .....	126
5.3.3 其他连续型概率分布 .....	134
思考与练习 .....	139
人物传记 .....	142
<b>第 6 章 抽样与抽样分布 .....</b>	<b>145</b>
6.1 概率抽样方法 .....	145
6.1.1 简单随机抽样 .....	145
6.1.2 分层抽样 .....	146
6.1.3 系统抽样 .....	146
6.1.4 整群抽样 .....	147
6.2 三种不同性质的分布 .....	147
6.2.1 总体分布 .....	148
6.2.2 样本分布 .....	148
6.2.3 抽样分布 .....	149

6.3	一个总体参数推断时样本统计量的抽样分布 .....	149
6.3.1	样本均值的抽样分布 .....	150
6.3.2	样本比例的抽样分布 .....	155
6.3.3	样本方差的抽样分布 .....	156
6.4	两个总体参数推断时样本统计量的抽样分布 .....	158
6.4.1	两个样本均值之差的抽样分布 .....	158
6.4.2	两个样本比例之差的抽样分布 .....	159
6.4.3	两个样本方差比的抽样分布 .....	159
	思考与练习 .....	161
	人物传记 .....	163
<b>第7章</b>	<b>参数估计 .....</b>	<b>165</b>
7.1	参数估计的一般问题 .....	165
7.1.1	估计量与估计值 .....	165
7.1.2	点估计与区间估计 .....	166
7.1.3	评价估计量的标准 .....	169
7.2	一个总体参数的区间估计 .....	171
7.2.1	总体均值的区间估计 .....	172
7.2.2	总体比例的区间估计 .....	175
7.2.3	总体方差的区间估计 .....	177
7.3	两个总体参数的区间估计 .....	178
7.3.1	两个总体均值之差的区间估计 .....	178
7.3.2	两个总体比例之差的区间估计 .....	184
7.3.3	两个总体方差比的区间估计 .....	185
7.4	样本容量的确定 .....	189
7.4.1	估计总体均值时样本容量的确定 .....	190
7.4.2	估计总体比例时样本容量的确定 .....	191
7.4.3	估计两个总体均值之差时样本容量的确定 .....	191
7.4.4	估计两个总体比例之差时样本容量的确定 .....	192
	思考与练习 .....	193
	人物传记 .....	196
<b>第8章</b>	<b>假设检验 .....</b>	<b>198</b>
8.1	假设检验的基本问题 .....	198

8.1.1	假设的陈述 .....	198
8.1.2	两类错误与显著性水平 .....	201
8.1.3	检验统计量与拒绝域 .....	203
8.1.4	利用 $P$ 值进行决策 .....	205
8.2	一个总体参数的检验 .....	208
8.2.1	总体均值的检验 .....	208
8.2.2	总体比例的检验 .....	216
8.2.3	总体方差的检验 .....	218
8.3	两个总体参数的检验 .....	221
8.3.1	两个总体均值之差的检验 .....	221
8.3.2	两个总体比例之差的检验 .....	231
8.3.3	两个总体方差比的检验 .....	234
	思考与练习 .....	238
	人物传记 .....	242
<b>第 9 章</b>	<b>方差分析与试验设计 .....</b>	<b>244</b>
9.1	方差分析引论 .....	244
9.1.1	方差分析及其有关术语 .....	244
9.1.2	方差分析的基本思想和原理 .....	246
9.1.3	方差分析中的基本假定 .....	248
9.1.4	问题的一般提法 .....	249
9.2	单因素方差分析 .....	250
9.2.1	数据结构 .....	250
9.2.2	分析步骤 .....	251
9.2.3	关系强度的测量 .....	257
9.2.4	用 Excel 进行方差分析 .....	258
9.3	方差分析中的多重比较 .....	259
9.4	双因素方差分析 .....	261
9.4.1	双因素方差分析及其类型 .....	261
9.4.2	无交互作用的双因素方差分析 .....	262
9.4.3	有交互作用的双因素方差分析 .....	269
9.5	试验设计初步 .....	272
9.5.1	完全随机化设计 .....	272

9.5.2	随机化区组设计 .....	274
9.5.3	因子设计 .....	275
	思考与练习 .....	277
	人物传记 .....	280
<b>第 10 章</b>	<b>一元线性回归 .....</b>	<b>283</b>
10.1	变量间关系的度量 .....	283
10.1.1	变量间的关系 .....	283
10.1.2	相关关系的描述与测度 .....	285
10.1.3	相关关系的显著性检验 .....	290
10.2	一元线性回归 .....	291
10.2.1	一元线性回归模型 .....	292
10.2.2	参数的最小二乘估计 .....	294
10.2.3	回归直线的拟合优度 .....	297
10.2.4	显著性检验 .....	301
10.3	利用回归方程进行估计和预测 .....	306
10.3.1	点估计 .....	306
10.3.2	区间估计 .....	307
10.4	残差分析 .....	310
10.4.1	用残差证实模型的假定 .....	310
10.4.2	用残差检测异常值和有影响的观测值 .....	313
	思考与练习 .....	315
	人物传记 .....	320
<b>第 11 章</b>	<b>多元线性回归 .....</b>	<b>323</b>
11.1	多元线性回归模型 .....	323
11.1.1	多元线性回归模型与回归方程 .....	323
11.1.2	估计的多元线性回归方程 .....	325
11.1.3	参数的最小二乘估计 .....	325
11.2	回归方程的拟合优度 .....	327
11.2.1	多重判定系数 .....	327
11.2.2	估计标准误差 .....	329
11.3	显著性检验 .....	329

11.3.1	线性关系检验 .....	330
11.3.2	回归系数的检验和推断 .....	331
11.4	多重共线性 .....	333
11.4.1	多重共线性及其所产生的问题 .....	333
11.4.2	多重共线性的判别 .....	334
11.4.3	多重共线性问题的处理 .....	335
11.5	利用回归方程进行估计和预测 .....	336
11.6	虚拟自变量的回归 .....	337
11.6.1	含有一个虚拟自变量的回归 .....	338
11.6.2	用虚拟自变量回归解决方差分析问题 .....	344
11.7	非线性回归 .....	347
11.7.1	双曲线 .....	347
11.7.2	幂函数曲线 .....	347
11.7.3	对数曲线 .....	348
	思考与练习 .....	350
	人物传记 .....	355
<b>第 12 章</b>	<b>时间序列分析和预测 .....</b>	<b>356</b>
12.1	时间序列及其分解 .....	356
12.2	时间序列的描述性分析 .....	358
12.2.1	图形描述 .....	358
12.2.2	增长率分析 .....	359
12.3	平稳序列的平滑和预测 .....	363
12.3.1	简单平均法 .....	364
12.3.2	移动平均法 .....	364
12.3.3	指数平滑法 .....	367
12.4	有趋势序列的分析和预测 .....	369
12.4.1	线性趋势分析和预测 .....	370
12.4.2	非线性趋势分析和预测 .....	371
12.5	复合型序列的分解 .....	382
12.5.1	季节性分析 .....	382
12.5.2	趋势分析 .....	385
12.5.3	周期性分析 .....	386

思考与练习 .....	388
<b>第 13 章 指数 .....</b>	<b>393</b>
13.1 引言 .....	393
13.2 加权指数 .....	394
13.2.1 权数的确定 .....	394
13.2.2 加权综合指数 .....	395
13.2.3 加权平均指数 .....	398
13.3 指数体系 .....	401
13.3.1 总量指数与指数体系 .....	401
13.3.2 指数体系的分析与应用 .....	402
13.4 几种常用的价格指数 .....	403
13.4.1 零售价格指数 .....	404
13.4.2 消费价格指数 .....	404
13.4.3 生产价格指数 .....	406
13.4.4 股票价格指数 .....	407
13.5 多指标综合评价指数 .....	408
13.5.1 多指标综合评价指数的构建 .....	408
13.5.2 几种常用的综合评价指数 .....	410
思考与练习 .....	413
<b>附录 1 各章练习题答案 .....</b>	<b>416</b>
<b>附录 2 常用统计表 .....</b>	<b>434</b>
表 1 累积二项概率表 .....	434
表 2 累积泊松分布表 .....	438
表 3 标准正态曲线下的面积 .....	439
表 4 $t$ 统计量的临界值 .....	440
表 5 $\chi^2$ 统计量的临界值 .....	441
表 6 $F$ 统计量的临界值 .....	443
表 7 随机数表 .....	451
<b>参考书目 .....</b>	<b>454</b>





# 第 1 章

## 导 论

---

不像其他科学,统计从来不打算使自己完美无缺,统计意味着你永远不需要确定无疑。

——Gudmund R. Iversen

在日常生活中,我们经常会接触到“统计”这一术语,在有关媒体中也经常会看见一些报导使用统计数据、图表等。在本章中,我们将讨论有关统计学的一些基本问题,包括统计学的含义、统计数据及其分类、统计中一些常用的基本概念等。

### 1.1 统计及其应用领域

#### 1.1.1 什么是统计学

统计学已发展成为具有多个分支学科的大家族。统计学家们给统计学下的定义众多,比较有代表性的是不列颠百科全书的定义:

▶ 定义 1.1 统计学(statistics)是收集、分析、表述和解释数据的科学。

统计学的定义告诉我们,统计学是关于数据的科学。其内容包括数据收集、数据整理、数据分析和数据解释等。

数据收集也就是取得统计数据。数据整理是将数据用图表等形式展示出来。数据分析则是通过统计方法研究数据,其所用的方法可分为描述统计方法和推断统计方法。数据解释是对分析的结果进行说明。

▶ 定义 1.2 描述统计(descriptive statistics)是研究数据收集、整理和描述的统计学分支。