第 20 卷第 4 期
2001 年 7 月

无锡轻工大学学报
Journal of Wuxi University of Light Industry

Vol.20 No.4
Jul. 2001

文章编号:1009-038X(2001)04-0416-02

整群比估计与 PPS 抽样的比较

张荷观
(无锡轻工大学 商学院,江苏 无锡 214036)

摘 要:对整群比估计与 PPS 抽样方法进行了比较,给出了整群比估计优于 PPS 抽样的条件.

关键词:整群抽样;比估计;PPS 抽样
中图分类号:O212.2 文献标识码:A

Comparison of the Ratio Estimates in Cluster Sampling and the PPS Sampling

ZHANG He-guan
(School of Business, Wuxi University of Light Industry, Wuxi 214036, China)

Abstract: In this paper, the ratio estimates in cluster sampling with the PPS sampling were compared and the condition that the ratio estimator is superior to PPS estimator was obtained in the paper.

Key words: cluster sampling; ratio estimator; PPS sampling

1 引言

在群大小不等时的整群抽样时,为了提高估计精度,常采用以群大小为辅助变量的比估计以及与群大小成比例的不等概抽样. W. G. 科克伦等[1]对这两种估计的精度作了比较,作者则进一步给出了比估计优于 PPS 抽样以及 PPS 抽样优于比估计的条件.

设总体的 N 个单元划分为 K 个群,第 i 群包含 M_i 个单元,Y_{ij} 为第 i 个群中第 j 个单元的标志值,$i=1,2,\cdots,K$;$j=1,2,\cdots,M_i$;$N=\sum_{i=1}^{K}M_i$. 记为:

$$\overline{Y}_i=\frac{1}{M_i}\sum_{j=1}^{M_i}Y_{ij},$$

$$\overline{\overline{Y}}_i=\frac{1}{N}\sum_{i=1}^{K}\sum_{j=1}^{M_i}Y_{ij}=\sum_{i=1}^{K}\frac{M_i}{N}\overline{Y}_i=\sum_{i=1}^{K}W_i\overline{Y}_i$$

对群进行不等概率抽样时,常采用按与 M_i 成比例的概率,从 K 个群中放回地抽取 k 个的 PPS 抽样,总体均值 $\overline{\overline{Y}}$ 的估计量及估计量的方差为[2,3]:

$$\overline{\overline{y}}_{PPS}=\frac{1}{k}\sum_{i=1}^{k}\overline{Y}_i$$
$$V(\overline{\overline{y}}_{PPS})=\frac{1}{k}\sum_{i=1}^{K}W_i(\overline{Y}_i-\overline{\overline{Y}})^2 \tag{1}$$

若把 M_i 作为辅助变量而采用放回的比估计,那么总体均值 $\overline{\overline{Y}}$ 的估计量及估计量均方误差的近似为[2,3]:

收稿日期:2001-01-15; 修订日期:2001-05-20.
作者简介:张荷观(1949-),男,江苏吴江人,教授.

$$\overline{\overline{y}}_R = \frac{\sum_{i=1}^{k} M_i \overline{Y}_i}{\sum_{i=1}^{k} M_i} \tag{2}$$

$$V(\overline{\overline{y}}_R) = \frac{K}{k}\sum_{i=1}^{K} W_i^2(\overline{Y}_i - \overline{\overline{Y}})^2$$

为了方便起见,作者仅对放回抽样方法进行了讨论.

2　比估计与 PPS 抽样的比较

定理 1　设 $W_i>0$, $W_i(\overline{Y}_i-\overline{\overline{Y}})^2>0$, $i=1,2,\cdots,K$. 若 $W_i\leqslant W_{i+1}$ 时, $W_i(\overline{Y}_i-\overline{\overline{Y}})^2\geqslant W_{i+1}(\overline{Y}_{i+1}-\overline{\overline{Y}})^2$, 或 $W_i\geqslant W_{i+1}$ 时, $W_i(\overline{Y}_i-\overline{\overline{Y}})^2\leqslant W_{i+1}(\overline{Y}_{i+1}-\overline{\overline{Y}})^2$, $i=1,2,\cdots,K-1$, 则

$$V(\overline{\overline{y}}_R) < V(\overline{\overline{y}}_{PPS}) \tag{3}$$

证　$W_i>0$, $W_i(\overline{Y}_i-\overline{\overline{Y}})^2>0$, $i=1,2,\cdots,K$. 根据契贝谢夫不等式[4], 当 $W_i\leqslant W_{i+1}$ 时, $W_i(\overline{Y}_i-\overline{\overline{Y}})^2\geqslant W_{i+1}(\overline{Y}_{i+1}-\overline{\overline{Y}})^2$, 或 $W_i\geqslant W_{i+1}$ 时, $W_i(\overline{Y}_i-\overline{\overline{Y}})^2\leqslant W_{i+1}(\overline{Y}_{i+1}-\overline{\overline{Y}})^2$, $i=1,2,\cdots,K-1$, 则

$$\begin{aligned} V(\overline{\overline{y}}_R) &= \frac{K^2}{k}\left[\frac{1}{K}\sum_{i=1}^{K} W_i^2(\overline{Y}_i-\overline{\overline{Y}})^2\right] < \\ &\frac{K^2}{k}\left(\frac{1}{K}\sum_{i=1}^{K} W_i\right)\left[\frac{1}{K}\sum_{i=1}^{K} W_i(\overline{Y}_i-\overline{\overline{Y}})^2\right] = \\ &\frac{1}{k}\sum_{i=1}^{K} W_i(\overline{Y}_i-\overline{\overline{Y}})^2 = V(\overline{\overline{y}}_{PPS}) \end{aligned}$$

定理 1 表明,当 M_i 愈大而相应的 $M_i(\overline{Y}_i-\overline{\overline{Y}})^2$ 愈小时,则采用比估计会优于 PPS 抽样.

定理 2　$W_i>0$, $W_i(\overline{Y}_i-\overline{\overline{Y}})^2>0$, $i=1,2,\cdots,K$. 若 $W_i\leqslant W_{i+1}$ 时, $W_i(\overline{Y}_i-\overline{\overline{Y}})^2\leqslant W_{i+1}(\overline{Y}_{i+1}-\overline{\overline{Y}})^2$, 或 $W_i\geqslant W_{i+1}$ 时, $W_i(\overline{Y}_i-\overline{\overline{Y}})^2\geqslant W_{i+1}(\overline{Y}_{i+1}-\overline{\overline{Y}})^2$, $i=1,2,\cdots,K-1$, 则

$$V(\overline{\overline{y}}_R) \geqslant V(\overline{\overline{y}}_{PPS}) \tag{4}$$

证　同样由于 $W_i>0$, $W_i(\overline{Y}_i-\overline{\overline{Y}})^2>0$, $i=1,2,\cdots,K$. 而根据契贝谢夫不等式,若 $W_i\leqslant W_{i+1}$ 时, $W_i(\overline{Y}_i-\overline{\overline{Y}})^2\leqslant W_{i+1}(\overline{Y}_{i+1}-\overline{\overline{Y}})^2$, 或 $W_i\geqslant W_{i+1}$ 时, $W_i(\overline{Y}_i-\overline{\overline{Y}})^2\geqslant W_{i+1}(\overline{Y}_{i+1}-\overline{\overline{Y}})^2$, $i=1,2,\cdots,K-1$, 则

$$\begin{aligned} V(\overline{\overline{y}}_R) &= \frac{K^2}{k}\left[\frac{1}{K}\sum_{i=1}^{K} W_i^2(\overline{Y}_i-\overline{\overline{Y}})^2\right] \geqslant \\ &\frac{K^2}{k}\left(\frac{1}{K}\sum_{i=1}^{K} W_i\right)\left[\frac{1}{K}\sum_{i=1}^{k} \overline{W}_i(\overline{Y}_i-\overline{\overline{Y}})^2\right] = \\ &\frac{1}{k}\sum_{i=1}^{K} W_i(\overline{Y}_i-\overline{\overline{Y}})^2 = V(\overline{\overline{y}}_{PPS}) \end{aligned}$$

定理 2 表明,若 M_i 愈小时而相应的 $\overline{Y}_i$ 愈接近 $\overline{\overline{Y}}$,则采用 PPS 抽样会优于比估计.

3　结　语

对群大小不等的整群抽样,当 M_i 差别较大时,采用简单随机抽样估计的精度一般较低. 这时可采用按与 M_i 成比例的不等概抽样,通常可以有效地提高估计精度. 但在简单随机抽样时,也可以把 M_i 作为辅助变量而采用比估计. 作者给出了比估计优于 PPS 抽样以及 PPS 抽样优于比估计的条件,这对于在抽样调查工作中如何根据总体的实际情况来选择比估计与 PPS 抽样方法提供了理论依据.

作者为了方便讨论,只给出了放回抽样时的结论.

参考文献:

[1] W. G. 科克伦. 抽样技术[M]. 张尧庭,吴辉译. 北京:中国统计出版社,1985.
[2] 冯士雍,施锡铨. 抽样调查——理论,方法与实践[M]. 上海:上海科学技术出版社,1996.
[3] 冯小筠,祝大平. 抽样调查的方法和原理[M]. 上海:华东师范大学出版社,1994.
[4] D. S. 密特利诺维奇. 解析不等式[M]. 张小萍,王龙译. 北京:科学出版社,1987.

(责任编辑:李春丽)