

“十二五”国家重点出版物出版规划项目



现代统计学系列丛书

现代基础 统计学

方开泰 彭小令

**Modern Elementary
Statistics**

高等教育出版社

“十二五”国家重点出版物出版规划项目

 现代统计学系列丛书

现代基础 统计学

**Modern Elementary
Statistics**

方开泰 彭小令

Xiandai Jichu Tongjixue

高等教育出版社·北京

内容提要

统计学是收集、整理和分析数据的科学和艺术。随着社会的发展,统计学的应⽤几乎遍及了自然科学、社会科学等所有的学科。本书依据编者多年在香港和珠海采⽤国际流⾏教科书讲授统计学的经验,⽣动全面地介绍了数据的收集与描述、统计推断和统计建模等主要内容。

本书可以作为高等学校经管类、文科类专业本科生和研究⽣的教科书或参考书。授课教师可根据学生的具体情况对本书的内容进⾏适当取舍。

图书在版编目(CIP)数据

现代基础统计学/方开泰,彭小令著. --北京:
高等教育出版社, 2014.11
ISBN 978-7-04-039674-4

I. ①现… II. ①方… ②彭… III. ①统计学 IV.
①C8

中国版本图书馆CIP数据核字(2014)第079724号

策划编辑 李蕊
插图绘制 杜晓丹

责任编辑 胡颖
责任校对 刁丽丽

封面设计 赵阳
责任印制 刘思涵

版式设计 赵阳

出版发行 高等教育出版社
社址 北京市西城区德外大街4号
邮政编码 100120
印刷 唐山市润丰印务有限公司
开本 787mm×960mm 1/16
印张 23
字数 430千字
购书热线 010-58581118

咨询电话 400-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.landaco.com>
<http://www.landaco.com.cn>
版次 2014年11月第1版
印次 2014年11月第1次印刷
定价 39.80元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换
版权所有 侵权必究
物料号 39674-00
审图号 GS(2013)5013号

现代统计学系列丛书编委会

(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

统计学是收集、整理和分析数据的科学和艺术。这里的“数据”泛指“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。统计学是一门独立的学科，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 and 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。

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“现代统计学系列丛书编委会”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。

现代统计学系列丛书编委会

前言

统计学是收集、整理和分析数据的科学和艺术。随着信息科学、风险管理、数据挖掘、资源环境、工程技术、经济金融和人文科学等学科的发展,对统计学提出了许多新问题,从而推动了统计学的理论和方法迅猛发展;同时统计学应用的领域也越来越广,几乎遍及自然科学、社会科学等所有学科。统计学发展的日新月异,也要求一个现代人要具备初步的统计知识,并对基本的统计思想有所了解,这些定会让他在生活和工作中受益良多。

目前,许多高校要求本科生掌握一定的统计学知识,修一至两门统计学和概率论的课程,我们感觉到统计教育的任务十分重要,特别是面向本科生的基础统计教育。在过去的许多年,我们在香港和珠海给理学院、工商管理学院、社会科学学院的学生系统讲授了基础统计学,采用国际流行的教科书。这些教科书图文并茂,例题丰富,附有统计软件介绍,深受读者的欢迎。但这些教科书具有浓厚的西方文化背景,有些例题的相关知识不易理解,同时,书本身往往过厚过重,达 600~700 页,让读者望而生畏。本书尝试用现代社会的例子来说明统计学的概念和方法,希望读者知道统计学和我们的生活、工作密切相关,从而对学习统计产生浓厚的兴趣。

通览国际流行的基础统计学教科书,我们发现内容大同小异,表明教师对基础统计学的内容体系是有共识的。本书两位作者在大学授课多年,尤其在教授人文社科和工商管理学生统计方面有多年的教学经验,在撰写时我们参考了国内外有关的教科书,同时考虑到非理工科学生的特点和需要,设计了 11 章内容。这 11 章主要包括 3 部分:数据的收集与描述(1、2、3 章),统计推断(4、5、6、7、9、10 章)和统计建模(8、11 章)。考虑到大多数中学已经将不少基础统计学的内容纳入在数学的课程中,如条形图、直方图等统计图表,均值、方差、中位数等数值特征,概率的计算,甚至线性回归的初步介绍。本书将这些内容包括进来是为了统计知识的连贯性和系统性,对于那些和高中数学课程有重复的内容,教师可以快速回顾相关章节,把更多的时间用于介绍新的或更深入的内容。本书可以作为大学经济管理类和文科类各专业本科生和研究生的教科书或参考书。授课教师可根据学生的具体情况对本书的内容进行适当取舍。

统计学课程在大学中越来越普及,会使用统计软件也逐渐成为大学生求

职的一项重要技能。考虑到实际应用需要，本书大部分章末都附加介绍了如何用统计软件 SPSS 实施该章提到的统计分析。书中的一些表格或图形也是由 SPSS 生成。在引用这些表格或图形时，我们尽量保持其格式与 SPSS 的输出一致，因而可能会出现其中一些名词与书中不符的情况，希望读者能够谅解。选择介绍 SPSS 是考虑到该软件是目前国内比较流行的统计软件，而且操作比较简单，适合非统计专业人士使用。对于一些比较适合使用 EXCEL 的内容，附录中也增加了相应的说明。

两位作者十分感谢高等教育出版社的马丽、李蕊、胡颖和蒋青编辑的大力支持以及高效率的工作。同时本书获得北京师范大学－香港浸会大学联合国际院校内的科研基金（代码 R201107）的支持，作者也感谢联合国际学院统计系老师张国秋、刘幼妹，学生莫莉莉、周妍、张飞龙、吕睿、刘少博、章恒璨、吴家驹、郑妍珣、郭莞等的帮助，书中部分插图由涂俊提供。本书有些数据和插图来自互联网，一些获得版权的允许，书中指明了出处，但有少数插图无法找到原始的来源，希望相关的作者给予原谅，或联系我们给予更正。因为我们的经验有限，请广大读者提出宝贵的建议。

两位作者特别感谢家人的理解和支持，没有她（他）们的支持，本书不可能顺利完成。

为方便教学，我们将本书的内容制作成 PPT 供大家参考；另外，也可提供书中所需的完整数据和部分习题的参考答案，有需要的读者可以发邮件索取。

方开泰，北京师范大学－香港浸会大学联合国际学院，珠海；

中国科学院应用数学研究所，北京，ktfang@uic.edu.hk

彭小令，北京师范大学－香港浸会大学联合国际学院，珠海，

xlpeng@uic.edu.hk

目录

第 0 章 什么是统计学	1
0.1 统计学的定义	2
0.2 统计学的几个基本概念	3
0.3 变量的类型	4
0.3.1 属性变量	4
0.3.2 数值型变量	5
0.4 基础统计学的主要内容	6
0.4.1 数据的收集与描述	6
0.4.2 统计推断	8
0.4.3 统计建模	10
0.5 统计软件	12
附录 0.1: 初识 SPSS	13
练习与讨论	15
第 1 章 抽样调查	17
1.1 什么是抽样调查	18
1.2 问卷调查的主要方式	20
1.3 抽样方法	21
1.3.1 概率抽样	22
1.3.2 非概率抽样	26
练习与讨论	28
第 2 章 数据的图形化描述	30
2.1 属性变量的图表描述	33
2.1.1 条形图	33
2.1.2 主次图	34
2.1.3 饼图	36
2.2 数值型变量的图表描述	37
2.2.1 茎叶图	37

2.2.2	频数分布表与直方图	39
2.3	二维数据的图表描述	41
2.3.1	两个属性变量分布的展示	42
2.3.2	散点图	44
2.3.3	混合类数据的图示	46
2.4	与时间有关的数据展示	48
2.5	空间数据的展示	50
附录 2.1:	如何用 SPSS 对数据进行图表描述	53
	练习与讨论	57
 第 3 章 描述数据分布的数值特征		62
3.1	对数据中心位置的描述	63
3.1.1	均值	63
3.1.2	中位数	64
3.1.3	众数	66
3.1.4	截尾平均	66
3.2	对数据波动程度的度量	67
3.2.1	极差	68
3.2.2	方差和标准差	68
3.2.3	变异系数	70
3.3	百分位数和箱线图	70
3.3.1	百分位数	71
3.3.2	四分位数与箱线图	73
3.4	加权平均	76
附录 3.1:	如何用 SPSS 计算描述数据分布的数值特征	79
	练习与讨论	81
 第 4 章 随机事件与概率		85
4.1	随机事件与概率	86
4.1.1	如何计算随机事件的概率	88
4.1.2	随机事件的运算及其图示	89
4.2	条件概率与事件的独立	91

4.3	全概率公式与贝叶斯公式	93
4.4	随机变量	96
4.4.1	离散型随机变量	98
4.4.2	连续型随机变量	101
4.5	随机模拟	105
	练习与讨论	109
第 5 章	几种重要的随机变量及其分布	112
5.1	二项分布	113
5.1.1	伯努利试验	113
5.1.2	二项分布	114
5.1.3	几何分布	118
5.2	泊松分布	120
5.3	正态分布	122
5.3.1	正态分布的定义	122
5.3.2	正态分布中的计算	125
5.3.3	正态性的验证	130
5.4	均匀分布	135
5.5	三个与正态分布相关的分布	136
5.5.1	卡方分布	136
5.5.2	t 分布	138
5.5.3	F 分布	140
	附录 5.1: 如何用 SPSS 验证数据的分布	141
	附录 5.2: 如何用 EXCEL 计算各分布中的概率	143
	练习与讨论	145
第 6 章	样本的统计推断	148
6.1	统计推断	149
6.2	样本均值的分布	152
6.3	总体均值的估计	156
6.3.1	总体均值的估计 (总体方差已知)	156
6.3.2	总体均值的估计 (总体方差未知)	159

6.4 总体比例的估计	162
6.5 样本量的确定	164
附录 6.1: 如何用 SPSS 计算均值的置信区间	166
练习与讨论	167
第 7 章 总体均值的比较	170
7.1 什么是假设检验	171
7.2 假设检验中的几个重要概念	175
7.2.1 两类假设	176
7.2.2 假设检验中的两类错误	177
7.2.3 p 值和假设检验的结论	180
7.2.4 假设检验与区间估计	182
7.3 总体方差未知时的假设检验 (t 检验)	183
7.4 两个总体均值的比较: 双样本假设检验	185
7.4.1 总体方差已知时的双样本检验 (Z 检验)	186
7.4.2 总体方差的比较 (方差齐次性检验)	188
7.4.3 总体方差未知时的双样本检验 (t 检验)	190
7.5 配对数据均值的比较	194
7.6 总体比例的比较 (Z 检验)	196
7.6.1 单总体比例的检验	196
7.6.2 两个总体比例的比较	197
7.7 多个总体均值的比较: 方差分析	199
7.7.1 方差分析	200
7.7.2 多重比较	206
附录 7.1: 用 SPSS 比较总体均值	209
练习与讨论	213
第 8 章 相关与回归分析	218
8.1 线性相关与相关系数	219
8.2 一元线性回归	223
8.2.1 一元线性回归模型	224
8.2.2 模型参数的估计: 最小二乘准则	225

8.2.3	对回归模型的评价	229
8.2.4	回归模型的诊断：残差分析	234
8.3	多元线性回归	237
8.3.1	多元线性回归模型	237
8.3.2	变量选择	246
8.4	其他类型的回归模型	250
8.4.1	可化为线性的回归模型	250
8.4.2	非线性最小二乘估计	253
附录 8.1：	用 SPSS 进行相关性和回归分析	256
	练习与讨论	259
第 9 章	拟合优度检验与列联表分析	266
9.1	拟合优度检验	267
9.2	属性变量之间的关联性检验	271
附录 9.1：	用 SPSS 进行卡方检验	276
	练习与讨论	279
第 10 章	不依赖于分布的统计推断	281
10.1	引言	282
10.2	符号检验	284
10.2.1	总体中位数的符号检验	284
10.2.2	匹配样本的符号检验	287
10.3	威尔科克森符号秩和检验	288
10.4	两个独立样本的秩和检验	291
10.4.1	威尔科克森秩和检验	292
10.4.2	曼 - 惠特尼 U 检验	295
10.5	比较多个总体的克鲁斯卡尔 - 沃利斯检验	296
10.6	斯皮尔曼秩相关系数	298
附录 10.1：	用 SPSS 进行非参数检验	300
	练习与讨论	304

第 11 章 时间序列分析	308
11.1 时间序列及其构成成分	309
11.2 平稳时间序列的预测	315
11.2.1 移动平均法	315
11.2.2 加权移动平均法	317
11.2.3 指数平滑法	318
11.3 时间序列的趋势预测	321
11.4 复合型时间序列的分解预测	323
附录 11.1: 用 EXCEL 计算移动平均和指数平滑	328
练习与讨论	329
附录	332
附录一 正态分布表	332
附录二 卡方分布表	334
附录三 t 分布表	336
附录四 F 分布表	337
附录五 威尔科克森符号秩和检验临界值表	341
附录六 威尔科克森秩和检验临界值表	342
统计词汇中英文对照表	344
参考文献	351

第0章 Chapter

0

什么是统计学

- 0.1 统计学的定义
 - 0.2 统计学的几个基本概念
 - 0.3 变量的类型
 - 0.3.1 属性变量
 - 0.3.2 数值型变量
 - 0.4 基础统计学的主要内容
 - 0.4.1 数据的收集与描述
 - 0.4.2 统计推断
 - 0.4.3 统计建模
 - 0.5 统计软件
-

在现代社会生活中，我们不断地和各种各样的数据打交道（图0.1）：一个班上学生的数学成绩，一天中各个时刻的温度，某大型企业员工的工资水平，不同类型汽车的耗油量，中国两大股市中股票的年收益率，等等。我们从这些数据中获取各种各样的信息，并通过分析这些信息来制定生活、生产中大大小小的决策。如人们根据温度的高低来决定穿衣，大学根据考生的高考成绩来决定是否录取，商家根据市场需求来决定进货量，基金公司根据股票的各项指标来决定要不要买入、何时买入某只股票，政府部门根据统计的社会经济数据来制定国家政策。如何从这些数据中获取有用的信息来做出科学有效的判断呢？统计学就是这样一门教我们如何同数据打交道，从中获取有用信息，并得出结论的学科。

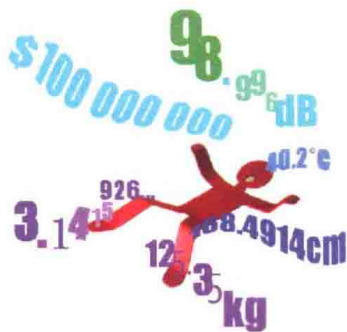


图 0.1

0.1 统计学的定义

什么是统计学？在《大不列颠百科全书》上，统计学的定义是：

统计学是用来收集和分析数据的一门科学和艺术。

统计学既是一门学问，也是一门艺术，更是一种看待分析事物的哲学观点。统计学经常简称为统计。用统计的思想来看世界，是用科学的方法来处理事件发生的随机性，并对随机性进行总结和归纳。用统计的思想来作决策，没有绝对的对与错，只有相对的适用与不适用。

具体说来，统计是一门关于数据的收集、整理、描述和分析的学科。

统计学早期被用于人口统计，中国是世界上最早进行统计调查的国家之一，最早的记录是周代《尚书·禹贡》，记录了当时土地、人口和财产的状况，被誉为国势统计学最早的萌芽。到了西周，统计工作进一步完善，距今已有3 000多年历史。而西方统计学于16世纪萌芽，17世纪形成。统计学的英文是“statistics”，与国家的英文“state”同一词根，也说明了统计学的出现与国家管理有关。

20世纪以来统计学逐渐变成了一门独立的学科，统计方法日益丰富。近代统计学的发展源于天文、农业、生物遗传的需要，例如统计中的回归分析（最小二乘法）就是

德国科学家高斯根据天文学的需要而创造的；孟德尔的豌豆试验开创了现代遗传学的先河，同时也大大促进了统计学的分支——试验设计的发展。随着现代科技的不断发展和计算机功能的日益强大，统计已经成为各类研究应用不可缺少的工具，其应用非常广泛，包括人口普查、政府部门统计、市场调查、风险管理、基因（DNA）研究、新药开发、生态监控、心理研究、海量数据的信息挖掘和大数据分析等。就连研究文学和历史都会用到统计，例如《红楼梦》前80回与后40回作者的鉴定。

0.2 统计学的几个基本概念

为进一步说明统计学的方法和应用，首先要介绍统计学的一些基本概念。为了便于理解，我们结合一个例子来说明。

例0.1 珠海是中国改革开放初期设置的四个经济特区之一，也是一个美丽的海滨城市（图0.2）。珠海政府或社会科学研究者若想了解珠海市居民的生活与工作状态（包括性别、年龄、婚姻状况、住房面积、年收入等），可以通过全面调查（又称普查）或抽样调查两种方式获取数据与信息。



图0.2

在调查研究中，所有被研究对象的集合（该例中为所有珠海市居民）称为**总体**，组成总体的各研究对象（该例中为每一个珠海市居民）称为**个体**。如果采用普查的方式，需要收集所有珠海市居民的相关信息，这对于一个拥有上百万人口的城市来说，并不是一件容易的事，要耗费大量的时间、人力和物力。因此，大部分的调查研究都是采用抽样调查的方式，被抽到的个体组成了一个**样本**，是总体的一个子集。被研究对象的特征或属性（如该例中的性别、年龄、婚姻状况、住房面积、年收入等）称**变量**。对个体进行测量得到的特征或属性的值称为**变量值**，也称为**观测值**，如某个珠海市女性居民今年35岁，对于该个体，变量“性别”的值是“女”，变量“年龄”的值是35。

当抽样是随机的，样本中包含的个体事先是不确定的，这时所得到变量值将依赖于抽样的结果。虽然变量的值事先不能确定，但有概率规律可循，这样的变量在统计学中称为**随机变量**。一个随机变量所有可能取值的集合构成了这个随机变量的**总体**。例如在