

“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

统计方法与 实验

张崇岐 李光辉

**Statistical Methods and
Experiments**

高等教育出版社

“十二五”国家重点图书出版规划项目



现代统计

统计方法与 实验

Statistical Methods and
Experiments

张崇岐 李光辉

高等教育出版社·北京

内容简介

本书共十二章,内容包括统计数据的表示,概率分布与随机抽样,描述性统计,函数结构与随机数,参数估计,区间估计,假设检验,方差分析,相关分析,回归分析,非参数统计简介,试验设计。本书以统计理论为基础,系统阐述统计方法的基本原理下,结合 R 软件操作,给出相应解决问题的步骤,有效帮助读者理解统计方法的思想精髓及应用技巧。

本书可作为高等学校统计学专业学生学习统计方法及软件应用的教材,也可作为理工、经济、管理、生物等专业学生数理统计课程的辅导教材或教学参考书。

图书在版编目(CIP)数据

统计方法与实验/张崇岐,李光辉编.--北京:
高等教育出版社,2015.9

ISBN 978-7-04-043098-1

I. ①统… II. ①张… ②李… III. ①统计学-高等学校-教材 IV. ①C8

中国版本图书馆 CIP 数据核字(2015)第 139220 号

策划编辑	张晓丽	责任编辑	张晓丽	封面设计	赵 阳	版式设计	余 杨 张 杰
插图绘制	杜晓丹	责任校对	陈 杨	责任印制	尤 静		

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 北京宏信印刷厂
开 本 787mm×960mm 1/16
印 张 24.75
字 数 450 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 2015 年 9 月第 1 版
印 次 2015 年 9 月第 1 次印刷
定 价 38.80 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 43098-00

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

统计学是一门收集、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,适用于统计专业和非统计专业的系列教材。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,一个面对统计专业,另一个面对非统计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

随着社会、经济和科学技术的发展,统计学在现代化管理和社会生活中的地位日益重要,人们的日常生活和一切社会活动都离不开统计。所以高等学校对学生的培养目标中,不论什么专业,都需要对统计学有所了解,掌握统计的一些基本方法,有的专业还需要学生能够使用统计软件包对数据进行处理和分析。

鉴于此,本书较为系统地介绍了统计学的一些基本原理和方法及其应用,而且例题大多来源于现实生活,方便学生理解和掌握相关知识。本书中所介绍的每一种统计方法都结合 R 软件,给出相应的解决问题的步骤,并对计算结果进行分析。在编排上本书是以介绍统计方法与理论为主,R 软件的实际操作为辅,在需要用到 R 软件的某项功能时才予以介绍,而不是开篇就将 R 软件各种功能解释殆尽。这样能使得读者在学习各种统计方法的过程中,逐步掌握 R 软件的实际操作。

本书共分 12 章,涵盖了统计学中的主要原理和方法。第 1 章简单介绍了统计学的主要研究内容以及 R 软件中向量和矩阵的运算,并介绍了一种在 R 软件中导入数据的方法,这样方便后续学习中对繁杂数据的输入。第 2 章作为概率统计的基础,介绍了常见的概率分布,并学习如何绘制各种分布的概率密度函数图和分布函数图。并在附录部分列出了常见的 R 绘图函数,以方便读者自学和操作。第 3 章介绍了描述性统计的基础,包括常见的描述性统计量、统计图以及对定性变量的描述统计。第 4 章介绍了一些常见的概率分布的构造,包括次序统计量的分布,混合分布以及一些其他的分布,以 R 软件中函数和控制流为基础,生成不同分布族下的随机数,并介绍了随机数的应用。第 5 章至第 7 章分别介绍了统计学的三大问题,即参数估计、区间估计和假设检验,这三章都省略了各种理论的推导和证明,相应地给出实际计算的程序,读者也可根据实际情况,对本书中所给出的程序做适当修改或者重新编写程序,以适应各种特殊的情况。第 8 章介绍了方差分析的基本原理与方法,这一章是作为后续学习的基础,特别在回归分析与试验设计中都有重要的应用,所以建议读者应充分理解本章的方法与步骤。第 9 章介绍了相关分析,包括单相关分析和多元相关分析、典型相关分析。第 10 章介绍了回归分析,包括了一元回归、多元回归、逐步回归和一些常见的非

线性回归等,本章内容是统计学中的一个重要的分支,回归分析的方法理论丰富,应用广泛,并且在 R 上易于操作。第 11 章介绍了非参数统计的一些基本概念和方法,并提供了一些常见的非参数检验的程序,在非参数回归部分也提供了相应的核回归程序。第 12 章介绍了三种常见的试验设计的方法:正交试验设计、均匀试验设计以及混料试验设计,并在附录部分列出了常见的正交试验设计表和均匀试验设计表。

本书省略了各类理论的推导和证明,但这些理论的证明与推导在一般教材中都能找到。有兴趣深入了解的读者可以参阅相关书籍和文献。

由于编者水平有限,书中难免有错漏以及欠妥之处,热诚地希望使用本书的读者指正,以便再版时修改。

张崇岐 广州大学经济与统计学院, cqzhang@gzhu.edu.cn

李光辉 凯里学院数学科学学院, liguanghui1985@126.com

郑重声明

高等教育出版社依法对本书享有专有出版权。任何未经许可的复制、销售行为均违反《中华人民共和国著作权法》，其为人将承担相应的民事责任和行政责任；构成犯罪的，将被依法追究刑事责任。为了维护市场秩序，保护读者的合法权益，避免读者误用盗版书造成不良后果，我社将配合行政执法部门和司法机关对违法犯罪的单位和个人进行严厉打击。社会各界人士如发现上述侵权行为，希望及时举报，本社将奖励举报有功人员。

反盗版举报电话 (010)58581897 58582371 58581879

反盗版举报传真 (010)82086060

反盗版举报邮箱 dd@hep.com.cn

通信地址 北京市西城区德外大街4号 高等教育出版社法务部

邮政编码 100120

目 录

第 1 章	统计数据的表示	1
1.1	统计数据的分类及特征	1
1.1.1	概述	1
1.1.2	统计数据的分类	3
1.1.3	统计数据的数字特征	3
1.1.4	R 语言简介	5
1.2	单变量数据	7
1.2.1	基本运算与常用函数	7
1.2.2	向量赋值	9
1.2.3	产生有规律的数列	9
1.2.4	向量运算	10
1.3	多维数组和矩阵	11
1.3.1	创建矩阵	11
1.3.2	矩阵的运算	12
1.3.3	矩阵的分解	19
1.3.4	矩阵的行、列之和与平均	23
1.3.5	矩阵合并	24
1.4	R 中读取 Excel 文件	25
习题 1		27
第 2 章	概率分布与随机抽样	28
2.1	随机变量与 R 中的内嵌分布	28
2.1.1	随机变量	28
2.1.2	常见分布函数的 R 表示	31
2.1.3	plot() 函数	32
2.2	常见离散型分布	34
2.2.1	二项分布	34
2.2.2	多项分布	37
2.2.3	几何分布	38

2.2.4	负二项分布	39
2.2.5	超几何分布	40
2.2.6	泊松分布	42
2.3	常见连续型分布	43
2.3.1	均匀分布	43
2.3.2	正态分布	44
2.3.3	指数分布	46
2.3.4	Γ 分布	48
2.3.5	χ^2 分布	50
2.3.6	β 分布	51
2.3.7	t 分布	52
2.3.8	F 分布	53
2.3.9	韦布尔分布	55
2.4	随机抽样	56
习题 2		59
第 3 章	描述性统计	60
3.1	描述性统计量	60
3.1.1	样本的基本特征	60
3.1.2	分散程度的度量	62
3.1.3	分布形状的度量	63
3.2	数据的标准化处理	67
3.3	频率分布与统计图	69
3.3.1	五数概括与箱线图	69
3.3.2	茎叶图	70
3.3.3	直方图	71
3.3.4	经验分布函数图	74
3.3.5	条形图	75
3.3.6	饼图	78
3.4	定性变量的描述性统计	79
3.4.1	单因素分析	80
3.4.2	两因素分析	81
3.4.3	三因素分析	82
习题 3		84

第 4 章	函数结构与随机数	85
4.1	构造概率分布	85
4.1.1	R 函数定义	85
4.1.2	次序统计量的分布	86
4.1.3	最大值和最小值分布	89
4.1.4	混合分布	90
4.1.5	其他分布	91
4.2	随机数的产生及应用	92
4.2.1	随机数的产生	92
4.2.2	随机数的应用 (Monte-Carlo 方法)	93
4.3	程序控制结构	98
4.3.1	if 结构	98
4.3.2	for 循环	100
4.3.3	while 循环	102
4.3.4	repeat 循环	103
习题 4		105
第 5 章	参数估计	106
5.1	统计量及其抽样分布	106
5.1.1	常用统计量	106
5.1.2	正态总体下的抽样分布	107
5.1.3	指数分布总体下的抽样分布	109
5.1.4	Γ 分布总体下的抽样分布	109
5.1.5	均匀分布下的抽样分布	109
5.2	总体参数的点估计	110
5.2.1	矩估计	110
5.2.2	最大似然估计	112
5.3	点估计的评价标准	118
习题 5		120
第 6 章	区间估计	121
6.1	枢轴量	121
6.2	正态总体参数的区间估计	122
6.2.1	单正态总体均值的区间估计	122

6.2.2	单正态总体方差的区间估计	126
6.2.3	两正态总体均值之差的区间估计	127
6.2.4	两正态总体方差之比的区间估计	129
6.3	指数分布下的区间估计	131
6.3.1	完全样本情形	131
6.3.2	截尾样本情形	132
6.4	非正态总体均值的大样本区间估计	134
6.4.1	单个非正态总体的均值的置信区间	134
6.4.2	两个非正态总体的均值之差的置信区间	137
6.5	样本容量的确定	138
6.5.1	总体均值估计的必要样本容量	139
6.5.2	总体成数估计的必要样本容量	140
习题 6		140
第 7 章	假设检验	142
7.1	假设检验的基本思想和概念	142
7.1.1	假设检验的基本思想	142
7.1.2	拒绝域	143
7.1.3	两类错误	144
7.2	单样本正态总体的假设检验	146
7.2.1	总体均值的假设检验	146
7.2.2	总体方差的假设检验	151
7.3	两个样本平均数差异的假设检验	154
7.3.1	两独立样本平均数差异的假设检验	154
7.3.2	两配对样本平均数的差异显著性检验	156
7.4	样本频率的假设检验	159
7.4.1	单个样本频率的假设检验	159
7.4.2	两个样本频率的假设检验	161
7.5	改进的 QQ 图	163
7.5.1	正态 QQ 图	163
7.5.2	非正态分布的 QQ 图	164
习题 7		168
第 8 章	方差分析	169
8.1	因子与数据框	169

8.1.1	因子	170
8.1.2	数据框	172
8.2	单因素方差分析	174
8.2.1	数学模型	174
8.2.2	方差齐性检验	179
8.2.3	均值多重比较	181
8.3	双因素方差分析	184
8.3.1	无交互影响的双因素方差分析	184
8.3.2	有交互影响的双因素方差分析	188
习题 8		193
第 9 章	相关分析	195
9.1	单相关分析	195
9.1.1	相关分析的基本概念	195
9.1.2	相关系数	197
9.1.3	相关性检验	199
9.2	多元相关分析	200
9.2.1	相关系数矩阵	200
9.2.2	相关系数矩阵的检验	204
9.3	典型相关分析	205
9.3.1	典型相关分析的原理及方法	205
9.3.2	样本典型相关分析	209
9.3.3	典型相关检验	212
习题 9		215
第 10 章	回归分析	216
10.1	一元线性回归	216
10.1.1	一元线性回归模型	216
10.1.2	回归系数的置信区间	219
10.1.3	一元线性回归模型的统计检验	219
10.1.4	一元线性回归的预测	223
10.2	多元线性回归	224
10.2.1	多元线性回归方程及基本假定	225
10.2.2	参数的最小二乘估计及其性质	226
10.2.3	总体回归方程的显著性检验	231

10.2.4	多元线性回归的预测	235
10.3	自变量的选择与逐步回归	236
10.3.1	自变量选择对估计和预测的影响	236
10.3.2	自变量的选择准则	237
10.3.3	逐步回归	241
10.4	曲线回归与残差分析	246
10.4.1	曲线回归	246
10.4.2	残差分析	253
10.5	岭回归	258
10.5.1	复共线性	258
10.5.2	岭估计及其性质	262
10.5.3	岭迹分析与岭参数选择	264
10.5.4	广义岭回归	268
习题 10		268
第 11 章	非参数统计简介	271
11.1	秩统计量	271
11.2	位置参数的符号检验	273
11.2.1	单样本符号检验	273
11.2.2	配对样本的符号检验	275
11.2.3	单样本 Wilcoxon 符号秩检验	276
11.2.4	两样本配对 Wilcoxon 检验	279
11.2.5	非配对样本秩和检验	280
11.3	总体分布函数的检验	281
11.3.1	Pearson 拟合优度检验	281
11.3.2	双样本分布的检验	288
11.3.3	Kolmogorov-Smirnov 检验	291
11.4	非参数回归简介	293
11.4.1	非参数回归模型与核函数	293
11.4.2	核估计方法	294
11.4.3	常用非参数回归函数包	298
11.4.4	核估计的大样本性质	299
习题 11		300

第 12 章 试验设计	301
12.1 试验设计的基本概念	301
12.2 正交试验设计	303
12.2.1 正交试验设计的基本原理	304
12.2.2 正交试验设计的基本步骤	306
12.2.3 正交试验设计的方差分析	312
12.3 均匀试验设计	314
12.3.1 均匀试验设计表	315
12.3.2 均匀试验设计的步骤与结果分析	318
12.4 混料试验设计	323
12.4.1 混料规范多项式	323
12.4.2 单纯形格子设计	324
12.4.3 单纯形中心设计	335
习题 12	342
 附录 1 常用正交试验设计表	 343
 附录 2 常用均匀试验设计表	 354
 附录 3 绘图函数命令	 371
 参考文献	 376

统计学是通过搜集、整理、分析数据等手段,以达到推断所测对象的本质,甚至预测对象未来的一门综合性科学.统计学着重于获取准确数据并对数据进行深层次的分析,从而得到一定的科学结论.其中,统计数据是统计推断与分析的基础,没有数据的统计推断是毫无意义的.所以如何准确、全面地表示统计数据是有效反映各类随机现象的关键.

1.1 统计数据的分类及特征

1.1.1 概述

由于人类的统计实践是随着计数活动而产生的,因此,统计发展史可以追溯到远古的原始社会,也就是距今有 5000 多年的漫长岁月.但是,能使人类的统计实践上升到理论上予以概括总结的程度,即开始成为一门系统的科学统计学,却是近代的事情,距今只有 300 余年的短暂历史.

中国早在周朝就设有 18 位“统计官员”——司书.到了汉代统计工作发展已经较为完善,皇帝十分注重统计数据的搜集与整理.刘邦占领秦都咸阳时,萧何首先收取秦的律令、图书,刘邦得知天下的地形,户口的多寡、强弱,以及人民的疾苦等.汉代继承了秦代的统计组织,基本上仍由三公九卿兼管统计工作,并开始由中央到地方形成了综合统计系统和业务统计系统.汉代对违反统计制度的制裁也是非常严重的,例如上计制度就是当时中央控制全国经济收支的重要制度.从西汉到东汉 400 年间,没有中断过.皇帝特别看重它,把它上升到法律高度,列入法律条文中,称为“上计律”.“上计律”规定上计的时间、报告的程序和内容,凡不按时上计,“上计不实”等情事,都要受到法律制裁.所以汉代的统计数据是相当准确的.

随着历史的发展,统计数据不断膨胀,统计方法也得到进一步的发展.到了唐代,分组法得到了广泛的应用.例如人口年龄的分组始于武德七年(624 年)令:“凡男女始生为黄,四岁为小,十六为中,二十一为丁,六十为老”.民户分组按是否本地人户为标志,把民户分为主户(或称“正户”“土户”)与客