

# 距离平衡排除临近单元抽样方法在 Warner 模型中的应用

董洪芹 蔡楠

(烟台南山学院基础教学部,烟台 264000)

**摘要** 随机化抽样方法对敏感性问题进行调查,已经成为社会敏感性问题调查的常用手段。在一定条件下,改进后模型的精度比 Warner 模型的高,说明距离平衡抽样排除临近单元设计方法比 SRSWR 抽样设计方法更能真实地反映实际情况。

**关键词** 距离平衡抽样排除临近单元 Warner 模型 Horvitz-Thompson 估计量

**中图法分类号** O212.2; **文献标志码** A

## 1 符号说明

总体中有  $N$  个单元,样本容量为  $n$ ,  $\pi$  表示为总体中敏感属性(A)的比例( $\pi$ 未知)。令

$$Y_i = \begin{cases} 1, & \text{如果第 } i \text{ 个体具有敏感性} \\ 0, & \text{否则} \end{cases} \quad i = 1, 2, \dots, N。$$

一阶包含概率:  $\pi_i = \frac{n}{N}$ ; 二阶包含概率<sup>1</sup>:

$$\pi_{ij} = \begin{cases} 0, & i \text{ 与 } j \text{ 相邻;} \\ \frac{n(n-1)}{N(N-3)}, & i \text{ 与 } j \text{ 不相邻。} \end{cases}$$

### 1.1 BSEC 抽样设计<sup>[1]</sup>

总体  $Y_1, Y_2, \dots, Y_N$  中有  $N$  个单元,规定距离等于 1 时为临近单元,即  $Y_i$  与  $Y_{i+1}$  为临近单元,这里令  $Y_N$  与  $Y_1$  为临近单元,从该总体中抽取一个样本量为  $n$  的 BSEC 样本  $y_1, y_2, \dots, y_N$ ,记为  $S$ 。

### 1.2 随机化装置及回答技术

首先制作外形上完全相同的两种卡片:一种卡片上写有“你有敏感属性 A”;另一种卡片上写有“你没有敏感属性 A 吗?”,然后以比例  $p$  和  $1-p$  混合均匀放入盒子里,接着让事先抽好的(BSEC)样本个体随机抽取盒子里的卡片,根据抽到卡片上的

问题作出回答,如果有此属性就回答“Yes”否则回答“No”,回答完问题后将卡片再放入盒子里供下个被调查者使用,设样本个体所产生的随机化回答为  $z_1, z_2, \dots, z_n$ 。目的是估计有限总体中敏感属性个体

所占的比例( $\pi$ ):  $\pi = \frac{1}{N} \sum_{i=1}^N Y_i = \bar{Y} = u$ , 而  $m = \sum_{i=1}^n z_i$  表示回答“Yes”的个数。

$z_i = d_i y_i + (1 - d_i)(1 - y_i)$ , 其中

$$d_i = \begin{cases} 1, & \text{若抽取到第一个问题;} \\ 0, & \text{若抽取到第二个问题。} \end{cases}$$

$$E_R(d_i) = p, V_R(d_i) = p(1-p)。$$

随机化回答的方差:  $V_R(z_i) = p(1-p)。$

## 2 改进后的新模型

把 Horvitz 与 Thompson(1952 年)提出的对总体和  $Y$  的估计量<sup>[2]</sup>:

$$\hat{Y}_{HT} = \sum_{i=1}^n \frac{y_i}{\pi_i} \text{ 称 } \hat{Y}_{HT} \text{ 为 Horvitz-Thompson 估计量,}$$

应用到 Warner 模型<sup>3</sup>中,从而得到新估计量模型:

$$\hat{\pi}_{\text{BSEC}} = \frac{\frac{1}{N} \sum_{i=1}^n \frac{z_i}{\pi_i} - (1-p)}{2p-1} \quad (1)$$

### 3 新模型估计量的方差

定理 3.1 在 BSEC 抽样设计<sup>[1]</sup>下,

$$V(\hat{\pi}_{\text{BSEC}}) = \frac{\pi(1-\pi)}{n} \left[ 1 - \frac{(n-1)(2\rho_1-1)}{N-3} \right] + \frac{p(1-p)}{n(2p-1)^2} \quad (2)$$

证明:

$$V(\hat{\pi}_{\text{BSEC}}) = V_d E_R(\hat{\pi}_{\text{BSEC}}) + E_d V_R(\hat{\pi}_{\text{BSEC}})。$$

其中

$$V_d E_R(\hat{\pi}_{\text{BSEC}}) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j \neq i}^N (\pi_{ij} - \pi_i \pi_j) \times \frac{[(2p-1)Y_i + (1-p)][(2p-1)Y_j + (1-p)]}{\pi_i \pi_j} + \frac{1}{N^2} \sum_{i=1}^N \frac{1-\pi_i}{\pi_i} [(2p-1)Y_i - (1-p)]^2 \quad (3)$$

$$\text{从而} \sum_{i=1}^N \frac{1-\pi_i}{\pi_i} [(2p-1)Y_i + (1-p)]^2 = \sum_{i=1}^N \sum_{j>i}^N (\pi_i \pi_j) \left[ \frac{(2p-1)Y_i + (1-p)}{\pi_i} \right]^2 + \sum_{i=1}^N \sum_{j>i}^N (\pi_i \pi_j) \left[ \frac{(2p-1)Y_j + (1-p)}{\pi_j} \right]^2 \quad (4)$$

将式(4)代入式(3)得

$$V_d E_R(\hat{\pi}_{\text{BSEC}}) = (2p-1)^2 \frac{\sigma^2}{n} \left\{ 1 - \frac{(2\rho_1+1)(n-1)}{N-3} \right\} \quad (5)$$

式(5)中  $\rho_1$  为一阶环形连续相关系数<sup>[1]</sup>;  $\sigma^2$  为有限总体的方差。

$$E_d V_R(\hat{\pi}_{\text{BSEC}}) = \frac{p(1-p)}{n(2p-1)^2} \quad (6)$$

故以证毕。

### 4 效率比较

通过比较新模型的方差式(2)与 Warner 模型<sup>[3]</sup>的方差,定理 4.1 当  $n$  固定为大于 1 的常数,  $N$

$> 3, \rho_1 > -\frac{1}{N-1}$  时,

$$V(\hat{\pi}_{\text{BSEC}}) < V(\hat{\pi}_w)。$$

证明:由 Warner 模型<sup>[3]</sup>的方差知

$$V(\hat{\pi}_w) = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2}。$$

由式(2)知

$$V(\hat{\pi}_w) - V(\hat{\pi}_{\text{BSEC}}) = \frac{(n-1)(2\rho_1+1)}{N-3} > 0。$$

当上述条件成立时,新模型的效率比 Warner 模型<sup>[3]</sup>的效率更好,由此可知在一定条件下,距离平衡抽样排除临近单元设计方法比 SRSWR 抽样设计方法更能真实地反映实际情况。

### 参 考 文 献

- 1 Mandal B N, *et al.* A family of distance balanced sampling plans. *Journal of Statistical Planning and Inference*, 2009; 139: 860—874
- 2 冯士雍,施锡铨. 抽样调查—理论,方法和实践. 上海: 上海科学技术出版社, 1994
- 3 Warner S L. Randomized response: a survey technique for eliminating evasive answer bias. *Am Statist Assoc*, 1965; 60: 63—69

## The Application of Distance Balanced Sampling Plans Excluding Contiguous in Warner Model

DONG Hong-qin, CAI Nan

(Yantai Nanshan University foundation department, Yantai 264000, P. R. China)

**[Abstract]** Using randomized sampling to research the sensitive questions has become commonly used methods in society. Under certain conditions, the improved new model of efficiency higher than Warner model of efficiency. Distance balanced sampling plans excluding contiguous units could more reflected the actual situation than SRSWR sampling design method.

**[Key words]** Distance balanced sampling plans excluding contiguous units Warner model Horvitz-Thompson estimator