第 20 卷第 2 期
2001 年 3 月
无锡轻工大学学报
Journal of Wuxi University of Light Industry
Vol. 20, No. 2
Mar., 2001

文章编号:1009-038X(2001)02-0199-03

回归估计量均方误差的近似值及其估计量的偏差

张荷观
(无锡轻工大学商学院,江苏无锡 214064)

摘　要:对回归估计,有下列结果:

(1) $\mathrm{MSE}(\overline{y}_{lr}) = V(\overline{y}_{lr}) + O(\frac{1}{n^2})$;　　(2) $E[\mathrm{mse}(\overline{y}_{lr})] = V(\overline{y}_{lr}) + O(\frac{1}{n^2})$.

其中

$V(\overline{y}_{lr}) = (\frac{1}{n} - \frac{1}{N}) S_y^2 (1-\rho^2)$,　　$\mathrm{mse}(\overline{y}_{lr}) = (\frac{1}{n} - \frac{1}{N}) s_y^2 (1-r^2)$.

关键词:回归估计;均方误差的近似值;均方误差近似值的估计;偏差
中图分类号:O212.2　　　　**文献标识码**:A

Bias of Approximate Mean Square Error of the Regression Estimator and Its Estimator

ZHANG He-guan
(School of Business, Wuxi University of Light Industry, Wuxi 214064)

Abstract: For the regression estimate, we abtain the following main theorem

(1) $\mathrm{MSE}(\overline{y}_{lr}) = V(\overline{y}_{lr}) + O(\frac{1}{n^2})$　　(2) $E[\mathrm{mse}(\overline{y}_{lr})] = V(\overline{y}_{lr}) + O(\frac{1}{n^2})$.

Where

$V(\overline{y}_{lr}) = (\frac{1}{n} - \frac{1}{N}) S_y^2 (1-\rho^2)$,　　$\mathrm{mse}(\overline{y}_{lr}) = (\frac{1}{n} - \frac{1}{N}) s_y^2 (1-r^2)$

Key words: regression estimation, approximate mean squar error, estimation of approximate mean square error, bias

1　引　言

设总体由 N 个单元组成,每个单元有两个数量指标 X、Y,且

$$Y_i = \overline{Y} + \beta(X_i - \overline{X}) + \varepsilon_i, i = 1,2,\cdots,N \quad (1)$$

其中

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i,\quad \overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i,\quad \beta = \frac{L_{11}}{L_{20}} \quad (2)$$

$$L_{kt} = \sum_{i=1}^{N} (X_i - \overline{X})^k (Y_i - \overline{Y})^t,\quad k、t = 0,1,2$$

设从总体中用不放回方式抽取样本量为 n 的简单随机样本时,记

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,\quad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,\quad b = \frac{l_{11}}{l_{20}}$$

收稿日期:2000-11-07;修订日期:2001-02-20.
作者简介:张荷观(1949-),男,江苏吴江人,教授.

$$l_{kt}=\sum_{i=1}^{n}(x_i-\overline{x})^k(y_i-\overline{y})^t,\quad k、t=0,1,2 \tag{3}$$

回归估计是用

$$\overline{y}_{lr}=\overline{y}+b(\overline{X}-\overline{x}) \tag{4}$$

作为 $\overline{Y}$ 的估计. 并已证明均方误差近似值 $V(\overline{y}_{lr})$ 及其估计量 $\mathrm{mse}(\overline{y}_{lr})$ 的偏差的阶为 $1/n^{\frac{3}{2}}$[1,2],即

$$\mathrm{MSE}(\overline{y}_{lr})=V(\overline{y}_{lr})+O(\frac{1}{n^{\frac{3}{2}}}) \tag{5}$$

$$E[\mathrm{mse}(\overline{y}_{lr})]=V(\overline{y}_{lr})+O(\frac{1}{n^{\frac{3}{2}}}) \tag{6}$$

其中

$$\begin{aligned}&V(\overline{y}_{lr})=(\frac{1}{n}-\frac{1}{N})S_y^2(1-\rho^2),\\&\mathrm{mse}(\overline{y}_{lr})=(\frac{1}{n}-\frac{1}{N})s_y^2(1-r^2)\\&S_y^2=\frac{L_{02}}{N-1},\quad s_y^2=\frac{l_{02}}{n-1},\\&\rho=\frac{L_{11}}{\sqrt{L_{20}L_{02}}},\quad r=\frac{l_{11}}{\sqrt{l_{20}l_{02}}}\end{aligned} \tag{7}$$

本文证明均方误差的近似值 $V(\overline{y}_{lr})$ 及其估计量 $\mathrm{mse}(\overline{y}_{lr})$ 的偏差的阶为 $\frac{1}{n^2}$.

2 $V(\overline{y}_{lr})$的偏差

引理 设 u_i、v_i 和 $w_i(i=1,2,\cdots,n)$ 是来自总体 U_i、V_i 和 $W_i(i=1,2,\cdots,N)$ 的简单随机样本,$\overline{u}$、$\overline{v}$ 和 $\overline{w}$ 为样本均值,且总体均值 $\overline{U}=\overline{V}=\overline{W}=0$,则对非负整数 k、t,有

$$(\text{Ⅰ})\ \mathrm{E}(\overline{u}^k)=\begin{cases}O(\frac{1}{n^{\frac{k}{2}}}),\text{若 } k \text{ 为偶数}\\ O(\frac{1}{n^{\frac{(k+1)}{2}}}),\text{若 } k \text{ 为奇数}\end{cases}$$

$$(\text{Ⅱ})\ \mathrm{E}(\overline{u}^k\overline{v}^t)=\begin{cases}O(\frac{1}{n^{\frac{(k+t)}{2}}}),\text{若 } k+t \text{ 为偶数}\\ O(\frac{1}{n^{\frac{(k+t+1)}{2}}}),\text{若 } k+t \text{ 为奇数}\end{cases}$$

$$(\text{Ⅲ})\ \mathrm{E}(\overline{u}\overline{v}\overline{w})=(\frac{1}{n}-\frac{1}{N})\frac{N-2n}{n(N-1)(N-2)}\sum_{i=1}^{N}U_iV_iW_i=O(\frac{1}{n^2})$$

证 (Ⅰ)、(Ⅱ)的证明见[3],现证明(Ⅲ). 由于

$$\begin{aligned}\overline{u}\overline{v}\overline{w}&=(\frac{1}{n}\sum_{i=1}^{n}u_i)(\frac{1}{n}\sum_{v=1}^{n}v_i)(\frac{1}{n}\sum_{w=1}^{n}w_i)=\\&\frac{1}{n^3}(\sum_{i=1}^{n}u_iv_iw_i+\sum_{i\neq j}^{n}u_iv_iw_j+\sum_{i\neq j}^{n}u_iv_jw_i+\\&\sum_{i\neq j}^{n}u_iv_jw_j+\sum_{i\neq j\neq l}^{n}u_iv_jw_l)\end{aligned}$$

于是

$$\begin{aligned}E(\overline{u}\overline{v}\overline{w})=&\frac{1}{n^3}\Big\{\frac{n}{N}\sum_{i=1}^{N}U_iV_iW_i+\frac{n(n-1)}{N(N-1)}\times\\&[\sum_{i\neq j}^{N}U_iV_iW_j+\sum_{i\neq j}^{N}U_iV_jW_i+\sum_{i\neq j}^{N}U_iV_jW_j]+\\&\frac{n(n-1)(n-2)}{N(N-1)(N-2)}\sum_{i\neq j\neq l}^{N}U_iV_jW_l\Big\}\end{aligned}$$

由

$$\begin{aligned}O&=\sum_{i=1}^{N}U_iV_i\sum_{i=1}^{N}W_i=\\&\sum_{i=1}^{N}U_iV_iW_i+\sum_{i\neq j}^{N}U_iV_iW_j\end{aligned}$$

得

$$\sum_{i\neq j}^{N}U_iV_iW_j=-\sum_{i=1}^{N}U_iV_iW_i$$

类似有

$$\sum_{i\neq j}^{N}U_iV_jW_i=\sum_{i\neq j}^{N}U_iV_jW_j=-\sum_{i=1}^{N}U_iV_iW_i$$

而

$$\begin{aligned}O&=\sum_{i\neq j}^{N}U_iV_j\sum_{i=1}^{N}W_i=\\&\sum_{i\neq j}^{N}U_iV_jW_i+\sum_{i\neq j}^{N}U_iV_jW_j+\sum_{i\neq j\neq l}^{N}U_iV_jW_l\end{aligned}$$

所以

$$\sum_{i\neq j\neq l}^{n}U_iV_jW_l=2\sum_{i=1}^{N}U_iV_iW_i$$

从而

$$\begin{aligned}\mathrm{E}(\overline{u}\overline{v}\overline{w})=&\frac{1}{n^3}[\frac{n}{N}\sum_{i=1}^{N}U_iV_iW_i-\frac{3n(n-1)}{N(N-1)}\times\\&\sum_{i=1}^{N}U_iV_iW_i+\frac{2n(n-1)(n-2)}{N(N-1)(N-2)}\sum_{i=1}^{N}U_iV_iW_i]=\\&(\frac{1}{n}-\frac{1}{N})\frac{N-2n}{n(N-1)(N-2)}\sum_{i=1}^{N}U_iV_iW_i\leqslant\\&\frac{1}{n^2}(\frac{1}{N}\sum_{i=1}^{N}U_iV_iW_i)=O(\frac{1}{n^2})\end{aligned}$$

定理 1 设按简单随机方法抽取样本时,对所有 C_N^n 个样本,$0<d_1\leqslant l_{20}/n\leqslant d_2<\infty$,$d_1$ 和 d_2 与 n 无关,则

$$\mathrm{MSE}(\overline{y}_{lr})=V(\overline{y}_{lr})+O(\frac{1}{n^2}) \tag{8}$$

证 由于

$$\begin{aligned}&\mathrm{MSE}(\overline{y}_{lr})=E(\overline{y}_{lr}-\overline{Y})^2=\\&E[(\overline{y}-\overline{Y})-(b-\beta)(\overline{x}-\overline{X})-\beta(\overline{x}-\overline{X})]^2=\\&E(\overline{y}-\overline{Y})^2+\mathrm{E}(b-\beta)^2(\overline{x}-\overline{X})^2+\\&\beta^2E(\overline{x}-\overline{X})^2-2E(b-\beta)(\overline{x}-\overline{X})(\overline{y}-\overline{Y})-\\&2\beta E(\overline{x}-\overline{X})(\overline{y}-\overline{Y})+2\beta E(b-\beta)(\overline{x}-\overline{X})^2\end{aligned}$$

根据(2)和(3)式

$$b-\beta=\frac{n}{l_{20}}[(\frac{l_{11}}{n}-\frac{L_{11}}{N})-(\frac{l_{20}}{n}-\frac{L_{20}}{N})\frac{L_{11}}{L_{20}}] \tag{9}$$

记

$$\begin{aligned}&U_i=X_i-\overline{X},\quad V_i=Y_i-\overline{Y},\\&W_i=U_i^2-\frac{1}{N}\sum_{i=1}^{N}U_i^2,\quad G_i=U_iV_i-\frac{1}{N}\sum_{i=1}^{N}U_iV_i\\&u_i=x_i-\overline{X},\quad v_i=y_i-\overline{Y},\end{aligned}$$

$$w_i = u_i^2 - \frac{1}{N}\sum_{i=1}^{N} U_i^2, \quad g_i = u_i v_i - \frac{1}{N}\sum_{i=1}^{N} U_i V_i$$

则 u_i、v_i、w_i 和 $g_i(i=1,2,\cdots,n)$ 分别是来自总体 U_i、V_i、W_i 和 $G_i(i=1,2,\cdots,N)$ 的简单随机样本，记 $\overline{u}$、$\overline{v}$、$\overline{w}$ 和 $\overline{g}$ 为相应的样本均值，则总体均值 $\overline{U}=\overline{V}=\overline{W}=\overline{G}=0$. 于是

$$b-\beta=\frac{n}{l_{20}}\left[(\overline{g}-\overline{u}\overline{v})-(\overline{w}-\overline{u}^2)\frac{L_{11}}{L_{20}}\right] \tag{10}$$

根据引理，有

$$E(b-\beta)^2(\overline{x}-\overline{X})^2 = E\left\{\frac{n}{l_{20}}\left[(\overline{g}-\overline{u}\overline{v})-(\overline{w}-\overline{u}^2)\frac{L_{11}}{L_{20}}\right]\overline{u}\right\}^2 \leqslant \frac{1}{d_1^2}E\left[(\overline{g}-\overline{u}\overline{v})\overline{u}-(\overline{w}-\overline{u}^2)\overline{u}\frac{L_{11}}{L_{20}}\right]^2 = \frac{1}{d_1^2}\cdot O\left(\frac{1}{n^2}\right)=O\left(\frac{1}{n^2}\right)$$

$$E(b-\beta)(\overline{x}-\overline{X})(\overline{y}-\overline{Y}) = E\left\{\frac{n}{l_{20}}\left[(\overline{g}-\overline{u}\overline{v})-(\overline{w}-\overline{u}^2)\frac{L_{11}}{L_{20}}\right]\overline{u}\overline{v}\right\} \leqslant \frac{1}{d_1}E\left[(\overline{g}-\overline{u}\overline{v})\overline{u}\overline{v}-(\overline{w}-\overline{u}^2)\overline{u}\overline{v}\frac{L_{11}}{L_{20}}\right] = \frac{1}{d_1}\cdot O\left(\frac{1}{n^2}\right)=O\left(\frac{1}{n^2}\right)$$

$$E(b-\beta)(\overline{x}-\overline{X})^2 = E\left\{\frac{n}{l_{20}}\left[(\overline{g}-\overline{u}\overline{v})-(\overline{w}-\overline{u}^2)\frac{L_{11}}{L_{20}}\right]\overline{u}^2\right\} \leqslant \frac{1}{d_1}E\left[(\overline{g}-\overline{u}\overline{v})\overline{u}^2-(\overline{w}-\overline{u}^2)\overline{u}^2\frac{L_{11}}{L_{20}}\right] = \frac{1}{d_1}\cdot O\left(\frac{1}{n^2}\right)=O\left(\frac{1}{n^2}\right)$$

而 $$E(\overline{y}-\overline{Y})^2=\left(\frac{1}{n}-\frac{1}{N}\right)S_y^2$$

$$\beta^2 E(\overline{x}-\overline{X})^2 = \beta^2\left(\frac{1}{n}-\frac{1}{N}\right)\frac{L_{20}}{N-1} = \left(\frac{1}{n}-\frac{1}{N}\right)\rho^2 S_y^2$$

$$\beta E(\overline{x}-\overline{X})(\overline{y}-\overline{Y}) = \beta\left(\frac{1}{n}-\frac{1}{N}\right)\frac{L_{11}}{N-1} = \left(\frac{1}{n}-\frac{1}{N}\right)\rho^2 S_y^2$$

所以 $$\mathrm{MSE}(\overline{y}_{lr})=\left(\frac{1}{n}-\frac{1}{N}\right)S_y^2(1-\rho^2)+O\left(\frac{1}{n^2}\right)=V(\overline{y}_{lr})+O\left(\frac{1}{n^2}\right)$$

3　mse$(\overline{y}_{lr})$的偏差

定理 2　在定理 1 的条件下，则

$$E[\mathrm{mse}(\overline{y}_{lr})]=V(\overline{y}_{lr})+O\left(\frac{1}{n^2}\right) \tag{11}$$

证　由于

$$r^2 s_y^2 = b^2\frac{l_{20}}{n-1}=\beta^2\frac{l_{20}}{n-1}+2\beta(b-\beta)\frac{l_{20}}{n-1}+(b-\beta)^2\frac{l_{20}}{n-1}$$

其中

$$E\left(\beta^2\frac{l_{20}}{n-1}\right)=\beta^2\frac{L_{20}}{N-1}=\rho^2 S_y^2$$

根据(10)式

$$(b-\beta)\frac{l_{20}}{n-1} = \frac{n}{n-1}\left[(\overline{g}-\overline{u}\overline{v})-(\overline{w}-\overline{u}^2)\frac{L_{11}}{L_{20}}\right]$$

于是

$$E\left[(b-\beta)\frac{l_{20}}{n-1}\right]=\frac{n}{n-1}\left[\left(\frac{1}{N}-\frac{1}{n}\right)\frac{L_{11}}{N-1}+\left(\frac{1}{n}-\frac{1}{N}\right)\frac{L_{20}}{(N-1)}\frac{L_{11}}{L_{20}}\right]=0$$

又由引理

$$E\left[(b-\beta)^2\frac{l_{20}}{n-1}\right]=O\left(\frac{1}{n}\right)$$

从而

$$E(r^2 s_y^2)=\rho^2 S_y^2+O\left(\frac{1}{n}\right)$$

所以

$$E[\mathrm{mse}(\overline{y}_{lr})]=\left(\frac{1}{n}-\frac{1}{N}\right)E(s_y^2-r^2 s_y^2)=V(\overline{y}_{lr})+O\left(\frac{1}{n^2}\right)$$

4　结　语

文献[1]、[2]给出了回归估计量的均方误差及均方误差估计量的偏差为 $O\left(\frac{1}{n^{\frac{3}{2}}}\right)$，本文改进了这一结论，证明回归估计量的均方误差及均方误差估计量的偏差为 $O\left(\frac{1}{n^2}\right)$.

参考文献：

[1] 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.
[2] 张荷观. 双重回归估计量的偏差和均方误差的估计量[J]. 生物数学学报，1991，6(4):98～110.
[3] DARID I P, SAKHATME B V. On the bias and mean square error of the ratio estimator[J]. **Jour Amer Stat Assoc**, 1974, 69: 464～466.

(责任编辑 朱 明　秦和平)