

“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

抽样调查

金勇进 编著

Sample Survey

高等教育出版社

“十二五”国家

现代

抽样调查

Sample Survey

金勇进 编著

Chouyang Diaocho

高等教育出版社·北京

内容简介

本书是为非统计学专业学生编写的抽样调查课程教材,主要讲述抽样调查的基本理论与方法。全书共分十一章,内容包括引论,抽样调查基础,简单随机抽样,分层随机抽样,比率估计,系统抽样,整群抽样,多阶段抽样,抽样推断的有关问题,问卷设计,调查的组织与管理。本书内容通俗易懂,侧重应用,略去了比较复杂的公式证明和推导,着重介绍不同抽样方法的特点和应用背景,对数学要求不高。每章后配有习题,书后附部分习题参考答案。

本书既可作为高等学校非统计学专业本科生和研究生抽样课程的教材,也可作为其他相关人员、实际工作者了解抽样方法、学习抽样理论的入门参考书。

图书在版编目(CIP)数据

抽样调查 / 金勇进编著. — 北京:高等教育出版社, 2015.5

ISBN 978 - 7 - 04 - 042410 - 2

I. ①抽… II. ①金… III. ①抽样调查统计 - 高等学校 - 教材 IV. ①C811

中国版本图书馆 CIP 数据核字 (2015) 第 067411 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

插图绘制 郝林

责任校对 刘娟娟

责任印制 张泽业

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 三河市潮河印业有限公司
开本 787 mm × 960 mm 1/16
印张 16.5
字数 300千字
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5月第1版
印次 2015年5月第1次印刷
定价 29.40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 42410 - 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”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

当我接受“现代统计学系列丛书”编委会的委托,为非统计学专业的学生写一本有关抽样的教科书时,就思考这本书的定位。我在高校任职近 30 年,讲授抽样课程的历史也有 20 年。教学过程中我做过不同尝试,包括讲授内容侧重点的不同,以及采用授课方式的不同。虽然上过我课的学生们有不同的专业背景,学习这门课程的动机也不尽相同,但还是存在一般的规律性。当进行公式的演绎、定理的证明时,盯着看你的学生不多,做其他事情的学生不少;而当进行案例的介绍,进行不同抽样方法适用性的研讨时,课堂气氛就十分活跃。抽样有一套科学完整的理论体系,同时又是一门实践性很强的应用性技术。对于不是专门从事抽样理论研究的学生而言,实用性的抽样技术是他们更喜欢的。

本书的名称为“抽样调查”,实际已经确定了书的定位,书中论及的不是抽样体系的完整理论,而是把抽样看成一项实践活动,介绍从事这项活动中所涉及的主要技术方法。基于这样的认识,本书有如下特点:

1. 立足基本,不追求抽样方法的全面与尖端。抽样有不同的方法,抽样课程也是讲这些不同的方法。本书的特点是介绍最基本、最常用的抽样方法,不追求抽样方法的全面与尖端。实践中经常采用的抽样方法,如简单随机抽样、分层随机抽样、系统抽样、整群抽样、多阶段抽样以及使用频率很高的比率估计方法,都在本书中得到展示。还有一些相对复杂的抽样方法和估计方法,如不等概抽样、二重抽样、回归估计、复杂样本方差估计等,都予以略去。由于篇幅限制,加之考虑到使用者的专业背景,对多种抽样方法和估计方法有必要进行筛选。

2. 本书增加了应用性较强的内容。与其他抽样教科书相比,本书增加了一些实用性内容,当把抽样调查视为一项实践活动时,这些问题通常都会遇到。例如,如何预防和处理调查过程中的缺失数据,如何考虑权数问题,怎样对特殊群体的单位数进行估计,对敏感性问题调查可以采用什么方法。此外,调查需要进行问卷设计,怎样设计一份好的调查问卷;为了保证采集数据的质量,也需要对调查的实施过程进行有效的组织与管理,这些问题在本书中都有不同程度的反映。

3. 内容通俗,侧重应用。本书的特色是将复杂的理论简单化,提取抽样理论中最核心的内容,用相对简单的方式陈述,着重论述不同抽样方法应用中的特

点,不同方法背后体现的统计思想,略去了复杂的公式推导和定理证明。在对不同抽样方法进行介绍时,列举大量的例题,通过对例题的解析,说明方法的应用。每章最后都给出了一定数量的习题或思考题,以利于读者对相关知识点的把握。对习题也给出参考答案,便于读者对照检查。

本书的完成得益于我的学生的帮助。参与本书编写的学生有:宋丹,王俊,朱新景,张婉璐,唐箫穗。他们在我拟写的书稿提纲和知识要点的基础上,结合我写的教材和课件,参考其他相关文献,提出了本书大部分内容的初稿;设计了习题,提供了习题参考答案,并对书中的公式进行了核阅。在此,我要向他们表示深深的谢意!同时,还要感谢审阅本书的专家,感谢高等教育出版社张晓丽编辑,本书中也包含着他们辛勤劳动的汗水。

为非统计学专业学生编写抽样理论与方法的教科书,以前还未曾有过,本书算作一个尝试,欢迎读者提出宝贵意见。

金勇进

2014年8月

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

第一章 引论	1
1.1 调查与抽样调查	1
1.1.1 调查	1
1.1.2 抽样调查	2
1.1.3 抽样调查的作用	6
1.1.4 抽样调查与普查	7
1.2 抽样调查步骤	7
1.2.1 调查策划	8
1.2.2 抽样方法选择	8
1.2.3 问卷设计	9
1.2.4 调查实施	9
1.2.5 数据录入	10
1.2.6 数据审核与整理	10
1.2.7 数据分析	11
1.2.8 数据发布	11
1.3 基本的抽样方法	12
1.3.1 简单随机抽样	12
1.3.2 分层抽样	13
1.3.3 系统抽样	13
1.3.4 整群抽样	13
1.3.5 多阶段抽样	14
1.4 数据采集方法	14
1.4.1 自填式	14
1.4.2 计算机辅助方式	15
1.4.3 面访式	16
1.4.4 网络调查	16
1.4.5 直接观测	17
1.4.6 方法的比较	18

总结	19
习题一	19
第二章 抽样调查基础	21
2.1 基本概念	21
2.1.1 目标总体与抽样总体	21
2.1.2 抽样框与抽样单元	22
2.1.3 总体参数与统计量	23
2.1.4 抽样误差和非抽样误差	25
2.1.5 总体方差,样本方差,估计量方差	26
2.2 抽样分布	28
2.3 两个极限定理	30
2.3.1 大数定理	30
2.3.2 中心极限定理	31
2.4 调查中的统计推断	31
总结	32
习题二	33
第三章 简单随机抽样	35
3.1 引言	35
3.1.1 定义	35
3.1.2 样本性质	36
3.1.3 抽样方法	37
3.1.4 符号说明	38
3.2 估计量	39
3.2.1 总体均值估计	39
3.2.2 总体总量估计	41
3.2.3 总体比例估计	41
3.2.4 置信区间	42
3.3 有放回抽样的估计	44
3.4 样本量	46
3.4.1 费用因素	46
3.4.2 精度要求	47
3.4.3 影响样本量的其他因素	49
3.4.4 总体方差的预估计	50
总结	51

习题三	52
第四章 分层随机抽样	54
4.1 引言	54
4.1.1 定义	54
4.1.2 应用特点	54
4.1.3 如何分层	56
4.1.4 符号说明	56
4.2 估计量	57
4.2.1 总体均值估计	57
4.2.2 总体总量估计	59
4.2.3 总体比例估计	61
4.2.4 置信区间	62
4.3 样本量	63
4.3.1 总样本量确定	63
4.3.2 样本量在各层分配	65
4.3.3 某些层大于 100% 抽样的情况	68
4.4 其他分层技术	69
4.4.1 目录抽样	69
4.4.2 事后分层	70
4.4.3 层的划分	72
4.5 效率分析	74
4.5.1 设计效应	74
4.5.2 比例分配与奈曼分配的比较	76
总结	77
习题四	77
第五章 比率估计	80
5.1 引言	80
5.1.1 定义	80
5.1.2 符号说明	81
5.2 估计量	81
5.2.1 总体比率估计	81
5.2.2 总体均值估计	84
5.2.3 总体总量估计	86
5.2.4 置信区间	88

5.3 样本量	90
5.3.1 比率估计样本量	90
5.3.2 均值估计样本量	91
5.3.3 总量估计样本量	92
5.4 比率估计应用条件	95
5.4.1 比率估计是有偏估计,要求样本量较大	95
5.4.2 研究变量与辅助变量之间有较高相关性	96
总结	97
习题五	98
第六章 系统抽样	102
6.1 引言	102
6.1.1 定义	102
6.1.2 系统抽样特点	103
6.1.3 排列规则与影响	103
6.1.4 不同系统抽样方法	104
6.2 估计量	106
6.2.1 总体均值及其估计	106
6.2.2 按无关标志排列时	109
6.2.3 按有关标志排列	111
6.3 有周期波动的系统抽样	112
6.4 效率分析	114
6.4.1 与简单随机抽样相比	114
6.4.2 与分层抽样相比	115
总结	115
习题六	116
第七章 整群抽样	120
7.1 引言	120
7.1.1 定义	120
7.1.2 应用特点	121
7.1.3 群的划分	122
7.1.4 整群抽样与分层抽样和系统抽样的关系	123
7.1.5 符号说明	124
7.2 群规模相等的情形	125
7.2.1 估计量	125

7.2.2 群内相关系数	128
7.2.3 整群抽样设计效应	130
7.3 群规模不等的情形	131
7.3.1 等概抽样,加权估计	131
7.3.2 等概抽样,比率估计	132
7.3.3 不等概抽样	135
7.4 群规模不等时的方法比较	138
总结	140
习题七	140
第八章 多阶段抽样	145
8.1 引言	145
8.1.1 定义	145
8.1.2 应用特点	146
8.1.3 推断原理	147
8.1.4 符号说明	147
8.2 初级单元规模相等的情形	148
8.2.1 总体均值和总体总量的估计	148
8.2.2 总体比例估计	150
8.3 初级单元规模不等的情形	152
8.3.1 等概率抽取初级单元	152
8.3.2 不等概率抽取初级单元	156
8.3.3 总体比例的估计	159
8.4 自加权设计	163
8.4.1 自加权含义	163
8.4.2 多阶段抽样中的自加权设计	163
8.5 样本的最优分配	164
总结	167
习题八	167
第九章 抽样推断的有关问题	172
9.1 缺失数据问题	172
9.1.1 缺失数据的产生和影响	172
9.1.2 降低无回答的措施与方法	173
9.1.3 处理缺失数据的权数调整	174
9.1.4 处理缺失数据的插补法	176

9.2 权数问题	178
9.2.1 权数定义与作用	178
9.2.2 权数计算步骤	179
9.2.3 多变量的权数调整	180
9.2.4 权数的影响与控制	182
9.3 对总体单元个数的估计	183
9.3.1 问题的提出	183
9.3.2 估计总体单元数 N	183
9.3.3 逆抽样	185
9.4 敏感问题调查	186
9.4.1 随机化回答技术	186
9.4.2 沃纳模型	186
9.4.3 西蒙斯模型	188
9.4.4 (采用随机化回答技术)需要注意的问题	189
总结	190
习题九	190
第十章 问卷设计	192
10.1 问卷简介	192
10.1.1 问卷的作用	192
10.1.2 问卷调查方式	194
10.1.3 问卷的类型	195
10.1.4 问卷的结构	196
10.1.5 问卷的设计流程	199
10.2 问卷设计	202
10.2.1 问题类型	202
10.2.2 封闭式问题的题型	203
10.2.3 问题的措辞	205
10.2.4 问题的顺序	209
10.3 量表	210
10.3.1 量表	210
10.3.2 常用的量表	213
10.4 问卷中的误差	216
10.4.1 问卷中的误差及其来源	216
10.4.2 误差的识别与控制	218

习题十	221
第十一章 调查的组织与管理	222
11.1 调查实施前的准备	222
11.1.1 组织准备	222
11.1.2 公共关系准备	224
11.1.3 文档准备	227
11.1.4 访员招募培训	230
11.2 数据收集	231
11.2.1 访员工作准备	232
11.2.2 说好开场白	232
11.2.3 提问	234
11.2.4 追问	235
11.2.5 复杂情形的问题处理	235
11.3 数据收集过程中的质量控制	236
11.3.1 数据收集过程的质量控制方法	236
11.3.2 访员作业管理	238
11.3.3 督导监控	238
11.3.4 数据收集过程的质量度量	240
习题十一	240
部分习题参考答案	242

本章的目的是向读者提供有关抽样调查的总体描述,使读者在学习具体的抽样方法之前对抽样调查活动有一个梗概的了解.本章首先从调查开篇,继而扩展到抽样调查,重点讨论非概率抽样和概率抽样的应用.在这一章中还将对抽样调查的整个过程进行描述,从最开始的调查目标的确定到最后环节的数据发布.本章还将对基本的抽样方式进行简要介绍,这些抽样方式都是在本书后面的篇幅中要比较详细地展开讨论的.最后,本章还将介绍采集数据的不同方法.总之,希望读者在阅读本章之后对调查活动能有一个总体的印象.

1.1 调查与抽样调查

1.1.1 调查

调查是获取信息的方式和过程.信息可以有多种,信息的类型或对信息的兴趣点不同,对调查的定义就可能不同.本书关注的是数据信息,所以这里把调查定义为:通过使用明确的概念、方法和程序,依据专门设计的调查方案指导的方式,从一个总体全部或部分单元中搜集感兴趣的指标信息,并将这些信息综合整理成数据系列的有关活动.

上面的定义表述了调查的两个重要特征:首先,调查是一种有策划、有方法、有程序的活动;其次,调查的结果表现为搜集到的数据.需要指出的是,调查并不是获取数据信息的唯一方式.

数据有两种类型,试验数据和调查数据.搜集数据的一种方法是通过试验,是指在试验方案的指导下,通过控制一个或多个变量,在有控制的条件下得到观测的结果.所以试验数据是在试验中控制试验对象,依据试验设计的指导完成的.在可控条件下得到的数据是试验数据的重要特征.例如,检验一个小麦的新品种是否具有更高的生产率,试验人员将不同的小麦品种按某种设计方式分种