卢宗辉 著

抽样方法的系统研究

中国统计出版社

抽样方法的系统研究

卢 宗 辉 著

中国统计出版社

（京）新登字 041 号

图书在版编目(CIP)数据

抽样方法的系统研究/卢宗辉著.
一北京：中国统计出版社，1998.6
ISBN 7-5037-2375-0

I. 抽…
II. 卢…
III. 抽样调查一研究
IV. C811
中国版本图书馆 CIP 数据核字(98)第 13193 号

中国统计出版社出版
（北京西城三里河月坛南街 75 号　100826）
新华书店经销
北京市顺义兴华印刷厂印刷
*
850×1168 毫米　32 开本　5.75 印张　11 万字
1998年8月第1版　1998年8月北京第1次印刷
印数：1-1500 册
*
定价：19.60 元

序　言

统计是一门认识方法论的科学，它是通过调查、汇总、综合、分析以达到对客观总体现象规律性的认识。从直观上说，要认识现象总体的数量特征，最理想的方法是对多个单位的标志表现加以一一观察登记，再汇总计算，但现实上却是不容易做到的。属于无限总体的现象、包括未来的时间序列总体、具有破坏性检验的总体等等都无法进行全面调查。有些由于总体范围太大，单位数很多，或由于时间要求的限制，条件不允许等等，进行全面调查虽然不是不可能，却也难以实现。通常我们所能够取得的只是有限的、部分的资料。这里存在着认识上的局部与整体、个体与总体之间的矛盾。抽样方法在它们之间架起了一道桥梁，为从样本认识总体的过渡奠定了科学的基础。基于这个特点，使抽样方法（包括抽样调查和推断）既不同于全面调查也有异于一般的非全面调查。全面调查虽然可以达到对总体的认识，但却没有非全面调查省时省力的好处，而一般非全面调查虽然省时省力，又难以用来推算总体的数量特征。只有抽样方法只付出非全面调查的代价，可以收到全面调查的效果。因此抽样方法的出现和推广大大提高了统计学的认识能力，也为哲学的认识论开辟了一条崭新的途径。挽近半世纪来在西方国家抽样方法的研究已经是统计学中最活跃的分支。

我国抽样调查的起步并不算晚，已有三十多年的历史。但在计划经济体制下，由于经济组织高度集中，经济类型单一，上下关系明确，调查项目相对稳定，各级信息需求一致，所以采用定期报表制度作为检查计划的基本手段，抽样调查充其量只作为辅助方法，活动空间受到限制。在社会主义市场经济条件下，情况发生了根本的变化。社会

各种经济类型同时发展，国家、集体、个人之间的关系更加复杂。而且经济决策多层，经济经营分散，经济变动频繁，国家调控系统、市场经济系统和社会保障系统，对信息的需求各不相同。依靠定期报表制度进行资料搜集，已难以适应新形势的需要。现在国家已建立以抽样调查为主体的统计调查方法体系，城市、乡村两抽样调查队逐步扩大调查项目和调查内容，并组建企业抽样调查队，把抽样方法有步骤地推广到商业、饮食、工业、交通、投资以及广播电视各个领域，这正显示着我国统计开始沿着现代化道路阔步前进。

但是，相对来说我国抽样理论和方法的研究还是滞后于形势。首先，抽样方法比较单一，不够灵活。联合国早在1960年就出版《统计抽样方法简要手册》第一卷《抽样调查理论基础》，向世界各国政府推荐行之有效的22种抽样方法。1972年修订本推荐方法增至32种。80年代以后联合国统计部门推广在普查中应用抽样技术，出版《工业普查实用抽样技术》、《农业普查的抽样方法》等。外国的成功经验我们应当吸收借鉴。其次，从我国统计实践中提出的问题，更宜加紧研究解决。例如抽样调查网点的合并与多主题抽样研究，样本轮换与动态抽样研究，多阶段抽样方法研究，不等概和不等群抽样方法研究，抽样与非抽样误差的度量和控制，总量指标的估计和统计分析等等。我们期望我国不仅是广泛应用抽样方法的大国，而且还是研究抽样理论成果显赫的大国。

我很高兴看到卢宗辉博士的新著《抽样方法的系统研究》一书出版，为我国抽样理论与方法的研究作出可称赞的贡献。作者长期从事抽样问题研究并参与广泛的抽样实践，涉及农业、工业、交通运输各个领域，先后发表《我国公路货运抽样方法研究》、《我国非交通系统民用载货汽车抽样调查修订方案抽样误差公式之我见》、《我国农产量抽样调查方法研究》、《论有关标志等距抽样下的总体推估问题》、《目录抽样刍议》等论文都很受欢迎。《抽样方法的系统研究》一书则是作者在博士论文的基础上，集中体现他的研究成果。我感到有以下

突出特点：

1. 研究的起点比较高。作者总是悉心研究国内外最新的研究成果，分析比较其异同，然后融会贯通，作出自己的论断。例如作者指出抽样误差的大小既与等概、非等概抽样方式有关，又与不同估计方法相联系，于是列出无偏估计、比率估计、回归估计六种抽样误差公式，经过比较之后令人信服地提出确定抽样方法的两个参数。又如作者对系统抽样研究，比较国外八种不同的抽样方法，然后归纳出选取系统抽样的若干原则，并得出我国应大力推广高效的麦多系统抽样的结论。读之得益良多。

2. 问题的针对性比较强。《抽样方法的系统研究》是一部学术著作，但所提出的问题却都是来自现实。例如多阶段抽样问题研究，作者分析了不等群的情况下，群体的变化对抽样方差的影响，并提出用群内相关系数来确定最优样本群的大小。又如样本轮换问题研究，作者在分析样本轮换的必要性基础上，提出样本轮换时间和样本轮换率的确定方法。这些都是当前抽样工作中急需解决的难题。

3. 严格的逻辑推理与直观的例证说明相结合。由于作者数学功底好，对于抽样方法的数量关系、抽样误差的影响因素、以及抽样效果分析优化求解等等都能加以严格的数学推导，但同时又辅以实例求证和直观说明，讲透问题的实质。即使数学程度不高的读者，读后仍能领会其思路的来龙去脉，这也是本书的一大优点。

当然，这是一本相当浓缩而又严肃的学术著作，有理由相信它能发挥举一反三的作用。我期盼本书的出版会吸引更多的同行专家和实际工作者参加这一研究行列，把我国统计工作提高到新的水平。

黄良文
1998 年 3 月
于厦门大学白城

前　言

建立社会主义市场经济体制历史地把抽样调查推到了统计调查方法体系的主体地位，抽样调查这一世界公认的统计调查方法也因此而在我国受到了前所未有的重视，关于抽样调查和国民经济核算方面问题的研究无疑已经并将在较长时期内成为我国统计研究领域的两大核心课题。鉴于此，多年来我就一直以抽样调查研究作为自己的主攻课题。在学习和参阅有关文献的过程中，我深感我国抽样理论研究相当薄弱，不仅落后于国外，而且本身也滞后于我国的抽样实践，既不能很有效地解决抽样实践中出现的问题，也适应不了我国日益拓宽抽样应用领域的迫切需要。之所以如此，原因是多方面的，但我认为其根本原因在于对抽样方法缺乏系统深入的研究，主要表现在以下七个方面：1. 在抽样的基本问题上，往往只讨论无偏估计下等概率抽样问题，而没有对各种总体推估方式与样本抽选方式相结合形成的多种抽样方法进行研究比较，从而使人们在实践中无法对总体推估方式和样本抽选方式进行最优的选择。2. 在系统抽样上，抽样方差的估算往往用简单随机抽样方差公式，各种提出的估算方法在理论上存在缺陷，尤其对各种系统抽样方法缺乏比较研究，使人们没法把握和明了对某项具体的抽样调查应采用哪种系统抽样方法为好。3. 在整群抽样上，往往把群不等的常见情况当成群相等的特殊情况来处理，没有考虑群不等对抽样误差的影响，对客观现象作了不切实际的抽象拟合。4. 在分层抽样上往往侧重最简单的分层，对较为深入的分层问题研究甚少。5. 在多阶段抽样上，

虽然我国普遍采用多阶段抽样，但对多阶段抽样的研究又出奇地少，即使有所涉及，也往往只讨论实际中可能性不大的最简单的群大小相等情况，尤其是在多阶段抽样方差的计算、各阶段最优样本群大小的确定方面更是如此，这使人们在具体的抽样实践中很难灵活正确地选择或运用合适的抽样方差公式来计算实际的抽样误差及事先确定出各阶段的样本容量。6. 在样本轮换方面，我国虽然实行了固定的样本轮换制度，但对样本轮换理论的研究明显不够，即使讨论样本轮换率也往往没有考虑抽样的总体推估方式，尤其是没有考虑样本轮换后的总体推估及其抽样方差的计算问题。7. 在非抽样误差上，我国虽然在个别误差上进行了研究，但很不全面、系统，尤其是在非抽样误差的估计与处理方法上仍非常欠缺，对重复抽样中出现的回答方面问题如何处理更是一片空白。对以上七个方面的研究就构成了本书的全部内容。本书旨在探究抽样方法中的各种“为什么?”，以深化我国抽样理论的研究，提高抽样理论对我国抽样实践的指导能力及其用于解决实际抽样问题的能力。

本书以博士论文为雏形，经过进一步提高、充实和提炼而成，期间得到了恩师黄良文教授和陈仁恩教授的大力支持与帮助。黄良文老师在指导我撰写博士论文、进行学术研究方面都给予了新的启迪和指教，使我受益匪浅。最后，借本书出版之机，谨向所有关心和帮助过我的老师、领导和朋友们表示衷心的感谢。

卢宗辉

1998 年 5 月

目　　录

第一章　抽样的基本问题研究

抽样(指概率抽样)可以分为等概率抽样和不等概率抽样，等概率抽样又可以分为简单随机抽样、分层等概率抽样、等概率整群抽样、等概率系统抽样和等概率多阶段抽样(即我们通常所讲的简单随机抽样、分层抽样、整群抽样、系统抽样和多阶段抽样)，不等概率抽样也可以同样分为简单不等概率抽样(即通常所讲的不等概率抽样)、分层不等概率抽样、不等概率整群抽样、不等概率系统抽样和多阶段不等概率抽样。但尽管如此，我们不难看出，这些抽样方式不论是等概率的还是不等概率的，都不过是简单抽样方式——简单随机抽样和简单不等概率抽样的变形或排列组合而已。如等概率整群抽样就是以群为最终抽样单位的简单随机抽样，分层等概率抽样就是几种简单随机抽样的并列，等概率系统抽样不论是看作分层抽样还是看成整群抽样，也都是简单随机抽样，等概率多阶段抽样是以上这些等概率抽样方式的纵横排列组合，归根到底也是简单随机抽样。不等概率抽样也同样可归结为简单不等概率抽样。所以，抽样的基本问题就是简单随机抽样和简单不等概率抽样的问题，对它们进行研究不论对抽样理论还是对抽样实践都具有基础意义和指南作用。

一、抽样方法的比较与选择

抽样从方法上讲包括两个方面：样本抽选和总体推估。样本抽

选指样本的抽选方式是等概率还是不等概率，重复还是不重复。总体推估是指总体推估方式是简单平均数推估还是比率推估或回归推估。由于样本的抽选与总体的推估方式是紧密结合在一起的，所以，下面综合考虑这两者进行探讨。为方便起见，假定样本都按重复抽样的方式抽取。

1. **必须注意的问题**

总体推估方式的选取是抽样中一个十分重要的方面，它关系到应选择哪个抽样误差公式来计算抽样误差的问题。因此，我认为很有必要区分两种极易混淆的总体推估方式：比率估计和无偏估计。这两者的区分是针对抽样单位的数目而言的，只有以抽样单位的个数计算的平均数进行的总体推估才是平均数推估，以抽样单位的其他标志值计算的样本值进行的总体推估就是比率推估。这两者在实践中常常容易被弄混淆，如《我国非交通系统民用载货汽车抽样调查修订方案》中规定的最终抽样单位是车队，却把按样本车队的车吨货物周转量进行的本属比率估计的总体推估当成了平均数推估，以致在抽样误差的计算上出现了问题。

2. **抽样方差公式的导出**

设总体线性回归方程为 $Y_i=\alpha+\beta X_i+\varepsilon_i$，Y 为调查变量，X 为辅助变量，$\varepsilon_i$ 为误差。由于对许多总体而言，$|\varepsilon_i|$和 ε_i^2 都随总体单位的大小测度 Z_i 的增加而增大，因此，我们可以认为 $|\varepsilon_i|$和 ε_i^2 是单位大小测度 Z_i 的某种函数。鉴于 ε_i 取决于特定的总体，相对 Z_i 增加，ε_i 也趋于增加但一般稍低于成比例，因此，可以认为这种函数的形式是 KZ_i^{2r}，从而取 $E(\varepsilon_i)=0$，$E(\varepsilon_i^2)=KZ_i^{2r}$，式中 K 为适当的常数，r 为异方差测度，它反映点(Y_i, Z_i)围绕其回归线的散布程度。

在简单不等概率抽样下，无偏估计、比率估计和回归估计分别表示为

$$\hat{Y}_P=\sum_{i=1}^{n}\frac{y_i}{nP_i}\qquad \text{（}P_i\text{ 为第 i 个样本单位的抽中概率）}$$

$$\hat{Y}_{PR}=\frac{\hat{Y}_P}{\hat{X}_P}X$$

$$\hat{Y}_{PL}=(\alpha+\beta\hat{X}_P)N$$

相应的抽样方差分别推导如下：

① 对无偏估计 $\hat{Y}_P=\sum_{i=1}^{n}\frac{y_i}{nP_i}$

$$\begin{aligned}
V(\hat{Y}_P)&=\frac{N^2}{n}\sum_{i=1}^{N}P_i\left(\frac{Y_i}{NP_i}-\overline{Y}\right)^2\\
&=\frac{N^2}{n}E\sum_{i=1}^{N}P_i\left(\frac{\alpha+\beta Z_i+\varepsilon_i}{NP_i}-\alpha-\beta\,\overline{Z}-\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i\right)^2\\
&=\frac{N^2}{n}E\sum_{i=1}^{N}P_i\left[\alpha\left(\frac{1}{NP_i}-1\right)+\beta\left(\frac{Z_i}{NP_i}-\overline{Z}\right)+\frac{1}{N}\left(\frac{\varepsilon_i}{P_i}-\sum_{i=1}^{N}\varepsilon_i\right)\right]^2\\
&=\frac{N^2}{n}\left\{\sum_{i=1}^{N}P_i\left[\alpha\left(\frac{1}{NP_i}-1\right)+\beta\left(\frac{Z_i}{NP_i}-\overline{Z}\right)\right]^2\right.\\
&\qquad\left.+\frac{1}{N^2}\sum_{i=1}^{N}P_i\left(\frac{\varepsilon_i}{P_i}-\sum_{i=1}^{N}\varepsilon_i\right)^2\right\}
\end{aligned}$$

$$\because\ E\left(\frac{\varepsilon_i}{P_i}-\sum_{i=1}^{N}\varepsilon_i\right)=0\qquad E\varepsilon_i\varepsilon_j=0$$

$$\begin{aligned}
E\sum_{i=1}^{N}P_i\left(\frac{\varepsilon_i}{P_i}-\sum_{i=1}^{N}\varepsilon_i\right)^2&=E\sum_{i=1}^{N}\left[\frac{\varepsilon_i^2}{P_i}-2\varepsilon_i\sum_{i=1}^{N}\varepsilon_i+P_i\left(\sum_{i=1}^{N}\varepsilon_i\right)^2\right]\\
&=E\left[\frac{1}{P_i}\sum_{i=1}^{N}\varepsilon_i^2-2\left(\sum_{i=1}^{N}\varepsilon_i\right)^2+\sum_{i=1}^{N}P_i\left(\sum_{i=1}^{N}\varepsilon_i\right)^2\right]\\
&=E\left[\frac{1}{P_i}\sum_{i=1}^{N}\varepsilon_i^2-\left(\sum_{i=1}^{N}\varepsilon_i\right)^2\right]\\
&=E\left[\frac{1}{P_i}\sum_{i=1}^{N}\varepsilon_i^2-\sum_{i=1}^{N}\varepsilon_i^2\right]
\end{aligned}$$

$$=\sum_{i=1}^{N}E\varepsilon_i^2\left(\frac{1}{P_i}-1\right)$$

$$=K\sum_{i=1}^{N}Z_i^{2r}\left(\frac{1}{P_i}-1\right)$$

$$\therefore\ V(\hat{Y}_P)=\frac{N^2}{n}\left\{\sum_{i=1}^{N}P_i\left[\alpha\left(\frac{1}{NP_i}-1\right)+\beta\left(\frac{Z_i}{NP_i}-\bar{Z}\right)\right]^2\right.$$

$$\left.+\frac{K}{N^2}\sum_{i=1}^{N}Z_i^{2r}\left(\frac{1}{P_i}-1\right)\right\} \qquad (1.1)$$

当样本单位是按照与大小测度 Z_i 成比例的概率抽取时，$P_i=Z_i/Z$，则在 N 很大时，可以认为 $N-1\doteq N$，$\sum_{i=1}^{N}\varepsilon_i=0$，$[(1/P_i)-1]\doteq(1/P_i)$，从而(1.1)式变为

$$V(\hat{Y}_P)\doteq\frac{N^2}{n}\left[\alpha^2\sum_{i=1}^{N}P_i\left(\frac{1}{P_i}-N\right)^2+\frac{K}{N^2}\sum_{i=1}^{N}\frac{1}{P_i}Z_i^{2r}\right]$$

$$=\frac{N^2}{n}\left[\alpha^2\frac{1}{N^2}\sum_{i=1}^{N}\left(\frac{1}{P_i}-2N+P_iN^2\right)+\frac{K}{N^2}\bar{Z}\sum_{i=1}^{N}Z_i^{2r-1}\right]$$

$$=\frac{N^2}{n}\left[\alpha^2\frac{1}{N^2}\sum_{i=1}^{N}\left(\frac{1}{P_i}-N\right)+\frac{K}{N}\bar{Z}\sum_{i=1}^{N}Z_i^{2r-1}\right]$$

$$=\frac{N^2}{n}\left[\alpha^2\bar{Z}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{Z_i}-\frac{1}{\bar{Z}}\right)+\frac{K}{N}\bar{Z}\sum_{i=1}^{N}Z_i^{2r-1}\right] \qquad (1.2)$$

② 对比率估计 $\hat{Y}_{PR}=\frac{\hat{Y}_P}{\hat{X}_P}X$，有

$$\hat{Y}_{PR}=\frac{N^2}{n}\delta_{PR}^2$$

$$=\frac{N^2}{n}\sum_{i=1}^{N}P_i\left[\left(\frac{Y_i}{NP_i}-R\frac{X_i}{NP_i}\right)\right]^2$$

$$=\frac{N^2}{n}E\sum_{i=1}^{N}\frac{1}{N^2P_i}(\alpha+\beta X_i+\varepsilon_i-RX_i)^2$$

$$\doteq\frac{N^2}{n}\cdot\frac{1}{N^2}\sum_{i=1}^{N}\frac{1}{P_i}E[\alpha+(\beta-R)X_i+\varepsilon_i]^2$$

$$=\frac{N^2}{n}\cdot\frac{1}{N^2}\sum_{i=1}^{N}\frac{1}{P_i}E\left[\frac{\alpha}{\overline{X}}(\overline{X}-X_i)+\varepsilon_i\right]^2$$

$$\left(N\text{ 很大时，}\beta-R=-\frac{\alpha}{\overline{X}}-\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i\approx-\frac{\alpha}{\overline{X}}\right)$$

$$=\frac{N^2}{n}\left[\frac{1}{N^2}\frac{\alpha^2}{\overline{X}^2}\sum_{i=1}^{N}\frac{1}{P_i}(X_i-\overline{X})^2+\frac{1}{N^2}\sum_{i=1}^{N}E\frac{1}{P_i}\varepsilon_i^2\right]$$

$$=\frac{N^2}{n}\left[\frac{1}{N^2}\frac{\alpha^2}{\overline{X}^2}\sum_{i=1}^{N}\frac{1}{P_i}(X_i-\overline{X})^2+\frac{K}{N^2}\sum_{i=1}^{N}\frac{1}{P_i}Z_i^{2r}\right] \tag{1.3}$$

若 α 很小，可以忽略，则可以得出比率估计的近似抽样方差

$$V(\hat{Y}_{PR})=\frac{K}{n}\sum_{i=1}^{N}\frac{Z_i^{2r}}{P_i} \tag{1.4}$$

在样本单位是按与大小成比例的概率抽取时，$P_i=Z_i/Z$，则(1.3)式就变为

$$V(\hat{Y}_{PR})=\frac{N^2}{n}\left[\frac{1}{N^2}\frac{\alpha^2}{\overline{X}^2}\sum_{i=1}^{N}\frac{Z}{Z_i}(X_i-\overline{X})^2+\frac{K}{N^2}\sum_{i=1}^{N}\frac{Z}{Z_i}Z_i^{2r}\right]$$

$$=\frac{N^2}{n}\left[\frac{\alpha^2}{\overline{X}^2}\overline{Z}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{Z_i}(X_i-\overline{X}_i)^2+\frac{K}{N}\overline{Z}\sum_{i=1}^{N}Z_i^{2r-1}\right] \tag{1.5}$$

③ 对回归估计 $\hat{Y}_{PL}=(\alpha+\beta\hat{X}_P)N$，在(1.3)式中令 $\alpha=0$ 即可得其抽样方差

$$V(\hat{Y}_{PL})=\frac{N^2}{n}\left[\frac{K}{N}\overline{Z}\sum_{i=1}^{N}Z_i^{2r-1}\right] \tag{1.6}$$

对简单随机抽样的三种推估方式 $\hat{Y}_s=\frac{\sum_{i=1}^{n}X_i}{n}N$(无偏估计)，$\hat{Y}_{SR}=\hat{R}X=\frac{\hat{Y}_s}{\hat{X}_s}X$(比率估计)和 $\hat{Y}_{SL}=(\alpha+\beta\hat{X}s)N$(回归估计)，在(1.1)式、(1.3)式和(1.6)式中令 $P_i=1/N$ 即可得相应的抽样方差：

$$V(\hat{Y}_s)\doteq\frac{N^2}{n}\beta^2\frac{\sum_{i=1}^{N}(Z_i-\overline{Z})^2}{N}+\frac{K}{N}\sum_{i=1}^{N}Z_i^{2r}$$

$$=\frac{N^2}{n}\left(\beta^2\delta_Z^2+\frac{K}{N}\sum_{i=1}^{N}Z_i^{2r}\right)$$

$$V(\hat{Y}_{SR})\doteq\frac{N^2}{n}\left[\frac{\alpha^2}{\overline{X}^2}\frac{\sum_{i=1}^{N}(X_i-\overline{X})^2}{N}+\frac{K}{N}\sum_{i=1}^{N}Z_i^{2r}\right]$$

$$=\frac{N^2}{n}\left(\frac{\alpha^2}{\overline{X}^2}\delta_X^2+\frac{K}{N}\sum_{i=1}^{N}Z_i^{2r}\right)$$

$$V(\hat{Y}_{SL})=\frac{N^2}{n}\left(\frac{K}{N}\sum_{i=1}^{N}Z_i^{2r}\right)$$

3. 抽样方法的比较和选择

根据上面的推导，我们把两种简单抽样下各种推估方式的抽样方差公式列表表示如下

简单抽样下的抽样方差

	简单随机抽样	简单不等概率抽样(PPS)
无偏估计	$\frac{N^2}{n}\left(\beta^2\delta_Z^2+\frac{K}{N}\sum_{i=1}^{N}Z_i^{2r}\right)$	$\frac{N^2}{n}\left[\alpha^2\overline{Z}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{Z_i}-\frac{1}{\overline{Z}}\right)^2+\frac{K}{N}\overline{Z}\sum_{i=1}^{N}Z_i^{2r-1}\right]$
比率估计	$\frac{N^2}{n}\left(\frac{\alpha^2}{\overline{X}^2}\delta_X^2+\frac{K}{N}\sum_{i=1}^{N}Z_i^{2r}\right)$	$\frac{N^2}{n}\left[\frac{\alpha^2}{\overline{X}^2}\overline{Z}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{Z_i}(X_i-\overline{X})^2+\frac{K}{N}\overline{Z}\sum_{i=1}^{N}Z_i^{2r-1}\right]$
回归估计	$\frac{N^2}{n}\left(\frac{K}{N}\sum_{i=1}^{N}Z_i^{2r}\right)$	$\frac{N^2}{n}\left(\frac{K}{N}\overline{Z}\sum_{i=1}^{N}Z_i^{2r-1}\right)$

从这些公式中可以看出以下几点：

① 除简单随机抽样下的无偏估计和回归估计外，其他估计量的方差中都包括α项。所以，在比率估计和按与大小成比例抽样的不等概率抽样(PPS)中，调查变量 Y_i 与辅助变量 X_i 的关系越密切地

成比例，对应的抽样方差就越小。

② 若$\alpha=0$，则回归估计与比率估计等效率，否则回归估计优于比率估计，且在不等概率抽样下，$\alpha=0$使三种总体推估方式效率相同。

③ 若X_i就是Z_i，则只要$\beta^2>(\alpha^2/\overline{X}^2)$即$\rho_{XY}\geqslant(C_X/2C_Y)$，C表示变异系数，则简单随机抽样下的比率估计比无偏估计有效；简单不等概率抽样下的无偏估计与比率估计效率差不多；$r\geqslant0.5$，不等概率抽样无偏估计比等概率抽样比率估计有效(加上$\beta^2>\alpha^2/\overline{X}^2$，则也比等概率抽样无偏估计有效)，不等概率抽样比率估计也比等概率抽样比率估计有效。证明如下：

∵ 不等概率抽样无偏估计和等概率抽样比率估计的抽样方差在$X_i=Z_i$下分别可写为

$$V(\hat{Y}_P)=\frac{N^2}{n}\left[\frac{\alpha^2}{\overline{Z}^2}\overline{Z}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{Z_i}(Z_i-\overline{Z})^2+\frac{K}{N}\overline{Z}\sum_{i=1}^{N}Z_i^{2r-1}\right]$$

$$V(\hat{Y}_{SR})=\frac{N^2}{n}\left[\frac{\alpha^2}{\overline{Z}^2}\frac{1}{N}\sum_{i=1}^{N}(Z_i-\overline{Z})^2+\frac{K}{N}\sum_{i=1}^{N}Z_i^{2r}\right]$$

$$\therefore V(\hat{Y}_P)-V(\hat{Y}_{SR})=\frac{N^2}{n}\left\{\frac{\alpha^2}{\overline{Z}^2}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{NP_i}(Z_i-\overline{Z})^2-\frac{1}{N}\sum_{i=1}^{N}(Z_i-\overline{Z})^2\right]+\frac{K}{N}\left(\overline{Z}\sum_{i=1}^{N}Z_i^{2r-1}-\sum_{i=1}^{N}Z_i^{2r}\right)\right\}$$

$$\because NP_i\geqslant1\qquad\therefore\frac{1}{N}\sum_{i=1}^{N}\frac{1}{NP_i}(Z_i-\overline{Z})^2\leqslant\frac{1}{N}\sum_{i=1}^{N}(Z_i-\overline{Z})^2$$

又 ∵ $r\geqslant0.5$时，Z_i与Z_i^{2r-1}的协方差是正数或零。

$$\therefore E(Z_i^{2r})=E(Z_i\cdot Z_i^{2r-1})\geqslant\overline{Z}E(Z_i^{2r-1})$$

即 $$\frac{\sum_{i=1}^{N}Z_i^{2r}}{N}\geqslant\overline{Z}\,\frac{\sum_{i=1}^{N}Z^{2r-1}}{N}$$

故：$V(\hat{Y}_P)-V(\hat{Y}_{SR})\leqslant0$，　　即$V(\hat{Y}_P)\leqslant V(\hat{Y}_{SR})$

不等概率抽样无偏估计比等概率抽样比率估计有效。又因为不等概