

抽样估计精度问题研究

李金昌 著



中国物价出版社



ISBN 7-80155-159-1



9 787801 551597 >

ISBN 7-80155-159-1
F · 124 定价 18元

浙江省重点扶植学科(统计学)基金资助
浙江省“151人才工程”入选人员资助经费

抽样估计精度问题研究

李金昌 著

中国物价出版社

图书在版编目(CIP)数据

抽样估计精度问题研究/李金昌著. —北京:中国物价出版社,

2000

ISBN 7-80155-159-1

I. 抽… II. 李… III. 抽样误差-研究 IV. C811

中国版本图书馆 CIP 数据核字(2000)第 76762 号

书名/抽样估计精度问题研究

著者/李金昌

出版发行/中国物价出版社(电话:68020336 邮编:100837

地址:北京市西城区月坛北小街 2 号院 3 号楼)

经销/新华书店

印刷/杭州商学院印刷厂

开本/850×1168 毫米大 印张/8 字数/215 千字

版本/2000 年 12 月第 1 版 印次/2000 年 12 月第 1 次印刷

书号/ISBN7-80155-159-1/F.124

定价/18 元

作者简介

李金昌,男,1964年12月出生,浙江义乌人。1986年获杭州商学院统计学专业学士学位,1990年获厦门大学统计专业硕士学位,1999年获厦门大学统计学专业博士学位。1988年至今,在杭州商学院从事统计教学和科研工作。1991年任讲师,1993年任副教授,1996年任教授。现为杭州商学院经济学院院长、统计系主任,硕士生导师。浙江省首批中青年学科带头人,浙江省“151人才工程”入选人员,原国内贸易部有突出贡献中青年专家,享受政府特殊津贴。

主要研究方向为:统计基本理论与方法,抽样调查,市场调查与分析等。先后主持完成国家级和省级基金项目3个,参与完成国家级和省级课题3个,主持完成其他课题多个。在《统计研究》等刊物发表学术论文50余篇,出版专著《抽样调查与推断》(独著)、《市场统计调查方法与应用》(4人合著,第一作者)、《调查统计学》(3人合著,第二作者)、《统计估算理论、方法和应用研究》(多人合著,第三作者)等多部,主编浙江省高等教育重点建设教材《新编统计学教程》。多项科研成果获各类奖励。曾先后被授予浙江省高校优秀青年教师和青少年英才等荣誉称号。

主要学术兼职有:中国统计学会常务理事,浙江省统计学会副会长,全国统计教材编审委员会委员等。

前 言

本书是在博士学位论文基础上形成的。

抽样估计作为一种科学的统计调查方法,已在实践中得到了广泛的应用,其优点可以概括为四个字:省、准、快、便。其中的“准”是指抽样估计只需通过对从研究现象总体中随机抽取的样本进行观测或调查,就可以在一定的概率保证下,据以对总体数量特征作出具有一定精确度的估计。这个精确度,就是抽样估计精度。在抽样估计中,追求尽可能高的估计精度,始终是我们的目标。因此,无论是抽样理论研究还是抽样实践应用,抽样估计精度问题都是一个关键的问题。

然而,什么是抽样估计精度?它受哪些因素影响?该如何加以测定与控制?如何在一定条件下使抽样估计精度达到最高?对于这些我们在讨论抽样估计精度问题时所必须面对的问题,目前还没有系统的结论。虽然国内外有关专家、学者曾从不同的方面和角度进行过一些研究,有些观点和结论也已在有关文献中得到了反映,但总的来说,把抽样估计精度问题作为一个专门的问题加以系统的探讨和论述,似乎尚未有见。正因为如此,对已有的研究成果加以归纳和总结,使之系统和完整化,并对相关的问题作更深入、更进一步的探讨和研究,应该是一项值得肯定的和有意义的工作。基于这样的考虑,我就把《抽样估计精度问题研究》作为自己的博士学位论文的选题。

全书共分八章。第一章为“抽样估计精度的一般问题”,将讨论抽样估计精度的含义与衡量指标、影响因素和控制原则等问题,是本书研究的基础。第二章为“抽样方法与抽样估计精度”,将讨论重复抽样与不重复抽样、等概率抽样与不等概率抽样对估计精

度的影响。第三章为“参数估计方法与抽样估计精度”,将讨论简单估计法、比率估计法与回归估计法三者估计精度的差异,小样本时比率估计法与回归估计法均方误差的高、低估现象,以及各种参数估计方法的结合应用与抽样估计精度的关系等问题。第四、五章为“抽样组织形式与抽样估计精度”,将讨论各种抽样组织形式下抽样估计精度的高低,以及多种抽样组织形式结合应用与抽样估计精度的关系等问题。第六章为“辅助信息的利用与抽样估计精度”,将讨论辅助信息的利用方式对抽样估计精度的影响,以及利用初样本辅助信息、前期样本信息和多辅助信息对提高抽样估计精度的作用等问题。第七章为“调查费用与抽样估计精度”,将讨论调查费用模型,调查费用与抽样估计精度的函数关系,以及费用约束条件下抽样估计精度的最优解等问题。第八章为“非抽样误差与抽样估计精度”,将讨论非抽样误差对抽样估计精度影响的特点,以及抽样框误差、无回答误差和计量误差对抽样估计精度影响的测定与控制等问题。全书的讨论始终围绕如何在一定条件下使抽样估计精度尽量达到最高这一中心,力求做到“涉及面广而有深度,理论性强但联系实际”,以期对今后同类问题的研究有参考价值。

当然,本书的研究只是初步的。虽然自己花费了不少精力想使研究成果尽量完美,但由于水平所限,书中定有不少方面需要进一步去研究,有些提法、观点和结论也可能需要进一步去推敲,因此,恳请同行专家、学者能对本书多加指正、不吝赐教,以使我今后对抽样估计精度问题的研究更加深入、更加完善。

李金昌

2000 年秋于杭州商学院

INTRODUCTION

This book is about "Sampling Estimate Precision".

The book is composed of 8 chapters. Chapter 1 is named "The Basic Problems of Sampling Estimate Precision". It discusses the meanings and the affection actors of sampling estimate precision, the ways to assess the precision and the principles of controlling the precision. Chapter 2 is named "Sampling Methods and Estimate Precision". It discusses the different affections to the precision between the sampling with replacement and without replacement, with equal probability and with unequal probability. Chapter 3 is named "Estimate Methods and Estimate Precision". It discusses the differences of precision among Simple estimate, Ratio estimate and Regression estimate, compares the MSE between Ratio estimate and Regression estimate under the condition of small sample, and observes the effects of precision by combining one estimate method with the others. Chapter 4 to 5 is named "Sampling Design and Estimate Precision". It discusses the different precisions to the different sampling designs (They are Pure Random Sampling, Stratified Sampling, Systematic Sampling, Cluster Sampling and Multistage Sampling), and tells whether the combined application of several sampling designs can raise the precision. Chapter 6 is named "Auxiliary Information and Estimate Precision". It discusses the relations between the using way of the auxiliary information and the estimate precision, and analyses the efficiency of using the auxiliary information that comes from pre-

liminary sample or previous sample. Chapter 7 is named "Survey Cost and Estimate Precision". It discusses the models of survey cost, the function relations between survey cost and precision, and the optimum precision under the given survey cost. Chapter 8 is named "Nonsampling Error and Estimate precision". It tells how the non-sampling errors affect the estimate precision, and how to assess and to control the estimate precision that affected by the sampling frame error, nonresponse error and measurement error.

This book is always concerned about how to get the highest sampling estimate precision under a definite condition.

目 录

前言	(1)
第一章 抽样估计精度的一般问题	(1)
第一节 抽样估计精度的意义	(1)
一、抽样估计精度的意义	(1)
二、抽样估计精度的表现形式	(3)
三、抽样估计精度的作用	(9)
第二节 抽样估计精度的影响因素	(11)
一、总体的客观条件	(11)
二、样本容量的大小	(13)
三、抽样方法	(14)
四、总体参数估计方法	(14)
五、抽样组织形式	(15)
六、辅助信息	(15)
七、调查费用	(15)
八、抽样框	(16)
九、概率保证程度	(16)
第三节 抽样估计精度的控制原则与要求	(17)
一、抽样估计精度的控制原则	(17)
二、控制抽样估计精度要注意的问题	(19)
本章主要参考文献	(21)
第二章 抽样方法与抽样估计精度	(23)
第一节 抽样方法概述	(23)

一、抽样方法所体现的估计精度的思想·····	(23)
二、重复抽样与不重复抽样·····	(24)
三、等概率抽样与非等概率抽样·····	(27)
第二节 是否重复抽样与抽样估计精度·····	(30)
一、两种抽样方法下的抽样估计量概率分布·····	(30)
二、两种抽样方法下的估计量方差·····	(31)
三、两种抽样方法下估计精度的比较·····	(35)
第三节 是否等概率抽样与抽样估计精度·····	(38)
一、等概率抽样的估计量方差·····	(38)
二、不等概率抽样的估计量方差·····	(41)
三、两种抽样方法下估计精度的比较·····	(44)
本章主要参考文献·····	(47)
第三章 参数估计方法与抽样估计精度·····	(49)
第一节 参数估计方法概述·····	(49)
一、参数估计方法所体现的估计精度的思想·····	(49)
二、两类参数估计方法及其特点·····	(50)
三、参数估计方法选择的原则·····	(51)
第二节 比率估计法的估计精度分析·····	(53)
一、比率估计量及其方差·····	(53)
二、比率估计法与简单估计法的效果比较·····	(55)
三、比率估计量偏差对估计精度的影响·····	(56)
第三节 回归估计法的估计精度分析·····	(59)
一、回归估计量及其方差·····	(59)
二、回归估计法与简单估计法、比率估计法的效果比较 ·····	(60)
三、回归估计量偏差对估计精度的影响·····	(62)
四、 $b \neq \beta$ 对估计精度的影响·····	(63)

第四节 小样本时复合估计量的均方误差分析	(64)
第五节 各种参数估计方法的结合应用与抽样估计精度	(70)
本章主要参考文献	(73)
第四章 抽样组织形式与抽样估计精度(上)	(75)
第一节 抽样组织形式概述	(75)
一、抽样组织形式所体现的估计精度的思想	(75)
二、各种抽样组织形式的原理与特点	(76)
第二节 分层抽样估计精度分析	(79)
一、分层抽样所体现的估计精度的思想	(79)
二、分层抽样的估计量方差及其与纯随机抽样的比较	(80)
三、各层应抽个体数的确定方法与抽样估计精度	(83)
四、层权偏差对抽样估计精度的影响	(86)
五、分层标志的选择与抽样估计精度	(88)
六、分层数与抽样估计精度	(90)
七、考虑费用增加时分层抽样必要性的精度分析	(91)
八、分层抽样三种精度控制方法的比较	(94)
第三节 等距抽样估计精度分析	(97)
一、等距抽样所体现的估计精度的思想	(97)
二、等距抽样的估计量方差及其与纯随机抽样的比较	(97)
三、等距抽样与分层抽样的效果比较	(101)
四、无序等距抽样与纯随机抽样的效果比较	(103)
五、一个例子	(104)
本章主要参考文献	(106)
第五章 抽样组织形式与抽样估计精度(下)	(109)

第一节 整群抽样估计精度分析	(109)
一、整群抽样所体现的估计精度的思想	(109)
二、整群抽样的估计量方差及其与纯随机抽样的比较	(110)
三、不等群等概率抽样效果分析	(113)
四、不等群不等概率抽样效果分析	(116)
五、群的大小对抽样估计精度的影响	(118)
第二节 多阶段抽样估计精度分析	(120)
一、多阶段抽样所体现的估计精度的思想	(120)
二、多阶段抽样的估计量方差及其与其它抽样组织形式的比较	(121)
三、不考虑费用因素的两阶段抽样最优解	(125)
第三节 多种抽样组织形式结合应用的估计精度分析	(127)
一、抽样后分层估计	(127)
二、分层等距抽样与等距分层抽样	(130)
三、分层整群抽样与整群抽样后分层	(130)
四、分层多阶段抽样	(132)
本章主要参考文献	(133)
第六章 辅助信息的利用与抽样估计精度	(135)
第一节 辅助信息的利用方式与抽样估计精度	(135)
一、辅助信息仅用于分层或仅用于比率估计	(136)
二、辅助信息既用于分层也用于比率估计	(138)
三、相关的几个问题	(139)
第二节 利用初样本辅助信息的估计精度分析	(140)
一、利用初样本辅助信息的分层抽样估计精度分析	(141)

二、利用初样本辅助信息的比率估计法估计精度分析	(146)
三、利用初样本辅助信息的回归估计法估计精度分析	(149)
第三节 轮换抽样时以前期样本资料为辅助信息的估计精度分析	(151)
一、估计量及其方差	(151)
二、最优解及其估计精度分析	(153)
第四节 多辅助信息的利用与抽样估计精度	(159)
本章主要参考文献	(166)
第七章 调查费用与抽样估计精度	(168)
第一节 调查费用与抽样估计精度的关系	(168)
一、调查费用函数模型	(168)
二、调查费用与抽样估计精度之间的函数关系	(170)
三、估计精度的费用效率分析	(172)
第二节 费用约束条件下的估计精度最优解	(179)
一、纯随机抽样下的最优解	(179)
二、分层抽样下的最优解	(181)
三、整群抽样下的最优解	(185)
四、多阶段抽样下的最优解	(188)
五、利用初样本辅助信息时的最优解	(190)
本章主要参考文献	(193)
第八章 非抽样误差与抽样估计精度	(194)
第一节 非抽样误差对估计精度影响的特点	(194)
一、抽样调查中的误差构成	(194)
二、非抽样误差的来源	(196)
三、非抽样误差对估计精度影响的特点	(198)

第二节 抽样框误差与抽样估计精度·····	(201)
一、抽样框误差的成因 ·····	(201)
二、抽样框误差对估计精度影响的测定 ·····	(204)
三、抽样框误差的控制 ·····	(215)
第三节 无回答误差与抽样估计精度·····	(216)
一、无回答误差的成因 ·····	(216)
二、无回答对估计精度影响的测定 ·····	(218)
三、无回答误差的控制 ·····	(220)
第四节 计量误差与抽样估计精度·····	(226)
一、计量误差的成因 ·····	(226)
二、计量误差对估计精度的影响 ·····	(228)
三、计量误差的控制 ·····	(230)
本章主要参考文献·····	(233)
后记·····	(236)

第一章 抽样估计精度的一般问题

第一节 抽样估计精度的意义

一、抽样估计精度的定义

抽样估计是指按照随机原则,从总体中抽取部分个体构成样本,然后以样本的调查或观测结果(形成抽样估计值)去估计总体数量特征(即总体参数)的一种推断方法。由于抽样的非全面性,这种估计必然会存在误差,即抽样误差。与此同时,在抽样调查过程中,还可能会存在其他一些误差,如估计量偏差、非抽样误差等。这样在抽样估计中,就存在一个估计误差有多大的问题,也就是我们所要讨论的抽样估计精度有多高的问题。

抽样估计精度即抽样估计精确度,是抽样估计中与估计误差相对应的一个概念。严格意义上讲,抽样估计精度是指在抽样估计中,抽样估计值对其均值的接近程度。这里的均值是指在相同的样本容量下,反复使用同一抽样方法所获得的各个抽样估计值的平均数。

然而,抽样估计的目的是为了对总体参数的真值作出估计,而不是为了获得抽样估计值的均值。我们总是希望抽样估计值尽量接近总体参数真值。因此,我们考虑抽样估计精度问题,必须把抽样估计值与总体参数真值两者联系起来。这样就有另一个与此相关的概念,叫做抽样估计准确度,即抽样估计值对总体参数真值的

接近程度。如果所有可能抽样估计值的均值等于总体参数真值,那么抽样估计精度就等同于抽样估计准确度。如果所有可能抽样估计值的均值不等于总体参数真值,那么抽样估计精度与抽样估计准确度之间就有一个偏差的区别。这个偏差可能由于估计量的不同构造方法所引起,也可能由各种非抽样误差所产生。

若设总体参数真值为 θ , 其抽样估计量为 $\hat{\theta}$ (根据某一样本得到的值即为估计值), 那么有:

$$\hat{\theta} - \theta = [\hat{\theta} - E(\hat{\theta})] + [E(\hat{\theta}) - \theta] \quad (1-1)$$

其中 $\hat{\theta} - \theta$ 即为抽样估计值与总体参数真值之差, $\hat{\theta} - E(\hat{\theta})$ 即为抽样估计值与其均值之差, $E(\hat{\theta}) - \theta$ 即为偏差。

对式(1-1) 等号两边同时平方并求期望, 可得:

$$E(\hat{\theta} - \theta)^2 = E[\hat{\theta} - E(\hat{\theta})]^2 + [E(\hat{\theta}) - \theta]^2 \quad (1-2)$$

其中

$E(\hat{\theta} - \theta)^2$ 称为估计量的均方误差, 常表示为 $MSE(\hat{\theta})$, 是衡量抽样估计准确度的基础指标;

$E[\hat{\theta} - E(\hat{\theta})]^2$ 称为估计量的方差, 常表示为 $V(\hat{\theta})$, 是衡量抽样估计精确度的基础指标;

$[E(\hat{\theta}) - \theta]^2$ 称为偏差的平方。

当 $E(\hat{\theta}) = \theta$ 时, 偏差为 0, $MSE(\hat{\theta}) = V(\hat{\theta})$ 。

如果说 $\hat{\theta}$ 越接近于 θ , 抽样估计越准确, $\hat{\theta}$ 越接近于 $E(\hat{\theta})$, 抽样估计越精确, 那么两者之间的差别就在于偏差的大小。因此, 在抽样估计中, 我们总是希望 $E(\hat{\theta}) = \theta$, 即偏差为 0, 从而使抽样估计准确度问题与抽样估计精确度问题合二为一, 更有利于控制抽样估计误差, 提高估计的准确度。然而, 在抽样估计中, 由于总体参