HZ BOOKS

PEARSON Prentice Hall

经 典 原 版 书 库

抽样理论与方法

（英文版）

本书只供在中国大陆销售

(美) 扎库拉·戈文达拉玉卢 著
肯塔基大学

机械工业出版社
China Machine Press

经 典 原 版 书 库

抽样理论与方法

（英文版）

(美) 扎库拉·戈文达拉玉卢
肯塔基大学 著

机械工业出版社
China Machine Press

本书版权登记号：图字：01-2005-1976

图书在版编目（CIP）数据

抽样理论与方法（英文版）/（美）戈文达拉玉卢（Govindarajulu, Z.）著. －北京：机械工业出版社，2005.6

（经典原版书库）

书名原文：Elements of Sampling Theory and Methods

ISBN 7-111-15889-X

I. 抽 … II. 戈… III. 抽样调查－英文 IV. C811

中国版本图书馆CIP数据核字（2004）第140998号

机械工业出版社（北京市西城区百万庄大街22号　邮政编码　100037）

责任编辑：迟振春

北京昌平奔腾印刷厂印刷·新华书店北京发行所发行

2005年6月第1版第1次印刷

787mm × 1092mm 1/16 · 27印张

印数：0 001-3 000册

定价：55.00元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

本社购书热线：（010）68326294

He hides his faults, the pseudo-wise,
and highlights the faults in others galore.
He knows naught of himself or his self
and the rest that he knows is best unknown.

—Sri Sathya Sai Baba

This book is dedicated to all the research workers who have enriched the theory and methods of sampling.

序

本书是作者在肯塔基大学多年为研究生讲授抽样调查的经验总结。他从中发展出一种方法，采用这种方法，在一学期内可以讲授抽样调查的全部概念，包括设计调查、收集数据以及用现代统计方法进行数据分析。在实现这个目标的同时并不影响对调查方法论的严格讨论，或者在按照调查数据进行的统计推断中省略对关键结果的重要证明。

我阅读过很多有关抽样调查理论与方法的书籍，在印度统计研究所工作期间也参加过多年实际调查数据的规划、收集和分析工作。本书在这一点上堪称卓越，因为它不仅为要成为一名抽样调查顾问的人提供了所需的理论知识，而且也提供了实际的指南。

一种正确实施的采用最优调查设计的抽样调查，可以用最小的成本提供足够精确的信息。也可以这样说，根据抽样数据作出的估计，即使容易出现误差，只要这些数据是由训练有素的调查人员在充分监督下收集起来的，其精度就比依据完整列表所得的估计更高。列表方法容易引起非抽样误差，而这种误差通常高于抽样误差。本书通过强调抽样调查设计的各个环节以及采用现代统计方法进行数据分析，把这个信息传达给读者。

本书包含的题材非常广泛，从简单抽样设计和相关的估计过程开始，直到更复杂的设计和采用现代统计方法（如刀切法、自助法和漏失值的估算等）。本书把与抽样调查有关的所有文献的讨论包容在一册书中，这一点非常吸引人。

本书可以作为研究生的教材，也可作为抽样调查从业人员的参考书。

C. R. Rao

宾夕法尼亚州立大学统计学系统计学Eberly讲席教授

前　言

在抽样中，人们的兴趣之一是对总体总和或总体均值的估计。所以，抽样的中心问题是统计推断。在一学期统计学课程中，可以按几种大纲讲授抽样。但是，如果采用已有的书作为教材，则不适合在一学期内学完抽样的理论与方法，例如，Hansen、Hurwitz和Madow的书（1953），Sukhatme等人的书（1984）以及Cochran的书（1977），都是按两学期课程设计的。有鉴于此，我以多年来在肯塔基大学为研究生讲授的一门课程为基础编写了此书。书中包含能够在一个学期内传授给学生的尽可能多的基本抽样思想。本书可以作为一学期课程的教材或参考书。书中给出了全部基本证明，从这个意义上说本书是自成一体的。

本书对不等概率放还抽样在第2章就给出介绍。有些读者可能愿意跳过第2章，在读第10章之前再返回这一章来。一学期课程可讲授第1~9章、第11~13章和第15、16、18章。可以把本书作为教授方法课程的教材，从而略去大部分证明。

与现有的抽样教材不同，本书包含许多新的题材，如贝叶斯方法、刀切法、自助法、小区域估计和估算法。不仅如此，多数概念是用数值实例说明的。每章后面备有习题，有些习题带有实际数据或人工数据。此外，每章后列出参考文献，并在书后提供总的文献目录。以本书为基本教材的课程所需的预备知识是大学的代数和基本统计推断。

由于篇幅有限，因此本书难以全面而客观地述及这一领域的专家们的各种观点。题材的选择无疑带有主观性，而且受到课程层次的制约。同时，本书对参考文献的选取也高度精炼。因此，在此谨向那些文献目录中未曾引用的著作或论文的作者表示歉意。非常欢迎读者指正书中存在的任何错误。

我是在明尼苏达大学师从Richard B. McHugh教授和I. Richard Savage教授学习抽样理论与抽样方法的，那时分别以Hansen、Hurwitz和Madow的书（1953）及P. V. Sukhatme的书（1954）为教材。我在肯塔基大学曾用Cochran的书（1977）作为一门课程的教材，并深受影响。我要感谢J. N. K. Rao教授和D. Raghavarao教授，承蒙他们阅读本书的一种早期原稿，并就主题的取舍提出许多建议，同时引起我对发表的许多重要论文的注意。同样感谢Xiao-Li Meng教授、C. R. Rao教授和John L. Wasik教授，他们阅读了本书的几种早期原稿，并对在书中加进实例、习题和新的主题提出富有建设性的和非常有益的建议。我要向Prentice-Hall公司的统计学编辑Ann Heath表示诚挚的谢意，还要感谢该公司的Mindy McClard、Linda Behrens、Bayani DeLeon、Robert Lenz和Jennifer Pan给予的巨大帮助与支持。Beverly A. Clayborne和Julie Smith承担了几种早期原稿的录入工作，Brian Moses出色地完成了后面几章和最后定稿的录入，对他们谨致谢意。我也十分感谢肯塔基大学统计学系提供的支持。最后感谢肯塔基大学的全体研究生，他们试用了本书几种早期的版本。

扎库拉·戈文达拉玉卢

于肯塔基州列克星敦

CONTENTS

CHAPTER 1

PRELIMINARIES

1.1 INTRODUCTION

Our behavior, attitudes, and sometimes actions are based on samples. We may generalize on the basis of a sample of size one—for instance, our likes or dislikes of foreign dishes, nationals, or countries. Such a sample is not likely to be representative of the whole population. What size sample should we draw, then? The size of the sample depends on the accuracy we need.

Sampling is widely used in the modern world. The statistical office of the United Nations has sample surveys conducted by member nations on topics of interest such as unemployment, size of labor force, and water consumptions. Sampling is useful in inventories and marketing of products. Airlines and Federal Reserve boards apportion money on the basis of samples of records. Public polls brought sampling techniques to the attention of the public, but surveys have proved to be a useful and important technique for the past several hundred years.

1.2 A BRIEF HISTORY OF SURVEY SAMPLING

An early contribution to sampling theory was made by the French mathematician P. S. Laplace, who tried to estimate the population of France using the theory of *ratio estimation* and the reported births for all areas and counts of inhabitants in a purposive sample of parishes. He also gave a measure of the sampling error under simplified

assumptions (see Cochran (1978)). Early in the twentieth century A. A. Tschuprow, a Russian statistician, was active in the International Statistical Institute's (ISI) discussions of the *representative method.* He also developed early theory of sampling and obtained a solution to the problem of *optimum allocation* of a sample to the various strata (see Tschuprow (1923)). In 1924 the ISI set up a commission which presented its report on the representative method in statistics in 1926. In its memoranda, the commission defined the method of *random selection* and *'purposive selection'* (the latter is named for using groups of elements as sampling units). The statistician who brought the concept of *randomization* to the fore was A. L. Bowley (1926), who developed (1) the notion of a *'frame'* (i.e., a complete list of the units in the population serving as sampling units); (2) a theory for *proportionate stratified sampling*, and (3) a theory for *purposive selection.*

R. A. Fisher developed the theory of *experimental design and analysis* for analyzing data from field trials. Fisher's theory consists of six principles, which include replication, local control, and randomization. Randomization was used to minimize the bias of the selection procedure. F. Yates (1946) and others at the Rothamsted Experimental Station developed the theory of *sampling clusters of elements* and of *multistage sampling.* J. Neyman's presentation of a paper in 1934 at a meeting of the Royal Statistical Society provided a catalyst for promoting theoretical research, methodological developments, and applications of *probability sampling.* He presented new sampling designs. He developed the optimum allocation of sampling units to strata quite independently of Tschuprow. At the invitation of W. E. Deming, Neyman gave a series of lectures in Washington, D.C., in 1937. During one of the lectures M. Friedman and S. Wilcox posed a problem which fell in the domain of *double-sampling* or *two-phase sampling* considered by Neyman (1938).

The above developments had a tremendous impact in the United States. The poor predictions made by the *Literary Digest* and the success of the Gallup Poll in forecasting the 1936 presidential election have focused attention on the value of *survey sampling.* In 1942 the sample survey of unemployment was transferred to the Bureau of Census, which made some far-reaching changes in survey design, introducing the use of *probability sampling* at all stages of sampling and ratio estimation. Also the Department of Agriculture launched a program of research and, as a part of it, a research group was established at the Statistical Laboratory of the Iowa State University. W. G. Cochran, who worked with F. Yates at the Rothamsted Experimental Station, joined that group in 1939 and made significant contributions to sampling theory and design. The new developments influenced the efforts of P. C. Mahalanobis at the Indian Statistical Institute and P. V. Sukhatme and V. G. Panse at the Indian Council of Agricultural Research. The Indian school made significant contributions to the survey sampling theory and methodology and, in particular, the control of nonsampling errors in sample surveys. The impact was also felt in Europe with contributions by P. Thionet in France, O. Anderson and H. Kellerer in Germany, and T. Dalenius in Sweden. For further details on the history and early developments of sample surveys the reader is referred to Hansen, Dalenius, and Tepping (1985) or Bellhouse (1988).

1.3 SAMPLING DESIGNS AND AN OVERVIEW OF SAMPLING

Sample design determines the precision of the estimates. Thus, the way a sample is drawn is as important as the mathematical form of the estimator. Sample design consists of both a sample selection plan and an estimation procedure. Each of these has several features: first we have to define sampling units (*primary*, *secondary*, etc.); sample schemes, such as *simple random sampling* (with or without replacement), *stratified sampling*, *double sampling*, *multistage sampling*, *cluster sampling*, and *systematic sampling*; optimum allocation of sampling units to various strata; and *sampling with varying probabilities*.

We try to estimate the population parameters by linear estimates. We can incorporate auxiliary information into the *ratio estimates* and *regression estimates*. We try to linearize nonlinear estimators (for example, ratio estimators) before we compute their variances. Also, an important effort in sample surveys is to control and minimize nonresponse errors. For a list of existing procedures, recommended alternatives, and a proposal for additional research the reader is referred to the panel on Incomplete Data of the Committee on National Statistics (1983). Also for further references on nonsampling errors see Neter and Waksburg (1964) and Mosteller (1978).

Regarding the foundations, Godambe (1955) formulated a general mathematical theory for survey sampling from finite populations. He showed that no uniformly best linear unbiased estimator exists. However, this does not invalidate the theory of Neyman allocation, because Godambe (1955) defines "linear" in a different fashion. The concept of a super population was devised in order to assert certain properties (such as unbiasedness) of estimators. Such a model can provide an effective guide to sample design within the premise of probability sampling. If the model is not correct, the estimates may be seriously biased and the confidence coefficient associated with the confidence interval may be less than the nominal coefficient. For an illustration see Hansen, Madow, and Tepping (1983). Some of the open problems are how to further control nonsampling errors, find the rate of convergence in asymptotic properties, and ensure privacy of response in sample surveys.

1.4 INGREDIENTS OF A SURVEY

Sample surveys can be classified as (1) *descriptive*, and (2) *analytical*. In a descriptive survey, for example, we study the proportion of the population watching a certain television program, or the proportion of people afflicted by a certain disease. In analytical surveys we compare groups and employ statistical techniques in order to estimate population parameters (pointwise or by intervals). Most of the surveys available at the United Nations fall in the descriptive category. Carrying out a survey may be easy if the people or subjects are well defined and well organized. For instance, taking a survey of inhabitants in a city is much easier than taking a survey of inhabitants who can be reached only by water or who are suspicious of strangers or who are homeless.