第 19 卷 第 1 期
2002 年 02 月
工 程 数 学 学 报
JOURNAL OF ENGINEERING MATHEMATICS
Vol. 19 No. 1
Feb. 2002

文章编号: 1005-3085(2002) 01-0080-09

整群抽样回归估计量*

俞纯权
(山东经济学院统计系, 济南 250014)

摘 要: 研究了有辅助指标的整群抽样, 构造了总体总值的回归估计量, 证明了回归估计量是可用的, 并且不劣于整群抽样下的简单估计量及整群抽样下的比估计量。

关键词: 整群抽样; 回归估计量; 可用性
分类号: AMS(2000) 62D05　　中图分类号: O212.2　　文献标识码: A

1 引 言

整群抽样具有实施便利、节省费用等优点, 因而是实际中广为应用的一种抽样方法。整群抽样的精度与群的划分有关, 当群划分不适当以致群内方差过小、群间方差过大时, 估计量精度就比较低。在实际问题中通常是基于某种方便的原则按自然形成的单位定义群, 例如按行政区划、部门行业等定义群, 这往往导致群内方差过小, 群间方差过大。对这种自然形成的群, 当群内方差过小, 群间方差过大时如何提高整群抽样的精度, 各文中尚未见报道。

抽样调查从本质上看是利用不完整的并且带有随机干扰的信息给出总体目标量的估计。在抽样调查中可资利用的信息有两类: 一类是基本信息, 指调查指标的样本信息, 这是估计总体目标量必不可少的信息; 另一类是辅助信息, 指除调查指标样本信息以外的一切有关总体及样本的信息, 主要有两种: 一种是抽样单元的规模信息, 另一种是相关性辅助指标的信息, 其中最常见的是调查指标的前期信息。在实际问题中, 我们在抽样调查之前对总体及抽样单元往往不是贫乏到一无所知的地步, 而是事先掌握某些可资利用的总体辅助信息, 例如与调查指标具有高度相关性的辅助指标总体信息。在抽样调查过程中除了获得调查指标的样本信息之外, 常可伴随获得相关性辅助指标的样本信息。这些辅助信息对提高估计精度有积极作用, 如不充分利用, 必将造成信息资源的浪费。基于这种考虑, 在整群抽样中如果存在与调查指标有着高度相关性的辅助指标, 就可利用辅助指标构造回归估计量以弥补因群内方差过小、群间方差过大所引起的精度损失, 提高整群抽样的效率。

* 收稿日期: 2000-06-06. 作者简介: 俞纯权(1947 年 11 月生), 男, 理学学士, 教授.

2　整群抽样回归估计量

设总体由 N 个群组成，第 i 群含有 M_i 个调查单元，$M_0 = \sum_{i=1}^{N} M_i$ 为总体包含的调查单元数。以 $\mathscr{Y}$ 表示调查指标，$\mathscr{X}$ 表示辅助指标，(Y_{ij}, X_{ij}) 表示第 i 群第 j 调查单元的两指标值，$i = 1, 2, \cdots, N, j = 1, 2, \cdots, M_i$。设调查的目标量是 Y。引入如下数字特征：

$Y_i = \sum_{j=1}^{M_i} Y_{ij}, X_i = \sum_{j=1}^{M_i} X_{ij}$ 为第 i 群两指标总值；$Y = \sum_{i=1}^{N} Y_i, X = \sum_{i=1}^{N} X_i$ 为两指标总体总值；$\bar{Y} = \frac{Y}{N}, \bar{X} = \frac{X}{N}$ 为两指标按群平均的总体均值；$S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2, S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$ 为两指标群间方差；$S_{yx} = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})(X_i - \bar{X})$ 为两指标群间协方差；$\rho = \frac{S_{yx}}{S_y S_x}$ 为两指标群间相关系数。

采用整群抽样从 N 个群中随机抽取 n 个，设抽中的群序号为 $\theta_1, \theta_2, \cdots, \theta_i, \cdots, \theta_n$，$\{\theta_1, \theta_2, \cdots, \theta_i, \cdots, \theta_n\}$ 是 $\{1, 2, \cdots, N\}$ 的一个子集，此处 i 表示入样的群在样本中的序号，其在总体中的序号应为 θ_i。为简化表达，在不致引起混淆的情况下将 $\theta_1, \theta_2, \cdots, \theta_i, \cdots, \theta_n$ 改记为 $1, 2, \cdots, i, \cdots, n$。对被抽中的群的调查单元全部调查，以 (y_{ij}, x_{ij}) 表示样本中第 i 群第 j 调查单元两指标观测值，$i = 1, 2, \cdots, n, j = 1, 2, \cdots, m_i$，$m_i$ 是群的大小，根据前面的约定，$m_i = M_{\theta_i}$。记

$$y_i = \sum_{j=1}^{m_i} y_{ij}, \qquad x_i = \sum_{j=1}^{m_i} x_{ij}$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2, \qquad s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$s_{yx} = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})$$

$$\hat{Y} = N\bar{y} = \frac{N}{n}\sum_{i=1}^{n} y_i, \qquad \hat{X} = N\bar{x} = \frac{N}{n}\sum_{i=1}^{n} x_i$$

则 $\hat{Y}, \hat{X}$ 分别为 Y, X 的简单估计量，且为无偏估计量。记

$$\hat{Y}_{lr} = \hat{Y} + b(X - \hat{X})$$

称 $\hat{Y}_{lr}$ 为 Y 的回归估计量，其中 b 为回归系数。当 $b = 0$ 时，$\hat{Y}_{lr} = \hat{Y}$ 即简单估计量；当 $b = \frac{\hat{Y}}{\hat{X}} = \frac{\bar{y}}{\bar{x}}$ 时 $\hat{Y}_{lr} = \hat{Y} + \frac{\hat{Y}}{\hat{X}}(X - \hat{X}) = \frac{\hat{Y}}{\hat{X}}X \triangleq \hat{Y}_R$

即比估计量。

3　$b = b_0$（b_0 为设定常数）时整群抽样回归估计量的性质

定理 1　当 b_0 是设定常数时，回归估计量

$$\hat{Y}_{lr} = \hat{Y} + b_0(X - \hat{X}) = \hat{Y} - b_0(\hat{X} - X)$$

是 Y 的无偏估计量。

$$V(Y_{lr}) = \frac{N^2(1-f)}{n}(S_y^2 + b_0^2 S_x^2 - 2b_0 S_{yx})$$

其中 $f = \frac{n}{N}$ 是抽样比。

证明 $E(Y_{lr}) = E(Y) - b_0[E(X) - X] = Y$, 得

$$V(Y_{lr}) = V[Y - b_0(X - X)] = V(Y) + b_0^2 V(X) - 2b_0 \mathrm{cov}(Y, X)$$

由文[3] 定理 2.2, 定理 2.3,

$$V(Y) = V(N\bar{y}) = N^2 V(\bar{y}) = \frac{N^2(1-f)}{n} S_y^2$$

$$V(X) = V(N\bar{x}) = N^2 V(\bar{x}) = \frac{N^2(1-f)}{n} S_x^2$$

$$\mathrm{cov}(Y, X) = \mathrm{cov}(N\bar{y}, N\bar{x}) = N^2 \mathrm{cov}(\bar{y}, \bar{x}) = \frac{N^2(1-f)}{n} S_{yx}$$

从而

$$V(Y_{lr}) = \frac{N^2(1-f)}{n}(S_y^2 + b_0^2 S_x^2 - 2b_0 S_{yx})$$

定理 2 记 $v(Y_{lr}) = \frac{N^2(1-f)}{n}(s_y^2 + b_0^2 s_x^2 - 2b_0 s_{yx})$, 则

$$E[v(Y_{lr})] = V(Y_{lr})$$

证明 由文[3] 定理 2.4 及定理 2.5,

$$E(s_y^2) = S_y^2,\ E(s_x^2) = S_x^2,\ E(s_{yx}) = S_{yx}$$

即得。

$$E[v(Y_{lr})] = V(Y_{lr})$$

定理 3 使 $V(Y_{lr})$ 极小化的 b_0 值为

$$b_0 = B = \frac{S_{yx}}{S_x^2} = \frac{\sum_{i=1}^{N}(Y_i - Y)(X_i - X)}{\sum_{i=1}^{N}(X_i - X)^2}$$

称 B 为群总体回归系数, 此时

$$V_{\min}(Y_{lr}) = \frac{N^2(1-f)}{n} S_y^2(1-\rho^2)$$

证明 $V(Y_{lr})$ 中 $S_y^2 + b_0^2 S_x^2 - 2b_o S_{yx}$ 是b_o 的二次函数, 二次项系数 $S_x^2 > 0$, 故必有最小值, 其最小值点

$$b_0 = -\frac{-2S_{yx}}{2S_x^2} = B$$

最小值为 $\frac{4S_y^2 S_x^2 - 4S_{yx}^2}{4S_x^2} = S_y^2(1-\rho^2)$。由此可得 $V_{\min}(Y_{lr})$。

4 b 需由样本估计时, 整群抽样回归估计量的性质

当 b 不能事先设定, 需由样本估计时, 根据定理 3 取 b 为

$$b=\frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

此时整群抽样回归估计量 Y_{lr} 是 Y 的有偏估计。

为研究它的性质, 先给出几个引理, 其证明见文[3] 引理 4.1 ~ 4.3。

引理 1　设 $\bar{y}$, $\bar{x}$ 分别是抽自均值为 $\bar{Y}$, $\bar{X}$ 的有限总体的简单随机样本的均值, 样本容量为 n, 则对非负整数 k, m 有

$E(\bar{y}-\bar{Y})^k=O(n^{-[(k+1)/2]})$, $E(\bar{y}-\bar{Y})^k(\bar{x}-\bar{X})^m=O(n^{-[(k+m+1)/2]})$。

引理 2　设 B 是有限总体回归系数, $\bar{x}$ 是简单随机样本的均值, 记

$D_i=(Y_i-\bar{Y})-B(X_i-\bar{X})$,

$d_i=(y_i-\bar{Y})-B(x_i-\bar{X})$。

则

1) $\bar{D}=0$;

2) $\sum_{i=1}^{N}D_i(X_i-\bar{X})=0$;

3) $E[\frac{1}{n}\sum_{i=1}^{n}d_i(x_i-\bar{x})]^2=O\left(\frac{1}{n}\right)$;

4) $E[\frac{1}{n}\sum_{i=1}^{n}d_i(x_i-\bar{x})]^4=O\left(\frac{1}{n^2}\right)$。

引理 3　在简单随机抽样下:

1) $b=B+\frac{\sum_{i=1}^{n}d_i(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$;

2) $E(b-B)^2=O(\frac{1}{n})$;

3) $E(b-B)=O(\frac{1}{\sqrt{n}})$;

4) $E(b-B)^4=O(\frac{1}{n^2})$。

定理 4　在简单随机抽样下, 整群抽样回归估计量 Y_{lr} 有下述性质

1) $B(Y_{lr})=E(Y_{lr}-Y)=E(Y_{lr})-Y$

$$=-\frac{N(1-f)}{(n-1)S_x^2}\frac{\sum_{i=1}^{N}D_i(X_i-\bar{X})^2}{N-1}+N\ O(\frac{1}{n^{3/2}})=N\circ O(\frac{1}{n});$$

2) $\mathrm{MSE}(Y_{lr})=V(Y)(1-\rho^2)+N^2\circ O(\frac{1}{n^{3/2}})$

$$=\frac{N^2(1-f)}{n}S_y^2(1-\rho^2)+N^2\circ O(\frac{1}{n^{3/2}})=N^2\circ O(\frac{1}{n});$$

3) 记　$s_e^2=\frac{1}{n-2}\sum_{i=1}^{n}[(y_i-\bar{y})-b(x_i-\bar{x})]^2$, 则

$$E(s_e^2) = S_y^2(1-\rho^2) + O(\frac{1}{\sqrt{n}})。$$

证明 1) $\hat{Y}_{lr} = \hat{Y} + b(X - \hat{X})$, 由引理 3,

$$E(\hat{Y}_{lr}) - Y = E[\hat{Y} + b(X - \hat{X})] - Y = E[b(X - \hat{X})]$$

$$= E\left[B + \frac{\sum_{i=1}^{n} d_i(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right](X - \hat{X})$$

因为 $E(s_x^2) = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right] = S_x^2$, 故当 n 足够大时

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = (n-1)s_x^2 \approx (n-1)S_x^2,$$

$$E(\hat{Y}_{lr} - Y) \approx B\ E(X - \hat{X}) + \frac{\left[\sum_{i=1}^{n} d_i(x_i - \bar{x})\right](\hat{X} - X)}{(n-1)S_x^2}$$

$$= -\frac{E\left[\sum_{i=1}^{n} d_i(x_i - \bar{x})\right](\hat{X} - X)}{(n-1)S_x^2}$$

$$= -\frac{E\left[\sum_{i=1}^{n} d_i(x_i - \bar{X}) - n\bar{d}(\bar{x} - \bar{X})\right](N\bar{x} - N\bar{X})}{(n-1)S_x^2}$$

$$= -\frac{N}{(n-1)S_x^2}\left\{E\left[\sum_{i=1}^{n} d_i(x_i - \bar{X})(\bar{x} - \bar{X})\right] - E[n\bar{d}(\bar{x} - \bar{X})^2]\right\}$$

记 $U_i = D_i(X_i - \bar{X})$, $u_i = d_i(x_i - \bar{X})$, 由引理 2

$$\bar{U} = \frac{1}{N}\sum_{i=1}^{N} U_i = \frac{1}{N}\sum_{i=1}^{N} D_i(X_i - \bar{X}) = 0$$

$$E\left[\sum_{i=1}^{n} d_i(x_i - \bar{X})(\bar{x} - \bar{X})\right] = nE[\bar{u}(\bar{x} - \bar{X})]$$

$$= nE[(\bar{u} - \bar{U})(\bar{x} - \bar{X})] = n\,\mathrm{cov}(\bar{u}, \bar{x})$$

由文[3:定理 2.3],

$$\mathrm{cov}(\bar{u}, \bar{x}) = \frac{1-f}{n}\frac{\sum_{i=1}^{N}(U_i - \bar{U})(X_i - \bar{X})}{N-1}$$

$$= \frac{1-f}{n}\frac{\sum_{i=1}^{N} U_i(X_i - \bar{X})}{N-1} = \frac{1-f}{n}\frac{\sum_{i=1}^{N} D_i(X_i - \bar{X})^2}{N-1},$$

由引理 1 及 Cauchy-Schwarz 不等式

$$E[n\bar{d}(\bar{x} - \bar{X})^2] \leq n\sqrt{E\bar{d}^2}\sqrt{E(\bar{x} - \bar{X})^4} = n\sqrt{E(\bar{d} - \bar{D})^2}\sqrt{E(\bar{x} - \bar{X})^4}$$

$$= n\sqrt{O(\frac{1}{n})}\sqrt{O(\frac{1}{n^2})} = O(\frac{1}{\sqrt{n}})$$

从而

$$B(Y_{lr}) = E(Y_{lr}) - Y = -\frac{N}{(n-1)S_x^2}\left[(1-f)\frac{\sum_{i=1}^{N} D_i(X_i - \bar{X})^2}{N-1} + O\left(\frac{1}{\sqrt{n}}\right)\right]$$

$$= -\frac{N(1-f)}{(n-1)S_x^2}\frac{\sum_{i=1}^{N} D_i(X_i - \bar{X})^2}{N-1} + N\,O\left(\frac{1}{n^{3/2}}\right) = N\,O\left(\frac{1}{n}\right)$$

(2) 记 $\bar{Y}^* = \bar{Y} + B(X - \bar{X})$, 由定理 3

$$E(\bar{Y}^*) = Y,\quad V(\bar{Y}^*) = V(\bar{Y})(1-\rho^2) = \frac{N^2(1-f)}{n}S_y^2(1-\rho^2)$$

$$MSE(Y_{lr}) = E(Y_{lr} - Y)^2 = E[(Y_{lr} - \bar{Y}^*) + (\bar{Y}_{lr}^* - Y)]^2$$

$$= E(Y_{lr} - \bar{Y}^*)^2 + E(\bar{Y}^* - Y)^2 + 2E(Y_{lr} - \bar{Y}^*)(\bar{Y}^* - Y)$$

$$= E[(b-B)^2(X - \bar{X})^2] + V(\bar{Y}^*) + 2E[(\bar{Y}^* - Y)(b-B)(X - \bar{X})]$$

由引理 3 及 Cauchy-Schwarz 不等式

$$E[(b-B)^2(X - \bar{X})^2] \leqslant \sqrt{E(b-B)^4}\sqrt{E(X - \bar{X})^4}$$

$$= \sqrt{O\left(\frac{1}{n^2}\right)}\sqrt{N^4 E(\bar{x} - \bar{X})^4} = \sqrt{O\left(\frac{1}{n^2}\right)}\sqrt{N^4\,O\left(\frac{1}{n^2}\right)} = N^2\,O\left(\frac{1}{n^2}\right)$$

$$E[(\bar{Y}^* - Y)(b-B)(X - \bar{X})] \leqslant \sqrt{E(\bar{Y}^* - Y)^2}\sqrt{E[(b-B)^2(X - \bar{X})^2]}$$

$$= \sqrt{\frac{N^2(1-f)}{n}S_y^2(1-\rho^2)}\sqrt{N^2\,O\left(\frac{1}{n^2}\right)}$$

$$= N^2\sqrt{O\left(\frac{1}{n}\right)}\sqrt{O\left(\frac{1}{n^2}\right)} = N^2\,O\left(\frac{1}{n^{3/2}}\right)$$

从而

$$\mathrm{MSE}(Y_{lr}) = V(\bar{Y}^*) + N^2\,O\left(\frac{1}{n^2}\right) + N^2\,O\left(\frac{1}{n^{3/2}}\right)$$

$$= \frac{N^2(1-f)}{n}S_y^2(1-\rho^2) + N^2\,O\left(\frac{1}{n^{3/2}}\right) = N^2\,O\left(\frac{1}{n}\right)$$

(3) 记 $S_D^2 = \frac{1}{N-1}\sum_{i=1}^{N}(D_i - \bar{D})^2$, $s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$, 在简单随机抽样下, $E(s_d^2) = S_D^2$, 而 $\bar{D} = 0$

$$S_D^2 = \frac{1}{N-1}\sum_{i=1}^{N}[(Y_i - \bar{Y}) - B(X_i - \bar{X})]^2$$

$$= \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2 + \frac{B^2}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2 - \frac{2B}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})(X_i - \bar{X})$$

$$= S_y^2 + B^2S_x^2 - 2BS_{yx}$$

$$= S_y^2 + \frac{S_{yx}^2}{S_x^4}\cdot S_x^2 - 2\frac{S_{yx}}{S_x^2}S_{yx} = S_y^2 - \frac{S_{yx}^2}{S_x^2} = S_y^2\left(1 - \frac{S_{yx}^2}{S_y^2S_x^2}\right)$$

$$= S_y^2(1-\rho^2)$$

故　$E(s_d^2) = S_y^2(1-\rho^2)$

又　$d_i - \bar{d} = (y_i - \bar{y}) - B(x_i - \bar{x}) = [(y_i - \bar{y}) - b(x_i - \bar{x})] + (b-B)(x_i - \bar{x})$

$$E(s_d^2) = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2\right]$$

$$= E\left\{\frac{1}{n-1}\sum_{i=1}^{n}[(y_i-\bar{y})-b(x_i-\bar{x})]^2\right\}+E\left[(b-B)^2\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}\right]$$

$$+2E\left[(b-B)\frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{n-1}\right]-2E\left[(b-B)b\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}\right]$$

当 n 大时, 由引理 3,

$$E\left[(b-B)^2\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}\right]=E[(b-B)^2 s_x^2]\approx S_x^2E(b-B)^2=O(\frac{1}{n})$$

$$E\left[(b-B)\frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{n-1}\right]=E[(b-B)s_{yx}]\approx S_{yx}E(b-B)=O(\frac{1}{\sqrt{n}})$$

$$E\left[(b-B)b\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}\right]\approx S_x^2E[(b-B)b]=S_x^2E[(b-B)^2+B(b-B)]$$

$$=S_x^2[E(b-B)^2+BE(b-B)]=S_x^2\left[O(\frac{1}{n})+O(\frac{1}{\sqrt{n}})\right]$$

$$=O(\frac{1}{\sqrt{n}})$$

故

$$E(s_d^2)=E\left\{\frac{1}{n-1}\sum_{i=1}^{n}[(y_i-\bar{y})-b(x_i-\bar{x})]^2\right\}+O(\frac{1}{n})+O(\frac{1}{\sqrt{n}})+O(\frac{1}{\sqrt{n}})$$

$$=E\left(\frac{n-2}{n-1}s_e^2\right)+O(\frac{1}{\sqrt{n}})$$

从而

$$E(s_e^2)=E(s_d^2)+O(\frac{1}{\sqrt{n}})=S_y^2(1-\rho^2)+O(\frac{1}{\sqrt{n}})$$

推论 1 Y_{lr} 是可用的。

证明
$$\frac{B(Y_{lr})}{\sqrt{MSE(Y_{lr})}}=\frac{O(\frac{1}{n})}{\sqrt{O(\frac{1}{n})}}=O(\frac{1}{\sqrt{n}})$$

从而 Y_{lr} 是可用的。

推论 2 n 大时

$$V(Y_{lr})\approx MSE(Y_{lr})\approx\frac{N^2(1-f)}{n}S_y^2(1-\rho^2)$$

证明 $MSE(Y_{lr})=V(Y_{lr})+[B(Y_{lr})]^2$

由定理 4

$$V(Y_{lr})=MSE(Y_{lr})+N^2O(\frac{1}{n^2})$$

$$=\frac{N^2(1-f)}{n}S_y^2(1-\rho^2)+N^2O\left(\frac{1}{n^{3/2}}\right)+N^2O\left(\frac{1}{n^2}\right)$$

$$= N^2\left[\frac{1-f}{n}S_y^2(1-\rho^2)+O\left(\frac{1}{n^{3/2}}\right)\right]$$

从而

$$V(Y_{lr})\approx MSE(Y_{lr})\approx\frac{N^2(1-f)}{n}S_y^2(1-\rho^2)$$

推论 3　记 $v(Y_{lr})=\frac{N^2(1-f)}{n}s_e^2$，则 $v(Y_{lr})$ 是 $V(Y_{lr})$ 的近似无偏估计。

证明　$E[v(Y_{lr})]=\frac{N^2(1-f)}{n}E(s_e^2)$

$$=\frac{N^2(1-f)}{n}\left[S_y^2(1-\rho^2)+O\left(\frac{1}{\sqrt{n}}\right)\right]$$

故　$E[v(Y_{lr})]\approx\frac{N^2(1-f)}{n}S_y^2(1-\rho^2)\approx V(Y_{lr})$。

5　整群抽样下回归估计量与简单估计量及比估计量的比较

当 n 大时，

$$V(Y_{lr})=\frac{N^2(1-f)}{n}S_y^2(1-\rho^2)$$

$$V(Y)=\frac{N^2(1-f)}{n}S_y^2$$

$$V(Y_R)=\frac{N^2(1-f)}{n}(S_y^2+R^2S_x^2-2R\rho S_yS_x)$$

其中 $R=\frac{Y}{X}$ 为总体比率，Y_R 是 Y 的整群抽样比估计量。由于 $0\leqslant\rho^2\leqslant 1$，故

$$V(Y_{lr})\leqslant V(Y)$$

这表明除非 $\rho=0$，整群抽样下回归估计量总是优于简单估计量。而

$$V(Y_{lr})\leqslant V(Y_R)$$
$$\Leftrightarrow -\rho^2S_y^2\leqslant R^2S_x^2-2R\rho S_yS_x$$
$$\Leftrightarrow R^2S_x^2-2R\rho S_yS_x+\rho^2S_y^2=(RS_x-\rho S_y)^2\geqslant 0$$
$$\Leftrightarrow\left(R-\rho\frac{S_y}{S_x}\right)^2=(R-B)^2\geqslant 0$$

这表明除非 $B=R$，整群抽样下回归估计量总是优于比估计量。

在一些定期进行的调查中，经常以上期调查指标 $\mathscr{X}$ 作为现期调查指标 $\mathscr{Y}$ 的辅助指标，它们是同质的。在一定条件下，由于事物发展的内在联系性，两期指标之间往往存在高度的正相关性。例如在农产量抽样调查中，正常年景下相邻两年的产量之间就存在高度的正相关性。此时在整群抽样中利用调查指标前期资料作为辅助信息构造回归估计量就可以显著地提高整群抽样的效率。另外，整群抽样回归估计量

$$Y_{lr}=Y+b(X-X)=Y+b\left(X-\frac{N}{n}\sum_{i=1}^{n}x_i\right)$$

只使用了辅助指标群总值 X_i 及总体总值 X 的信息，并没有直接使用调查单元的辅助指标信息 X_{ij}。因此，在已具备 X_i 及 X 的资料时，对调查单元进行调查只需收集调查指标的信息 Y_{ij} 即可。从调查角度讲，这种有辅助指标的整群抽样比通常的整群抽样几乎没有增加工作量，

但由于使用了相关性辅助指标, 可望估计精度有较大提高。

参考文献:

[1] Cochran W G. Sampling Techniques[M]. 3rd ed. Now York: John Wiley and Sons, 1977
[2] 许宝马录. 抽样论[M]. 北京: 北京大学出版社, 1982
[3] 冯士雍, 施锡铨. 抽样调查——理论、方法与实践[M]. 上海科学技术出版社, 1996
[4] Kish L. Survey Sampling[M]. New York: John Wiley and Sons, 1965
[5] 梁小筠, 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
[6] 俞纯权. 整群抽样比估计量[J]. 统计研究, 增刊, 1996
[7] 俞纯权. PPS 整群抽样下的比估计量[J]. 应用概率统计, 1999; 15(1)

The Regression Estimator in Cluster Random Sampling

YU Chun-quan
(Department of statistics, Shandong Institute of Economics, Jinan 250014)

Abstract: The cluster random sampling with auxiliary variable is investigated and a regression estimator of population total is constructed. It is proved that the estimator is feasible and the estimator is not worse than eiter mpared with simple estimator in cluster random sampling ratio estimator in cluster random sampling.
Keywords: Cluster random sampling; Regression estimators; feasibility

———————————————————————————————

(上接 36 页)

The Advances and Perspective in the Basis Function Network Approximation

JIAO Li-cheng, HOU Biao, LIU Fang
(Key Lab for Radar Signal Processing, Xidian Univ, Xi' an 710071)

Abstract: The states of art of the theories and applications of A revuew fir aookucatuib if in approximating to the approximation theory of multivariate functions. are reviewed, especially for ridgelet. The applications in image processing and the related problems on ridgelet transforms are discussed too.
Keywords: Single hidden-layer feedforward neural networks; The greedy algorithm; Projection pursuit regression; Fourier analysis; Wavelet analysis; Wavelet neural networks; Ridgelet; Frame theory; Image compression