人口普查质量评估的三系统模型*

胡桂华

【摘　要】迄今为止，所有国家和地区在人口普查后组织的用于评估人口普查质量的事后调查均使用双系统估计量估计实际人口数和人口普查覆盖误差。由于人口普查与事后调查不独立而引起的交互作用偏差使双系统估计量低估或高估人口数。文章依据对数线性模型和最大似然估计构造的三系统模型能显著减少交互作用偏差，同时通过一个实证案例演示了三系统模型的选择和实际人口数估计及其方差估计的过程。实证结果表明，人口普查与行政记录独立但均与事后调查相关的三系统模型最适合于既定的数据描述。

【关键词】对数线性模型　三系统模型　交互作用偏差　渐近方差估计

【作　者】胡桂华　重庆工商大学数学与统计学院，教授。

一、引　言

目前世界上所有进行人口普查的国家和地区对其质量评估工作的一个主要内容是计算人口普查覆盖误差，即目标人口总体实际人口数与普查登记人口数之差。目标总体的实际人口数是无法确切知道的，因此在进行人口普查质量评估时，要用这样或那样的方法计算目标总体实际人口数的一个估计值，把这个估计值与人口普查登记人口数之差当做人口普查覆盖误差。获得目标总体实际人口数估计值的方法，一种是用人口行政记录资料来推算；另一种是抽取一个以地理小区为单位的样本，对样本中的每个小区追溯登记本小区在普查日的人口，以此构造目标总体人口数的估计量。在后一种方法中，如果在构造估计量时把样本各个小区人口普查登记资料的信息也包括进来，无疑会大大提高估计的精度。在此种思路下，一种常用的估计量是“双系统估计量”，即通过移植有名的捕获—再捕获模型（胡桂华，2010）构造而成。目前，用双系统估计量（胡桂华，2011）估计实际人口数并据此计算人口普查覆盖误差被认为是人口普查质量评估领域较前沿的方法。这里的双系统是指人口普查及其在人口普查登记工作结束不久采取抽样方法进行的事后调查（贺本岚等，2010）。

然而，双系统估计量有一个不可忽视的缺陷，那就是由于人口普查与事后调查不独立

* 本文为全国统计科学研究计划重点项目“多系统模型在人口普查质量评估中的应用”（批准号：2012LZ044）的阶段性成果。

而引起的交互作用偏差(Wachter 等,2000)。交互作用偏差中的"交互作用"是指总体中的某些人倾向于被普查和事后调查遗漏,而"偏差"是指一些人口被双重遗漏,也就是既未包括在普查人口数中,也未包括在双系统估计量估计的人口数中。导致交互作用偏差的原因很多。对难以计数群体(无家可归人口、伤残人口、吸毒人口、精神病患者、流动人口等),双系统估计量的交互作用偏差更大。

有两种测算交互作用偏差的方法(Scott,2012)。一是在假设女性不存在交互作用偏差的情况下,利用人口平衡方程估计的全国性别比率(男性数量 / 女性数量)和使用双系统估计量估计的女性数量的乘积作为估计的全国男性人口数,并减去使用双系统估计量估计的男性数量来测算全国男性的交互作用偏差,即被双系统估计量低估的男性人口数。二是使用关系式"估计的事后调查人口数 - 估计的因数据处理误差而引起的正偏差人口数 + 因交互作用偏差引起的负偏差人口数 = 使用人口平衡方程估计的人口数"来测算全国人口的交互作用偏差。第一种测算方法的精度取决于女性不存在交互作用偏差的假设条件与现实情况符合的程度。第二种测算方法的精度受人口平衡方程估计的人口数,以及估计的事后调查人口数和估计的数据处理误差人口数质量的影响。

交互作用偏差有其自身形成的原因(Bell,2001)。交互作用偏差源于统计相依性,即如果被普查登记就会有更大或更小的机会被事后调查登记。这里的统计相依性有两种可能的情况:一种情况是,被普查登记的人口很可能充分意识到了普查的重要性,因而比起那些被普查遗漏的人更可能参与事后调查,从而导致两个系统匹配人数(相对于随机情形来说)倾向于偏大,由于匹配人数是双系统估计量的分母,因而造成双系统估计量低估实际人口数。另外一种情况是,被普查登记的人,认为已经回答了普查问题,因而比起那些没有被普查登记的人更不愿意参加事后调查,结果导致两个系统匹配人数(相对于随机情形来说)倾向于偏小,同理造成双系统估计量高估实际人口数。可见,统计相依性会导致两种不同方向上的偏差。当总体中不存在异质性时(由于年龄、性别、是否拥有属于自己的房屋、所处地理位置等因素而导致的总体中的一些人口比另外一些人口被人口普查或事后调查登记的概率不同),人们被登记的概率相同,相应发生两种不同方向的统计相依性的概率也会大致相同(因为不会出现登记概率大或小的那些人聚集在前面所说的两种情况中,于是统计相依性所导致的两种不同方向上的偏差会倾向于相互抵消。然而,当总体中存在异质性的时候,情况就不同了。这时,人们被登记的概率不同,相应的,统计相依性所导致的发生在一种方向上的偏差的概率会系统地偏大(或偏小),两种方向上的偏差不会倾向于相互抵消。所以,异质性加剧了双系统估计量中的交互作用偏差效应。

上面的分析告诉我们,设法消除总体内的异质性是降低双系统估计量中的交互作用偏差效应的一种可供考虑的途径。目前在各国实践中都是采用抽样后分层技术来实现总体内同质性的目标(Hogan,1992)。但是,由于在此技术中不得不舍弃一些分层标志,因此很难做到同一层内的单位完全同质。罗吉斯蒂回归模型技术容许把所需要的分层标志尽可能地纳

入模型，因而在实现总体内同质的目标这方面的效果更好一些。美国普查局在 2010 年人口普查质量评估中试用了这种技术（Bell 等，2008）。遗憾的是，由于该技术具体操作比较复杂，所以美国普查局不打算在 2020 年人口普查质量评估中继续采用这种方法。这就意味着，通过追求总体内同质的途径来降低双系统估计量中的交互作用偏差效应没有多大的努力空间。于是，人们把注意力更多地放在设法通过资料校正的途径来降低交互作用偏差。一是从人口行政记录系统中获取事后调查缺失的人口名单及其人口统计特征，并补充到事后调查中，以减少同时被人口普查与事后调查遗漏的人口（Zaslavsky等，1993）；二是依据人口行政记录系统、人口普查及其事后调查构造三系统模型，也称为三系统估计量（Zaslavsky等，1993）。

相比双系统估计量，三系统估计量有 4 个方面的优势：一是覆盖范围广，能覆盖被人口普查和事后调查同时遗漏的人口（在逃犯人等）；二是能够包括大部分同时被人口行政记录系统、人口普查和事后调查这 3 个系统遗漏的人口；三是允许在各资料系统之间存在统计相依关系的情形下构造实际人口数估计量；四是能以较高精度测算双系统估计量交互作用偏差的幅度。

由于三系统估计量的这些优势，美国普查局可能在 2020 年人口普查质量评估中首次使用三系统估计量替代双系统估计量（包括罗吉斯蒂回归模型双系统估计量）估计实际人口数，并用以计算人口普查覆盖误差。美国普查局此前未使用三系统估计量的原因有两个：一是三系统估计量也不能完全消除交互作用偏差；二是构造覆盖范围广泛又不重复的行政记录系统有一定的难度。然而，经过多年努力，这两个问题已经基本得到解决。

美国普查局的 Richard 在 2012 年的一次国际统计学研讨会上提交了一篇“利用行政记录构造三系统模型估计 2020 年人口普查覆盖误差”的学术论文。这篇论文可归纳为 3 个部分。一是对双系统估计量高估或低估实际人口数原因的简单分析，即异质性和统计相关性引起的个人被登记概率不同而使双系统估计量估计的人口数偏离实际人口数，对难以计数人口，这种偏差更大。二是根据 3 个系统之间的相关或独立性关系及利用对数线性模型构造了未被任何一个系统登记的人口数估计量（共 8 个）。三是模拟分析（个人被 3 个系统登记的概率和总体规模估计量）及其对模拟结果的分析。

然而，Richard 的这篇论文存在一些明显的缺陷：（1）它使用的是模拟数据资料。很显然，如果模拟数据与实际数据存在差异，那么模拟分析得出的结论与实证分析得出的结论就会有明显的差距，从而难以发挥对实际工作的指导作用，有时甚至起误导作用。事实上，模拟数据总是或多或少地偏离实际数据。（2）只简单地列出了不同假设条件下的三系统模型。对每个假设条件的含义、提出的背景，以及违背该假设条件的后果分析甚少。这不利于学者和实际工作者理解、掌握和运用三系统模型。（3）没有给出总体规模估计量的方差计算公式。因为对任何一个估计量必须给出其方差估计量，否则这个估计量是不可用的。这是由于方差估计量是度量总体参数估计量对总体参数估计精度的重要指标。对数据使用者来

说，他们不只是需要数据，还想知道数据的精度或准确度。(4)只给出了三维列联表缺失组格(未被 3 个系统中任何 1 个系统登记的人口数)期望值的估计量。由于三系统模型是依据 3 个系统构造的，因此不同的系统组合可以构造出不同的三系统模型。换句话说，三系统模型的形式不是唯一的。这就存在需要从若干个三系统模型中选择一个最优的三系统模型的问题。选择三系统模型需要进行模型假设检验，而这需要知道每个组格人口数目期望值的估计量。(5)只简单分析了双系统估计量估计人口数目精度不高的因素(异质性和统计相依性)，但对这两个因素如何影响双系统估计量却并未做任何分析。

针对上述缺陷，本文将在以下几个方面进行改进。一是深入分析统计相依性和异质性影响双系统估计量的形式或方式及其程度。二是指出不完备列联表、对数线性模型和最大似然估计是构建三系统模型的主要统计工具。三是交代清楚每个三系统模型的统计意义及其对应的对数线性模型。四是不仅给出缺失组格人口数的期望值估计量，还给出其他组格的期望值估计量，为三系统模型假设检验提供基本数据。五是给出三系统模型总体规模估计量的方差估计量的计算公式，并使用实证资料进行三系统模型及其方差估计值的计算与分析。本文结合使用对数线性模型和最大似然估计构造不完备的三维列联表各个组格及总体的人口数目估计量(Fienberg，1972)。

二、三系统模型的基本理论

三系统模型(又称三系统估计量)指的是由人口普查、事后调查和行政记录这 3 个系统构造的用来估计总体实际人口数和未被任何一个系统登记的人口数的模型。根据估计的总体实际人口数计算人口普查覆盖误差。为了便于阐述基本理论，本研究假定人口普查、事后调查和行政记录这 3 个资料系统的人口名单是对同一个人口总体在同一时点上的状态进行全面登记的结果。另外，假定在独立进行人口普查或事后调查或行政记录登记的时候，人口总体中所有的人具有相等的、不为 0 的被登记概率。

本研究把总体中的个人分别按其在这 3 个系统中的登记情况进行分组：(1)是否被人口普查系统登记。是(i=1)，否(i=2)。(2)是否被事后调查系统登记。是(j=1)，否(j=2)。(3)是否被行政记录系统登记。是(k=1)，否(k=2)。将 3 个分组进行交叉组合，形成 8 个复合组。它们分别是(122)；(212)；(221)；(211)；(121)；(111)；(112)；(222)。相应的用 $n_{ijk}(i,j,k=1,2)$ 表示各个组的实际观察人口数，即 n_{122}，n_{212}，n_{221}，n_{211}，n_{121}，n_{111}，n_{112}，n_{222}。其中，n_{122}= 只被人口普查登记的人口数，n_{211}= 同时被事后调查和行政记录登记但未被人口普查登记的人口数，n_{222}= 未被任何系统登记的人口数(未知)，等等。

总体中的人口包括两个部分：一是被至少一个系统观察到的人口；二是未被任何系统观察到的人口。我们用 n 来表示研究总体中至少被一个系统所观察到的人口数，即最低人口数。很显然，$n=n_{122}+n_{212}+n_{221}+n_{211}+n_{121}+n_{111}+n_{112}$，总体实际人口数 $N=n+n_{222}$。用 μ_{ijk} 表示三维列联表第(ijk)组格人口数的期望值，μ_{222} 为没有被任何系统观察到的人口数目期望值，称为缺失

组格人口数的期望值。用$\hat{\mu}_{222}$表示n_{222}的期望值μ_{222}的估计量。不难看出，总体实际人口数估计量$\hat{N}=n+\hat{\mu}_{222}$。把上述有关数据列在表中，即三维列联表（见表1）。如果表中组格n_{222}数据缺失，那么此表为不完备的三维列联表。

表1　不完备的三维列联表

	在行政记录中		不在行政记录中	
	在事后调查中	不在事后调查中	在事后调查中	不在事后调查中
在人口普查中	n_{111}	n_{121}	n_{112}	n_{122}
不在人口普查中	n_{211}	n_{221}	n_{212}	n_{222}

（一）对数线性模型

使用三系统估计量替代双系统估计量带来的一个最大好处是，由于3个系统比2个系统覆盖人口总体更多的人口（尤其是容易被普查和事后调查同时遗漏的人口。如在逃犯人、不相信政府的人等），因而估计的人口数更接近实际人口数。然而，3个系统之间仍然可能存在交互作用和登记概率的异质性问题。对数线性模型能放宽统计相关性和异质性条件的限制，对多系统模型尤其如此。因此，使用对数线性模型构造三系统估计量。

对数线性模型是处理多维列联表的重要工具。现在所面对的是n_{222}未知的三维列联表。三系统估计量（作为其来源的多样本捕获—再捕获模型）所要应用的是针对不完备列联表的对数线性模型理论。Fienberg（1972）指出，对多个系统的捕获—再捕获实验，可以使用标准层次的模型技术选择适宜的对数线性模型。为了与后面使用的符号一致，这里约定对数线性模型的3个系统为人口普查（C），事后调查（S）和行政记录（A）。在完备表的情形下，饱和对数线性模型为：

$$\log\mu_{ijk}=\lambda+\lambda_i^{(C)}+\lambda_j^{(S)}+\lambda_k^{(A)}+\lambda_{ij}^{(CS)}+\lambda_{ik}^{(CA)}+\lambda_{jk}^{(SA)}+\lambda_{ijk}^{(CSA)} \quad (1)$$

式（1）中，λ是各个$\log\mu_{ijk}$（$i=1,2;j=1,2;k=1,2$）的基础水平，$\lambda_i^{(C)}$，$\lambda_j^{(S)}$，$\lambda_k^{(A)}$分别是人口普查、事后调查和行政记录的主效应项，$\lambda_{ij}^{(CS)}$，$\lambda_{ik}^{(CA)}$，$\lambda_{jk}^{(SA)}$分别是人口普查和事后调查、人口普查和行政记录及事后调查和行政记录的二维交互作用项，$\lambda_{ijk}^{(CSA)}$是人口普查、事后调查和行政记录的三维交互作用项。为了解释饱和对数线性模型参数的统计意义，需要增加限制条件：$\sum_i\sum_j\sum_k\lambda_{ijk}^{(CSA)}=0$，$\sum_i\lambda_i^{(C)}=\sum_j\lambda_j^{(S)}=\sum_k\lambda_k^{(A)}=0$，$\sum_i\sum_j\lambda_{ij}^{(CS)}=\sum_i\sum_k\lambda_{ik}^{(CA)}=\sum_j\sum_k\lambda_{jk}^{(SA)}=0$。

对表1给出的不完备三维列联表，还需要增加另外的限制条件，即三维交互作用项设定为零。这样，针对3个系统的捕获—再捕获实验数据不完备表的对数线性模型就变为：

$$\log\mu_{ijk}=\lambda+\lambda_i^{(C)}+\lambda_j^{(S)}+\lambda_k^{(A)}+\lambda_{ij}^{(CS)}+\lambda_{ik}^{(CA)}+\lambda_{jk}^{(SA)} \quad (2)$$

当式（2）中某一个或某几个交互影响项等于0时所形成的模型叫做简约模型，这样就形成了不完备表情形下的对数线性模型体系。在这个体系中一共有8个模型。对数线性模型1：在式（2）中，$\lambda_{ij}^{(CS)}=0$，$\lambda_{ik}^{(CA)}=0$，$\lambda_{jk}^{(SA)}=0$，其统计意义是，人口普查、事后调查和行政记录两两相互独立；对数线性模型2：在式（2）中，$\lambda_{ik}^{(CA)}=0$，$\lambda_{jk}^{(SA)}=0$，其统计意义是，人口普查与事后调查相关但均独立于行政记录；对数线性模型3：在式（2）中，$\lambda_{ij}^{(CS)}=0$，$\lambda_{jk}^{(SA)}=0$，其统计意义是，人口普查与行政记录相关但均独立于事后调查；对数线性模型4：在式（2）中，$\lambda_{ij}^{(CS)}=0$，$\lambda_{ik}^{(CA)}=0$，

其统计意义是，事后调查与行政记录相关但均独立于人口普查；对数线性模型5：在式（2）中，$\lambda_{ik}^{(CA)}=0$，其统计意义是，人口普查与行政记录独立但均与事后调查相关；对数线性模型6：在式（2）中，$\lambda_{ij}^{(CS)}=0$，其统计意义是，人口普查与事后调查独立但均与行政记录相关；对数线性模型7：在式（2）中，$\lambda_{jk}^{(SA)}=0$，其统计意义是，事后调查与行政记录独立但均与人口普查相关；对数线性模型8，在式（2）中，没有为0的假设项，其统计意义是人口普查、事后调查和行政记录两两相关。

（二）最大似然估计量

三系统估计量的目标是，使用最大似然估计实现对μ_{222}从而对N的估计。为完成这个任务，首先要弄清楚，来自现实世界的数据适宜用哪一个对数线性模型来描述。毫无疑问的是，数据一定能够用饱和对数线性模型8来描述。不过，此时我们还关心的是，数据是否同时还能够用其他更简约的模型来描述。若果真如此，我们将会采用简约的对数线性模型下的最大似然估计来构造μ_{222}和N的估计量，这会提高估计量的有效性，而且方差的计算也会相应简化。为此，在构造μ_{222}和N的估计量的任务之前提出了一个先行的任务，即在8个对数线性模型中进行选择，找出最适宜描述数据的模型。

模型检验统计量要用到μ_{ijk}的最大似然估计量$\hat{\mu}_{ijk}$。于是，首先要分别讨论1～8每一个对数线性模型下各个组格最大似然估计量$\hat{\mu}_{ijk}$的构造方法。对于每一个对数线性模型，除了给出最大似然估计量$\hat{\mu}_{ijk}$的构造方法，还要给出假若选定该对数线性模型是最适宜的模型，那么如何在该对数线性模型下构造μ_{ijk}和N的最大似然估计量。这里需要特别强调的是，使用最大似然估计要求对数线性模型有较少的参数，否则为估计这些参数而需要收集更多的数据，从而增加每个参数估计的难度。

（三）组格估计量$\hat{\mu}_{ijk}$、缺失组格估计量$\hat{\mu}_{222}$及总体实际人口数估计量$\hat{N}$

1. 对数线性模型1。人口普查、事后调查和行政记录两两相互独立，其对数线性模型表达式为：

$$\log\mu_{ijk}=\lambda+\lambda_i^{(C)}+\lambda_j^{(S)}+\lambda_k^{(A)} \qquad (3)$$

在对数线性模型1下，无法直接得到最大似然估计量$\hat{\mu}_{ijk}$的代数解析式，此时的最大似然估计量$\hat{\mu}_{ijk}$只能通过迭代算法得到。结合实际来考虑，人口普查、事后调查和行政记录相互独立在实际中几乎是难以实现的，换句话说，模型1难以较好地拟合数据。因此，本文把对数线性模型1下的最大似然估计量$\hat{\mu}_{ijk}$、μ_{222}和N的估计量略去不论。

2. 对数线性模型2～4。这3个模型分别意味着，人口普查与事后调查相关但均与行政记录独立；人口普查与行政记录相关但均独立于事后调查；事后调查和行政记录相关但均独立于人口普查。我们以对数线性模型2为例，其对应的对数线性模型为：

$$\log\mu_{ijk}=\lambda+\lambda_i^{(C)}+\lambda_j^{(S)}+\lambda_k^{(A)}+\lambda_{ij}^{(CS)} \qquad (4)$$

在该模型下，组格期望最大似然估计量$\hat{\mu}_{ijk}$可以直接得到。它们是$\hat{\mu}_{221}=n_{221}$；当组格$(ijk)\in$

$\{111,121,211\}$时，$\hat{\mu}_{ijk}=\{(n_{ij+})(n_{111}+n_{121}+n_{211})/(n-n_{221})\}$；当组格$(ijk)\in\{112,122,212\}$时，$\hat{\mu}_{ijk}=\{(n_{ij+})(n_{112}+n_{122}+n_{212})/(n-n_{221})\}$。这6个组格人口数的期望估计量$\hat{\mu}_{ijk}$具体写出为：

$$\hat{\mu}_{111}=\frac{(n_{111}+n_{112})(n_{111}+n_{121}+n_{211})}{n-n_{221}},\hat{\mu}_{121}=\frac{(n_{121}+n_{122})(n_{111}+n_{121}+n_{211})}{n-n_{221}}$$

$$\hat{\mu}_{211}=\frac{(n_{211}+n_{212})(n_{111}+n_{121}+n_{211})}{n-n_{221}},\hat{\mu}_{112}=\frac{(n_{111}+n_{112})(n_{112}+n_{122}+n_{212})}{n-n_{221}}$$

$$\hat{\mu}_{122}=\frac{(n_{121}+n_{122})(n_{112}+n_{122}+n_{212})}{n-n_{221}},\hat{\mu}_{212}=\frac{(n_{211}+n_{212})(n_{112}+n_{122}+n_{212})}{n-n_{221}}$$

而μ_{222}的最大似然估计量$\hat{\mu}_{222}$为：$\hat{\mu}_{222}=\frac{n_{112}+n_{122}+n_{212}}{n_{111}+n_{121}+n_{211}}\times n_{221}$。总体实际人口数估计量$\hat{N}=n+\hat{\mu}_{222}$，其方差估计量（Bernard，2009）为：

$$\hat{Var}(\hat{N})=(\hat{\mu}_{222})^2\left(\frac{1}{n_{111}+n_{211}+n_{121}}+\frac{1}{\hat{\mu}_{112}+\hat{\mu}_{122}+\hat{\mu}_{212}}+\frac{1}{n_{221}}+\frac{1}{\hat{\mu}_{222}}\right)$$

3. 对数线性模型5～7。这3个模型分别意味着人口普查、行政记录与事后调查相关；人口普查和事后调查独立但均与行政记录相关；行政记录和事后调查独立但均与人口普查相关。下面以对数线性模型5为例，其对应的对数线性模型为：

$$\log\mu_{ijk}=\lambda+\lambda_i^{(C)}+\lambda_j^{(S)}+\lambda_k^{(A)}+\lambda_{ij}^{(CS)}+\lambda_{jk}^{(AS)} \quad (5)$$

其组格期望估计量$\hat{\mu}_{ijk}$能直接得到，即：

$$\hat{\mu}_{122}=n_{122},\hat{\mu}_{221}=n_{221},\hat{\mu}_{121}=n_{121}$$

$$\hat{\mu}_{111}=\frac{(n_{111}+n_{112})(n_{111}+n_{211})}{n_{111}+n_{212}+n_{112}+n_{211}},\hat{\mu}_{211}=\frac{(n_{211}+n_{212})(n_{111}+n_{211})}{n_{111}+n_{212}+n_{112}+n_{211}}$$

$$\hat{\mu}_{112}=\frac{(n_{111}+n_{112})(n_{112}+n_{212})}{n_{111}+n_{212}+n_{112}+n_{211}},\hat{\mu}_{212}=\frac{(n_{211}+n_{212})(n_{112}+n_{212})}{n_{111}+n_{212}+n_{112}+n_{211}}$$

而μ_{222}的期望值的最大似然估计量则为$\hat{\mu}_{222}=(n_{221}n_{122})/n_{121}$，总体实际人口数估计量$\hat{N}=n+\hat{\mu}_{222}$，其方差估计量为：$\hat{Var}(\hat{N})=(\hat{\mu}_{222})^2\left(\frac{1}{n_{121}}+\frac{1}{n_{221}}+\frac{1}{n_{122}}+\frac{n_{121}}{n_{221}n_{212}}\right)$。

4. 对数线性模型8。人口普查、事后调查和行政记录两两相关，其对数线性模型表达式为：

$$\log\mu_{ijk}=\lambda+\lambda_i^{(C)}+\lambda_j^{(S)}+\lambda_k^{(A)}+\lambda_{ij}^{(CS)}+\lambda_{jk}^{(AS)}+\lambda_{ik}^{(CA)} \quad (6)$$

在该模型下，各个组格的$\hat{\mu}_{ijk}$等于该组格的实际频数观察值。而μ_{222}的期望值的最大似然估计量$\hat{\mu}_{222}$为：$\hat{\mu}_{222}=\frac{\hat{\mu}_{111}\hat{\mu}_{221}\hat{\mu}_{122}\hat{\mu}_{212}}{\hat{\mu}_{121}\hat{\mu}_{211}\hat{\mu}_{112}}=\frac{n_{111}n_{221}n_{122}n_{212}}{n_{121}n_{211}n_{112}}$。总体实际人口数估计量为$\hat{N}=n+\hat{\mu}_{222}$，其方差估计量为：

$$\hat{Var}(\hat{N})=(\hat{\mu}_{222})^2\left(\frac{1}{n_{111}}+\frac{1}{n_{112}}+\frac{1}{n_{121}}+\frac{1}{n_{122}}+\frac{1}{n_{221}}+\frac{1}{n_{212}}+\frac{1}{n_{211}}+\frac{1}{\hat{\mu}_{222}}\right)$$

（四）模型选择

为进行对数线性模型选择，需要进行下列 7 个统计检验。检验所针对的原假设分别是：模型 l 与模型 8 没有差异。其中，l=1，2，3，4，5，6，7。由于在前面已经认定，在实际问题中，模型 1 是不可能很好地拟合数据的，所以我们把第一个检验略去不做。

用于拟合优度检验的统计量是对数似然比率估计量 G^2。其公式为：

$$G^2=2\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}(n_{ijk})\log\left(\frac{n_{ijk}}{\hat{\mu}_{ijk}}\right) \qquad (7)$$

其中，$(ijk)\neq(222)$。这个统计量服从自由度为 q 的卡方分布。q 值分别为：模型 1，q=3；模型 2～4，q=2；模型 5～7，q=1；模型 8，q=0。如果在上述 7 个统计检验中，同时有两个检验的原假设未被拒绝，这意味着同时有两个简约模型被选中。例如，现在同时有模型 M_1 和 M_2 被选中，即：M_1：$\log\mu_{ijk}=\lambda+\lambda_i^{(C)}+\lambda_j^{(S)}+\lambda_k^{(A)}+\lambda_{ij}^{(CS)}+\lambda_{jk}^{(SA)}$；$M_2$：$\log\mu_{ijk}=\lambda+\lambda_i^{(C)}+\lambda_j^{(S)}+\lambda_k^{(A)}+\lambda_{ij}^{(CS)}$。

现在需要进一步检验它们当中的哪一个能够更好地描述数据。检验统计量是：

$$\Delta G^2=G_2^2-G_1^2 \qquad (8)$$

其中，G_2^2，G_1^2 分别是模型 M_2 与模型 M_1 的似然比检验统计量。ΔG^2 近似服从皮尔逊卡方分布，其自由度为对数线性模型 M_1 与对数线性模型 M_2 的自由度之差。

三、实证分析

为证实双系统估计量严重低估黑人成年男性数目，美国学者曾经对密苏里州圣路易斯地区 70 个调查小区进行过调查。调查对象是该区域内全体黑人成年男性。对调查对象登记了 3 份人口名单：1988 年人口普查预演调查名单、1988 年人口普查预演调查的事后调查名单、人口行政记录名单。考虑到年龄和是否拥有住房这两个标志的不同取值会导致被调查者登记概率不同，为了满足在登记概率相同的总体内构造三系统估计量的要求，将调查对象按年龄和是否拥有住房两个标志进行了交叉分层。前文中所提到的所有诸如构造组格期望频数的最大似然估计量、选择最适宜拟合数据的对数线性模型，以及估计人口总体实际人口数等工作，在该实证分析中都是分别在各个交叉层内进行的。交叉层的划分是：20～29 岁有房者；20～29 岁租房者；30～44 岁有房者；30～44 岁租房者。各个组格的观察人口数在 4 个事后层的分布如表 2 所示。

表 2　密苏里州圣路易斯 1988 年试点人口普查的 3 个数据来源在各个事后层的分布　　人

事后层	n_{222}	n_{221}	n_{212}	n_{211}	n_{122}	n_{121}	n_{112}	n_{111}	n
20～29 岁有房者	–	59	08	19	31	19	13	79	228
20～29 岁租房者	–	43	34	11	41	12	69	58	268
30～44 岁有房者	–	35	10	10	62	13	36	91	257
30～44 岁租房者	–	43	24	13	32	07	69	72	260

资料来源：Zaslavsky 等，1993。

本文选择此实证数据主要基于 3 个方面的考虑：一是中国国家统计局和其他国家的国家统计局迄今为止尚

未在人口普查质量评估中使用三系统模型，因而不存在这方面的实证数据；二是中国统计法规定，任何个人无权出于任何目的进行入户调查，因而本文作者没有资格自行组织调查获得三系统模型所需要的家庭或个人资料（具有隐私性和保密性）；三是美国学者 Glenn 在美国普查局工作，并且是在普查局的授权下进行了 3 个系统数据资料的采集或获取工作，其资料的准确性应该是有保障的。

利用前面的有关公式和表 2 中的数据，可以估计未被任何系统登记的人口数、总体实际人口数，以及进行对数线性模型拟合优度检验。有关计算结果如表 3、表 4 和表 5 所示。

表 3　不同模型下缺失组格数目（$\hat{\mu}_{222}$）和似然比（G^2）的估计值

模型	20～29 岁				30～44 岁				自由度
	有房者		租房者		有房者		租房者		
	$\hat{\mu}_{222}$	G^2	$\hat{\mu}_{222}$	G^2	$\hat{\mu}_{222}$	G^2	$\hat{\mu}_{222}$	G^2	
模型 2	26.22	34.46	76.43	12.19	33.16	59.27	58.42	15.71	2
模型 3	7.86	68.55	23.65	52.80	8.03	84.54	12.78	70.73	2
模型 4	24.02	59.01	25.96	54.23	24.35	62.54	17.30	76.06	2
模型 5	96.22	3.15	146.78	6.53	166.77	3.55	196.23	3.04	1
模型 6	19.07	58.71	20.20	51.58	17.20	61.25	11.13	69.99	1
模型 7	24.84	34.44	132.79	8.78	34.99	59.25	79.34	14.73	1
模型 8	245.11	0	379.69	0	418.83	0	378.68	0	0

从表 3 可以看出：（1）20～29 岁年龄组租房者比有房者缺失组格的人口数都多。这是因为，这个年龄组的无固定住所的黑人男性流动性很大，调查员很难找到他们，有时即便找到，他们也可能拒绝合作。另外，当地行政管理机构对该年龄组的租房黑人男性的登记工作重视不够，认为他们不久会离开其管辖区域。从 30～44 岁年龄组来看，在 7 个模型中，有 3 个模型（模型 4、模型 6 和模型 8）租房者的缺失组格人口数少于有房者的缺失组格人口数，而另外 4 个模型（模型 2、模型 3、模型 5 和模型 7）租房者的缺失组格人口数多于有房者的缺失组格人口数。这表明，这个年龄组的有房者和租房者的流动性和调查登记的难度大致差不多。（2）不同对数线性模型估计的缺失组格人口数差别大。从 20～29 岁有房者来看，估计的缺失组格人口数最多的是模型 8 的 245.11 人，其次是模型 5 的 96.22人，最少的是模型 3 的

表 4　使用似然比（G^2）值进行的模型拟合优度检验

模型	20～29 岁				30～44 岁			
	有房者		租房者		有房者		租房者	
	G^2	p 值	G^2	p 值	G^2	p 值	G^2	p 值
模型 2	34.46	<0.01	12.19	<0.01	59.27	<0.01	15.71	<0.01
模型 3	68.55	<0.01	52.80	<0.01	84.54	<0.01	70.73	<0.01
模型 4	59.01	<0.01	54.23	<0.01	62.54	<0.01	76.06	<0.01
模型 5	3.15	>0.01	6.53	>0.01	3.55	>0.01	3.04	>0.01
模型 6	58.71	<0.01	51.58	<0.01	61.25	<0.01	69.99	<0.01
模型 7	34.44	<0.01	8.78	<0.01	59.25	<0.01	14.73	<0.01
模型 8	0	>0.01	0	>0.01	0	>0.01	0	>0.01

表 5 总体规模$\hat{N}$及其渐近标准误差和置信区间的估计值 人

	$\hat{N}$	渐近标准误差	95%置信区间	
			下限	上限
20～29 岁				
有房者	324.22	32.25	261.01	387.43
租房者	414.78	54.29	308.37	521.19
30～44 岁				
有房者	423.77	59.63	306.90	540.64
租房者	456.23	88.45	282.87	629.59

7.86 人。从 20～29 岁租房者来看，估计的缺失组格人口数最多的是模型 8 的 379.69 人，最少的是模型 6 的 20.2 人。从 30～44 岁有房者来看，估计的缺失组格人口数最多的是模型 8 的 418.83 人，最少的是模型 3 的 8.03 人。从 30～44 岁租房者来看，估计的缺失组格人口数最多的是模型 8 的 378.68 人，最少的是模型 3 的 12.78 人。面对差别如此大的这些估计值，我们该如何选择呢？一是进行模型之间的选择，取最优模型的估计值；二是与实地调查或经验判断的结果进行比对，取最接近实地调查或经验判断结果的估计值。(3)不同对数线性模型的似然比估计值差异大。其中，模型 5 的 4 个事后层的似然比估计值最小(饱和模型除外)，分别是 3.15，6.53，3.55 和 3.04，而模型 3 的 4 个事后层的似然比估计值分别为 68.55、52.80、84.54 和 70.73。

从表 4 可以看出，若给定显著性水平 0.01，虽然模型 5 和模型 8 均通过了拟合优度检验，但由于后者模型参数多，所以选择前者。也就是说，最适合用来拟合表 2 数据的是对数线性模型 5，即人口普查与行政记录独立但均与事后调查相关的模型。

对数线性模型 5 中的 4 个事后层的总人口数、标准差和 95%概率把握程度下的置信区间如表 5 所示。

从表 5 可以看出，在 20～29 岁有房者和租房者的总体规模估计值的置信区间长度分别是 126.42 人和 212.82 人；在 30～44 岁，有房者和租房者的总体规模估计值的置信区间长度分别是 233.74 人和 346.72 人。可见，无论是 20～29 岁还是 30～44 岁，有房者的总体规模的置信区间总是小于租房者，也就是说，有房者比租房者的估计精度高。

四、结 语

(一) 结论

双系统估计量要求人口普查与事后调查独立。但这两项调查其实是很难做到独立的。如果某人在人口普查中得到了好处，那么就会乐于参加事后调查，这使得双系统估计量低估实际人口数；如果某人在人口普查中受到了不公正的待遇，那么就会躲避事后调查，这使得双系统估计量高估实际人口数。多年来，双系统估计量一直遭到国外学者批评的原因也正是如此。如果在人口普查与事后调查的基础上再增加一个系统，例如行政记录系统，那么因前两个系统不独立而引起的交互作用偏差的影响将会显著减少。例如，在逃犯人同时被人口普查及其事后调查遗漏的可能性较大，但其户籍档案或监狱档案则往往会包括他们。也就是说，三系统覆盖人口总体的范围要大于双系统，因而三系统估计量估计的人口数的

精度显著高于双系统估计量。因此,在未来的人口普查质量评估中用三系统估计量替代双系统估计量是必然的,而且是有效的。

(二)相关问题的若干说明

"捕获—再捕获"模型的典型问题是,对封闭的池塘中的鱼进行二次或三次或更多次独立重复捕获,用每次捕到的鱼的数量和在某二次、某三次……捕获中同时出现的鱼的数量来估计池塘中鱼的总数。本文是把这个模型移植到人口调查当中,用来自 3 个来源的对人口总体进行登记的名单中被登记的人口的数量和同时出现在某两个名单中,以及同时出现在 3 个名单中的人员的数目来估计总体的人口数量。对这种移植,需要进行以下说明。

第一,对人口总体的三次登记并不是在同一时点上进行的,而在 3 个不同时点上所观察到的并不是同一个人口总体。在实际应用中,为了与"捕获—再捕获"模型的理论背景相吻合,必须要对第二次登记、第三次登记的结果进行校正,使其成为对第一次登记时点人口状态的追溯结果。

第二,对人口的登记与对鱼的捕获在机制上并不相同。必须要对人口登记做若干假定,在这些假定成立的条件下,由"捕获—再捕获"模型移植而来的估计量才是无偏的。对此,胡桂华(2011)针对双系统估计量的无偏性提出了它必须要满足的 3 个假设条件,且这 3 个假设条件等价于下面的两个要求:(1)人口普查与事后调查相互独立;(2)人口总体中的每一个人在人口普查中被登记的概率相同,并且在事后调查中被登记的概率也相同(在人口普查中被登记的概率与在事后调查中被登记的概率可以不同)。

第三,本文讨论中所提到的三次登记资料都指的是对人口总体进行全面登记的结果。然而,在事实上,人口普查质量评估工作是在各省按城乡分层后以普查小区为单位抽取样本的,所以在构造三系统估计量时所依据的三系统资料是以普查小区为单位的样本资料而不是以人为单位的全面调查资料。与之相应,本文讨论中所用到的组格实际频数显然应代之以用有限总体概率样本构造的估计量;本文讨论中所给出的人口总体实际人口数估计量方差的计算公式在这里亦不宜应用,而应当采用复杂抽样方案条件下方差估计量的近似算法(如简单随机抽样刀切法、分层随机抽样刀切法和泰勒膨胀法等)。

第四,在本文所引用的实证案例中,考虑到年龄和是否拥有住房这两个标志的不同取值会导致被调查者登记概率不同,为了满足在登记概率相同的总体内构造三系统估计量的要求,将调查对象按年龄和是否拥有住房两个标志进行了交叉分层。事实上,可能会导致被调查者登记概率不同的标志并不仅仅是年龄和是否拥有住房。于是,寻找和选择可能会导致被调查者登记概率不同的标志就成为一个需要认真研究的问题。特别是,由于人口普查(及其事后调查)与行政记录在政府工作职能上有明显的区别,因此影响二者登记概率的标志也不会完全一样,这就更增加了寻找和选择这种标志的难度。

(三)对中国的启示

中国从 2000 年起开始使用存在交互作用偏差的双系统估计量进行人口普查质量评

估，但尚无应用三系统模型的考虑。有关部门应尽快开展这方面的研究工作，为中国人口普查质量评估工作引入三系统模型做好准备。在中国使用三系统模型具有可行性。首先，具备了构建人口行政记录系统的基本资料。在中国目前的条件下，人口行政记录系统资料来源较多，包括户籍资料、暂住人口资料、出生人口资料、死亡人口资料、人口迁移资料等。虽然这些人口资料的登记机构、登记时间、登记用途、登记方法存在差异，但只要对这些资料进行综合、廓清，形成一份定格在某个标准时点上的人口名单，就能够形成一个自成体系的行政记录资料系统。其次，三系统模型理论已经比较成熟。西方学者利用对数线性模型通过最大似然估计构建了若干三系统模型，给出了选择最优三系统模型的具体操作方法，并通过实证案例论证了三系统模型优于双系统估计量。再次，美国普查局计划在 2020 年人口普查质量评估中首次使用三系统模型替代双系统估计量估计实际人口数及人口普查净误差。

在中国人口普查质量评估工作中使用三系统模型意义重大。一是有助于人口普查质量评估工作走在世界人口普查质量评估领域的前列。二是能显著提高中国未来人口普查净误差及其普查遗漏估计值的精度。由于净误差是普查遗漏估计的基础，所以只要提高了净误差估计的精度，就能间接提高普查遗漏估计值的精度。三是三系统模型尤其适合于特殊人群（患有某种流行性疾病的人口、吸毒与贩毒人口、某种信仰的人口、无家可归人口等）数量的估计，而双系统估计量则往往低估了他们的数量。

参考文献：

1. 贺本岚等（2010）：《人口普查事后质量抽查的有关问题：国外经验及借鉴》，《商业经济与管理》，第 9 期。
2. 胡桂华（2010）：《美国 2000 年和 2010 年人口普查质量评估方法解读》，《数理统计与管理》，第 2 期。
3. 胡桂华（2011）：《人口普查净误差构成部分的估计》，《统计研究》，第 3 期。
4. Zaslavsky A.M. and Wolfgang G.S.（1993），Triple System Modeling of Census，Post-Enumeration Survey，and Administrative-List Data. *Journal of Business and Economistatistics*. 11（3），279-288.
5. Hogan，H.（1992），The 1990 Post-Enumeration Survey：An Overview. *The American Statistician*. 46（4），261-269.
6. Bell R.M. and Cohen M.L.（2008），Coverage Measurement in the 2010 Census. National Academies Press.
7. Scott Konicki.（2012），2010 Census Coverage Measurement Estimation Report：Adjustment for Correlation Bias. U.S.Census Burea.
8. Fienberg S.E.（1972），The Multiple Recapture Census for Closed Populations and Incomplete 2^k Contingency Tables. *Biometrika*. 59（3），591-603.
9. Wachter K.W. and Freedom D.A.（2000），The Fifth Cell Correlation Bias in U.S.Census Adjustment Evaluation. *Review*. 24（2），191-211.
10. Bell W.R.（2001），Accuracy and Coverage Evaluation Survey：Correlation Bias. U.S.census Bureau.

（责任编辑：朱　犁）