第20卷第1期 无锡轻工大学学报 Vol.20,No.1
2001年1月 Journal of Wuxi University of Light Industry Jan.,2001

文章编号:1009-038X(2001)01-0095-03

分子已知情况下的比率估计量

张荷观
(无锡轻工大学商学院,江苏无锡 214036)

摘　要:给出了分子已知情况下比率估计量的偏差和均方误差以及均方误差的估计,并对两类比率估计量进行了比较.

关键词:比率估计量;分子已知;偏差;均方误差
中图分类号:F224　　　**文献标识码**:A

The Ratio Estimator with Known Numerator

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

Abstract: In this paper, the bias of the ratio estimator and the estimate of mean square errors of this estimator with known numerator have been compared and discussed.

Key words: the ratio estimator; known numerator; bias; mean square error.

一般,比率估计量的分子与分母都是随机变量.但在社会经济统计中,常会遇到分子是已知的常数,而分母是随机变量的比率估计量.即这种比率估计量的分子是由全面调查所得的常量,而分母则是采用抽样调查由样本所给出的统计量.

假定总体由 N 个单元组成,每个单元有两个标志值(X_i, Y_i),$i=1,2,\cdots\cdots,N$.记

$$\overline{X}=\frac{1}{N}\sum_{i=1}^{N}X_i,\ \overline{Y}=\frac{1}{N}\sum_{i=1}^{N}Y_i,\ R=\frac{\overline{Y}}{\overline{X}} \quad (1)$$

称 R 为总体比率.

设从总体抽取一个样本量为 n 的简单随机样本时,只调查 X 标志.第 i 个抽到的单元标志值为 x_i,$i=1,2,\cdots\cdots,n$.记

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,\hat{R}=\frac{\overline{Y}}{\overline{x}} \quad (2)$$

称 $\hat{R}$ 为样本比率,其中 Y 标志的总体均值 $\overline{Y}$ 已知.这就是分子已知情况下的比率估计量.

本文给出了这种比率估计量的偏差、均方误差以及均方误差的估计.同时给出了这种比率估计量的均方误差小于一般比率估计量的条件.

1　估计量的偏差

定理1　设 $X_i>0$,$i=1,2,\cdots,N$,$\overline{Y}>0$.则对简单随机抽样,有

$$E(\hat{R})=R+O(\frac{1}{n}) \quad (3)$$

证　根据(2)式,得

$$\hat{R}=\overline{Y}[\frac{1}{\overline{X}}-\frac{\overline{x}-\overline{X}}{\overline{X}^2}+\frac{(\overline{x}-\overline{X})^2}{\overline{x}\overline{X}^2}] \quad (4)$$

由于 $\overline{x}$ 是简单随机抽样的样本均值,从而

$$E(\overline{x})=\overline{X}$$

$$E(\overline{x}-\overline{X})^2=(\frac{1}{n}-\frac{1}{N})S_x^2=O(\frac{1}{n})$$

$$S_x^2=\frac{1}{N-1}\sum_{i=1}^{N}(X_i-\overline{X})^2$$

收稿日期:2000-09-03;修订日期:2000-10-10.
作者简介:张荷观(1949-),男,江苏吴江人,教授.
万方数据

于是

$$E(\hat{R}) = R + \overline{Y}E\left[\frac{(\overline{x}-\overline{X})^2}{\overline{x}\,\overline{X}^2}\right]$$

记 $X_{\min}=\min\{X_1,X_2,\cdots,X_N\}$，则 $\overline{x}\geqslant X_{\min}>0$.

所以

$$\left|E(\hat{R})-R\right| \leqslant \frac{\overline{Y}}{X_{\min}\overline{X}^2}E(\overline{x}-\overline{X})^2 = O\left(\frac{1}{n}\right)$$

2 估计量的均方误差

定理2 在定理1的条件下,比率估计量 $\hat{R}$ 的均方误差为

$$\begin{aligned}\mathrm{MSE}(\hat{R}) &= E(\hat{R}-R)^2\\ &= \frac{R^2}{\overline{X}^2}\left(\frac{1}{n}-\frac{1}{N}\right)S_x^{\,2} + O\left(\frac{1}{n^2}\right)\end{aligned} \tag{5}$$

证 根据(4)式,则

$$(\hat{R}-R)^2 = \overline{Y}^2\left[\frac{\overline{x}-\overline{X}}{\overline{X}^2} - \frac{(\overline{x}-\overline{X})^2}{\overline{x}\,\overline{X}^2}\right]^2 = \overline{Y}^2\left[\frac{(\overline{x}-\overline{X})^2}{\overline{X}^4} + \frac{(\overline{x}-\overline{X})^4}{\overline{x}^2\overline{X}^4} - \frac{2(\overline{x}-\overline{X})^3}{\overline{x}\,\overline{X}^4}\right]$$

由于[1]

$$E(\overline{x}-\overline{X})^3 = O\left(\frac{1}{n^2}\right),\ E(\overline{x}-\overline{X})^4 = O\left(\frac{1}{n^2}\right) \tag{6}$$

从而

$$E(\hat{R}-R)^2 = \frac{\overline{Y}^2}{\overline{X}^4}E(\overline{x}-\overline{X})^2 + O\left(\frac{1}{n^2}\right) = \frac{R^2}{\overline{X}^2}\left(\frac{1}{n}-\frac{1}{N}\right)S_x^{\,2} + O\left(\frac{1}{n^2}\right)$$

3 均方误差的估计量

定理3 在定理1的条件下,有

$$E[\mathrm{mse}(\hat{R})] = \frac{R^2}{\overline{X}^2}\left(\frac{1}{n}-\frac{1}{N}\right)S_x^{\,2} + O\left(\frac{1}{n^2}\right) \tag{7}$$

其中

$$\mathrm{mse}(\hat{R}) = \frac{\hat{R}^2}{\overline{x}^2}\left(\frac{1}{n}-\frac{1}{N}\right)s_x^{\,2},$$

$$s_x^{\,2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2 \tag{8}$$

证 由于

$$\overline{x}^4 = (\overline{x}-\overline{X})^4 + 4(\overline{x}-\overline{X})^3\overline{X} + 6(\overline{x}-\overline{X})^2\overline{X}^2 + 4(\overline{x}-\overline{X})\overline{X}^3 + \overline{X}^4$$

根据(8)式,得

$$\begin{aligned}\mathrm{mse}(\hat{R}) &= \frac{\overline{Y}^2}{\overline{x}^4}\left(\frac{1}{n}-\frac{1}{N}\right)s_x^{\,2} = \\ &\overline{Y}^2\left(\frac{1}{n}-\frac{1}{N}\right)s_x^{\,2}\left(\frac{1}{\overline{X}^4} - \frac{\overline{x}^4-\overline{X}^4}{\overline{x}^4\overline{X}^4}\right) = \\ &\overline{Y}^2\left(\frac{1}{n}-\frac{1}{N}\right)s_x^{\,2}\left\{\frac{1}{\overline{X}^4} - \frac{4(\overline{x}-\overline{X})}{\overline{X}^5} - \right.\\ &\frac{(\overline{x}-\overline{X})^4+4(\overline{x}-\overline{X})^3\overline{X}+6(\overline{x}-\overline{X})^2\overline{X}^2}{\overline{x}^4\overline{X}^4} + \\ &\frac{4(\overline{x}-\overline{X})^2}{\overline{x}^4\overline{X}^5}\left[(\overline{x}-\overline{X})^3+4(\overline{x}-\overline{X})^2\overline{X} + \right.\\ &\left.\left.6(\overline{x}-\overline{X})\overline{X}^2+4\overline{X}^3\right]\right\}\end{aligned} \tag{9}$$

显然

$$E(s_x^{\,2}) = S_x^{\,2} \tag{10}$$

而

$$(\overline{x}-\overline{X})s_x^{\,2} = \frac{(\overline{x}-\overline{X})}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2 = \frac{\overline{x}-\overline{X}}{n-1}\left[\sum_{i=1}^{n}(x_i-\overline{X})^2 - n(\overline{x}-\overline{X})^2\right]$$

记

$$u_i = x_i-\overline{X},\ U_i = X_i-\overline{X}$$

则

$$\overline{u} = \frac{1}{n}\sum_{i=1}^{n}u_i = \overline{x}-\overline{X},\ \overline{U} = \frac{1}{N}\sum_{i=1}^{N}U_i = 0$$

于是

$$(\overline{x}-\overline{X})s_x^{\,2} = \frac{\overline{u}}{n-1}\sum_{i=1}^{n}u_i^{\,2} - \frac{n}{n-1}\overline{u}^3$$

其中

$$\begin{aligned}E\left(\overline{u}\sum_{i=1}^{n}u_i^{\,2}\right) &= \frac{1}{n}E\left(\sum_{i=1}^{n}u_i^{\,3} + \sum_{i\neq j}^{n}u_iu_j^{\,2}\right) = \\ &\frac{1}{n}\sum_{i=1}^{N}U_i^{\,3} + \frac{n-1}{N(N-1)}\sum_{i\neq j}^{N}U_iU_j^{\,2} = \\ &\frac{1}{N}\sum_{i=1}^{N}U_i^{\,3} - \frac{n-1}{N(N-1)}\sum_{i=1}^{N}U_i^{\,3} = \\ &\frac{N-n}{N(N-1)}\sum_{i=1}^{N}U_i^{\,3}\end{aligned}$$

$$E(\overline{u}^3) = \frac{(N-n)(N-2n)}{n^2N(N-1)(N-2)}\sum_{i=1}^{N}U_i^{\,3}$$

从而

$$E[(\overline{x}-\overline{X})s_x^{\,2}] = \frac{N-n}{n(N-1)(N-2)}\sum_{i=1}^{N}U_i^{\,3} = O\left(\frac{1}{n}\right) \tag{11}$$

同样可得

$$E\left\{s_x^{\,2}\left[\frac{(\overline{x}-\overline{X})^4+4(\overline{x}-\overline{X})^3\overline{X}+6(\overline{x}-\overline{X})^2\overline{X}^2}{\overline{x}^4\overline{X}^4}\right]\right\} = O\left(\frac{1}{n}\right) \tag{12}$$

$$E\left\{s_x^{\,2}\frac{4(\overline{x}-\overline{X})^2}{\overline{x}^4\overline{X}^5}\left[(\overline{x}-\overline{X})^3+4(\overline{x}-\overline{X})^2\overline{X} + 6(\overline{x}-\overline{X})\overline{X}^2+4\overline{X}^3\right]\right\} = O\left(\frac{1}{n}\right) \tag{13}$$

由(9)、(10)、(11)、(12)和(13)式,即得

$$E[\mathrm{mse}(\hat{R})] = \frac{R^2}{\overline{X}^2}\left(\frac{1}{n}-\frac{1}{N}\right)S_x^{\,2} + O\left(\frac{1}{n^2}\right)$$

万方数据

4　与一般比率估计量均方误差的比较

对于分子与分母都为随机变量的一般比率估计量 $\hat{R}_2 = \overline{y}/\overline{x}$,相应的均方误差为[2,3]

$$\mathrm{MSE}(\hat{R}_2) = \frac{1}{\overline{X}^2}\left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{N-1}\sum_{i=1}^{N}(Y_i - RX_i)^2 + O\left(\frac{1}{n^2}\right) \tag{14}$$

记

$$V_1 = \frac{R^2}{\overline{X}^2}\left(\frac{1}{n} - \frac{1}{N}\right)S_x^{\,2}$$

$$V_2 = \frac{1}{\overline{X}^2}\left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{N-1}\sum_{i=1}^{N}(Y_i - RX_i)^2 \tag{15}$$

根据(15)式,当

$$\rho < \frac{1}{2}\sqrt{\frac{\sum_{i=1}^{N}(Y_i - \overline{Y})^2/\overline{Y}^2}{\sum_{i=1}^{N}(X_i - \overline{X})^2/\overline{X}^2}} = \frac{C_Y}{2C_X} > 0 \tag{16}$$

成立时,必有 $V_1 < V_2$.其中

$$\rho = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2\sum_{i=1}^{N}(Y_i - \overline{Y})^2}}$$

因此只要 ρ 值较小时,就会有 $V_1 < V_2$.特别当 $\rho < 0$ 时,一定有 $V_1 < V_2$.这在直观上也是容易理解的,当 $\rho < 0$ 时,$\hat{R}$ 的波动就会小于 $\hat{R}_2$,从而有 $V_1 < V_2$.

而当

$$\rho > \frac{C_Y}{2C_X} > 0 \tag{17}$$

成立时,必有 $V_1 > V_2$.

5　实　例

某单位家属区共有 2159 户职工,由于每户的住房面积(单位:m^2) Y 是已知的,从而 $\overline{Y} = 41.9$ 已知.从该单位家属区按简单随机方法抽取 30 户,只调查家庭人口 X 如下,但为了比较的需要,同时也列出了相应的住房面积.试求该单位家属区人均住房面积的估计及均方误差的估计.

表 1　样本数据

Tab.1　Sampled data

住房面积/m^2	人口	住房面积/m^2	人口	住房面积/m^2	人口
53.5	5	28.9	3	35.2	4
31.4	3	28.9	2	31.4	3
37.5	4	37.5	3	48.4	4
28.9	3	35.2	3	43.5	4
31.4	4	43.5	3	53.5	3
59.1	2	35.2	2	22.6	2
59.1	3	43.5	5	35.2	3
22.6	3	53.5	2	35.2	2
59.1	5	22.6	2	41.9	3
31.4	3	48.4	3	43.5	3

解　已知 $N = 2159$,$n = 30$,$\overline{Y} = 41.9$.记第 i 户家庭人口为 x_i,根据表 1 的样本数据得

$$\overline{x} = 3.13,\ s_x^{\,2} = 0.81$$

于是根据(2)式,人均住房面积的估计

$$\hat{R} = \frac{41.9}{3.13} = 13.4$$

而根据(8)式,均方误差的估计

$$\mathrm{mse}(\hat{R}) = \frac{13.4^2}{3.13^2}\times\left(\frac{1}{30} - \frac{1}{2159}\right)\times 0.81 = 0.49$$

对于分子与分母都为随机变量的一般比率估计量 $\hat{R}_2$,则均方误差的估计[2,3]

$$\mathrm{mse}(\hat{R}_2) = \frac{1}{\overline{x}^2}\left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \hat{R}_2 x_i)^2 \tag{18}$$

因此,假定同时对每户家庭人口与住房面积作调查,就可以采用一般的比率估计方法.记第 i 户住房面积为 y_i,根据表 1 的样本数据得

$$\hat{R}_2 = \frac{\overline{y}}{\overline{x}} = \frac{39.39}{3.13} = 12.6$$

$$\mathrm{mse}(\hat{R}_2) = \frac{1}{3.13^2}\times\left(\frac{1}{30} - \frac{1}{2159}\right)\times 168.31 = 0.56$$

6　结　语

对分子是已知常数,而分母是随机变量的比率估计,在 X 与 Y 为负相关,或虽为正相关但相关系数较小时,这种比率估计都能比分子与分母均为随机变量的一般比率估计有更高的估计精度。本文的数值例子就是这种比率估计的实际应用。

参考文献:

[1] 许宝騄.抽样论[M].北京:北京大学出版社,1982.7~11.
[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,1996.111~117.
[3] 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,1994.95~132.

(责任编辑:秦和平)

万方数据