

外国优秀统计学教材系列丛书 · 影印版

作者：
Sharon L.Lohr

全国统计教材编审委员会组织引进

Sampling: Design and Analysis

抽样：设计与分析

中国统计出版社
China Statistics Press





抽样：设计与分析

Sampling: Design and Analysis

Sharon L. Lohr

Arizona State University



英文版
附光盘



中国统计出版社
China Statistics Press

Sampling : design and analysis / Sharon L. Lohr


(ISBN 0-534-35361-4)

Original Edition Copyright © 1999 by Duxbury Press, a division of Thomson Learning. All Rights Reserved.

Authorized Reprint Edition by Thomson Learning and China Statistics Press. No part of this book may be reproduced in any form without the express written permission of Thomson Learning and China Statistics Press.

This edition is only for sale in the People's Republic of China (excluding Hong Kong, Macau SARs and Taiwan).

981-243-597-2

本书英文影印由汤姆森学习出版集团  授权中国统计出版社独家出版。未经出版者书面许可，不得以任何方式翻印、拷贝、仿制或转载。

此版本仅限在中华人民共和国境内（不包括香港、澳门特别行政区及台湾）销售版权所有，侵权必究。

北京市版权局著作权合同登记号：图字：01-2002-6312

(京)新登字 041 号

图书在版编目(CIP)数据

抽样：设计与分析

/(美)洛尔(Sharon L. Lohr)著. - 影印本

北京：中国统计出版社，2002.12

ISBN 7-5037-3953-3

I. 抽… II. 洛… III. 抽样调查-教材-英文 IV. C811

中国版本图书馆CIP数据核字(2002)第086854号

责任编辑 / 刘国宁

封面设计 / 刘国宁 张建民

E-mail / cbsebs@stats.gov.cn

出版发行 / 中国统计出版社

通信地址 / 北京市西城区月坛南街75号

邮政编码 / 100826

办公地址 / 北京市丰台区西三环南路甲6号

电 话 / (010) 63459084 63266600-22500(发行部)

印 刷 / 北京市顺义兴华印刷厂

经 销 / 新华书店

开 本 / 850 × 1092 毫米 1/16

字 数 / 950 千字

印 张 / 32.25

印 数 / 1-3000 册

版 别 / 2002 年 12 月第 1 版

版 次 / 2002 年 12 月第 1 次印刷

书 号 / ISBN 7-5037-3953-3/C. 2011

定 价 / 60.00 元

中国统计版图书，如有印装错误，本社发行部负责调换。

引进外国优秀统计学教材专家委员会

主任：贺 铿 国家统计局副局长

副主任：王吉利 国家统计局统计教育中心主任
谢鸿光 中国统计出版社社长

委员：（按姓氏笔画排序）

文兼武 国家统计局统计科学研究所 所长
孙山洋 北京大学数学科学学院概率统计系 教授
严建辉 中国统计出版社 副总编
吴喜之 中国人民大学统计学系 教授
何书元 北京大学数学科学学院概率统计系 教授
张尧庭 上海财经大学经济学院 教授
张润楚 南开大学数学学院统计学系 教授
邱 东 东北财经大学 教授
陈 江 美国雪城大学管理学院 教授
陈家鼎 北京大学数学科学学院概率统计系 教授
杨振海 北京工业大学应用数理学院 教授
孟晓犁 哈佛大学统计系 教授
林正炎 浙江大学理学院数学系 教授
茆诗松 华东师范大学统计系 教授
郑 明 复旦大学管理学院统计系 副教授
柯惠新 北京广播学院 教授
徐国祥 上海财经大学统计学系 教授
袁 卫 中国人民大学 教授
蒋 萍 东北财经大学统计系 教授
谢邦昌 台湾辅仁大学统计资讯学系 教授
濮晓龙 华东师范大学统计系 副教授

办公室：

温 明 国家统计局统计教育中心教材处处长
刘国宁 中国统计出版社第二编辑部主任
孙洪娟 国家统计局统计教育中心教材处统计师
吕 军 中国统计出版社第二编辑部编辑

出版说明

21

世纪的竞争是人才的竞争，是全球性人才培养机制的较量。

如何培养面向现代化、面向世界、面向未来的高素质的人才成为我国人才培养的当务之急。为此，教育部发出通知，倡导在全国普通高等学校中使用原版外国教材，进行双语教学，培养适应经济全球化的人才。

为了响应教育部的号召，促进统计教材的改革，培养既懂统计专业知识又具备较高英语语言能力的统计人才，全国统计教材编审委员会在国家统计局领导的大力支持下，组织引进了这套“外国优秀统计学教材”。

为了做好“外国优秀统计学教材”引进工作，全国统计教材编审委员会将其列入了“十五”规划，并成立了由海内外统计学家组成的专家委员会。在对国外统计学教材的使用情况进行了充分了解，对国内高等院校使用外国统计学教材的需求情况进行了仔细分析，并对从各种渠道推荐来的统计教材进行了认真审定的基础上，制定了引进教材书目。在确定引进教材书目的过程中，我们得到了国内外有关专家、有关院校和外国出版公司及其北京办事处的支持和帮助，在此致谢。中国人民大学统计学系的吴喜之教授不仅推荐了大量的优秀候选书目，而且校译了影印教材的翻译目录，为这套教材的及早出版作了大量的工作，我们表示衷心的感谢。

这套引进教材多数是国外再版多次、反响良好，又比较适合国内情况、易于教学的统计教材。我们希望这套引进教材的出版对促进我国统计教材的改革和高校统计学专业双语教学的发展能够起到重要的推动作用。

全国统计教材编审委员会

2002年8月28日

Preface

Surveys and samples sometimes seem to surround you. Many give valuable information; some, unfortunately, are so poorly conceived and implemented that it would be better for science and society if they were simply not done. This book gives you guidance on how to tell when a sample is valid or not, and how to design and analyze many different forms of sample surveys.

The book concentrates on the statistical aspects of taking and analyzing a sample. How to design and pretest a questionnaire, construct a sampling frame, and train field investigators are all important issues, but are not treated comprehensively in this book.

I have written the book to be accessible to a wide audience, and to allow flexibility in choosing topics to be read. To read most of Chapters 1 through 6, you need to be familiar with basic ideas of expectation, sampling distributions, confidence intervals, and linear regression—material covered in most introductory statistics classes. These chapters cover the basic sampling designs of simple random sampling, stratification, and cluster sampling with equal and unequal probabilities of selection. The optional sections on the statistical theory for these designs are marked with asterisks—these sections require you to be familiar with calculus or mathematical statistics. Appendix B gives a review of probability concepts used in the theory of probability sampling.

Chapters 7 through 12 discuss issues not found in many other sampling textbooks: how to analyze complex surveys such as those administered by the United States Bureau of the Census or by Statistics Canada, different approaches to analyzing sample surveys, what to do if there is nonresponse, and how to perform chi-squared tests and regression analyses using data from complex surveys. The National Crime Victimization Survey is discussed in detail as an example of a complex survey. Since many of the formulas used to find standard errors in simpler sampling designs are difficult to implement in complex samples, computer-intensive methods are discussed for estimating the variances.

The book is suitable for a first course in survey sampling. It can be used for a class of statistics majors, or for a class of students from business, sociology, psychology, or biology who want to learn about designing and analyzing data from sample surveys. Chapters 1 through 6 treat the building blocks of sampling, and the sections without asterisks in Chapters 1 through 6 would provide material for a one-quarter course on

sampling. In my one-semester course, I cover sections without asterisks in Chapters 1 through 8, and selected topics from the other chapters. The material in Chapters 9 through 12 can be covered in almost any order, and topics chosen from those chapters to fit the needs of the students.

Exercises in the book are of three types: exercises involving critiquing and analyzing data from real surveys, or designing your own surveys, expose you to a variety of applications of sampling; mathematical exercises (indicated by asterisks) develop your theoretical knowledge of the subject; and exercises using SURVEY allow you to experiment with different sample designs without having to collect all the data in the field. The computer program SURVEY, developed by Professor Ted Chang of the University of Virginia (Chang, Lohr, and MacLaren, 1992), allows you to generate samples on the computer from a hypothetical population. The SURVEY exercises allow you to go through all the steps involved in sampling, rather than just plug numbers into a formula found earlier in the chapter. A disk that includes the data sets and the SURVEY program is provided with the book .

You must know how to use a statistical computer package or spreadsheet to be able to do the problems in this book. I encourage you to use a statistical package such as Splus, SAS, or Minitab, or to use a spreadsheet such as Excel, Quattro Pro, or Lotus 1-2-3 for the exercises. The package or spreadsheet you choose will depend on the length and level of the class. In a one-quarter class introducing the basic concepts of sampling, a spreadsheet will suffice for the computing. Some exercises in the later chapters require some computer programming; I have found that Splus is ideal for these exercises as it combines programming capability with existing functions for statistical analysis. Sampling packages such as SUDAAN (Shah et al., 1995) and WesVarPC (Brick et al., 1996), while valuable for the sampling practitioner, hide the structure behind the calculations from someone trying to learn the material. I have therefore not relied on any of the computer packages that exist for analyzing survey data in this book, although various packages are discussed in Section 9.6. Once you understand why the different designs and estimators used in survey sampling work the way they do, it is a small step to read the user's manual for the survey package and to use the software; however, if you have only relied on computer packages as a black box, it is difficult to know when you are performing an appropriate analysis.

Six main features distinguish this book from other texts intended for students from statistics and other disciplines who need to know about sampling methods.

- The book is flexible for content and level. Many sampling courses have students with a wide range of statistical knowledge. By appropriate choice of sections, this book can be used for an audience of undergraduates who have had one introductory statistics course or for a first-year graduate course for statistics students. The book is also useful for a person doing survey research wanting to learn more about the statistical aspects of surveys and to learn about recent developments. The exercises are flexible as well. Some of the exercises emphasize mastering the mechanics. Many, however, encourage the student to think about the sampling issues involved, and to understand the structure of the sample design at a deeper level. Other exercises are open-ended, and encourage the student to explore the ideas further.
- I have tried to use real data as much as possible—the Acme Widget Company never appears in this book. The examples and exercises come from social sciences,

engineering, agriculture, ecology, medicine, and a variety of other disciplines, and are selected to illustrate the wide applicability of sampling methods. A number of the data sets have extra variables not specifically referred to in text; an instructor can use these for additional exercises or variations.

- I have incorporated model-based as well as randomization-based theory into the text, with the goal of placing sampling methods within the framework used in other areas of statistics. Many of the important results in the last twenty years of sampling research have involved models, and an understanding of both approaches is essential for the survey practitioner. The model-based approach is introduced in Section 2.8 and further developed in successive chapters; however, those sections could be discussed at any time later in the course.
- Many topics in this book, such as variance estimation and regression analysis of nonidentical complex surveys, are not found in other textbooks at this level. The comprehensive sampling reference *Model Assisted Survey Sampling*, by Särndal, Swensson, and Wretman is at a much higher mathematical level.
- This book emphasizes the importance of graphing the data. Graphical analysis of survey data is often neglected because of the large sizes of data sets and the emphasis on randomization theory, and this neglect can lead to flawed data analyses.
- Design of surveys is emphasized throughout, and is related to methods for analyzing the data from a survey. The philosophy presented in this book is that the design is by far the most important aspect of any survey: no amount of statistical analysis can compensate for a badly-designed survey. Models are used to motivate designs, and graphs presented to check the sensitivity of the design to model assumptions. For example, in Chapter 2, the usual formula for calculating sample size is presented. But a graph is also given so that the investigator can see the sensitivity of the sample size to the assumed population variance.

Many people have been generous with their encouragement and suggestions for this book. I am deeply in their debt, although I reserve any credit for the book's shortcomings for myself. The following persons reviewed or used various versions of the manuscript, and provided invaluable suggestions for improvement: Jon Rao, Elizabeth Stasny, Fritz Scheuren, Nancy Heckman, Ted Chang, Steve MacEachern, Mark Conaway, Ron Christensen, Michael Hamada, Partha Lahiri, and several anonymous reviewers: Dale Everson, University of Idaho; James Gentle, George Mason University; Ruth Mickey, University of Vermont; Sarah Nusser, Iowa State University; N. G. Narasimha Prasad, University of Alberta, Edmonton; and Deborah Rumsey, Kansas State University. I had many helpful discussions with, and encouragement from, Jon Rao, Fritz Scheuren, and Elizabeth Stasny. David Hubble and Marshall DeBerry provided much helpful advice on the National Crime Victimization Survey. Ted Chang first encouraged me to turn my class notes into a book, and generously allowed use of the SURVEY program in this book. Many thanks go to Alexander Kugushev, Carolyn Crockett, and the production staff at Brooks/Cole for their help, advice, and encouragement. Finally, I would like to thank Alastair Scott, whose inspiring class on sampling at the University of Wisconsin introduced me to the joys of the subject.

Sharon L. Lohr

目录

第1章	引言	1
1.1	关于样本的争论	1
1.2	优良样本的必要条件	2
1.3	选择偏差	4
1.4	测量偏差	8
1.5	调查表设计	10
1.6	抽样和非抽样误差	15
1.7	练习	17
第2章	简单概率抽样	23
2.1	概率抽样的类型	23
2.2	概率抽样的框架	25
2.3	简单随机抽样	30
2.4	置信区间	35
2.5	样本量估计	39
2.6	系统抽样	42
2.7	简单随机抽样的随机化定理结果*	43
2.8	简单随机抽样模型*	46
2.9	什么时候使用简单随机抽样?	49
2.10	练习	50
第3章	比率和回归估计	59
3.1	比率估计	60
3.2	回归估计	74
3.3	区域中的估计	77
3.4	比率和回归估计模型*	81
3.5	比较	88
3.6	练习	89

第4章	分层抽样	95
4.1	什么是分层抽样?	95
4.2	分层抽样定理	99
4.3	抽样权重	103
4.4	将观察值分配到层中	104
4.5	定义层	109
4.6	分层抽样的一个模型*	113
4.7	事后分层	114
4.8	定额抽样	115
4.9	练习	118
第5章	等概率整群抽样	131
5.1	整群抽样的符号	134
5.2	一阶段整群抽样	136
5.3	二阶段整群抽样	145
5.4	在整群抽样中使用权数	153
5.5	设计一个整群抽样	154
5.6	系统抽样	159
5.7	整群抽样模型*	163
5.8	总结	168
5.9	练习	169
第6章	不等概率抽样	179
6.1	抽选一个初级样本单元	181
6.2	一阶段放回抽样	184
6.3	二阶段放回抽样	192
6.4	无放回不等概率抽样	194
6.5	不等概率抽样例子	199
6.6	随机化定理结果和证明*	204
6.7	模型和不等概率抽样*	211
6.8	练习	213
第7章	复杂调查	221
7.1	设计成分的整合	221
7.2	抽样权重	225
7.3	估计一个分布函数	229
7.4	来自于复杂调查的点数据的图示	235
7.5	设计效应	239
7.6	全国犯罪受害者调查	242
7.7	抽样和试验设计*	247
7.8	练习	249

第 8 章	无响应	255
8.1	忽视无响应的效果	256
8.2	设计调查以减少非抽样误差	258
8.3	回访和两相抽样	262
8.4	无响应机制	264
8.5	无响应的权数方法	265
8.6	估算	272
8.7	无响应的参数模型*	278
8.8	什么是可以接受的的响应率?	281
8.9	练习	282
第 9 章	在复杂调查中的方差估计*	289
9.1	线性化(Taylor 系列)方法	290
9.2	随机组方法	293
9.3	再抽样和重复方法	298
9.4	广义方差函数	308
9.5	置信区间	310
9.6	总结和软件	313
9.7	练习	315
第 10 章	复杂调查的分类数据分析*	319
10.1	多项抽样的 χ^2 检验	319
10.2	关于 χ^2 检验的调查设计效应	324
10.3	χ^2 检验的校正	329
10.4	对数线性模型	336
10.5	练习	341
第 11 章	复杂调查数据回归*	347
11.1	简单随机抽样的基于模型的回归	348
11.2	复杂调查中的回归	352
11.3	在回归中应该使用权数吗?	362
11.4	整群抽样的混合模型	368
11.5	Logistic 回归	370
11.6	对总体总数的广义回归估计	372
11.7	练习	374
第 12 章	抽样中的其他话题*	379
12.1	两相抽样	379
12.2	捕获和再捕获估计	387
12.3	回访的区域估计	396
12.4	稀有事件的抽样	400
12.5	随机化响应	404

	12.6 练习	407
附录 A	调查计划	413
附录 B	抽样中使用的概率概念	423
	B.1 概率	423
	B.2 随机变量和期望值	426
	B.3 条件概率	430
	B.4 条件期望	432
附录 C	数据集	437
附录 D	例子中使用的计算机代码	449
附录 E	统计表	457
	参考文献	459
	作者索引	485
	主题索引	489



Contents

CHAPTER 1	Introduction	1
	1.1	A Sample Controversy 1
	1.2	Requirements of a Good Sample 2
	1.3	Selection Bias 4
	1.4	Measurement Bias 8
	1.5	Questionnaire Design 10
	1.6	Sampling and Nonsampling Errors 15
	1.7	Exercises 17
CHAPTER 2	Simple Probability Samples	23
	2.1	Types of Probability Samples 23
	2.2	Framework for Probability Sampling 25
	2.3	Simple Random Sampling 30
	2.4	Confidence Intervals 35
	2.5	Sample Size Estimation 39
	2.6	Systematic Sampling 42
	2.7	Randomization Theory Results for Simple Random Sampling* 43
	2.8	A Model for Simple Random Sampling* 46
	2.9	When Should a Simple Random Sample Be Used? 49
	2.10	Exercises 50

CHAPTER 3	Ratio and Regression Estimation	59
3.1	Ratio Estimation	60
3.2	Regression Estimation	74
3.3	Estimation in Domains	77
3.4	Models for Ratio and Regression Estimation*	81
3.5	Comparison	88
3.6	Exercises	89
CHAPTER 4	Stratified Sampling	95
4.1	What Is Stratified Sampling?	95
4.2	Theory of Stratified Sampling	99
4.3	Sampling Weights	103
4.4	Allocating Observations to Strata	104
4.5	Defining Strata	109
4.6	A Model for Stratified Sampling*	113
4.7	Poststratification	114
4.8	Quota Sampling	115
4.9	Exercises	118
CHAPTER 5	Cluster Sampling with Equal Probabilities	131
5.1	Notation for Cluster Sampling	134
5.2	One-Stage Cluster Sampling	136
5.3	Two-Stage Cluster Sampling	145
5.4	Using Weights in Cluster Samples	153
5.5	Designing a Cluster Sample	154
5.6	Systematic Sampling	159
5.7	Models for Cluster Sampling*	163
5.8	Summary	168
5.9	Exercises	169

CHAPTER	6	Sampling with Unequal Probabilities	179
	6.1	Sampling One Primary Sampling Unit	181
	6.2	One-Stage Sampling with Replacement	184
	6.3	Two-Stage Sampling with Replacement	192
	6.4	Unequal-Probability Sampling Without Replacement	194
	6.5	Examples of Unequal-Probability Samples	199
	6.6	Randomization Theory Results and Proofs*	204
	6.7	Models and Unequal-Probability Sampling*	211
	6.8	Exercises	213
CHAPTER	7	Complex Surveys	221
	7.1	Assembling Design Components	221
	7.2	Sampling Weights	225
	7.3	Estimating a Distribution Function	229
	7.4	Plotting Data from a Complex Survey	235
	7.5	Design Effects	239
	7.6	The National Crime Victimization Survey	242
	7.7	Sampling and Experiment Design*	247
	7.8	Exercises	249
CHAPTER	8	Nonresponse	255
	8.1	Effects of Ignoring Nonresponse	256
	8.2	Designing Surveys to Reduce Nonsampling Errors	258
	8.3	Callbacks and Two-Phase Sampling	262
	8.4	Mechanisms for Nonresponse	264
	8.5	Weighting Methods for Nonresponse	265
	8.6	Imputation	272
	8.7	Parametric Models for Nonresponse*	278
	8.8	What Is an Acceptable Response Rate?	281
	8.9	Exercises	282

CHAPTER	9	Variance Estimation in Complex Surveys*	289
	9.1	Linearization (Taylor Series) Methods	290
	9.2	Random Group Methods	293
	9.3	Resampling and Replication Methods	298
	9.4	Generalized Variance Functions	308
	9.5	Confidence Intervals	310
	9.6	Summary and Software	313
	9.7	Exercises	315
CHAPTER	10	Categorical Data Analysis in Complex Surveys*	319
	10.1	Chi-Square Tests with Multinomial Sampling	319
	10.2	Effects of Survey Design on Chi-Square Tests	324
	10.3	Corrections to Chi-Square Tests	329
	10.4	Loglinear Models	336
	10.5	Exercises	341
CHAPTER	11	Regression with Complex Survey Data*	347
	11.1	Model-Based Regression in Simple Random Samples	348
	11.2	Regression in Complex Surveys	352
	11.3	Should Weights Be Used in Regression?	362
	11.4	Mixed Models for Cluster Samples	368
	11.5	Logistic Regression	370
	11.6	Generalized Regression Estimation for Population Totals	372
	11.7	Exercises	374
CHAPTER	12	Other Topics in Sampling*	379
	12.1	Two-Phase Sampling	379
	12.2	Capture-Recapture Estimation	387
	12.3	Estimation in Domains, Revisited	396

12.4	Sampling for Rare Events	400
12.5	Randomized Response	404
12.6	Exercises	407
APPENDIX A	The SURVEY Program	413
APPENDIX B	Probability Concepts Used in Sampling	423
B.1	Probability	423
B.2	Random Variables and Expected Value	426
B.3	Conditional Probability	430
B.4	Conditional Expectation	432
APPENDIX C	Data Sets	437
APPENDIX D	Computer Codes Used for Examples	449
APPENDIX E	Statistical Table	457
	References	459
	Index	485