招生考试研究

Admission & Examination Research

2013 1

总第17辑

上海市教育考试院主办

上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

巍巍交大 百年书香
www.jiaodapress.com.cn
bookinfo@sjtu.edu.cn

策划编辑 腾 飞
责任编辑 腾 飞
封面设计 陈燕静
责任营销 刘依萱

Admission & Examination Research

ISBN 978-7-313-10034-4

9 787313 100344 >

定价：25.00元

招生考试研究

上海市教育考试院　主办

（2013-1）

上海交通大學出版社

图书在版编目(CIP)数据

招生考试研究. 2013 年. 第 1 辑/上海市教育考试院主编. —上海:上海交通大学出版社,2013

ISBN 978-7-313-10034-4

Ⅰ. 招... Ⅱ. 上... Ⅲ. 招生—考试—中国—文集 Ⅳ. G473. 2-53

中国版本图书馆 CIP 数据核字(2013)第 142508 号

招生考试研究 2013 年

第 1 辑

上海市教育考试院 **主编**

上海交通大学出版社出版发行

(上海市番禺路 951 号 邮政编码 200030)

电话:64071208 出版人:韩建民

上海交大印务有限公司 印刷 全国新华书店经销

开本:787mm×1092mm 1/16 印张:6.75 字数:159 千字

2013 年 7 月第 1 版 2013 年 7 月第 1 次印刷

ISBN 978-7-313-10034-4/G 定价:25.00 元

招生考试研究

（上海市教育考试院主办）

通讯地址： 上海市钦州南路500号 《招生考试研究》

邮　　编： 200235

电　　话： 021-64946609　021-64946068

传　　真： 021-64946679

电子邮箱： zkyj@shmeea.edu.cn　xdllf2000@163.com

目　　录

Contents

特稿

群配对刀切法伪估计值中之异常值及其解决方法

钱家和

[内容摘要] 在整群抽样的样本中，如果群内样本呈现高度相关，对估计值的精度会有影响，尤其在方差估计中采用了群配对刀切法的时候，对方差估计会有较大影响。本文分析了造成群配对不平衡的几个因素，在刀切法伪估计值中出现异常值的机制，以及分析了配对群不平衡是如何影响方差估计并造成方差超估的。本文用公式表述了不平衡配对对伪估计值的影响，并用实例加以说明。本文对原方法作了改进，提出了部分群配对刀切法，并用实际数据对改进前后的两种方差估计方法的效果进行了比较。

[关 键 词] 整群刀切法；群配对刀切法；伪估计值；重复样本估计值中的异常值

一、概　　述

在教育测量的数据分析中，重复抽样方法，如刀切法和 bootstrap 法，常被用来估计统计量的方差。之所以用重复抽样方法而不常用公式，是因为教育调查的数据通常是由复杂抽样方法搜集的，而方差估计的公式往往过于复杂，或者缺少运用公式所必需的有关数据。重复抽样方法虽然运算量大，但可靠易行（Wolter，2003）。整群刀切法（grouped jackknife replicate resampling method，GJRR）是典型的重复抽样方法（Shao & Tu，1995；Tukey，1958；Wolter，2003）；群配对刀切法（paired grouped jackknifing，PGJRR）是对整群刀切法的一个改进，其主要改动是在抽样的过程中，将两个（或两个以上）的样本群配对并置于同一个方差层（variance stratum）之中（Kish & Frankel，1974；Krewski & Rao，1981）。群配对刀切法广泛应用于教育测量调查的分析过程中，比如 NAEP，PISA，PIRLS，以及 TIMSS①（Allen，Donoghue，& Schoeps，2001；Martin，Mullis，& Kennedy，2003；Neidorf，Binkley，Gattis，& Nohara，2006；Nohara，2001；Rust，1985）。

作者简介：钱家和，男，博士，美国教育考试中心（ETS）统计理论与计量心理研究所高级研究员。译者为钱家和和李淑红，李淑红博士是美国教育考试中心（ETS）英语语言测试中心计量心理学家。

① NAEP：The National Assessment of Educational Progress，美国国家教育进展评估；PISA：the Program for International Student Assessment，国际教育评估计划；PIRLS：the Progress in International Reading Literacy Study，国际阅读教学进展；TIMSS：the Trends in International Mathematics and Science Study，国际数学与科学教育进展。

教育测量抽样调查的一个典型设计是分层多阶段抽样。比如 NAEP 的州样本采用了与群样本量成正比的二阶段抽样(probability-proportional-to-size,PPS),这里群通常指学校,而学生则用简单随机抽样方法从被抽中的学校中抽取。由于抽取概率有所不等,有些个人或学校会拒绝参加,要取得无偏估计,在计算估计量和运用重复抽样方法中,必须采用加权的样本估计量(Cochran,1977;Hansen & Tepping,1985)。

虽然群配对刀切法应用广泛,但是其伪估计值(pseudo estimate)中经常会出现异常值(或称离群值,outlier),这是应用中的一个问题。每当异常值出现,方差的估计就会偏大。本文旨在分析在群配对刀切法的伪估计值中出现异常值的原因,并提出解决出现的异常值的改进方法,即局部群配对刀切法(partially paired grouped jackknifing,PPGJRR)。

在分层抽样中应用群配对刀切法时,我们通常将同一层中的群进一步排序,即作隐性分层(implicitly stratified)(Kish,1965);然后 PPS 抽取一定数量的群,将相邻近的两个群配成对,每一对形成一个方差层(variance stratum)。在分析中,配对群会被当作是从现显分层结构(explicitly stratified)中抽取的。计算时,将两个群随机去掉一个,再将留下的群中的样本权加倍(Allen *et al.*,2001)。有时一个方差层会分到三个群,这时可以去掉一个群,将剩下的两个群中样本的权加半倍(Allen *et al.*,2001)。

在方差估计中,我们要求配对的两个群的样本量相似,并且两个群中的样本基本服从同一个分布,之间差别仅仅由于抽样造成。在这种条件下,群配对刀切法通常被认为是适当的(Westat,2007)。但是,配对的两个群的样本量和样本值往往会很不平衡或很不相似。并且这种不均匀往往是在群的自然形成过程中产生的。比如,一个学区往往只有一个大型学校,重点学校也只侧重某些方面的教学。当出现异常值,其影响往往很大,造成估计量的不稳健性(Tukey,1960)。方差估计是二阶中心矩,对异常值非常敏感(Barnett & Lewis,1978;Efron & Gong,1983)。特别是群配对刀切法,对异常值是十分敏感的。

本文的前面部分阐明配对的两个群之间的不平衡,以及其对群配对刀切法方差估计的影响,然后用经验数据的例子来说明这种不平衡的存在和影响。经验数据表明,将群隐性排序,再抽群配对,并不能保证配对的两个群达到满意的平衡。为了减少这种不平衡,可以采用改进的局部群配对刀切法来解决异常值的问题。当配对群不平衡并造成伪估计的异常值时,表明两个群并不服从同一个分布,因而应该认为两个群分属于不同的方差层。这时,我们将群随机平均分成两个子群,让子群配成一对,再按群配对刀切法的程序进行方差估计。采用局部群配对刀切法,会去除原来的异常值,偏大的方差估计就比较正常。这种方差估计的结果是稳健的。

本文第一节概述在应用群配对刀切法时,伪估计值中出现异常值的问题,以及解决这个问题的思路。第二节介绍整群刀切法和群配对刀切法,分析配对群的不平衡对方差估计的影响,介绍判断刀切伪估计中的异常值的准则,并且提出改进了的局部群配对刀切法。第三节用实际数据表明群配对不平衡的存在,以及刀切伪估计中的异常值的存在,并且用实例对群配对刀切法改进前后的结果进行比较。第四节是有关的总结和一些评论。

二、问题及其解决方法

(一) Hurwitz-Thompson 估计量

一个调查的样本可以用不同的抽样方法抽取,比如分层抽样、整群抽样、多阶段抽样和等

距抽样(Cochran,1977)。复杂数据就是指用这些方法及其组合抽取的一个样本(Kish,1965)。NAEP 的州公立学校的样本就是用二阶段 PPS 抽取的(Allen,Donoghue,& Schoeps,2001)。

虽然许多调查的样本设计都想采用等概率抽样方法,但由于实际情况的限制,我们往往对总体的不同部分要采用不同的抽样方法。比如,公立学校与私立学校要用不同的抽样框,个别学校拒绝参加调查等其他因素。结果样本的抽取概率会不相等,这时要得到无偏估计,就应该应用加权估计量。

本文将总体定义为$R=\{x_i \mid i=1,2,\cdots,N\}$,其样本量为 N。从$R$中抽取的样本定义为$r=\{(x_i,w_i) \mid i=1,2,\cdots,n\}$,样本量为$n$;其中$w_i$为样本$i$的权,其基本部分等于样本被抽取概率的倒数(Allen *et al.*, 2001)。总体统计量$\widetilde{X}$的一个 Hurwitz-Thompson 估计量$\tilde{x}=\sum_{i=1}^{n} w_i x_i$则是无偏的(Cochran,1977)。总体平均量$\overline{X}$则可用比估计量

$$\bar{x}=\frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \tag{1}$$

来估计。虽然$\bar{x}$是有偏的,偏误为$O\left(\frac{1}{n}\right)$,其随着$n$的增大而减少,因而是渐近无偏的(Cochran,1977)。

(二) 整群刀切法

整群刀切法的一个优点就是可以提高方差估计的运算效率,而且这种方法可以用于某些对于刀切法无效的估计量的方差估计,比如百分位数的估计量(Shao & Wu,1992)。整群刀切法的方差估计和其他重复抽样方法(如 bootstrapping)是一致的(Hansen & Tepping,1985)。近来,整群刀切法被拓展到对一个统计过程的方差误的估计,这些过程包括对 IRT 模型的估计和 IRT 的联接过程等。这种整群刀切法被称为完过程整群刀切法(complete grouped JRR method,CGJRR)(Haberman,Lee, & Qian,2009)。

应用整群刀切法,所有样本要被分成大小相似的G个群(通常$G\geqslant 50$)。对抽样调查的样本,整群刀切法里的群(group)往往就是初级的抽样单位(也称为样本群,cluster)。在用二阶段抽样得到的州教育测试的样本中,这种群往往就是学校,而学校也就是第一阶段的抽样单位。

将n_l定义为群l的样本量。样本r的总样本量则等于各群样本量之和,$n=\sum_{l=1}^{G} n_l$。在样本中去除一个群就形成了整群刀切法的一个重复样本(replicate samples)。重复样本$r_{(g)}$就是在样本r中去除群g而得到,其样本量为$n_{(g)}=\sum_{l\neq g} n_l$。全部可以形成$G$个重复样本。对每一个重复样本$r_{(g)}$都可以用 Hurwitz-Thompson 估计量得到一个平均数的伪估计值

$$\bar{x}_{(g)}=\frac{\sum_{l\in r_{(g)}} w_l x_l}{\sum_{l\in r_{(g)}} w_l}。$$

这样整群刀切法的方差估计就可以表达为

$$v_{JRR}(\bar{x})=\frac{(G-1)}{G}\sum_{g=1}^{G}(\bar{x}_{(g)}-\bar{x}_{(.)})^2, \tag{2}$$

此处 $\bar{x}_{(.)}$ 是所有 $\bar{x}_{(g)}$ 的平均值，不过这仅是一种估计，在(2)中的 $\bar{x}_{(.)}$ 还可以是全样本的加权平均估计（Haberman *et al.*，2009；Wolter，2003）。

（三）群配对刀切法

群配对刀切法是对整群刀切法的一个改造（Allen *et al.*，2001；OECD，2009）。比如对于NAEP，群配对刀切法就是将在同一方差层的两个学校配成一对，这里要求两个学校均衡，即大小相似，影响成绩的人口统计特征也相似（Westat，2007）。当所有群都配好对，各个方差层也就形成了。在方差估计计算时，方差层被认作一般的层来处理。这种方法的重复样本就是在样本的某个对子中随机去除一个群而得到，比如第 j 个重复样本就是在第 j 对群中随机去除一个群而形成，这时在留下的一个群中，每个样本的权都要加倍。如果总共有 J 个方差层，就形成 J 套重复样本。为叙述方便，假定每个方差层有两个学校；而在实际中，会有一个方差层内有两个以上群的情况（Allen *et al.*，2001）。

在第 j 个方差层中，定义学校 k 的样本量为 n_{jk}，w_{jki} 为此层学校 k 中样本 i 的权，$N_{jk}=\sum_{i=1}^{n_{jk}} w_{jki}$ 为学校 k 中所有样本权的和，并定义 $N_j=N_{j1}+N_{j2}$。第 j 层的层权 W_j 等于比例将 N_j 除以 N，层权的和 $\sum_{j=1}^{J} W_j=1$。我们可以估计第 j 层学校 k 的平均成绩

$$\bar{x}_{jk}=\frac{\sum_{i=1}^{n_{jk}} w_{jki}x_{jki}}{N_{jk}}$$

以及第 j 层的平均成绩

$$\bar{x}_j=\frac{N_{j1}}{N_j}\bar{x}_{j1}+\frac{N_{j2}}{N_j}\bar{x}_{j2}。$$

而总体平均数估计则可表达成为

$$\bar{x}=\sum_{j=1}^{J} W_j\bar{x}_j=\sum_{j\neq l} W_j\bar{x}_j+W_l\bar{x}_l。$$

在第 l 个重复样本的第 l 层中，如果学校 2 被去掉的话，学校 1 里的样本的权都要加倍 $w_{j1i}^{*}=2w_{j1i}$。由于对样本权会作不同的调整，比如标准化，事后分层，或修整（trimming），实际的加倍因子可能会偏离 2。那么第 l 层的平均成绩为

$$\bar{x}_l^{*}=\frac{\sum_{i=1}^{n_{lk}} w_{lki}^{*}x_{lki}}{2N_{lk}}=\frac{\sum_{i=1}^{n_{lk}} w_{lki}x_{lki}}{N_{lk}}。$$

因为学校 2 被去除，所以 $\bar{x}_l^{*}=\bar{x}_{l1}$。这样在第 l 个重复样本中，$W_l^{*}=\sum_{i=1}^{n_{j1}} w_{l1i}^{*}$。在第 l 个重复样本的第 l 个伪估计平均值等于

$$\bar{x}_{(l)}=\sum_{j=1}^{J} W_j\bar{x}_j=\sum_{j\neq l} W_j\bar{x}_j+W_l^{*}\bar{x}_l^{*}。$$

因此 $\bar{x}_{(l)}$ 和 $\bar{x}$ 之间的离差等于

$$\begin{aligned}\bar{x}_{(l)}-\bar{x}&=W_l^{*}\bar{x}_l^{*}-W_l\bar{x}_l\\&=\frac{2N_{l1}}{N}\left(\frac{N_{l1}}{2N_{l1}}\bar{x}_{l1}+\frac{N_{l1}}{2N_{l1}}\bar{x}_{l1}\right)-\frac{N_l}{N}\left(\frac{N_{l1}}{N_l}\bar{x}_{l1}+\frac{N_{l2}}{N_l}\bar{x}_{l2}\right)\\&=\frac{1}{N}(N_{l1}-\bar{x}_{l1}-N_{l2}\bar{x}_{l2})。\end{aligned}\tag{3}$$

这样群配对刀切法的方差估计就是

$$v_{PGJRR}(\overline{x})=\frac{(J-1)}{J}\sum_{l=1}^{J}(\overline{x}_{(l)}-\overline{x})^2 \tag{4}$$

(Allen *et al.*, 2001)。

1. 配对的群不平衡所引起的问题

$\overline{x}_{(l)}$ 和 $\overline{x}$ 的离差，即公式(3)所示，可以包括抽样误差和偏误两重因素。如果 $N_{l1}\approx N_{l2}$，离差由两个学校平均成绩的差别决定。定义 $E(\overline{x}_{l1})=\overline{X}_{l1}$ 以及 $E(\overline{x}_{l2})=\overline{X}_{l2}$。如果 $\overline{X}_{l1}\approx\overline{X}_{l2}$，离差由两个学校的大小(样本权和 N_{l1} 与 N_{l2})之差决定。伪估计平均值 $\overline{x}_{(1)}$ 的偏误(bias)可用下式求得，因为依据(3)

$$E(\overline{x}_{(l)}-\overline{X})=\frac{1}{N}(N_{l1}\overline{X}_{l1}-N_{l2}\overline{X}_{l2})$$

假定第 l 层的总体是调和的，而 $\overline{x}_{l1}$ 和 $\overline{x}_{l2}$ 抽到的两个学校的平均估计，它们的期望 $\overline{X}_{l1}$ 和 $-_{2}$ 应该相等，这时其伪估计平均值$^{-}_{()}$是无偏的。这证明了要避免在伪估计平均值出现异常值，要求配对的两个群的样本量大小相似并且方差层中群的平均值同分布。换言之，两个群之间的差异是由抽样所造成的。这也证明了配对群的样本量和平均值的平衡，是群配对刀切法有效的一个充分条件。在实际中，配对群的不平衡与异常值会经常出现，下一节会用实际例子来说明。

2. 对群配对刀切法的一个改进

当配对的两个群很不一致的时候，我们必须认定它们实际上分属于不同的方差层，并将这个方差层拆为两个新方差层。每层包括其中一个现存群，还有一个没有参加调查的虚拟群，但其数据缺失。为了在层中形成群的配对，一个办法是将现存群随机分成两个子群，将它们配成对子。我们依照群配对刀切法的原理，去掉一个子群，将留下的子群的样本权加倍，再进行方差估计。这样的方法在其他的调查中也有应用，例如 PIRLS 就是用这种方法来处理方差层中只有一个学校的情形(Gonzalez & Kennedy, 2003)。这种改进的整群刀切法称为部分群配对刀切法(partially paired GJRR)，或非全部群配对刀切法，其中拆开的学校称为准学校(quasi-schools)。这样在新的重复样本中，准学校的成绩用来计算伪估计值，这时异常值可被消除，因而消除不平衡的群配对的影响。

在教育测量调查中，一个基本任务就是对子总体(domains)作各种估计并进行方差估计(Cochran 1977; Kish, 1980 & 1988; U. N. Statistical Office, 1950)。一个子总体的平均值，如男生的平均成绩，可以用专门定义的样本权来估计。其样本权可这样定义，一个样本如果是男生，他的权等于他原来的样本权，否则为 0。对子总体估计也可以推出和(3)相类似的关系。但是，在作子总体估计，一对学校的成绩的差别往往大于总体之估计。因为学校中的子样本大小和成绩水平，都难以受设计的控制。如对(3)的讨论，当一对学校的子样本大小或成绩水平很不一致时，用群配对刀切法就会超估方差。这种情形在子总体很小时更容易发生，例如在美国，子总体为亚裔学生或印地安学生。

3. 探测伪估计值中的异常值

因为样本的平均值服从渐近正态分布(Cochran, 1977)，伪平均估计值也服从渐近正态分布。每当有异常值存在时，伪平均估计值的直方图和正态分布曲线就不吻合；而当异常值去除时，伪平均估计值的直方图和正态分布曲线就应该吻合。伪估计值的标准离差(standard deviation, SD)是测量分布离散度的。作为经验法则，伪估计值$^{-}_{()}$离中位数$^{-}_{()}$3 倍标准离差就

可被认定为异常值。即$|\bar{o}-\bar{o}|\geqslant 3$SD。不过如果数据的分布中心对称，标准离差本身对异常值会很敏感；而四分位数间矩（inter-quartile range，IQR）对异常值，则比 SD 较不敏感，所以在分析异常值时，SD 往往用 IQR 代替。当一个伪估计平均值离中位数 3 倍 IQR 时（即$|\bar{o}-\bar{o}|\geqslant 3$IQR），就被认定为异常值（Hoaglin，Mosteller & Tukey，1983；Mosteller & Tukey，1977）。但是数据如果是偏态的（skewed），用 3IQR 来判断异常值则会发生误判，因为在这种情形下 IQR 会很小或者被低估。

三、经验数据的结果及比较

（一）数据来源

在研究中，我们用真实的州教育评估调查的样本，来表明在应用群配对刀切法时，由伪估计值中的异常值而引起异常的方差估计。各州样本在抽样时，是用二阶段整群抽样方法，从各州的抽样框中独自抽取。在第一阶段，每州抽 80 个左右的学校，然后再从抽中的学校抽 30 个左右的学生，所以一个州样本大约有 2～3 千个学生。依群配对刀切法的设计，共计有 62 个重复样本。学生成绩用五组似然值（plausible values，PV）来测估，其标度范围（scale range）为 0～300 分。这些似然值由条件推估过程（conditioning procedure）（Mislevy，1991；Rubin，1987）来推算，在估算时先产生有不同特征学生群体的成绩后验分布，然后由不同的后验分布随机产生似然值，最后再将成绩的似然值转换到报告的成绩标度范围。本研究中，由于有五组似然值，可以检验结果的可靠性。

（二）鉴别伪估计值中异常值（离群值）

表 1 中是五个州样本的每组似然值的一些基本统计量和用群配对刀切法估计的方差，并包括其伪估计值的标准差（SD）及四分位数间矩（IQR）。在第一个州样本中，运用 3IQR 法则，第 8 个和第 57 个平均数的伪估计值都鉴别为异常值。其他四个州的伪估计值中，各发现一个异常值。由于每个州样本有 62 个重复样本，每组似然值可以得到 62 个伪估计值。图 1 和图 2 分别是第三个州的第一组和第二组的似然值得到的伪估计值所形成的曲线，两图中的最小伪估计值都是异常值。

表 1　基本统计量以及伪平均成绩中的异常值

（五个州公立学校样本）

	重复样本	PV1	PV2	PV3	PV4	PV5
州 1						
州平均成绩		153.64	153.44	153.29	153.31	153.33
有异常值的州伪平均值 1	8	153.14	152.91	152.71	152.75	152.81
有异常值的州伪平均值 2	57	153.06	152.84	152.69	152.67	152.81
州伪平均值的 SD		0.208	0.202	0.207	0.192	0.202
州伪平均值的 IQR		0.158	0.174	0.149	0.137	0.155

（续表）

	重复样本	PV1	PV2	PV3	PV4	PV5
州 2						
州平均成绩		159.20	158.73	159.16	159.50	158.76
有异常值的州伪平均值	11	158.74	158.31	158.74	159.11	158.32
州伪平均值的 SD		0.120	0.126	0.121	0.120	0.125
州伪平均值的 IQR		0.148	0.141	0.130	0.137	0.136
州 3						
州平均成绩		154.74	154.68	154.49	154.84	154.98
有异常值的州伪平均值	10	153.96	153.89	153.69	154.07	154.25
州伪平均值的 SD		0.186	0.181	0.188	0.181	0.184
州伪平均值的 IQR		0.158	0.174	0.149	0.137	0.155
州 4						
州平均成绩		154.00	154.05	154.26	154.15	154.67
有异常值的州伪平均值	8	154.36	154.47	154.67	154.53	155.06
州伪平均值的 SD		0.124	0.118	0.120	0.120	0.121
州伪平均值的 IQR		0.148	0.141	0.130	0.137	0.136
州 5						
州平均成绩		136.68	136.63	136.89	136.31	136.70
有异常值的州伪平均值	24	137.34	137.35	137.64	137.04	137.43
州伪平均值的 SD		0.179	0.182	0.187	0.181	0.180
州伪平均值的 IQR		0.180	0.206	0.183	0.209	0.177

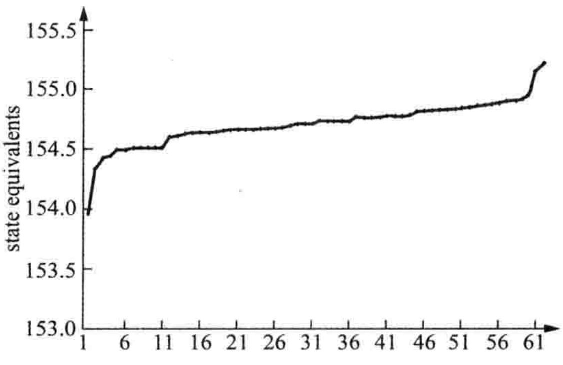

图 1 Jackknife pseudo means of PV1

(Sample of 州 3)

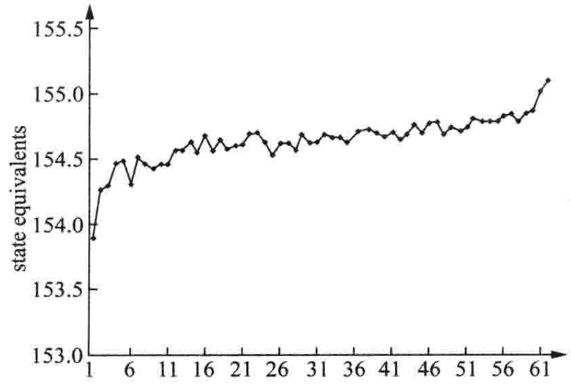

图 2 Jackknife pseudo means of PV2

(Sample of 州 3)

在图 1 中，PV1 的伪平均值依大小排序；图 2 中的伪平均值的排序依照图 1 的次序排列。

表 2 列出了在伪估计值中造成异常值的不平衡配对学校的平均成绩。这些造成异常值的配对学校,或者平均成绩有很大差异,或者样本大小有很大差别。第一个州的第 57 对学校样本大小相差很大,其余的在平均成绩上差别显著。总之,在伪估计值产生异常值是由配对学校不平衡所造成的,与第二节中的分析结论相一致。这种方法中权的加倍往往使得上述不平衡更加严重。

表 2　PGJRR 方法中不平衡配对引起伪估计值异常值的学校平均成绩

(五个州公立学校样本)

	重复样本	PV1	PV2	PV3	PV4	PV5	样本权之和①
州 1							
去除学校的平均成绩	8	114.23	116.95	108.30	118.60	115.70	192.31
权加倍学校的平均成绩	8	110.39	108.23	101.62	105.45	107.42	2553.86
去除学校的平均成绩	57	157.88	157.36	155.43	161.12	156.80	1087.23
权加倍学校的平均成绩	57	104.98	103.04	101.57	103.01	109.63	2477.18
州 2							
去除学校的平均成绩	11	178.49	171.89	175.28	174.53	173.13	361.10
权加倍学校的平均成绩	11	122.57	122.57	123.69	127.33	120.66	998.33
州 3							
去除学校的平均成绩	10	161.09	158.05	162.14	161.49	159.04	1026.36
权加倍学校的平均成绩	10	89.68	86.27	89.46	90.52	92.01	2237.32
州 4							
去除学校的平均成绩	8	142.32	136.30	140.80	141.79	140.24	416.81
权加倍学校的平均成绩	8	176.49	175.78	180.34	178.11	177.19	991.97
州 5							
去除学校的平均成绩	24	127.50	124.54	124.91	124.93	125.03	14830.30
权加倍学校的平均成绩	24	157.03	154.84	157.19	157.07	156.38	9863.04

表 3 中的两组方差估计值,V_{PGJRR} 和 V_{PPGJRR} 分别由群配对刀切法和部分群配对刀切法推算。很明显,由改进的方法推算的 V_{PPGJRR} 值全部小于 V_{PGJRR} 的值。原来伪估计值中的异常值超过了 3IQR,正如表 2 中所示这是由配对学校不平衡所致,并非由于抽样误差造成的,因而 V_{PGJRR} 的值是超估的。在部分群配对刀切法中,配对的学校很不平衡时,认定这两个学校分属两个不同的方差层。在这种层中将一个学校随机分成配对的两个准学校,再依照群配对刀切法估计方差。据表 3,V_{PGJRR} 的值要比 V_{PPGJRR} 值大 13%到 35%不等。即这种部分群配对刀切法估计的方差比原来估计的方差要小。例如,根据第一个州样本的第一组成绩似然值,其 V_{PGJRR} 的值等于 1.89,而 V_{PPGJRR} 的值为 1.61。

① 在计算学校样本权之和时,各州样本的总样本权之和都规范化为州的总学生数之和。

表 3 州平均成绩的两刀切法方差估计

（五个州公立学校样本）

	PV1	PV2	PV3	PV4	PV5
			州 1		
V_{PGJRR}	1.89	1.87	1.75	1.97	1.92
V_{PPGJRR}	1.61	1.57	1.47	1.63	1.69
V_{PGJRR} 与 V_{PPGJRR} 比率	1.17	1.19	1.19	1.21	1.13
			州 2		
V_{PGJRR}	0.86	0.95	0.88	0.87	0.94
V_{PPGJRR}	0.67	0.79	0.71	0.73	0.77
V_{PGJRR} 与 V_{PPGJRR} 比率	1.29	1.20	1.24	1.20	1.23
			州 3		
V_{PGJRR}	2.07	1.96	2.12	1.98	2.03
V_{PPGJRR}	1.57	1.45	1.60	1.49	1.59
V_{PGJRR} 与 V_{PPGJRR} 比率	1.31	1.35	1.33	1.33	1.28
			州 4		
V_{PGJRR}	0.92	0.84	0.87	0.87	0.87
V_{PPGJRR}	0.81	0.69	0.71	0.73	0.74
V_{PGJRR} 与 V_{PPGJRR} 比率	1.13	1.22	1.22	1.18	1.17
			州 5		
V_{PGJRR}	1.92	1.99	2.10	1.96	1.95
V_{PPGJRR}	1.60	1.65	1.69	1.53	1.51
V_{PGJRR} 与 V_{PPGJRR} 比率	1.20	1.21	1.24	1.28	1.29

四、结　论

在应用刀切法时，两个配对的群，如果在成绩上或样本大小上很不平衡，伪估计值中则会出现异常值，这时群配对刀切法会超估方差。本文分析了群配对刀切法重复样本的结构，用公式表示了不平衡配对对伪估计值的影响。对经验数据的评估也证实了分析的结论。

在设计评估的样本时，很难做到让配对的群在样本大小与测量值在分布上适当平衡，总会有配对的群不平衡，这往往会导致伪估计值中出现异常值。在实际进行评估调查时，我们经常发现学校有自己的特征，互相之间很不相同。同一学校或同一小区的学生的人口特征往往相近，成绩之间相关度很高（Johnson & Rust，1992；Nohara，2001；Neidorf *et al.*，2006；Turner & Adams，2007）。由群配对刀切法和部分群配对刀切法分别得到的两组结果，可以有 13％到 35％的差别，这些是不平衡的配对对群配对刀切法方差估计影响的证据。