文章编号：1002-1566 (2013) 03 - 0425 - 08

Fay 平衡半样本的方差估计的基本理论和应用

吕萍

(北京大学中国社会科学调查中心，北京 100871)

摘要：方差估计是抽样调查的重要组成部分，在实际的抽样设计中，往往通过复杂的细分层提高样本的代表，平衡半样本方法成为方差估计的主要方法之一。本文主要阐述 Fay 平衡半样本方法的基本原因和应用，并对针对该方法存在的问题进行改进，最后通过一个实例模拟阐述其在实际调查中的应用。

关键词：方差估计；平衡半样本方法；Fay 平衡半样本方法；Fay 分组平衡半样本方法；Fay 部分平衡半样本方法

中图分类号：C811, O212 **文献标识码**：A

The Theory and Application of Fay Balanced Repeated Replication in Variance Estimation

LV Ping

(Institute of Social Science Survey, Peking University, Beijing 100871, China)

Abstract: The variance estimation is very important part in sample survey. In practice, we often use many stratums to improve representativeness of sample in sample design. Balanced repeated replication method is one of important methods in variance estimation. This paper expatiates the theory of Fay balanced repeated replication, then we improve this method. In the end, we use a practical example to show the application of this method.

Key words: variance estimation, balanced repeated replication, fay balanced repeated replication, fay grouped balanced repeated replication, fay partially balanced repeated replication

0 引言

现在越来越多的领域需要通过抽样调查获取信息，日益受到社会各个领域的关注。方差估计 [1] 是是抽样调查中评价调查数据质量的重要指标，方差估计方法包含方差公式计算法、泰勒级数展开法和重抽样方法，其中重抽样方法 [2-3] 主要有随机组、刀切法 (Jackknife)、平衡半样本 (Balanced Repeated Replication) 和自助法 (Bootstrap) 等方差计算方法。由于实际的大型调查一般采用分层多阶段不等概的复杂抽样设计，目的是为了目标变量的均值、总量、百分数、相关系数等多种类型的参数估计，若采用传统的公式法，计算过程十分复杂，重抽样方法以其稳健性和简便性得到了广泛的应用。

收稿日期: 2011 年 1 月 27 日 **收到修改稿日期**: 2012 年 1 月 18 日
基金项目: 论文是国家社科基金 (11CTJ005) 和全国统计科学研究重点项目 (2012LZ055) 的阶段性成果。

中国的抽样调查起步较晚，大型的抽样调查同样采用分层多阶段不等概的复杂抽样设计。在进行抽样设计时，往往利用所有可以获得的辅助信息对抽样框进行分层，进而提高估计量的精度。但在实际调查中，过多的分层会导致部分层中有较少的初级单元落入，使用随机组和刀切法往往无法得到有效的方差的估计量。而平衡半样本方法 [4] 适用于这种每个层中有较少初级单元的情况，有效提高了方差估计量的精度。平衡半样本方法的实质是通过调整每个层中的样本单元的抽样权数，利用不同权数得到目标参数的估计量，进而利用目标估计量的变异性得到方差估计量。这种方法已经在国内的几个大型调查中使用，但是该方法使每个层的权数是 2 倍或零倍的权数，一方面在每次的计算目标变量的估计量时舍去了一部分样本信息造成样本信息的浪费，另一方面权数的变异较大，容易高估方差估计量。当目标变量是比例估计量且每个层中的样本量比较小时，利用平衡半样本方法得到的方差估计量倾向于高估，因此，Robert Fay[5] 于 1989 年对这种平衡半样本方差估计方法进行了修正，通过减弱权数的变化提高方差估计量的精度，这种方法更适合中国抽样调查的国情。Andy Sadler 和 Helen Chen (2010)[6] 用 PPI (The Producer Price Index) 数据比较 Fay 平衡半方法以及其它重抽样方法并进行比较，Jonathan J. Lisic 和 Omolola E. Ojo (2008)[7] 在 NCS (National Compensation Survey) 中应用 Fay 平衡半样本方法计算方差，Van L. Parsons (2010)[8] 将 Fay 平衡半方法应用到国际健康调查 (National Health Interview Survey, NHIS) 中，并利用 1997–2005NHIS 的成人样本的方差。Fay 平衡半样本方法也用于无回答数据的方差计算中，Eric V. Slud 和 Yves Thibaudeau (2008)[9] 比较了无回答情况下平衡半样本方法和逆概率方法，以及结合 Laplace 和平衡半方法 [10]，并应用于 IPP (Income and Program Participation) 中。在国内，这种方法使用并不普遍，近几年逐渐开始应用，例如中国青少年儿童发育特征的综合抽样调查 [11] 等。赵馨 (2011)[12] 介绍了平衡半样本方法在 Warner 模型中的应用，高歌、范玉波 (2010)[13] 推导出二分类敏感随机应答技术 Warner 模型在整群、分层整群抽样总体比例的方差计算公式，吕萍 (2011)[14] 介绍重权数方法时简单提及 Fay 平衡半样本方法，金勇进 (2009)[15] 介绍了缺失数据中的平衡半样本方法. 但是，在实际调查中，由于实际抽样设计和实际调查的复杂性，Fay 平衡半样本方法的实际应用更加复杂，本文将主要介绍 Fay 平衡半样本方法的基本理论和实际应用，并对该方法进行改进，使之更加符合中国实际调查。

1 Fay 平衡半样本方法的基本原理

平衡半样本方法是针对每个层中有两个初级单元的复杂调查的方差估计方法，假设对总体 $N=\sum_{h=1}^{L}N_h$ 采用分层随机抽样，在每个层中有放回随机抽取 2 个样本单元，设 y_{h1} 和 y_{h2} 是第 h 层的样本观测值 $(h=1,\cdots,L)$。从每个层中抽取一个样本单元组成一个半样本，则共有 2^L 个半样本。利用 2^L 个半样本的均值的平均值得到目标变量的均值估计量 (本文以均值估计量为例，其它估计量的计算与此相同) 是

$$\overline{y}_{st,\alpha}^{BRR}=\sum_{h=1}^{L}W_h(\delta_{h1\alpha}y_{h1}+\delta_{h2\alpha}y_{h2}),$$

其中,

$$\delta_{h1\alpha}=\begin{cases}1, & \text{如果第 } h \text{ 层中的第一个单元被选入第 } \alpha \text{ 个半样本}, \\ 0, & \text{其它},\end{cases}\qquad \delta_{h2\alpha}=1-\delta_{h1\alpha}.$$

利用 $\overline{y}_{st,\alpha}$ 之间的变异来构造方差估计量可得到 $V(\overline{y}_{st})$ 的无偏估计量。但是，当层数 L 较

大时, 计算过于庞大, 呈几何级数增长。为简化计算, 并尽量保留原有信息, Plackett 和 Burman (1946) 给出 $k \times k$ 阶正交 Hadamard 矩阵 $(\delta_{kh})_{K\times K}$ 的方法, 其中 K 为 4 的倍数, 第 α 行第 h 列的 $+1$ 表示第 α 个半样本是层 h 的第一个单元, -1 表示第 α 个半样本是层 h 的第二个单元, 且满足 $\sum_{\alpha=1}^{k}\delta_h^{(\alpha)}=0$ 且 $\sum_{\alpha=1}^{k}\delta_h^{(\alpha)}\delta_{h'}^{(\alpha)}$, $h'\neq h$。平衡半样本的正交选择矩阵是通过去掉正交矩阵 $(\delta_{kh})_{K\times K}$ 的全 1 列后随机选取 H 列得到, 即 $(\delta_{kh})_{H\times K}$ $(H+1\leq k\leq H+4)$。

由上可知, 平衡半样本方法的实质上是通过调整每个层中的样本单元的权数, 利用不同权数计算目标参数的估计量, 进而利用估计量的变异性得到方差估计量。但是, 这种方法使每个层的权数是 2 倍或零倍的权数, 一方面在每次的计算目标变量的估计量时舍去了一部分样本信息造成样本信息的浪费, 另一方面权数的变异较大, 容易高估方差估计量。当目标变量是比例估计量且每个层中的样本量比较小, 利用平衡半样本方法计算的容易高估方差估计量。

Robert Fay 通过减弱权数变化的方法对传统的平衡半样本方法进行修正, 进而提高目标估计量的方差估计量的精度, 引入 Fay 因子 ε $(0\leq\varepsilon<1)$, 使每个层中的初级单元的权数的变化由 $(2w_{l_1},0\times w_{l_2})$ 变化减弱为 $\big((1+\varepsilon)w_{l_1},(1-\varepsilon)w_{l_2}\big)$, 利用所有的样本信息提高了样本的利用率, 同时也避免在计算比例估计量时半样本的样本量过少带来的一系列问题。所以, Fay 平衡半样本方法在方差估计中具有更为广泛的作用。以均值的估计量为例, 首先均值的目标变量的估计量是

$$\overline{y}_{st,\alpha}^{FBRR}=\sum_{h=1}^{L}W_h(\delta_{h1\alpha}y_{h1}+\delta_{h2\alpha}y_{h2}, \tag{1}$$

其中

$$\delta_{h1\alpha}=\begin{cases}1+\varepsilon, & \text{如果第 } h \text{ 层中的第一个单元被选入第 } \alpha \text{ 个半样本},\\ 1-\varepsilon, & \text{其它},\end{cases}\qquad \delta_{h2\alpha}=1-\delta_{h1\alpha},$$

W_h 是第 h 样本单元的权数。

因为 Fay 平衡半样本方法利用所有的样本信息导致得到的估计量的差异较小, 在方差计算时利用 $1/\varepsilon^2$ 系数来调整方差, 即均值的方差估计量是

$$v^{FBRR}(\overline{y}_{st}^{FBRR})\triangleq\frac{1}{k\varepsilon^2}\sum_{\alpha=1}^{k}(\overline{y}_{st,\alpha}-\overline{y}_{st})^2, \tag{2}$$

其中 ε 的选择是十分重要的, 当 ε 趋向于零时, 该方法近似于泰勒线性方法, 当 ε 趋向于 1 时, 近似于标准的平衡半样本方法。在实际调查中, 选择 $\varepsilon=0.5$ 可以得到比较好的估计结果。由于平衡半样本的余集也是平衡半样本, 因此在实际使用时可以用平衡半样本的余集提高目标变量的精度, 即

$$\overline{y}_{st,\alpha}^{c}=\sum_{h=1}^{L}W_h(\delta_{h2\alpha}y_{h1}+\delta_{h1\alpha}y_{h2}),$$

其中

$$\delta_{h1\alpha}=\begin{cases}1+\varepsilon, & \text{如果第 } h \text{ 层中的第一个单元被选入第 } \alpha \text{ 个半样本},\\ 1-\varepsilon, & \text{其它},\end{cases}\qquad \delta_{h2\alpha}=1-\delta_{h1\alpha}.$$

其方差估计量是

$$v^{c}(\overline{y}_{st})\triangleq\frac{1}{k\varepsilon^2}\sum_{\alpha=1}^{k}(\overline{y}_{st,\alpha}^{c}-\overline{y}_{st})^2.$$

由此，得到充分利用平衡半样本和余集半样本的目标变量的估计量是

$$\overline{\overline{y}}_{st,\alpha}=\frac{\overline{y}^{c}_{st,\alpha}+\overline{y}_{st,\alpha}}{2},\qquad \overline{v}(\overline{y}_{st})=\frac{v^{c}(\overline{y}_{st})+v(\overline{y}_{st})}{2}.$$

Fay 平衡半样本方法同样要求每个层中的样本量是 2，但是，在实际中往往无法满足这个要求，即有些层中的样本量多于 2 个，有些层中却只有 1 个初级样本单元。在实际调查中，可以利用划分虚拟层的方法，将每个层中的初级样本单元随机分成两个虚拟层，或将邻近的两个层合并为一个虚拟层，然后利用 Fay 平衡半样本方法。在实际调查中包含层内初级单元的个数大于 2 或小于 2 的情况。

2 Fay 分组平衡半样本方法

在用 Fay 平衡半样本进行方差估计时，若某些层中的样本个数大于 2，无法直接用 Fay 平衡半样本方法。此时，可以用 Fay 分组平衡半样本方法 [16] (Fay Grouped Balanced Repeated Replication) 或混合正交矩阵方法 [17] (Mixed Orthogonal Matrix) 进行方差估计。

2.1 Fay 分组平衡半样本方法

Fay 分组平衡半样本方法是将层内的初级样本单元随机分为两个虚拟组，其中初级但愿的个数分别为 $m_h[n_h/2]$，([ ] 表示取整) 和 n_h-m_h，然后按照 Fay 平衡半样本方法得到目标参数的均值和方差估计量。则均值估计量是

$$\overline{y}^{FGB}_{st,\alpha}=\sum_{h=1}^{L}W_h(\delta_{h1\alpha}y_{h1}+\delta_{h2\alpha}y_{h2}),$$

其中

$$\delta_{h1\alpha}=\begin{cases}1+\varepsilon\Big(\dfrac{n_h-m_h}{m_h}\Big)^{1/2}, & \text{如果第 } h \text{ 层中的第一个单元被选入第 } \alpha \text{ 个半样本，}\\ 1-\varepsilon\Big(\dfrac{m_h}{n_h-m_h}\Big)^{1/2}, & \text{其它,}\end{cases}$$

$$\delta_{h2\alpha}=1-\delta_{h1\alpha}.$$

均值的方差估计量是：

$$v^{FGB}(\overline{y}^{FGB}_{st})\triangleq\frac{1}{k\varepsilon^2}\sum_{\alpha=1}^{k}(\overline{y}_{st,\alpha}-\overline{y}_{st})^2. \tag{3}$$

层内虚拟组的划分产生了虚拟组的组内方差和组间方差，导致方差估计量的高估。为了提高估计精度，应尽量减小随机组内的组内方差，即划分的两个虚拟组组内的差异要尽量小，组间的差异要尽量大。在实际调查中，当层内样本单元结构的信息无所获得时，此时为了提高目标变量的估计精度，将层内虚拟组的划分过程重复 r 次，则均值估计量是

$$\overline{y}^{FGBR}_{st,\alpha}=\frac{\sum_{h=1}^{L}W_h(\delta_{h1\alpha}y_{h1}+\delta_{h2\alpha}y_{h2})}{r},$$

其中

$$\delta_{h1\alpha}=\begin{cases}1+\varepsilon\Big(\dfrac{n_h-m_h}{m_h}\Big)^{1/2}, & \text{如果第 } h \text{ 层中的第一个单元被选入第 } \alpha \text{ 个半样本，}\\ 1-\varepsilon\Big(\dfrac{m_h}{n_h-m_h}\Big)^{1/2}, & \text{其它,}\end{cases}$$

$$\delta_{h2\alpha}=1-\delta_{h1\alpha}.$$

均值的方差估计量是：

$$v^{FGBR}(\overline{y}_{st,\alpha}^{FGBR}) \triangleq \frac{1}{rk\varepsilon^2}\sum_{\alpha=1}^{k}(\overline{y}_{st,\alpha}-\overline{y}_{st})^2.$$

上述方法有效地提高了方差估计量的精度，在实际中，也可以利用平衡半样本的余集更加有效的利用样本信息提高估计精度。

2.2 Fay 混合正交矩阵的平衡半样本方法

Fay 分组平衡半样本方法在一定程度上解决了层内初级样本单元多于 2 个的问题，但是虚拟分组的划分不可避免地导致了一定的误差，为了提高估计的精度和有效性，可以利用混合正交的平衡半样本方法。假设一个分层的抽样设计，某些层中的抽样单元的个数大于 2，利用 $(K, n_1\times n_2\times n_2\times\cdots\times n_l)$ 的混合正交矩阵替代正交矩阵，其中第 k 列包含 n_k 个不同的样本单元，每一行表示一个重复的混合正交样本，得到的方差估计量是

$$v^{FMB}(\overline{y}_{st}^{FMB}) = \frac{1}{k\varepsilon^2}\sum_{\alpha=1}^{k}(\overline{y}^{\alpha}-\overline{y}^{(\cdot)})^2, \tag{4}$$

其中，$\overline{y}^{\alpha}=\sum_h(w_h/(n_h-1))y_{i(k,h)}$，$\overline{y}^{(\cdot)}=\sum_\alpha\overline{y}^{\alpha}/K$。

该方法提高了估计的精度，但是计算过程比较复杂，对估计量的有效性有较高的要求，在实际中使用较少。相反，分组平衡半样本方法由于简便性和有效性得到了广泛的应用。

在一个复杂调查中，因为在抽样设计中没有考虑方差的估计问题，导致出现过多的层，此时的计算量是十分大的，为了提高计算效率需要对层进行合并，尤其是当层内样本单元个数是 1 时，合并层是一种比较好的选择。

3 Fay 部分平衡半样本方法

在实际调查中，有效的分层可以极大地提高估计的精度，但是当层数太多时计算量十分庞大，为了操作的有效性和经济性，需要将 L 个层进行适当合并，从而降低计算量。例如，将 L 个层合并成 G $(L<K)$ 个虚拟层，设 L_g 是虚拟层集，利用 $R\times G$ 正交矩阵得到平衡半样本，这种方法称为 Fay 部分平衡半样本 [18](Fay Partially balanced Repeated Replication) 方法，从而得到的估计量是

$$\overline{y}_{st,\alpha}^{FPB} = \frac{\sum_{h=1}^{L}W_h(\delta_{h1\alpha}y_{h1}+\delta_{h2\alpha}y_{h2})}{r},$$

其中

$$\delta_{h_G1\alpha}=\begin{cases}1+\varepsilon, & \text{如果第 } h_G \text{ 层中的第一个单元被选入第 } \alpha \text{ 个半样本},\\ 1-\varepsilon, & \text{其它},\end{cases}\qquad \delta_{h2\alpha}=1-\delta_{h1\alpha}.$$

方差估计量是：

$$v^{FPB}(\overline{y}_{st,\alpha}^{FPB}) \triangleq \frac{1}{k\varepsilon^2}\sum_{\alpha=1}^{k}(\overline{y}_{st,\alpha}-\overline{y}_{st})^2. \tag{5}$$

这种方法不仅提高计算效率，同时有效地解决了层内只有一个初级样本单元的方差估计问题。但是，这种方法容易导致方差估计量的高估，因为在合并层的过程中引入了层间偏差。

为提高估计的精度，应尽可能的合并相似的层，例如利用有效的辅助信息，合并邻近相似的层。

4 改进 Fay 平衡半样本方法

平衡半样本方法一般假定层内的样本是简单随机抽样或有放回的不等概率抽样得到的。但是，在实际调查中一般用不放回的系统抽样或不放回的 PPS 抽样方法来提高样本的代表性。此时，通常假设抽样设计是不放回的抽样使用平衡半样本方法估计方法，因此常常导致估计量的高估。利用辅助信息对每次重复抽取的半样本进行校准，即利用校准权数的方法提高估计的精度。

权数调整是抽样调查的重要内容之一，平衡半样本方法实质上就是在基础权数的基础上通过利用重抽样方法对权数进行调整得到估计量。调查数据的权数调整一般包含各阶段的基础权数、无回答调整和事后分层调整权数等，在实际中，为了简化运算，首先对调查数据进行基础设计权数调整、无回答权数调整和事后分层权数调整，然后使用上述 Fay 平衡半样本方法计算方差估计量，即利用的第 r 个平衡半样本的最终权数是

$$w_{end(r)} = w_{base} \times w_{nor} \times w_{post}, \tag{6}$$

其中 $w_{base}, W_{nor}, w_{post}, w_{end(r)}$ 表示基本设计权数、无回答调整权数、事后分层调整权数和最终的平衡半样本权数。

但是，由式 (6) 得到的估计量往往不是无偏估计量。因为每次得到的 Fay 平衡半样本的样本的结构发生了变化，此时原来的无回答和事后分层调整的权数是错误的。所以 Fay 平衡半样本方法需要在原始样本的基础设计权数基础上，对每次的 Fay 平衡半样本进行无回答权数和事后分层权数调整或利用其他辅助信息进行校准，即

$$w'_{end(r)} = w_{base} \times w'_{nor} \times w'_{post}, \tag{7}$$

其中 $W'_{nor}, w'_{post}, w'_{end(r)}$ 表示第 r 个平衡半样本的无回答调整权数、事后分层调整权数和最终的平衡半样本权数。从而提高各个 Fay 平衡半样本的目标变量的估计精度，进而提高方差估计量的精度。本文将这种方法称为修正 Fay 平衡半样本方法。同样的可以将用在分组 Fay 平衡半样本方法和 Fay 部分平衡半样本方法中。

5 实例

在实际调查中，由于实际情况的复杂性，通常需要结合各种 Fay 平衡半样本方法。下面以实例模拟说明 Fay 平衡半样本方法在实际中的应用。

5.1 数据说明

采用分层三阶段不等概率的复杂抽样设计对中国 6–15 岁青少年儿童发育特征进行抽样调查。第一阶段，将全国的区县分成 14 层，在每个层中抽取 100 个区县样本；第二阶段，在样本区县中抽取学校样本；第三阶段，在每个学校样本中随机抽取学生样本。采用 Fay 平衡半样本方法计算各个年级各个区域的某测试得分的均值估计量和方差估计量。

5.2 数据分析

一些软件中已经出现了平衡半样本方法的程序包或模块，例如 R 软件中的程序包 "survey" 的 svrepdesign 命令可以进行平衡半样本方法的方差估计。但是，实际调查往往比较复杂，无法简单、直接的用现成的软件进行计算。在本次调查中，需要结合前面讲述的各种 Fay 平衡半样本方法，即对于层中初级单元个数少于 2 的需要与邻近的层进行合并，初级单元的个数大于 2 的则需要随机划分法为 2 个虚拟组，增加正交矩阵的维数，并重复执行这个过程以提高估计的精度。这个过程的计算比较复杂，下面用 R 软件编程计算各个年级各个区域的目标变量的估计量。

(1) 将 14 层中的区县进行修正，即将区县个数小于 2 的层就近合并，将区县个数大于 2 的层随机划分为 2 个虚拟组。

(2) 用 64 阶正交矩阵 (由于 (1) 中多次划分虚拟组，所以需要增加正交矩阵的维数) 使用 Fay 分组平衡半样本方法对目标变量的均值估计量和方差估计量进行计算。

(3) 重复执行 (1)(2)100 次，将每次执行的结果进行平均即为最终的估计量，结果如下表 1 所示。

表 1 利用各种 Fay 平衡半样本得到的结果

类别	均值	方差	类别	均值	方差
第一类区域的一年级	8.4895	1.9088	第三类区域的一年级	5.6416	0.0259
第一类区域的二年级	10.3346	0.0979	第三类区域的二年级	7.8216	0.0757
第一类区域的三年级	12.2267	0.2613	第三类区域的三年级	10.2319	0.0580
第一类区域的四年级	13.9607	0.5074	第三类区域的四年级	11.3056	0.0548
第一类区域的五年级	15.2732	0.3394	第三类区域的五年级	13.1257	0.0481
第一类区域的六年级	16.2247	0.0137	第三类区域的六年级	14.0743	0.0778
第一类区域的七年级	16.8702	0.2010	第三类区域的七年级	14.4021	0.0553
第一类区域的八年级	17.5474	0.1863	第三类区域的八年级	15.1380	0.0569
第一类区域的九年级	17.8189	0.1017	第三类区域的九年级	16.1583	0.0648
第二类区域的一年级	6.2660	0.0530	第四类区域的一年级	5.6811	0.1121
第二类区域的二年级	8.7801	0.0799	第四类区域的二年级	7.7421	0.1165
第二类区域的三年级	11.3997	0.1036	第四类区域的三年级	10.0858	0.0442
第二类区域的四年级	12.8911	0.0606	第四类区域的四年级	11.7227	0.0790
第二类区域的五年级	14.4682	0.0654	第四类区域的五年级	13.0938	0.0294
第二类区域的六年级	15.2963	0.0712	第四类区域的六年级	14.2030	0.0308
第二类区域的七年级	16.0752	0.0712	第四类区域的七年级	14.5514	0.0339
第二类区域的八年级	16.3978	0.0310	第四类区域的八年级	15.6508	0.0247
第二类区域的九年级	17.1689	0.0760	第四类区域的九年级	16.5035	0.0216

由表 1 可知，利用 Fay 平衡半样本方法得到目标变量的均值估计量和方差估计量，其估计量的效果相对较好，而且这个过程可以重复的计算其它目标变量的方差估计量。

6 总结

Fay 平衡半样本方法主要针对分层多阶段抽样设计的复杂调查，更加适用于中国的国情。本文主要介绍 Fay 平衡半样本方法的基本理论及其在实际调查中的应用，并对这种方法提出了改进，最后用一个实例模拟说明 Fay 平衡半样本方法在实际中的应用和优势。Fay 平衡半样本方法具有十分广泛的应用，其在缺失数据中的应用及改进将是作者今后研究的重点。

[参考文献]

[1] Wolter K. Introduction to Variance Estimation (2 Ed.) [M]. New York: Springer-Verlag, 2007.

[2] Rao J N K, Wu C F J, and Yue K. Some recent work in resampling methods [J]. Survey Methodology, 1992, 18: 209–217.

[3] Krewski D, Rao J N K. Inference from stratified samples: Proper-ties of the linearization, jackknife, and balanced repeated replication methods [J]. Annals of Statistics, 1981, 9: 1010–1019.

[4] Rao J N K, Shao J. On balanced half sample variance estimation in stratified sampling [J]. Journal of the American Statistical Association, 1996, 91: 343–348.

[5] Judkins D R. Fay's method for variance estimation [J]. JOS, 1990, 6: 223–229.

[6] Andy Sadler, Helen Chen. Comparison of Variance Estimation Methods Using PPI Data [C]. JSM 2010 Proceeding, 2010.

[7] Jonathan J Lisic, Omolola E Ojo. Application of Fay's Method for Variance Estimation in the National [C]. Compensation Survey Benefits Products [A]. JSM 2008 Proceeding, 2008.

[8] Van L Parsons. Application of a Fay-Modified Balanced Repeated [C]. Replication Method to the National Health Interview Survey (NHIS) [A]. JSM 2010 Proceeding, 2010.

[9] Eric V Slud, Yves Thibaudeau. BRR versus Inclusion-Probability Formulas for Variances of Nonresponse Adjusted [C]. Survey Estimates [A]. JSM 2008 Proceeding, 2008.

[10] Yves Thibaudeau, Eric V Slud. The Method of Laplace and BRR: A Hybrid Variance Estimation Method in survey [C]. JSM 2008 Proceeding, 2008.

[11] 董奇. 中国儿童青少年心理发育特征调查项目技术报告 [M]. 北京: 科学出版社, 2011.

[12] 赵馨, 闫在在, 魏福红, 唐俊, 张景. 平衡半样本的方差估计方法在 Warner 模型中的应用 [J]. 科技创新导报, 2011,(13): 96–96.

[13] 高歌, 范玉波. 分层整群抽样的 Warner 模型 RRT 技术及其对大学生婚前性行为调查中的应用 [J]. 数理统计与管理, 2010, (01): 185–190.

[14] 吕萍. 重权数在复杂调查的方差估计中的应用 [J]. 统计研究, 2011, 28(2): 93–97.

[15] 金勇进, 邵军著. 缺失数据的统计处理 [M]. 北京: 中国统计出版社, 2009.

[16] Rao J N K, Shao J. Modified balanced repeated for complex survey data [J]. Biometrika, 1999, 86: 403–415.

[17] Sitter R R. Balanced repeated replications based on orthogonal multi-arrays [J]. Biometrika, 1993, 80: 211–221.

[18] Lee K H. The use of partially balanced designs for half-sample replication method of variance estimation [J]. Journal of the American Statistical Association, 1972, 67: 324–334.