

抽 样 技 术

[美] W·G·科克伦 著

张尧庭 吴 辉 译

中国统计出版社

抽样技术

（第三版）陈希孺、袁见龙、袁方
编 著 袁方、袁见龙

中国统计出版社

抽 样 技 术

[美] W·G·科克伦 著

张尧庭 吴 辉 译

中国统计出版社

William G. Cochran
SAMPLING TECHNIQUES

Third edition

John Wiley & Sons, 1977

本书根据1977年版本译出

抽 样 技 术

【美】W·G·科克伦 著

张尧庭 吴 辉 译

中国统计出版社出版

新华书店北京发行所发行

北 京 印 刷 一 厂 印 刷

787×1092毫米 32开 23印张 50万字

1985年4月第1版 1985年4月北京第1次印刷

印数：1—15,000

统一书号：4006·054 定价：3.90元

译者的话

威廉·G·科克伦是美国著名的统计学家，曾先后担任美国约翰·霍普金斯大学、北卡罗来纳州立大学、衣阿华州立大学教授和哈佛大学统计学名誉教授，他还是美国国家科学院和美国艺术科学院的院士。1980年他与世长辞，但人们永远不会忘记他在统计学术上所作出的重要贡献。他所写的《抽样技术》一书，从50年代出版以来，一直是这一领域中的权威性著作。全书包含了国际上抽样技术的最新发展成果，并引用了大量的、各种类型的例子来说明各种抽样方法的效果。他的论述条理清楚，对于数学定理的证明，尽量采用简明初等的方法，易于为广大读者接受。尤其可贵的是，书中还叙述了丰富的实际调查经验，并对各种不同的方法作出了有实际分析的评价和比较。每章最后都有一些练习题，读者通过练习可以加深对原理的理解，掌握抽样的方法。

抽样调查在我国日益为人们重视，并已在许多实际领域中得到应用。我国统计部门和其他有关部门对抽样调查方法的研究也正在向深度发展。但目前我国出版的有关抽样调查的书籍较少，一些统计原理书籍在介绍抽样调查知识时，往往只列出计算公式，而没有进行必要的数学推导，这就妨碍读者充分理解抽样原理与各种方法。在我国的有关书籍中，对国际上最新发展的抽样技术也介绍较少。本书的翻译出版正好可以弥补这方面存在的缺陷。

目前，统计术语在我国尚未标准化，其中有许多术语已

经有了比较一致的译名，还有一部分术语则很不统一。我们在翻译中为了照顾社会经济统计工作者与数理统计工作者的习惯，尽可能采用了两方面都常用的译名；对于某些有分歧的译名，则在它第一次出现时附以英文，并注明其他较流行的译名，供读者参考。还有少数术语在英文中是一个词，在汉译中则在不同的场合有不同的习惯译法（如 median 有时译为中值，有时译为中位数），我们遵从习惯，不予改动。Estimator 与 Estimate 这两个术语，尽管作者一开始就说明它们在意义上的区别（我们把前者译为估计量，后者译为估计值），但他在以后的叙述中却并未严格遵守他自己的规定，在不少场合是混用了。我们在翻译时尽可能加以区别，在两种意义都有的场合则译成“估计”。为了便于读者比较，我们在书末附了一个统计术语英汉对照表，供大家参考。原书中有个别印刷错误，我们发现时都给以纠正，并写明译注。

本书根据 1977 年出版的原著第三版译出。1981 年冬天，瑞典统计学家、世界著名抽样调查专家托·戴伦纽斯 (Tore Dalenius) 教授来我国讲学时，曾把这一最新的版本赠送给国家统计局。原国家统计局副局长、现任顾问常诚同志听到本书内容的介绍后，认为它对我国开展抽样调查很有参考价值，所以决定组织翻译。第一章至第八章由武汉大学数学系张尧庭翻译（翻译过程中曾参考了湖北财经学院张璋同志和西安交通大学方刚同志的一些译稿），第九章至第十三章由国家统计局统计科学研究所吴辉翻译。我们衷心感谢常诚、黄海等同志对我们翻译工作的关怀与鼓励，并感谢统计出版社的大力支持。没有这些，是不可能顺利完成这本书的翻译、出版的。

由于我们水平的限制，译文中很可能还会有不妥或错误的地方，欢迎大家批评指正。

张尧庭

吴 绎

1984 年 3 月

序 言

象前几版一样，这本教科书对抽样理论作出了全面的叙述，这些理论是为适应抽样调查中实际应用的需要而发展起来的。书中有一些例子用来说明如何把理论应用于实际，还有一些给学生做的练习题。这本书不论用作侧重于理论的抽样调查课程的教材，还是供学生个人阅读，都将是有用的。

为了理解书中的大部分材料，读者在数学知识方面至少必须熟悉代数，特别是比较复杂的代数表达式，还需要掌握有限样本空间的概率(包括组合概率)知识。这本书假定读者已学过导论性的统计课程，在这种课程中，对均值，标准差，正态分布，二项分布，超几何分布与多项分布，中心极限定理，线性回归和较简单类型的方差分析，都是有论述的。由于经典抽样调查理论中很多是处理由抽样调查方案所确定的随机化中估计量的分布，所以具有一些非参数知识是有益的。

在这一版中，论题的排列次序基本上与以前各版相同。书中新增加了或改写了一些节，主要是出于以下三个理由：(1) 对在抽样领域中比较新的一些论题(抽样方案或估计方法)加以介绍；(2) 在过去的 15 年中，为了改进较老的抽样方法或从与之竞争的一些方法中借鉴更多有效的东西，已对这些老方法作了进一步的发展，本书论述了这些发展；(3)

将前几版中的一些证明加以压缩，阐明或简化。

这一版中论述了以下一些新的问题：当采用复杂的抽样调查方案时，从调查结果中得出的非线性估计量的标准误或置信限如何确定，是一个困难的问题，现已研究出近似的解决方法；对调查结果进行统计分析（例如，回归分析）时，愈来愈需要这些方法。有些调查含有一些敏感性的问题，被调查者不大愿意作出真实的答复，在这种调查中采用的一种新的方法，是向被调查者或者提出那个敏感性的问题或者是提出一个无关的非敏感性问题，而由随机化所作出的具体选择是调查员所不知道的。在某些问题的调查中，采用两份有部分重迭的名单（或称为抽样框）来包括全部总体，在经济上可能是有吸引力的，在没有充分人力、资金进行抽样的国家中这也是一种基本的做法。当对总体内一些子群的均值进行比较时，已推广应用了双重抽样方法。如果可以假定，有限总体本身是从无限超总体中抽取的一个随机样本，而在这个无限超总体中适合于比率估计量或回归估计量的数学模型是成立的，那么，比率估计量和回归估计量就具有一些引人注意的性质，现已有著作对这些性质进行了有趣的论述。上面的假定并非是新的（我最近注意到，拉普拉斯在 1800 年左右曾在一个抽样的问题中应用过这个假定），但它阐明了抽样调查理论与标准的统计理论之间的关系。

本版还对以前版本中的一些论题作了进一步的论述，其中的一个例子是第 9 A 章，它有一部分是根据以前第 9 章中的材料写成的。这主要是为了对不等概率不放回抽样中那些在我看来是主要的方法作出更为充分的叙述，其中包括布鲁尔、J. N. K. 拉奥、德宾各自独立提出的类似的方法，

莫蒂的方法，拉奥、哈特利、科克伦方法，以及麦多的与等距抽样有关的方法，书中还把这些方法应用于自然总体时的效率作了比较。统计学家们已对调查中计量误差的各组成部分的大小的确定方法作出了一些新的研究，它们可以由各不同调查员用重复计量的方法来确定，也可以用交叉子样本的方法来确定，或结合采用这两种方法来确定。关于比率估计量，当应用标准大样本公式来计算方差和估计方差时是会产生小样本偏差的，从自然总体取得的数据已被用来对这种小样本偏差进行估计。统计学家们也作了一些尝试，创造出偏差较小的比率估计量以及计算它的抽样方差的公式。在分层抽样方面，当重要的调查项目在一个以上的情况下如何在各层中分配样本含量，以及当每层只抽选一个单位时如何估计抽样误差，对这些问题也有了进一步的论述。为了处理具有线性趋势的总体而采用新的等距抽样方法也是有意思的。

阿尔瓦·L·芬克纳和埃米尔·H·杰布曾为我准备了大部分讲稿，本书的第一版就是根据这些讲稿写成的。提供背景材料的一些调查曾得到海军部海军研究室的支持。托尔·戴伦纽斯，大卫·J·芬尼，丹尼尔·G·霍维茨，莱斯利·基什，P·S·K·桑巴西瓦·拉奥，马丁·桑德留斯，约瑟夫·塞德兰斯克，阿莫德·R·森，特别是乔恩·N·K·拉奥在对抽样的最近发展的讨论中或在对这一版的建议中，曾给了我很大的帮助，他们费了不少心血来阅读本版中新的和修改过的各节，对存在的缺陷、弱点，论述得不够清楚的地方以及在论题的选择方面向我提出了许多建设性的建议。我还要感谢罗温娜·福斯，霍利·格兰诺和伊迪

丝·克洛茨在准备打字稿过程中为我打字和进行其他工作。
我谢谢大家。

威廉·G·科克伦

1977年2月于马萨诸塞州南奥尔良

目 录

第 1 章 引言	(1)
1.1 抽样方法的优点	(1)
1.2 抽样调查的若干应用	(3)
1.3 抽样调查的主要步骤	(7)
1.4 抽样理论的作用	(11)
1.5 概率抽样	(13)
1.6 与概率抽样不同的方法	(15)
1.7 正态分布的应用	(17)
1.8 偏差和它的影响	(19)
1.9 均方误差	(22)
练习	(24)
第 2 章 简单随机抽样	(26)
2.1 简单随机抽样	(26)
2.2 简单随机样本的抽选	(27)
2.3 定义和记号	(30)
2.4 估计量的性质	(31)
2.5 估计量的方差	(33)
2.6 有限总体校正系数	(36)
2.7 根据一个样本来估计标准误	(38)
2.8 置信限	(40)

2.9	证明的另一个方法	(42)
2.10	放回的随机抽样	(43)
2.11	比率的估计	(45)
2.12	子总体均值的估计	(50)
2.13	子总体总值的估计	(52)
2.14	域均值的比较	(56)
2.15	正态近似的正确性	(57)
2.16	总体均值的线性估计量	(64)
	练习	(66)
第 3 章	抽样比例及百分比	(73)
3.1	品质标志	(73)
3.2	样本估计量的方差	(74)
3.3	P 对标准误的影响	(77)
3.4	二项分布	(79)
3.5	超几何分布	(80)
3.6	置信限	(82)
3.7	多于两类的分类	(86)
3.8	多于两类情况下的置信限	(87)
3.9	p 的条件分布	(88)
3.10	子总体的比例与总值	(90)
3.11	不同域之间的比较	(91)
3.12	对整群抽样中的比例进行估计	(92)
	练习	(98)

第 4 章 样本含量的估计	(103)
4.1 一个假设的例子	(103)
4.2 问题的分析	(105)
4.3 精确度的确定	(106)
4.4 为求比例而进行的抽样中求 n 的公式	(107)
4.5 稀少项目——逆抽样	(109)
4.6 在连续数据的情况下求 n 的公式	(110)
4.7 总体方差的预先估计	(112)
4.8 在多于一个调查项目的情况下的样本 含量.....	(116)
4.9 当需要对总体的各部分作出估计时的样本 含量.....	(118)
4.10 决策问题中的样本含量.....	(119)
4.11 设计效果 (Deff)	(122)
练习.....	(123)
第 5 章 分层随机抽样	(128)
5.1 概述	(128)
5.2 记号	(129)
5.3 估计量的性质	(130)
5.4 估计量的方差和置信限	(136)
5.5 最优分配	(138)
5.6 分层随机抽样与简单随机抽样的相对精确 度	(142)
5.7 什么时候分层使精确度有大的提高	(144)

5.8	要求进行大于 100% 的抽样的分配	(148)
5.9	在连续数据的情况下估计样本含量	(150)
5.10	求比例的分层抽样.....	(153)
5.11	求比例的分层抽样中精确度上的提高.....	(155)
5.12	求比例时样本含量的估计.....	(157)
	练习.....	(158)

第 5 A 章 分层抽样的其他方面..... (165)

5 A.1	偏离最优分配的影响	(165)
5 A.2	层含量的误差的影响	(168)
5 A.3	当调查项目超出一个时, 样本含量在各层 之间的分配问题	(171)
5 A.4	当调查项目超出一个时, 样本含量在各层 之间的其他分配方法	(174)
5 A.5	小样本情况下的两种方式分层	(178)
5 A.6	有控制的抽选	(181)
5 A.7	层的构成	(183)
5 A.8	层数	(190)
5 A.9	抽样后分层(后分层)	(194)
5 A.10	定额抽样.....	(195)
5 A.11	从一个样本来估计分层的得益.....	(196)
5 A.12	每层抽一个单位时方差的估计.....	(200)
5 A.13	作为研究域的层.....	(203)
5 A.14	估计子总体的总值及均值.....	(205)
5 A.15	从两个抽样框中抽样.....	(209)
	练习.....	(211)

第 6 章 比率估计量	(219)
6.1 估计方法	(219)
6.2 比率估计量	(220)
6.3 比率估计量的近似方差	(222)
6.4 从一个样本估计方差	(226)
6.5 置信限	(227)
6.6 比率估计量与单位均值的比较	(229)
6.7 比率估计量成为最佳线性无偏估计量的条件	(230)
6.8 比率估计量的偏差	(234)
6.9 方差和估计方差公式的准确度	(237)
6.10 分层随机抽样中的比率估计量.....	(239)
6.11 组合比率估计量.....	(241)
6.12 组合估计量与分别估计量的比较.....	(243)
6.13 估计方差的简捷计算.....	(246)
6.14 采用比率估计量时的最优分配.....	(250)
6.15 无偏的比率型估计量.....	(254)
6.16 这些方法的比较.....	(258)
6.17 改进的方差估计.....	(261)
6.18 两种比率的比较.....	(263)
6.19 两个比率的比率.....	(268)
6.20 多元比率估计量.....	(270)
6.21 乘积估计量.....	(272)
练习.....	(273)

第7章 回归估计量..... (277)

7.1 线性回归估计量 (277)

7.2 b 已预先确定情况下的回归估计量 (279)

7.3 从样本算出 b 时的回归估计量 (282)

7.4 方差的样本估计 (285)

7.5 与比率估计量及单位均值的大样本比较 ... (286)

7.6 计算 $V(\bar{y}_{lr})$ 和 $V(\bar{y}_{lr})$ 的大样本公式的准确性 (288)

7.7 线性回归估计量的偏差 (289)

7.8 线性回归模型的线性回归估计量 (291)

7.9 分层抽样中的回归估计量 (292)

7.10 从样本估计回归系数..... (294)

7.11 两类回归估计量的比较..... (296)

练习..... (297)

第8章 等距抽样..... (300)

8.1 概述 (300)

8.2 等距抽样与整群抽样的关系 (303)

8.3 估计均值的方差 (303)

8.4 等距抽样与分层随机抽样的比较 (311)

8.5 按“随机”次序排列的总体 (311)

8.6 有线性趋势的总体 (314)

8.7 处理有线性趋势的总体的方法 (316)

8.8 有周期变异的总体 (318)

8.9 自相关总体 (320)