



21世纪高等院校经济管理类规划教材

应用统计学

□ 潘鸿 张小宇 吴勇民 主编

ECONOMICS & MANAGEMENT

- 突出统计方法和技术的应用
- 提供课件、教案、习题答案、补充阅读资料



人民邮电出版社
POSTS & TELECOM PRESS



★ 21世纪高等院校经济管理类规划教材

应用统计学

□ 潘鸿 张小宇 吴勇民 主编



人民邮电出版社

北京

图书在版编目(CIP)数据

应用统计学 / 潘鸿, 张小宇, 吴勇民主编. — 北京:
人民邮电出版社, 2011.5
21世纪高等院校经济管理类规划教材
ISBN 978-7-115-24982-1

I. ①应… II. ①潘… ②张… ③吴… III. ①应用统计学—高等学校—教材 IV. ①C8

中国版本图书馆CIP数据核字(2011)第038815号

内 容 提 要

本书系统讲述应用统计学基本知识和基本技能, 融入电子表格(Excel)的实际应用, 介绍参数估计、假设检验、方差分析、相关与回归、时间序列分析、指数分析等应用统计方法。本书突出体现应用统计学的应用性、层次性和趣味性, 坚持体系的完整性, 注重传统与创新的统一及统计理论在实践中的应用, 便于授课与学生自习使用。

本书提供电子课件、电子教案、习题库、上机操作数据、实验指导等资料。

本书可作为经济管理类专业本科生的教材, 也可作为从事相关工作人员的培训教材。

21 世纪高等院校经济管理类规划教材

应用统计学

-
- ◆ 主 编 潘 鸿 张小宇 吴勇民
责任编辑 李育民
执行编辑 万国清
 - ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号
邮编 100061 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
大厂聚鑫印刷有限责任公司印刷
 - ◆ 开本: 787×1092 1/16
印张: 13.25 2011 年 5 月第 1 版
字数: 317 千字 2011 年 5 月河北第 1 次印刷

ISBN 978-7-115-24982-1

定价: 28.00 元

读者服务热线: (010)67170985 印装质量热线: (010)67129223

反盗版热线: (010)67171154

广告经营许可证: 京崇工商广字第 0021 号

21 世纪高等院校经济管理类规划教材

编 委 会

编委会主任：李随成

编委会秘书组：万国清 李海峰 宣颖

（排名不分先后，按姓名拼音排序）

- | | | | |
|-----|--------------|-----|--------------------------------|
| 陈 良 | 南京财经大学，教授 | 秦成德 | 西安邮电学院，教授 |
| 段兴民 | 西安思源学院，教授 | 卿 涛 | 西南财经大学，教授 |
| 高忠庭 | 河北经贸大学，研究员 | 任 远 | 西安交通大学，教授 |
| 郝晓雁 | 山西财经大学，教授 | 宋瑞卿 | 山西财经大学，教授 |
| 何元贵 | 广东外语外贸大学，教授 | 宋 玮 | 中国人民大学，副教授 |
| 侯荔江 | 西南财经大学，副教授 | 王飞跃 | 贵州财经学院，教授 |
| 侯 颖 | 齐齐哈尔大学，副教授 | 王辅东 | 广东工业大学，高级经济师 |
| 胡日东 | 华侨大学，教授 | 王汝林 | 中国电子商务协会移动商务专家咨
询委，常务副主任，教授 |
| 黄 勇 | 对外经贸大学，教授 | 王胜洲 | 河北经贸大学，副教授 |
| 姬会英 | 南阳理工学院，副教授 | 王文博 | 西安交通大学，教授 |
| 戢守峰 | 东北大学，教授 | 韦 苇 | 西安翻译学院，教授 |
| 李宝瑜 | 山西财经大学，教授 | 吴声怡 | 福建农林大学，教授 |
| 李 兵 | 对外经济贸易大学，副教授 | 吴战勇 | 南阳理工学院，副教授 |
| 李海峰 | 西安外事学院，副教授 | 徐 黎 | 南阳理工学院，副教授 |
| 李红霞 | 西安科技大学，教授 | 许永兵 | 河北经贸大学，教授 |
| 李 丽 | 安徽工业大学，副教授 | 闫秀荣 | 哈尔滨师范大学，教授 |
| 李淑平 | 中原工学院，教授 | 于翠华 | 齐齐哈尔大学，副教授 |
| 李随成 | 西安理工大学，教授 | 查会琼 | 安徽工业大学，副教授 |
| 李雪欣 | 辽宁大学，教授 | 张世君 | 河南理工大学，教授 |
| 李子猷 | 陕西师范大学，教授 | 张伟峰 | 宝鸡文理学院，教授 |
| 梁 军 | 宁波工程学院，教授 | 张宗和 | 浙江工商大学，教授 |
| 马 洪 | 上海财经大学，教授 | 周 苏 | 浙江大学，教授 |
| 马元生 | 西安思源学院，教授 | 朱金生 | 武汉理工大学，教授 |
| 米子川 | 山西财经大学，教授 | | |
| 乔 瑞 | 南阳理工学院，教授 | | |

丛 书 序

本丛书根据普通本科高校的教学需求设计并编写，丛书涉及的书目包含经济管理类的专业基础课教材及部分专业课教材。

众所周知教材质量的核心是内容质量，为了将本丛书中每种教材均打造成为优秀教材，由众多教育一线的专家学者组建成的编委会为本套丛书审纲、审稿。可以说从书中每种教材均凝结了众多编委会成员的心血。我们相信通过大家共同的努力，本套丛书将有希望涌现出几种堪称优秀的、能适应普通本科高校教学需求的、高质量的立体化教材。

为满足社会对人才的需求，普通本科的教学改革持续进行，不少教学改革已经取得令人瞩目的成果。当前高校教师对教学改革配套教材的需求呼声很高，但我们考虑教材出版周期较长和其教书育人的特性使其只适合作为已成熟教学改革方案的载体，未经过较长时间检验的教学改革成果直接体现在教材上并不合适，故而本丛书的所有教材均基于较为成熟的教学改革成果。

教材虽然是传播教学改革成果的最佳工具之一，但仅靠教材本身很难全面地将教改思想传达到位，需要有与教材配套的教案及其他辅助资料。为此我们将本丛书打造成主教材与配套电子资料包相结合的立体化教材，提高教材的应用性和实用性。

本丛书中每种教材所配的电子资料包内均含作者精心制作的电子课件、电子教案、习题答案，有些教材还提供了案例分析、学习指导等更为丰富的教学素材或学习素材。配套电子资料包可参照教材所附的“配套资料索取说明”索取。立体化教材建设与精品课程建设的内容高度相似，本丛书中一些教材源自于已建成的精品课程，部分教材正在进行相关的精品课程建设，这些教材都有精品课程网站支持。

为使教材更加适应当前的教学需求，我们在保证理论体系完整性和系统性的同时，坚持以活泼的创作风格和贴近工作、社会的内容提升教材的可读性和可用性。

尽管我们力图为高校提供高质量的、立体化的、符合未来两三年教改趋势及教学需求的优秀教材，但正如一位国家级教学名师所说：“教材不是编出来的，是教出来的，来回反复修改，来回‘磨’出来的。”我们这套丛书还未经过“打磨”，再加上我们的水平有限，尚存在已知的和未知的一些不足，我们有决心持续地“打磨”这套教材，也希望读者给予反馈以资我们修正，使本套教材尽早达到“优秀”的水准（编委会联系方式 wanguoqingljw@163.com 或 goodbook2010@tom.com）。

丛书编委会

2010年4月

应用统计学是高等院校经济管理类专业的核心课程之一,但学生却普遍反映比较难学,原因是应用统计学用到的数学知识较多,计算工作量大而复杂。为了解决这一矛盾,在近二十年的教学实践中,笔者一直致力于探索统计学的实际应用,曾为专科生、本科生、研究生开设了“社会经济统计学原理”、“农业统计学”、“国民经济统计学”、“应用统计学”、“经济计量软件应用”等课程,在教学活动中积累了大量的经验,特别是在应用统计学教学方面进行了多次有益的尝试。

全书共分十章,系统讲述应用统计学基本知识和基本技能,并结合计算机应用介绍抽样推断、假设检验、方差分析、相关与回归、时间序列分析、指数分析等应用统计方法,本书将本领域的最新研究成果和教学实践结合起来,融入电子表格的相关内容(考虑界面友好性和功能齐全性,本书电子表格软件选用 Excel 2007,较早版本 Excel 也可参考使用;除数据分析外,其他任务也可选用 WPS 表格),实现了理论讲授与实验操作的有机结合。

第一章绪论,对“统计”一词的内涵进行分析,并梳理统计学发展历史,重点介绍应用统计学中常用的概念,包括总体、样本、标志、指标、指标体系、参数、统计量和变量等。

第二章至第四章分别介绍了统计数据的收集、整理及描述性统计。第五章至第七章为统计推断的内容,主要包括抽样、参数估计和假设检验。第八章为方差分析,第九章为时间序列分析、第十章为统计指数。

在编写过程中,我们力图使本教材突出以下几个特点:

其一,突出应用性。在系统阐述统计学有关理论的前提下,突出统计方法和技术的应用,并努力通过与电子表格相结合,使之程序化。

其二,体现层次性。无论是内容结构还是叙述手法都力求言简意赅,层次分明,方便学生掌握。

其三,立体化建设。本教材不仅提供电子课件、电子教案、习题答案和上机操作数据、补充阅读等资料,还提供教学网站支持(www.jluyytj.com),理论讲授与教学实验,方便教师授课及学生学习(配套资料的索取方式参见本书末页“配套资料索取说明”)。

本书适合经济管理类专业的本科生使用,也可作为从事相关工作人员培训教材使用。

本书由多位教师共同编写完成,潘鸿负责全书的内容设计和统稿,同时负责第三、八、九章的编写和实验指导书的编写;张小宇负责第二、四、六、七章的编写;吴勇民负责第一、五、十章的编写。

参与收集数据、文字校对和课件修改的有兰欢、刘刚、刘志强、罗甜、马杰、乔俏、覃多贵、王卓识、吴坤、赵锴、周惠珺、罗侠。

本书在编写过程中,参考和采纳了主要参考文献所列的著作中的一些观点,在此向各位作者表示诚挚的谢意。

鉴于我们的水平有限，编写的教材肯定会存在不少错误和不妥之处，恳请广大读者提出宝贵意见。

潘 鸿
2011 年 2 月

目 录

第一章 绪论	1
第一节 统计与统计学	1
一、统计学的含义	1
二、统计学的研究对象	2
三、统计学的产生与发展	4
第二节 统计学的应用与分科	6
一、统计学的应用	6
二、统计学的分科	9
第三节 统计学的基本概念	11
一、总体和样本	11
二、标志、指标和指标体系	12
三、参数、统计量和变量	14
本章小结	14
思考与练习	15
第二章 统计数据的收集	17
第一节 统计数据的来源	17
一、统计数据的间接来源	17
二、统计数据的直接来源	17
第二节 统计数据的类型	18
一、分类数据、顺序数据和数值型数据	19
二、截面数据、时间序列数据和面板数据	19
第三节 调查方案设计	20
一、调查目的	20
二、调查对象	20
三、调查单位	20
四、调查表	21
第四节 调查问卷设计	21
一、问卷及问卷设计标准	21
二、问卷的开发程序	22

三、问卷措辞应注意的问题	22
四、回答项目的设计	23
本章小结	27
思考与练习	28
第三章 统计数据的整理	30
第一节 统计数据的预处理	30
一、统计数据的审核	30
二、统计数据的排序	30
三、统计数据的筛选	31
四、统计数据的插补	34
五、数据透视表和透视图	36
第二节 统计表	37
一、统计表的分类	37
二、统计表的设计规则	39
第三节 统计分组和频数分布	40
一、统计分组的概念	40
二、频数与频数分布	42
三、利用电子表格绘制不变和可变直方图	43
第四节 统计图	46
一、统计图的类型	46
二、利用电子表格绘制常用的统计图	53
本章小结	55
思考与练习	55
第四章 描述性统计	58
第一节 集中趋势的度量	58
一、平均数	58
二、众数	61
三、中位数	61
四、分位数	62
五、众数、中位数和平均数的比较	63

第二节 离散程度的度量	64	一、无偏性	90
一、极差	65	二、有效性	90
二、平均差	65	三、一致性	91
三、方差和标准差	66	第四节 总体均值的区间估计	91
四、变异系数	67	一、正态总体、方差已知(或非正态总体、大样本)	91
第三节 偏态和峰态的度量	68	二、正态总体、方差未知小样本	93
一、偏态及其测度	68	第五节 总体比例的区间估计	96
二、峰态及其测度	69	第六节 总体方差的区间估计	97
本章小结	71	第七节 两个总体均值之差的区间估计	98
思考与练习	72	一、大样本	98
第五章 抽样	74	二、小样本	100
第一节 抽样概述	74	第八节 两个总体比例之差的区间估计	102
一、全及总体	74	第九节 两个总体方差比的区间估计	103
二、样本	74	第十节 样本容量的确定	104
三、抽样框	75	一、估计总体均值时样本容量的确定	104
四、抽样单位	75	二、估计总体比例时样本容量的确定	105
五、总体参数和样本统计量	75	本章小结	105
第二节 抽样方法及其分类	76	思考与练习	107
一、简单随机抽样	76	第七章 假设检验	108
二、分层抽样	78	第一节 假设检验的相关概念	108
三、系统抽样	78	一、原假设与备择假设	108
四、整群抽样	80	二、显著性水平和两类错误	109
第三节 抽样分布	80	三、检验统计量和拒绝域	110
一、抽样分布的概念和种类以及几种常见的分布	80	四、显著性概率	112
二、样本均值(平均数)的抽样分布	81	第二节 假设检验的基本步骤	113
三、样本比例(成数)的抽样分布	82	一、构造原假设和备择假设	113
四、样本方差的抽样分布	84	二、选择合适的统计量	113
本章小结	84	三、计算临界值和拒绝域	113
思考与练习	85	四、作出判断	114
第六章 参数估计	87	第三节 总体均值的检验	114
第一节 参数估计的相关概念	87	一、单个总体均值的检验	114
第二节 参数估计方法	88	二、两个总体均值之差的检验	117
一、点估计与大数定律	88	第四节 总体比例的检验	120
二、区间估计	88	一、单个总体比例的检验	121
第三节 参数估计量的评价标准	90	二、两个总体比例之差的检验	122
		第五节 总体方差的检验	123

一、单个总体方差的检验	123	一、长期趋势分析	160
二、两个总体方差比的检验	126	二、季节变动分析	167
本章小结	128	三、循环变动与不规则变动的分析	171
思考与练习	130	本章小结	173
 第八章 方差分析	132	思考与练习	173
第一节 方差分析的基本问题	132	 第十章 统计指数	176
一、方差分析及相关术语	132	第一节 统计指数概述	176
二、方差分析的基本思想和原理	133	一、统计指数的概念	176
三、方差分析的基本假定	135	二、统计指数的分类	176
第二节 单因素方差分析	136	三、统计指数的作用	177
一、单因素方差分析的数据结构	136	第二节 综合指数和平均指数	178
二、单因素方差分析步骤	136	一、综合指数	178
三、用 Excel 进行单因素方差分析	140	二、平均指数	180
四、方差分析中的多重比较	142	第三节 指数体系与因素分析	181
本章小结	143	一、总量指标变动的因素分析	182
思考与练习	143	二、平均指标变动的因素分析	182
 第九章 时间序列分析	146	第四节 几种常用的经济指数	184
第一节 时间序列概述	146	一、居民消费价格指数	184
一、时间序列的概念	146	二、生产者价格指数	187
二、时间序列的分类	147	三、工业生产指数	187
三、时间序列的编制原则	148	四、股票价格指数	188
第二节 时间序列的描述性分析	149	本章小结	189
一、发展水平	149	思考与练习	189
二、平均发展水平	150	 附录 《应用统计学》实验指导书	192
三、增长量与平均增长量	153	说明	192
四、发展速度与增长速度	154	实验一 统计工作过程实验	193
五、平均发展速度与平均增长速度	156	实验二 电子表格在统计分析中的应用	193
六、发展速度分析应注意的问题	157	实验三 统计学知识综合运用	194
第三节 时间序列的构成要素	157	 主要参考文献	198
第四节 加法模型与乘法模型	158	 配套资料索取说明	199
第五节 时间序列的分解分析	159		

第一章 绪 论

作为社会生活中的一项重要活动——统计对于大家可能是既陌生又熟悉。在日常生活中，我们常常接触到这样的词汇：据统计或据不完全统计；我们也常常在媒体看到或听到：国家统计局公布的国内生产总值的年度增长率，物价指数的上涨率等类似的描述。在没有系统学习统计学这门科学之前，大家可能都认为统计学不过是一些枯燥乏味的数据和资料，是一些单调和没有人情味的表格和图表。其实，这是对统计学的片面认识。那么，究竟什么是统计学呢？我们为什么要学习统计学呢？统计学究竟在我们的社会生活中扮演什么样的角色呢？本章将介绍统计学的一些基本问题，包括统计学的含义，统计学的产生和发展，统计学的研究对象、过程和方法以及统计学学习过程中常用的几个基本概念，最后，将介绍本教材各章节所使用的统计软件。

第一节 统计与统计学

一、统计学的含义

统计作为一种古老的社会实践活动，已在人类社会存续数千年，而统计学作为这种社会实践活动的经验总结和理论概括，也有了三百多年的历史。可以说，自从有了人类文明社会，就有了统计实践活动。无论是在人们的日常生活中还是在国家治理中，统计都作为一种重要的管理手段被广泛地重视，从而演变成一种系统的理论工具。

当今人们对“统计”一词赋予诸多含义，因此，统计有多种多样的理解。它可以指统计数据的收集活动，即统计工作；也可指统计活动的结果，即统计数据；还可以指分析数据的方法和技术，即统计学。

与其他学科一样，统计学是随着人类社会的发展和社会管理的需要而发展起来的。伴随着人类社会由低级向高级演进，统计活动也逐渐变得复杂起来，仅仅用数字计量的简单统计活动已不能满足社会的需要，人们开始对客观现象进行定量分析，并不断地在数据的处理和进行分析中进行抽象、概括和总结，于是，统计学的产生就顺理成章了。

统计学家给统计学（statistics）下了不同但相似的定义，《不列颠百科全书》将统计学定义为“一门收集、分析、表述和解释数据的科学”，《韦伯斯特国际辞典》对统计学的定义是“统计学是一门收集、分析、解释和提供数据的科学”；美国学者特里奥拉在他所著的《初级统计学》教科书中认为，“统计学指的是一组方法，用来设计实验、获得数据，然后在这些数据的基础上组织、概括、演示、分析、解释和得出的结论”；我国学者贾俊平在其所编著教科书《统计学》（第二版）给统计学下的定义是“统计学是收集、处理、分析、解释数据并从数

据中得出结论的科学”。上述关于统计学的概念有一个共同的特点，那就是与数据的解释与分析相关。因此，总结上述定义，我们可以将统计学的含义概括如下。

定义 1.1 统计学，是一门研究大量社会现象和自然现象的总体数量方面包括收集、整理和分析统计数据的方法论科学。

为了能对统计学的定义有一个清楚的理解，我们将统计学归纳为如下几个特点：

1. 数量性

统计学的认识力首先表现在它以准确性和无可争辩的事实为基础，同时，这些事实用数字加以表现，具有简短性和明显性。数量性是统计学研究对象的重要特点。

2. 总体性

统计研究社会现象和自然现象的数量方面指的是总体的数量方面。从总体上研究社会现象和自然现象的数量方面，是统计学区别于其他社会科学的一个重要特点。社会现象是各种社会规律相互交错作用的结果，它呈现出一种复杂多变的情景。统计学对社会现象总体数量方面的调查研究，用的是综合研究方法，而不是对单个事物的研究，但其研究过程是从个体到总体，即必须对足够大量的个体（这些具体都表现为一定的差别、差异）进行登记、整理和综合，使它过渡到总体的数量方面，从而把握社会现象的总规律、总水平及其变化发展的总趋势。比如，了解市场物价情况，统计着眼于整个物价指数（如居民消费价格指数 CPI）的变动，而不是某一种商品价格的变动，但物价统计必须从了解相关代表性商品的价格变动情况开始，才能经过一系列的统计工作过程，达到对于物价总体数量变动情况认识的目的。

3. 具体性

统计学研究的数量方面是指社会现象和自然现象的具体的数量方面，而不是抽象的数量关系，这是它不同于数学的重要特点。统计学与数学相比，有如下几个不同。首先，从研究对象上看，数学研究的是没有量纲或没有单位的抽象的数，而统计学研究的是有具体实例或计量单位的数据；其次，统计学与数学研究中所使用的逻辑方法也是不同的，数学研究中使用的是纯粹的演绎，而统计学则是演绎与归纳相结合，占主导地位的是归纳。数学家可以坐在屋里，凭借聪明的大脑从假设命题出发，推导出漂亮的结果，而统计学家则要深入实际收集数据，并与具体的实际问题相结合，经过大量的归纳才能得出有益的结论。

4. 指导性

统计是研究大量社会现象和自然现象数量规律的一门方法论科学，所以统计学方法可以帮助其他学科探索内在的数量规律性，为人们的实践活动提供方法论依据，从而指导人们的实践活动。例如，美国季度国内生产总值增长率、月度通胀率或失业率一公布，马上会引起银根松紧、利率调整甚至华尔街股市涨跌等一系列的连锁反应。不仅如此，这些数据往往能在几分钟内传遍全世界，一些国家不得不迅速做出相应的反应。统计数据对经济政策的影响和指导性作用显而易见。

5. 广泛性

统计学研究的数量方面非常广泛，它指全部社会现象的数量方面。广泛性这一特点，可区别于研究某一特定领域的其他社会科学（如政治学、经济学、社会学、法学等）。

二、统计学的研究对象

统计学将大量社会现象和自然现象的总体数量方面作为研究对象，并提供一套科学的统

计方法来分析和探索数量的内在规律性。我们可以通过以下几个具体的例子来说明统计学研究的对象。

对于人类的身高，人们通常认为，父母的身高越高，孩子的身高就会越高；反之，父母的身高越矮，孩子的身高就会越矮。统计学家费朗西斯·高尔顿（Francis Galton）发现了这是一个错误的认识！在他的生物统计实验室里，他收集人类的身高数据，并对这些数据进行分析，发现非常高的父亲，其儿子往往要比父亲矮一些；而非常矮的父亲，其儿子往往要比父亲高一些。似乎是某种神秘的力量，使得人类的身高从高矮两极移向所有人的平均值，这个现象被他称之为“向平均回归”（regression to the mean）。他将这一思想纳入了所构建的统计模型中，从而对人类身高进行仔细研究。他说，假设不发生这种向平均值的回归，那么从平均意义上看，高身材父亲的儿子将与他们的父亲一样也会很高，在这种情况下，一些儿子的身高必须高于他们的父亲，以抵消身材比父亲矮小者的影响，使平均值不变。高身材者这一代人的儿子也将如此，那么会有一些儿子身材更高。这个过程将一代一代传下去。同样地，将会有一部分儿子身材比他们的父亲矮小，而且有一部分儿子将更加矮小，如此下去，不用多少代，人类种族就将由特别高和特别矮的两极构成。显然，上述的情形并没有发生，人类的身高在平均意义上趋向于保持稳定。

在现实生活中被人们饶有兴趣地观察的另一个例子是关于人类性别的结构。众所周知，一个家庭新生婴儿的男女性别是随机分布的。从表面上看，新生婴儿的性别比例似乎没什么规律可循。但如果对新生婴儿的性别进行大量观察统计，就会发现性别比例还是有规律可寻的，即婴儿总数中男孩的数量要多于女孩，大致为每生育 100 个女孩，就有 107 个左右的男孩。这个 107 : 100 的比例就是新生婴儿男女性别的数量规律性，古今中外大致相同，它是由人类社会长期遗传与发展的结果。人类社会要延续、要发展，就要保持男女人数的大致相同。但从新生儿的性别比例看，却出现了男女的失衡。这难道不违背人类永续生存的自然规律吗？统计学家们对这一现象进行了分析解释，他们认为，尽管从新生婴儿来看，男性要高于女性，但男孩的死亡率高于女孩，到了中青年时，男女人数就大致相同了。进入中老年后，男性的死亡率仍然高于女性，导致男性的平均预期寿命比女性短，老年男性要少于女性。从一个国家乃至全人类看，如果没有人为的 B 超、堕胎等干扰，其规律是：婴幼儿时男性略多于女性，中青年时男女人数大致相同，老年时女性又略多于男性。这样既保证人类在中青年结婚生育时性别的大致平衡，又使得在人口总数上男女也大体相当，有利于人类社会的进化和发展。统计学家往往通过对大量数据的分析，来挖掘数据现象背后的规律。

上面两个例子说明，通过多次观察或试验得到大量的统计数据，利用统计方法是可以探索出其内在数量规律性的。客观事物本身的特点再加上科学设计的统计方法使得我们能够在探索数据的规律方面不断前进。

从客观事物特点来说，任何客观事物都是必然性与偶然性的对立统一，同样，任何一个数据，也都是必然性与偶然性共同作用的结果。必然性反映了事物本质的特征和联系，是比较稳定的，因而它决定了事物的内在本质是有规律可循的；偶然性则反映了该事物每个表现形式的差异。如果客观事物只有必然性一个方面的特征，事物的表现形式会比较简单，我们可以比较容易把握它的规律性。正是由于偶然性的存在，造成了事物的表现形式与必然性和规律性发生偏移，从而形成了表面形式的千姿百态和数据表现形式的千差万别，这样，就导致必然性的数量规律性被掩盖在表面的差异之中了。

前面所说的两个例子，个别家庭父母和子女的身高和每个新生儿的性别都是随机现象，其表现形式也是充满偶然性的，但每个例子本身可以通过对大量数据的收集整理和研究得出其内在的规律性。应用统计方法可以从偶然性中探索到内在的、本质的数量规律：从统计方法来看，统计学提供了一系列的方法，专门用来收集数据、整理数据、显示数据的特征，进而分析和探索（或推断）出事物总体的数量规律性。当然，如果事物本身的规律比较简单，所用的统计方法也就相对容易；如果事物本身的规律错综复杂，所用的统计方法也就相对复杂。而这些又都是统计学所重点关注的问题。

三、统计学的产生与发展

尽管统计活动在人类社会已存续数千年，但统计学作为一门独立的学科却只有 300 多年的历史。一般认为，统计学产生于 17 世纪中叶，其形成过程是从几个不同的领域开始的，在各自的领域又产生了许多不同的学派，如形成于德国的“国势学派”，形成于英国的“政治算术学派”，产生于 19 世纪中叶的“数理统计学派”、“社会统计学派”等。然而，考察统计学产生的历史，无论是古典统计学、近代统计学，还是现代统计学，其发展过程是沿袭两条主线展开：其一是以“政治算术学派”为开端形成和发展起来的以社会经济问题为主要研究对象的社会经济统计；其二是以概率论的研究为开端并以概率论为基础形成和发展起来的以方法和应用统计研究为主的数理统计。我们对统计学历史的考察就是以这两条主线展开的。

1. 政治算术——社会经济统计

政治算术学派学产生于 17 世纪中叶的英国，主要代表人物是威廉·配第（William Petty, 1623—1687）和约翰·格朗特（John Graunt, 1620—1674）。

威廉·配第在其代表性著作《政治算术》（1676 年）一书中，运用大量的数字资料和定量的研究方法对英国、法国、荷兰三国的经济实力进行了分析与比较，为统计学的产生奠定了基础。配第的这些开创性工作，得到了马克思的高度评价，马克思在《资本论》中评价配第“是政治经济学之父，在某种程度上也可以说是统计学的创始人”。

另一个代表性人物约翰·格朗特，他在 1662 年出版了《关于死亡表的自然观察与政治观察》。当时，伦敦瘟疫流行，死亡情况严重，引起社会不安。他根据“死亡率公报”对伦敦人口的出生率、死亡率、性别比例，作了分类计算和预测，证明没有悲观的必要。另外，在该书中，他通过大量的观察，研究并发现了人口与社会现象中重要的数量规律性。如新生儿的性别男女比例稳定在 14 : 13，即我们在前面的例子中谈到的 107 : 100，并解释了这一比例出现的原因。他得出了男性在各年龄组中死亡率高于女性；新生儿的死亡率较高；一般疾病与事故的死亡率较稳定，而传染病的死亡率波动较大等很多富有启发性的结论。更为重要的是，格朗特在研究中运用多种方法对统计资料进行间接的推算，并相互印证。由于约翰·格朗特的这些研究成果，该书被许多统计学家誉为“真正统计科学的肇端”。

政治算术学派的统计学家们基本上沿着威廉·配第开创的政治经济统计和约翰·朗特开创的人口统计而展开进一步的研究。在经济统计方面，出现了农业统计、工商统计和物价指数计算方法的研究。在人口统计方面，由于研究对象出现了扩展，出现了保险统计、卫生统计或医疗统计等分支。除此之外，人口调查、社会调查、道德统计等也逐渐发展起来。

19 世纪中叶以后，包括政治统计、人口统计、经济统计、犯罪统计、道德统计、社会统

计等多方面内容的“社会统计学”开始发展起来，与之相适应的社会调查与社会研究也有了较大发展，并成为社会科学研究的重要方法之一。人们试图通过社会调查，收集、整理、分析资料，以揭示社会现象和问题，并提出解决问题的具体办法。例如，法国经济学家和社会改良家李·普莱（Le Play, 1806—1882）曾利用“家庭预算表”来研究社会生活水平和贫困问题；德国统计学家恩格尔（Ernest Engel, 1821—1896）在他的《比利时工人家庭的生活费》（1895年）一文中，提出了著名的“恩格尔法则”，即“家庭收入越多，则饮食消费支出在家庭收入中所占百分比越少；家庭收入越少，则饮食支出在家庭收入中所占百分比越大”。在这一法则基础上，他又引申出至今仍得到人们广泛应用的“恩格尔系数”，用于衡量人们生活水平。

与此同时，经济调查和经济统计学也迅速发展起来。比如，早在1835年挪威就与人口普查相结合从事农业统计调查；1939~1840年法国进行第一次农业普查；1846年比利时进行了农业普查；1790年美国进行了第一次人口普查等。在这期间，美国、德国等还进行了工业普查。到了20世纪，这些调查活动日臻成熟。为满足国家和社会客观需要而进行的各种经济调查，不仅取得了大量的社会经济数据，为经济学家据以证实或提出新的经济理论提供了依据，也为统计学家从中概括和提出新的统计方法提供了数据材料。到了19世纪中叶，德国经济学和统计学家克尼斯（K.G.A.Knies）在他的论文《独立科学的统计学》（1850年）中提出统计学是一门独立的科学。可见，随着统计实践和学科门类的发展和分工，统计学作为一门对社会现象和自然现象进行数量对比分析的方法论科学，已为社会所公认。

此外，20世纪对国民收入的计算和研究、指数的编制及其方法研究、时间序列分析、经济预测和计量经济学等，也都取得了长足的进展。而这些成就，其思想渊源都可追溯到政治算术学派所开创的统计学的理论和方法，这为后来的社会经济统计的发展奠定了基础。

2. 概率论——数理统计

统计学产生与发展的第二条主线是以概率论为基础形成和发展起来的以随机现象为主要研究对象的数理统计。

在西方，概率论的思想渊源最早可以追溯到意大利的文艺复兴时代，最初的研究是为赌徒们找出掷骰子取胜的一套办法，从而合理地算出成功的机率。这一时期的标志性著作是意大利数学家卡尔达诺（Kiriolamo Cardano, 1501—1576）出版的《论赌博》一书，他的研究成果为运用数学理论研究概率论开辟了道路。著名的天文学家伽利略（Galilei Galileo, 1564—1642）写了一篇同样题目的论文，提出了概率论的基本原理，奠定了数理统计的基础。

几乎与政治算术学派的研究一致，概率论的真正研究历史是从17世纪中叶开始的，这一时期是古典统计学的黄金时代，其主要奠基人是法国的帕斯卡（B. Pascal, 1623—1662）和费马特（P.Fermat, 1606—1665）。他们将赌博中出现的具体问题归纳为一般的概率原理，为后来概率论和统计学的发展奠定了重要的基础。到了18世纪，积极研究概率论的还有瑞士数学家贝努里（J.Bernoulli, 1654—1705），全面论述了概率论原理并将概率论建立在数学的基础上，在其所著的《推算法》一书中，提出了著名的“贝努里定理”，这是大数法则的早期形式。

古典统计时期的概率论基本上是独立发展的，它与社会经济统计没有太多的联系，但这一时期也有学者将概率论尝试着应用于社会现象的研究。到了19世纪，用概率论研究社

会经济现象的人日益增多,在这方面做出重大贡献的是法国数学家拉普拉斯(P.S.Laplace, 1749—1827)和比利时统计学家凯特勒(A.Quetelet, 1796—1874)。拉普拉斯在其所著的《概率论分析》一书中,总结了前人的研究成果,并以大数法则为桥梁,将概率论与社会经济现象联系起来;凯特勒在发展和应用概率与统计方面做出了重要贡献,其著作主要有《概率论书简》和《社会物理学》等,其主要贡献是将自然科学的研究方法引进社会现象的研究中,他发展了大量观察法并为数理统计学的发展奠定了基础。

从19世纪中叶到20世纪中叶,概率论的进一步发展为数量统计学的形成和发展奠定了基础。英国生物学家和统计学家高尔顿(F.Galton, 1822—1911)首次提出并阐述了“相关”的概念,并首次提出了“相关系数”的概念;英国数理统计学家卡尔·皮尔逊(K.Pearson, 1857—1936)提出了计算复相关和偏相关的方法,并将复相关和回归理论扩展到许多领域。

20世纪初期,大工业的发展对产品质量检验问题提出了新的要求,即只抽取少量产品作为样本对全部产品的质量好坏做出推断。因为大批量产品要做全面的检验,既费时、费钱,又费人力,加之有些产品质量的检验要做破坏性检验,全部检验已不可能。数理统计学派的先驱、英国统计学家戈塞特(W.S.Gosset, 1876—1937)建立了“小样本理论”,即所谓的“t分布”。通过这个理论,人们可以从大量的产品中只抽取较小的样本完成对全部产品质量的检验和推断,这样就使统计学进入了现代统计学(主要是推断统计学)的新阶段。以后著名统计学家R.A.费希尔(R.A.Fisher, 1890—1962)研究出了F统计量、极大似然估计、方差分析等方法,奈曼(J.Neyman, 1894—1981)创立了区间估计理论,并和伊根·夏普·皮尔逊(E.S.Pearson, 1857—1936)发展了假设检验理论。20世纪杰出的统计学家不胜枚举,正是他们的努力,推动了统计理论的发展和应用。到了20世纪中叶,现代统计学的基本框架已经确立。

从20世纪50年代以来,统计理论、方法和应用进入了一个全面发展的新阶段。一方面,统计学受计算机科学、信息论、混沌理论、人工智能等现代科学技术的影响,新的研究领域层出不穷,如多元统计分析、现代时间序列分析、贝叶斯统计、非参数统计、线性统计模型、探索性数据分析、数据挖掘等。另一方面,统计方法的应用领域不断扩展,几乎所有的科学研究都离不开统计方法。因为无论是自然科学、工程技术、农学、医学、军事科学,还是社会科学都离不开数据,对数据进行研究和分析就必然要用到统计方法,现在连纯文科领域的法律、历史、语言、文学、新闻等都越来越重视对统计数据进行分析,国外的人文与社会学科普遍开设统计学的课程,可见,统计学已经成为一门极其重要的基础性学科。

第二节 统计学的应用与分科

一、统计学的应用

统计方法是适用于所有学科领域的通用数据分析方法,只要有数据的地方就会用到统计方法。目前,随着定量研究的日趋重要,统计方法已被应用到自然科学和社会科学的众多领域,统计学也已发展成为若干分支学科组成的科学体系。可以说,几乎所有的研究领域都要用到统计方法,比如日常生活、公司或企业的生产经营管理、政府部门和学术研究领域等。

下面列出一些统计方法在不同领域中应用的例子。

1. 生活中的统计学

我们在日常生活中，无时无刻不遇到与统计相关的问题，只要我们自觉地运用统计学的角度看问题，就能更容易看清事物的本质，以在我们生活中最为常见的彩票为例。人们购买彩票都想中大奖，希望能用 2 元钱获得 500 万元的大奖。殊不知，大家在抱着这样的心态购买彩票时，却很少理性计算中奖的概率。以国内流行的体彩七星彩和福彩双色球为例，只要稍加计算，就能知道我们有多少胜算。体彩七星彩的玩法如下：该类体育彩票每注由 7 位自然数组成，每一位数字都有 0~9 共 10 种变化，其排列组合共有 $10 \times 10 \times 10 \times 10 \times 10 \times 10 \times 10$ 种，即 10 的 7 次方，也就是共有 1 000 万注，中一等奖的概率是 $\frac{1}{10\,000\,000}$ 。而福彩双色球

(33 选 6、16 选 1) 中头奖的概率是

$$\frac{1}{C_{33}^6 \times C_{16}^1} = \frac{1}{17\,721\,088}$$

显然，用 2 元钱博得 500 万元的大奖这样的小概率事件在统计学中被认定是几乎不可能发生的事情。即便有的人实在受不住大奖的诱惑，也可以比较这两种彩票中奖的概率，理性地选择购买体彩七星彩，从而增加自己中奖的机率。

2. 企业中的统计学

在企业的生产经营活动中，无时无刻不在利用统计信息作为其行为的指南，充分有效的统计数据能够使企业在了解市场供求状况、把握投资机会、选择投资方向、降低投资风险、调整产品结构、加快产品开发、制定发展战略、加强产品质量管理、进行财务分析以及进行市场前景预测等方面都具有积极的作用。

以我国影碟机市场为例来说明统计学在企业决策方面的指导作用。20 世纪 90 年代末，我国影碟机市场已经进入“微利时代”，影碟机厂商之间的竞争已达到白热化，影碟机厂商是要在这个微利市场继续经营，还是要改变投资方向选择退出呢？他们在做出决策之前会对这个市场进行详细地调研，考虑众多因素后再对其经营进行抉择。蓝田市场研究公司 2000 年 3 月对全国 31 个城市的影碟机市场中消费者未来两年购买影碟机的意向进行了统计抽样调查，调查样本家庭用户总计 22 429 个。图 1.1 说明了全国城市居民未来购买影碟机的意向。

从图 1.1 中的统计数字可以看出，城市的影碟机市场发展空间有限，在所有被访家庭中，仅有 14.4% 的家庭未来两年内有打算购买影碟机的意向，影碟机在城市市场的发展空间狭小，这些信息为影碟机厂商经营决策提供了重要的参考。

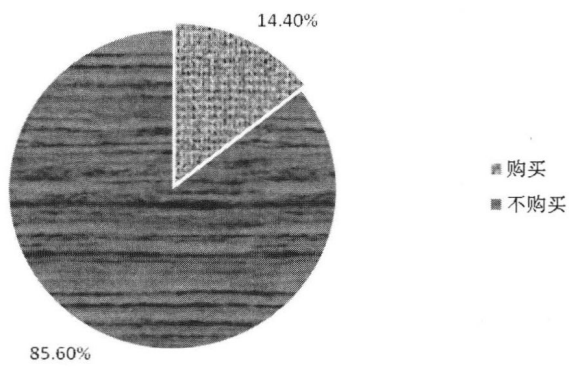


图 1.1 全国消费者两年购买影碟机的意向（2000 年）

3. 军事中的统计学

统计在军事中也起着极为重要的作用。要打赢高技术条件下的局部战争，必须对有限的