

 高等学校现代统计学系列教材

# 应用 统计学

贾俊平 编著

Applied Statistics

高等教育出版社

高等学校现代统计学系列教材

# 应用统计学

Applied Statistics

Yingyong Tongjixue

贾俊平 编著

高等学校现代统计学系列教材编委会 (按姓氏笔画排序)

主 编: 方开泰

副主编: 史宁中 何书元 陈 敏 耿 直

编 委: 马 洪 方开泰 史宁中 杨 虎

何书元 何晓群 张爱军 张崇岐

陈 敏 郑 明 赵彦云 耿 直

曾五一 缪柏其

# 总 序

统计学是一门收集、整理和分析数据的科学和艺术。这里的“数据”通指“信息的载体”，涵盖了大千世界中的文本、图像、视频、时空数据、基因数据等。统计学是一个独立的学科，在历史上曾隶属于数学，但统计学与数学有着本质的区别，因此统计学教育有其自身的特点和要求，这些特点表现为：（1）统计学研究的是随机现象，而数学研究的是确定性的规律；（2）统计学是一门应用性很强的学科，许多概念和原理来自于实际的需要，不是数理逻辑的产物；（3）数据在统计学中扮演了重要的角色。目前，统计学已被列为一级学科。

在过去的 30 年中，随着生命科学、信息科学、物质科学、资源环境、认知科学、工程技术、经济金融和人文科学等众多学科的发展，产生了许多新的统计学分支，如风险管理、数据挖掘、基因芯片分析等。此外，计算机及其有关软件在统计教育和应用中扮演了越来越重要的角色，它们提供了越来越多的图形表达和分析的方法，使得许多原来教科书中重要的内容，现在已变得无足轻重。统计教育必须要改革才能适应高速发展的形势。

大学的统计教育可分为两大类，一类是非统计学专业的课程，另一类是统计学专业的教学设计。非统计学专业的学生学习统计的目的是为了应用，在大学阶段，课程不多，主要是学习基础的统计概念和方法，学会使用统计软件，培养其解决实际问题的能力。统计学专业的课程设置十分重要，应向国际靠拢，对教师队伍的要求也较高。虽然这两类学生的教育有很多共同点，但在课程设置中必须加以区分。

我国的统计教育在过去受苏联的影响很深，把统计学作为数学的一个分支，在内容上偏理论，少应用，过于强调概率论在统计中的作用。统计学是一门应用性很强的学科，应从实际问题、从数据出发，通过统计的工具来揭示数据内部的规律。用“建模”的思路来教统计，使学生能更加容易理解统计的概念和方法，知道如何将实际问题抽象为统计模型，反过来又指导实践。对非统计学专业的学生，要强调统计的应用。学生要能熟练地使用至少一个统计软件包。对于统计学专业的学生，要培养学生对实际问题的建模能

力。有些实际问题可直接应用现有的统计方法来解决,如问卷调查的统计分析。有些问题在初次接触时并不像一个统计问题,必须有坚实的统计基础和对实际问题的洞察力,才能从中发掘出统计模型。要培养学生的这种能力及统计思想(统计思想是统计文化的一部分,是用统计学的逻辑思考问题)。教师在授课中要结合较多的应用例子,要求学生做案例研究,鼓励学生参加建模比赛,参加企业的实际项目。

为满足我国统计教育发展的需要,我们计划编写一套面向高校本科生、特别是一般院校,适用于统计学专业和非统计学专业的系列教材。系列教材的编写宗旨是:突出教学内容的现代化,重视统计思想的介绍,适应现代统计教育的特点及时代发展的新要求;以统计软件为支撑,注重统计知识的应用;内容简明扼要,生动活泼,通俗易懂。编写原则为:(1)从数据出发,不是从假设、定理出发;(2)从归纳出发,不是从演绎出发;(3)强调案例分析;(4)重统计思想的阐述,弱化数学证明的推导。系列教材分为两个方向,一个面对统计学专业,另一个面对非统计学专业和应用统计工作者。

高等学校现代统计学系列教材是适应形势的要求,由高等教育出版社邀请专家组成“高等学校现代统计学系列教材编委会”负责选题、审稿,由高等教育出版社出版。

以上是我们编写这套教材的背景和理念,希望得到读者的支持,特别是高校领导和教学一线的教师的支持。我们希望使用这套教材的师生和读者多提宝贵意见,使教材不断完善。

高等学校现代统计学系列教材编委会

# 前 言

为适应社会发展的需求及教学改革对课程建设调整的新形势,编者结合多年在课程教材建设方面的思考和探索编写了这本教材。主要内容包括数据的描述性分析、统计推断的基本原理和方法以及实际中常用的一些统计方法。全书共包括 10 章,第 1 章介绍统计的一些基本概念、统计数据及其收集方法以及常用的一些统计分析软件等。第 2 章和第 3 章介绍数据的描述性方法,包括数据的图表展示和一些常用统计量的计算方法。第 4 章介绍常用的一些概率分布以及样本统计量的概率分布。第 5 章和第 6 章介绍推断统计的基本原理和方法,包括参数估计和假设检验。第 7 章介绍类别变量的分析方法,包括一个类别变量的拟合优度检验以及两个类别变量的独立性检验。第 8 章介绍相关与回归分析方法,包括相关分析和一元线性回归分析。第 9 章介绍时间序列的预测方法。第 10 章介绍统计指数,包括指数的编制及其应用。

本书以纸质教材配套数字课程形式出版,数字课程可以补充、拓展纸质版教学内容,加强教学效果,进一步增强教材的基础性、实用性、先进性和启发性,提高教材使用的实效性。本书的内容和特色如下:

1. 摆脱理论式、公式化教学,将重心放在对统计思想、方法的理解及应用上,培养读者将统计用于实处。
2. 结合使用 SPSS (19.0 中文版) 和 Excel 两个软件,对每章涉及的方法在章末给出了具体的操作步骤,读者按此操作就会得到所需的统计分析结果。除部分方法展示了计算过程外,多数统计计算都使用软件直接给出结果,并对输出结果做了详细解读。
3. 数字课程分 8 个模块:统计英语、Excel 统计函数、拓展阅读、释疑解难、配套数据、名词解释、知识导图、基本公式,从知识的深度和广度上对教材内容进行补充,辅助教与学。可登录易课程网站“<http://abook.hep.com.cn/1243709>”获取资源,更多资源不断更新中。
4. 内容编排上打破常规,创新设计。在不影响传统阅读学习习惯下,在纸质教材相应位置进行资源标注,引导读者配合数字课程进行个性化学

习，同时资源以二维码形式呈现，扫描后即可实时学习。

统计的精髓是使复杂问题简单化，而不是把简单问题复杂化；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；学好统计的关键是掌握如何运用统计思维来思考问题，而不是简单地记住那些死的统计知识。本书力求按照这样的思路来写作，但未必能在内容中全部得到体现。限于编著者水平有限，教材中缺点和错误在所难免，恳请各位专家、同仁和读者指正。

贾俊平

2013 年 6 月

于中国人民大学统计学院

# 目 录

|                   |        |
|-------------------|--------|
| 第1章 统计、数据和计算机     | 1      |
| 开篇案例：怎样理解这样的一些结论？ |        |
| 1.1 统计及其应用领域      | 3      |
| 1.1.1 统计学研究什么     | 3      |
| 1.1.2 统计的应用       | 5      |
| 1.2 怎样获得统计数据      | 9      |
| 1.2.1 变量与数据       | 9      |
| 1.2.2 数据的来源       | 10     |
| 1.3 统计与计算机        | 11     |
| 主要术语              | 15     |
| 软件应用              | 15     |
| 思考与练习             | 15     |
| <br>第2章 用图表展示数据   | <br>17 |
| 开篇案例：用哪些图形展示奖牌？   |        |
| 2.1 用图表展示类别数据     | 19     |
| 2.1.1 生成频数分布表     | 19     |
| 2.1.2 类别数据的图示     | 23     |
| 2.2 用图表展示定量数据     | 27     |
| 2.2.1 生成频数分布表     | 27     |
| 2.2.2 定量数据的图示     | 29     |
| 2.3 合理使用图表        | 40     |
| 主要术语              | 41     |
| 软件应用              | 41     |
| 思考与练习             | 42     |

|       |                   |    |
|-------|-------------------|----|
| 第3章   | 用统计量描述数据          | 45 |
|       | 开篇案例：哪名运动员的发挥更稳定？ |    |
| 3.1   | 位置的描述             | 47 |
| 3.1.1 | 平均数               | 47 |
| 3.1.2 | 中位数和分位数           | 49 |
| 3.1.3 | 用哪个值代表一组数据的位置     | 51 |
| 3.2   | 离散程度的描述           | 52 |
| 3.2.1 | 极差和四分位差           | 53 |
| 3.2.2 | 方差和标准差            | 53 |
| 3.2.3 | 比较几组数据的离散程度：变异系数  | 56 |
| 3.3   | 分布形状的描述           | 58 |
|       | 主要术语              | 59 |
|       | 软件应用              | 59 |
|       | 思考与练习             | 59 |
| 第4章   | 概率分布              | 61 |
|       | 开篇案例：彩票中奖的概率有多大？  |    |
| 4.1   | 度量事件发生的可能性        | 63 |
| 4.2   | 随机变量的概率分布         | 64 |
| 4.2.1 | 随机变量及其概括性度量       | 64 |
| 4.2.2 | 随机变量的概率分布         | 67 |
| 4.2.3 | 其他几个重要的统计分布       | 71 |
| 4.3   | 样本统计量的概率分布        | 74 |
| 4.3.1 | 统计量及其分布           | 74 |
| 4.3.2 | 样本均值的分布           | 75 |
| 4.3.3 | 样本比例的分布           | 79 |

|                        |     |
|------------------------|-----|
| 4.3.4 统计量的标准误差         | 79  |
| 主要术语                   | 81  |
| 软件应用                   | 81  |
| 思考与练习                  | 81  |
| <br>                   |     |
| 第5章 参数估计               | 83  |
| 开篇案例：科学家在哪个年龄段易取得重大突破？ |     |
| 5.1 参数估计的基本原理          | 85  |
| 5.1.1 点估计与区间估计         | 85  |
| 5.1.2 评价估计量的标准         | 90  |
| 5.2 总体均值的区间估计          | 92  |
| 5.2.1 大样本的估计           | 92  |
| 5.2.2 小样本的估计           | 94  |
| 5.3 总体比例的区间估计          | 96  |
| 5.4 总体方差的区间估计          | 97  |
| 5.5 样本量的确定             | 99  |
| 5.5.1 估计总体均值时样本量的确定    | 99  |
| 5.5.2 估计总体比例时样本量的确定    | 100 |
| 主要术语                   | 102 |
| 软件应用                   | 102 |
| 思考与练习                  | 102 |
| <br>                   |     |
| 第6章 假设检验               | 105 |
| 开篇案例：乙肝疫苗有效吗？          |     |
| 6.1 假设检验的基本原理          | 107 |

|                |     |
|----------------|-----|
| 6.1.1 怎样提出假设   | 107 |
| 6.1.2 怎样做出决策   | 109 |
| 6.1.3 怎样表述决策结果 | 115 |
| 6.2 总体均值的检验    | 118 |
| 6.2.1 大样本的检验   | 118 |
| 6.2.2 小样本的检验   | 121 |
| 6.3 总体比例的检验    | 123 |
| 6.4 总体方差的检验    | 124 |
| 主要术语           | 126 |
| 软件应用           | 126 |
| 思考与练习          | 126 |

## 第7章 类别变量分析 129

开篇案例：男孩子更喜欢玩游戏吗？

|                                |     |
|--------------------------------|-----|
| 7.1 一个类别变量的拟合优度检验              | 131 |
| 7.1.1 期望频数相等                   | 131 |
| 7.1.2 期望频数不等                   | 134 |
| 7.2 两个类别变量的独立性检验               | 136 |
| 7.2.1 列联表与 $\chi^2$ 独立性检验      | 136 |
| 7.2.2 应用 $\chi^2$ 检验应注意的问题     | 139 |
| 7.3 两个类别变量的相关性度量               | 140 |
| 7.3.1 $\phi$ 系数和 Cramer's V 系数 | 140 |
| 7.3.2 列联系数                     | 141 |
| 主要术语                           | 143 |
| 软件应用                           | 143 |
| 思考与练习                          | 143 |

|                     |         |
|---------------------|---------|
| 第 8 章 相关与回归分析       | 145     |
| 开篇案例：子代与父代一样吗？      |         |
| 8.1 变量间关系的度量        | 147     |
| 8.1.1 变量间是什么样的关系    | 147     |
| 8.1.2 用散点图描述相关关系    | 148     |
| 8.1.3 用相关系数度量关系强度   | 151     |
| 8.2 一元线性回归          | 154     |
| 8.2.1 一元线性回归模型      | 154     |
| 8.2.2 参数的最小二乘估计     | 155     |
| 8.2.3 回归直线的拟合优度     | 158     |
| 8.2.4 显著性检验         | 161     |
| 8.3 利用回归方程进行预测      | 163     |
| 8.3.1 平均值的置信区间      | 163     |
| 8.3.2 个别值的预测区间      | 164     |
| 主要术语                | 168     |
| 软件应用                | 168     |
| 思考与练习               | 168     |
| <br>第 9 章 时间序列预测    | <br>171 |
| 开篇案例：如何预测社会消费品零售总额？ |         |
| 9.1 时间序列的成分和预测方法    | 173     |
| 9.1.1 时间序列的成分       | 173     |
| 9.1.2 预测方法的选择与评估    | 176     |
| 9.2 平稳序列的预测         | 178     |
| 9.2.1 移动平均预测        | 178     |
| 9.2.2 简单指数平滑预测      | 178     |

|                     |         |
|---------------------|---------|
| 9.3 趋势预测            | 181     |
| 9.3.1 线性趋势预测        | 181     |
| 9.3.2 非线性趋势预测       | 186     |
| 9.4 多成分序列的预测        | 190     |
| 9.4.1 分解预测          | 190     |
| 9.4.2 Winter 指数平滑预测 | 194     |
| 主要术语                | 198     |
| 软件应用                | 198     |
| 思考与练习               | 199     |
| <br>第10章 指数         | <br>201 |
| 开篇案例：如何理解 CPI?      |         |
| 10.1 指数的含义与分类       | 203     |
| 10.2 加权指数的编制        | 203     |
| 10.2.1 加权综合指数       | 204     |
| 10.2.2 加权平均指数       | 205     |
| 10.3 几种常用的价格指数      | 207     |
| 10.3.1 居民消费价格指数     | 208     |
| 10.3.2 工业生产者出厂价格指数  | 208     |
| 10.3.3 股票价格指数       | 208     |
| 主要术语                | 210     |
| 思考与练习               | 210     |
| <br>参考书目            | <br>211 |

