



“十三五”普通高等教育本科部委级规划教材

陈 磊◎主编

应用统计学

★ 阐明核心原理 ★ 构建知识体系 ★ 引导实践操作 ★ 提升应用技巧



国家一级出版社



中国纺织出版社

全国百佳图书出版单位



“十三五”普通高等教育本科部委级规划教材

应用统计学

陈磊◎主编

国家一级出版社



中国纺织出版社

全国百佳图书出版单位

内 容 提 要

统计学在现代科学和日常生活中的应用十分广泛。得益于信息技术的飞速发展,大量的数据可以被记录和保存下来。如何更有效地使用这些数据?如何从纷杂的数据中累积经验?如何通过丰富的数据进行验证假设?这些问题直接影响了个人的认知高度。本教材主要面向管理专业学生,由浅入深,内容涵盖了描述统计、参数估计、假设检验、方差分析、时间序列等现代统计学核心的方法,分别介绍了抽样、样本分布、假设检验、方差分析等重要的统计学概念,再结合案例和实际操作强化了对概念的理解,更积累了处理和分析类似问题的经验。本教材引入了案例思考形式的问题,把问题留给学生,让学生设身处地去思考对策。本教材强化了专业统计软件的操作,注重培养学生使用工具解决问题的能力。本教材还配有微信公众号和微视频讲解,欢迎得到广大学生和教师的关注和订阅。

图书在版编目(CIP)数据

应用统计学 / 陈磊主编. —北京:中国纺织出版社, 2018.3

“十三五”普通高等教育本科部委级规划教材

ISBN 978-7-5180-4113-8

I. ①应… II. ①陈… 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(2017)第232474号

策划编辑:曹炳镛 责任印制:储志伟

中国纺织出版社出版发行

地址:北京市朝阳区百子湾东里A407号楼 邮政编码:100124

销售电话:010-67004422 传真:010-87155801

http://www.c-textilep.com

E-mail: faxing@c-textilep.com

中国纺织出版社天猫旗舰店

官方微博 http://weibo.com/2119887771

三河市宏盛印务有限公司印刷 各地新华书店经销

2018年3月第1版第1次印刷

开本:710毫米×1000毫米 1/16 印张:22

字数:325千字 定价:48.00元

凡购本书,如有缺页、倒页、脱页,由本社图书营销中心调换

高等院校“十三五”部委级规划教材经济管理类编委会

主任：

倪阳生：中国纺织服装教育学会会长

赵宏：天津工业大学副校长、教授、博导

郑伟良：中国纺织出版社社长

赵晓康：东华大学旭日工商管理学院院长、教授、博导

编委：（按姓氏音序排列）

蔡为民：天津工业大学管理学院院长、教授、硕导

郭伟：西安工程大学党委常委、教授、博导

胡剑峰：浙江理工大学经济管理学院院长、教授、博导

黎继子：武汉纺织大学国际教育学院院长、教授、博导

琚春华：浙江工商大学计算机与信息工程学院院长、教授、博导

李晓慧：北京服装学院教务处处长兼商学院院长、教授、硕导

李志军：中央财经大学文化与传媒学院党总支书记、副教授、硕导

林一鸣：北京吉利学院执行校长、教授

刘晓喆：西安工程大学高教研究室主任、教务处副处长、副研究员

刘箴言：中国纺织出版社工商管理分社社长、编审

苏文平：北京航空航天大学经济管理学院副教授、硕导

单红忠：北京服装学院商学院副院长、副教授、硕导

石 涛：山西大学经济与工商管理学院副院长、教授、博导

王核成：杭州电子科技大学管理学院院长、教授、博导

王进富：西安工程大学管理学院院长、教授、硕导

王若军：北京经济管理职业学院院长、教授

乌丹星：国家开放大学社会工作学院执行院长、教授

吴中元：天津工业大学科研处处长、教授

夏火松：武汉纺织大学管理学院院长、教授、博导

张健东：大连工业大学管理学院院长、教授、硕导

张科静：东华大学旭日工商管理学院副院长、教授、硕导

张芝萍：浙江纺织服装职业技术学院商学院院长、教授

赵开华：北京吉利学院副校长、教授

赵志泉：中原工学院经济管理学院院长、教授、硕导

朱春红：天津工业大学经济学院院长、教授、硕导

PREFACE | 前言

人类社会进步的源动力是知识的积累。哲学中的经验主义（Empiricism）学派指的就是人类的知识主要来自感知经历，通过经历的累积凝结成了知识。这种经验主义获取知识的过程正是符合统计学收集数据，分析数据，解释数据从而累积为经验和知识的逻辑过程。可以说统计学是人类获取知识和积累知识的重要手段，正因为漫漫历史长河中人类不断趋利避害，积累经验才成就了今天的辉煌。

随着科技的不断进步，“大数据”“互联网+”等理念不断融入日常生活中，数据收集过程正以前所未有的速度大幅跃进，数据的数量倍增，质量也更有保障。另外，随着计算机技术的发展，数据分析也取得了长足的进步，新版本的统计软件可以在很短的时间内处理更多更复杂的统计计算，阻碍统计学发展的瓶颈被一一突破。如今，数据处理与分析已经成为现代社会不可或缺的技能，无论在经济规划、科学研究还是在商业发展、日常生活中，数据分析技术已经广泛渗透至各个领域。

本书迎合了信息化时代的特点，旨在从实际统计软件操作出发，提高同学的应用和分析能力。编者希望同学们通过本书的学习掌握多种统计学分析方法，并学会运用软件更科学更有效地分析、解释各种现象，进而可以用所学的方法解决实际问题，学以致用。编者更希望同学在信息爆炸的未来掌握信息数据筛选、处理、分析、凝练的手段，让这些信息数据为我所用，通过信息的分析辨认真伪，避免在信息浪潮中随波逐流，迷茫失措。

本书主要面向管理专业学生。教材的编写主要遵循三个原则：第一，从实践中学习。本课程大幅度加强软件操作的讲解，理想的上课环境是计算机房或者学生自带电脑。第二，语言精练。对复杂的概念逻辑精练归纳，尽量用简明达意的语言讲述。第三，鼓励学生思考。教材引入大量案例思考形式的问题，

把问题留给学生，让学生设身处地去思考对策。本书与其他统计学教材相比有三个特色：

第一，软件多样化。与其他重理论、轻实践的教材相比，本书提供了大量软件操作练习题。这些软件包括 SPSS，Excel 统计分析库，MiniTab，R 语言等。其中 R 语言作为免费的开源软件更适合学生获取。丰富的插件，完善的功能和全球范围的沟通支持平台也促使 R 语言突飞猛进，成为全球主流的统计开发语言。为此本书特用一个章节介绍 R 语言的操作和应用。由于本书加入了大量的软件操作内容，编者更推荐在计算机房等环境下使用本书授课，这样可以获得更佳的学习效果。

第二，注重思考和解释。本书与一般的统计学教材不同，不再以公式和计算作为学习重点，对于类似公式推导、例题演算等笔算内容点到为止。本书更侧重于培养学生统计建模能力和对结果分析、解释的能力。其原因是人类笔算能力十分有限，处理数据量略大的计算时耗时耗力，而且容易出错，而这些计算工作完全可以交给计算机软件处理，过程便捷，结果精确。编者也期望学生将学习的重心放在思考问题和解决问题上，不需要在背公式和计算上耽误精力。

第三，解决问题。本书中列举大量实际问题，以例题和案例等形式出现。其目的是让学生直接面对真实的问题，运用所学的方法解决问题，思考分析结果，进而寻求更佳解决方案。这些案例中使用的数据大部分是真实数据，与其他教材中凑变量、代公式的模式完全不同。通过这种案例形式的解决问题培训，学生可以更好地掌握所学的知识，在未来生活中用所学的方法建立统计模型，分析并解决问题。

这套教材由东华大学旭日工商管理学院统计学教研组主编。本书还配有微信公众号和微视频讲解，欢迎广大学生和教师关注和订阅。如有疑问也可以在微信公众号中提出，编者会尽量回答。编者水平有限，书中错误和疏漏在所难免，敬请同学和老师提出宝贵意见，编者一定认真听取并作改进和修订。

感谢同学和老师采用本书，感谢出版社的大力支持。

陈磊、王满、王维红

2016 年于东华大学

CONTENTS | 目 录

第一章 绪论

第一节 统计学与统计数据	002
一、统计数据 (data)	002
二、统计学 (statistics)	003
三、统计研究的过程	005
四、统计学的产生和发展	006
五、统计学在商务和经济管理中的应用	008
第二节 统计学中的基本概念	010
一、总体和样本	010
二、参数和统计量	010
三、数据的类型	010
第三节 常用统计软件介绍	012
一、SPSS 软件概述	012
二、其他软件	012
【本章小结】	014
【本章习题】	014

第二章 数据搜集和预处理

第一节 数据收集	019
一、数据来源分类	020
二、从数据采集的具体手段分类	021
三、基于网络的新方法	024
第二节 数据的预处理和软件操作	026
一、数据收集的技巧和数据审查	026
二、利用软件快速剔除	027
三、数据预处理及软件操作	028
四、数据的转化及软件实现	036
五、数据的编码和录入	039
【本章小结】	043
【本章习题】	043

第三章 数据探索和特征描述

第一节 集中趋势统计量及软件操作	047
一、数据类型描述	047
二、集中趋势（数据的中心）的描述	051
第二节 偏离程度的描述和计算	060
一、离散程度的描述	060
二、偏度和峰度	064
三、离散程度的软件操作	066
第三节 描述统计的图形生成及软件操作	070
一、箱体图	070
二、直方图	072
三、茎叶图	074

【本章小结】	075
【本章习题】	078

第四章 参数估计

第一节 参数估计及软件实现	082
一、参数估计原理	082
二、总体均值的估计	085
第二节 总体比例和方差的区间估计	097
一、总体比例参数估计	097
二、总体方差的参数估计	101
【本章小结】	106
【本章习题】	108

第五章 假设检验

第一节 假设检验概述	112
一、假设检验的基本思想	112
二、假设检验中两种类型的错误	113
三、假设检验的步骤	115
四、检验功效	119
五、检验的 p 值	120
第二节 单个总体的参数假设检验	123
一、总体均值的检验	123
二、总体方差的假设检验	132
三、总体比例的检验	136
第三节 两个总体的参数检验	138

一、两个总体均值之差的检验	138
二、两个总体比率之差的检验	149
三、两个总体方差比的检验	150
【本章小结】	151
【本章习题】	154

第六章 方差分析

第一节 方差分析的原理和单因素方差分析	160
一、方差分析的原理	160
二、单因素方差分析过程	165
三、SPSS 实现单因素方差分析	172
第二节 双因素方差分析	176
一、双因素方差分析的软件操作	176
二、双因素方差分析的计算逻辑	187
三、有交互作用的双因素方差分析的计算逻辑	190
【本章小结】	193
【本章习题】	196

第七章 线性回归分析方法及其软件实现

第一节 一元线性回归	200
一、线性回归的含义	200
二、回归的种类	203
三、定义直线回归方程式	204
四、回归方程式的软件操作和参数解释	207
第二节 多元线性回归	215
一、多元线性回归的概念	215

二、多元线性回归的软件操作	217
三、自变量个数与模型质量的关系	219
四、线性关系的检验	221
五、回归系数的检验	222
六、共线性和模型选择方法	225
【本章小结】	231
【本章习题】	234

第八章 非线性回归模型

第一节 常用的非线性回归模型	238
一、非线性回归模型与线性回归	238
二、常用非线性回归模型及其特点	239
三、非线性回归模型估计	243
四、相关指数	244
五、非线性回归在 SPSS 中的实现	245
第二节 Logistic 回归模型介绍	260
一、二分类变量 Logistic 回归模型 (Binary Logistic regression model)	260
二、二分类变量 Logistic 回归模型的性质和特征	262
三、利用计算机处理 Logistic 回归模型	263
【本章小结】	270
【本章习题】	270

第九章 时间序列分析

第一节 时间序列的分类	273
一、时间序列的定义	273

二、时间序列分类	274
第二节 时间序列的构成要素	278
一、时间序列的影响因素	278
二、时间序列分解模型	279
三、影响因素分析	280
第三节 平稳时间序列分析与预测	286
一、预测精度	286
二、平稳时间序列	287
三、平稳时间序列的预测	288
第四节 非平稳时间序列分析与预测	293
一、时间序列的分解	293
二、趋势模型的预测	294
三、具有季节效应模型的预测	295
四、具有循环趋势的预测	299
【本章小结】	299
【本章习题】	300

第十章 统计指数

第一节 统计指数及其种类	301
一、统计指数概述	301
二、统计指数分类	302
三、指数编制中的问题	304
第二节 综合指数	305
一、指数编制方法	305
二、简单指数	306
三、加权指数	308

第三节 常用统计指数介绍	315
一、消费价格指数	315
二、工业生产指数	319
三、股票价格指数	320
【本章小结】	321
【本章习题】	322

附 录

附录 1 : Excel 常用统计函数	324
附录 2 : 检验量表格	330
一、左侧 Z 检验表格	330
二、Z 右侧检验表格	331
三、t 值检验表格	333
四、卡方检验	334
五、F 检验	335
附录 3 : 学习资料下载	337
一、微信公众号	337
二、课程微视频	337
三、数据文件	338
附录 4 : 参考文献	338

第一章 绪论

【本章学习目标】

1. 理解统计学的含义
2. 理解描述统计和推断统计
3. 了解统计学的应用领域
4. 了解数据的类型
5. 理解统计中的几个基本概念

【引导案例】

第一财经（CBN）是中国非常有影响力、品种较完整的财经媒体集团，隶属于中国第二大传媒集团——上海东方传媒集团（SMG）。

第一财经拥有第一财经电视、第一财经日报、第一财经广播、第一财经周刊、第一财经网站、第一财经研究院，并在积极探索数字媒体业务（如无线业务）和金融商业信息服务业务（如实时财经新闻业务和数据库业务）。第一财经致力于为中国广大投资者和商界、经济界人士，以及全球华人经济圈提供实时、严谨、高质的财经新闻，打造具有公信力和强大影响力的全媒体金融与商业信息服务集团。

第一财经金融价值榜（CFV）利用第一财经专业财经资讯平台优势，集聚智库和专家合力，权威客观反映中国金融行业公司和产品价值，打造在学术界、商业界和媒体界具有影响力的金融价值评估标准，提供具有行业权威性及导向性的竞争力排名榜单。

“第一财经金融价值榜”在上海启动，每年评选经过3个多月的调研、面访，于11月进行数据汇总统计，公布最终榜单和报告，并同期举办“第一财经年会”。第一财经特邀中国社会科学院金融研究所，并依托第一财经研究院，借鉴国内外优秀的研究体系，结合中国金融行业的特点，共同制定了符合中国国情发展现状、具有一定研究价值和中国特色的金融价值研究模型，并在此基础上开展“第一财经金融价值榜”评选活动。

评选指标主要分为客观指标和主观指标两部分，并按照每个奖项的特性设置一定的加权比例。客观指标数据通过数据搜集、问卷调查、新闻跟踪得出。而主观指标则主要通过第一财经专业团队评分和专家顾问委员会评分等方式得出。

CFV 还将邀请中国知名经济学家和金融研究人员组成专家顾问委员会，参与评选体系的制定和主观指标的评定，并为评选过程提供谋划、指导与监督。

在上述案例中需要调研和分析许多数据，比如金融公司的资产规模、产品种类、盈利能力、风险状况等变量数据。本章介绍统计学中的基本概念，讨论可用于统计分析的数据类型，并说明如何取得这些数据。我们还将介绍描述统计和推断统计，它们是将数据转化为有意义的且易于解释的统计信息的方法。

第一节 统计学与统计数据

一、统计数据 (data)

我们知道，统计学是通过对统计数据进行分析来发现统计规律的一门学科。在日常工作和生活中，我们经常看到媒体上的各种报道中有很多数据，例如：

2015年7月14日（周二）收盘，沪指跌1.16%，报收3924.49点，成交金额8301亿元。金融、运输设备、建筑和资源类股票拖累大盘，券商板块领跌；其中，华泰证券、中国中车、中煤能源和中国交建跌幅居首；

据彭博新闻社调查得到的分析师预估中值，2015年中国第二季度经济同比增幅为6.8%，低于第一季度7%的同比增速；

中国人民银行授权中国外汇交易中心公布，2015年5月28日银行间外汇市场人民币汇率中间价为：1美元对人民币6.1202元，1欧元对人民币6.6882元，100日元对人民币4.9619元，1港元对人民币0.78912元，1英镑对人民币9.4148元，1澳大利亚元对人民币4.7456元，1新西兰元对人民币4.4501元，1新加坡元对人民币4.5459元，1加拿大元对人民币4.9150元，人民币1元对0.58642林吉特，人民币1元对8.4000俄罗斯卢布……

有时，为了研究某一问题，我们也常常通过调查或实验得到各种统计数

据,如消费调查中消费者的年龄、性别、收入、文化程度、消费金额……企业效益调查中企业规模、企业类型、所属行业,销售收入、利润总额、销售利润率、资产总额……纺织品特性实验中的洗涤方式、洗涤水温、缩水率……这些都是统计数据。

统计数据不仅仅是数值型和分类型数据,也包括文字型数据、音频数据、图像数据以及文本数据等。特别是在当前大数据时代背景下,数据的体量、传播速度、非结构化数据的比例都急速增加,与传统的数据的有了质的变化,这给传统的统计分析提出了很大的挑战,也给统计学的发展提供了空前的机遇。目前,文本挖掘、关联分析、决策树等数据挖掘方法都日益得到越来越多的应用。

二、统计学 (statistics)

观察和统计是我们认识世界的重要途径。在很多自然和社会现象中,我们往往是首先通过大量的观察和统计,发现某种现象存在一定的规律,进而由相关学科去揭示这些规律存在的原因,这样逐步去认识这个世界的。当然,有时候人们观察统计到的现象也有可能是“伪”的,这就需要进一步用统计方法去检验,或者用专业知识去证明这些现象并不是科学规律而是一种错觉,这也使得我们对世界的认识更加深刻。

科学的统计方法还能够使我们智慧地解读我们身边的数据和信息。在我们日常的生活中,每天都充斥着大量的信息,哪些信息是“真”的?哪些信息是“假”的?怎样解释身边的信息或数据?你是否常常会错误地解读数据呢?以下就是几个常见的错误地解释数据的例子:

(1) 某年5月投放了某种冰淇淋广告,在接下来的三个月中,冰淇淋的销售额增加了30%,因此,该广告效果非常好。

在这个例子中,对数据的解读显然是不合理的。因为不管是否投放广告,冰淇淋的销售额在6月、7月、8月通常都会增加,这是一种由天气变化导致的效应。人们却把这个结果解释成广告的影响,而事实上,这个结果是受时间 t 的影响。

(2) 城市的教堂越多,犯罪案件也越多,因此,教堂数量增加导致犯罪案件的增加。

在这个例子中,对数据的解读显然也是不合理的。事实上,在大城市里,