

经济及社会事务部
统计司

方法研究

F 辑 第 82 号

人口和住房普查 编辑手册



联合国
纽约，2001 年

说 明

本出版物中使用的名称以及材料的编排方式，并不意味着联合国秘书处对任何国家、领土、城市或地区或其当局的法律地位、或者对其边界或界线的划分表示任何意见。

本出版物中使用的“国家”一词在适当情况下亦指领土或地区。

使用“发达地区”和“发展中地区”等名称是为了统计上的方便，未必表示对某一国家或地区在发展进程中所处发展阶段作出的判断。

联合国文件都用英文大写字母附加数字编号。

ST/ESA/STAT/SER.F/82

<http://www.un.org/Depts/unsd/>

联合国出版物

出售品编号：C.00.XVII.9

版权©联合国，2001年

版权所有，翻印必究

前 言

多年来，联合国发布了一系列手册和技术报告，目的是协助各国规划和进行更具成本效益的人口和住房普查工作。这些手册和报告经过反复审查和更新，反映了普查中新出现的发展动态和问题。为协助各国作好 2000 年和今后几轮的普查筹备工作，已经编写了一系列手册，本出版物是这些手册的一部分。本系列中的其他手册包括：

(a) 《人口和住房普查管理手册》，F 辑，第 83 号(联合国出版物，出售品编号 E.00.XVII.15)；

(b) 《地理信息系统和数字绘图手册》；F 辑，第 79 号(联合国出版物，出售品编号 E.00.XVII.12)；

(c) 《未来人口普查经济特点收集指导》(即将出版)。

《关于人口和住房普查的原则和建议，第一次修订本》，(联合国，1998 年)审查了普查初期阶段采用质量控制和改进系统的好处，这一系统对有效进行整个普查工作至关重要。该手册还审查了编辑计划的重要性，该计划的制订应成为整个普查方案的一部分，并应与其他普查计划和程序兼收并蓄。《人口和普查编辑手册》应参照《原则和建议》的内容，这对该手册的使用很有帮助，因为《原则和建议》为第三、四和五章中概述的编辑程序提供了大量背景资料。

本出版物旨在使各国广泛了解普查和调查数据编辑方法，并向有关官员提供各种普查编辑方法使用情况资料，同时鼓励各国保留历来的编辑经验，促进主题专家和数据处理专家之间的交流，并翔实地介绍目前正在开展的普查或调查活动，以免下一次的普查或

调查出现重复现象。

本手册审查了人工编辑和计算机辅助编辑的利弊。在大规模的普查中，人工改错并不划算，因此在专门设计的计算机程序中通常具体规定纠正这类误差的条件。这些程序能够根据有关个人或住户或其他个人或住户的情况，自动进行查错和推算。许多普查手册都介绍了数据自动校正技术。

在查找和校正误差的过程中，计算机编辑发挥了重要作用。在计算机编辑阶段，可以通过与主题专家磋商，确定对一致性进行检查的细节。所查到的误差可参照原调查表校正，也可自动校正。自动编辑虽然加快了数据处理速度，但需要确实保障要处理的数据质量。

本出版物由导言和五个章节组成。导言部分介绍了普查过程，以及普查中会出现的各种误差。第一章涉及普查编辑的基本问题。第二章至第五章介绍数据处理的各个阶段，进行普查数据编辑的程序和方法。技术上的考虑，特别是与编程有关的考虑在附件部分作了说明。

尽管本手册主要涉及人口和住房普查编辑，但其中的许多概念和方法也适用于各种调查活动。

美国普查局为本手册的编写作出了贡献，其中 Michael J. Levin 提供了各种服务，对此我们表示衷心的感谢。我们还要感谢世界各地的许多主题专家和程序员，他们运用丰富的普查和调查经验，审查并协助完成了本出版物。

目 录

	页 次
前言	iii
缩略语	xii
导言	1
A. 本手册的目的	1
B. 普查过程	1
C. 普查过程中出现的误差	2
1. 覆盖误差	2
2. 内容误差	2
D. 本手册的结构	3
一、 普查和调查编辑	5
A. 对编辑工作的历史回顾	5
B. 编辑小组	5
C. 编辑实践：编辑与未编辑的数据	6
D. 编辑要则	8
1. 过分进行编辑的弊端	10
2. 未知数据的处理	11
3. 失真的更正	11
4. 容限的确定	11
5. 总结编辑工作经验	11
6. 质量保证	11
7. 编辑成本	11
8. 推算	12
9. 存档	12
二、 编辑应用系统	13
A. 人工校正与自动校正	13
B. 数据校正准则	15
C. 有效性和一致性检查	17
1. 自顶向下编辑法	17
2. 多变量编辑法	18
D. 编辑考虑	21
E. 改正和推算数据的方法	22
1. 静态推算或“非流水”技术	22
2. 动态推算或“流水”技术	22

3. 动态推算（流水）问题.....	25
4. 检验推算矩阵.....	30
5. 推算标记.....	32
F. 其他编辑系统.....	34
三、结构编辑.....	35
A. 地理编辑.....	35
1. 住所的位置（地区）.....	35
2. 城市和农村常居地.....	35
B. 覆盖面检查.....	36
1. 事实上和法律上的查点.....	36
2. 住户和住房单元的分级.....	36
3. 调查表的片断.....	36
C. 住房记录的结构.....	36
D. 住房记录与人口记录之间的对应.....	36
1. 空置和入住住房.....	36
2. 重复的住户和住房单元.....	37
3. 遗漏的住户和住房单元.....	37
4. 居住者数目与居住者之和的一致.....	37
5. 居住者与建筑物/住户类型之间的一致.....	38
E. 重复的记录.....	38
F. 特殊人口.....	38
1. 集体单位的人.....	38
2. 无住房人口.....	39
G. 确定户主和配偶.....	40
1. 编辑户主变量.....	40
2. 编辑配偶.....	42
H. 年龄和出生日期.....	42
1. 在有出生日期但无年龄时.....	42
2. 在年龄和出生日期不一致时.....	42
I. 计点无效登录项.....	42
四、人口项目的编辑.....	43
A. 人口特点.....	43
1. 关系（P2A）.....	44
2. 性别（P3A）.....	45
3. 出生日期和年龄（P3B）.....	46
4. 婚姻状况（P3C）.....	49
5. 初婚年龄（P4F）.....	50

6. 生育率：所生子女数（P4A）和健在子女数（P4B）	50
7. 生育率：最后一个活产儿的出生日期（P4C）	56
8. 生育率：第一胎时的年龄（P4G）.....	56
9. 死亡率（P4D）	57
10. 丧母和丧父情况（P4E）和母亲的行号	57
B. 迁移特点	57
1. 出生地（P1C）	58
2. 公民资格（P3D）	59
3. 持续居住时间（P1D）	60
4. 以前居住地（P1E）	61
5. 在以前某个订定日期的居住地（P1F）	61
C. 社会特点	62
1. 读写能力（识字情况）（P5A）	62
2. 就学情况（P5B）	62
3. 教育程度（学完的最高年级或程度）（P5C）	63
4. 学习科目和学历（P5D）	63
5. 宗教（P3E）	63
6. 语言（P3F）	64
7. 民族（P3G）	64
8. 残疾（P8A）	65
9. 缺陷和障碍（P8B）	66
10. 残疾的原因（P8C）	66
D. 经济特征	66
1. 活动状况（P6A）	66
2. 工作时间（P6B）	69
3. 职业（P6C）	69
4. 产业（P6D）	69
5. 就业状况（P6E）	69
6. 收入（P6F）	70
7. 机构部门（P6G）	70
8. 工作地点（P6H）	70
五、住房情况编辑	71
A. 基本题目	72
1. 建筑物：建筑物的说明（H01）	72
2. 建筑物：外墙建筑材料（H02）	72
3. 建筑物：建筑年份或时期（H03）	72
4. 住所：住所位置（H04）	72
5. 住所：住所类型（H05）	73

6. 住所：住用状况 (H06)	73
7. 住所：所有权类别 (H07)	73
8. 住所：房间数目 (H08)	74
9. 住所：楼面面积 (H09)	74
10. 住所：供水系统 (H10)	74
11. 住所：厕所和排污设备 (H11)	75
12. 住所：洗澡设备 (H12)	75
13. 住所：厨房设备 (H13)	76
14. 住所：照明 (H14)	76
15. 住所：固体废物处理 (H15)	76
16. 住所：一个或多个住户居住 (H16)	76
17. 住所：居住人数 (H17)	76
18. 住户：户主的特征 (H18)	77
19. 住户：保有权 (H19)	77
20. 住户：租金和业主自己居住的住房费用 (H20)	77
B. 其他题目	77
1. 建筑物：住所数目 (A01)	77
2. 建筑物：电梯 (A02)	77
3. 建筑物：农场 (A03)	78
4. 建筑物：屋顶的建筑材料 (A04a)	78
5. 建筑物：地板的建筑材料 (A04b)	78
6. 建筑物：修理情况 (A05)	78
7. 住所：卧室数目 (A06)	78
8. 住所：烹调燃料 (A07)	79
9. 住所：暖气类型和暖气所用能源 (A08)	79
10. 住所：有无热水 (A09)	79
11. 住所：管输送煤气 (A10)	79
12. 住所：电话 (A11)	79
13. 住所：住房单元的利用 (A12)	80
14. 居住者：车辆数目 (A13)	80
15. 居住者：耐用设备 (A14)	80
16. 居住者：可供住户使用的室外空间 (A15)	80
C. 住用和空置的住房单元	80

附 件

一、导出变量	83
A. 住房记录的导出变量	83
1. 住户收入	83

2. 家庭收入.....	83
3. 家庭类型.....	83
4. 有亲属关系的人.....	84
5. 家庭的工作者.....	84
6. 全套管道.....	84
7. 全套厨房.....	85
8. 租金总额.....	85
B. 人口记录的导出变量.....	85
1. 经济活动状况.....	85
2. 小家庭数目和小家庭的亲属.....	85
3. 自己的子女.....	87
4. 住房内的父母.....	87
5. 目前在校年数.....	87
二、调查表格式与键入的关系.....	89
三、键入的考虑因素.....	93
A. 输入数据.....	93
1. 扫描.....	93
2. 埋头键入.....	93
3. 交互式键入.....	94
B. 测试键入指令.....	95
C. 验证.....	95
1. 依附式验证.....	95
2. 独立式验证.....	95
四、流程图样本.....	97
五、推算方法.....	103
六、计算机编辑软件包.....	107
词汇表.....	109
参考文献.....	113

表

1. 利用编辑和未编辑的数据，按 15 岁年龄组和性别分列的抽样人口.....	7
2. 1990 年和 2000 年按 15 岁年龄组未知人数确定的人口和人口变化.....	8
3. 1990 年和 2000 年按无未知数据的 15 岁年龄组确定的人口和人口变化.....	8

图

图 1. 假设的典型住户，包括住户成员之间的关系、性别和生育情况	15
图 2. 户主和配偶出现同一性别的住户实例	15
图 3. 住户某些成员出现同一年龄的实例	15
图 4. 所报年龄可能有出入的住户实例	17
图 5. 选定人口特征多变量编辑规则的实例	19
图 6. 未经编辑数据集中户主和配偶性别相同及其解决的实例	19
图 7. 采用伪代码方式，更正性别变量的编辑说明实例	20
图 8. 有 3 名子女的极年轻寡妇多变量编辑分析的实例	20
图 9. 选定项目共同代码的实例	21
图 10. 作为动态推算输入实例的抽样住户	23
图 11. 基于性别和关系的年龄初始静态矩阵	24
图 12. 一项变动后动态推算矩阵的实例	24
图 13. 多项变动后动态推算矩阵的实例	24
图 14. 户主和户主父亲未分配语言情况下的实例	26
图 15. 语言动态推算矩阵初始值	27
图 16. 住户成员未分配语言情况下的实例	27
图 17. 户主和子女年龄数据遗漏的子女的实例	28
图 18. 户主和子女年龄和年级数据遗漏的子女的实例	29
图 19. 非流水阵列的抽样组值和抽样推算代码	30
图 20. 每个错误推算次数摘要报告的实例	31
图 21. 调查表错误抽样报告	31
图 22. 按包括多个变量的调查表显示的补充错误列表的实例	31
图 23. 带推算值标记的抽样人口记录	33
图 24. 生育率空白和增加标记的年轻女性作标记的实例	33
图 25. 户主列为第一人的住户的实例	40
图 26. 户主列为第五人的住户的实例	41
图 27. 拥有生育信息的住户的说明	52
图 28. 在年龄和所生子女数有效时确定健在子女的初始值	53
图 29. 拟为各对已知信息发展的抽样推算矩阵	55

方 框

1. 普查编辑须知	6
2. 校正数据要则	15
3. 结构编辑的指导原则	35

4. 年龄的编辑和推算	47
-------------------	----

附 图

A.一.1. 大家庭图解.....	86
A.一.2. 有两个小家庭的住户实例.....	87
A.二.1. 附有个人页的调查表实例.....	89
A.二.2. 附有个人页的调查表流程实例.....	89
A.二.3. 调查表实例，所有人列在同一页上的住户页.....	90
A.二.4. 住户页调查表流程实例，每页多个人.....	90
A.二.5. 多人住户页实例，没有键入问题.....	91
A.二.6. 多人住户页实例，可能有键入问题.....	92
A.四.1. 确定户主的流程图样本.....	98
A.四.2. 确定住户是否存在配偶的流程图样本.....	99
A.四.3. 编辑户主和配偶性别变量的流程图样本.....	100

缩 略 语

CD	死亡儿童
CEB	所生子女人数
CLA	生活在外的子女人数
CLH	住在家中的子女人数
CS	健在的子女人数
EA	查点区
GIS	地理信息系统
ILO	国际劳工组织（劳工组织）
ICIDH	缺陷、残疾和障碍的国际分类
NIM	新的推算方法
OCR	光学认字
OMR	光学标志辨认器
PES	查点后调查
SAS	统计分析系统
SNA	国民核算体系
SPSS	社会科学统计包
UNESCO	联合国教育、科学及文化组织（教科文组织）

导 言

A. 本手册的目的

1. 计划周全的普查或调查能够提供误差很少的最后产品，因此是一个国家的无比重要的资料来源。为使普查或调查取得准确无误的结果，应尽量避免数据出现差错或不相一致的情况，这在数据处理阶段尤为重要。在进行和完成数据收集和输入后，要利用有关程序查找数据记录中的差错，并调整个别项目，这一程序叫作人口和住房普查编辑。

2. 任何普查或调查数据都不是完整无缺的。各国早已认识到，从普查和调查中得来的数据并不可靠准确。它们采取各种办法，弥补数据上的不足，并解决答复不一致的问题。但是，由于普查间隔很长，因此往往难以提供有关文件，详细介绍用于编辑数据的程序。因此，各国在制订新一轮普查或调查计划时，须调整最初进行数据收集活动时采取的方法。

3. 《人口和住房普查编辑手册》的主要目的，是使人们真正了解普查和调查数据编辑方法，并为有关官员提供使用各种普查编辑方法的信息。同时还鼓励各国保存其历来积累的编辑经验，加强主题和数据处理专家之间的交流，并翔实地介绍本次普查或调查期间开展的活动，以免今后重复以前的工作。

4. 主题专家¹和数据处理专家小组在制订编辑说明及普查和调查方案时，可将本手册作为参考书使用。本手册采用了“细则手册”的形式，使各国能够完全根据本国的统计现状进行编辑。本出版物还旨在进一步促进这些专家在制订和实施编辑方案的过程中进行交流。

5. 导言部分介绍了普查过程，普查中会出现的各种差错，以及普查编辑的基本原则。随后各章介绍了在数据处理各个阶段编辑普查数据的程序和方

法。虽然本手册重点论述的是人口和住房普查编辑问题，但其中的许多概念和方法也适用于调查活动。

B. 普查过程

6. 人口和/或住房普查的全过程包括：收集、汇编、评价和发布与所有个人及其住所有关的人口和/或住房及经济和社会数据（联合国，1998年）。普查是在某一特定时间内，在一国全国上下，或其明确划定的部分地区进行的。这种普查能够及时提供某个地区人口和住房的简要情况。

7. 普查的基本目的，是提供有关一个国家人口规模、分布和特点的资料。普查数据用于各种用途，如决策、规划和行政管理，以及教育、劳动力、计划生育、住房、保健、运输和农村发展等方案的管理和评价。普查数据在行政上的一个基本用途，是划分选区，并分配管理机构代表权。普查还是了解调查研究情况的重要渠道，它所提供的数据可用来对人口组成和分布进行科学分析，并制订预测未来经济增长的统计模式。普查还为工商业提供各种基本数据，它们需要利用这些数据评估住房、学校、家居、衣食、娱乐设施、医疗用品及其他货物和服务方面的需求。

8. 所有普查和调查都具有某种重要特征，这包括：(a) 筹备工作；(b) 查点；(c) 数据处理，包括数据输入、编辑和制表；(d) 建立数据库和传播结果；(e) 评价结果；及(f) 分析结果。

9. 筹备工作包括多项内容，如确定普查的法律基础；编制预算；编制日历；作出行政安排；绘制图表；编制住房单元清单；拟订制表计划；编制调查表；以及制订计划和培训从事查点、试行、数据处理和宣传工作的人员。

10. 查点工作取决于如下若干因素，如所选定的查点方法、查点期的时间安排和时限；监督水平；以及是否和如何使用抽样方法。数据输入后，必须进

¹按本手册中的定义，主题专家包括在人口、住房和其他有关领域工作的专家。

行编码、输入、编辑和制表。数据处理能够产生微观和宏观数据基。国家普查/统计局利用这些数据基进行制表、时序分析及制图和绘图，同时还用于地理信息系统的专题绘图和其他传播技术。评估内容和覆盖结果采用各种方法，包括人口分析和查点后调查。最后，结果分析从各个方面入手，包括概要说明普查结果，从政策角度进行普查结果分析，以及从一个或多个方面对有关国家的人口和社会状况进行详细的分析研究。

C. 普查过程中出现的误差

11. 普查数据中有多种误差来源，一般可分为两种，即覆盖误差和内容误差。

1. 覆盖误差

12. 覆盖误差是普查查点中遗漏或重复计数个人或住房单元所引起的误差。覆盖误差除其他外，还源于另外一些原因：如地图或查点区清单不完整或不精确；查点员未详细检查指定地区内的所有单位；重复计数；遗漏不愿接受查点的人；对某类人，如对访问者或非居留的外国人处理不当；以及查点后普查记录遗失或被销毁。对覆盖误差应尽可能在实地加以纠正。统计局编辑程序能消除实际重复的记录。但是，必须注意确定是否存在重复的个人或住户。例如，双胞胎的资料就可能相似，除非采用数序列。因此，在普查过程中适用了编辑原则，此项原则可确定何时接受和拒绝可能重复的信息，以及何时按推算方法进行更正。

13. 第三章介绍了结构编辑。这种方法用来检查住户确切人数记录，并修改顺序排列和重复人数。

2. 内容误差

14. 内容误差是因报告或记录的个人、住户和住房单元属性有误引起的误差。此外，造成这种误差的原因可能包括，问题的设计或排序不合理，被调查者与查点员之间缺少交流，编码和数据输入有误，人工和计算机编辑不精确，结果造成汇总表有误差。在普查过程的初期阶段，应当编制和储存编辑系列（亦

称审计系列），以免造成数据遗失。以下各节对上述每种误差作了说明。

(a) 调查表设计中出现的误差

15. 造成内容误差的原因之一，是问题或指令表述不明确。由于查点期间和查点后，还会出现设计不合理的调查表，因此对调查表的类型、格式、准确措词及项目编排应十分仔细地加以考虑。应采取预测的办法，尽量减少因调查表设计不合理而有可能产生的误差。例如，如果跳行方式不明确，或安排不当，查点员就有可能错误地略过调查表中的某些部分，而且无法收集到全部有关信息。

(b) 查点员造成的误差

16. 查点员和调查对象相互影响，除非在进行普查时，使用自行管理的调查表。查点员在提问时有可能造成两种误差，如省略或改变有关问题的措词，或不能向调查对象充分说明有关问题的含义。查点员在记录各种答复时，也有可能造成一些误差。查点员培训质量对所收集的数据质量具有决定性作用。因此，应在普查程序的各个方面，对查点员进行适当的培训，并使他们了解其在普查过程中所起的作用的重要性，以及如何使查点工作适合普查其他阶段的情况。此外，由于查点员具有许多不同的背景，而且受教育程度有高有低，因此必须对他们进行培训，以使他们真正懂得如何通过提问，获得适当的答复。

(c) 调查对象造成的误差

17. 如果调查对象对某类项目理解有误，就有可能使数据出错。误报或由他人代替作出答复（某个人，而不是有关资料所涉及的人对调查表所作的答复），也会造成误差。可以采取一些方法增强个人答复的准确性，如进行普查宣传，并对查点员进行培训，使其能够解释普查的目的和提出各种有关问题的原因。一些国家使用了自行管理调查表，因此不存在查点员与调查对象相互影响的问题。使用自行管理表格时，调查对象若对有关问题或指示产生误解，就会造成误差。

18. 调查对象和查点员造成的误差最好在查点阶段纠正，同时仍可借助调查表格、调查对象和查点员。监督员必须能够对查点员进行培训，并核对查点员在查点期间定期收集的数据，以免查点员在数据中掺入常有的偏见。在向地区或中心部门发送调查表之前，监督员应当在实地处理查点员和调查对象造成的误差。

(d) 编码误差

19. 由于编码员有可能编错信息代码，因此编码过程中会出现一些误差。在输入数据期间，键控出错可能会给数据造成误差。一般而言，在此阶段若不进行监督和核查，就会拖延数据的发布，而且会使今后查找和校正误差的工作更加困难。在进行编码前或编码期间，经常要进行人工编辑。

(e) 数据输入误差

20. 在数据输入软件中，可以设置区域检查和某种基本的一致性检查程序。情报资料输入系统能够将字段值或数据项值归入该项值允许的范围，从而使数据输入操作员能有更多的机会键入合适的数字，而且还可以在数据编制过程的后几个阶段减轻数据编辑方面的某些负担。不过，这些检查会使数据输入速度放慢。因此，在输入数据期间，应根据需要仔细考虑在多大程度上进行一致性检查，以保持适当的数据输入速度。应预先进行权衡，以免负责输入数据的办事员在这上面花太多时间。对键控进行核查，无疑会提高数据质量。对键控表格进行核查时，可重新键入相同的资料，这项工作经常在抽样的基础上进行。

(f) 计算机编辑误差

21. 编辑在普查数据处理中是一个关键的步骤。在编辑过程中，可以利用拟真数据辨识无反应或不统一的信息。但一种自相矛盾的情况是，任何这类编辑工作都有可能引起新的误差。

(g) 制表误差

22. 由于数据处理错误或“未知”（尚未提供）数据的采用，在制表阶段有可能出现误差。在此阶段校正已有的误差时，难免会出现新的误差。重要的是

要维护处理系统，而不是设法对表格本身进行校正，以便在表格内容不统一时，还能进行编辑。如果从数据处理到公布的所有阶段误差始终存在，它们就会显现出来，而且有关结果的价值会受到质疑。在表格发布前，有必要进行全面检查，以便能够为预期的地理单位准备全部计划编制的汇总表。在编辑阶段进行区域检查和一致性检查，可以减少大部分误差，但还有必要在制表后进行汇总检验。经过培训并富有经验的人员应当查阅各种表格，以弄清不同单位报告的数字是否与所了解的当地情况相符。在某些情况下，快速参阅普查明细表，可以查明编码误差。还有一种有效的方法，是计算选定比率和增长率，并与以前抽样调查公布的普查数字或其他数字相比较。不过，若准备与其他调查数字进行比较，所使用的概念应具有可比性。如果在最后制表中发现误差，应首先对数据集进行校正。

23. 很重要的一点是，数据处理员在纠正数据集中的问题时，不能更改制表。这类更改不能在微观数据中进行，因此在制订和运用其他方案时亦不能重复。该小组应当对微观数据集进行各种更改，以使国家普查/统计局的其他数据处理程序能够在一定程度上编制可比表格。此外，由于国家普查/统计局有时向公共部门和私人部门的研究人员和其他用户公布部分微观数据文件，因此表格应能复制。

24. 上文已经指出，在普查过程中，还要进行一些有序的相关活动，而且每次活动都有可能出现误差。重要的是应当记住，计算机编辑是反馈系统的一部分，它既能协助进行制表，又能协助进行数据收集和整理。国家普查/统计局防止计算机编辑出现问题的最好办法，是尽可能进行实地编辑。该局还需要确保编码和数据输入的精确性，并应在所有活动中，包括在数据输入、编辑及制表中，继续进行反馈。

D. 本手册的结构

25. 第一章审查了编辑工作在普查和调查中所起的作用。其他各章述及了一些特定的专题。第二

章介绍了编辑和推算方法的实际用途。第三章介绍了结构编辑，即同时审查住房和人口项目情况的编辑，以及协助进行其他编辑的某种程序，如确定是否有或

只有一个户主。第四章审查了人口编辑问题，第五章涉及住房编辑。最后，一系列附件审查了与人口和住房普查编辑及推算有关的各种具体问题。

一、普查和调查编辑

A. 对编辑工作的历史回顾

26. 计算机问世前，要雇用大批半熟练的办事员，为大部分普查工作编辑单项表格。但是，由于极少数项目之间的复杂关系，简单的核对不可能发现数据中可能存在的全部不一致之处。不同的办事员对有关规则的解释各有不同，甚至同一名办事员也会作出不相一致的解释。

27. 随着计算机的推广使用，普查编辑有了明显的改变。与人工编辑相比，计算机能查出更多的一致之处。普查过程并行规则已使自动推算成为可能（Nordbotten, 1963年；Naus, 1975年）。同时，在普查过程中，能够接触到越来越多的调查对象，或至少是这些调查对象已填写好的调查表。许多编辑小组开始感到，编辑工作做得越多越好，编辑工作做的越周密细致，所取得的结果就越精确。各种方案产生出数以千计的错误信息，需要对原始表格进行人工检查，就某些调查而言，还需要重新询访调查对象。

28. 随着计算机的使用，对数据集进行更改变得更加容易了。有时这些更改还能校正各种记录或项目。许多记录经过了计算机多次处理，而且每次都由不同人员对其中的误差和不一致之处进行审查（Boucher, 1991年；Granquist, 1997年）。

29. 在整个普查过程中，开发了一些通用普查编辑组件，其中有些组件迄今仍在沿用。最初，这些组件是为计算机主机研制的，后来有些组件经过修改，已用于个人计算机。在此期间，Fellegi 和 Holt（1976年）提出了一种新的通用编辑和推算方法，这些方法虽未马上得到应用，但今天已日益为人们所采纳，这是因为国家普查/统计局的编辑工作变得越来越复杂。

30. 1980年代普查编辑工作取得了重大进展，国家普查/统计机构开始利用个人计算机对数据进行输入、编辑和制表。在输入数据阶段，或这之后不久，数据处理机很快就能对数据进行联机处理。在开展调

查工作时，工作人员可以编制一些程序，在收集数据期间查找误差，或在输入数据的同时，直接在机器中查找误差。进行计算机编辑，可以不断与调查对象进行更多的接触，从而解决编辑过程中遇到的各种问题（Pierzchala, 1995年）。

31. 在最初几年，对普查和调查数据进行了更加复杂而全面的核对，这项工作似乎十分成功。编辑小组制订了比以往更加复杂的编辑说明，数据处理专家用数月时间，编制了流程图和程序编码。分析人员很少对软件包进行评价。编辑似乎能够纠正数据收集、编码和键控最初阶段产生的各种问题。但是，许多分析人员了解到，在不少情况下，所有这类额外的编辑都损坏了数据的可靠性，或至少延长了取得结果的时间，或使结果出现了偏差。有时上述程序对有关数据进行三番五次的处理，对项目进行接二连三的校正，致使结果与最初未经编辑的数据不能相互比较。

32. 在普查或调查的过程中，进行这种面面俱到的编辑，严重拖延了许多普查和大型调查工作的开展。办事人员要花很多时间寻找人工表格，数据处理专家要为查看极少数案例而不断开发应用系统。Granquist（1997年）指出，许多研究表明，由于需要许多这类额外工作，因此“提高普查质量很难，或办不到，或根本不可能，而且许多严重的系统误差靠编辑是辨别不出来的”。

33. 由于国家普查和调查组织要不断开展普查和调查活动，因此进行广泛的计算机编辑是完全可行的。不过，各国家普查/统计局必须解决的一个问题是，哪种程度的计算机编辑适合实现其目的。

B. 编辑小组

34. 国家统计机构在进行普查筹备工作时，需要考虑有可能提高其工作质量的各种因素，其中之一是建立编辑小组。该小组要由普查管理人员、主题专家和数据处理人员组成，负责进行编辑工作。编辑小组应在普查筹备工作之初建立，在拟订调查表时建立

更为合适。在编辑过程中，编辑小组从始至终发挥重要作用。注意将小组成员聚集在一起，制订并实施编辑和推算规则，这可以使普查工作以更快的速度和更高的效率进行。

35. 普查官员与用户界举行会议，讨论制表和其他数据产品问题，有助于深入了解需要进行的编辑工作。用户时常要求编制特殊表格或各类表格，这需要进行额外编辑，以消除潜在的不一致之处。编辑小组应当订出计划，在进行初步编辑阶段采用这些表格，而不要在普查后，再将这些表格作为特殊表格采用。在提出异议或进行总编排期间，应制订编辑规则和编制计算机程序，这有助于这些程序本身的应用，并可加快编辑和推算各个部分的周转次数。随后要由编辑小组确定这方面各个过程的影响，并酌情采取补救行动。

36. 主题和数据处理专家应当共同制订各种编辑和推算规则。编辑小组要在普查筹备工作初期制订误差审查措施和编辑计划。普查或调查编辑小组要编制几套书面统一规则和校正方法。

37. 除制订编辑和推算规则外，主题和数据处理专家在普查或调查的所有阶段，包括在分析阶段，都必须相互协作。编辑内容过多同编辑内容过少一样有弊无利，而且还会在数据集中掺入未经编辑的或虚假的信息。因此，两个小组都必须负责妥善保存元数据基。编辑小组还必须有效利用现有行政资源和调查登记册，来改进随后的普查或调查活动。

38. 国家统计局/普查机构对巨型计算机的使用，限制了主题专家和数据处理专家之间的交流。在微型

计算机问世前，这种不协调的情况持续了很长时间。不过计算机程序包越来越适合用户的需要，目前不少主题人员都能制订和检验他们自己的编辑和制表计划。主题专家通常不处理有关数据，但他们往往了解处理这些数据的步骤。

C. 编辑实践：编辑与未编辑的数据

39. 各国开展普查编辑工作，是为了改进数据及其编排方式。本手册这一部分强调说明了国家普查/统计局在公布未经编辑的普查数据时所遇到的问题。为说明这些问题，采用了一套假设的数据。

方框 1. 普查编辑须知

普查编辑应当做到以下几点：

- (1) 向用户提供优质普查数据；
- (2) 确定误差的类型和来源；
- (3) 提供经调整的普查结果。

40. 虚拟国家的国家普查/统计局力求为多重用户提供服务，但力不从心。一些用户可能希望在表格中列入未知数据，以进行分析或调查研究，还有些用户可能希望提供“干扰”（可能的误差）最少的数据，以利于他们进行规划或决策。如果国家普查/统计局散发的是表 1 左边那种未经编辑的表格，分析人员和决策者在使用有关数据时，就必须作出假设。表 1 具体说明了这一点，但只涉及到少数个人。从此表中可以看出，在该国，有 23 人的性别未填报，15 人的年龄未填报。

表 1. 利用编辑和未编辑的数据, 按 15 岁年龄组和性别分列的抽样人口

年龄组	未编辑的数据				编辑的数据		
	总数	男	女	未填报的人数	总数	男	女
总数	4 147	2 033	2 091	23	4 147	2 045	2 012
15 岁以下	1 639	799	825	15	1 743	855	888
15 至 29 岁	1 256	612	643	1	1 217	603	614
30 至 44 岁	727	356	369	2	695	338	357
45 至 59 岁	360	194	166	0	341	182	159
60 至 74 岁	116	54	59	3	114	53	61
75 岁以下	34	12	22	0	37	14	23
未填报	15	6	7	2			

41. 大多数用户都会自己决定如何处理未知数据。一种可能不很成熟但合乎逻辑的方法, 是按已知数值的比例分配未知数值。如果国家普查/统计局选择通过推算得出未知值的办法, 编辑小组可确定 12 名男子和 11 名女子。在这一数字中, 男女人数基本相等, 但有所偏重, 因为在人口普查中, 查点较多的是女性。因此, 有关结果与表 1 右边载列的编辑数据相符。

42. 其他选择方法可用来处理未知值。例如, 编辑小组可决定只根据性别分布情况进行推算, 而不考虑现有其他信息, 如配偶之间的关系, 无论性别不明的个人是否如报告的那样, 是另一个人的母亲, 或性别不明的个人所生子女人数是否有正式记录。一种可供选择的推算方法, 是考虑在这些其他变量中, 采用一种或多种变量。

43. 国家普查/统计局还可采取另一种选择办法, 即根据年龄分布情况进行推算。例如, 在表 1 列示的抽样人口中, 共有 15 个示例未报告年龄。这些数据可再按与已知数值相同的比例进行分配, 这是一种合乎逻辑的推算方法。不过, 编辑小组可以考虑另外一些变量和组合数字, 如丈夫与妻子、父母与子女、祖父母与儿孙之间的相对年龄, 这或许能够取得更好的结果。

44. 在表 1 右边, 由于未知数据已经取消, 因此编辑数据“更加明确”(见“编辑数据”项下的竖行)。该表的这一边之所以无未知数据, 是因为有关程序将把它们归入其他答复中。不过, 许多人口学家和其他主题专家历来希望将未知数据列入有关表格, 如同表 1 中未编辑的数据那样。他们认为, 这种程序有助于他们对数字进行各种评估, 借以衡量普查程序的效力, 或协助制订未来普查和调查计划。在进行制表时, 无论有无未知数据, 如下两个目标都能够实现, 即编制出供实际用户使用的编辑表格和供评价用的未编辑表格。

45. 在发表的表格中使用未知数据将会遇到的另一个问题是, 未知数据有可能影响趋势分析。新技术使这种分析比以往容易了许多。例如, 表 2 显示了从连续两次普查中了解到的年龄分布情况。该小国未知人数有所减少, 1990 年为 217 人, 据报告约占答复人数的 6.5%, 而到了 2000 年, 只剩下 15 人, 不到答复人数的 1%。

46. 在这方面, 国家普查/统计局必须了解未知人数的不一致是如何影响单项普查和两次普查之间的变化的。例如, 在 1990 年的普查中, 未知人数占 6.5%, 这使得人们难以比较两次普查中, 15 岁年龄组的百分比分布变化。在这十年中, 15 至 29 岁的人所占百分比似乎只从 27% 增至 30%, 但所分布的未知人数有可能改变这种分析。

47. 经修订的表 3 按比例或按某种推算方法，显示了分布的未知人数。这里不难看出两次普查中数字和百分比的变化以及年龄组的分布情况。诚然，为

获得精确可靠的结果，编辑小组必须能使两次普查和/调查之间进行的编辑相互统一，这包括内部的统一。注明“未填报”的横行被取消。

表 2. 1990 年和 2000 年按 15 岁年龄组未知人数确定的人口和人口变化

年龄组	人数		人数 变化	百分比 变化	百分比	
	2000 年	1990 年			2000 年	1990 年
总数	4 147	3 319	828	24.9	100.0	100.0
15 岁以下	1 639	1 348	291	21.6	39.5	40.6
15 至 29 岁	1 256	902	354	39.2	30.3	27.2
30 至 44 岁	727	538	189	35.1	17.5	16.2
45 至 59 岁	360	200	160	80.0	8.7	6.0
60 至 74 岁	116	89	27	30.3	2.8	2.7
75 岁以上	34	25	9	36.0	0.8	0.8
未填报	15	217	-202	-93.1	0.4	6.5

表 3. 1990 年和 2000 年按无未知数据的 15 岁年龄组确定的人口和人口变化

年龄组	人数		人数 变化	百分比 变化	百分比	
	2000 年	1990 年			2000 年	1990 年
总数	4 147	3 319	828	24.9	100.0	100.0
15 岁以下	1 743	1 408	335	23.8	42.0	42.4
15 至 29 岁	1 217	952	265	27.8	29.3	28.7
30 至 44 岁	695	578	117	20.2	16.8	17.4
45 至 59 岁	341	230	111	48.3	8.2	6.9
60 至 74 岁	114	109	5	4.6	2.7	3.3
75 岁以上	37	42	-5	-11.9	0.9	1.3

D. 编辑要则

48. 编辑工作是根据预定的规则，对各种答复进行系统的检查和校正（或更正）。有些编辑业务需要由人进行人工校正，还有些则要利用计算机进行电子校正。如果国家普查/统计局不对普查或调查结果进行编辑，普查出版物就有可能出现某些无用的数据。编辑工作可以减少失真的评估，并有助于进行数据处理和增强用户的信心。此外，根据 Pullum、Harpham 和 Ozsever(1986 年)的看法“……要初步完

成编辑和精简程序，必须首先查明各种答复是否相互吻合，以及是否适合调查文书的基本格式”。

49. 原始普查数据文件包含多种误差。数据处理将这些误差分为两类。一类误差有可能妨碍进一步进行数据处理，另一类虽会产生无效或不一致的结果，但不会干扰随后数据处理操作的逻辑流程。正如《人口和住房普查的原则和建议第一次修订本》（联合国，1998 年，第 195 段）中说明的，第一类误差必须全部校正，第二类误差应尽量校正。因此，在数

据处理阶段，普查编辑的基本目的是尽量确定误差，并对数据集进行更正，以增强数据项的效用和一致性。不过，数据处理不可能校正所有普查误差，包括调查表中的各种答复，这些答复从内部讲虽具有一致性，但实际上存在两个问题，一是调查对象报告有误，二是查点员记录有误。

50. 编辑工作一般分两类：(1) **决定性编辑**，即确定实际存在的误差，和(2) **疑问性编辑**，即指出令人质疑的数据项目（Granquist, L. 和 Kovar, 1997年：第420页）。决定性编辑要确定实际存在误差的数据项目，而疑问性编辑则要指出有可能无效或不一致的数据。严重误差是通过决定性编辑查明的误差，包括无效或遗漏的条目，以及因内容不统一而造成的误差。相形之下，疑问性编辑要确定的，则是不属于带有明显主观性的编辑范畴的数据项目，这类项目与同一调查表中的其他数据和其他可疑条目相比，属于较高或较低一类。为了使人们保持对普查的信心，在进行编辑的过程中，必须查明和处理决定性编辑确定的误差，这在国家普查/统计局决定发布微观数据的情况下尤为重要。与决定性编辑相比，疑问性编辑较难校正误差，而且查找和消除误差的效用不大，同时还增加了整个编辑过程的成本。

51. 普查包括所有项目，这主要是因为规划者和决策者需要它们，因此，在进行普查编辑和推算期间，应当解决疑问性编辑确定的误差，但在调查期间无需这样做。不过，在确定最后普查编辑内容时，主题人员应对两类编辑内容进行调查，一是为试验性普查确定的编辑内容，二是在进行数据处理期间确定的编辑内容，以使单项编辑工作取得预期的成本效益。这类调查必须成为普查评价的一部分。正如Granquist和Kovar(1997年,第422页)指出的，“在使原始数据发生变化的标记数目中，精确数据属于其中的一部分，但在对编辑工作进行评价或研究时很少报告这类数据”。

52. 还有一套方法和术语与微观编辑和宏观编辑有关。正如已说明的那样，普查和调查编辑要查明数据记录中和记录之间存在的各种误差。本手册介绍

了微观编辑，这类编辑涉及如何确保单项数据记录的有效性和一致性，以及住户记录之间的关系。另一种编辑方法是宏观编辑，它的作用是核对汇总数据，以确保这些数据合理适用。例如，一个国家未报年龄的人可能占很大百分比。在为获得完整数据基进行年龄推算后，可进行宏观和汇总检查，这样，即使老年人报告年龄的人数较少，也不会降低推算价值。编辑小组可根据分析结果，专门采取一些措施，以减少有可能出现偏差的危险。

53. 进行编辑时，应尽量保留原始数据。编辑小组不仅需要优质明确的数据，还需要保留有关组织在实地收集的数据。在进行计算机处理的所有阶段都需要保存原始数据，因为说不定编辑小组会认为，它需要对编辑过程进行重新审查。有时，编辑小组在编辑过程中发现系统误差后，就会对原始数据进行审查。还有时进行这种审查，是因为发现数据集有遗失或重复，而且数据集必须重新组合和编辑。

54. 有时，误差是在数据处理机构以外产生的。Banister(1980年,第2页)指出，“我们知道，如果某个小组的人大多未回答一个特定的普查问题，就说明他们没有理解这个问题，拒绝回答这个问题，或对配合进行普查不感兴趣”。因此，她坚持认为，各小组的未答复率应列入普查存储媒介和公布的表格中。普查/统计局目前更有可能用小型光盘或其他媒介为研究人员保存这些数据。

55. 越来越多的证据表明，任何计算机编辑都不能替代更加优质的普查数据收集工作。国家普查/统计局了解到，在某种情况下，计算机编辑不仅具有局限性，而且还会产生反作用：较之其改正的误差而言，编辑增加了数据集的误差。更正普查项目与校正此项目并非一回事。因此，编辑小组必须相互协作，共同确定编辑工作从开始进行，到进行之中，到最后完成的全部过程。

56. 编辑和推算工作可能或不可能改进数据质量，但精确的数据集对进行分析十分有利。这项工作首先是设计普查调查表。人口学家和其他主题专家通

常要确定此类调查表的内容，这当中，他们往往要与各类用户进行磋商。最终产生的普查数据“主要是供广大学者、决策者和基层人民使用，而不是只供纯人口学家使用”（Banister, 1980年，第17页）。但是，在普查工作和国家普查/统计局的可信度有可能降低的情况下，所进行的普查须避免无效而不一致的记录。正如Banister所指出的，“普查组织会援引各种事例，如记者所写的诙谐文章，或居民就表格中出现的三岁祖父和上下班旅客乘坐虚无列车等误差，书面向普查官员提出的责问”。

57. 问题是要确定在多大程度上才能获得优质数据集。正如上文所指出的，巨型计算机和微型计算机的先后问世，为最终全自动进行编辑工作提供了可能。在许多国家普查/统计局，主题专家实际上十分热衷于编辑工作。因此，这些机构现正对一致性进行大量过去难以进行的检测，特别是记录间核对和住户间检查。遗憾的是，微型计算机的这一特点也造成了许多问题，过分进行编辑是其中最大的一个问题。

1. 过分进行编辑的弊端

58. 过分进行编辑给编辑工作造成了多方面的不利影响，包括准时性降低，费用增加，价值失真等。它还会使人错以为数据质量无问题。下文对这些令人关注的情况作了审查。

(a) 时间性

59. 国家普查/统计局编辑的内容越多，整个编辑过程就越长。主要的问题在于确定增加的时间能使普查产品增加多少价值。每个编辑小组都必须根据实际情况，连续评估增加时间和资源使整个普查产品取得的净效益。经常出现的情况是，虽然投入了时间，但收效甚微，因此，即使数据中出现一些小“差错”，也比不能及时为主要用户提供有关信息强。

(b) 资金

60. 同样，普查工作的费用随着时间的延长而增加。在其编辑量日益增加，内容更加复杂的情况下，各国家普查/统计局都必须确定，它是否真有必要为额外工作增加费用，以及它能否支付得起这些附加费用。

(c) 价值失真

61. 虽然编辑工作的意图在于积极改进数据质量，但编辑量的增加和编辑工作的进一步复杂化，会产生一种消极影响。有时编辑小组会改错项目，这有各种原因，如主题专家与数据处理专家之间交流有误；错综复杂的方案存在差错；或在编辑中反复处理一种普查项目等。国家普查/统计局需要尽量避免这类问题的产生。例如，Granquist和Kovar（1997年）指出，推算夫妻年龄时，不妨使用一套区分其年龄的有效方法，不过这类方法有可能人为地造成数据上的偏差，这方面的事例还是很多的。

(d) 准确无误的假象

62. 过分进行编辑，会使国家普查/统计局工作人员和其他用户误以为普查内容准确无误，在这些机构未实施和制订质量保证措施的情况下尤其如此。此外，无论编辑小组作出多大努力，普查制表中都免不了会出现不相一致的结果，因此，重要的是要告诫用户，一些小的误差有可能产生，这是一个特别实际的问题，因为许多国家都会发布抽样微观数据。国家普查/统计局不希望发布有碍规划进程的数据，因此，应特别注意确保适当编辑所有关键变量，并利用这些变量制订计划。例如，任何国家普查/统计局都不希望发布性别或年龄不详的微观数据或制表。另一方面，减少编辑量，残疾或识字变量仍会产生效用。交叉制表中会出现一些不相一致的内容，这是因为国家普查/统计局不可能编辑所有成双变量。因此，编辑小组应当检查最重要的组合因素。编辑小组发现不一致的内容时，应利用校正程序加以校正。

2. 未知数据的处理

63. 在制订普查计划的过程中，编辑小组必须提前决定处理“不明”或未知数据的办法。如上所述，表格中未知数据横竖各栏既无信息可查，亦无依据可用，因此，大多数国家的规划者更赞成对这些数据进行推算。许多用户没有对未知数据进行处理，不过它们在结果表格中列入了与已知数据相等的未知数据，以便根据实际情况推算未知数据。编辑小组需要决定如何系统处理未知数据。

3. 失真的更正

64. 国家普查/统计局在制订编辑规则时，通常不采用范例。编辑小组应制订适合实际人口或住房特点的规则。有关数据应按编辑规则进行处理。例如，一套规则可要求户主的子女至少应比户主小 15 岁。但是，户主的子女实际上可能属于社会或生物范畴的子女，即该子女可能是某配偶的亲生子女，而不是户主的亲生子女。因此，年龄有可能相差不到 15 岁。大多数国家的规划者都不是分别制订有关于子女和过继子女的计划。在此情况下，如果编辑规则改变了有关子女的年龄，就会造成教育程度、劳动力参加率及其他方面的不一致。因此，应对这项规则进行检验，以使在该规则得到充分实施前，能取得一些结果。

4. 容限的确定

65. 编辑小组必须确定每个项目的“容限水平”，有时对项目组合也需要这样处理。容限水平表示编辑小组采取补救行动前，允许存在的无效或不一致答复数量。例如，就大部分普查项目而言，一些调查对象因各种原因，不会作出“可接受的”答复，不过这类调查对象所占百分比很小。某些项目，如年龄和性别，与很多其他项目合并，用来制订计划，这类项目的容限水平可能相当低。如果遗漏或不一致的答复所占百分比很低（不到 1-2%），则任何合理的编辑规则都不可能影响数据的使用。如果这方面的百分比很高（占 5-10% 或更高，依情况而定），则进行简单的，甚至是复杂的推算，都有可能使普查结果失真。

66. 为最大限度地减少遗漏的答复，国家普查/统计局应当促使普查工作者尽全力获得实地信息。如果某个特定国家认为，某些项目如识字率或残疾率，不需要十分的精确性，则这些项目的容限水平可能会高的多。有时，编辑小组可能会通过某种办法，来纠正误差过多的项目，如让查点员返回实地，在电话中进行实地采访，或应用他们在某一领域的知识。不过，返回实地或采取其他后续措施往往费用过高，而且国家普查/统计局还有可能决定不采用该项目，或只将它同所附的慎重说明一同使用。

67. 所产生的一个问题是，应由谁来确定项目

的容限水平。可能须由编辑小组，包括主题专家和数据处理专家来决定容限水平。主题人员对有关项目必须使用很长时间，因此，他们在专业上不一定能够获得最优质的数据。数据处理专家会发现，他们实际上无法制订出合适的编辑方案，使容限降至可接受的水平，或因数据本身的原因，任何方案都不可能包含有效容限。

5. 总结编辑工作经验

68. 在对数据进行编辑时，需要记录经过详细分析的积极和消极反馈，以提高现在和今后普查或调查的质量。编辑小组必须不断作出努力，以确定哪些普查正在发挥适当的作用，哪些未发挥作用。他们还必须确定，正在发挥适当作用的普查工作是否有一些方面能够改进和精简，以使用户能够更快获得有关数据。在普查过程中，国家普查/统计局越早查明误差，就越有可能校正这些误差。

6. 质量保证

69. 质量保证对所有普查活动都是重要的，因此应当确实建立正式的质量保证机制，对计算机编辑和推算阶段的进展情况进行监测。审计跟踪、业绩评估和判断统计，这三个方面对分析编辑质量和数据处理速度至关重要(Granquist 和 Kovar, 1997 年；加拿大统计局, 1998 年)。

7. 编辑成本

70. 本手册有助于各国在完成普查或调查数据的编辑和推算方面，降低昂贵的时间和资源成本。正如 Granquist 和 Kovar (1997 年, 第 418 页) 指出的，甚至到了“1990 年代，编辑成本仍基本上与 1970 年代一样昂贵。不过，由于对技术发展的不断利用，此项工作已在很大程度上实现了合理化”。在大多数国家，编辑活动占用的时间和资金过多，因此，各国必须确定投资收益率。据 Granquist 和 Kovar 估计 (1997 年)，1990 年代初期，编辑住户调查情况资料的成本约占世界各国普查预算总额的 20%。

71. 进行面面俱到的编辑会拖延发表普查结果的时间。虽然国家普查/调查工作人员可能只掌握叙

述性的普查，Pullum、Harpham 和 Ozsewver 进行的一项调查研究（1986 年）发现，机器编辑世界融资调查，使发表调查结果的时间拖延了约一年。因此，国家普查/统计局最好首先将其资金用来进行更加有效合理的普查或调查查点工作。

8. 推算

72. 推算工作是要解决与编辑期间确定的遗漏、无效或不一致答复相关的各种问题。它的作用在于更正一项或若干项记录中的一种或多种答复，或遗漏的数值。对这些记录正在进行编辑，以确保记录的可靠性和内部一致性。与调查对象接触，或对调查表进行人工研究，可以在该过程中使某些问题提早得到解决。不过，由于答复多、成本高和时间性强，在这些初期阶段，一般不可能解决所有问题。推算工作还要处理其余编辑误差，这对编制完整统一的，包括推算数据在内的文件很有必要。编辑小组成员可充分采用微观数据，并掌握可靠的辅助信息，因此他们能够作出最精确的推算。

(a) 推算记录应与未编辑的记录大致相同。只推算极少数变量通常是最有效的方法，据此可以充分保

留调查对象提供的数据。基本假设（并非始终符合实际情况）是，调查对象更有可能出现一两个差错，而不是多个差错；

(b) 推算记录应符合所有编辑内容；

(c) 编辑小组应标明推算值，并明确确定推算方法和来源。编辑小组还应保留记录的各个方面未推算或已推算的数值，以评估推算的程度和效力。

9. 存档

73. 在普查或调查中，质量保证工作的部分内容，是编制各种程序文件，然后将这类文件存档。国家普查/统计局必须保存编辑或未编辑的数据文件，以供随后进行分析之用。有些程序如扫描，能自动保存原始图象。在键控输入大量数据后，应将这些数据并置并保存，以供可能进行的分析之用。

74. 文件编制应当充分全面，以便普查或调查规划者能够在今后对相同程序进行调整，使之与正考虑进行的普查或调查相适应。过程和结果必须具有复现性。最后，未编辑的数据与编辑的数据必须保存在若干个地方，并应采取适当措施，确保这些数据能够长时间连续使用。

二、编辑应用系统

75. 本章对编辑和推算过程应用系统进行了一般概述，其中列举了某些实例，以具体说明未编辑的数据有可能给用户带来的问题，以及编辑的数据之所以更为有用的原因。本章认为键控和编码是初步编辑过程的一部分。本章还提出了根据有效性和一致性规则进行计算机编辑的一般问题，并详细审查了各种通用的计算机编辑方法，如静态推算（非流水式）和动态推算（流水式）。

76. 编辑普查和调查资料的目的，是查找遗漏和不一致之处。推算方法用来校正这些问题。进行编辑可以确定处理遗漏和各种不可用记录的具体程序。进行推算可以更正无效的记录，并解决数据集中发现的不一致之处。此产品是一种经编辑的微观制表数据文件，其中包括可接受的全部适用数据项目统一记录，这些记录是为每个住房单元和被查点的个人编制的。

77. 重要的是，必须再次指出，无论进行多少编辑，都代替不了高质量的查点。在利用推算方法处理随机遗漏和不一致的问题时，编辑工作能够发挥有益的作用。但是，如果在收集数据期间出现系统误差，则进行编辑也无法改进数据质量，无论编辑程序多么先进。选择待调查的论题，对提高现有数据质量至关重要。在进行采访时，调查对象必须自愿并能够提供充足的信息。因此，有必要避免对某些论题进行调查，因为这些论题有可能引起恐慌、地方偏见或迷信，还有可能造成一些过于复杂和难解的问题，一般调查对象在人口普查范围内不容易作出答复。为了获得最可靠的答复，必须对每个问题采用准确的措词，当然，这些措词要依国家情况而定，并应在普查前加以检验。因此，最重要的是，国家普查/统计局应为获得更加优质的普查数据拨出充足的资源。

78. 在普查过程中，为了进行计算机编辑阶段的工作，编辑小组要制订书面编辑指令或说明、决策表、流程图以及伪编码。² 流程图有助于主题专家了解有关变量之间的各种联系，并便利书写编辑指令。

附件四中载有各种抽样和流程图。主题专家与计算机专家合作，共同编写编辑指令，并说明就每个数据项目采取的行动。编辑指令应当简明扼要，一目了然，因为这些指令是编辑程序包的依据。

79. 整个普查编辑小组成员，即各位人口学家和数据处理人员，都应当广泛进行人口数据处理和分析。在普查工作中，不合格的人员有可能无意中造成更多的误差和偏差。

A. 人工校正与自动校正

80. 对普查情况进行人工编辑，可能需要数月或数年，这很有可能造成人为误差。人工编辑难以替代计算机编辑，从某种角度讲，这是因为它不可能对人工校正过程进行编辑追踪，或对这种追踪进行调整。计算机编辑或自动编辑缩短了所需时间，并减少了人为误差的产生。计算机编辑和人工编辑都要通过审查可接受的数值，检验一项记录的效用。计算机程序还根据有关的一致性记录，查对记录的数值。最后，也是最重要的一点，自动编辑使编辑追踪成为可能，因此可以再现，而人工编辑则不能。

81. 普查和调查要收集大量数据，在此过程中，工作人员不可能始终参照原始文件校正误差。即使有现成的原始调查表，上面所载列的数据有时可能也不准确或不一致。使用计算机编辑和推算系统，可以直接校正或更正有误差的数据，并将所发现的误差和所作的全部更正编成报告。进行计算机编辑应订出周密计划，以使工作人员能有时间从事其他数据处理活动。通过计算机系统处理大量数据可能耗费时间，但又比人工校正节省时间。

82. 人工校正有多种形式。可以举一简单实例，说明性别答复中存在的误差，如一名监督人员对查点员的工作进行检查，发现了一个明显的差错，查点员将名叫“玛丽”的某个人归入了“男性”类别。监督人员在将此人改成“女性”时采用了人工编辑方法。如果监督人员不对调查表进行校正，而是将它直接寄给外地办事处，办事处工作人员也许能注意到这个问

²图7表明，伪编码是一套书面编辑指令或说明。

题，并进行人工校正。在中心办公室，编码员在编码时，可能会发现姓名与性别不一致，从而进行人工校正。编码员也许发现不了这个问题，但在键控员将数据输入调查表时，他们就会注意到姓名与性别的不一致，并在输入数据之前进行人工校正。

83. 但是，如果编码员未注意到这类误差，而且键控员输入了“男性”编码，就可以针对这种情况采取一些不同的程序。对于与性别有关的项目如生育信息组，编辑程序可标明，此男性的资料是女性的。并在键控员输入数据时，编制有关这一内容的信息。这样，键控员通过查看调查表，就可以发现这实际上是名女性，并进行人工校正。另外，如果国家普查/统计局使用不带键控功能的编辑程序，计算机程序就能标明，此人是男性，但情况介绍是女性的。其次，办公室工作人员使用地理信息系统，可以在收集器中找到原始调查表，将此表调出，并确定名叫“玛丽”的调查对象被误报成了男性，而不是女性。在此基础上，办公室工作人员可将此资料交还键控员，由其调出该记录，并进行人工校正。

84. 这一实例表明了人工编辑的利弊。在采取以上概述的任何步骤时，普查人员都会注意到姓名与性别的不一致，并会进行校正。但是，使用人工编辑方法的普查/统计局可能在每个阶段都需要工作人员对这种关系进行检查。此项活动耗费了大量精力，如果指示工作人员不进行人工编辑，结果或许没有什么不同，特别是在汇总方面。

85. 直到最近，对数据集进行校正仍主要采取人工更改办法。许多国家依然认为使用自动校正方法不方便，因此，它们在上述其中一个阶段还是使用人工校正办法。如果是小型数据集，而且时间安排不紧要，或劳动力为密集型，一般可以进行人工校正。这样做的好处是，如果调查表上的信息完整精确，而且不一致之处可以通过检查有关表格得以实际解决，则普查或调查基本上不存在质量问题（编辑小组必须假设，例如“玛丽”不是“加利”，如果表格上列有生育一项，实际上可以推测出，有关数据就是为此人收集的，没有出现误差）。事实上，编辑和推算程序一般不会改进数据收集质量。它们只能更正某些内容。

86. 有时为进行人工校正查阅调查表会一无所获。由于各种原因，调查表中没有需要的信息。有的人往往不愿透露自己的年龄，因此，不会在调查表上填写这一项。在此情况下，检查调查表解决不了问题。对此，编辑小组必须决定如何处理这种情况。如果是使用人工校正，国家普查/统计局必须选定“未知项”，或利用某些数值选定年龄项。

87. 如果不与调查对象接触，人工校正难免会降低质量和一致性，并会耗费更多的时间和费用。计算机不存在疲劳问题，因此速度更快。计算机也没有人所面临的问题，因此不会影响质量或一致性的保持，而且在大多数情况下，计算机会使数据处理费用更加低廉。大多数国家现都采用了某种自动校正技术。

88. 答复缺漏或不一致，会降低数据质量，并给提供易于理解的普查数据造成了困难。一些用户主张将缺漏和不一致的答复作为“未报”情况编列成表。还有些用户主张将这些情况均衡地列入所报告的统一项目中。另有一些用户建议利用“类似”答复推算规则，解决答复缺漏和不一致的问题。利用计算机，可以根据调查表中的其他信息，或根据具有类似特点的个人或住户上报的情况，对答复进行灵活而有效的推算。

89. 计算机能够检查多种指数，编辑加工应利用这一性能。就此而言，编辑程序因涉及多种有关指数，因此可以对各种答复进行更加合理的推算，这是简单的编辑方法所办不到的。另一方面，设计不合理的编辑程序有可能产生不精确的普查数据。编辑小组应由各有关学科有经验的主题专家和数据处理员组成。在对一致性进行验证时，编辑小组应仔细选择要审查的各种变量，以确定编辑和推算标准。程序输出应包括要改变或推算的答复百分比。这样，分析人员就能够更准确地判断数据质量。例如，推算出的百分比高，就是告诫用户，使用数据时要慎重。

90. 编辑或审计追踪装置显示了对每个变量进行的更正。该装置用来在所获取的编辑和推算加工数据中，跟踪各种答复的来龙去脉。

B. 数据校正准则

91. 无论是人工编辑，还是自动编辑，都应尽

量反映实际情况，消除缺漏和无效记录，并更改不一致的项目。

方框 2. 校正数据要则

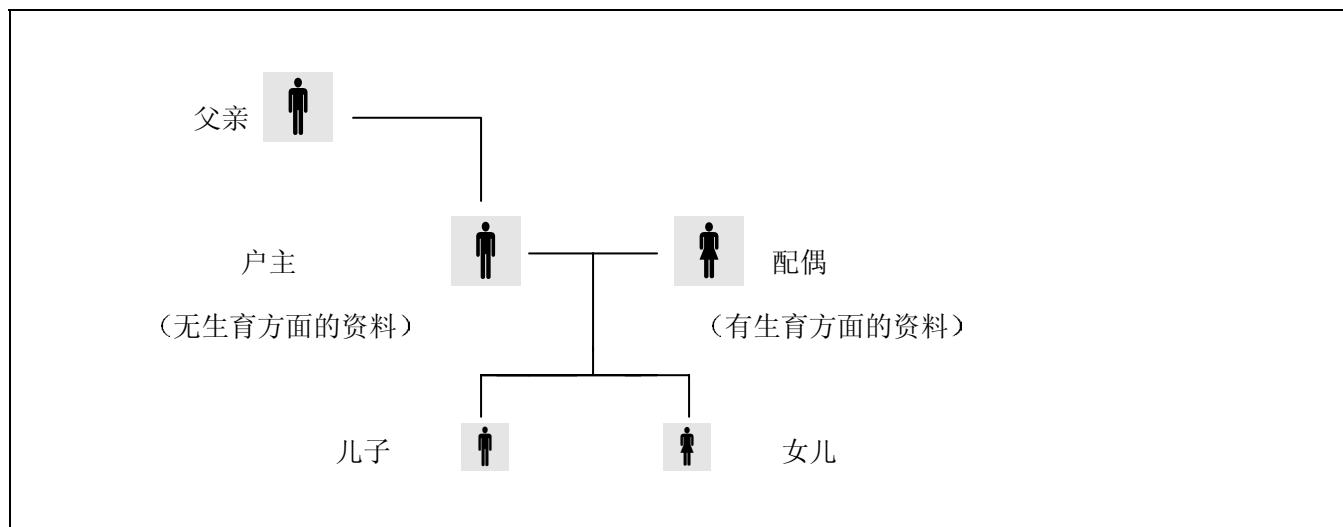
在确定编辑程序后，最好铭记如下数据校正建议：

- (1) 尽量少改动最初记录的数据；
- (2) 消除项目中明显的不一致；
- (3) 提供有误或缺漏项目记录，这方面可以参照其他有关住房单元、个人或住户和类似群体中其他人的记录，但始终要依据具体的程序。在某些情况下，“未报”一项适用于某些项目。

92. 请参看如下特定住户图表（图 1）。此图表明住户的各种关系和性别记录一致。户主是男性，

无生育方面的资料；配偶是女性，有生育方面的相关资料。

图 1. 假设的典型住户，包括住户成员之间的关系、性别和生育情况



93. 但是，在许多情况下，有关资料并不统一，这就造成了如下一些问题：对不一致的住户登记应进行哪些编辑加工？如果按图 2 所显示的，户主和配偶都报成了男性，编辑小组应如何进行编辑？按过去的典型编辑规则，配偶第一方被设想为男性，特别是在此人为户主的情况下，配偶第二方被视为女性。

94. 在此情况下，如果户主恰巧是妻子，而不是丈夫，则所采用的编辑规则就不适用了，而且国家

普查/统计局最终会遇到如下四种误差：

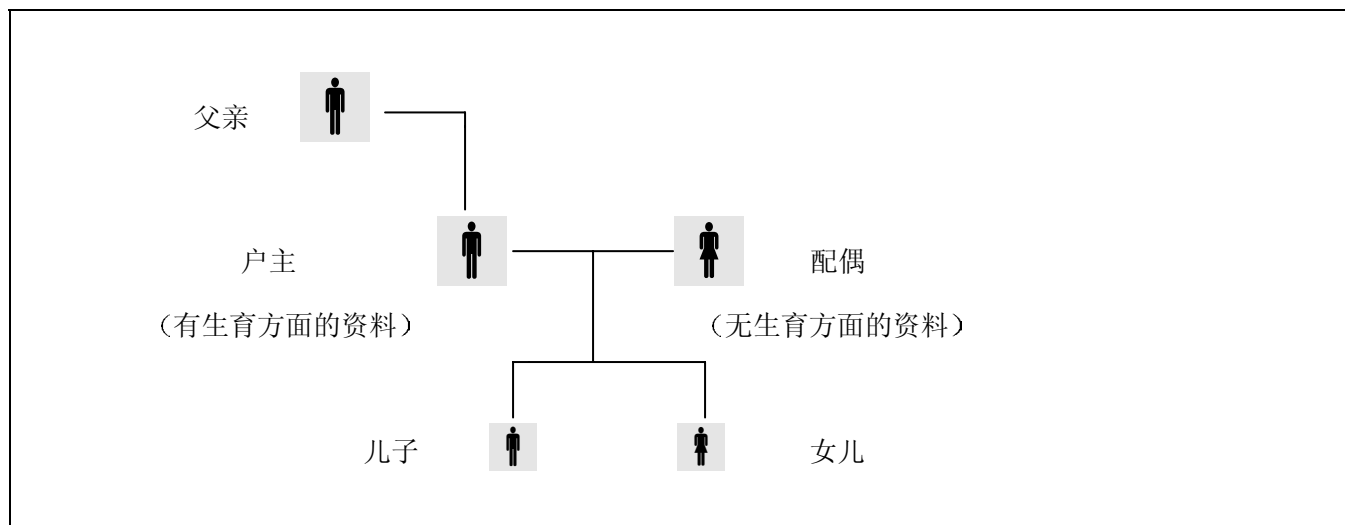
- (a) 户主的性别不对；
- (b) 配偶的性别不对；
- (c) 户主生育方面的资料漏失；
- (d) 错将生育方面的资料归入男性配偶栏内。

这种编辑程序显然不适用。

95. 相比之下，有效的编辑程序能够发现户主和配偶出现同一性别的情况，并会核查双方生育方面

的资料。由于只有户主有这方面资料，因此户主是女性。在此情况下，就可以适用这些项目的编辑规则。

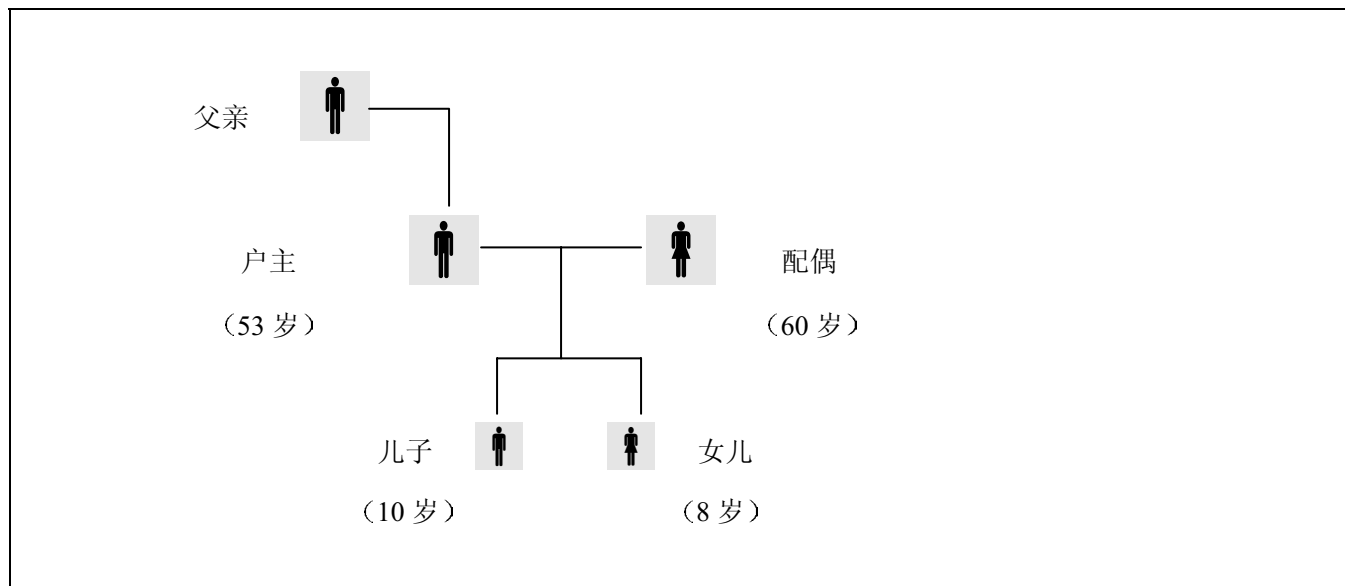
图 2. 户主和配偶出现同一性别的住户实例



96. 图 3 中例举的另一个实例也表明了这一点。大多数国家都认为生育年龄在 15 至 49 岁之间。假设一名妇女报告说她生育子女的年龄是在 52 岁，根据表明子女母亲年龄或计算年龄差别的直接证明，母亲与子女之间的年龄差不能超过 50 岁。编辑小组必须确定，这种年龄差是否能够接

受，或是否应加修改，并通过编辑，替代其中的一个或另一个年龄。如果通过编辑，扩大了可接受的生育子女年龄范围，而且有关妇女报告的育龄较大，那么在年龄本身被错报的情况下，就会有更多的反常情况被输入数据集。编辑小组必须确定所报告的年龄是否符合特定的变量。

图 3. 住户某些成员出现同一年龄的实例



97. 图 4 提供了另一种可能的情况。假设通过编辑，在图 4 中发现一名 70 岁的妇女有两个分别为 10 岁和 8 岁的子女。这种情况或许存在，因为丈夫可能有前妻生的子女。在此情况下，子女与户主有关，但与丈夫的配偶本人无关。即使很有可能出现如下情况，即键控员本想键入“4”，来取代 40，但却错键入了“7”。设想在母亲与子女的年龄相差 50 多岁的情况下，出于各种理由，主题专家要求数据处理员更改母亲或子女的年龄。由于这一要求，需要进行另一项更加复杂的编辑工作。在有关妇女为 70 岁，其所生的第一个子女为 10 岁的情况下，编辑小组必须决定更改谁的年龄。该小组有可能将第一个子女的年龄改成 20 岁，这可以解决第一个子女的年龄问题，并可更改配偶的年龄。但第二个子女的年龄问题依然存在，而且也需要进行编辑。

98. 在只考虑母亲和一个子女的年龄时，推算

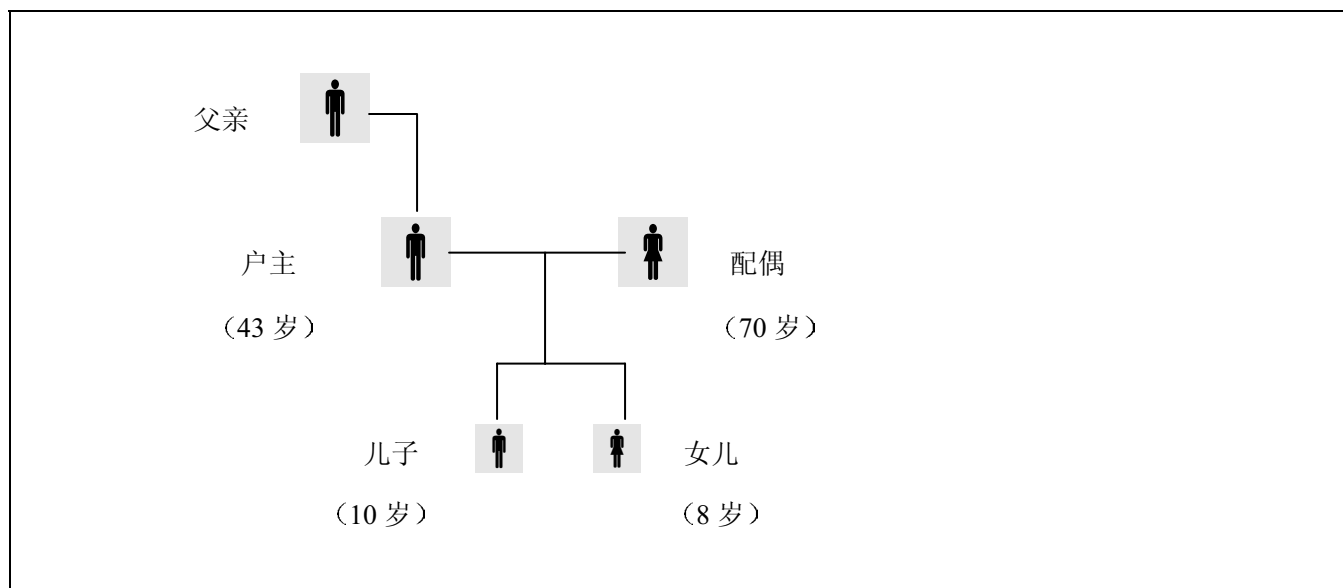
会随机选定年龄，而且正确率占 50%。但是，如果进行编辑还要核对丈夫的年龄，编辑小组很可能会根据这类补充资料更改配偶的年龄。这种更改使整个家庭的年龄更加相符。

C. 有效性和一致性检查

99. 编辑工作的一项主要要求是，任何项目中都不应出现无效的数值。此外，记录上对各种有关项目的答复必须相一致。

100. 推算应尽可能同时考虑全部有关的变量资料，而没有必要接二连三地进行。不过，在某些情况下，编辑工作可先对一致性进行核查，然后再确定条目的有效性。如果推算根据对一致性的核查选定一个数值，则必须将此数值与原条目进行比较，以确定是否有实际变化。如果没有，仍旧可以保留原输入条目。

图 4. 所报年龄可能有出入的住户实例



101. 例如，在对婚姻状况进行编辑期间，首先要对婚姻关系进行检查，以确定记录的是否为配偶，如果是，而且未报告配偶已婚，则要把婚姻状况写成“已婚”。在输入已婚编码时，程序系统会进行检查，以确定原始调查对象过去的情况。如果输入的是已婚编码，程序将不会更改记录，因而不会出现误差。

1. 自顶向下编辑法

102. 这种过程先从第一项开始编辑（“顶端项”），它通常是调查表上的第一个变量，然后顺序逐项编辑，直至完成所有项目。自顶向下法不保存数据之间的关系，但它确为完成编辑提供一个适当的框架。

103. 在编辑过程中，有些编辑不止一次变动某项的值。这种步骤会给数据集引入一个或多个错误。推算值与其他数据可能不一致。甚至在按顺序处理变量时，在可能的情况下也应同时参照所有其他的变量编辑某个特定的变量。例如，依据母亲年龄推算的子女年龄，就可能与子女所报的上学年数或在该地区居住的年数相矛盾。在这种情况下，将重新推算年龄直至不矛盾为止。推算年龄是一个中间变量，直至最后赋值。在编辑过程中，推算的中间变量不应记录为变化，直至最后赋值为止。

104. 虽然就少数数据项和情况而言，编辑程序可以接受空白或“未报”条目，但相关的信息能够为大多数空白或登记错误的数据项提供登录值。以这种方式提供的登录值不一定个个都正确。但是计算机具有比较不同存储值的巨大能力和速度，因而可以确定合理说明情况的置换登录值。有时候，多数情况下产生的列表比来自未经编辑的记录或估算的列表更一致，后一种列表将所有不能接受的登录项都转换为“未报”项。

105. 编辑程序还必须进行结构检验（见第三章）。编辑应当检验人口项目（见第四章）和住房项目（见第五章）。此外，如附件一所说明，编辑过程也许还应在制表所需的个人记录上生成一个或数个重新编码的变量。

106. 极端重要的是应避免循环编辑——变动一个或数个数据项，然后在某个晚些时点又以原来的方法将它们变回去。本手册在其他地方指出，工作人员必须进行数次运行以确保对所有项目全都进行编辑。可以建立在首次运行期间改变数据的编辑标准，但在第2次运行期间适用于变动后的数据时，它们又

将该数据变回到原始的配置。这一过程可以连续经历多次运行。编辑小组应避免在编辑过程中采用这种标准。

2. 多变量编辑法

107. “自顶向下”的普查和调查编辑法在第1节中作了介绍，它不一定总能给出最好的结果——最接近变量实际分布的结果。如上所述，自顶向下法使用时没有恰当的防范规定，在编辑时经常引起问题。

108. 另一种方法是多变量法，它以 Fellegi-Holt 系统为基础。这种方法要求具备更多的计算专门知识和更强大的计算机能力，但获得的结果也许更接近“现实”。不同类型的多变量编辑方法载于附件五“推算方法”。在多变量编辑系统中，需要确定一组肯定语句和变量之间的关系。然后，编辑参照住户中的数据测试每个语句，以弄清所有语句是否都是真实的。对于任何虚假的语句，编辑将逐项地跟踪无效登录项或矛盾项。经过所有测试后，编辑和推算系统必须评估如何能够最可靠地改变记录以便通过所有编辑。编辑小组通常使用最低限度变动法，尽量少改动变量数目以获得可接受的记录。

109. 图5中的11个说明语句提供一个实例，说明在选定人口特征多变量编辑中可以适用的规则。在此例中，户主必须为15岁或更大年龄。就广义编辑而言，较好的做法将是利用“×”岁，其中×是该国确定的最低年龄。例中的各个语句如关系、性别、年龄、婚姻状况和生育率等，专注于其他重要的初始变量。各变量密切相关，因此编辑小组应当统观它们以便最有效编辑数据。这里还应指出，虽然所有的变量都重要，但对于数据来说，有些变量较之其他变量更为关键。

图 5. 选定人口特征多变量编辑规则的实例

编号	规 则	关系	性别	年龄	婚姻状况	生育率
1	户主应为 15 岁或更大					
2	配偶应为 15 岁或更大					
3	“配偶”应已婚					
4	如有配偶，户主应已婚					
5	如有配偶，户主和配偶的性别应相反	1	1			
6	15 岁以下人应未婚					
7	男性应无生育率		1			1
8	15 岁以下女性应无生育率					
9	15 岁及以上女性生育率登录不应为空白					
10	“子女”应比户主年轻					
11	“父母”应比户主年大					
	合 计	1	2			1

110. 在图 6 的实例中，配偶双方来自与图 5 相同的人口。双方都报为男性。这里的编辑过程简单和直观。计算错误项数最大的变量是将首先编辑的变量。在图 6 中，编辑程序执行“性别”的推算过程，因为根据图 5 的数据，该变量相对于（1）关系和性

别和（2）生育率和性别错误最多。当编辑程序检验生育率并发现户主具有生育率信息但配偶不具有时，推算将“女性”分配给户主。最后，在编辑小组重新检验计算系列而且所有肯定语句都真实时，就不需要作进一步的编辑。

图 6. 未经编辑数据集中户主和配偶性别相同及其解决的实例

人	关系	性别	生育的子女
<u>未经编辑的数据</u>			
1	户主	男性	03
2	配偶	男性	空白

<u>性别经编辑后的数据</u>			
1	户主	女性	03
2	配偶	男性	空白

111. 关于这种编辑的编辑说明可以写成如图 7 所示。如果双方的生育率完整，编辑将行得通。然而

编辑显然是不完整的，因为它只照顾到户主和配偶双方的生育率都完全和准确的情况。

图 7. 采用伪代码方式，更正性别变量的编辑说明实例

```

如果户主性别=配偶性别
  如果户主生育率不是空白
    如果配偶生育率为空白
      (如果户主的性别不是女性) 使性别=女性
      (如果配偶的性别不是男性) 使性别=男性
    或者 另作处理，因为他们性别相同且都有生育率!!!
  结束-如果
或者 这属于户主生育率为空白的情况
  如果配偶的生育率不空白
    (如果户主的性别不是男性) 使性别=男性
    (如果配偶的性别不是女性) 使性别=女性
  或者 另作处理，因为双方都无生育率!!!
  结束-如果
结束-如果
    
```

112. 下图（图 8）为一实例，其中编辑过程考虑一个 13 岁的女户主，按键入的信息称，她已守寡，但有 3 名子女。在程序通过编辑规则时，产生下列结果：

图 8. 有 3 名子女的极年轻寡妇多变量编辑分析的实例

编号	规 则	关系	性别	年龄	婚姻状况	生育率
1	户主应为 15 岁或更大	1		1		
2	配偶应为 15 岁或更大					
3	“配偶”应已婚					
4	如果配偶存在，户主应为已婚				1	
5	如果配偶存在，户主和配偶的性别应相反	1				
6	15 岁以下的人不应已婚			1	1	
7	男子不应有生育率					1
8	15 岁以下的女性不应无生育率		1	1		
9	15 岁或以上女性的生育率登记项不应空白					
10	“子女”应年轻于户主					
11	“父母”应年长于户主					
	合 计	2	1	3	2	1

113. 根据肯定语句系列，年龄变量多数有错，它是应首先改动的项目。在年龄变动时，如果解决所有矛盾，编辑将结束。否则，程序编辑不一致数目其次多的变量。

D. 编辑考虑

114. 当国家普查/统计局为编辑程序和为随后的制表编制代码表时，它们不妨为部分项目确立共同的代码。例如在许多国家，地点代码（出生地、父母出生地、以前居住地、工作地点）、语言、民族/种族和公民资格等非常类似。可将“地点”的共同编码计划确定为三位数代码，第一位表示洲，第二位表示区域，第三位表示具体国家。国家普查/统计局还可利用国际组织如联合国统计司开发的国别数字代码（联合国，1999年）。密切相关变量的一组共同代码能够减少编辑错误，并在编辑过程中对数据处理程序有所帮助。共同的代码也允许数据处理程序视情况将登录项从一个项目调至另一项。

115. 编码表的结构能够便利编辑进程及稍后

编辑、制表和分析期间的处理。对于移民或民族群体众多的大国而言，代码以洲、区域和国家为基础，各自分配不同的代码或位数，这将比简单的列举可取。

116. 图9给出实例说明诸如出生地、公民资格、语言和民族等项目共同的代码。就菲律宾人而言，说伊洛卡诺语和他加禄语的人的代码不同于菲律宾通用语言的代码。依据特定国家的具体情况，这些代码也可彼此不同。虽然英语有单一的代码，但说英语的不止一个群体。因此，加拿大和美国的出生地、公民资格和民族的代码略有不同。就法国出生、拥有法国公民资格、说法语和具有法国民族渊源的人而言，使用相同的代码。因此，其中一个项目遗失，而且如果编辑小组决定这种解决方案适当，数据处理程序能够从其他登录项之一移来代码。

117. 如果调查表上的一组项目并不相互独立，国家普查/调查人员不应要求提供全部这些数据。编辑小组必须视情况个别地决定何时将项目直接用于分配，以及何时利用其他可获的变量。

图9. 选定项目共同代码的实例

群 体	出生地	公民资格	语言	民族渊源
法国/法语	10	10	10	10
西班牙/西班牙语	20	20	20	20
拉丁美洲	25	25	20	25
菲律宾/菲律宾语	30	30	30	30
伊洛卡诺语			32	
他加禄语			32	
英格兰/英语	40	40	40	40
加拿大	50	50	40	50
美国	52	52	40	52

118. 不同普查之间（或普查与调查之间）诸如工作或民族渊源等变量的定义不同，还产生另一个问题。无论对于目前编辑的数据还是对于来自以前普查的数据集，国家普查/统计局必须决定如何将这变化考虑进去，以便显示趋势。如果能够利用原始的未经编辑的数据，数据处理程序能够对有关的编辑作出

变动并重新处理这些数据。

119. 例如，当只确定少数情况时，某个欧洲国家可能使用单一代码代表所有南亚国家的原籍国。然而，由于移民格局发生变化，下次普查或调查在处理的全过程可能需要单独的代码代表印度、孟加拉国、巴基斯坦、斯里兰卡和其他南亚国家。

E. 改正和推算数据的方法

120. 如上所述，在所有的普查和调查中，都出现“未报”、“不详”或其他信息遗漏造成的数据记录的空白。答卷人、查点员或数据输入错误也造成无效的登录项。进行改正的方法各不一样，需视项目而定。在多数情况下，能给数据项分配有效的代码，并有相当的把握保证正确，其方法是利用来自个人或住户记录内其他数据项的答案，或来自其他住户或人员的记录的答案。

121. 本手册介绍两种校正错误数据的计算机技术。一种是静态推算或“非流水”法，主要用于遗漏或不详项目。另一种是动态推算或“流水”³法，主要用于遗漏的数据及用于不一致或无效的项目。

1. 静态推算或“非流水”技术

122. 采用静态或非流水推算法时，编辑程序从某个预定的数据集给一个遗漏的项目分配特定的答案，或按比例根据有效答案的分布推算一个答案。采用非流水法时，程序并不更新原始变量集。在处理第一人、第二人、第十人或任何其他人的记录后，这些值与初始静态矩阵中的值相比并无变化。利用原始数据可估算任何遗漏数据。

123. 静态推算是一种随机法，与动态推算一样，但值并不随时间发生变化。这一方法在附件五中说明。

124. 有时静态推算采用比率法，依据预定的比例分配答案。作为答案按比例分布的一个实例，假设一张关于从事农业活动的 33 岁男子工作时间的有效数据表，即数据来自填全而不是遗漏的项目，该表显示 25%的人每周工作 50 小时；40%的人每周工作 60 小时；而 35%的人每周工作 70 小时。不过，除非可从以前的普查、调查或其他来源获得可靠的数据，这项技术要求对来自当前普查的有效答案预列表，这从经济上或业务上看或许不一定可行。

³联合国统计司从大约 130 个国家和地区收到的调查结果表明，用语“流水”和“非流水”不为人们普遍所知。然而有些国家已经适用这些方法编辑数据。

2. 动态推算或“流水”技术

125. 消除数据不详项的另一个办法是动态估算或流水技术，将它用来给不能获得、不详、不正确或不一致的登录项分配数值。美国普查局最早开发了这项方法，但自那时以来其他机构进一步完善了它。当数据集中出现未知项（或在某些情况下多个未知项）时，动态推算法使用一个或多个变量估计可能的答案。普查编辑越来越多地利用动态推算法，因为它容易，而且产生清楚和可复制的结果。除此之外，通过消除未知项，更容易取得普查与调查之间的趋势，因为分析人员不需要逐项地处理未知项。

126. 就动态推算而言，当某人的某些信息不详时，具有类似特征的其他个人的已知数据便可确定可加利用的最合适的信息。这些特征包括性别、年龄、与户主关系、经济状况和教育程度。推算矩阵本身为一组数值，类似于卡片叠中的卡片。这些矩阵存储数据，当遇到未知数时提供利用的信息。通过更新和（或）按逻辑“组合卡片叠”，卡片叠不断变化，这样在数据处理过程中答案估算发生变化；因此使用术语“流水”。

127. 作为一个简单的实例，一个单一的值可作为卡片叠存储。例如，如果由于某个原因一人的性别无效，就任意给卡片叠分配一个初始值（男性或女性），这样确定一个初始值。种子值成为遇到不明性别的第一个人的性别。不过，如果第一人的性别有效，第一人的性别置换种子值。如果第二人的性别不详，那么推算矩阵分配存储的性别。在这种情况下，推算的性别是第一人的性别。实质上，当编辑发现一个项目可接受的值时，它将值输入推算矩阵。在它发现不可接受的值时，推算用推算矩阵的有效值置换它。

128. 这里说明的动态推算（流水）法的问题之一是，如果两个不同的项目有未知值，不得使用同一“提供者”个人分配有效答案。每个值可来自一个“实际的”人，但这些可能是不同的人。更好的办法将是同时分配两个变量，它们来自同一人。不过，这些复杂矩阵的程序编制可能造成某些困难。

129. 下文数据（图 10）说明一个由 10 个人组成的住户。代表性别和年龄的数目 9 和 99 表示遗

漏的信息。虽然其他的变量可用于推算，例如教育和职业，但未将它们列入这一简短的实例。

图 10. 作为动态推算输入实例的抽样住户

标识符 编号	关系	性别	年龄
1	1	1	39
2	2	2	35
3	3	1	13
4	3	9	10
5	4	2	40
6	4	1	99
7	4	2	13
8	5	9	99
9	5	1	44
10	5	2	36

注：9、99 = 遗漏的信息

130. 如果叫作性别阵列 (SEXARRAY) 的推算矩阵的初始值是男性 (代码=1)，推算矩阵将显示为这样：SEX=1

131. 人 1 处理后，值仍将为 1。不过，处理第二人后，值将变为 2，因为此人是女性。变量现将显示为：SEXARRAY=2

132. 对于一个处理个人的性别的每项有效登录，该人性别的代码置换估算矩阵值。在处理第三人时，估算再次将值变为 1 或男性。

133. 第四人的性别不详，所以编辑查找推算矩阵值，在这一情况下它是男性，编辑用推算矩阵值置换未知值。人 5 是女性，所以它置换推算矩阵中以前来自人 3 (男性) 的值。这个过程继续至人 8 为止。

134. 编辑再次利用推算，而且人 8 变为女性，因为从人 7 取得的推算矩阵值是女性。编辑两次使用推算矩阵取得值：一次获得男性和一次获得女性。由于性别大致以相等的频率出现，经过长时间的运行，

估算大约以一半的时间使用每个性别。经对所有 10 人处理后，变量将显示为：SEXARRAY=2

135. 虽然推算矩阵以这种方式分配值，但也存在利用此种过程的其他更复杂的方式。例如，编辑程序能够利用与户主的关系和性别以协助确定某个个人的年龄。考虑下列关系代码的部分清单：

- 1 = 户主
- 2 = 配偶
- 3 = 子女
- 4 = 其他亲属
- 5 = 非亲属

136. 数据处理人员能够生成可能接近按年龄列示的关系的实际情况的初始年龄值。这些值并不非常重要，因为在使用它们前编辑将几乎肯定置换它们。还有，编辑要求推算许多值，所以影响最后制表的只是很少数的初始值。这些值可如图 11 所示。

图 11. 基于性别和关系的年龄初始静态矩阵

	关 系				
	户主 (1)	配偶 (2)	儿子/女儿 (3)	其他亲属 (4)	非亲属 (5)
男 (1)	35	35	12	40	40
女 (2)	32	32	12	37	37

137. 再考虑图 10 介绍的 10 人。由于我们抽样中的第一人列为户主 (代码=1) 而且他是男性 (代码=1), 推算期间他的年龄 (39) 置换第一个元素 (坐标 1, 1)。然后卡片叠载有图 12 显示的值。

图 12. 一项变动后动态推算矩阵的实例

	关 系				
	户主 (1)	配偶 (2)	儿子/女儿 (3)	其他亲属 (4)	非亲属 (5)
男 (1)	39*	35	12	40	40
女 (2)	32	32	12	37	37

138. 第二人是配偶 (代码=2) 和女性 (代码=2), 所以她的年龄 (35) 置换第二纵列第二横行中的值, 将卡片叠改至这些值。住户其他个人的年龄类似地置换推算矩阵值, 直至第五人。

139. 注意以前的性别推算过程将性别 1 分配给人 4。由于编辑要求推算性别的值, 编辑并不以该人的年龄更新阵列。编辑将只以来自记录的值更新, 其中性别和关系在初始时都是正确的。不过, 当编辑进入到人 6 时, 它发现年龄是未知数。该人是男性而且他是户主的一名“其他亲属”。因此, 编辑将推算矩阵元素用于其关系组是“其他亲属” (第一横行第四纵列) 的男性, 并分配该类年龄的值 (“男性其他亲属”——在本例中是 40)。

140. 第八人的性别和年龄均未报。编辑将性别推算为女性, 然后根据这种分析的年龄和关系代码 (5) 分配年龄。本例中年龄为 37。

141. 虽然编辑根据已知关系推算年龄值, 但它将以前分配的性别值用于其他变量。这里, 将分配的值用于进一步的估算是糟糕编辑过程的一个实例 (见下文 3(d) 节)。更好的办法将是寻找其他已知数据项如婚姻状况用于推算。

142. 在第十人次之后, 图 13 给出推算矩阵的值。在本例中, 两种推算都使用初始静态矩阵。通常将只有少量 (如有的话) 的初始值用于推算。多数情况将使用从查点人口分配的值。

图 13. 多次变动后动态推算矩阵的实例

	关 系				
	户主 (1)	配偶 (2)	儿子/女儿 (3)	其他亲属 (4)	非亲属 (5)
男 (1)	39	35	13	40	44
女 (2)	32	35	12	13	36

3. 动态推算（流水）问题

(a) 地理考虑

143. 如果编辑程序使用动态推算方法估算遗漏的值，它应尝试使用按最小的地理界定的区域分类的数据。这一过程应提高获得正确答案的概率，因为居住在同一小地域的人的人口特点和其他特点通常有些相似。如果人口不同质，将不存在相关性，所以编辑小组必须逐例查找变量。

(b) 相关项的利用

144. 在使用动态推算获得遗漏值以前，应努力利用相关项分配可能是正确的值。例如，如果某人婚姻状况遗漏，编辑程序将确定此人在住户中是否有配偶。如果有，程序将分配已婚代码而不使用估算矩阵。不过，如无此种证据存在，程序可能不得不依靠推算矩阵值。

(c) 变量的次序如何影响矩阵

145. 利用推算矩阵的国家普查/统计局应当考虑在它们确定编辑秩序时需要哪些变量。就人口项目而言，统计局将希望一开始先编辑性别和年龄，以便将这些用于推算矩阵。总体编辑不应将未经编辑的变量用于推算矩阵，尽管多数计算机程序包将接受“未知”横行或纵列。后继编辑矩阵可使用编辑后的数据阵。不过，凡有可能，统计局应考虑排除来自推算矩阵经过编辑的数据。

146. 例如，如果编辑根据性别和关系推算年龄，这一推算矩阵阵列中的各个单元（性别×关系），如果性别或年龄是推算的，就不应更新。通常，只有在年龄、性别和关系全都有效且一致时，编辑程序包才应将年龄登入相关性别和关系的单元。不过，由于其他因素，有时候难免要使用经编辑的数据。

(d) 推算矩阵的复杂性

147. 通过使推算矩阵变得较为详细，国家普查/统计局提高了获得一致的，“正确”推算矩阵值的概率。例如，程序能够只利用关系推算婚姻状况。不过，孀居或离婚的可能性随年龄增加。因此，凭年龄和关系推算婚姻状况是讲得通的。编辑程序利用目前

的人的年龄和关系，从存储在推算矩阵中的上一个有效记录，取具有相同特点的一个人的婚姻状况的值。

148. 尽管如此，上述过程可能产生新的问题。国家普查/统计局通常以固定顺序编辑调查表项目，采用自顶向下的方法将年龄排在婚姻状况之后编辑。如属此种情况，在记录中遗漏婚姻状况和年龄二者时，就不可能从上个具有相同年龄和关系值的记录取婚姻状况的值。因此，程序就不一定能确定这个记录的年龄类别。另一个解决方案将是让推算阵列设有一个横行或纵列用于“未报”项。这个过程将允许程序利用上个具有相同关系和“未报”年龄的记录的婚姻状况类别分配婚姻状况的值。不过，有两个因素不支持这种做法。一是同一组合中“未报”情况极少，因此将难以更新遗漏项的估算阵列。其次，实质上不可能获得这些“未知”值组合的恰当的非流水值，因为在“实际”世界上它们并不存在。

149. 上述问题的解决方案给数据处理人员产生更多的工作，但可产生统计上更明确的结果。编辑程序首先测试以确定项目是否具有有效的代码。如果目前的人的记录不拥有该项目有效的代码，推算矩阵不将该项目用于记录。数据处理器通过建立更简单的推算阵列能够便利该进程。现继续早先的实例，如果因为值遗漏程序必须推算婚姻状况，推算阵列一般将有两个维数：年龄和关系。如果经测试后程序未找到年龄的有效代码，它将会凭关系推算婚姻状况。由于对关系的编辑先于对婚姻状况的编辑，关系代码将为有效。程序将这些相同的原则用于所有动态推算过程。

(e) 推算矩阵的发展

150. 主题工作人员与数据处理人员协作，也可编制适当的推算矩阵（有些编辑小组利用多个推算矩阵）。只有有效的答案才更新推算矩阵；编辑小组不使用分配或推算的值。主题专家和数据处理人员双方必须检验编辑说明和流水法以便保持一致和完整。

151. 应当投入大量的时间和思考以发展推算矩阵，包括研究利用行政记录和以前普查或调查的效果。即使经过研究和发展后，编辑人员仍不应随机适

用推算矩阵。在推算矩阵内部不一致时，需要作出很大努力使它们保持一致。在推算矩阵不利用标准约定时，工作人员必须分别考虑每个约定。

152. 虽然就本手册的例子而言，推算矩阵的每个单元都有一个值，但有些编辑小组为每个单元保留一个以上的可能性。这些单元提供一个额外维数。现举例说明，如果一个家庭所有子女的年龄均不知，例如，对一个有 4 个男孩的家庭，计算机将不会同个值分配 4 次，产生四胞胎。相反，它将分配 4 个不同的年龄。不过，即使这里，也可能同一值分配一次以上，这将取决于矩阵中存储的内容。

(f) 标准化的推算矩阵

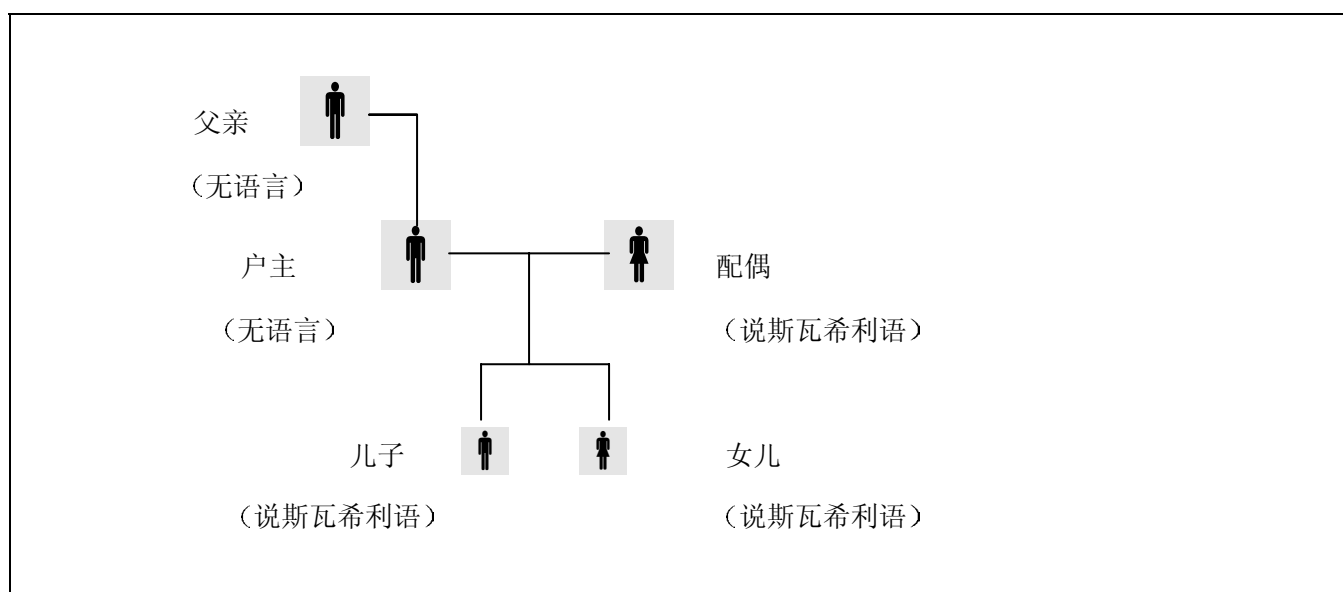
153. 标准化的推算矩阵能够简化编辑进程。能够快速测试和适用各种社会和经济变量如年龄组和

性别的带标准维数的估算矩阵。

154. 例如，国家普查/统计局可能希望发展推算矩阵，以在未给出语言代码时确定这种代码。编辑程序寻找的第一个地方几乎肯定是住户内报告说某一给定语言的另一个人。如果未能找到，该程序能够选择同一性别和年龄组的前一人的语言（在所有三项有效时已更新了推算矩阵）。这个过程将给出一种可能的语言，因为说同一或类似语言的人通常地理位置相互靠近。

155. 图 14 中的变量“语言”未载列户主的任何信息。不管什么原因，光扫描器或键控器可能未拾取到语言输入或代码，或可能出了其他故障。不过，由于配偶和子女全都说斯瓦希利语，可能把该语言分配给户主和户主的父亲，后者的语言登录项也遗漏。

图 14. 户主和户主父亲未分配语言情况下的实例



156. 在住户任何人均未报语言时，编辑程序必须另想办法。首先，编辑应查找其他变量间接估计使用的语言。有时种族、民族或出生地能表明可适当推

算的语言。如果此种标识项能加以利用，编辑小组可选择用它来确定户主的语言。如果不能，编辑可利用年龄和性别来计算。推算矩阵可能如图 15 所示。

图 15. 语言动态推算矩阵初始值

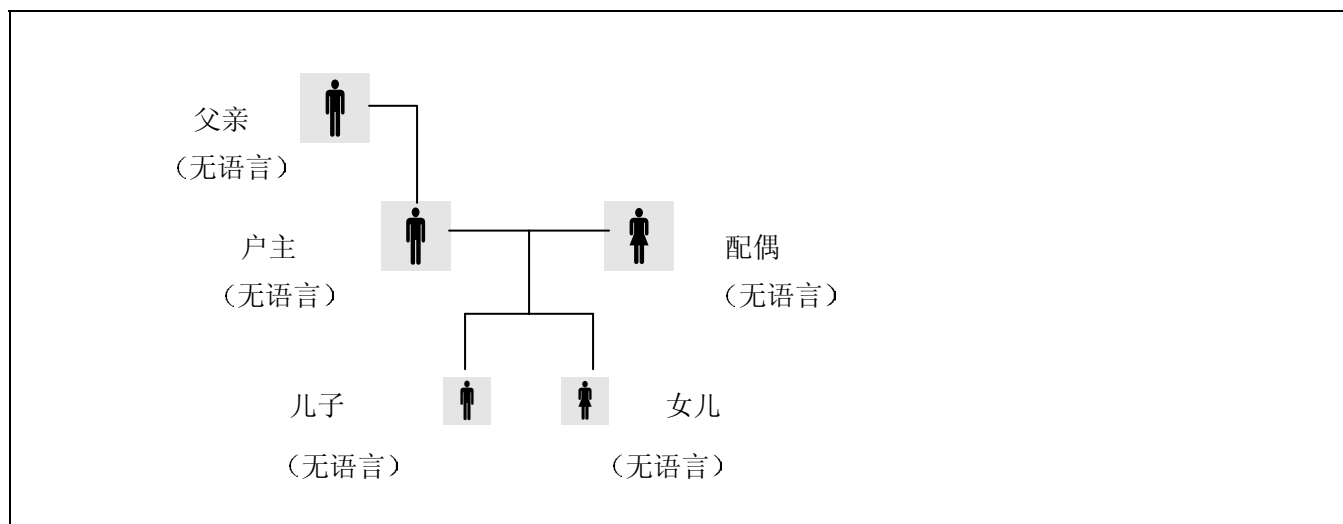
性别	年 龄					
	15 岁以下	15-29 岁	30-44 岁	45-59 岁	60-74 岁	75 岁及以上
男	语言 1	语言 1	语言 1	语言 1	语言 1	语言 2
女	语言 1	语言 1	语言 1	语言 1	语言 1	语言 2

157. 如果决定推算，程序根据年龄组和性别将一种语言分配给户主。在本例中，推算矩阵的登录项将只是前面户主的登录项，因为给定住户中所有其他的人都得到与户主相同的语言代码。

158. 在此时刻，如果住户中仍无人报告讲何种

界定的语言，编辑程序就利用推算矩阵根据户主的年龄和性别将一种语言分配给户主。分配的语言是数据文件中相同年龄和性别的另一个户主最近所说的语言。由于推算矩阵随着遇到可接受的情况连续不断地更新，分配的语言可能是一般群众的所说的语言。

图 16. 住户成员未分配语言情况下的实例



159. 在编辑工作的开头，将会发生编辑规则的例外。工作人员在从一个地区转至另一地区时必须小心谨慎地注意可能发生的语言变化。有些国家还必须关注说不同语言的人的局部化混居。不过，即使在这种情况下，除非存在某些语言选择性少报的情况，否则产生于推算的分配和未分配值的百分比应当大体相同。

160. 另一项编辑可能观察宗教。宗教的答案再次可以凭年龄和性别推算。在所有信息可以获得时，编辑程序将继续更新，并将从推算矩阵取出答案代表“未知”信息。这种推算矩阵像语言推算矩阵一样，但各单元中是宗教而不是语言。

161. 注意这种解释假定采用自顶向下的顺序法。使用复杂方法如 Fellegi-Holl 法和新推算法(NIM)（见附件五）的编辑小组同时适用所有有关的编辑。本过程还假定存在着适当的编辑次序。

162. 许多经济特点如劳动力参加、上周或上几周工作的时间和去年工作的时间等，能够利用类似的特点进行推算。通过利用类似的推算矩阵，编辑程序能够快速检验变量特点的值，而且编辑程序也应全面提高运行速度。

163. 注意，有时难以为一个系列的头几个推算矩阵获得适当编辑的特点。统计局通常不想包括未经编辑的项目作为推算矩阵的维数；在经编辑前，编辑

将不把性别或年龄用作推算矩阵的维数。因此，头几个推算矩阵将使用不需要编辑的不同变量或值无法改变的变量。对于人口项目的第一个推算矩阵而言，编辑可以使用住房单元的人数，包括用零代表空置单元。

164. 就一般住房编辑来说，第一个推算矩阵也可将住房单元的人数用作初始维数，但编辑小组可以修改住房项目的作用以说明空置单元。例如，如果第一次住房编辑是有关“外墙建筑材料”或“墙的类型”，初始值可以依据住房单元的人数，包括有关单位何时空置的值。

165. 在单元空置但“墙的类型”有效时，编辑用外墙类型更新第一个单元。在墙的类型已知时，对于入住的单位而言，编辑更新对应于单位人数的单元。不过，在外墙建筑材料为未知数时，推算矩阵将根据单位的人数提供外墙建筑材料的值。

166. 在初始使用推算矩阵后，编辑小组可能想转到某些其他住房特点，例如“屋顶类型”和“保有权”。不论选择何种特点，都必须明确区分单位并提

供足够的多样性，以便将不重复选择同一属性。同一属性的反复选择会给出半非流水值而不是动态推算（流水）值。例如，将动态推算用于部队营房（“集体住房”）可能造成同一值反复使用，如果选择的唯一特点是年龄和性别的话。在这种情况下，所有的居民也许都是男性，而且多数将处于有限的年龄范围内。因此，该特定矩阵可能不会给出最有效的结果。如果“保有权”具有足够的多样性，自有住房者和租赁者有着足够的比例，这一变量能够起到作用。否则，该国可使用不同类型的屋顶。

167. 一般来说，许多编辑小组发现，通过将可比较的维数用于推算矩阵，他们确实减少了检验工作，得出结果的速度加快，并且也许更加准确。

(g) 在不使用动态推算时

168. 如果编辑小组决意不使用任何动态推算，编辑的顺序仍很重要。例如，年龄与许多项目相关，包括与户主的关系、教育程度、就业和生育率（对于女性）。考虑图 17 认定的住户成员：

图 17. 户主和子女年龄数据遗漏的子女的实例

人	关系	年龄	年级	工作	职业	所生子女数
1	1	40	12	1	33	空白
3	3	X	7	空白	空白	空白

注：X=不正确的数据
空白=不适用

169. 人 3 的记录有关系 3（个子女），但未报年龄。要找出年龄，编辑程序可以使用户主与子女的年龄之差（非流水值或凭推算从以前单位得出的值）。例如，如果该差为 25，则子女的年龄变为 15 岁（户主年龄 40 减年龄之差 25）。

170. 上学年数也已知，本例为 7 年。15 岁与这一年级可能十分对应。由于特定年龄的适当上学年数范围小于户主与子女之间年龄差的年龄范围，更合

适的做法是先检验教育程度是否适当。如果教育程度已报，由静态（非流水）或动态（流水）推算确定的年龄能用来提供适当的年龄。如果教育程度未知，那么就可用户主与子女的年龄差来分配年龄。

171. 不过，甚至年龄差的信息也可能遗漏。实际上在多数国家，教育程度遗漏的可能性比年龄的大。如果年龄和年级都遗漏，下例说明编辑小组可采取的步骤。

图 18. 户主和子女年龄和年级数据遗漏的子女的实例

人	关系	年龄	年级	工作	职业	所生子女数
1	1	40	12	1	33	空白
3	3	X	X	空白	空白	空白

172. 图 18 中年齡和年级数据都缺，但存在其他的信息。人 3 未大到足可就业，而且也太年轻因而无子女（或为男性）。利用就业信息，一组非流水值可以获得年龄，但它将是一个低于最低可接受工作年龄的年龄。或者，如果编辑小组利用动态计算，推算矩阵值给出一个年龄值。选定的年龄也许应将户主的年龄用作保持一致的变量之一。例如，如果户主的年龄是 20 而不是 40，将 14 岁分配给人 3 显然将是不合适的。在年龄确定后，也能确定年级，后者因此应与年龄和工作状况保持一致。

173. 如果编辑小组决定推算所有或多数项目，它应制定一项以合乎逻辑的方式实施编辑的战略。就人口项目而言，编辑应先从可能具有未知数的所有项目开始。编辑小组应利用任何其他可获的信息帮助确定每个项目是否列入第一个推算矩阵。各国发展推算矩阵的细节都非常具有特定性，但它们的国家普查/统计局都可能得到可用于这一目的的某些信息。

174. 许多编辑软件包都深入跟踪住房单元内的人数。例如，未知性别的推算矩阵能够依据住房单元内居住者的数目分配男性或女性。因此，为一入住所内一性别未知或无效的人选定的初始值可以是男性。就两人住所而言，初始值可以是女性。对于 3 人住所，该值将为男性，如此等等。在测试并解决所有一致性编辑如户主和配偶的性别及生育信息的存在等之后，矩阵将只用作最后手段。

(h) 推算矩阵应多大？

175. 多数计算机软件包能够接受多维计算矩阵。在建立推算矩阵前应考虑下列几点。

(一) 推算矩阵太大时出现的问题

176. 在主题专家和数据处理专家联合班子携手合作时，某些国家普查/统计局遇到的最大问题之

一是过于热切。在开发编辑软件包时很容易失去控制，因而程序编制所花的时间大大超过所需的时间并减慢普查或调查处理速度。例如，编辑小组可能决定，为了确定年龄，除了“性别”、“教育程度”和“劳动力参加”外，女性还必须包括“所生子女数”。增加“所生子女数”可能会使年龄估计略为准确一些，但程序编制增加的复杂性可能不值得这样做。编辑小组还必须决定从准确和效率两个角度看推算矩阵维数多少才将给出最好的结果。推算矩阵太大（单元太多），就不能彻底更新，而且可能会反而过多地利用非流水值。

(二) 了解推算矩阵在干什么

177. 除了推算矩阵太大外，路径也可能混乱不清。重要的是应确保主题人员及数据处理人员能够遵循所有路径。他们必须共同确保推算矩阵履行预定的任务。主题人员和数据处理人员必须再次联手合作，核实推算矩阵的每个变量或维数都正确实现。此外，他们还必须确保所有的组合恰当运作。

(三) 推算矩阵太小时产生的问题

178. 如果推算矩阵维数太小，或如果由于分组（例如年龄组或教育程度分得太少），一推算矩阵值在更新前反复使用，那么该矩阵就太小。例如，在年龄阵列中无性别维数的情况下，如果年龄未知，一个家庭的所有子女更有可能获得同一年龄。主题人员应与数据处理人员合作测试所有不同组合的推算矩阵，并确保任何一种都不发生得太频繁。

(四) 对矩阵而言困难的项目

179. 事实证明有些项目如“职业”和“产业”等极难编辑。职业和产业的单独推算矩阵可能产生不一致的结果，而交叉检验各对职业和产业登录项可能是又费钱又费力的工作。例如，如果有理发师或高级理发师在鱼类加工厂工作，就需要某种其他类型的编

辑。此外，职业和产业类别数量多，可能使得很难进行动态推算。就某些项目而言，编辑小组可能断定编辑的效果适得其反，并代之以选择使用“来说明”或“未报”。或者，使用静态（非流水）法就可能够了。

4. 检验推算矩阵

180. 编辑软件包中推算矩阵的基本结构应如图 19 所示。编辑说明必须确定用于推算的阵列，并

使用非流水值代表初始组的值。

(a) 设置初始静态矩阵

181. 下文概述的过程在每次发现某人在所有 3 个项目——在本例情况下是“关系”、“性别”和“年龄”——中都有有效值时更新推算矩阵。不过，在编辑程序发现无效（或空白）性别时，推算矩阵根据有效的关系和性别代码（已经编辑的变量）选择一个值。

图 19. 非流水阵列的抽样组值和抽样推算代码

```

.
.
.
22 A01-AGE-FM-SEXRL (2, 6)
23. 户主      配偶      子女      其他亲属      父母      未报
24.  40      40      10      20      65      20      男
25.  40      40      10      20      65      20      女
.
.
.
40 如果年龄=0:98
41 设 A01-AGE-FM-SEXRL (性别、关系)=年龄
42 或
43 信息‘年龄未知，所推算’年龄
44 写‘年龄未知，所推算年龄=’年龄
45 推算年龄=A01 (性别、关系)
46 信息‘年龄现已知’年龄
47 结束-如果
.
.
.

```

(b) 错误信息

182. 编辑软件包应该提供数种方法确保它们正确实施编辑和推算。下文审查其中两个特点即信息命令和写入命令。

183. 一种信息源是信息显示，如上文图 19 所示。这项命令为各级地理单位（例如查点区、大小民政区划）及每份调查表（产生具体信息和信息发生总次数）。所有调查表的摘要报告可如图 20 所示：

图 20. 每个错误推算次数摘要报告的实例

计 数	错误数	信 息	行号
-	14-1	每名妇女太多子女	2629
-	14-2	每名妇女太多子女	2645
2	14-3	男孩未申报	2669
2	14-4	女孩未申报	2678
33	14-5	最近出生月份未报	2723
7	14-6	未生育子女；母子年龄差合适	2892

注：这里“14”仅指某个给定序列中的项目 14；错误按顺序计数。

184. 按调查表安排的报告（图 21）可以给出调查表编号，包括所有确定的地理代码。报告然后列出程序发现的错误，按项目（本例中为年龄）和按软件程序中的行号列示，如下

文右方所见。在本例中，年龄空白，但依据此人的关系和性别推算，推算矩阵提供岁数为 48。就本例而言，具体年龄未知，但如果希望，信息命令也可写入该信息。

图 21. 调查表错误抽样报告

身份调查表: 01 01 017	行号
AGE (1) = 年龄未知，所以推算	#46
AGE (1) = 48 年龄现已知	

(c) 定制错误列表

185. 软件也可提供另一种命令，允许更详细地分析编辑说明和编辑流。该命令可用来显示作出变化前的信息，然后再显示作出的所有变化。最后再次显

示带所作变化的记录。分析人员采用这种方法可以查明编辑是否正确遵循所有路径。结果如图 22 所示。输出的第 1 行给出变量（例如，省、关系、性别、年龄）。然后显示输入数据，接着是错误（本例中为无年龄），再接着是变化作出后显示的数据。

图 22. 按包括多个变量的调查表显示的补充错误列表的实例

	省	区	户主	关系	性别	年龄
输入数据	01	01	17	1	1	
错误	年龄未知，所以推算年龄=空白					
经编辑的数据	01	01	17	1	1	48

186. 这个过程协助编辑小组确定编辑是否沿着正确路径进行。

187. 这种测试是普查和调查编辑的重要部分。下列方法代表测试编辑过程的一个可能的方法。该进

程开始时可由专家通过建立一个“完美的”住户进行系统分析。完美的住户指一个完整的住户，具有所有的特点，包括户主、配偶、子女、其他亲属和非亲属。完美的住户必须通过所有的编辑而无任何错误。

(a) 数据处理人员按顺序向每个住户引入一个单一错误，对应于编辑说明和编辑程序的顺序；

(b) 然后分析人员检查编辑进程初期所有的路径；

(c) 一旦编辑正确遵循所有的路径，数据处理人员运行整个数据集抽样，查找实际数据集中特异情况并按需要作修改；

(d) 最后，数据处理人员运行整个数据集。

188. 在证实信息正确工作而且适当的修改已作出时，数据处理人员可以决定在调查表一级退出它们。如果大国运行整个数据集并留下每份调查表的信息语句，产生的行数和纸张量的代价将会过分高昂。不过，这些信息的总结报告应继续保留，因为它为不同级别地理区域给出有益的信息。输出情况将如图 22 所示。

189. 计算机编辑通常包括一个保护过程。对于变动和替代值，编辑线索显示所有的数据变化和计算。查阅编辑线索将确定变动数是否低得足以使记录组可被接受。

190. 如果某个特定项目错误太多，该项目可能未自行或相对于其他项目作了适当的预先测试，表明查点员或答表人不理解该项目。例如，有时候查点员搞糊涂了，只从男性成年人而不从女性成年人收集生育信息。如果此类数据收集具有系统性，编辑小组可能得让程序编制人员的生育数据从已婚夫妇的男性移至女性。否则，编辑小组在此阶段再纠错已无能为力。

191. 编辑程序通常需在数个不同的文件上运行以覆盖所有的情况。此外，由于错误的语法或逻辑，数据处理人员将需要作出变动。即使最有经验的数据处理专家偶尔也键入一个“大于”符号而不是“小于”符号，而且错误只是在数次运行后才被发现，因为特定的问题不一定会立即显现。相类似，逻辑小缺陷一开始也不一定显现。如有可能，主题专家和数据处理专家再次需要合作在编辑进程之前解决这些问题。

(d) 运行编辑多少次？

192. 一般来说，运行编辑程序三次是一个不错

的主意，现解释如下。

193. 第一次编辑运行提供带实值而不是带初始静态矩阵产生的值的推算矩阵。许多国家利用其他来源——以前的普查或调查或者行政记录——的数据为阵列提供非流水值。数据处理人员运行完整的数据集或其一大部分以提供推算矩阵的值。来自实际数据集的非流水值可能更为准确和当前。编辑只利用这种初始静态矩阵的大致为百分之二；其余为动态推算值。

194. 第二次编辑运行实际编辑。第二次编辑运行包括数项重复运行以覆盖所有情况。到了此时，数据处理人员将需要进行变动以便更正错误语法或逻辑引起的错误。此外，即使最有经验的数据处理专家也可能出错，而且由于特定的问题不一定立即显现，只有经过数项运行才能发现错误。同样，逻辑方面的小缺陷也不一定一开始就显现。

195. 第三次运行确保 (1) 数据集中不留错误，和 (2) 编辑程序不引入新错误。在处理人员最后这一次运行编辑程序时，错误列表中不应再出现错误。如果仍留有错误，编辑逻辑也许有错误，因此数据处理人员需要修改它。此外，此次运行通常告诉数据处理人员编辑是否偶然引进了编辑逻辑造成的新错误。

5. 推算标记

196. 推算标记这种方法用来保留关于未经编辑数据的信息。如前所述，许多编辑小组对于在变动未经编辑的答案时可能丢失潜在信息之事感到关切。在由于不一致而改变一个值的情况下，编辑小组不妨保留原始值以便在普查后作进一步的人口分析或错误分析。主题专家和程序编制人员双方都将希望分析遗漏、无效或不一致数据的各个方面。编辑小组的成员需要保证推算和未估算的分布保持一致，看看编辑和推算计划中是否出现任何系统错误。例如，有时因为程序忽略更新推算矩阵，数据处理专家偶然只利用非流水值。如果国家进行了普查预测试，编辑小组可能需要在预测试后调查某些变量之间的关系，以便最后确定调查表。在以前的普查中，带大容量硬盘的微机尚不多见时，许多统计局的磁带或其他存储媒体没

有空间保持额外数据；不过，近来多数国家保留关于未经编辑的数据的信息已不再是个问题。

197. 有些国家决意保持简单的二进制会计变量作为每个项目的标记。这个方法简单，而且为每个变量占用单个比特。例如，在 1990 年普查中，美国普查局在每个记录的末尾为住房和人口记录的每个变量设置推算标记。例如，就每个住房变量而言，标记的变量初始为“0”，如果原始项目以任何方式改变，将“0”变为“1”。程序不保留原始值，尽管统计局有时为每个记录或在合计中编制这些值。

198. 其他方法也可用来保留未经编辑的答案。在图 23 的例子中，国家普查/统计局利用推算矩阵将配偶的年龄从 70 改为 40。国家普查/统计局能够很容易地将推算前的值（本例中为 70）置入为推算标记保留的区域，并保留用于已分配值分布列表的变量，在本例情况下为 40。为了审议数据集中的变化，统计局可对已分配和未分配的值进行频率分布或交叉制表。在对编辑对于数据集的影响进行这种分析后，如果基于编辑的制表看起来可疑或异常，编制小组可能希望考虑改变编辑或编辑流的一部分。

图 23. 带推算值标记的抽样人口记录

人	性别	年龄	所生子女	性别标记	年龄标记	所生子女标记
1	1	40	空白	0	0	1
2	2	40	7	0	70	0

199. 图 24 说明一名 13 岁女性的情况，她被记录为生有一名子女（已生子女 1）。然而编辑小组已经断定，生头胎的最低年龄将为 14 岁，而且 14 岁以

下女性的生育更可能是错误而不是事实。一如既往，这引起了这种情况属于数据集中的噪声还是实际值的问题。

图 24. 生育率空白和增加标记的年轻女性作标记的实例

人	性别	年龄	所生子女	性别标记	年龄标记	所生子女标记
生育率空白						
4	2	13	1	0	0	
生育率空白和增加标记						
4	2	13	空白	0	0	1

200. 根据编辑规则，为已生子女推算“空白”信息。注意，已生子女标记稍较复杂，因为它必须说明推算的空白以及数值登录项，假定主题人员想研究报告为有子女的 13 岁女性的人数和特点。数据处理人员可将原始信息记录在留给标记的记录区内，通常在记录的末尾。然后，该组公布的表将不列入该女性已生子女的信息，但该信息仍可用于以后研究。在某个晚些时候，特别是在规划后续调查或下次普查时，编辑小组可以利用关于 13 岁女性生育子女的信息，决定他们是否需要降低包括的年龄。

201. 使用推算标记的一个问题是刚才所说的过程在计算机内占去相当大的空间。在标记重复每个变量时，编辑后数据集所占空间将接近未编辑数据集的两倍。对许多国家来说，如果长期存储，那将不是接受的。不过，可将原始数据和编辑加以存储以便日后重建。

202. 人口数量十分庞大的国家为研究目的可能更倾向于以抽样方式利用推算标记。例如，某个国家可能希望每隔 100 个住房单元建立一个数据集。然后编辑将按该较小数据集上的推算标记运行，帮助评价编辑如何影响数据的质量并确定未经编辑与已经编辑的数据之间存在什么差异。

F. 其他编辑系统

203. 本手册的大部分篇幅说明如何利用自顶向下法进行普查和调查的计算机编辑。有少数国家实施另一种较复杂的过程进行计算机编辑,叫做多变量编辑(见上文 C.2 节)。Fellegi 和 Holl (1976 年)是这些过程的首创人,这些过程通常适用于普查(或调查)中最重要的变量:年龄、性别、关系和婚姻状况。不过,它们也能适用于变量的任何组合,或普查或调查表上所有的变量。采用这种方法时,编辑程序同时观察一个住户中一个人或所有人的这些项目的答案,以便确定遗漏或矛盾的答案。

204. 加拿大统计局发展了 Fellegi-Holt 法并将它用于加拿大 1976 至 1991 年的普查。在加拿大 1996 年普查中,对这种方法进行了完善并称它为新推算法(NIM)。它首次允许“同时对大的(编辑和推算)问题的数据变量和定性变量进行最低限度变动的推算”(Bankier, Houle 和 Luc, 无日期)。

205. 如果编辑过程利用传统的动态推算或流水法进行,一系列调查表项目的推算信息可能来自许多不同的个人,这取决于用来更新推算矩阵的信息。例如,如果人 A 的性别、关系和婚姻状况是正确的,这些值将更新有关的推算矩阵。如果 A 的年龄遗漏或无效,当然将不把它用来更新推算矩阵。实际上,其他的项目将更新该值。所以,如果下一个人有矛盾的性别而且推算“性别”,人 A 将提供性别。如果年龄也未知,编辑程序将利用某个他人的年龄。

206. 新推算法利用项目的提供者,希望所有遗漏或矛盾的信息能够来自单一提供者或少数提供者。为了从单一提供者获得所有或大部分的信息,必须将全部数据记录存储在计算机的存储器中。这样,在年龄和性别均未知或无效时,存储的这些变量便可为这两个项目提供值。

207. 自动化流水推算法的目标应为下述几条:

(a) 推算的住户应与未能通过编辑的住户十分相似;

(b) 如果可能,住户的推算数据应来自单一提供者而不是两个或多个提供者。此外,推算的住户应十分近似于该单一提供者;

(c) 基于可利用提供者的同样可靠的推算动作应有类似的中选机会,以避免虚假地夸大人口中数量较少但很重要的群体的规模(Bankier, Houle 和 Luc, 无日期)。

208. 根据新推算法,这些目标首先通过确定尽可能近似于未通过编辑住户的已通过编辑住户来实现。这就是说,两个住户在尽可能多的定性变量上应相配合,只有数值变量之间存在小的差别。具有这些特点的住户叫做“最近邻”。就每个最近邻而言,下一步应是确定非配合变量(包括数值和定性变量)的最小子集,如果推算,它们允许住户通过编辑。然后随机选择一种通过编辑而且类似于未通过编辑住户和已通过编辑住户二者的推算动作。(Bankier, Houle 和 Luc, 无日期)。

三、结构编辑

209. 结构编辑检验覆盖面和确定各种记录应如何匹配起来。这些结构编辑必须保证 (a) 一个查点区内的所有住户和集体宿舍记录全都存在而且备妥; (b) 所有入住住房单元都有人的记录, 但是空置房无人的记录; (c) 住户必须既无重复的人的记录也无遗漏的人的记录; 以及 (d) 查点区必须既无重复的住房记录也无遗漏的住房记录。因此, 经结构编辑检验后保证一般调查表完整无缺。

210. 用于一次普查或调查的具体结构编辑随着时间推移可能需要变动, 因为用于确定和更正结构错误的技术变化非常快。因此, 本章审议项目有效性的比较一般的问题和记录之间及记录内部各项目的关系。第四章和第五章论述特定各别人口项目和住房项目。

A. 地理编辑

1. 住所的位置 (地区)

211. 按照《人口和住房普查的原则和建议, 第一次修订本》(联合国, 1998年, 第64页)的定义, 一个地区指“明确的人口群……在这里居民住在邻近住所, 而且具有一个名称或当地承认的地位”。有关住所位置的补充信息可在《原则和建议》第2.49-2.59段中的“地区”和“城市和农村”定义项下找到。对于负责进行住房普查的人来说, 至关重要是应研究这种信息, 因为在进行住房普查时用来说明住所位置的地理概念对于实施普查和日后对普查结果制表都极端重要(联合国, 1998年, 第2.312段)。

212. 在为位置进行编辑时, 地理代码必须绝对准确。为数据处理获得地理等级的完整而准确的代码, 是整个普查工作最艰难的任务之一。如果地理代码编错, 数据录入人员就可能将住房单元分配给国家另一个地区。这类错误纠正起来往往是非常困难的。

2. 城市和农村常居地

213. 一国之内城市与农村地区之间的传统区分基于这样一种假设: 不管如何定义, 城市地区生活

方式不同, 而且生活水准通常高于农村地区。在许多工业化国家, 这种区分已模糊不清, 而且城市与农村地区在居住环境方面的主要差别往往只是居民集中程度的差异。发展中国家城市与农村的生活方式和生活方式依然差别明显, 但这些国家快速城市化产生了对有关不同规模城市地区的信息的巨大需要(联合国, 1998年, 第2.53段)。

方框 3. 结构编辑的指导原则

结构编辑应管理下列任务:

- (1) 确保每个查点区整批记录具有正确的地理代码(省、县、查点区等), 而且采用共同的做法命名各批记录;
- (2) 保证包括每个住房单元; 而且登录查点区内全部住户;
- (3) 将住户并入有关查点区, 并将查点区并入适当的上级地理区域;
- (4) 根据人口多少和调查表的布局, 帮助决定调查表小册子内外的人口页与住户页;
- (5) 将每个个人记录分配至有效的记录类型;
- (6) 将集体宿舍或集体住房记录与住房单元的记录分开处理;
- (7) 确保各类记录之间存在对应关系; 例如, 空置单元无人, 入住单元至少有一人。确保每个住户个人记录数与住房记录上住户总计数对应。确保当多个文件用于单一住户时调查表数目正确, 并且正确联系;
- (8) 消除住户内部重复记录(重复的人)和住户之间的重复记录(重复住户或住户的各个部分), 以避免过多覆盖;
- (9) 在一个记录类型内处理空白记录;
- (10) 处理遗漏的住房单元。

214. 多数国家在普查前先确定哪些地区是“城市”地区, 哪些是“农村”地区, 并在收集普查数据后再作所需调整。如果国家给城市和农村常居地分配

代码（例如城市为 1 和农村为 2），这些代码可在键入时输进去，或可根据编辑小组规定的标准在编辑期间确定。如果编辑小组提供一份城市地理单位和农村地理单位的清单，数据处理人员就能容易地将适当的代码分配给住房记录。

215. 应作出努力确保居民特点与查点区普遍一致。例如，在有些国家，除了医生、教师和类似职业的人外，农村地区不应有很多专业人员，而城市地区也不应有多少农业工人。编辑小组应进行检查以查明地理区域分类是否正确。

B. 覆盖面检查

1. 事实上和法律上的查点

216. 国家普查/统计局收集普查信息时往往按事实上（人员在普查夜所在地）或法律上（人员通常所在地）的方式进行。检查住房记录之间，特别是住所内人员计数与个人记录之间关系的编辑必须考虑普查的类型。有时各国同时收集事实上和法律上的信息。每个人的一个项目可以表明他（或她）是否（1）始终居住，（2）临时来访但通常住宅在别处，或（3）通常居住在该住户内但临时外出。以事实为基础的汇总表只使用（1）和（2）项，如果所有三项都存在；以法律为基础的汇总表只使用（1）和（3），如果所有三类都存在。⁴ 设计编辑程序时应确保，在所有三类记录都存在时，对应要适当。如果事实上的记录答案很少，这可能表明他们实际上是缺席居民，或存在着可能需要特别处理的查点问题。

2. 住户和住房单元的分级

217. 第五章审议住户、住房单元与住所之间的关系。这些概念的实施取决于个别国家普查/统计组

⁴ 采用这些类别的国家普查/统计局在使用时必须十分谨慎小心，不仅在数据收集和处理时，而且在以后分析时也是如此。在使用这三个类别时，用户必须知道所选定的人口，因为分析所有数据集时将会导致某些人包括两次。如果要求事实上的人口，汇总表必须将类别（3）即临时外出的人除外；如果要求法律上的人口，汇总表必须将类别（2）除外。初始制表即有关印刷报告和补充媒体的制表期间，编辑小组可能选择制作总数据集的一个子集以便处理。至于后来的制表，文件资料应明确指出如何处理各种可能的汇总表。

织。不过，在着手个别住房编辑前，编辑小组必须制定检查方法用以确保在数据收集和键入期间遵守分级。

3. 调查表的片断

218. 在逐项编辑前，计算机程序必须检查有效的记录、遗漏的记录和重复的行号，作为结构编辑的组成部分。它还必须断定正在编辑的记录是否属于居住在集体宿舍中的人。数据录入人员在录入记录时可能出错，而且有时将会忘记删除片断信息（记录的部分）。初步编辑的一个职能应是审查文件的片断记录，以便删除它们。最常见的情况将是一份记录载有地理代码但无人口或住房信息。

C. 住房记录的结构

219. 可以列入全国住房普查或调查信息收集范围的专题之一是一座建筑物内住所的数目。在这种情况下，查点的单位是建筑物，而且收集关于其中平常和基本住所数目的信息（见联合国，1998 年，第 2.418 段）。

220. 用语“一般编辑”指这样一种做法，即确保作为建筑物组成部分的住房单元数匹配住房记录中住房单元的总数。在匹配不一致的情况下，作为建筑物特点输入的住房单元数应加以更正，以匹配住房单元数记录。如果有关建筑物编辑为拥有 5 个住房单元，但该建筑物个别住房单元记录的实际计数为 4，编辑小组必须决定作出何种调整：（a）根据个别记录计数改变第一个数字（多数情况下证明这将较可接受）；（b）利用关于现有记录的信息引入另一项记录（此种做法应加以避免）。

D. 住房记录与人口记录之间的对应

221. 如果普查或调查包括住房和人口记录二者，就需要结构编辑以弄清两类记录是否相吻合。

1. 空置和入住住房

222. 空置住房单元应无人口记录，而入住住户必须有人口记录。存在人口记录但住房列为空置时，将把空置状况改为入住状况。有时记录布局将空置状况和保有权合列同一项内，所以在作决定时还必须将

此信息考虑进去。还有，如果自有住房单元的“单元价值”或租用单元的“支付租金”有答案可用，那么编辑程序在作决定时使用这项信息；否则，就可能需要推算矩阵。

223. 如果假定为入住单元的住房无人口记录出现，编辑小组就必须决定将它计为空置单元还是用另一单元的人置换。如果该单元空置，推算可容易地将变量改为空置状态。不过，如果该单元入住，那么编辑小组就必须决定是否和如何分配人数相等特点类似的另一单元的人，如果能够这样做的话。由于不可能知道遗漏人员的特点，只有编辑小组断定无其他办法可想时，才应利用这种方法。下文概述三种可能的备选方法。

(a) 选择让住房单元空置

224. 在这种情况下，编辑小组决定，来自实地的空置住房单元应留为空置，所以不推算值。空置单元的住房编辑在第五章中说明。

(b) 数次重访住房单元以完成调查表

225. 国家普查/调查局可以选择实施这样的过程，即要求查点人员重新审查关于空置单元的数据直至他们弄清这些单元空置还是入住，并直至他们至少收集了最低限度的特点为止。在此种情况下，编辑小组应制定编辑方法并进行检查，以查明该单元是空置还是具有足够的特点可被认为“入住”。依据编辑小组决定哪些信息为“最起码的”信息而定，适用第四章说明的正规编辑，或为“遗漏的”人提供来自提供者记录的数据，如上文第 205-207 段所述。

(c) 用另一住房单元代替遗漏的人

226. 替代整个住户或遗漏个人的过程在本章其他地方说明。这些步骤要求假定遗漏的人具有与替代的人相同的特点，通常情况下几乎绝不会有这种事情，而且步骤本身也非常复杂。尽管如此，如不采取这些步骤，人数和按特点分类的人的计数可能减少。

2. 重复的住户和住房单元

227. 由于多种原因而出现重复的住房单元。有时候个别数据输入人员将会两次输入同一住房单元。

有时候不同的数据输入人员将会偶然地重新键入同一一些住房单元，或由于国家普查/统计局缺乏质量保证而重新键入整个查点区。第三，查点员可不正确地记录一个住房单元的地理代码，由于向它分配与另一住房单元相同的地理身份而产生重复的信息。

228. 如果统计局监管人员成批键入，就将不会发生重复问题。尽管如此，应开发一个编辑程序确保不会由于数据录入人员两次键入同一或同一些住户而出现重复的住户。各国不应在完成结构检查和消除重复记录问题前对数据进行分类。在分类前，工作人员能够以人工方式更正成批记录；分类后，工作人员就无法找出问题了。在分类数据时，编辑能够检查重复的住户并利用推算消除随后的重复录入项。

3. 遗漏的住户和住房单元

229. 同样，经分类后，遗漏的住户可能变得明显了。例如，编辑程序预计在最低一级地域内住户是有顺序的，例如 1、2、3、4，但只收到 1、2、4。此时必须作出决定，或者对各单元重新编号，或者找出某种“可接受的”方法用另一单元取代单元 3。

4. 居住者数目与居住者之和的一致

230. 记录在住户记录上的居住者数目应与住户中人数之和完全相等。编辑程序将人数相加，然后将此值与住房记录上居住者的数目作比较。如果和不同于居住者数目之值，就必须或者调整居住者数目之值以等于人数之和，或者必须调整个别登录项。

(a) 在居住者人数大于居住者之和时

231. 如果住房记录上居住者数目之值大于个人之和，编辑小组就面临一个实际问题。谁也不清楚遗漏的人的特点。因此，选择按特点推算遗漏人员特点或用类似住户人员取代的编辑小组可能面临两难境地。遗漏人员不应取代。然而，如果接受居住者人数的值，备选方法就是缩小查点人口的数目。编辑小组必须分析整个情况，然后决定合适的路径。

232. 有几个方法可用来找出和取代遗漏的记录，但无一完全令人满意。可以保留具有不同的但却是很重要的特点的整个住户。在找出一个具有某些但不

是所有个人的住户时，可以检索文件找出一个所有或多数已知特点相匹配的住户，然后根据提供住户的其他人员可以调整遗漏的人。不过，这种作业的程序编制工作非常复杂，所以利用这种方法的普查/统计局应为这种作业提前很长时间开始规划。

233. 这种过程的变通办法是对记录有遗漏的所有住户作出标记，并接着进行其他编辑。在编辑进程结束时，并在更正所有个人登录项之后，编辑小组可以选择让数据专家检查文件，利用充分编辑的数据集进行增补和变动。通过采用这种自顶向下法，编辑小组可找到可以接受的提供者。

(b) 按性别检查人数

234. 有时候住房记录按性别报告居住者人数。在这种情况下，编辑必须按性别分开合计人数。如果和再次不同于居住者数目，每种情况下都应调整值之一。通常情况下调整住房记录上的总数而不是增加“遗漏的”记录或删除具有有用信息的记录。

(c) 按顺序编号

235. 人口记录应按次序编号。这些编号应作为变量出现，例如调查表上的行号或顺序号。另外，顺序号应按数字出现。可能出现错误：有时候由于查点员按错序汇总信息而使调查表或人员表格弄乱，或者他们可能跳过页面而在数据集中留下空白页。虽然缺乏顺序通常不影响编辑或制表，但许多国家普查/统计局选择按正确的次序对人员重新编序。因此，编辑程序必须能够找出乱序的人并重新编序。由于重新编序有时将会影响与户主的关系，编辑说明中应考虑这一点。重新编序将肯定影响有关变量如母亲的行号或丈夫的行号。

5. 居住者与建筑物/住户类型之间的一致

236. 住户成员之间关系的类型与住房单元的类型应保持一致。有时候住户成员出现在申报为集体住所的住宅中或相反。在这些情况下，关系类型或住房单元的类型必须考虑到住户的规模和其他变量。

E. 重复的记录

237. 重复加行号不可能出现在光阅读调查表

或其他扫描调查表中。对于键入的表格，国家普查/统计局可能选择检查住户清单与人工键入的住户行号之间的一致。这种人口检查可以提高键入数据的质量，特别是对(1) 出现在列有住户全体人员的一页上的人名与(2) 关于人的纵列、横行或页面上的数据进行比较。一开始看上去重复的两个人，在查阅其名字时可能发现，他们实际上是双胞胎。

238. 如果正确设置数据屏幕和跳跃模式，键入的表格不应有重复行号。当代多数软件包自动产生顺序号作为数据录入进程的组成部分。在工作人员录入一人的重复记录时可能引入错误，或错误的行号可能产生重复记录。当处理每份记录时，编辑程序将它与住房单元以前的人口记录作比较。编辑必须弄清每个行号已正确捕捉。重复加行号是错误，而且必须变更。

239. 各国可能选择开发自己的键入计划，而不是使用现成的软件包。此时编辑小组必须确定可接受的错误率。许多方法可用来作出这些决定。一种方法可能是遵循下列指导原则：

(a) 如果两份不同记录的行号一致而且不同的特点数为2或小于2，编辑将消除其中一份记录，因为它可能是重复的；

(b) 如果不同的特点有3个或超过3个，将变更行号。

F. 特殊人口

1. 集体单位的人

240. 结构编辑处理居住在集体单位如机构、营房或疗养院中的人的方法，应不同于处理居住在正规住房单元中的人。由于集体单位通常将无户主，各国必须确定如何最有效地区分单位类型。一种方法是集体单位采用不同的记录类型。另一种方法是为关系分配一个特定的代码，它代表“集体”（“group”或“collective”）住所。

(a) 当集体单位是不同记录类型时

241. 当国家普查/统计局选择使用单独的记录类型时，编辑小组将不难确定哪些记录为集体单位的记录。只要直接参照这些记录就能容易地完成集体单

位的制表。集体记录独有的变量如集体类型等可单独编辑和推算。不列入集体记录的变量能够容易地进行检查，以确认它们实际上是空白项。不过，产生的文件体积较大，因为这些记录可能短于正规的人口记录，但占用的空间将与长方形的文件一样大。另外，在编辑和推算期间，有些程序可能得因某些项目而检查人口记录和集体记录二者。

(b) 在一个变量区分集体记录与其他记录时

242. 在使用单独的变量而不是单独的记录类型时，编辑小组可能较难确定哪些记录是集体单位或集体记录。在这些情况之下，只要参照变量本身，仍能很容易制作集体单位的汇总表，因为这种变量说明哪些记录是集体单位个人的记录。能够容易地检查不列入集体记录的变量，以便通过参照集体单位的代码而弄清它们实际上是空白项。产生的文件体积较小，因为不需要集体单位个人的补充记录，而是将它们简单地包括为带有代表住户/集体单位变量的不同代码的人口记录。在编辑和推算期间，程序将只需检查人口记录，而不是因某些项目而检查人口记录和集体记录二者。

(c) 在“集体类型”代码遗漏时

243. 表明集体单位的代码可能遗漏或无效，或集体单位代码与关系代码之间可能发生不匹配。在集体单位代码遗漏但关系代码表明集体单位时，建议的解决方案是相应地变更集体单位代码。如果集体单位代码存在但关系代码遗漏时，关系代码可根据集体单位类型确定。

(d) 在集体单位代码存在但所有的人均有关系时

244. 如果集体单位代码存在但根据关系代码住房单元内所有的人都有亲属关系，那就应改变代码以表明一个住房单元。另一方面，如果单位编码为住户但单位内没有两个人有亲属关系，就可能需要将单位改为集体住所。住户可拥有五六个无亲属关系的人而且仍不是集体单位。如上文所强调，编辑小组的成员可能需要协商以解决特定和异常情况问题。

(e) 区分各类集体单位

245. 多数国家区分各类集体单位。它们经常将信息进一步细分为特定类型的集体宿舍。这种信息可以或者单独编码为“集体宿舍类”项目，或者作为多种可能性包括在住户关系代码中。

2. 无住房人口

(a) 季节性迁移

246. 在有季节性迁移的某些国家，访问人将需要了解是否由于参照时间关系单元空置或有人居住。所以，即使住户有完整的信息，也可将该住户在另一地点计数（查点）。当然，相反的情况也成立。如果不小心谨慎，就可能遗漏在不同地点有两个住所的住户。

247. 有时候，非常经常的情况是，整个住户在一年的部分时间居住在一个地方，在一年的其余时间又居住在另一个地方。国家普查/统计局和编辑小组必须决定如何处理各类情况。例如，有些人在一年的部分时间呆在另一个家，例如有人在一年的暖和季节居住在一国较寒冷的地方，而在一年寒冷的季节居住在该国较暖和的地方。另一种情况是游牧民，他们在一年的部分时间移徙，部分时间定居——也许正赶上该国选择普查的那段时间。

(b) 无家可居的人

248. 依照定义，无家可居的人将无住房信息。不过，建立“虚拟”记录（一种新记录，初始时包括某些变量的空白值）将使结构检查变得较为容易，而且使记录与其他住户单元的结构保持一致。编辑小组将必须决定是否建立这种虚拟住房记录以协助数据处理和制表过程。

(c) 难民

249. 同样，难民可能住在临时住所，并可能需要以一个特定的变量、一种独立的记录或一份虚拟的住房记录加以表示，以说明他们的状况。编辑小组将需要开发和实施适当的程序。

G. 确定户主和配偶

1. 编辑户主变量

250. 在确定住户成员时，历来的做法是先确定户主或参照人，然后再按与户主或参照人的关系确定住户其余成员。户主定义为住户其他成员公认为户主的那个住户成员。各国可以使用它们认为最合适的用语来确定此人（户主、住户参照人等），只要唯一这样确定的人用来确定住户成员之间的关系。建议各国在发表的报告中说明所用的概念和定义（联合国，1998年，第2.67段）。

(a) 关系的次序

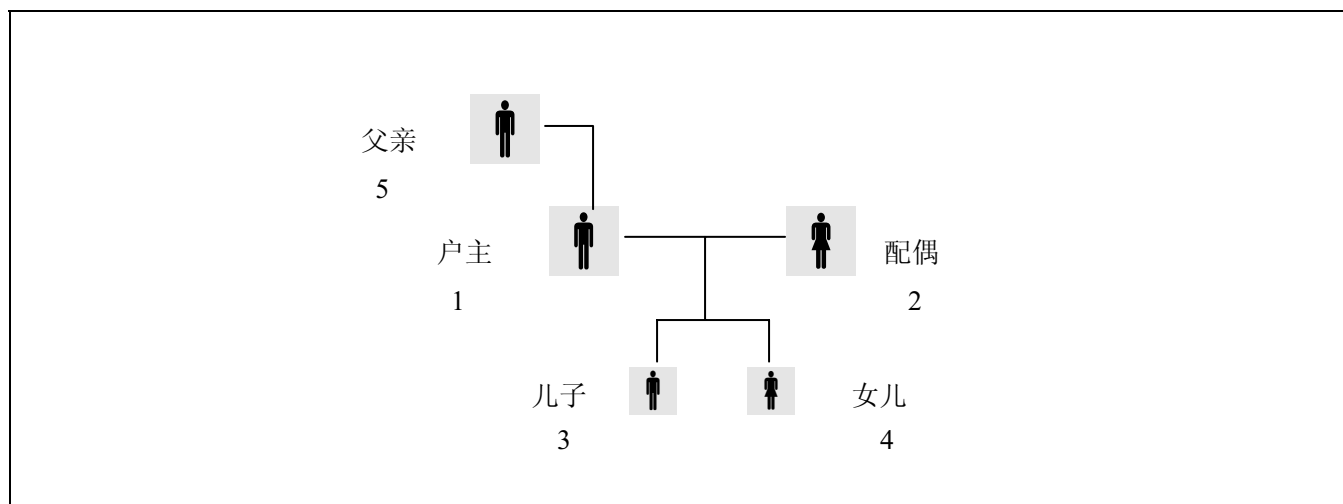
251. 单元内关系的次序影响编辑，因为许多编

辑假定户主是第一人，而且他（或她）的数据先编辑。例如，首先在对户主的编辑中检查诸如语言、民族和宗教等变量。如果户主拥有其中任何变量的有效信息，则住户的任何其他成员如出现遗漏、代码错误或键入出错，就为其推算此种信息（参阅第四章）。户主需要先编辑，因为将他（或她）的特点用来分配或推算住户其他成员的值。

(b) 在户主不是第一人时

252. 查点人员依据在指定户主时遇到的不同情况而在实地采取的行动影响编辑过程。为理解这个问题，首先考虑图 25 说明的住户。

图 25. 户主列为第一人的住户的实例

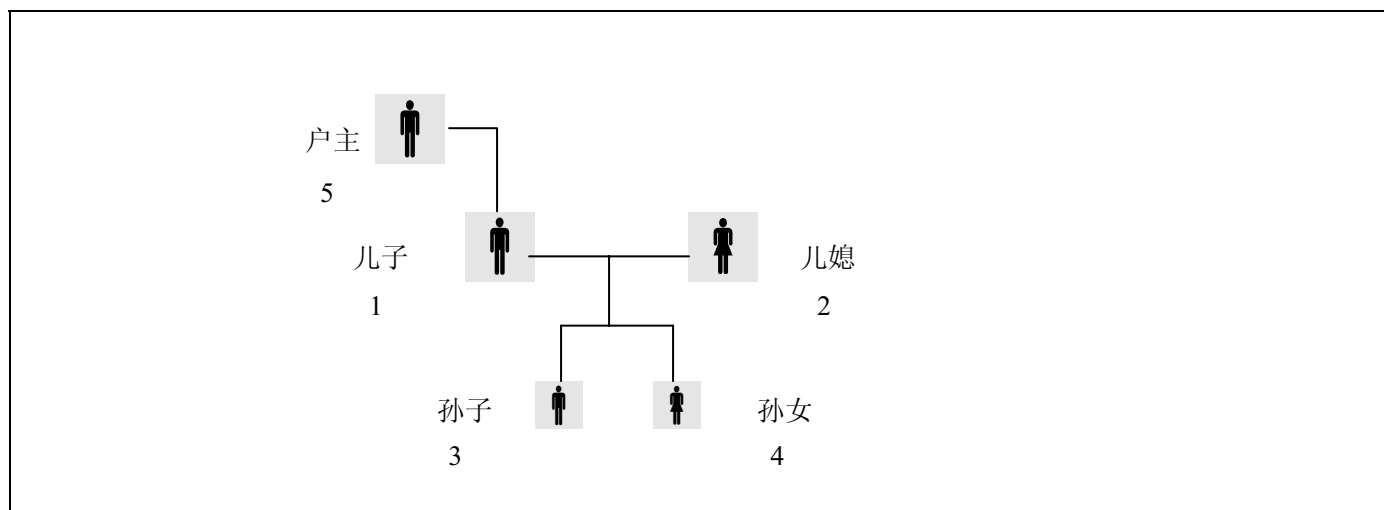


253. 该住户显示实地遇到的典型情况：户主和配偶、他们的子女及户主的父亲。如果查点员以这种方式收集信息，基于户主处于住户第一位置的编辑将会顺利进行。

254. 然而，如果查点时将祖父指定为户主，就

重新配置关系，如图 26 的第二次描述。如果查点人员跨进住所，发现由夫妇和两个子女组成的核心家庭，访问期间户主的父亲进入房间并声称他是户主，就将出现上述情况。依据一般推定的户主的协议，人 5 将成为户主，而人 1 成为儿子，人 2 为儿媳依次类推。

图 26. 户主列为第五人的住户的实例



255. 显然，按此两个住户所示，基于不同指定的户主的编辑路径将是不同的。为了余下的编辑和制表而确定实际户主有三种不同的可能性：(a) 可用一个指示符来说明哪个人是户主，而且指示符在编辑和制表过程中可自始至终地使用；(b) 如果户主不列为第一人，可将他（或她）移至第一位；名单上位置较高的人可各下移一位；或(c) 可变更关系代码，使第一人成为户主，而不管其他关系。

(一) 分配指示符给户主记录

256. 在关于户主的编辑过程中，使用指示符来确定单元内户主的行号。如果户主留在收集的位置上，可将指示符设定在该位置，每当特定的编辑或制表需要时始终能容易找到户主。可将变量“户主指示符”设定于户主的行号，并在编辑中用来分配或推算单元内其他人遗漏或无效的特点。如果户主是住户内第一人，变量“户主指示符”的值为“1”。

(二) 使第一人为户主

257. 编辑小组可选择将户主移至住户的第一位。此种编程工作略比上文第(一)种要求的复杂。数据处理专家必须编写一个程序，将户主移至表上第一位，后跟以前处于第一位的人，再后跟原先处于第二位的人，以此类推，直至达到原为户主前一位的人。因此，如果户主处于第5位，各人的次序将从1、2、

3、4、5变为5、1、2、3、4，作此变动后，户主将处于第一位，这将使其余编辑变得较易进行，因为户主将始终在该位置。尽管如此，如果进行此种操作，会给数据集的完整性造成某种“损害”。因为各人的次序已发生变动，分析人员可能难以确定实地收集各人的次序和这种次序对结果解释的潜在影响。

(三) 重新分配关系代码以使第一个为户主

258. 如果编辑小组决定，所列的第一人应为户主，那么在编辑中需要遵循过程(a)和(b)。

(a) 向第一人分配户主的值；

(b) 执行一个例行程序，将值分配给住户中其他的人以调整住户。例如，在图26中，父(母)以户主作开头。在将人1定为户主时，人2将需要重新分配为“配偶”。人3和人4将分配为“子女”，而人5将重新分配为父(母)(如图25所示)。将需要一个子例行程序载有一个矩阵以保存初始值和变动后的值。

259. 采用这一过程，影响数据集完整性的程度甚至更大。各人的次序不像上例那样变动，而且分析人员将不难确定从实地收集的各人的实际次序。但是，所有的关系将改变，而且分析人员将不知道最初哪人选为户主。另一方面，户主处于第一位置时制表通常将较为容易。与上例不一样，就此过程而言，程

序编写人员将不必实际调动记录。

(c) 一个以上户主

260. 在发现一个以上户主时，编辑小组必须确定将谁定为户主。必须根据主题专家确定的特点和按编辑流程进行编辑。编辑程序然后必须重新分配被认定为户主的其他人的关系。

(d) 无户主

261. 同样，如果未发现户主，编辑必须确定应将谁指定为户主。在本例中，住户中其他人之间的关系将可能需要通过编辑进行调整。

2. 编辑配偶

(a) 在一夫一妻制社会确实发现一个配偶时

262. 如果确实发现一个配偶，变量“配偶指示符”为以后的编辑跟踪配偶的行号。这些编辑必须可能包括寻找户主和配偶相反的性别，寻找适当的年龄差或寻找其他有关的特点。

(b) 在一夫一妻制社会发现一个以上配偶时

263. 在一夫一妻制社会中，如果数据集中出现一个以上配偶，编辑必须确定谁是配偶并重新分配原被确定为配偶的其他人的关系。主题专家必须再次确定编辑的特点和流程应是什么。

(c) 一夫多妻制社会的配偶

264. 如果在一夫多妻制社会发现一个以上配偶，编辑小组可能想使信息保留原样不变，或进行某种一致性检验。例如，至少每名一夫多妻制配偶应有户主相反的性别。如果发现相同性别的配偶，应对相

同性别的配偶先进行编辑。

H. 年龄和出生日期

1. 在有出生日期但无年龄时

265. 在收集出生日期而不收集年龄时，可从普查或调查日期减去出生日期获得年龄信息。有些国家普查/统计局选择只根据普查年份或出生年份获得年龄，给出有可能发生偏差的值。如果使用年和月，年龄将更精确，但使用年月日时结果将最精确。

2. 在年龄和出生日期不一致时

266. 在普查或调查获得年龄和出生日期二者时，从参照日期减去出生日期获得“计算”的年龄。如果此值与报告的年龄之差超过1年，编辑小组可能希望采取补救行动。通常，出生日期优先于报告日期，并用计算的年龄取代报告的年龄。

I. 计点无效登录项

267. 有些编辑小组可能选择实施计点主要变量例如年龄和性别等（或所有变量）无效和不一致登录的数目，然后才开始实际编辑。如果编辑小组事先作了准备或利用这同一些项目进行定期调查，它们可以获得几种不同的动态推算阵列加以利用。如果无效或不一致登录项的比例非常小，编辑小组可以决定只将少数变量用于推算。如果错误百分比比较大，编辑小组可能需要利用更多的变量来说明所需的大量推算。

268. 较小的推算阵列通常较好，因为随着编辑和推算的发展它们较易检查，而且在实际编辑期间也较易利用。不过，如果数值反复利用，就将需要较大和变化较多的推算矩阵。

四、人口项目的编辑

269. 第四章涉及编辑人口项目，包括与人口统计、迁移、社会和经济特点有关的项目。这些编辑的说明考虑到了个别项目的有效性和人口项目之间以及人口项目与住房项目之间的一致性。对各项目之间的关系有所了解，使得能够规划一致性编辑，以保证获得更高质量的数据用于制表。例如，人口记录不应有15岁的女性生有10个子女或有7岁的子女上大学。

270. 在给人口项目分配值时，编辑小组必须决定是否分配“未申报”；分配一个静态推算（非流水）值用于“不明项”或其他值；或根据其他人或住房单元的特点分配一个动态推算（流水）值。在许多情况下，优先采用动态推算，因为它消除了制表阶段编辑的必要性，届时只利用汇总表本身的信息便能就未知数作出决定。推算矩阵提供空白登录项，无效登录项，或在不存在其他具有有效答案的相关项时解决不一致问题。有些国家不同地区的人口特点具有某些变化，但多数个别地区内部差异很小。另一些国家不同地区之间差异很大，特别是有关城市和农村常居地。在发展推算矩阵时必须考虑这种差异，在确定初始“流水”值时尤其如此。编辑小组应当确立应为空白提供登录项的情况。这一登录项应来自上一个具有类似特点的住房单元。

271. 所有人的记录都应具有序列号以有助于数据处理。第三章说明的结构编辑检查顺序号与序列号次序之间的一致性。

272. 编辑小组只应为可适用的项目编辑每个人的记录。经编辑后的项目可能不同，这取决于城市/农村、气候和其他条件。可取的做法是有选择地进行编辑，依这些条件而定，但在实践中没有几个国家有时间或专门知识去发展和执行多个阵列，以变更遗漏或不一致的数据。实际执行这种增加的程序的国家的国家甚至更少。

273. 调查表上收集的信息也取决于选定的人口特点。例如，只问及女性的生育率和只问及成年人的经济活动。

274. 有时候编辑小组应允许某些项目有“未报”登录项。编辑小组可能缺乏可靠的依据来推算某些特点的答案。作出“未报”答案的决定必须根据为规划和政策用途产生适当的列表特点的要求来权衡。只要“未报”情况的分布与报告情况的分布相同，在规划人员需要选定的信息时分配“未报”情况不应造成问题。但是，如果“未报”情况有所偏斜，编制后的推算可能造成问题，特别是就小地区或特定类型条件而言。例如，如果少年女性回答人拒绝披露其生育信息而且未收集到生育率，编辑进程将无法协助获得此种信息。

275. 人口编辑往往比住房编辑复杂，因为交叉列表一般复杂得多。多数国家只按各级地理区划编制个别住房特点，但人口项目可能有多层定义列表。如上文解释，选择不利用动态估算的国家应为“未知”值确定一个标识符，以便在出现无效或不一致答复时使用。

276. 对于使用动态推算的国家而言，编辑小组应当发展具有区分人口特点的简单推算矩阵。对于多数国家而言，年龄组和性别是动态推算最佳的初始变量，而且应首先编辑。利用多变量编辑的国家统计/普查局应同时编辑年龄、性别和其他变量如关系和婚姻状况。可能有助于动态推算的其他项目包括教育程度和就业状况。

277. 编辑小组必须非常小心谨慎以便在推算期间不使数据偏斜。不应假定未推算的数据和推算的数据将一定具有相同的分布。经常的情况是，未知数据本身就已偏斜。例如，老年人报告年龄的可能性小于年轻人。

A. 人口特点

278. 每个人的关系、性别、年龄和婚姻状况的数据是普查的基本数据，而且也许应合在一起编辑。人口或分组人口的年龄和性别结构对于基于人口普查的几乎所有规划工作来说都是基本的要素。这些项目对于产生富有意义的汇总表来说也至关重要，因为

几乎所有其他的分析都基于按年龄和性别分列的其他变量的交叉列表。

279. 编辑人口与住房数据的多变量(Fellegi-Holt)法在本手册第二章中介绍。由于人口变量是所有普查规划工作不可分割的组成部分，如果时间和专门知识允许，应使用这一方法。总体数据集的质量几乎肯定会受益于着眼于年龄和性别及其他选定变量以确定错误或不一致的优先编辑。错误最多的项目首先编辑，接下来是错误或不一致较少的项目。

1. 关系 (P2A)⁵

280. 关系项目用来帮助确定住户和家庭结构。它出现于多数普查和调查表开头附近，并且协助确保计点住房单元内的每个人。查点人员和答卷人利用住户成员之间关系的信息来确保无人遗漏。关系项也有助于检查住户成员之间的性别和年龄的一致性。单一户主和（非一夫多妻制社会）不多于一个配偶的确定由第 250-264 段论述。

(a) 关系编辑

281. 由于关于关系的统计资料日益重要，在开发允许家庭和分家庭构成适用于各种列表的编辑时应适当小心。首先发展适当的关系代码显然将有助于这种努力（关于家庭类型记录见第 607-610 段）。

282. 在不能分配关系和不使用动态推算时，必须为无效或不一致的答复分配“未知”。如使用动态推算，可从按年龄和性别或其他适当特点构成的推算矩阵分配关系。推算矩阵不应推算将与住户内既定关系相冲突的关系。例如，甚至在一夫多妻制住户，也不应推算第二和第三配偶，除非编辑小组决定实施此种编辑。

(b) 在户主必须首先出现时

283. 如果户主不作为第一人出现，第三章介绍

的结构编辑表明可利用指示符跟踪户主位置。如果编辑小组想让户主成为第一人，可将户主置于第一位置，其方法是或者重新安排各人次序，或者使住户位置保持不变但重新安排关系，如结构编辑一章所说明的那样。前一种方法要求很多的程序编写专门知识；而后一种方法如果不极端小心谨慎的话，可能会损害数据集。

(c) 在关系颠倒编码时

284. 有时候查点人员颠倒收集关系信息：不是收集住户中每个人与户主的关系，而是收集户主与每个人的关系。因此，第三人的关系为“父母”而不是“子女”。住户最终可能出现 4 个父母而不是 4 个子女。在编辑小组发现这类系统性问题时，它必须开发某种解决方案不给住户造成过多损害。

(d) 在存在多配偶时

285. 如按第三章说明的那样进行，结构编辑将已为一夫一妻住户检查了“一个和唯一的”户主及“不超过一个”配偶。就一夫多妻制住户而言，编辑小组应决定何时允许一夫多妻制关系和何时不允许。有时候，似乎拥有一夫多妻制关系的住户实际上是错误。例如，一个住户可能已有一个确定的户主和配偶，但另一对又相互报为“配偶”，使配偶总数达到 3 个。编辑应作检查，以弄清第二对配偶实际上不是父亲和母亲、儿子和儿媳、女儿和女婿或某种其他组合。有时候，可以一定的把握确定这些关系，而有时候不能。当对上述详细关系编码时，编辑小组应预期看到适当的推算。在一夫多妻制住户中补充配偶为实际配偶时，编辑应当检查性别，也许还应检查年龄。

(e) 当多个父母出现时

286. 住户不应报告两个以上的父母，而且父母应当性别相反。当两个以上的父母出现时，也许应当使追加的父母成为“其他亲属”。有时候普查或调查有一个代码代表“父母”或“岳父母”，它将允许多达 4 个“父母”而不是两个，每个性别的父母不超过两个。

(f) 在普查收集限定性别的关系时

287. 有些普查或调查收集限定性别的关系：

⁵第四和第五两章括号中的数目指《人口和住房普查的原则和建议，第一次修订本》，统计文件，M 辑，第 67/Rev.1 号（联合国出版物，出售品编号：E.98.XVII.8）出现的项目，因此，“P2A”对应于该出版物第 59-60 页第二部分第五章“普查题目清单”人口项目 2a。住房项目附有从《原则和建议》第二部分第六章第 2.293 和 2.416 段确定的基本和补充住房普查题目衍生而来的建议的代码。

“丈夫”和“妻子”分开而不是“配偶”；“儿子”和“女儿”分开而不是“子女”；如此等等。如果不对这些答案编辑，汇总表可能就有带“男性”女儿或“女性”丈夫的数据。编辑必须决定编辑的优先次序——是关系优先还是性别优先。在有些情况下，例如对于丈夫和妻子，其编辑较之对其他人员如年幼子女的编辑更为重要。

(g) 在关系和婚姻状况不匹配时

288. 关系和婚姻状况重叠时应相互一致：报告关系为“配偶”的人，在婚姻状况项目内应为“已婚”。在两个项目不一致时，由编辑小组作出关于改变哪个变量的选择。有时候，关系含糊不清，因此在编写编辑说明时应加小心。例如，在许多国家，brother-in-law 既是配偶的兄弟（而且无须结婚），也是同胞姐妹的配偶（而且必须结婚）。

289. 在关于关系的报告中还经常出现若干其他当代问题。当两个不同性别的未婚者婚外同居时，关系代码可以是“未婚伴侣”，或可以是“配偶”。如果普查或调查有一个代码代表未婚伴侣，那么适当的婚姻状况就不应是“已婚”，除非此人与同居之人以外的人结婚。

290. 同样，同性别的人现在也同居，或者保持罗曼蒂克的关系，或者保持非罗曼蒂克的关系。保持非罗曼蒂克关系的人的代码可写为“同室人”或“非亲属”。就处于罗曼蒂克关系的人而言，有些国家定为“未婚伴侣”可能是一种合适的分类。此时，编辑小组还必须决定适当的对应婚姻状况。

2. 性别 (P3A)

291. 性别是最易收集的特点之一，但要求在编辑时作些思考。它属于最重要的变量之列，因为大部分人口特点是根据性别来分析的。性别推算要求与其他变量作些比较。在有些情况下，性别应当基于有关人员通常是户主与配偶之间性别的差异，但也基于父母与姻亲之间的差异。性别也许不应留为“无效”或“不明”，因为它是一个极其重要的变量。因此，应当作些思考如何最有效地获得符合国家实际情况的结果。

292. 如果某人不是户主或户主的配偶，就不存在其他的人可用于参照；因此，此人记录中的其他项目应加以检查。如果出现任何生育率项目，就应分配女性代码。不过，例如此人的性别遗漏或无效但配偶存在，如果配偶的性别已表明，编辑可将相反的性别推算给此人。

(a) 在性别代码有效但户主和配偶性别相同时

293. 在矛盾证据看来可靠的情况下，即使存在有效代码，也应改变性别代码。例如，记录表明，在住户已有户主和配偶或在分支家庭中已有已婚夫妇时，存在着第二对夫妇。如果第二对夫妇中的两人报告同一性别，就用关于生育率和其他项目的信息来确定谁是男性谁是女性。然后就能改变错误之人的记录。

(b) 在一男性拥有生育信息或一成年女性无此信息时

294. 编辑可能发现住宅内有一个拥有生育信息和（或）子女的男性，这种错误可能基于母亲的个人编号或类似的变量。如果无配偶存在，可将性别改为女性而不是删除生育信息。同样，根据由编辑小组确定的某些情况，可将无生育信息和陪伴子女的成年女性改为男性。

(c) 在性别代码无效而且存在配偶时

295. 如果性别的登录项为空白或无效，在配偶性别有效的情况下，编辑程序应当利用与户主的关系和配偶的性别的登录项来编辑正确的代码。如果与户主的关系是“户主”，那么程序就检查以看看是否存在配偶（通过检查住户中关系为配偶的另一人）。通过确定配偶的性别代码，将相反的性别代码分配给户主。

(d) 在配偶的性别代码无效时

296. 如果此人与户主的关系是“配偶”，而且户主的性别已给出，程序就将与户主性别相反的性别分配给此人。

(e) 在性别代码无效而且女性信息存在时

297. 调查表中众多的线索表明答卷人是否为

女性。如果程序尚未确定此人的性别而且存在任何女性指标，那么此人为记录应分配女性性别。例如，如果接受编辑的人包括一个或数个生育项目，可将性别定为女性。生育项目包括已生子女、居住在住户中的子女、居住在其他地方的子女、死亡子女和过去 12 个月中的活产子女。另一个可能性是此人可能是住户中另外某人的母亲，从而此人的行号等于住户中另一人母亲的行号。

(f) 在性别代码无效而且此人是配偶的丈夫时

298. 如果根据一个显示丈夫行号的项目，此人是住户中某个另一人的丈夫，就应向性别的登录项分配男性。

(g) 在性别代码无效而且确定性别的信息不充分时

299. 如果编辑完全未用动态推算，就必须分配不明性别的值。不幸的是，这意味着所有的汇总表都将为不明性别的人载列一个或一组额外的纵列或横行。由于性别是一个二进制变量，可以更替分配值，从任一值开始，第二个无效登录项使用相反的性别，并以这种方式继续下去。

(h) 推算性别比率的说明

300. 在利用非流水法时，女性性别的分配较为经常。成年女性是带有生育登录项的唯一性别，因此，她们的选择与随机分布有所偏斜。由于这个原因，如果没有充分的信息，一个没有信息的人更可能是男性而不是女性。因此，重要的是应考虑开发出考虑到两性之间总体比例的推算矩阵。

3. 出生日期和年龄 (P3B)

301. 年龄是最难收集和编辑的特点之一。不过，它也许是最重要的变量，因为几乎所有的人口特点均依据年龄进行分析。编辑年龄要求与其他变量进行广泛的比较。在多数情况下，推算的年龄应当基于有关人员年龄之间的存储之差。如果年龄不能据此推算，就应使用个人记录上的其他变量。编辑也许应当要求进行一系列的推算矩阵，包括按性别、婚姻状况、关系和上学情况推算的年龄；母子之间的年龄差；夫妻之间的年龄差，以及户主与配偶的年龄差。

(a) 年龄与出生日期

302. 结构编辑根据出生日期计算年龄，不过，有益的做法是首先审查年龄与出生日期之间的差。如《原则和建议》所指出的（联合国，1998 年，第 2.88 段），关于年龄的信息可通过获取出生日期（年月日）来获得或直接问及此人上个生日时的年龄来获得。

303. 出生日期产生的信息更加准确，只要情况许可就应使用。如果不知道确切的出生日或甚至出生月，可代之以说明该年的季节。如果人们知道其出生日期，提出出生日期的问题是适当的，此种日期可按照阳历或阴历来确定，或以传统民间文化中固定周期中的名称计算或认定的年数来表示。但是极为重要的是查点人员与答卷人之间应清楚了解出生日期依据的是哪个日历系统。如果有可能，某些答卷人会参照不同于其他答卷人的日历系统回答，调查表上必须规定说明使用的日历系统。查点人员不可取的做法是试图将一个系统的日期转换至另一个系统的日期。所需的转换作为计算机编辑工作的一部分来进行可能最可靠（联合国，1998 年，第 2.89 段）。

304. 由于若干原因，直接问及年龄可能产生不太准确的答案。即使所有答案都基于相同的年龄计算方法，答卷人也不一定了解所要的年龄是上个生日、下个生日还是最近的生日。此外，也可能将年龄四舍五入到以零或五结尾的最近的整数；估计数不一定如此认定；而且故意谎报也不难（联合国，1998 年，第 2.90 段）。

305. 许多国家普查/统计局或者收集出生日期或者收集年龄，但不会二者都收集。如《原则和建议》所指出的（联合国，1998 年），实足年龄非常重要：它用于许多编辑，而且作为许多推算矩阵的一个维数。更重要的是，许多国家政策基于年龄，所以必须作出努力来获取最高质量的年龄报告。不过，基于在理想的情况下，有些年龄也不会报告。因此，必须作出努力确保正确计算年龄，并使之与住户个别成员的其他答案相一致。

(b) 出生日期与年龄的关系

306. 在结构编辑期间，如果年龄不与出生日期分开单独收集，就应计算年龄。个人编辑期间的年龄

编辑将是记录内部和记录之间一致性的一项彻底测试，但第一步是根据出生日期和普查日期计算年龄，重要的是应测试根据出生日期计算的年龄，以便确认它属于普查日期的界限范围内。普查年内普查日期出生的子女将计算为-1 并必须加以纠正。出生日期后查点的婴儿也许应从普查中除去。不过，如果经审查后发现由于查点或处理的关系，出生日期搞错了，就应利用其他变量获得更可靠的年龄估计值。

(c) 在计算的年龄超出上限时

307. 对于 2000 年及以后的普查，多数国家将选择记录出生年的所有 4 位数字。可接受的范围将自 1800 年代某个时候——也许 1880 年及以后——直至普查年。虽然 3 位数便足可使计算机开展工作，但 3 位数字年份的使用可能会把查点员和办事处工作人员搞糊涂。有时候计算的年龄将超出普查界定年龄的上限，并需要作出调整。如果普查在 2000 年，而某个人报告生于 1850 年，计算的年龄 150 岁就可能超出可接受的范围，因而将需要被更改。

(d) 年龄编辑

方框 4. 年龄的编辑和推算

年龄的编辑和推算应如下进行：

- (a) 年龄为空白时分配年龄；
- (b) 检查已婚人员的最低年龄；
- (c) 检查户主的最低年龄；
- (d) 检查父母的最低年龄；和
- (e) 任何其他国家规定的检查。

308. 编辑程序应当检查该人的报告年龄与该人的母亲、父亲或子女的报告年龄是否一致。编辑应规定母亲或父亲的年龄与子女年龄之间的最低岁数差。在推算年龄时，应与有关登录项进行一致性检查，例如，在该地区居住的年数（居住期限）和读完的最高年级（教育程度）。所有此类检查应在年龄变更前或分配推算年龄前进行。

309. 编辑开始时应先检查有效性。如果年龄有效，专家可能希望进行检查，以便查清此人的年龄与

他（她）母亲的年龄（如果此人母亲在住户内）和与此人子女的年龄（如果此人是妇女而且在住户中有子女）是否一致。如果年龄不一致，就应注明此人的年龄并在以后加以变更。

(e) 在户主和配偶存在时的年龄编辑

310. 编辑的下一步是确定配偶是否存在。如果如此，应检查配偶年龄是否有效（至少×岁，这取决于该国规定的最低结婚年龄）。如果年龄不一致而且如果利用动态推算，现在程序将利用从丈夫年龄与妻子年龄差导出的一个推算值。年龄差的变动性小于年龄本身的变动性，所以程序中的推算矩阵将存储丈夫和妻子（来自以前记录的）年龄差。从此人配偶的年龄加减此值以形成计算的年龄。

311. 为确保这一计算的年龄与其他特点相一致，推算矩阵还应包括婚姻状况、居住期限和完成的最高年级学业。如将这些变量排除在外，就可能造成计算的年龄小于此人在当地居住年数或小于教育程度隐含的年数。例如，推算矩阵可能给出年龄为 8 岁，但此人可能已经报告说他们在该地已居住了 10 年。在编辑程序进行就地居住年数编辑时如果没有其他变量，另一推算矩阵将把居住年数从正确的值更改为不正确的值。

(f) 在户主配偶不在但子女在时户主的年龄编辑

312. 在不可能与配偶年龄进行比较以确定户主年龄时，程序可检查关系。如果关系是“户主”，编辑程序就能检查住户有关具有已知正确的年龄的儿子或女儿的其他记录（如有的话）。程序检查儿子或女儿的年龄，并利用类似于上文就夫妻说明的方法的“年龄差”动态推算法计算此人的年龄。如前所述，计算的年龄顾及到居住期限和最高级别的教育程度。通过将在该地区居住的年数和读完的最高年级作为推算矩阵的组成部分，算出的年龄就将与这些变量相一致并将避免明显的错误。

(g) 户主父母在时户主的年龄编辑

313. 当某人不归于上述类别之一时，程序能检索此人在住户中的父母。如果找出此人的父母，依靠利用年龄差的推算矩阵就能计算年龄。子女与父母之

间的年龄差一般比夫妻之间的年龄差大得多。因此，只有在夫妻年龄差的方法失效时程序才适用这种编辑。计算的年龄应顾及教育特点、在当地居住的年数包括已完成的最高年级学业及婚姻状况、生育率和经济活动等。程序应假定，如果某人已结过婚，有子女或报告参加过任何类别的经济活动，他（或她）至少有一个最低限度可接受的年龄。

(h) 户主孙子女存在时他的年龄编辑

314. 当某人不属于上述类别之一时，程序能检索此人在住户中的孙子女。如果找出此人的孙子女，依靠利用年龄差的推算矩阵就能计算年龄。户主与孙子女之间的年龄差比夫妻之间或户主与子女之间的年龄差大得多。因此，只有在夫妻年龄差和户主与子女年龄差的编辑失效的情况下，程序才适用这种编辑。计算的年龄应顾及教育特点包括已完成的最高年级学业、在当地居住的年数、婚姻状况、生育率和经济活动。程序应假定，如果某人已结过婚，有子女或参加任何类型的经济活动，他（或她）至少有一个最低限度可接受的年龄。

(i) 无其他年龄可加利用时户主的年龄编辑

315. 当某人不属于上述类别之一时，程序能检索户主的另一亲属或非亲属。如果找出此人，而且此人有报告的年龄，编辑小组必须决定是否依靠利用年龄差的推算矩阵使用任何可加利用的信息。不过，户主与其他亲属或非亲属之间的这样年龄差非常大，以致于编辑可能决定完全放弃努力，并只是简单利用其他变量进行户主年龄的动态推算。无论如何，只有在夫妻、户主/子女和户主/孙子女的年龄差技术都失效的情况下，程序才适用这种编辑。不管计算年龄如何确定，它都应顾及教育特点包括完成的最高年级学业，以及在当地居住的年数、婚姻状况、生育率和经济活动。如果某人已结过婚，有子女或参加任何类型的经济活动，他（或她）至少有一个最低限度可接受的年龄。

(j) 户主年龄已定时配偶的年龄编辑

316. 配偶的年龄编辑通常与户主的年龄编辑同时进行，因为联合编辑需要两人的信息。然而，如果分开编辑，当配偶的年龄无效或与其他变量不一致

时，就应使用利用与户主的年龄差和其他变量的动态推算矩阵确定配偶年龄的最可靠估计。如前所述，计算的年龄应顾及教育特点包括已完成的最高年级的学业、在当地居住的年数、婚姻状况、生育率和经济活动。程序应当假定，如果某人已结过婚，有子女或参加任何类型的经济活动，他（或她）至少有一个最低限度可接受的年龄。

(k) 在住户中其他已婚夫妇两人之一年龄已知时他们的年龄编辑

317. 编辑应首先确定这一记录是某人已婚的记录。如果是这样，程序能在住户的其他记录中检索此人的配偶。如果未找出配偶，程序进入编辑的下一部分。如果找出配偶，应检查配偶的年龄是否有效（至少×岁，依据该国规定的最低结婚年龄而定）。如果年龄不一致而且如果使用动态推算，程序现将利用从夫妻年龄之差导出的一个特别推算值。年龄差小于年龄本身，所以程序中的推算矩阵将存储夫妻（以前记录的）年龄差。从此人配偶的年龄加减此值以形成计算的年龄。

318. 为确保这一计算的年龄与其他特点相一致，推算矩阵还应包括婚姻状况、居住期限和最高教育程度。如将这些变量除外，就可能造成计算的年龄小于此人在当地居住的年数，或小于教育程度隐含的年数。

(l) 在户主年龄已定时子女的年龄编辑

319. 如果这是户主的儿子和女儿，就能利用户主的年龄、年龄差、居住期限和教育程度等导出计算年龄。计算年龄应再次顾及教育特点包括完成的最高年级学业、在当地居住的年数、婚姻状况、生育率和经济活动。程序应假定，如果某人已结过婚，有子女或参加任何类型的经济活动，他（或她）至少有最低限度可接受的年龄。

(m) 在户主年龄已定时父母的年龄编辑

320. 如果这是户主的父母之一，就能利用户主的年龄、年龄差、居住期限和教育程度等导出计算的年龄。计算年龄应顾及教育特点包括已完成的最高年级的学业、在当地居住的年数、婚姻状况、生育率和

经济活动。程序应假定，如果某人已结过婚，有子女或参加任何类型的经济活动，他（或她）至少有最低限度可接受的年龄。

(n) 户主年龄已定时孙子女的年龄编辑

321. 如果这是户主的孙子女，就能利用户主的年龄、年龄差、居住期限和教育程度等导出计算的年龄。计算年龄再次顾及教育特点包括已完成的最高年级的学业、在当地居住的年数、婚姻状况、生育率和经济活动。程序应假定，如果某人已结过婚，有子女或参加任何类型的经济活动，此人至少 12 岁。

(o) 所有其他人的年龄编辑

322. 对于住户中其他有亲戚关系和无亲戚关系的人，编辑小组必须确定合适的推算矩阵。指导原则将取决于特定的普查或调查，以及取决于国家的社会和经济特点。例如，一个结过婚、有子女或已参加经济活动的人，有可能已至少达到某国规定的最低年龄。根据此种信息，如果利用动态推算，从推算矩阵得到的值不应该低于最低年龄。同样，如果某人上学，受过教育，或能读能写，但不是户主，从未结过婚，而且无经济活动，那么就应将此人列入其年龄小于成年人最低年龄但大于或等于最低上学年龄的年龄组。然后可为小于最低上学年龄的那些人找出推算矩阵值。虽然这项技术不完善，但它限制推算矩阵能取值的范围。

4. 婚姻状况 (P3C)

323. 在《原则和建议》（联合国，1998 年，第 2.96-2.98 段）中，婚姻状况定义为每个人相对于该国婚姻法或习俗的个人状况。拟确定的婚姻状况的类别包括但不限于：(a) 单身（从未结过婚）；(b) 已婚；(c) 丧偶并未再婚；(d) 离婚并未再婚；和 (e) 已婚但分居。在有些国家，(b) 类可能需要增加一个次级类别，即已正式结婚但尚未作为夫妻生活在一起的人。在所有国家中，(e) 类应包括依法和事实上的分居，如有必要可列为单独的次级类别。尽管已分居的夫妻仍可被算为已婚人（因为他们不能再婚），但 (e) 类的两个次级类别不应列入 (b) 类。在有些国家，将有必要考虑符合习俗的结合（根据习

惯法合法而且具有约束力）和法外结合，后者常称事实上的（自愿）结合。

(a) 婚姻状况编辑

324. 编辑小组必须决定普查或调查的合适的初婚最低年龄。初婚的最低年龄（某个×年龄）在一国的不同地区或不同民族群体之间可能有所不同。例如，如果农村人口比城市人口早结婚，编辑小组应包括这一事实。通常情况下，国家普查/统计局在查点前确定最早结婚的年龄，以便只向超过确定年龄的人提问。更年轻的人自动进入“未婚”类别。不过，如向每个人询问婚姻状况，编辑小组就必须为整个人口编定一个编辑程序。

(b) 在不使用动态推算时的婚姻状况分配

325. 虽然只为×岁和更大年龄的人编列婚姻状况表，其中×为初婚的最小年龄，编辑小组必须确定是否编辑和进行多大程序的编辑。如果某国对无效或不一致的答复只使用“未报”或“未知”，那么在发现无效或不一致的登录项时，“未报”的代码应取代不适当的答案。对于低于年龄×的人而言，如果答案“未婚”遗漏，就应进行推算。

(c) 在使用的动态推算时的婚姻状况分配

326. 如果使用动态推算，婚姻状况的编辑应(a) 在登录超出范围时推算一个值，和 (b) 检查报告婚姻状况与关系和年龄之间是否一致。

(d) 配偶应已婚

327. 在关系类别中编码为“配偶”的所有人应编码为已婚。

(e) 已婚夫妻对的配偶

328. 如果人 A 的配偶（人 B）的行号是一个变量，那么人 B 应以人 A 作为配偶；此外，A 和 B 均应为已婚（如第 327 段所提及），而且性别相反。

(f) 如有配偶，户主应为已婚

329. 如果婚姻状况无登录项出现，但是与户主关系的登录项是“户主”，程序应作检查以看清配偶是否存在（通过检查住户其他成员的关系）。如果配

偶存在，程序给户主的婚姻状况分配为“已婚”。

(g) 户主无配偶，有或无子女

330. 如果配偶不存在，而且此人为男性有子女，程序按有子女存在的年龄推算婚姻状况。如果无子女存在，程序可能按无子女的年龄推算婚姻状况。一名男性为户主，但其妻子不在住户内，极有可能已离婚、分居或丧偶。

(h) 如果所有其他方法无效，就推算

331. 对于带超出范围的代码而又不能根据上述测试分配代码的人而言，下一步应检查年龄。如果年龄有小于年龄×的有效登录项，应分配“未婚”。在所有其他情况下，均应利用推算矩阵分配登录项。推算矩阵的设置应按性别和年龄（二维）；按性别、年龄和关系（三维）；或按性别、年龄、关系和已生子女数（四维）进行。编辑小组应再次已经确定了编辑次序，所以在发展推算矩阵时重要的是应记住哪些项目已经编辑和哪些项目未经编辑。如在婚姻状况前只编辑了性别和关系，推算矩阵必须允许其他各项“未报”。

(i) 年轻人年龄与婚姻状况的关系

332. 对于报告除“未婚”以外有效婚姻状况的所有的人，应与年龄作一致性检查。所有已婚的人必须年满×岁或更大，其中×是特定国家允许人们结婚的最低年龄。如果年龄小于×岁或空白，需根据其他有关变量（如已生子女数或经济活动）作进一步的一致性检查。如果这些项目的登录无效，应给婚姻状况分配“未婚”，在所有其他情况下，婚姻状况不应改变。

5. 初婚年龄（P4F）

333. 按照《原则和建议》（联合国，1998年，第2.142段），“初婚日期”包括首次结婚的年月日。在初婚日期难以获得的国家，可取的做法是收集关于结婚年龄的信息和关于多少年前结婚（婚姻存续期间）的信息。包括在内的不仅有契约性初次结婚和事实上的结合，而且有符合惯例的婚姻和宗教婚姻。对于普查时丧偶、分居或离婚的妇女，应获得“初婚的日期/初婚的年龄/初婚解体以来的年数”。关于初婚

解体的信息（如果有关）提供了计算作为处理阶段派生专题的“初婚存续期间”所需的数据。在婚姻存续期间的报告比年龄报告更为可靠的国家中，婚姻存续期间所生子女汇总表产生的生育率估计，比根据按妇女年龄分类的活产子女数据所作的估计更可靠。通过从目前年龄减去结婚年龄或直接从结婚以来度过的年数，就能得出关于婚姻存续期间的数据。

334. 应为所有已婚者输入初婚日期（或依照《原则和建议》只对女性）。程序应检查是否一致：未婚者应无信息，但是已婚的人应有有效的年月日。编辑小组需要决定月日是否必须有效：不利用动态推算的国家可为月日分配“不明”；利用动态推算的国家在月日遗漏时可推算它们。

(a) 未婚人员的婚姻状况应为空白

335. 未结过婚的人不应报告初婚年龄。如果未婚的人出现有效登录项，编辑小组必须决定是改变该人的婚姻状况还是使年龄空白。如果打算改变婚姻状况，只利用“未说明”的国家将适用该代码。利用动态推算的国家也许应利用年龄和性别来获得适当的婚姻状况答案。

(b) 已婚人员应有登录项

336. 就初婚年龄而言，不利用动态推算的国家可以分配“未说明”或“不明”。利用动态推算的国家可以利用其他变量如配偶年龄或配偶之间的年龄差、子女数和去年所生的子女等来确定初婚的适当年份。

6. 生育率：所生子女数（P4A）和健在子女数（P4B）

337. “所生子女数”指活产子女总数，因而除去死胎、流产和堕胎。有时候，人口学家使用“活产子女数”这种表述方式，但是本文将使用术语“所生子女数”。

338. 对本节所载每个专题应收集数据的总体包括15岁（或某个最低限度可接受的年龄）及以上的妇女，不管其婚姻状况如何，也不管诸如曾婚妇女等特定次级分类。在不使用50岁及以上年龄妇女数据的国家，应集中精力只收集15岁至50岁年龄妇女的数据；在调查最近生育率时，有些国家适宜的做法可能是将年龄低限降低

几岁（联合国，1998年，第2.121段）。

(a) 收集的生育率项目

339. 联合国在《人口和住房普查的原则和建议，第一次修订本》（1998年）中建议获取关于3个生育率项目的信息：所生子女数、最近活产子女出生日期和母亲头胎活产时的年龄。对于结婚年龄、日期或存续期间等项目的答案可改进基于所生子女数的生育率估计。还有，许多国家继续收集关于尚存活子女的信息，这特别有助于生育率的追溯性分析。

340. 普查和调查利用国家界定的最低年龄及有时候还利用最高年龄收集关于所有女性生育率的信息。

(b) 生育率编辑的一般原则

341. 生育率编辑的目的是使登录项彼此一致和与年龄一致：

(a) 活产子女总数不可能大于该人的年龄加某个国家界定的最低年龄。见下文关于“生第一胎时年龄”一节，以便编辑确定母亲与最大活产子女之间的最小年龄差；

(b) 活产子女总数不可能大于生活在住房单元内、生活在其他地方的死亡子女数之和。在总数大于部分之和时，编辑小组必须决定哪项优先，以便能进行调整；

(c) 如果对尚存活的子女和死亡子女都收集数据，这些子女的总数不可能大于所生子女数；

(d) 所生子女数不可能小于“过去12个月所生子女数”的登录值；

(e) 依据国家而定，并依据所生子女和尚存活的子女的实际数而定，可将一个推算矩阵用于过去12个月所生子女数项目，以便按年龄和所生子女数分配答案。不过，在出现空白时向过去12个月中所生子女数分配一个值的过程必须十分小心。就多数国家而言，这个项目的空白意味着未生育子女。分配的值可能使数据偏斜；

(f) 有时候，各国按性别收集所生子女数、健在

子女数和其他生育率项目。在这些情况下，本文介绍的编辑以总量工作，但是各国可能希望增加检查以说明可加利用的补充信息。这些增加的检查包括弄清所生的男性子女数是健在的男性子女数和死亡男性子女数之和，以及所生的女性子女数是健在的女性子女数与死亡的女性子女数之和。关于不按性别区分的子女的编辑，在和不等部分相加总数时需要采取适当的行动。

(c) 所生子女与健在子女之间的关系

342. 所生子女和健在子女的数据用于间接估计生育率和死亡率。普查或调查的结果按女性的单年或5年的年龄组编排。各种算法获得不变或变化的死亡率估计。不过，为了获取最佳结果，编辑小组确定可用数据的适当编辑时必须小心谨慎。

343. 开发通用编辑法的部分问题是不同国家要求不同类型的信息。例如，不同国家收集下列各集信息：

(a) 仅所生子女数

(b) 所生子女数和健在子女数（两性合在一起或分开）

(c) 所生子女数、健在子女数和死亡子女数（两性合在一起或分开）

(d) 所生子女数、居住在家子女数、居住他处子女数和死亡子女数（两性合在一起或分开）。

(d) 只报告所生子女数时编辑

344. 如果国家不使用动态推算，“所生子女数”的无效或遗漏值应分配“不明”。在使用动态推算的国家，专家们必须决定他们是否想将动态推算用于所有项目。如果专家们使用这种方法，所生子女数可以根据女性单一年龄或至少一个其他特点来获取。也可以仅仅利用母亲单一年龄的单一维数阵列。其他特点可能有诸如教育程度或宗教等项目。因为已知在许多国家，不同的教育程度或不同的宗教信仰存在着不同的生育率。

(e) 在报告所生子女数和健在子女数时的编辑

345. 如果“所生子女数”和“健在子女数”两项的答案都有，程序需要确定下列方面：

- (a) 各项目内部是否一致（所生子女数等于或大于健在子女数吗？）；
- (b) 每个项目与女性的年龄是否一致；
- (c) “所生子女数”与“去年所生子女数”（或最近一次生育）是否一致，如果收集的话。

346. 人口学家利用所生子女数和健在子女数的项目来获取间接死亡率估计数。由于这一点，编辑必须保持这两个项目之间的关系。有时候只有其中一个项目报告，另一个不明。方便的编辑将是假定所生子女无死亡并且两项相同。不过，在使两项相同的过程中，间接死亡率估计必须顾及出生后可能死亡的婴儿，从而低估死亡率和高估预期寿命。如果普查或调查中此类情况极少出现，则损害不大。但如果此种情况较频繁地出现，如在使用间接法的国家中预期的那样，效应可能是很大的。图 27 给出一个例子。

图 27. 拥有生育信息的住户的说明

人	关系	性别	年龄	所生子女数	健在子女数
1	户主	1	60		
2	配偶	2	60	5	99
3	女儿	2	40	3	3
4	孙女	2	20	1	1
5	孙女	2	18	0	0
6	孙女	2	1		

注：99=遗漏或无效

347. 在本例中，配偶报告所生子女数共 5 名，但不知由于什么原因未记录健在子女数。答卷人或查点人员未报告数值，或数据录入员错键了信息。许多国家开发一种编辑法，根据所生子女数将值“5”分配给健在子女数。然后这种做法使数据偏斜。

348. 实际上，该值无需作任何变动。不使用动态推算的那些国家可选择将“不明”值留着不动。当然，这种决定也产生偏斜，因为编辑决定“不明”和“已知”答案在制表时分布相同。如果某个国家需要关于所生子女数和健在子女数的数据，以确定死亡率的间接估计，这个国家在数据报告方面也许存在问题。在这种情况下，使数据保持不明状态可能会使最后分析发生偏斜。不能将所生子女数或健在子女数不明的女性用于确定死亡率估计，因为所生子女数与健在子女数之差无法确定。

349. 利用动态推算法的国家应当考虑至少根据其他生育率项目和女性年龄来确定遗漏的信息。如果存在有关女性年龄、所生子女数和健在子女数的有

效信息，推算矩阵可以更新，并在项目遗漏时可加利用。在所生子女数遗漏时，推算矩阵将拥有女性年龄和所生子女数。

350. 此外，在开发推算矩阵时，重要的是应记住所生子女数和健在子女数必须符合母亲与最大子女之间的年龄差（如果存在此种信息），并符合特定年龄母亲的所生子女总数。

351. 例如，所生子女数推算数与母亲年龄之差必须至少为 12。然后，利用女性 5 岁年龄组的推算矩阵在某些情况下将几乎肯定推算出不相容的信息。

352. 图 28 所附的推算矩阵上部显示女性年龄，底侧列示所生子女总数。登录项是健在子女的推算值。有时候答案将是适当的，但有时候并不适当。如果程序遇到一个生有 5 名子女的 19 岁的女性，5 名健在子女的值也许应通过年龄差标准（根据健在子女和报告的年龄，年龄差为 15）。不过，对于 15 岁的人来说，无论 5 名所生子女（年龄差为 10）还是 4 名健在子女（年龄差为 11）都将是不可接受的。

图 28. 在年龄和所生子女数有效时确定健在子女的初始值

所生子女数	年 龄												
	15	16	17	18	19	20	21	22	23	24	25-29	30-34	35+
0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	0	0	0
2		2	2	2	2	2	2	2	2	2	1	1	1
3			3	3	3	3	3	3	3	3	2	2	2
4					4	4	4	4	4	4	3	3	3
5					5	5	5	5	5	5	4	4	4

353. 在单一年的年龄适用于年轻女性时，推算矩阵更有效。此时只有该特定年龄的有效年龄差答案才输入推算矩阵，而且只有有效的答案才能从推算矩阵取出。

(f) 在报告所生子女数、健在子女数和死亡子女数时的编辑

354. “所生子女数”是“健在子女数”与“死亡子女数”之和。任何不一致可按下文解释的方法解决。

(一) 在所有 3 个项目都报告时

355. 如果存在所有 3 种信息，程序需要确定：

(a) 3 个项目内部是否一致（所生子女数是健在子女数与死亡子女数之和吗？）；

(b) 其中每项都与女性的年龄是否一致；

(c) 所生子女数与去年所生子女数（或最近的生育）是否一致，如果收集的话。

356. 如果所有项目都一致，编辑结束。然而，如果任何一项不一致，编辑必须加以解决。3 项可能内部不一致：例如，一名女性可能生有 5 名子女，但只有两名健在和两名死亡。编辑小组应当决定哪个变量优先于其他变量。在许多情况下，虽然女性可能忘记死亡的确切数，但她可能记住所生子女数。然后，

编辑小组可以选择接受所生子女数和健在子女数，二者相减得出死亡子女新的相一致的值。

(二) 在两个项目报告时

357. 由于类别所生子女数（CEB）是健在子女数（CS）与死亡子女数（CD）之和，如果其中任何两项信息可以获取，计算机程序能够确定第三个变量：

如果 CEB 和 CS 已知， $CD=CEB-CS$ 。

如果 CS 和 CD 已知， $CEB=CS+CD$ 。

如果 CEB 和 CD 已知， $CS=CEB-CD$ 。

这些测试通常将首先进行。一旦程序确定所有 3 项信息均有效和一致，编辑结束。

(三) 在 1 个项目报告时

358. 在 3 个项目中只有一个已知时，如果国家不利用动态推算，应将其他两项定为“不明”。如果国家利用动态推算，编辑小组需要确定至少再得到一个项目的方法，然后应通过加减法求得第三项。可以利用二维矩阵，根据第一项和女性单一年的年龄来获得第二个生育率值。例如，如果所生子女数已知，健在子女数可从推算矩阵获得，就如上文所述，然后应通过减法求得死亡子女数。同样，如果健在子女数已知，所生子女数从女性单一年的年龄和健在子女数的推算矩阵获得，而且死亡子女数通过减法求得。

(四) 在无一次报告时

359. 在3项中任何一项都得不到时,编辑小组必须就如何处理作出决定。如果国家不利用动态编辑,所有3项都应变为“不明”,而且不得应用于死亡率或生育率的间接估计法。在使用动态推算的国家,专家们必须决定他们是否想将动态推算用于所有项目。

360. 如果专家们决定利用动态推算,所生子女数可根据女性单一年的年龄和至少另一个特点来获取。也可能只利用母亲单一年的年龄的单维阵列。其他特点可为诸如教育程度或宗教等项目。

361. 一旦第一项确定,如果获取第二个生育率项目,就可遵循上文就只有一项报告时如何编辑所概述的步骤。然后,可从头两项获取第三项。3个项目应相容,因为只有所有3项都相容时才应更新推算矩阵。获取的生育率还应与该地区其他女性相容,因为要使用来自这些女性的信息更新矩阵。

(g) 在报告所生子女数、居住在家子女数、居住他处子女数和死亡子女数时的编辑

(一) 在所有4项都报告时

362. 如果所有4项信息都存在,程序需要确定:

(a) 所有4项内部是否一致,以便所生子女数为居住在家子女数、居住在外子女数和死亡子女数之和;

(b) 4项中每项是否与女性年龄相一致;

(c) “所生子女数”与“去年子女数”(或最近的生育)是否一致,如果收集的话。

363. 如果各项都一致,编辑结束。不过,如果任何一项不一致,编辑就需要解决不一致问题。如上述3项的情况一样,所有4项内部可能不一致。编辑小组应再次决定哪个变量优先于其他变量。在许多情况下,女性答卷人可能记住她所生的所有子女,但可能忘记移居他处的某些人或死亡的确切数。然后,编辑小组可能选择接受所生子女数和健在子女数(居住他处子女数和居住在家子女数之和),并通过减法求

出其他变量的新的相一致的值。编辑小组可能需要为事件的各种组合开发算法。

(二) 在其中3项报告时

364. 所生子女数(CEB)是居住在家子女数(C LH)、居住在外子女数(C LA)和死亡子女数(C D)之和。如果其中任何3项可以获取,计算机程序可以确定第4个变量:

如果 CEB、CLH 和 CLA 已知,
 $CD=CEB-CLH-CLA$ 。

如果 CLH、CLA 和 CD 已知,
 $CEB=CLH+CLA+CD$ 。

如果 CEB、CLA 和 CD 已知,
 $CLA=CEB-CLH-CD$ 。

如果 CEB、CLA 和 CD 已知,
 $CLH=CEB-CLA-CD$ 。

(三) 在其中两项报告时

365. 如果只有其中两项已知,编辑小组就必须决定下一步怎么办。如前所述,在许多国家,妇女并不报告死亡子女的确切数。最有可能略去的另一项是关于居住在住房单元外子女的信息,这通常也不能直接获得。因此,在编辑调查表、实施查点和进行处理时必须谨慎小心,以便获得所有生育率项目的优质数据。

366. 将住房单元内子女相加,就能求出居住在单元内子女(C LA)的数据。只要单元内只有一名女性有适当的关系,简单的计算就应给出居住在单元内的子女数。如果一名以上女性具有这种关系,根据在数据收集中子女将紧随母亲的假定,仍能利用编辑程序。在所有其他办法无效时,利用动态推算的国家可根据母亲年龄和其他已知变量之一推算居住在单元内的子女数。(见下文根据其他项目和母亲年龄推算个别生育率项目的一般规则。)重要的是,凡有可能均应利用女性单一年的年龄,以及所生子女、居住在单元内子女、居住他处子女或死亡子女的单列数。

367. 作为例子,所生子女数和死亡子女数可能是有效登录项,但是居住在住户内子女数和居住他处子女数可能是无效登录项。在这种情况下,将与母亲有适当关系的子女相加就能求出居住在家子女数(假

定母亲是户主)。然后 4 项中的 3 项将可获得, 再通过减法: $CLA = CEB - CLH - CD$ 就能确定第 4 项即居住他处的子女数。

368. 不过, 在只有两项已知时, 更有可能的情况是将需要记录所生子女数和居住在家子女数。女性通常愿意报告所生子女数, 而且关于居住在家子女的信息通常通过观察或在查点时与答卷人合作就能获得, 但是对于居住他处子女或死亡子女, 这些解决方案就用不上了。然后, 编辑就能利用女性年龄和所生子女数 (CEB) 推算矩阵, 或甚至更好的办法是利用女性年龄、所生子女数 (CEB) 和居住在家子女数 (CLA) 推算矩阵。这些变量将从具有居住他处子女数 (CLA) 相同特点的类似女性获取信息。

369. 有的国家只利用女性年龄和所生子女数 (CEB) 的两维矩阵, 也不包括第三维即居住在家子女数 (CLH), 它们有可能获得与另两个值不相容的

居住他处子女数 (CLA) 的值。例如, 如果女性年龄为 25 而 CEB 是 5, 就有可能从居住他处子女数推算矩阵获得值 3。如果居住在家子女数的值是 2, 那么编辑无问题。死亡子女数的值应为 0, 而且各生育率项目应为: $CEB=5, CLH=2, CLA=3, CD=0$ 。

370. 不过, 居住在家子女数的值实际可能是 4, 只有女性年龄和所生子女数用来确定居住他处子女数的值。于是居住他处子女数的值 3 将在各项之间造成不相容。所生子女数的值 (5) 将小于存活子女数之和 (在家 4 人和在他处 3 人, 或合计 7 人)。因此, 应利用三维矩阵: 如果 CEB 为 5 和 CLH 为 4, 推算矩阵中居住他处子女数的值为 1 (而且通过减法应确定死亡子女数的值为 0)。或者, 推算矩阵中居住他处子女数的值为 0 (而且通过减法应确定死亡子女数的值为 0)。如图 29 那样, 可能需要为其他各对已知信息发展类似的推算矩阵。

图 29. 拟为各对已知信息发展的抽样推算矩阵

如果这些已知.....		为其中之一利用推算矩阵 (并然后加或减)	
所生子女数	居住在家子女数	居住他处子女数	死亡子女数
所生子女数	居住他处子女数	居住在家子女数	死亡子女数
所生子女数	死亡子女数	居住在家子女数	居住他处子女数
居住在家子女数	居住他处子女数	所生子女数	死亡子女数
居住在家子女数	死亡子女数	所生子女数	居住他处子女数
居住他处子女数	死亡子女数	所生子女数	居住在家子女数

371. 在每种情况下, 4 项中有两项可以获取。第 3 项采用动态推算获取, 而第 4 项采用加减法求得。编辑小组必须决定, 根据文化背景情况, 哪条是最佳路径。

(四) 在只有一项报告时

372. 在 4 项中只有 1 项为已知时, 情况就更成问题。各国必须决定在这么多信息遗漏时它们应怎么办。如果利用动态推算, 如上文所述, 第一个推算矩阵将利用一个诸如女性单一年的年龄这样的项目和一个已知项, 以产生一个两维推算矩阵用于任何其他

一项。一旦两项确定, 另两项按定义仍是未加项。因此, 继续将动态推算用于第三项不应造成与其他各项不相容, 因为它们是未知项。利用上文讨论的有关两个已知项和两个未知项的方案来获得第三项。然后通过减法求出第四项。所有 4 项应相容。见第 376 段关于只有一项已知时可能的解决方案。

(五) 在无一项报告时

373. 4 项中无 1 项可加利用时, 编辑小组必须决定在无任何已知项的情况下应如何办。如果国家不利用动态推算, 所有项目应变为“不明”而且不应用

于间接估计死亡率或生育率的方法。在确定利用动态推算的国家，专家们必须决定他们是否想将推算法用于所有项目。见第 376 段以了解在没有任何已知项时的另一种解决方案。

374. 如果专家们决定利用动态推算，所生子女数的值可以根据女性单一年的年龄和至少另一个特点来获得。也可只利用母亲单一年的年龄的单维阵列。其他特点可以是教育程度或宗教等项，因为据了解在许多国家，各种教育程度或宗教信仰有着不同的生育率。

375. 一旦第一项确定，就能利用上文只有一项已知时使用的方法求出第二个生育率项目。然后可从头两项求得第三项，而且采用减法求得第四项。4 项应相容，因为只有所有项目都相容时才应更新推算矩阵。求得的生育率也应与该地区其他女性相容，因为要使用来自这些女性的信息更新推算矩阵。

(h) 所有生育率项目单一提供源的重要性

376. 凡有可能，非常重要的一点是，在不了解任何情况时从同一名妇女推算所有项目。为了确保所有信息均来自同一女性，可能需要发展利用所有生育率信息的推算矩阵。在这种情况下，只有在编辑程序确定所有生育率项目一致时才能更新推算矩阵。

7. 生育率：最后一个活产儿的出生日期 (P4C)

377. 利用关于最后一个活产儿出生日期（年月日）及其性别的信息来估计目前的生育率。其后，在处理阶段，可得出普查日期前 12 个月内的活产子女人数，作为过去 12 个月活产人数的估计。对于估计目前年龄性别生育率和其他生育率计量，这个办法所提供的数据比妇女在普查日期前 12 个月内的生育数信息更为准确（联合国，1998 年，第 2.134 段）。

378. 应当注意到，关于最后一个活产儿出生日期的信息并未提供关于 12 个月期间活产子女总数的数据。即使所报最后一个活产儿数据没有错，这个项目只确定在此 12 个月期间至少有一活产儿的妇女人数，而不是出生人数。少部分妇女会在一年内生一个以上孩子（联合国，1998 年，第 2.135 段）。

379. 这项信息仅需向据报一生中至少有一个活产儿的 15 至 50 岁妇女收集。此外，这项信息应向已按性别收集到其所生子女数据的各类婚姻状况妇女收集。如果仅向抽样妇女收集一生的生育率数据，则关于目前生育率的信息也应向同一批抽样妇女收集（联合国，1998 年，第 2.136 段）。

380. 应将下列编辑包括在编辑程序中。应为国家规定的最低年龄至国家规定的最高年龄之间的所有女性输入最后一个子女的出生日期。程序应检查是否对应。例如，男性和未在选定年龄组的女性应无信息出现。还有，选定年龄组中已产子女数大于零的女性应有最后生育的有效年月日。编辑小组需要决定月日是否必须有效：利用动态推算的编辑小组在月日遗漏时可推算它们；不利用动态推算的小组将为月日分配“不明”。

381. 如果信息遗漏或无效，对于最后一个子女的出生年份而言，不利用动态推算的国家可以分配“未申报”或“不明”。利用动态推算的国家可利用其他变量如年龄和所生子女数等获得最后子女的出生日期。

8. 生育率：第一胎时的年龄 (P4G)

382. 生第一胎活产儿时母亲的年龄被用来根据第一胎数据安排间接估计生育率并提供关于开始生育的信息。如果将此专题包括在普查中，应向每名至少已有一个活产子女的妇女获取信息（联合国，1998 年，第 2.143 段）。

383. 生育第一胎年龄的确定或者直接通过一个明示的项目“生育第一胎的年龄”进行，或者在最大子女年龄已知情况下依靠母亲目前年龄与最大子女年龄之间的年龄差进行。国家规定的生育子女的最早年龄并不是生理上最早的生育年龄。例如，如果国家规定的第一胎生育最早可接受的年龄是 13 岁，答卷人可能报告或查点员可能记录某人的生育年龄是 11 或 12 岁。然后，编辑小组必须决定是改变最早可接受的年龄，还是改变母亲的年龄或她生第一胎的年龄（利用子女的年龄或她的年龄，依据用来确定年龄差的变量而定）。同样，编辑小组必须决定什么样的

“最高年龄”是第一胎生育的最大年龄。虽然女性进入 50 岁以后仍有能力生育子女，但这种事情并不多见，而且为了纠正错误，编辑必须确定范围以外的情况是否真实。

384. 重要的是应该记住，第一胎生育的最早或最迟年龄（以及母亲与居住在住户内最大子女的年龄差）必须符合国家的习俗和传统。专家们必须确定何时一个值是干扰，而不是合法的第一胎生育年龄。在规则确定后，专家们就必须决定如何纠正问题。如果不利用动态推算，程序应当分配“不明”。在利用动态推算时，它可根据年龄类似和所生子女数类似的其他女性确定第一胎生育年龄。确定推算矩阵的专家们可能希望考虑到各种有关的因素如城乡常住地（如果两个地区生育率不同）、女性参加劳动力的情况（虽然目前劳动力状况与她生育第一胎时的状况未必相同）和教育程度。

9. 死亡率 (P4D)

385. 对于那些从公民登记中得不到满意和持续的死亡情况统计的国家，它们用过去 12 个月死亡情况的信息来估计按性别和年龄分列的死亡率水平和模式。为了使根据这一项目作出的估计可靠，必须尽量准确地全面报道过去 12 个月按年龄和性别分列的死亡情况。过去数十年来死亡率问题已普遍列入普查调查表内，这大大有助于利用间接程序估计成人死亡率（联合国，1998 年，第 2.137 段）。

386. 关于死亡率，最理想的是收集每户在普查日期前 12 个月期间的死者总数。如果不可能获取过去 12 个月期间关于死亡情况的信息，建议至少收集 1 岁以下儿童的死亡数据。对于所报的每名死者，也应收集其姓名、年龄、性别和死亡日期（年月日）。应注意向调查对象清楚地说明基准时期，以避免因误解而出错。例如，每一个国家可以节庆或历史性日期界定确切基准时期（联合国，1998 年，第 2.138 段）。

387. 《原则和建议》（联合国，1998 年）提议收集普查前一年内死者的姓名、年龄、性别和死亡年月日。不利用动态推算的国家，在其中每个变量无效时，可向其分配“不明”。利用动态推算的国家可将

年龄（年龄组内的）、性别和死亡年份作为其他变量推算矩阵的维数。实际的估算矩阵也许因国而异，而且编辑小组将需要通力合作以发展适当的推算矩阵。

10. 丧母和丧父情况 (P4E) 和母亲的行号

388. 为收集关于丧母或丧父的信息，应提出以下两个直接问题：(a) 住户内被查点之人的生母在普查时是否仍然健在；(b) 住户内被查点之人的生父在普查时是否仍然健在。调查应能得到关于生身父母的信息。因此，应当注意排除收养和寄养父母。由于回答丧母丧父情况问题的健在子女通常不止一个，因此必须编拟问题，以克服兄弟姐妹报告父母情况时的重复问题。为此，还应增提两个问题：(c) 答问者是否母亲最年长的健在子女；和(d)答问者是否父亲最年长的健在子女。制表时只应参照最年长的健在子女进行（联合国，1998 年，第 2.140 段）。

389. “母亲活着”和“母亲行号”的编辑是相互关联的，应合在一起进行。对于就母亲活着报告非“是”的人，应检查母亲行号是否为有效的登录项；如果出现有效登录项，应检查母亲行号以看清它是否是 00，或它是否等于年龄大于或等于 12 岁的女性的行号。如果其中任一种情况都不真实，程序假定此人有母亲并向母亲的人口动态分配是。如果母亲行号中的登录项无效，而且母亲活着编码为“否”或“不知道”，就应消除母亲行号中的登录项。在所有其他情况下，应把“不知道”的代码分配给母亲活着，而且行号中的任何登录都应消除。

390. 如果母亲活着，国家可能选择不为报告“否”或“不知道”的人的母亲行号进行编辑。在所有其他情况下，可检查行号是否一致，或利用人的关系和被报告为母亲的人的行号、性别、关系和年龄进行推算。如果存在不一致或母亲不能确定，可分配“居住他处”的代码。

B. 迁移特点

391. 由于自然增长（生育率和死亡率）和净迁移，一国的人口统计数据随时间发生变化。可分长期迁移（自出生以来）或短期迁移，用以前居住地和居住时间或按以前订定的时点计量。由于这些项目经常

相互关联，类似于就基本人口变量说明的联合编辑可能适用于部分国家。如果使用自顶向下法，编辑的次序就变得重要了，因为某些项目必须先于另一些项目编辑。

392. 迁移项目所需的代码经常比其他项目的详细，因为地理单位可能需要用于计划和政策目的。规划新学校或新卫生诊所的工作人员可能需要关于小区的详尽信息。还有，国内和国外的不同地方也可能需要不同的编码方案和不同的编辑。

393. 应检查关于出生国和当地居住年数的数据是否一致，因为两个项目之间存在着明显的关系。此外，住户各个成员的答案之间也存在某些合理的关系。例如，如果某子女当地居住年数无答案，就能从母亲的答案推算，而且编辑程序将检查推算的值不超过子女的年龄。

1. 出生地 (P1C)

394. 出生地首先指人们出生的国家。应该指出，出生国不一定与公民资格有关，它是另一个专题（见第405-409段）。对于出生在普查所在国的人（当地人）来说，出生地的概念也包括该人出生时期母亲居住国的特定类型的地理单位。不过，在有些国家，当地人出生地定义为实际出生所在的地理单位。各国应解释普查使用哪种定义（联合国，1998年，第2.29段）。

(a) 出生国和当地居住年数登录项的关系

395. 可以检查出生地 (P1C) 和持续居住时间 (P1D) 的登录项是否一致，因为两项之间存在着牢固的关系。还有，住户的不同成员之间也存在关系，而且可以根据家庭其他成员就有关人员是否迁移作出假定。

(b) 为出生地的无效登录项分配“不明”

396. 如果一国选择不利用动态推算，出生地的任何无效答案应变为“不明”。除非编码出错，否则一国通常不应编辑家庭成员之间或有关地理区域的不一致答案。

(c) 利用出生地的静态推算

397. 只有出生国登录项超出范围才应加以变更。如果当地居住年数的代码是“始终”，就应分配

国家代码。如果登录项不是“始终”，就可利用前一个人的信息。例如，如果前一人是母亲，母亲在当地的居住年数可与此人年龄作比较。如果母亲的登录项大于或等于此人岁数，就应分配“本国”代码；否则，应分配“母亲出生国”。如果根据母亲的登录项不能分配出生国，可以同样的方法利用其他有关个人的登录项。如经这些测试后仍不能分配登录项，可将出生国分配为“不明”。

(d) 利用出生地的动态推算

398. 如前所述，只有超出范围才应变更出生国的登录项。如果当地居住年数的登录项为“始终”，应将“本国”代码分配给出生国。如果登录项不是“始终”，就应研究住户内其他人员的信息作为弄清此人出生国的线索。

(e) 在一人母亲存在时分配出生地

399. 如果出生国为空白或无效，而且持续居住时间不是“始终”，就可检索此人的母亲。如果发现母亲在住户内，审议母亲持续居住时间的登录项。如果她当地居住年数的登录项为“始终”，可将此人的出生国分配为“本国”。如果此人母亲不始终居住在当地，但此人年龄小于或等于母亲在当地居住的年数，程序也能将“本国”分配给出生国。如果此人年龄大于母亲在当地居住年数，而且母亲出生国有效，将此人的出生国分配为与母亲相同的出生国。

(f) 分配户主子女的出生地

400. 如果此人母亲不在住户内，但此人是户主的儿子或女儿，那么为了获得出生地，可以利用户主记录上的信息进行数项检查。如果户主在当地居住年数的登录项是“始终”，程序应将“本国”作为出生国分配至此人的记录。如果户主在当地居住年数的登录项不是“始终”，但此人的年龄小于或等于户主在当地居住的年数，程序也应将“本国”分配为此人的出生国。不过，如果此人年龄大于户主在当地所住年数，程序应分配户主的出生国，如果它有出生国的有效代码。

(g) 分配非户主子女的出生地

401. 可以进行颇不相同的推算，这取决于某人

是大于还是小于该国编辑小组确定的某个给定年龄（X岁）。如果此人小于X岁，应从一个不满X岁的子女的前第一份记录凭年龄和性别推算出生国。

(h) 分配有丈夫成年女性的出生地

402. 如果此人年满X岁或更大而且为女性，程序就应进行检查以确认她在住户内是否有丈夫。如果该妇女有丈夫，而且他有出生国的有效代码，程序应将他的出生国代码分配至她的记录。如果丈夫无有效的出生国代码，就应观察他在当地居住年数的登录项。如果他当地居住年数编为“始终”，妇女的出生国应分配“本国”。如果丈夫在当地居住年数不是“始终”，那么就应凭年龄和性别推算妇女的出生国。

(i) 分配无丈夫成年女性的出生地

403. 超过编辑小组规定的某个最低年龄的妇女在住户中没有丈夫，但她可能是住户内子女的母亲。在这种情况下，程序应检索她最年长的子女。如果不能找出该子女，程序能凭年龄和性别推算出生国。如果子女拥有有效的出生国代码，而且母亲报告的当地居住年数大于子女岁数，程序应凭年龄和性别推算出生国。但是如果母亲的当地居住年数小于或等于子女岁数，程序应给她分配子女的出生国。

(j) 分配男性出生地

404. 为获得一男性出生地，编辑程序可设法找出他的妻子，或如果他是户主，程序应设法找出他的子女。首先，程序会试找出该男子的妻子。如果找到她，而且他在当地居住年数小于或等于她的年数，将妻子的出生国分配给该男子的记录。如果男子在当地居住的年数大于妻子居住年数，应凭年龄和性别利用推算矩阵推算出生国。在男子是家庭户主，在住户中有儿子或女儿，而且在当地居住的年数等于或小于子女年龄时，程序应分配与子女相同的出生国。如果他在当地的时间大于子女年龄，程序应凭年龄进行推算。

2. 公民资格 (P3D)

405. 应收集有关公民资格的信息，从而可将人口分为(a) 出生公民(b) 通过声明、选择、婚姻或其

他手段归化的公民；和(c) 外国人。此外，应收集有关外国人公民资格的信息。必须按此来记录公民资格国，而不是将其作为表示公民身份的修饰词，因为其中有些修饰词与说明族裔群体的修饰词是相同的。务必十分详细地将公民资格国的信息编码，以便能个别地查明在国内的外国人所代表的各个公民资格国。为了编码，建议各国使用《供统计使用的国家或地区的标准编码》（联合国，1999年）中所列的数字编码系统。使用标准编码按公民资格对外国人口进行分类将使这类数据更加有用，并使各国能够在国际上交换有关外国人口信息。如果一国决定公民资格国并成各大组，建议采用上述出版物中所确定的区域和次区域标准分类（联合国，1998年，第2.105段）。

(a) 公民资格编辑

406. 公民资格取决于每个国家的定义。在多数国家但并不是所有国家，出生于该国的人凭出生自动成为公民。因此，编辑时应观察出生地与公民资格之间的关系，而且可能需要将“出生公民”分配给该国出生的人。

(b) 民族/种族与公民资格的关系

407. 有些国家也收集“民族”或“种族”，它们可能提供补充信息用于确定公民资格，特别是收集的答案无效时。就许多国家而言，第一代和第二代移民的族裔与公民资格应几乎完全一致。对于国际移民历史悠久的国家而言，这一特点的价值可能较低，但仍可与其他变量一并考虑。

(c) 归化与公民资格的关系

408. 在发生归化的国家，普查项目不一定会覆盖归化的要求。例如，如果需要有一个居住期，就能利用“持续居住时间”的项目测试归化期是否满足。然后，如果某人生在国外而且有一个无效或不一致的公民资格答案，编辑小组可选择为公民资格分配“归化”。未满足归化的持续居住时间要求的其他人，将利用非流水推算法分配为“外国人”。

(d) 持续居住时间与公民资格的关系

409. “持续居住时间”项目不一定出现在调查

表上，或在确定公民资格方面可能模棱两可，或者编辑小组可能选择不用它。那么，如果公民资格的值无效或与出生地不一致，在不利用动态推算时应分配“不明”。为无效值选择动态估算的国家至少应利用两个特点（而且其中一个也许应是出生地），从该地区类似的人获取“已知”信息。

3. 持续居住时间 (P1D)

410. 持续居住时间指截至普查之日的整段期间，以整年计算，在此期间每个人(a) 在普查之时已经住在为其常居住的地区；和(b) 住在该地区所在的主要行政单位或较小的行政单位（联合国，1998年，第2.35段）。

(a) 持续居住时间的编辑

411. 像出生国一样，在编制人口流动性的统计资料时，持续居住时间是重要条件。在有些情况下，人口的某个分群体与全国相比其流动性可能大得多。本项目的编辑考虑到该人的出生地和住户其他成员的答案。“持续居住时间”应与“以前居住地”或“过去某个订定日期的居住地”结合起来编辑。

(b) 事实上/法律上的居住地和持续时间

412. 普查不论是事实上的普查还是法律上的普查都可能影响编辑。由于法律上的普查收集常居地的信息，而事实上的普查对在普查之夜居住地的人查点，因此持续居住时间的编辑可能获得互不相同的信息。此外，代码和编辑必须顾及“始终”居住在当地或“从未离开”的人。就这些个人而言，编辑程序应跳过一致性编辑和其他编辑。

(c) 年龄与持续居住时间的关系

413. 编辑的第一部分应检查年龄与出生地之间是否一致，并应检查居住在地区或行政单位年数的有效登录项。一人居住在某个地区或行政单位的年数不能大于该人的年龄。此外，出生在国外的人不可能始终居住在该地区或行政单位。程序应将“始终”分配给居住在当地或行政单位的年数，如果居住在当地或行政单位的年数大于岁数而且出生国是本国。如果居住在当地或行政单位的年数大于岁数但出生国不

是本国，该人的年龄应分配为居住在当地或行政单位的年数。在此种情况下，假定虽然出生在本国之外，但此人在不足1岁时移居当地或行政单位。

(d) 出生地与持续时间的关系

414. 如属超出范围的登录项，应利用与出生地相同的测试。应检索以前有关的人（母亲、户主、丈夫、子女）。应根据查到的信息进行推算。不过，分配值前，它必须与其记录接受编辑的人的年龄和出生地一致。

(e) 对于始终居住此间的人

415. 如果有关某人居住在当地或行政单位多少年数的答案是“始终”，但出生国不是“本国”，编辑小组可能希望将该人的年龄分配给在当地或行政单位持续居住的时间。专家们将会假定，虽然此人出生于本国之外，但在不足1岁时就移居当地或行政单位。编辑的下一部分将检查居住在当地或行政单位年数的登录项是否有效，由于某人居住在当地或行政单位的时间不能大于此人年龄，在这种情况下，将把年龄分配给在当地或行政单位的年数。

(f) 某人持续时间来自母亲的持续时间

416. 如果该类别没有有效的代码，程序能够进行记录间的检查，在住户中检索此人的母亲。如果找到母亲的记录，则能够提供有助于分配遗漏值的信息。如果此人母亲始终居住在当地或行政单位，而且她的出生国是“本国”（就如应该的那样），程序将把“始终”分配给此人在当地或行政单位居住年数类别。如果母亲的出生国不是“本国”，即使当地或行政单位居住年数的登录项是“始终”，这也表明母亲的类别有错。然后程序将无视母亲的出生国并将年龄分配给当地或行政单位持续居住时间。如果母亲在当地或行政单位居住年数的登录项不是“始终”但是是一个有效的代码，而且此人年龄小于母亲在当地或行政单位居住年数，编辑将回头检查母亲的出生国。如果母亲的出生国是“本国”，程序将把此人的年龄分配给在当地或行政单位居住的年数。不过，如果某人年龄等于或大于母亲在当地或行政单位居住的年数，程序将把母亲在当地或行政单位的年数分配给此人

在当地或行政单位的年数。

(g) 某人持续时间来自子女的持续时间

417. 如果有关人员是子女（儿子或女儿），编辑程序检查户主记录寻找可能的信息，来协助分配关于持续居住时间的遗漏数据的值。如果户主出生于“本国”而且始终居住在当地或行政单位，程序将把“始终”分配给子女在当地或行政单位的年数。如果户主始终居住在当地或行政单位但不出生于“本国”，将把子女的年龄分配给当地或行政单位居住年数。如果户主在当地或行政单位居住年数的登录项不是“始终”但为有效代码，要是与接受编辑子女的记录中的年龄一致，这一信息可加利用。如果子女年龄等于或大于在当地或行政单位居住的年数，程序将把户主在当地或行政单位的年数作为儿子或女儿在当地或行政单位的年数。如果子女的年龄小于户主在当地或行政单位的年数，程序将依据户主的出生国分配一个值。如果户主生于“本国”，值将为“始终”；如果不是，程序将把儿子或女儿的年龄分配给在当地或行政单位居住的年数。

(h) 在无其他信息可加利用时某人的持续时间

418. 当上述所有努力均未能产生有效值时，程序能将“未报”或“不明”分配给此人在当地或行政单位居住的年数。如果值仍然无效，在不利用动态推算时应分配“不明”。为无效值选择动态推算的国家应至少利用两个特点从在该地区类似人员得出“已知”信息。

4. 以前居住地 (P1E)

419. 以前居住地指移居目前常居的行政单位之前居住的主要或较小行政单位或外国（联合国，1998年，第2.38段）。

(a) 以前居住地的编辑

420. “以前居住地”项目应与“持续居住时间”一起编辑。如果该人出生于本地（国家、地区或行政单位，依据普查项目而定）而且从未迁移过，应或者将此项留为空白，或分配“从未离开”的特定代码。不过，空白在制表期间可能引起问题，所以编辑小组

需要决定处理这种情况的最佳方法。

(b) 在边界变动时的以前居住地

421. 国界随时间推移而变动，所以应小心确保适当的对应关系体现在编码方案中。此外，设置代码的方法应允许逻辑组合。例如，如前所述，三位数代码中第一位代表居住的洲，第二位代表洲内的区域，第三位代表区域内的国家。

(c) 在某人自出生以来未迁移过时

422. 数据处理人员对某些个人项目制表。因此专家们应确保除了其他地点代码外，还使用“出生本地”的特殊代码。程序采用这种方法能够将出生于一个地点的人和出生于一个地点但迁移至同一地区内另一地点的人加以区分。

(d) 利用单元内其他的人

423. 在“以前居住地”无效或不一致时，通常适用类似于就“持续居住时间”执行的编辑。如果母亲在住房单元内，编辑程序能够审查她以前的居住地。然后程序能够为成人不经常流动国家中的子女和成人查找户主以前的居住地。

(e) 对于以前居住地不适当的其他人

424. 如果上述任何方法都不产生有效值，程序能将“未报”或“不明”分配给此人以前居住地的年数。如果值仍然无效，在不利用动态推算时应分配“不明”。为无效值选择动态推算的国家至少应利用两个特点从该地区类似人员获得“已知”信息。

5. 在以前某个订定日期的居住地 (P1F)

425. 过去某订定日期的居住地，系指在普查之前订定日期该人所居住的主要行政单位或较小的行政单位或外国。所选的基准日期应该是对国家目的来说最有用的日期。在多数情况下，这个日期被认为是普查前一年或五年。前者提供一年中关于迁移的当前统计数据；对于收集供分析国际迁移用的数据，后者可能更为恰当，但对于分析当前内部迁移情况也许不那么适合。在选择基准日期时还应考虑到个人能否准确回忆普查日期前一年或五年的常居地。在每五年进

行一次人口普查的国家，对大多数人来说，五年前的日期正好是上次普查的时间。在其他情况下，回顾一年前的情况也许比回忆五年前的情况准确。然而，一些国家也许得利用另一种基准时间，不同于普查前一年或五年，因为这两个间隔时间可能会造成回忆方面的困难。根据国家具体情况，也许有必要将基准时间定在与大多数人会记得的一个重要事件发生的日期。此外，关于抵达该国的年份的信息也许有助于了解国际移民的情况（联合国，1998年，第2.40段）。

426. “过去某个订定日期的居住地”类似于以前居住地的编辑。通常，各国将提问“持续居住时间”和“以前居住地”或简单地提问“过去某个订定时间的居住地”。如果此人生于查点地（国家、地区或行政单位，依据普查项目而定）而且从未流动过，可将本项留为空白，或可使用“从未离开”的特定代码。如前所述，空白可能在制表时造成问题。然后，适用前三段就以前居住地说明的相同过程。

C. 社会特点

427. 社会特征因国而异，但一般指说明国家社会文化情况各个方面的项目。教育项目包括识字、就学情况和教育程度以及学习科目和学历，可按照联合国教育、科学及文化组织（教科文组织）编制的《国际教育标准分类》1997年修订本的类别分类（联合国，1998年，第2.144-2.164段）。本节还覆盖残疾、缺陷和障碍及残疾原因。有关残疾问题的共同框架和定义可在《缺陷、残疾和障碍国际分类》（《残疾分类》）中找到，它由世界卫生组织于1980年出版。下文审查的其他社会特点包括宗教、语言和民族。

1. 读写能力（识字情况）（P5A）

428. 关于识字情况的数据，最好向所有10岁和10岁以上的人收集。不过，在若干国家内，某些10岁至14岁的人可能才开始入学识字。这一年龄组的识字率可能误导人。因此，对识字情况进行国际比较时，应制表说明所有15岁和15岁以上的人的识字情况。凡收集较年轻之人的有关数据的国家，关于识字情况的列表至少应将15岁以下的人与15岁和15岁以上的人区分开来（联合国，1998年，第2.147段）。

429. 每个国家都应为识字制表确定最低年龄；同样，编辑小组必须决定对识字进行编辑的最低年龄，因为可能需要内部使用的补充汇总表。在调查表编制的过程中，编辑小组应决定收集的最低年龄和到哪种教育程度便不再需要提问。因此，如果调查对象已达到某个就学水平，查点员可不必再提识字的问题。

430. 识字情况的编辑首先检查已学完的最高年级；如果最高年级有基于说明的“识字”登录项，就应分配“是”的代码。应将达到规定就学水平的人视为识字。在发现无效的识字代码的情况下应分配一个值。该登录项或者应为“不明”，或者利用基于订定变量的推算矩阵确定，例如最高年级和性别。

2. 就学情况（P5B）

431. 原则上关于就学情况的信息应向所有年龄的人收集。它特别与正式入学年龄的人口有关，这个人口群体的年龄一般从5岁至29岁，但可因国家而异，要视国家教育结构而定。如果数据收集范围扩大到包括学前教育和（或）在生产和服务企业、社区组织和其他非教育机构为成人组织的其他系统化教育和培训课程的就学情况，则年龄范围可酌情予以调整（联合国，1998年，第2.151段）。

(a) 就学编辑

432. 每个国家的编辑小组必须决定什么年龄适合于上学读书。由于多数国家将就学分为几个等级，如果这些等级按年龄编制，专家们还必须决定哪些年龄组适合于各种等级的学校教育。所有其他人的登录项必须改变。如果编制程序为这一类别产生不一致的答案，就需要或者改变年龄，或者改变就学情况。如果某一特定国家确定了某个预定的年龄，就应指示查点人员略去超过此年龄人员的就学情况。在有人到了中年继续参加中等或高等教育的学习的情况下，确定就学情况的高限就可能不适合了。也许可在普查前通过预测试来测试各种答案和答案的组合，以便可在实际普查前作出这些决定。

(b) 全日制入学或非全日制入学

433. 有些国家可能希望获得关于全日制入学

或非全日制入学接受教育的信息。如果包括本项，它可能需要成为就学情况编辑的组成部分，或可成为一项单独的编辑。

(c) 就学情况与经济活动的一致性

434. 与其他主要项目如主要经济活动等的一致性编辑应首先进行。如果入学接受教育是主要经济活动的登录项之一，而且某人报告他（或她）的主要经济活动是上学读书，就应将“是”的代码分配给就学，而且主要经济活动应是“学生”。这就是说，答案应一致。在所有其他情况下，任何有效的答案均应接受。

(d) 就学情况无效或不一致登录项的分配

435. 如果登录项超出范围而且学完的最高年级有效，就应利用基于年龄、性别和最高年级的推算矩阵分配一个登录项。如果最高年级无有效代码，就应利用识字登录项来给就学情况赋值。如果识字情况无有效代码，就应仅根据年龄和性别给就学情况登录项赋值。

436. 推算矩阵可能需要反映按性别和年龄列示的不同就学格局（有时候按单一年的年龄或小跨度年龄组）。

3. 教育程度（学完的最高年级或程度）（P5C）

(a) 教育程度的编辑

437. 教育程度（最高年级或程度）的编辑应包括(a)有效登录项与年龄之间的一致性检查，和(b)在原始登录项超出范围时登录项的推算。如前所述，在不利用动态推算的国家，值应为“未说明”。在利用动态推算的国家，年轻人将需要性别和单一年的年龄，稍大的儿童将需要性别和小跨度年龄组。在其数据包括最高年级和最高程度的国家，可能需要多个推算矩阵。

(b) 教育程度的最低年龄

438. 每个国家的编辑小组必须决定入学的最低年龄。最低年龄确定后，学完的最高程度通常不应超过某人年龄加某个常数（它代表入学的最低年龄）。同样，重要的是应利用儿童单一年的年龄，因为如果

年龄组范围非常大，更新推算矩阵可能引入错误。

(c) 年龄与教育程度的关系

439. 编辑小组还必须决定数据集中将允许多少干扰。通常更好的办法是在年龄与教育程度相冲突时改变少量例外情况，而不是接受大量实际不一致的答案。因此，对于原始登录项超出范围或与年龄不一致的情况，可分配一个登录项。对于不利用动态推算的国家，可录入“未说明”。对于利用动态推算的国家，可根据年龄（包括学龄人员单一年的年龄）、性别和就学情况获取登录项。教科文组织承认识字不同于教育程度，所以也许不应将“读写能力”作为一个值用于推算矩阵。

4. 学习科目和学历（P5D）

440. 关于各人按教育水平和学习科目分列的信息，对检查劳动力市场中学有专长合格人力的供求是否配合相当重要。它对于规划和管理各级和各类教育机构和培训课程及其分机构的生产能力也同样重要（联合国，1998年，第2.158段）。

441. 未满15岁（或其他预定年龄）的人不应拥有关于学习科目和（或）学历的信息。15岁和15岁以上的人，教育程度与学习科目和（或）学历之间应存在某种关系。在每种情况下，在发生无效登录时，不利用动态推算的国家可将该登录项变为“不明”。利用动态推算的国家可能希望考虑利用年龄、性别、教育程度及可能还有职业等分配学习科目和（或）学历。

5. 宗教（P3E）

442. 为人口普查的目的，宗教可确定为(a)所选择的宗教或精神信仰，不论这一信仰是否由一个有组织的团体为代表，或(b)附属于信奉具体宗教或精神信条的一个有组织的团体。每个在人口普查中调查宗教情况的国家都应利用最适合其需要的定义，并应在普查出版物中列明所使用的定义（联合国，1998年，第2.109段）。

(a) 宗教编辑

443. 宗教是适合第二章介绍的例子的变量之

一。为某个人获得一个有效值，或者直接从住户的另一成员，如果一个值可加利用，或者从具有类似特征的另一户主。编辑小组应当确定用于其他社会变量的逻辑编辑方案。户主应首先指定和编辑，不论他（或她）是否单元内的第一人。如果具有无效或不明宗教的某人是户主，应采取下列步骤：

(b) 户主无宗教，但单元内另一人存在宗教

444. 第一步是确定住房单元内是否另有人具有有效的宗教，并分配第一个有效的宗教。

(c) 户主无宗教，或单元内任何其他人都无宗教

445. 如果住户内任何人均未报宗教，或者分配“不明”（如果该国不利用动态推算），或者从具有类似特点的最近的户主推算宗教，其中包括年龄和性别，以及语言、出生地和其他有关变量，视情况而定。

(d) 对于无宗教的户主以外的人

446. 如果此人不是户主而且报告无宗教，分配户主的宗教。

6. 语言 (P3F)

447. 在普查中可收集三种语言数据（联合国，1995年，第2.112段）：

(a) 母语，其定义是个人在其幼年时家中常说的语言；

(b) 常用语言，其定义是个人目前在其现居家中所说或最经常说的语言；

(c) 说一种或多种指定语言的能力。

(a) 语言编辑

448. 在可能列在调查表上的三种不同语言量度（母语、常用语言和说一种或多种指定语言的能力）中，头两种即母语和常用语言是关联的。在二者均出现于调查表上时，编辑小组应考虑合起来编辑。如果任何一种无效，可利用另一种提供登录项。

(b) 语言编辑：户主

449. 语言是适合于第二章介绍的例子的另一个变量。编辑小组应当确定用于其他社会变量的逻辑

编辑方案，并首先编辑户主。如果具有无效或不明语言（母语或常用语言）的人是户主，首先确定住房单元内是否另有他人具有有效的语言并分配第一种有效的语言。在无任何人时，如果不利用动态推算就分配“不明”，或从具有类似特点的最近的户主推算一种语言，其中包括年龄和性别，以及其他语言变量、出生地和这些情况下的其他变量。

(c) 语言编辑：户主以外的人

450. 如果此人不是户主而且语言无效，就分配户主的语言。

(d) 语言编辑：利用族裔或出生地

451. 语言与族裔及有时候与出生地密切相关，而且就有些国家而言可以结合在一起编辑。还有，编辑小组应考虑编制代码以反映这些变量之间的关系。依据代码数字的位数和该国语言和族裔群体的分布，可以形成对应关系以帮助分配不明或不一致的答案。

(e) 语言编辑：母语

452. 如果母语不明，但此人是菲律宾人并生于菲律宾，可分配相应同等语言——他加禄语、伊洛卡诺语或菲律宾其他语言。通常，只有户主以这种方法分配语言，而且此种语言的代码也分配给住户其他成员，但是每个国家的编辑小组需要考虑特定情况，包括地理（例如城市或农村居住地）、年龄或其他项目。

(f) 语言编辑：说一种指定语言的能力

453. 说一种指定语言的能力是适合于第二章介绍的例子的第三个变量。再次应先对户主进行编辑。如果户主语言的值无效或不明，第一步是看看住房单元内是否另有他人有说一种语言的有效能力，并分配第一种有效的能力。然后，如果不存在这种人，要是这个国家不利用动态推算，就分配“不明”，或从具有类似特点的最近的户主推算语言能力（例如，年龄和性别，但也包括出生地和其他有关变量，如考虑到其他的情况）。如果此人不是户主，而且说指定语言的能力无效，就分配户主的能力。

7. 民族 (P3G)

454. 对有关人口中的民族和（或）族裔群体的信息的需要取决于各国的情况。确定族裔群体所依据

的一些基础包括：族裔本源（即原籍国或地区，以别于公民资格或法定国籍国）、种族、肤色、语言、宗教、衣着或饮食习惯、部落或这些特征的各种组合。此外，所使用的一些词汇，例如“种族”、“血统”或“部落”有一些不同的含义。为此，调查人口中的族裔特征的每个国家所使用的定义和标准，必须由其所希望查明的群体来决定。根据此项题目的性质，这些群体因国而异；因此无法建议制订有关的国际标准（联合国，1998年，第2.116段）。

(a) 民族编辑

455. 在民族无效或不明时，其他几个变量如收集的话能够有助于“确定”它。在许多国家，无论在国内还是外国的出生地与民族之间存在着某种关系。同样，“母语”经常能有效地表明许多国家的民族，因为类别，因此也包括代码，将是类似的，如果说不相同的话。

(b) 民族编辑：户主

456. 族裔也适合第二章介绍的例子。编辑小组应当遵循已为其他社会变量说明的方案。户主应先编辑。如果族裔无效或不明的人是户主，首先寻找住房单元内任何其他有效的民族，并分配第一个有效的民族。如果无此种人存在，下一步或者是分配“不明”，或者如果该国不使用动态推算，则从具有类似特点（年龄和性别，以及语言、出生地和其他可能合适的变量，视情况而定）的最近的户主推算民族。

(c) 民族编辑：户主以外的人

457. 如果此人不是户主而且族裔无效，就分配户主的族裔。

(d) 民族编辑：语言和出生地的利用

458. 族裔和语言，有时候还和出生地密切相关，而且就有些国家而言，可合起来进行编辑。还有，编辑小组应考虑编排其代码以反映这些变量之间的关系。依据代码数位和该国族裔群体和语言的分布而定，可将对应关系发展得将有助于分配不明或不一致的答案。

459. 例如，如果族裔不明，但此人说一种菲律宾语言，而且出生于菲律宾，可分配有关等同的族裔即菲律宾族裔。通常，只有户主将以这种方法分配民

族（而且将向其他成员分配该代码），但每个国家的编辑小组需要考虑特定的情况，包括地理（例如城市或农村居住地）、年龄或其他项目。

8. 残疾 (P8A)

460. 为了进行人口残疾方面的计量，应将残疾人定义为这样的人，他们由于长期身体精神或健康不好而使得其可以做的活动种类和数量受到限制。断腿和生病等临时状况所造成的短期残疾不包括在内。只有长达六个月以上的残疾状况才包括在内（联合国，1998年，第2.262段）。

461. 用来识别残疾人的问题应列出主要残疾类别，以便人人可以查看每一种残疾是否出现。建议采用以《缺陷、残疾和障碍国际分类》为根据的下列主要残疾类别表：视障碍甚至包括戴眼镜）；听障碍（甚至包括用助听器）；讲障碍（说话）；行动/活动障碍（行走、上楼梯、站立）；身体移动障碍（伸、蹲、跪）；握/拿障碍（用手指紧握或手持物体）；学习障碍（智力障碍、智力迟钝）；行为障碍（心理问题、情绪问题）；自理障碍（洗澡、穿衣、进食）；其他（具体指明）。如果某人表明他有清单所列的一项或多项残疾，则被视为残疾人（联合国，1998年，第2.264段）。

(a) 残疾编辑

462. 在某人不同意这个问题时，难以断定这个项目空白是由于无残疾还是因为出于任何原因调查对象不愿回答。一国的编辑小组必须决定他们是否想以通常的方法编辑这个项目，在不利用动态推算时分配不明，或在利用动态推算时利用其他个人的答案。抑或专家们可决定只有具有说明残疾存在的答案才应接受，而且任何有效的答案应为“无残疾”。在后一种情况下，将不利用动态推算。

(b) 多种残疾

463. 收集多种残疾信息的国家将需要修改编辑方法。编辑程序将需要跟踪有多少残疾人口总数能够容纳以及这些残疾的重复和分布情况。如前所述，多数国家将会发现利用他人数据分配残疾的做法是不适宜的，因此在无效的情况下可能需要“不明”类别和甚至“是否存在残疾不明”类别。

9. 缺陷和障碍 (P8B)

464. 可以按照其他特征例如教育和就业将据报有残疾的人和据报没有残疾的人进行比较性分析,以估计障碍程度。各国也可能有兴趣收集关于残疾人在什么具体情况下体验障碍的数据,例如,使用公共交通工具、干活或参加社会活动。收集这类信息的目的是减少造成残疾人孤立的具体因素(实物障碍、缺乏服务、社区的消极态度、对残疾人的偏见)。提出关于身体缺陷的问题时应当指明阻止残疾人在平等基础上参与社会活动的困难种类。为了了解残疾人所处的环境,物理和社会因素都应考虑(联合国,1998年,第2.276段)。

(a) 缺陷和障碍编辑

465. 一国的编辑小组必须再次决定他们是否想以通常的方法编辑这个项目,在不利用动态推算时分配不明类别,或在利用动态推算时利用他人的答案。抑或专家们可以决定只有具体说明障碍存在的答案将被接受,而且任何无效的答案将为“无障碍”。

10. 残疾的原因 (P8C)

466. 关于残疾原因的信息对预防方案的规划和评估都很重要。由于普查篇幅有限,询问引起残疾的一般情况,而不是提出关于具体疾病或损伤的详细信息,可以获得有关残疾原因的信息。建议在收集关于残疾原因的数据时采用五个主要类别:(a) 先天/产前;(b) 疾病,如小儿麻痹症、麻风、白内障;(c) 损伤/事故/创伤,如:交通事故、意外跌伤、火烧、战争或意外中毒所引起的损伤;(d) 其他;(e) 不详(联合国,1998年,第2.277段)。

(a) 残疾原因编辑

467. 一国的编辑小组必须决定是否以通常的方法编辑这个项目,在不利用动态推算时分配不明类别,或在利用动态推算时利用他人的答案。抑或专家们可决定仅有具体说明残疾原因存在的答案将被接受,而且将不利用推算矩阵。

D. 经济特征

468. 原则上应收集整个人口经济活动状况的信息,但实际上收集根据各国情况所确定的最低年龄或最低年龄以上每个人经济活动状况的信息。不应该

自动将最低离校年龄作为收集活动状况信息时所用的最低年龄限数。有些国家通常有许多儿童从事农业或其他类别经济活动(如采矿、编织、小买卖),这些国家需要选择比幼年儿童就业不普遍的国家较低的最低年龄。编制经济特征表时应至少将不满15岁者和15岁及15岁以上者相区别;最低离校年龄高于15岁以及从事经济活动的儿童低于这一年龄的国家,应努力保证获得关于这些儿童的经济特征,以便至少可对15岁和15岁以上者实现国际可比性。超过正常退休年龄的老年男女从事经济活动的情况也常常被忽视。因此,在计算从事经济活动人口时应注意这一点。计算从事经济活动人口时通常不应使用最高年龄限数,因为可能有相当多超过退休年龄的老人通常或偶然从事经济活动(联合国,1998年,第2.172段)。

469. 每个国家都必须确定参加经济活动的最低年龄。想收集关于童工数据的国家可能需要选择一个较低的最低年龄,但必须记住,在未参加劳动力的儿童被错误查点为属于劳动力时将会出现一些干扰。在确定最低年龄后,编辑经济活动项目以便为X岁或更大的人制表;因此,只是为了弄清所有登录项为空白项才将需要对X岁以下儿童进行编辑。为了便利所有制表,可能已为15岁以下儿童录入的任何答案都应删除。

1. 活动状况 (P6A)

470. 经济活动状况由数个经济变量组成,其中有些在下文说明。这些变量可满足数据收集的需要,但为了进行数据处理和分析,可能需要重新分类。

471. “现时活动状况”是指较短的基准时期(例如一星期)内一人与经济活动的关系。现时活动最适用于一年内人们的经济活动不受会引起变动的季节因素或其他因素较大影响的国家。这种一星期的基准时期可以是明确规定的最近一星期、最近一个完整的日历星期或查点之前的最后七天(联合国,1998年,第2.180段)。

472. 按照联合国(1998年,第2.182段)的规定,就业者包括在规定年龄以上并在一周或一天的较短基准时期内属于以下一类情况的所有人:

(a) 有薪酬就业:

(一) 在工作者：在基准时期内从事某种工作以取得以现金或实物支付的工资或薪金的人。

(二) 有职业但不在工作者：已经从事目前职业，而在基准时期内暂时不在工作，但由于例如持续得到工资/薪金、保证可在意外事件结束后恢复工作、已商定在短期离职后恢复工作的日期等情况证明其有正式的工作委派的人。

(b) 自营职业：

(一) 在工作者：在基准时期内从事一些工作以取得以现金或实物支付的利润或家庭收益的人。

(二) 经营一家企业但并不在工作者：经营一家企业（可以是工商企业、农场或服务性企业），但由于某种具体原因在基准时期内暂时不在工作的人。

473. “通常不从事经济活动”人口包括既不列为就业也不列为失业的所有人。应将这些人口列为以下四类（联合国，1998年，第2.202段）：

(a) 学生：基准时期内在任何公立或私立正规学校接受任何等级的系统教育，但通常不列为从事经济活动的男女；

(b) 操持家务者：通常不从事经济活动而在自己家中从事家务的男女，例如家庭主妇和负责照料家务和儿童的其他亲属（但挣工资的家庭工人列为从事经济活动者）；

(c) 退休金或资本收入领取者：通常不从事经济活动，但因财产或投资、利息、租金、版税获得收入或因以前经济活动获得退休金，而且不能列为学生或操持家务者的男女；

(d) 其他：通常不从事经济活动但接受公共援助或私人资助的男女，以及不属于上述任何类别的所有其他人。

(a) 有关活动状况的类别

(一) 失业人口（P6A1）

474. 按照联合国（1998年，第2.194段）的规定，失业人口包括所有超过规定年龄并在基准时期内处于下列状况的人：

(a) 没有工作：他们未从事有薪酬就业或自营职业；

(b) 目前等待工作：在基准时期内等待从事有薪酬就业或自营职业；

(c) 寻求工作：在最近一个指定时期采取步骤寻求有薪酬就业或自营职业。具体步骤可包括在公立或私立职业介绍所登记；向雇主提出申请；在工作地点、农场、工厂门房、市场或其他集会场所进行询问；刊登或答复报纸广告；争取亲友帮助；寻找土地、房屋、机器或设备以建立自己的企业；安排资金；申请许可证和执照等。在失业者分类中区分首次寻求工作者和其他寻求工作者是有益的。

475. 大体上，被归类为失业者的人必须符合上述三项标准。但是，在谋求工作的常规手段具有局限性，劳动力市场在很大程度上没有组织或规模有限，当时劳动力吸收不足或劳动力基本上是自营职业的情况下，可通过放宽谋求工作这一标准而适用失业的标准定义。这种放宽的做法是针对发展中国家的，在这些国家中，谋求工作这一标准并不能完全反映失业的程度。在极为特殊的情况下，可以放宽到完全取消这一标准，这样放宽后，仍然适用的两个基本标准就是“没有工作”和“目前等待工作”（联合国，1998年，第2.195段）。

476. 失业——“被解雇”，“找工作”，不论此人能否找到工作，以及“最近工作的年份”（如有的话）——应合起来一并编辑。而且，它们还需与经济活动的答案相容，在多数情况下，如果工时、产业、职业、工人级别和工作场所等项目已填，就不应再填它们。如果主题专家确定在答案空白或无效时“被解雇”项需要登录，那就可以执行利用年龄和性别也许还有此人教育程度的推算矩阵。

(二) 求职（P6A2）

477. “求职”的编辑应与“被解雇”和“不求职的原因”的编辑一并进行。主题人员应当开发利用这些项目的登录项的编辑程序以推算其他项目。编辑应考虑当地和区域情况及普查或调查变量。

(三) 现时不参加经济活动者（P6A3）

478. “现时不从事经济活动人口”，或称“不属于劳动力”的人口，包括在用测定目前活动的较

短基准时期内既没有就业也没有失业的所有人（联合国，1998年，第2.205段）。根据目前“不从事经济活动”的原因，可将其划归为下列任一类别：

- (a) 在校就学；
- (b) 从事家务；
- (c) 以领取退休金或资本收入为生；
- (d) 由于其他原因不工作，包括残疾或缺陷。

“现时不参加经济活动”的编辑已纳入上文经济活动的编辑。

(四) 不求职的原因（P6A4）

479. 本项目只应为记录为“不在求职”的人编辑；所有其他人应有空白登录项。或者，如果职业、产业和就业状况等项出现有效登录，应录入“有职业但不在工作者”代码。这一代码说明在基准期内就业但不在工作的参加经济活动的人。在所有其他情况下，如果不利用动态推算，可分配“不明”类别。就利用动态推算的国家而言，可利用年龄、性别和主要活动分配一个登录项。

(b) 经济活动状况的编辑

480. 经济活动一般有以下各类：

- (1) 就业并在工作者；
- (2) 就业但不在工作者；
- (3) 自营职业并在工作者；
- (4) 自营职业但不在工作者；
- (5) 求职者；
- (6) 学生；
- (7) 操持家务者；
- (8) 退休金或资本收入领取者；
- (9) 其他未列入劳动力者。

481. 关于这一变量，头四种可能性属于参加经济活动的人，其次四类属于不参加经济活动的人。“在工作”的人（1和3类）是就业者，“不工作”的那些人（2和4类）可能失业或未列入劳动力，依据对失业项目（“被解雇”；“求职”；“过去工作

的年份”）的答复而定。

(一) 就业人员

482. 如果选择参加经济活动人员类别之一（1至4类），工作时间、职业、产业、经济活动状况和工作场所等变量均应当填写。如果未填，应予编辑并补填，或者填为不明，或者填入非流水值或流水值。如果选择1至4类中的一类，有关被解雇、求职和过去工作年份等变量应为空白。如果填写，应改为空白。

(二) 失业人员经济活动

483. 如果选择不参加经济活动人员的类别之一（5至9类），“被解雇”、“求职”和“过去工作年份”等变量均应填写。如果未填有效登录内容，应加以编辑和补填，或者填上“不明”，或者填上非流水值或流水值。如果选择5至9类，工作时间、职业、产业、经济活动状况和工作场所等变量均应为空白。如果填写，应改为空白。

(三) 学生和退休人员的经济活动

484. 如果选择类别6即学生，主题人员需要决定上学读书的变量是否必须为“是，在校”。如果选择类别8即退休人员，主题人员需要决定人员是否必须为某个年龄才退休。

(四) 在经济活动无效而且就业变量报告时

485. 如果经济活动登录项无效，而且如果有关工作时间、职业、产业和工作场所等有些变量报告，就应将调查对象的经济活动编为代码值1至4。也许将需要一个推算矩阵选择适当答案。

(五) 在经济活动无效而且失业变量报告时

486. 如果报告有关“被解雇”、“求职”和“过去工作时间”等任何变量，经济活动登录项应编为代码值5至9。如果此人正在上学，值也许应为6。如果此人为老年人，值也许应为8。否则，主题专家可利用推算矩阵分配适当的答案。

(六) 在经济活动无效而且无经济变量报告时

487. 如果任何经济活动项目均未出现答案，主题专家也许将希望利用推算矩阵确定最适当的答案，并然后推算其他经济项目。

2. 工作时间 (P6B)

488. 工时是在人口普查中为经济活动所选的一段基准时期中生产的货物和服务实际所用的总时数,包括正常工作的时间和加班的时间。假如基准时期较短,例如人口普查之前一周,建议工时以小时为计算单位;如果基准时期较长,例如人口普查之前12个月,则以星期为计算单位,可能的话则以日为单位。工时还应当包括从事这样的活动所用的时间,例如准备、修理或维修工作场所或工作设备所用的时间。这些活动与所生产的货物或服务无直接关系,但按其定义仍然是属于这项工作的任务和责任。在实践中,这些时间还包括在从事这些活动时的空闲时间,例如等候时间、待命时间或其他短休时间。不过不应包括时间较长的午饭时间和因休假、假日、病假或冲突(例如罢工和闭厂)而没有工作的时间(联合国,1998年,第2.210段)。

489. 本项目只应为其经济活动答案为“就业并在工作”或“自营职业并在工作”的人编辑。对于某些国家,也应为操持家务者包括工时。编辑小组预定的类别应予接受。如果不利用动态推算,应将空白、零或非数字代码改为“未报”,而且主题专家可能希望将经济活动变量改为“不在工作”,如果报告的工时等于零的话。

490. 如果利用动态推算,推算矩阵最低限度的变量包括年龄组和性别,但其他变量如教育程度、职业或产业主要类别也可利用。

3. 职业 (P6C)

491. 职业是指在基准时期就业者(或失业者在失业前)所做的工作的种类,不论这个人应归入的产业或就业状况为何(联合国,1998年,第2.214段)。

492. 本项目只应为其经济活动为“就业并在工作”或“自营职业并在工作”的人编辑。如果不利用动态推算,应将空白、零或无效答案改为“未报”。

493. 职业代码编制得往往使不同的数字表示主要和次要职业代码。就职业而言,几乎不可避免的自填将会增大编码工作量。

494. 如果利用动态推算,推算矩阵的最低限度的变量包括年龄组和性别,但其他变量如教育程度或产业主要类别也可加以利用。

4. 产业 (P6D)

495. 按照联合国(1998年,第2.221段)的规定,产业是指一名就业者在为经济特征数据而确定的基准时期工作所在企业单位的活动(或失业者最后的工作)。关于如何选择拟进行分类的工作/活动,见《原则和建议》第2.212段。

496. 本项目只应为其经济活动是“就业并在工作”或“自营职业并在工作”的人编辑。如果不利用动态推算,应将空白、零或无效答案改为“未报”。

497. 产业代码编制得往往使不同的数字表示主要和次要产业代码。本项目几乎不可避免的自填将会增大编码工作量。

498. 如果利用动态推算,推算矩阵的最低限度的变量包括年龄组和性别,但其他变量如教育程度或产业主要类别也可利用。

5. 就业状况 (P6E)

499. 就业状况是指从事经济活动的人的就业状况。也就是这个人与雇用他/她的人或组织机构之间的明示或默示的雇用合同种类。用来界定分类各个组别的基本标准是经济风险的种类,其中一个要素是这个人忠于这份工作的程度,以及这个人在这份工作中对企业单位和其他工人有或将有什么样的权力。应小心谨慎,以确保从事经济活动的人按就业状况来分类,分类时所依据的工作与按职业、产业和部门来将人分类时所用的一样(联合国,1998年,第2.226段)。

500. 从事经济活动的人口应按就业状况分类(联合国,1998年,第2.227段)如下:

- (a) 雇员,能够从中区别有固定合同的雇员(包括正常雇员)和其他雇员;
- (b) 雇主;
- (c) 自营工作者;
- (d) 作贡献的家庭工人;
- (e) 生产者合作社成员;
- (f) 不能按其状况分类的人。

501. 法人企业的业主——经理，他们通常将被归入雇员类，但是为了便于作出某种说明和分析，将他们归为雇主一类。

502. 本项目只应为其经济活动是“就业并在工作”或“自营职业并在工作”的人编辑。如果不利用动态推算，可将空白、零或无效答案改为“未报”。如果利用动态推算，推算矩阵最低限度的变量包括年龄组和性别，但其他变量如教育程度或产业主要类别也可利用。

6. 收入 (P6F)

503. 有关《人口和住房普查的原则和建议，第一次修订本》介绍的人口经济特征的普查题目，专注于国际劳工组织（劳工组织）建议界定的参加经济活动人口，而经济生产的概念是相对于《国民帐户体系》（联合国，1998年，第2.165段）确立的。参加经济活动的人口包括在某一订定的基准时期内，按《国民帐户体系》定义，为经济货物和服务的生产提供或可提供劳动力的所有男女。在此框架内，收入可以用以下来界定：(a) 每一个从事经济活动的人每月工作所得的现金或实物收入，或(b) 每户每年从各种来源得到的现金和（或）实物收入总数。一般的实地调查，特别是人口普查，要收集可靠的收入数据，特别是从自营职业和财产所得的收入，是极端困难的。如再涉及非现金收入，则更是难上加难了。人口普查收集收入数据，即使只限于现金收入，也会在工作量、答复错误等方面遇到特别的问题。因此，一般认为这个问题，包括更广泛的收入定义，比较适用于抽样调查。由于本国的需要，一些国家还是想得到有关现金收入的有限信息。如果这样定义的话，所得到的信息除了有助于人口普查之外，还会有助于关于住户收入的分配、消费和积累等的统计（联合国，1998年，第2.236段）。

504. 《原则和建议》认定两类收入：个人收入和住户收入。两个项目均需要类似的编辑。就个人收入而言，如果不利用动态推算，无效的收入答案应分配“未说明”或“不明”。如果利用动态推算，可利用年龄、性别、教育程度、产业、职业和其他限定条件来形成收入的推算矩阵。

505. 就住户收入这一变量而言，它是住户挣得的全部收入之和，并登录在住房记录上。动态推算的

编辑大致相同，不过利用的是户主的年龄、性别和教育程度而不是每个个人的这些情况。

7. 机构部门 (P6G)

506. 就业机构部门是指职业所属企业的法律组织形式和主要职能、行为和目的（联合国，1998年，第2.239段）。

507. 某些可能的产业和职业与就业的机构部门（法人公司、政府、非营利部门、住户或其他）之间存在着关系。有些国家可能选择检查变量之间的这些关系，以确保在对这些变量交叉列表时汇总表不显示不一致。

508. 就编辑而言，不利用动态推算的国家将需要在不明的情况下给机构部门分配“不明”。利用动态推算的国家应考虑地域内类似人员的年龄和性别，也许包括主要的产业或职业。

8. 工作地点 (P6H)

509. 工作地点是现时就业的人工作的地点，也是一个通常就业的人做其主要工作的地点，这个主要工作是用来确定调查对象的其他经济特征如职业、行业、就业状况。虽然有关工作地点的信息可用来建立地区就业劳动力档案（相对于按居住地建立的人口资料档案），但是主要的目的是把工作地点信息与居住地信息连接起来（联合国，1998年，第2.245段）。

510. 由于“工作地点”用于公交系统的统计资料，因此重要的是应使报告信息的任何变动反映所考虑的具体地区。这样，各国编辑小组可能希望考虑给无效情况分配“不明”，而且只分析“已知”情况。

511. 如果接受自填并必须为它编码，本项的编码作业将会增加时间和复杂度。如为数字确定等级，例如，第一个数字表示省、第二个表示区，以此类推，编码作业也许将更为有效和准确。

512. 就推算矩阵而言，数据处理人员需要确保只将可能的地理位点分配给矩阵。明智的做法可能是为每个行政单位或其他地理区域开始采用新的非流水法，确保不能选择以前的值。对于推算矩阵本身而言，年龄和性别，也许还有修改的主要职业或产业主要类别可以包括在内。还有，国内和国外的工作可能需要不同的推算矩阵。

五、住房情况编辑

513. 住房情况编辑规范考虑到各个项目的有效性以及项目间的一致性。了解特定国家项目间的具体关系，就可以制定进行前后一致编辑的计划，从而保证制表使用较高质量的数据。例如，在墙壁用竹子建造时，住房单元不应有水泥房顶。同样，单元的房屋内应有自来水，以便结构内有抽水马桶或浴缸或淋浴设施。

514. 关于人口项目，编辑小组必须对遗漏或不能解决的项目决定分配“未说明”——对于“未知”或其他值分配的静态推算值（非流水式），或者基于其他住房单元的特点分配一动态推算值（流水式）。以前，在许多情况下，人们偏爱动态推算，因为在只有表格本身的信息可用于就未知项目作出决定时，它消除了制表阶段所需的推算。这样建立的推算矩阵为空白、无效登录项提供登录值，或者在不存在其他具有有效答复的相关项目时解决不一致情况。一些国家全国的住房特点可能有所不同，但在大部分地区内却差异很小。另一些国家地区之间特定项目可能有很大不同，特别是在城市和农村地区之间。在编制推算矩阵特别是最初的非流水式值时，必须考虑到这种差异。编辑小组可能想具体说明在什么情况下应从以前具有其他类似特点的住房单元中为空白项提供一个登录值。

515. 除一国缺乏有关集体（集团）住所的住房信息外，应为每个序号分配一个（且唯一的一个）住房记录（见第三章“结构编辑”）。有关结构编辑的一章概述一系列质量保证程序。如果住房记录缺失，编辑程序可以生成一个住房记录，这取决于编辑小组的决定。同样，当出现两个或多个记录时，该程序可以去掉一个或几个记录。

516. 在理想的情况下，应只对有关的项目选择性地编辑每个住房记录。根据城市/农村、气候和其他条件，编辑的项目可以有所不同。然而，在实践中，很少哪个国家有时间或专门知识去研究和实施多重系列，以改变遗漏或不一致的数据。实际实施选择性编辑的国家就更少了。

517. 收集的调查表的资料还视住所类型（住房单元或集团住所）及住房单元是空置还是住用而定。对于集体或集团住所，编辑可能仅限于在集团住所收集的项目或同时在集团住所和其他住房单元收集的项目。

518. 顾名思义，无家可归者通常没有住房记录。如果一国选择为他们指定标识而存在这些记录，该国处理这些记录的方式可以与处理集体住所记录的方式相同，或者可以要求用一个完全不同的编辑值，包括完全没有值。

519. 有时，应允许对特定项目登录“未报告”值。当国家的编辑小组没有可靠的依据来推算对特定特点的答复时，可能发生这种情况。决定填写“未报告”答复时必须考虑到，在规划人员需要部分信息时要求提供适当的表格特点以供规划和制定政策之用。只要“未报告”情况与已报告情况分布相同，分配“未报告”情况就不应产生问题。然而，如果“未报告”情况在某种程度上有偏差，汇编后的推算就可能会有问题，特别是对于小地区或特定类型的条件。例如，如果居住在国家界定的“不达标”住房的答卷人拒绝透露一些住房特点，查点员就不能报告这些特点，规划人员就不能采用补救程序来减轻不达标状况。

520. 住房编辑往往比人口编辑简单，因为交叉制表往往没有那么复杂。多数国家只按照不同的地理级别编辑个别住房特点。如上所述，选择不使用动态推算的国家应当确定，在出现无效或不一致答复时对于“未知”情况使用一标识。

521. 对于使用动态推算的国家，编辑小组应编制简单的推算矩阵，所使用的维数应区分住房特点。对于多数国家来说，有关“住所类型”的变量是用于动态推算的最佳初始变量，不管是住房单元还是集体住所，包括这些类别之内的单元类型。

522. 对于一些国家来说，地区还用作这些推算矩阵的一个维数。也可以使用保有权。例如，如果国家的房屋一半为租赁单元，一半为自有单元，就适合

将保有权作为推算矩阵的一个维数列入。不过，如果只有 5% 的单元是租赁单元，就更适合使用另外一些特点。保有权常常是推算矩阵中使用的一个有用变量，特别是保有权主要类型占较大百分比的国家。考虑的其他特点包括墙壁的类型和是否有电。

523. 对于每个国家来说，作为推算矩阵的维数包括在内的特定变量必须与数据集中的变量相符，因此，对于住房项目来说，必须注意各个项目以及项目组合能对特点进行区分。在《原则和建议》中，对基本项目和附加项目进行了区分；下文 A 节介绍的编辑项目强调了列入矩阵的基本项目。

A. 基本题目

524. 住房普查的查点单位是(a) 建筑物；(b) 住所；和(c) 住户的居住者。联合国已经编制了一个人们普遍感兴趣和具有普遍价值的包含 20 个基本编辑题目的清单，这些题目对于全面的统计公司掌握方法也很重要。为了方便使用者，下文按字母顺序介绍建议的这些题目和若干附加题目的编辑代码。这些题目按查点单位的类型分组。

1. 建筑物：建筑物的说明 (H01)

525. 联合国建议 (1998 年，第 2.299 段) 对于有些空间用于居住目的的建筑物采用下列分类方法。

1. 与一个单独住房单元共有空间的建筑物；
 - 1.1. 独立式；
 - 1.2. 附联式；
2. 包含多个住房单元的建筑物；
 - 2.1. 高达两层；
 - 2.2. 从三层到十层；
 - 2.3. 十一层以上。
3. 为住在公共机构内的人修建的建筑物。

526. 如果该值无效，在不使用动态推算时应分配“未知”值。对于无效值选择动态推算的国家至少应使用两个特点，其中可能包括外墙的建筑材料、建筑期和（或）建筑物内住房单元的类型（见联合国，1998 年，第 2.300 段），以便从该地区类似住房单元

中获得“已知”信息。

2. 建筑物：外墙建筑材料 (H02)

527. 这个题目是指住所所在建筑物外部（外层）墙壁的建筑材料。如果这些墙壁是由多种材料修建，则应报告所用的主要材料。所区分的类型（例如，砖、混凝土、木材、粘土）有赖于有关国家最常使用的材料，并有赖于从建筑耐久性或评估持久性的角度来看待这些材料的重要性（1998 年，第 2.304 段）。

528. 如果该值无效，在不使用动态推算时应分配“未知”值。对于无效值选择动态推算的国家至少应使用两个特点，如建筑时期和 / 或建筑物的住房类型，以便从该地区类似住房单元获得“已知”信息。

3. 建筑物：建筑年份或时期 (H03)

529. 该题目是指住所所在建筑物的房龄。如果刚过去的两次普查之间的时期未超过 10 年，建议应调查该时期内修建的建筑物的准确建筑年份。如果两次普查之间的时期超过 10 年，或者以前从未进行过普查，则应调查在过去 10 年里修建建筑物的准确年份。对于在 10 年以上的时间里修建的建筑物来说，应从能提供评价现有房屋房龄的有用办法的时期收集资料。在收集关于该题目的数据时可能会遇到困难，因为在一些情况中，居住者可能不知道建筑日期（联合国，1998 年，第 2.307 段）。

530. 一些国家，即使使用动态推算的国家，接受关于建筑年份或建筑时期的项目的“未知”答复。当这种情况发生时，该国可以选择对该项目不使用动态推算，即使对于其他变量使用推算矩阵。对于这种情况，如果该值无效，在不使用动态推算时也许仍应分配“未知”值。对于无效值选择动态推算的国家应使用至少两个特点，其中包括建筑物类型、外墙建筑材料和 / 或建筑物内的住房类型，以便从该地区类似住房单元获得“已知”信息，

4. 住所：住所位置 (H04)

531. 住所位置是一个地理变量，与结构编辑一起在第三章中介绍。

5. 住所：住所类型 (H05)

532. 下面概述的分类说明一种三位数字的编码系统,联合国设计这种系统(1998年,第2.328-2.365段)是用来把有类似建筑特点的住房单元以及集体住所分门别类。居住者(人口)在各个类别中的分布情况对普查时的住房情况提供了宝贵资料。这种分类还提供了抽样调查中分层办法的有用依据。住所可分为下列类别:

1. 住房单元;
 - 1.1. 传统寓所;
 - 1.2. 基本寓所;
 - 1.3. 临时住房单元;
 - 1.4. 流动住房单元;
 - 1.5. 简陋住房单元;
 - 1.5.1. 凑合住房单元;
 - 1.5.2. 在不准备供人居住的永久性建筑物中的住房单元;
 - 1.5.3. 其他不准备供人居住的房舍;
2. 集体住所;
 - 2.1. 旅馆、寄宿舍和其他寄宿公寓;
 - 2.2. 公共机构;
 - 2.3. 营地;
 - 2.4. 其他。

533. 编辑小组制定编辑项目时应确保,所有集体住所和住房单元都有内部一致的信息。如果住所类型的值未知或无效,编辑小组可能想拟定一个编辑项目,参照其他住所来分配住所类型。否则,如果该值无效,在不使用动态推算时应分配“未知”。对无效值选择动态推算的国家普查/统计局应至少使用两个特点以便从该地区类似住房单元获得“已知”信息,如建筑物类型、保有权、房间数、楼面面积或空置状况。

6. 住所：住用状况 (H06)

534. 决定住所的住户不论是临时不在还是暂时居住,这些住所都应被记录为“住用”或“空置”,这需要根据进行的是法律上的还是实际上的人口普查来考虑。不论在何种情况下,尽可能地把当作主要住所的住所与当作第二住所的住所区别开来似乎是

有益的。在第二住所与主要住所的特点有明显区别时,这样做非常重要。例如,农业住户从他们在乡村中的永久住所迁到在一年中某个季节在农业地点上的基本建筑中就属于这种情况(联合国,1998年,第2.369段)。建议对传统和基本住房分类如下:

1. 住用;
2. 空置;
 - 2.1. 季节性空置;
 - 2.2. 非季节性空置;
 - 2.2.1. 出租;
 - 2.2.2. 销售;
 - 2.2.3. 待拆;
 - 2.2.4. 其他。

535. 如果住房单元被占用,居住者数目(代码H17)和人口记录数不得为零。如果记录没有人,要么该单元空置要么人失踪了。如以前在结构编辑中所述,专家必须制定确定单元是否空置的程序。如果列出单元被住用,但实际上是空置着的,必须制定一种办法来确定空置的类型,列为“未知”或者使用动态推算。如果单元列为空置,但由于可得到的居住者资料或人口记录的记数可以确定该单元实际被住用,那么住用状态必须改为“住用”。

536. 如果该值无效,居住者数目为零并且没有人口记录,在不使用动态推算时应分配“未知空置”。如果该值无效,但居住者数目不是零或存在着人口记录,应分配“住用”。对于无效值选择动态推算的国家(以推算空置类型)应至少使用两个特点,以便从该地区类似住房单元获得“已知”信息,或者,可以分配“未知空置”。

7. 住所：所有权类别 (H07)

537. 这个题目是指住所本身的所有权,而不是它们所在土地的所有权(联合国,1998年,第2.370段)。不应把所有权与保有权相混淆。应收集有关资料来说明下列情况:住所是否由公共部门(中央政府、地方政府、公营公司)拥有;住所是否私有(由住户、私营公司、合作社、住房协会等所拥有)。这个问题有时需要加以充分阐述,说明如果是分期付款或抵

押形式购买，则款额是否已全部付清。住所的类别按所有权分类如下：

1. 由所有者居住；
2. 由非所有者居住；
 - 2.1. 公有；
 - 2.2. 私有；
 - 2.3. 其他。

538. 如果所有权与保有权相关，在拟定编辑项目时应考虑到这一点；如果不相关，那么所有权类型可能独立于其他住房变量。如果“所有权类别”的值无效，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两种特点，其中可以包括墙壁的建筑材料、保有权、住房单元类型和房间数，以便从该地区类似住房单元获得信息。

8. 住所：房间数目（H08）

539. 一个房间的定义是，在四面有从地面到天花板或房顶的围墙的住房单元或其他住所中的一块空间，或至少有2米高，或有足够的空间能放置一张成人床，也就是说，至少有4平方米的面积。因此，房间总数包括卧室、餐房、起居室、书房、可居住的阁楼、佣人的房间、厨房、用于专业或商业目的的房间，以及其他用于或准备用于居住的单独空间，只要它们能满足墙壁或楼面面积的标准。通道、阳台、大厅、卫生间和厕所不能算作房间，即使它们符合标准。如果认为需要的话，可以为国家统计目的，单独收集有关面积小于4平方米、与“房间”定义的其他方面相同的面积的资料（联合国，1998年，第2.375段）。

540. 由于房间数目可能独立于其他住房变量，如果该值无效，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应当至少使用两个特点，如住房单元类型、墙壁建筑材料、保有权和空置状态，以便从该地区类似住房单元获得“已知”信息，

9. 住所：楼面面积（H09）

541. 楼面面积是指住房单元中可用的楼面面积，即根据住房单元外墙内侧测量而得的楼面面积，同时不包括不能居住的地下室或阁楼。在多住房的建筑物中，所有共用空间都不包括在内。对住房单元和集体住所应

采取不同的办法（联合国，1998年，第2.378段）。

542. 楼面面积可以涉及房间数目和/或卧室数目，因此国家的编辑小组在拟定编辑项目时不妨考虑到这一点。否则，楼面面积便会独立于其他住房编辑数据。可能需要指定测量单位，如平方米。如果该值无效，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点，其中包括住房单元类型、墙壁建筑材料、保有权和空置状况，以便从该地区的类似住房单元获得“已知”信息。

10. 住所：供水系统（H10）⁶

543. 根据联合国的建议（1998年，第2.381段），普查将获取的有关供水系统的基本资料是住房单元是否有自来水装置。这种资料将表明是否经由整个社区系统的管道向住区供水，或是用诸如高压箱、抽水机等个别供水装置。这个题目的查点单位是住房单元。还需要指出，在该单元内是否有水龙头，如果没有，水龙头距大门有多远。建议的距离是200米，因估计在这一距离内取得自来水可使住房单元内的住户不必过分费劲地取得用水。除了水龙头位置外，对供水来源也特别关注。因此，建议按供水系统将住房单元分类如下：

1. 单元内有自来水；
 - 1.1. 由社区系统供水；
 - 1.2. 由私人来源供水；
2. 在单元外但在200米范围内有自来水；
 - 2.1. 由社区系统供水；
 - 2.1.1. 专用；
 - 2.1.2. 合用
 - 2.2. 由私人来源供水；
 - 2.2.1. 专用；
 - 2.2.2. 合用；
3. 无自来水（包括距住所200米之外的来源取得的自来水）。

544. 社区系统受到政府机构的检查和管制。这种系统一般由政府机构经营，但在某些情况下由合作

⁶对于下列变量，查点单位实际上是住房单元：供水系统、厕所和排污设备、洗澡设备、厨房设备、照明和固体废物处理。

社或私营企业经营。

545. 与水有关的设施项目——供水系统、厕所和排污设备、洗浴设施和热水供应——可能应一起编辑。因为这些项目密切相关，一个项目遗漏或无效，可以使用其他项目来生成一个值。在没有自来水的地区，专家可能对这些单位使用专业性编辑。否则，该地区的其他单位也许有类似特点，建议在使用动态推算时对这些项目进行动态推算。

546. 如果与水有关的系统的值无效，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点。作为一个规则，这些特点也许包括住房单元类型，然后是厕所和排污设备，用来从该地区类似住房单元取得“已知”信息。

11. 住所：厕所和排污设备（H11）

547. 有些国家认为应将非抽水马桶的分类加以扩充，以区别某些广泛使用的类型，并显示一定程度的卫生情况。联合国（1998年，第2.386段）建议按厕所设备将住房单元分类如下：

1. 住房单元内有厕所；
 - 1.1. 抽水马桶；
 - 1.2. 非抽水马桶；
2. 在住房单元外有厕所；
 - 2.1. 抽水马桶；
 - 2.1.1. 专用；
 - 2.1.2. 合用；
 - 2.2. 非抽水马桶；
 - 2.2.1. 专用；
 - 2.2.2. 合用；
3. 无厕所。

548. 厕所设备和排污设备的类型是另一个与水有关的住房项目，应成为其他与水有关项目的共同编辑项目。与“私人”、“合用”、“专用”等有关的值可以用于确定值是否一致，如果它们不一致，确定采取什么编辑方法来解决。当存在一个或多个与水有关的变量时，可以确定未知或不一致信息的估

计数，而不使用“未知”或动态推算。然而，如果这样做不能提供一个有效的值，在不使用动态推算时应当分配“未知”。对于无效值选择使用动态推算的国家应至少使用两个特点，以便从该地区类似住房单元得到“已知”信息，作为一项规则，这些特点要包括住房单元类型，以及供水、墙壁的建筑材料、保有权和空置状态。

12. 住所：洗澡设备（H12）

549. 根据联合国的建议（1998年，第2.390段），应取得关于每一住所房地范围内是否具有固定的浴缸或淋浴装置的资料。这一题目的查点单位是住房单元。不妨收集补充资料以显示设备是否为住所住户专用以及是否有供洗澡用的热水或者只有冷水。然而，上文提议的区分也许不大适合世界上有些地区的国家需求。相反，也许应区别是否在住所内有单独的浴室、在建筑内有单独的浴室、在建筑内有共用浴室以及公用浴室。建议按有无洗澡设备及其类型，将住房单元分类如下：

1. 在住房单元内有固定浴缸或淋浴设备；
2. 在住房单元内无固定浴缸或淋浴设备；
 - 2.1. 在住房单元外有固定浴缸或淋浴设备；
 - 2.1.1. 专用；
 - 2.1.2. 合用；
 - 2.2. 无固定浴缸或淋浴设备。

550. 洗澡设备的类型应成为其他与水有关项目的共同编辑项目的一部分。可以用“私人”、“合用”、“专用”等值来确定值是否一致，如果不一致，应确定采取什么编辑方法来处理问题。当出现一个或多个与水有关的其他变量时，可以确定未知或不一致资料的估计数，而不使用“未知”或动态推算。然而，如果所有其他办法都不成功，如果该值无效，在不使用动态推算时应当分配“未知”。对于无效值选择使用动态推算的国家应至少使用两个特点，以便从该地区类似住房单元得到“已知”信息，作为一项规则，这些特点包括住房单元类型，然后是供水、墙壁的建筑材料、保有权和空置状态等。

13. 住所：厨房设备（H13）

551. 按照《原则和建议》（联合国，1998年，第2.395段），收集是否具备厨房这一方面的资料，将便于有关烹调所用设备种类（如火炉、煤气炉或明火）以及是否具有厨房洗涤槽和防止食物腐坏的食物贮存空间等资料的收集。建议按有无厨房或专用厨房用的其他空间，将住房单元分类如下：

1. 住房单元内有厨房；
2. 住房单元内有作厨房用的其他空间；
3. 住房单元内无厨房或作厨房用的其他空间；
 - 3.1. 在住房单元外有厨房或作厨房用的其他空间；
 - 3.1.1. 专用；
 - 3.1.2. 合用；
 - 3.2. 无厨房或作厨房用的其他空间。

552. 为了编辑厨房设备的资料，可以使用“私人”、“合用”、“专用”等值确定这些值是否一致，如果不一致，应确定采用什么编辑方法来处理问题。如果存在一个或两个厨房设备变量，可以确定未知或不一致资料的估计数，而不使用“未知”或动态推算。然而，如果该值无效，在不使用动态推算时应当分配“未知”。对于无效值选择使用动态推算的国家应至少使用两个特点，以便从该地区类似住房单元得到“已知”信息，作为一项规则，这些特点要包括住房单元类型，然后是供水、墙壁的建筑材料、保有权和空置状态等。

14. 住所：照明（H14）

553. 应收集关于住所内照明类别的资料，例如电、煤气或油灯提供的照明或其他一些来源提供的照明。如果以电照明，有些国家不妨收集资料以显示其电力是由社区供应、发电厂供应还是若干其他来源（工业装置、矿场等）。除照明类别外，各国应评估关于电力是否用于照明以外的用途（例如烹调、热水、暖气等）的资料。如果该国的住房条件允许从照明种类取得这项资料，则无需进行进一步调查（联合国，1998年，第2.398段）。

554. 如果该值无效，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点，以便从该地区类似住房单元获得“已知”信息，作为一项规则，这些特点包括住房单元类型，以及电力、墙壁的建筑材料、保有权和空置状态。

15. 住所：固体废料处理（H15）

555. 根据《原则和建议》（联合国，1998年，第2.401段），这一题目是指收集和处理来自住房单元住户的固体废料。查点单位为住房单元。按固体废料处理方法对住房单元分类的准则如下：

1. 由指定的收废料者定期收集固体废料；
2. 由指定的收废料者不定期收集固体废料；
3. 由自行指定的收废料者收集固体废料；
4. 住户将固体废料倒入当局监督的当地垃圾堆场；
5. 住户将固体废料倒入非当局监督的当地垃圾堆场；
6. 其他安排（包括由住户焚烧固体废料）。

556. 固体废料项目独立于其他住房变量。如果该值无效，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点，作为一项规则，这些特点可能包括住房单元的类型，然后是墙壁的建筑材料、保有、空置状态或厨房设备等，以便从该地区类似住房单元获得“已知”资料。

16. 住所：一个或多个住户居住（H16）

557. 多于一个住户居住项目独立于其他住房项目。如果该值无效，国家应对户主数进行计数，然后使用该数目。

17. 住所：居住人数（H17）

558. 通常居住在住房单元或其他住所的每个人应算作一名住户。因此，这一题目的查点单位是住所。然而，由于住房普查通常与人口普查同时进行，这一定义是否适用，应取决于在人口普查中收集和记录的个人资料是否指出当事人所在的地点，或指出其通常住址。应注意区别居住流动单元者（例如船、篷车和拖车式活动房屋）与利用这些单元作为交通工具

者（联合国，1998年，第2.407段）。

559. “居住人数”与人口记录相关，两者应是相同的。如果不同，应采取措施纠正居住人数项目或人口数量记录。通常，将调整居住人数使其等于单元内的人数。这一项目不应是“未知”，也不应推算。

18. 住户：户主的特征（H18）

560. 户主的特征通常从人口记录中获得，以帮助确定用于制定计划和进行分析的交叉列表使用的信息。这些项目，包括种族、宗教和收入，有助于确定不同的社会状况或需要。因为这些特征已经编辑到人口项目中，在此不需要进一步编辑。

19. 住户：保有权（H19）

561. 根据联合国的建议（1998年，第2.410段），保有权是指关于住户居住整个或部分住房单元的安插。查点单位是居住一住房单元的住户。按保有权将住户分类如下：

1. 住户成员拥有一住房单元；
2. 住户成员租用整个或部分住房单元；
 - 2.1. 住户成员作为主要租户租用整个或部分住房单元；
 - 2.2. 住户成员作为转租租户租用整个或部分住房单元；
3. 其他安排。

应单独考虑那些经房主许可或未经房主许可居住房屋而不付租金者，尤其是在这种情况普遍的地区。

562. 保有权可能与所有权类型(H12)有关，因此编辑小组在确定编辑项目时可能需要考虑这两个项目之间的关系。否则，如果保有权的值无效，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点，如住房单元类型和空置状况，以便从该地区类似住房单元获得“已知”信息。

20. 住户：租金和业主自己居住的住房费用（H20）

563. 租金和业主自己居住的住房费用项目独立于其他住房变量，除非很显然只对于租住单元发生

租金，以及只对于业主居住的单元发生业主费用。编辑小组必须审查每一种情况，并确定这些变量之间最适当的关系。如果该值无效，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点，以便从该地区类似住房单元获得“已知”信息。

B. 其他题目

564. 本节确定若干其他题目，这些题目可能对于许多国家规划其国家普查和调查有用。这些类别也是按查点单位类别（建筑物、住所和住户）分组，建议按字母顺序给出其编辑代码（联合国，1998年，第2.416-432段）。

1. 建筑物：住所数目（A01）

565. 建筑物内住房单元数目的编辑作为结构编辑的一部分已在第三章第219-220段解释。

2. 建筑物：电梯（A02）

566. 这一题目是指在多层建筑物内是否具备电梯（一封闭的平台可升降运送人和货物）。收集关于是否在多数时间内可用电梯（也就是电梯在多数时间内运行、但须定期维修）的资料（联合国，1998年，第2.419段）。

567. 如果建筑物只有一层或者是一个附属单元，即不应具备电梯。如果具备电梯，编辑人员必须确定哪一项占先，是层数还是具备电梯。如果电梯占先，就必须改变层数，或者通过将值改为“未知”或者通过使用动态推算获得另一个值。如果层数占先，而建筑物只有一层，对于“有无电梯”的答复必须改为“无”。

568. 如果有电梯，如果它用电，必须进行核对以确保建筑物内有电源供应。

569. 最后，如果电梯项目的值无效，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点，如建筑物类型和外墙的建筑材料，以便从该地区类似住房单元获得“已知”信息。

3. 建筑物：农场（A03）

570. 一些国家认为其国家普查有必要具体说明查点建筑物是否为农场建筑物。农场建筑物为农业用地的一部分，用于农业和/或住房用途（联合国，1998年，第2.420段）。

571. 农场建筑物项目独立于其他住房项目。各国可以选择与职业和行业人口项目是否一致。如果该值无效，在不使用动态推算时则分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点，以便从该地区类似住房单元获得“已知”信息。

4. 建筑物：屋顶的建筑材料（A04a）

572. 在若干情况下，用于建筑屋顶和/或地板的材料也许值得特别注意，并用以进一步评估建筑物内住所的质量。这一题目是指用于屋顶和/或地板的材料（虽然也许涉及建筑物的其他部分，例如框架或地基，视各国的特定需求而定）。查点单位是建筑物。仅列举主要的材料，就屋顶而言，可能是瓦片、水泥、金属板、棕榈、禾秆、竹或类似的植物材料、泥、塑料板或其他材料（联合国，1998年，第2.421段）。

573. 有时，有关外墙建筑材料的答复与有关屋顶建筑材料的答复不一致；例如，如果查明墙壁的建筑材料强度不足以支撑屋顶，就可能发生这种情况。如上所述，如果发生这种情况，专家必须决定是改变其中一个变量，还是使用“未知”。如果一个值无效，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点，如建筑物类型、住房单元类型、墙壁建筑材料、保有权和空置状态等，以便从该地区类似住房单元获得“已知”信息。

5. 建筑物：地板的建筑材料（A04b）

574. 所报告的地板建筑材料与屋顶和墙壁的建筑材料可能一致也可能不一致。如果国家编辑小组发现不一致或者无效的组合，就必须决定是否分配“未知”或使用推算矩阵来改变一个或多个答复。如果该值无效，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点，如建筑物类型、外墙建筑材料、住房单元类型、保有权和空置状态等，以便从该地区的类似住房单元获得“已知”信息。

6. 建筑物：修理情况（A05）

575. 这一题目显示该建筑物是否需要修理以及需要修理的种类。查点单位是建筑物。根据修理情况的建筑物分类包括：“不需修理”、“需要略为修理”、“需要一般修理”、“需要大修”和“不能修理”。略为修理是指建筑物及其组成部分的定期维修，例如修理破裂的窗子等。一般修理是指矫正一般缺点，例如屋顶沟槽缺失、大片灰泥破裂、楼梯无安全扶手等。在建筑物有严重结构缺陷时需要大修，例如屋顶板或瓦片缺失、外部墙面破裂和穿孔、缺少楼梯等。“不能修理”一词是指无法修复的建筑物：由于许多结构特点，这种建筑物应予拆除而非修复。该词多数情况下用于指那些只剩下框架，没有完整外墙和/或屋顶的建筑物（联合国，1998年，第2.422段）。

576. 建筑物的修理情况项目独立于其他住房变量。因此，如果该值无效，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点，如建筑物类型、外墙建筑结构和住房单元类型等，以便从该地区的类似住房单元获得“已知”信息。

7. 住所：卧室数目（A06）

577. 除了列出房间数目外，一些国家普查收集关于住房单元内卧室数目的资料，住房单元为这一题目的查点单位。卧室的定义是备有床铺、供晚间休息用的房间（联合国，1998年，第2.423段）。

578. 有时查点员报告的卧室数目的值大于房间数目的值。⁷如果这种情况发生，如果国家对于无效或不一致答复使用“未说明”，卧室数目一项应使用“未说明”。如果使用动态推算，应从以卧室数目为其因素之一的推算矩阵“估计”卧室数目。这样，卧室数目将不会大于房间数目，因为只有房间和卧室值相吻合时才更新卧室的值。最简单的情况将是一个线性数组，以房间数目作为单元，卧室的值位于单元内。更加复杂的推算矩阵可能包括住房单元内的人数和结构的类型。

⁷如果房间和卧室都存在，这两项应一起编辑，卧室数目不应超过房间数目。由于卧室数目是“附加”题目，只在两者都存在时才进行这项编辑。

579. 否则，如果卧室的值无效，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点（其中一个为房间数目），以便从该地区类似住房单元获得“已知”信息。

8. 住所：烹调燃料（A07）

580. 基于密切监测自然资源利用的需要，许多国家住房普查将烹调燃料这一题目包括在内。查点单位是住房单元；“烹调燃料”是指主要用于准备主餐的燃料。如果用两种燃料（例如电和煤气），应将最常用的一种列入。烹调燃料的分类取决于各国情况，可能包括电、煤气、石油、煤、木材、动物废料等。还应收集集体住所这方面的资料。特别是在集体住所数目大的国家内（联合国，1998年，第2.424段）。

581. 应该将对厨房设备的答复与对烹调燃料的答复一起进行编辑。编辑小组确定两个变量之间的关系，并确定一种编辑方法来核对两者是否一致。可能利用与“私人”、“合用”、“专用”等有关的价值确定值是否一致，如果不一致，确定采用什么编辑方法来纠正问题。当存在一个或两个烹调燃料变量时，可以对于未知或不一致信息作出估计，而不使用“未知”或动态推算。然而，如果该值无效，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点，其中包括厨房设备、建筑物类型、墙壁建筑材料、保有和空置状态等，以便从该地区类似住房单元获得信息，

9. 住所：暖气类型和暖气所用能源（A08）

582. 这一题目涉及住所暖气类型及所用能源。查点单位为所有住所。这一题目同一些国家无关，因为地理位置和气候的缘故，这些国家的住所不需要暖气。“暖气类型”是指为绝大部分空间提供暖气的系统种类：可能是为住所专用或一系列住所合用的中央暖气，或者不是中央暖气系统而是在住所内分别以炉、壁炉、或若干其他取暖装置提供暖气。“暖气所用能源”则与暖气类型密切相关并涉及主要的能源，例如固体燃料（煤、褐煤和煤与褐煤产品、木材）、石油、气体燃料（天然气或液化煤气）和电（联合国，1998年，第2.425段）。

583. 暖气类型和暖气所用能源相互有关，并与有无热水和住房单元内使用的其他公用设施如电和

管输送煤气有关。编辑小组在制定暖气类型和暖气所用能源的编辑规范时应考虑到是否能得到这些项目。暖气类型可能独立于其他住房项目，因此可能必须单独编辑。然而，在不知道“暖气所用能源”或其不一致时，程序能够核查照明所用能源类型。最后，如果该值无效，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点，如住房单元类型、墙壁建筑材料、保有和空置状况等，以便从该地区类似住房单元获得“已知”信息。

10. 住所：有无热水（A09）

584. 这一题目涉及住所内是否有热水供应。热水是指加热至一定温度以管道和龙头提供住所的用水。收集的资料可能说明在住所内是否有热水供应，或在住所外有专用或合用的热水或根本没有热水供应（联合国，1998年，第2.426段）。

585. 有无热水可能与给水加热的方式有关，尽管使用太阳能给水加热可能与其他住房单元无关。编辑小组必须视其他住房项目和地理位置决定适当的编辑方法。最后，如果值无效，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点，如有无自来水，以便从该地区类似住房单元获得“已知”信息。

11. 住所：管输送煤气（A10）

586. 这一题目涉及住区内是否有管输送煤气。管输送煤气通常是指经由管道运送的天然气或人工煤气，其消耗量加以记录。这一题目也许同一些国家无关，因当地未发展管道系统或缺乏天然气。

587. 管输送煤气与照明和烹调燃料以外的其他住房项目无关。编辑小组必须确定适当的编辑办法以及如何检查一致性。如果该值无效或不一致，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点，如暖气所用能源、建筑物类型、住房单元类型、墙壁建筑材料、保有和空置状态等，以便从该地区类似住房单元获得“已知”信息。

12. 住所：电话（A11）

588. 这一题目涉及住房单元内有无电话。电话

是指电话线而非电话机，因为一条电话线可连接多部电话机（联合国，1998年，第2.428段）。

589. 有无电话可能与编辑时使用的其他住房项目无关。然而，如果某些地区没有电话，编辑小组在拟定编辑项目时应考虑到这一点。如果“电话”项目的值无效，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点，如住房单元类型、墙壁建筑材料和保有权，以便从该地区的类似住房单元获得“已知”资料。

13. 住所：住房单元的利用（A12）

590. “住房单元的利用”是指该住房单元是否完全用作居住（住宅）用途。住房单元可用作居住，并可用作商业或制造业或其他用途（联合国，1998年，第2.429段）。

591. “住房单元的利用”独立于其他住房项目。如果该值无效，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点，如住房单元类型、墙壁的建筑材料、保有权和所有权等，以便从该地区类似住房单元获得“已知”信息。

14. 居住者：车辆数目（A13）

592. “车辆数目”是指住房单元居住者通常可使用的汽车和运货车数目。“通常可使用”一词是指居住者拥有根据若干其他较长期协议（如租约）可使用的汽车和运货车（联合国，1998年，第2.430段）。

593. 车辆数目独立于其他住房变量。如果国家有些地区没有任何车辆，专家不妨考虑特定地区的特殊编辑项目。否则，如果该值无效，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点，如住房单元类型、墙壁的建筑材料、保有权和所有权等，以便从该地区类似住房单元获得“已知”信息。

15. 居住者：耐用设备（A14）

594. 收集关于是否具备诸如洗衣机（A14a）、洗碗机（A14b）、冰箱（A14c）、冷藏箱（A14d）、电视（A14e）和个人计算机（A14f）等耐用设备的资料，视国家情况而定（联合国，1998年，第2.431段）。

595. 对于多数设备来说，单元内必须拥有电力以便可以使用设备。当出现这些项目时，编辑小组应当考虑一种编辑方法检查是否有电（可能有以煤气作动力的冰箱或“冰柜”）。此外，如果特定国家需要自来水以便运行洗衣机或洗碗机，编辑也需对此予以说明。还可利用编辑方法确定是否应有特定项目，视是否有电和水而定，并在出现不一致情况时采取适当的行动。还有，当一个国家的特定地区没有电力或自来水时，专家在拟定编辑项目时可能需要承认这一点。如果该值无效或不一致，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点，如住房单元类型、电力、墙壁的建筑材料和保有权等，以便从该地区类似住房单元获得“已知”信息。

16. 居住者：可供住户使用的室外空间（A15）

596. 这一题目是指是否具备供居住住房单元住户成员休闲活动使用的户外空间。分类包括户外空间为住房单元的一部分（例如，附联式住房的后院）、房屋邻近的户外空间（例如，公寓房屋毗连的后院和游乐场）、距住房单元10分钟步行距离为公共休闲地区一部分的户外空间（例如，公园、体育中心和类似场地），或在10分钟步行距离内没有户外空间（联合国，1998年，第2.432段）。

597. 可供住户使用的户外空间量独立于其他住房项目。然而，在某些地区或某些类型的建筑物中，可能没有户外空间。编辑小组可能需要在制定编辑方法时考虑具体情况。如果该值无效，在不使用动态推算时应分配“未知”。对于无效值选择动态推算的国家应至少使用两个特点，例如建筑物类型和住房单元类型，以便从该地区类似住房单元获得“已知”信息。

C. 住用和空置的住房单元

598. 上述编辑项目针对的是被住用的住房单元。然而，空置住房单元和住用住房单元可能具有不同的特点，因而不使用相同的编辑项目。国家普查/统计局编辑小组将需要对各类单元制定不同的编辑项目，因为对于空置住房单元并不收集所有住房项目。编辑小组将需要特别注意推算矩阵变量，因为这些变量很有可能不同。

附 件

附件一

导出变量

599. 为了最佳地利用普查或调查数据，国家常常需要将其他变量组合在一起的变量。例如，关于经济活动状况的项目（第四章，D.1节）已经是收集的若干普查变量的组合。不必在国家普查/统计局每次想进行特殊制表时都编写一个程序对信息进行再编码，数据处理专家可以编写一个程序进行一次再编码，在个人记录上存储再编码的资料，然后在以后的制表中使用。国家普查/统计局需要决定，在他们确定是否生产和存储信息时，在制表中使用再编码的频率如何及一个特定的再编码的相关程度如何。重要的是要记住，再编码也占个人记录的空间。人口数目越大，使用的空间越多。

600. 许多变量可以这样生成。例如，如果报告了出生日期，但未报告年龄，那么第一次可以通过从普查参照日期减去出生日期来确定年龄，将这一资料存储在记录中。同样，住户收入可以将每个人的收入相加得到，将这一总数放入住房记录供以后使用。

601. 有时，导出变量来自单个记录中一个或几个输入项目的组合，有时来自几个记录。例如，“不从事经济活动——上学”分类可能要求查看至多四个项目的答复。在制作表格或规划补充表格时，使用导出变量将使编程更容易和更有效率，并有助于使各个时期的数据可以比较。下文介绍一些可能的导出记录的实例。

A. 住房记录的导出变量

1. 住户收入

602. 住户收入的导出变量是得到的一个住户内所有个人各类收入的总和。收入资料类别可能包括工资、个人营业收入、利息和股息、社会保障和退休收入、汇款、特许权使用费和租金。如果也收集了总收入，在编辑期间应通过将各个类别相加核对每个人的总收入。然后，将这一总数与记录的总收入核对。如果相加得出的收入与报告的总收入不相等，编辑小

组必须制定一项改正计划。要么改变总数以反映各个部分的总和，要么必须改变一个或几个个别收入类别。当住户内个人的总收入已确定时，可以通过将个人收入相加得到住户收入变量。

603. 编辑小组必须考虑到住户内一人或多人因业务失败或其他原因而有负收入的情况。在这种情况下，该人的收入将降低而不是提高住户总收入。

2. 家庭收入

604. 家庭收入的导出变量是得到的家庭内所有个人各类收入的收入总和。家庭与住户不同，家庭通常只包括有关的个人，但这一定义视特定国家的情况而定。对于一些国家，住户和家庭是相同的，因此家庭收入的导出变量是不必要的。家庭收入信息类别可能包括工资、自己营业收入、利息和股息、社会保障和退休收入、汇款、特许权使用费或租金。如果也收集了总收入，编辑期间应通过将各个类别相加核对每个人的总收入。然后将这一总数与记录的总收入进行核对。如果相加得到的收入与报告的总收入不相等，编辑小组必须制定纠正的计划。要么必须改变总数以反映各部分的总和，要么必须改变一项或多项个人收入种类。当所有个人的总收入已确定时，通过将家庭内部个人收入相加得到家庭收入。

605. 编辑小组必须考虑到家庭中一人或几人因业务失败或其他原因而有负收入的情况。在这种情况下，该人的收入将降低而非提高家庭总收入。

3. 家庭类型

606. 有时，确认“家庭类型”对于某些制表是有用的。例如，家庭类型的导出变量可能是从1到8，表示一个家庭的类型和组成。家庭类型的导出变量可用于察看各种特点对于家庭结构的影响。

607. 如《人口和住房普查的原则和建议，第一次修订版》（联合国，1998年，第2.60-2.66段）所

述，家庭的定义在各国有所不同。一种定义是，家庭由户主和一个或几个因血缘、婚姻和收养而与户主有亲属关系的住在同一住户的其他人组成。住户内与户主有亲属关系，所有人都是他或她的家庭的成员。不过，并非所有住户内都有家庭，因为住户可以由一组无亲属关系的人或一个独居的人组成。

608. 主题专家可能按户主性别和是否有亲属将家庭类别分为“已婚夫妇家庭”或“其他家庭”。例如，美利坚合众国使用下列基于家庭类型的数据代码，这些代码是从对性别和关系问题的答复中推导出来的。

(a) 代码 1 和 2: 已婚夫妇家庭。户主及其配偶被作为同一住户的成员查点的家庭；户主是男性时用代码 1，户主是女性时用代码 2。

(b) 代码 3: 其他家庭: 男性户主，无妻子。户主为男性且户主没有配偶的家庭。

(c) 代码 4: 其他家庭: 女性户主，无丈夫。户主为女性且户主没有配偶的家庭。

(d) 代码 5 和 6: 非家庭住户，住户非家庭。因此没有户主的配偶或其他亲属；户主为男性时用代码 5，户主为女性时用代码 6。

(e) 代码 7 和 8: 单人住户。单独居住的一人被视为一个住户，但不是一个家庭，因为没有其他亲属。户主为男性时用代码 7，户主为女性时用代码 8。

609. 确定上述类别比例的一个较简单办法是获得称作“有配偶的已婚户主”的导出变量。可以根据户主住户是否有配偶，对户主的婚姻状况进行再编码。在每个住房单元，对人口记录中有配偶的人进行扫描。将单独的代码“是”或“否”放在住房记录适当的栏内。对于集体住户，对本变量可以留空白或分配另一个代码。然后，对于人口表格，有配偶的已婚者将在制表时被识别出来。

4. 有亲属关系的人

610. 有亲属关系的人是与户主有某种亲属关系的人。有亲属关系的人的导出变量是与户主有亲属

关系的所有人之和。在大量没有亲属关系的人在住房单元内住在一起的情况下，该值特别重要。当许多没有亲属关系的人以这种方式住在一起时，常常将他们划类为住在“集体住所”或“集团住所”。

611. 在发展数据集时，国家统计局常常编制按年龄分列的不同组别有亲属关系的人的导出变量。例如，可以编制 0-5 岁、5-17 岁、6-17 岁、0-17 岁有亲属关系儿童、65 岁以上有亲属关系的人和 75 岁以上有亲属关系的人的导出变量。

612. 例如，家庭内“有亲属关系的儿童”可能包括住户内 18 岁以下的户主自己的子女和与户主有亲属关系的其他人，而不管婚姻状况如何，但户主配偶除外。有亲属关系的儿童可以也可以不包括养子女，因为他们与户主没有亲属关系，但这项决定视国家的情况而定。

5. 家庭的工作者

613. 有时，国家想按工作者数目比较住户变量，如按住户规模分列的收入分配情况和与每个受抚养人相对应的工作者数目。国家可以通过将在参照期如一周或一年（或者一个日历年或最后 12 个月）内至少工作一周的人的数目相加得到导出家庭内工人数目。如果仅收集了“上一周”的资料，国家可以利用这个期间工作的人数。

6. 全套管道

614. 普查调查表上的一些项目用来获得有关管道设施的数据。这些项目通常与有无自来水、抽水马桶和浴缸或淋浴有关，通常获得有关住用和空置住房单元的这些项目。全套管道的导出变量可以有助于比较一个时间点两个地区或群体之间或随着时间推移的社会经济条件。例如，在三种设施——自来水（热水或冷水）、抽水马桶和浴缸或淋浴——都有的情况下（或在单元内或者在单元所在的建筑物外），可以获得全套管道的导出变量。编辑小组将需要确定最合适的全套管道变量组。

615. 在本例中，如果三个项目分别设问，就可

以获得导出变量，在编辑作业期间，将确定是否存在所有三个项目的总结。如果住房单元有自来水、抽水马桶和浴缸或淋浴，那么就有“全套管道”。不同时具备三个项目，它应缺乏“全套管道”。

7. 全套厨房

616. 普查用来从调查表有关厨房设备、冰箱和水槽的项目中获得关于厨房设备的数据，这些项目既为住用单元又为空置单元收集。当厨房设备（电、煤油或煤气炉、微波炉和非便携烤箱或烹调用炉）、冰箱和有自来水的水槽位于作为住所被查点的同一建筑物内时，该单元可能被认为具备“全套厨房设备”。这些设备不必位于同一房间。

617. 在上述三个项目分别设问时可以获得导出变量，在编辑作业期间，确定是否存在所有三个项目。“缺乏全套厨房设备”包括下列条件：存在所有三种特定的厨房设备，但设备位于不同的建筑物内；存在一些但不是所有的设备；或者在作为被查点住所的同一建筑物内三种特定的厨房设备都不具备。

8. 租金总额

618. 国家通常收集有关现金或合同租金的数据。现金租金通常不包括公用设施的费用。有时国家也需要有关租金总额的资料。租金总额是现金或合同租金加上公用设施（电、煤气和水）和燃料（包括石油、煤、煤油和木材）平均每月费用的估计数，如果是由出租人负责支付这些费用的话。租金总额意在消除对于是否将设施和燃料作为支付租金的一部分纳入的不同做法所产生的差异。不支付现金租金而住用的出租单元，在制表时可以单独列为“无现金租金”。

619. 租金总额的导出变量通过将支付的租金数额和为公用设施支付的数额相加得出，如果这些款项是单独收取的话。

B. 人口记录的导出变量

1. 经济活动状况

620. 经济状况导出变量可能对于制表有用，但

它要求几个变量的资料。按照《人口和住房普查的原则和建议，第一次修订本》中的类别，重新安排一些变量是必要的。该导出变量可以包括九个类别：

从事经济活动

就业

1. 在工作
2. 有职业，但不工作
3. 在武装部队

未就业

4. 找工作
5. 失望的工人

不从事经济活动

6. 主妇
7. 学生
8. 不能工作
9. 其他

621. 由于在许多表中使用各种从事经济活动状况的类别，编辑小组应考虑在数据记录中插入导出变量，而不是让数据处理员在制表过程中重新对经济状况进行分类。制表过程中的重新分类可能引入差错，因为不同的数据处理员可能以略微不同的方式重新分类；即使一个程序也可能视编辑或制表的特定要求而进行不同的重新分类。从事经济特点研究的专家应制定关于导出变量的规范。

2. 小家庭数目和小家庭的亲属

622. 所有国家都像核心家庭一样发展。考虑下面的大家庭，三角形代表男性，八角形代表女性。住户有一个户主和配偶（数字1和2），有两个孩子（数字4和7）。他们的儿子（人4）与他们的儿媳妇（人3）结婚，他们因儿子而有两个孙辈（人5和人6）。女儿（人7）与他们的女婿（人8）结婚，他们因女儿而有两个孙辈（人9和人10）。

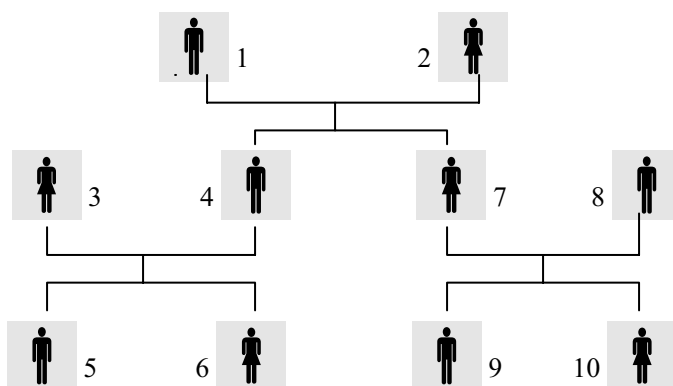
623. 在多数普查中，如果编辑小组想研究类似图A.1所示家庭的结构，可能难以区分孙子孙女，因为人5、6、9和10与户主的关系都记录为“孙辈”。小家庭和小家庭成员的重新编码将允许更详细地分

析家庭结构。

624. 小家庭的一个定义是“已婚夫妇（丈夫和妻子作为同一住户成员查点），有或没有 18 岁以下未婚的子女”。编辑小组可能想加上“有或没有 18

岁以下从未婚子女的单亲，与户主或户主的配偶住在一个住户并有亲属关系，但不包括户主或户主的配偶”。小家庭的数目不包括在家庭计数之内，因为小家庭成员已作为户主家庭的一部分计算。

图 A.一.1. 大家庭图解



625. 小家庭的导出变量，包括亲属的数目和类型，可以在处理数据期间定义。可以拟定一个特殊的编辑项目，以帮助基于住户内的关系确定小家庭。随着每个小家庭被确定——非户主 / 配偶的一对（有或没有子女）或非户主 / 单亲和孩子——按顺序给每个小家庭指定数字。然后为各种关系分配代码：家庭“户主”是代码 1，家庭“户主”的配偶是代码 2，家庭“户主”的孩子是代码 3。小家庭存在的条件是，必须至少有一对小家庭亲属存在：或者“户主”和“配偶”（代码 1 和 2），或者“户主”和“孩子”（代码 1 和 3）。当家庭户主、配偶和孩子都住在一起时，对于该小家庭代码 1、2 和 3 都存在。小家庭分类为：已婚夫妇小家庭，有孩子或没有孩子；母子小家庭；父子小家庭。单亲包括维持单亲家庭或单亲小家庭的人。已婚夫妇包括已婚夫妇家庭和已婚夫妇小家庭的丈夫和妻子。

626. 在编制导出变量时，使用与户主的关系来确定家庭内的关系；因此，普查中关系编码越详细，对于小家庭的匹配性就越好。例如，如果“孩子的配偶”关系有自己的代码，程序将能够用相反性别的“儿媳 / 女婿”来匹配“儿子 / 女儿”，从而产生一个小家庭。没有这条补充资料，匹配仍可进行，但在匹配时“其他亲属”可能模糊或可能被错误匹配。同样，该程序将能够为有“配偶”和“侄儿 / 侄女”的“户主的兄弟姐妹”匹配代码。

627. 图 A.一.2 中给出的示例表明有两个小家庭的住户：小家庭 1 包括人 3（小家庭 1 的户主）、人 4（配偶）和人 5 和人 6（小家庭 1 的子女）。小家庭 2 包括人 7（小家庭 2 的户主）、人 8（配偶）、人 9 和人 10（小家庭 2 的子女）。

图 A.一.2. 有两个小家庭的住户实例

人的编号	关系	性别	小 家 庭	
			编号	关系
1	户主	男性		
2	配偶	女性	1	1
3	儿子	男性	1	2
4	儿媳妇	女性	1	3
5	孙辈	男性	1	3
6	孙辈	男性	2	1
7	女儿	女性	2	2
8	女婿	男性	2	3
9	孙辈	男性	2	3
10	孙辈	女性		

3. 自己的子女

628. 有时，国家想生成关于“自己子女”即从生物学上讲户主和 / 或配偶的子女。例如，“自己的子女”可能是户主亲生的 18 岁以下从未结婚的儿子或女儿、（父母一方的）继子女或户主的养子女。表格可以列明进一步分为与父母或只与父母一方生活在一起的“自己的子女”。在这种情况下，顾名思义，与父母双方生活在一起的户主自己的子女只是在已婚夫妇家庭才有。

629. 在小家庭内，“自己的孩子”可能是 18 岁以下未婚的子女，可能是儿子、女儿或继子女；或（1）母子小家庭中的母亲，（2）父子小家庭中的父亲，或（3）已婚夫妇小家庭中任一方配偶的收养子女。

630. 按照编辑小组选择的定义，导出变量“自己的子女”可能是某人（通常是妇女）自己子女的总和。有时，用户可能需要自己子女的按年龄分列的更详细资料。例如，在美国，编制 6 岁以下和 6 至 17 岁的自己子女的数目的导出变量。这些值放入所有女性的记录中。这种资料特别用于确定有“自己子女”的劳动妇女的特点。

4. 住房内的父母

631. 这些数据审查单亲家庭的子女与父母住同一住房单元的子女相比所具有的特点。编辑通过利用关系代码确定某人有多少父母在住房内，获得该导出变量。程序考虑每个子女的关系代码，利用该资料与有关小家庭的资料一起确定有多少父母住在住房单元内。

5. 目前在校年数

632. 有些国家可能问两个有关教育的问题：

- (a) 该人目前是否上学；
- (b) 最高教育程度。

633. 在这些国家，编辑小组常常发现，在查点时一个人实际在上学的情况下有两个项目不相匹配。有时，有可能引起该人的最高教育程度比目前在校年数少一年。如果该人处于一系列年级或级别的中间阶段，统计数据将不受影响。然而，如果该人正上特定级别系列的第一年级，则与其他来源的数据匹配也许是不可能的。例如，一个上一年级的人将被记录为在校但没有教育程度。同样，进入中学的人将被记录为在校，但教育程度将为小学最高年级（或级别）。

634. 可以为此项目组合编制称为“目前在校年数”的导出变量。如果该人目前没有上学，代码将与最高教育程度相同。如果该人目前上学，编辑将对教育程度年级（或级别）加 1，分配到“目前在校年级”中。

635. 有些国家问三个有关教育的问题，即上述两个项目和关于是否完成最高年级的第三个项目。如果也获得了这一信息，也应在确定“目前在校年数”时使用这一信息。

附 件 二

调查表格式与键入的关系

636. 在普查或调查中人口项目的两个最常见的调查表格式是个人页和住户页。

637. 个人页包括一个或两个人口信息封面页，每个人有单独的一页。这种方法之所以有用是因为一

个人的所有信息出现在一页上，使信息容易收集。这种格式还使得较容易在查点期间核验其内部一致性。

个人页可以合并在一个装订在一起的小册子内，以便易于在现场进行处理，如图 A.二.1 所示。

图 A.二.1. 附有个人页的调查表实例

人 X 的个人页		人 X+1 的个人页	
项目 1	项目 10	项目 1	项目 10
项目 2	项目 11	项目 2	项目 11
•	•	•	•
•	•	•	•
•	•	•	•

图 A.二.2. 附有个人页的调查表流程实例

个人页			
项目 1		项目 11	
项目 2		项目 12	
项目 3		项目 13	
等等		等等	
•			
•			
•			

638. 对个人页上项目的编码和键入基本上是一种机械操作，在此过程中，不指望编码员 / 数据输入操作员对所提供信息的有效性进行评价，而是分配适当的码或按键。图 A.二.2 说明了一页上记录的特定个人的信息流程。输入一页上该人的数据比起一边键入一边翻页要容易些。随后在计算机编辑阶段核验其有效性。

639. 住户页在一页上列出住户的所有信息，或者可能情况下在若干页上列出住户的所有信息，其中在每一页上列出所有住户成员。以这种方式列出住户成员之所以有用，是因为不必为每个人打印调查表，因而节省了空间。此外，查点员可以在收集数据时比较住户成员的输入内容。

图 A.二.3. 调查表实例，所有人列在同一页上的住户页

住 户 页					
人	项目 1	项目 2	项目 3	项目 4	等等
1					
2					
3					
4					
5					
•					
•					
•					

640. 第三种方法是每个人有单独的表格，然后查点员在查点期间或查点后汇集成一个松散的小册子。这种方法之所以有效，是因为查点员只收集住户所必需的确切数目的表（页）。缺点是在转移或其他处理期间这些表可能分开，如果普查部门不能够将同一住户的表重新汇集在一起，就会造成许多潜在的编辑问题和覆盖面问题。

641. 调查表页的实际尺寸也是一个考虑因素，不仅是为了查点，而且是为了键入。在编码和键入阶段，文件必须平放在工作桌表面，编码员或数据输入操作员必须能够容易地找到表格上项目的位置并处理项目。

642. 当所有信息都在一页上时，工作人员可以容易地键入住户页，显然这样速度会快些，因为数据

输入操作员不必翻页。图 A.二.4 说明住户页上的信息流。

643. 在键入超出一页的人口或住房信息时，可能会出现这个问题。为了解决这个问题，国家统计局有可能采取下述两种方法之一。

644. 一次输入一个人的数据。数据输入操作员可以键入一个人在若干页中第一页上的信息行，然后翻到第二页和随后的页。在第一个人在各页上的输入完成之后，数据输入操作员再返回该住户的住户页的第一页，键入第二个人、第三个人等的信息。只要数据输入操作员在整个键入过程中不串行，这类键入方法就行得通。虽然可以编写计算机编辑程序，以便在个人项目错误键入到另一个人的行时对信息进行调整，但程序本身很难编写。

图 A.二.4. 住户页调查表流程实例，每页多个人

住 户 页				
个人	项目 1	项目 2	项目 3	等等
1	→			↺
2	→			↺
3	→			↺
4				
5				
•				
•				
•				

645. 可以一次输入一页的数据。数据输入操作员可以键入整页的信息，然后移到下一页。在这里，数据输入操作员键入本页的所有信息而不管涉及多少人。然后，数据输入操作员翻页，键入下一页所有人的信息。这里，视键入的类型（带计算机编辑或不带计算机编辑）而定，可以包括或不包括跳越模式。

无论如何，在计算机编辑阶段，来自各组键入数据的记录都将被汇集在一起，然后必须对个人编号的任何错误键入进行处理。

646. 在下面的实例（图 A.二.5）中，住户的人口统计信息带来不常见的键入问题，因为普查为所有人的所有项目都获得一个答复。

图 A.二.5. 多人住户页实例，没有键入问题

住 户 页					
个人	关系	性别	年龄	等等	
1	户主	男	40		
2	配偶	女	35		
3	孩子	女	18		
4	孩子	男	12		
5	兄弟姐妹	男	35		
6	兄弟姐妹的配偶	女	30		
7	兄弟姐妹的孩子	男	5		
8	兄弟姐妹的孩子	女	3		
等等					

647. 然而，同一住户的第二页(图 A.二.6)可能产生一些键入问题。例如，如果该国选择仅收集 5 岁以上的人使用语言的情况，第八个人为 3 岁，她的信息将为空白。数据输入操作员应使这个孩子的该项目保持空白，否则计算机编辑就得在以后努力予以纠正。

最低年龄的人参加劳动力情况，低于最低年龄的妇女的生育及所有男性的生育情况。在图 A.二.6 中，数据输入操作员可能错误地把人 6 有关已生育孩子的信息（此例中为 4）键入到人 5 的位置。然后计算机编辑将删除该男子的生育率项目并推算该妇女的生育率，但可能不会推算出正确的值。

648. 同样，其他项目也可能是空白的，如低于

图 A.二.6. 多人住户页实例，可能有键入问题

住户页 2				
个人	语言	劳动力	生育孩子数	等等
1	语言	是		
2	语言 1	否	3	
3	语言 1	否	0	
4	语言 1			
5	语言 1	是		
6	语言 1	否	4	
7	语言 1			
8				
等等				

649. 由于费用或空间限制，一个国家必须多次使用住户表格。然而，在人口较少时，或该国拿得出

额外费用时，个人页表格由于错误键入出现的匹配错误可能要少于住户表格出现的错误。

附件三

键入的考虑因素

650. 许多国家使用扫描设备，或者为光学标志辨认设备（OMR）或者为光学识字设备（OCR）。当作业顺利高效，费用不多时，这些都比键入有优势。然而，许多国家，即使一些承诺使用扫描的国家，也可能无法负担最初的启动费用或查点期间和查点后持续不断的维护费用。此外，许多国家把为普查而获得的扫描器用于许多目的，包括其他调查，诸如输入和输出表等行政记录。不过，将文件键入到扫描器中所涉及的基本技能只在使用相同或类似的机器时才能转移。

651. 键入的好处之一在于，键入过程中学习的技能可以转移到国家普查/统计局和其他政府机构的其他活动中。在普查培养了专门的数据输入操作员之后，这些数据输入操作员可以为各类后续调查键入数据。这些调查可能包括查点后调查和其他调查，如生育率或住户收入和支出调查。工作人员还可以键入行政记录，诸如生命记录和贸易、移民入境和移民出境和海关等记录。

652. 因此，当一个国家决定在普查中是否使用扫描设备时，它也应当决定该国是否将继续使用这些机器。多用途机器在普查之后仍可长期有用。不过，键入数据的国家普查/统计局将会发现，学习的技能能够转移，机器本身在国家统计部门或政府其他部门可以继续发挥作用。在就键入或扫描作出决定时，可能应当考虑到对设备的继续使用问题。

A. 输入数据

1. 扫描

653. 使用光学或其他扫描装置来采集数据的国家，通常在采集过程中并不对其采集的数据进行校正，尽管修改可能依赖嵌入系统的跳越模式。然而，选择键入数据的国家可能有几个选择，视它们需要键入的速度有多快和需要多少手工检验而定。每一种选择均依赖于编辑小组的要求、数据输入操作员的技能

和编辑程序是否先进。在大型普查业务中，最大的问题是将数据全部键入。必须使用能更快地产生结果的办法。《原则和建议》指出，有足够的机器、良好的培训和专门的数据输入操作员的国家可实现较快的周转时间。但是，其中每一项都需要充足的资金，而资金并不总是可以得到的。

654. 所需数据输入设备的数量和类型取决于选定何种数据输入方法、可用于该普查阶段工作的时间、国家面积、数据采集工作的下放程度以及一系列其他因素。就键盘数据输入而言，平均输入速度通常为每小时 5 000 至 10 000 次按键，有的操作员速度远低于此，有的则大大超过。影响操作员速度的因素有：(a) 辅助软件和程序；(b) 操作员任务的复杂程度；(c) 设备的工效特点、可靠性和速度；(d) 是否一直有工作的问题；(e) 所应聘工作人员的训练情况和才能；以及(f) 工作人员的积极性（联合国，1998 年，第 1.193 段）。

2. 埋头键入

655. 埋头键入有两种形式。第一种是在遇到什么数据项目就键入什么数据项目，采取不跳越的模式。在这种情况下，键入的速度更快，因为数据输入操作员在遇到无效或不一致的信息时不必停下来。这样还可以更准确，因为他们是更加机械地执行任务。第二种埋头键入形式是对于无效或不一致的结果核查调查表，这样过程将较慢，并需要键入人员拥有更多的专门知识。必须认真地考虑速度所代表的价值。似乎矛盾的是，这种方法也可以提高准确度，如果数据输入操作员发现数据记录得正确却被错误编码的话。有时错误键入本身也可能立即被发现和纠正，因为编辑程序包可以自动核对。

(a) 不带跳越模式的埋头键入

656. 当所有输入项目被键入或被手工跳过时，可以保持一定的节奏，某些跳越模式将不排除有效但

暂时不一致的信息。例如，如果一个人被记录为男性，多数编辑小组将要求有关生育的整个部分被跳过。在这种情况下，数据输入操作员将击键跳过（使用空格键或箭头跳过男性的记录或年幼女性的记录），因为所有域将是空的。然而，这需要时间，空格可能不完全正确。例如，数据输入操作员输入的空格可能太多或太少，其他项目因为对齐不正确可能被错误键入。如果所有域都以这种方式键入，那么在没有跳越模式时可以键入这一信息。例如，当数据输入操作员遇到育龄成年女性时，该女性已出生的孩子、存活的孩子或去年出生的孩子等项目已经收集并编码，就键入所有项目。如果键入了生育信息，计算机编辑程序可以确定哪一个项目或哪些项目有效，哪些项目必须修改。当编辑过程确定该人是成年女性，但生育信息空白，必须使用动态推算或其他适当方式来获得用于制表的生育信息。如果由于跳越模式实际信息丢失，数据输入操作员仍可以将屏幕退回到他们没有键入的项目，这种形式的键入能够在编辑过程而非在键入过程发现和纠正性别、年龄和生育情况之间的不一致。

(b) 带跳越模式的埋头键入

657. 第二个埋头键入方法涉及使用跳越模式的键入。如果编辑小组要求跳越模式通常反映查点员收集数据的方式，如果跳越模式容易采用，数据输入操作员能迅速学会键入模式，键入就容易和快捷。如果键入模式非常复杂，数据输入操作员就会被搞糊涂，总是键入到错误的位置。采用涉及所键入的大部分记录的有限模式时，带有跳越模式的键入的成本效益就最高。

658. 编辑小组需要确定适合他们国家的普查或调查的跳越模式。例如，对儿童也就是低于该国界定的潜在就业年龄而言，跳过所有就业项目是有道理的。这些项目常常占人口项目的一半，因此对儿童而言跳过这些项目可以提高效率，年龄处于分界线的儿童或该国对童工问题感兴趣的特殊情况除外。

659. 编辑小组逐个项目决定哪些年龄组将包括哪些项目。工作人员可以将项目编组以便容易管理跳越模式。

660. 并不总是很容易就跳越模式作出明确的决定。例如，考虑下面的顺序：

-
1. 该人的公民身份如何？
 - 在该国出生（跳到项目 3）
 - 归化
 - 非公民
 2. 该人在哪一年进入？
 3. 下一个项目
-

661. 对于在该国出生的人，可以建立一种跳越模式，从 1 跳到 3，即跳过进入年份这一项。然而，有时数据操作员违反该跳越模式，也许是由于查点员或编码员犯了错误，也许是由于键入错误。这涉及许多因素，其中包括数据输入操作员的技能水平、文化环境、调查表的布局和屏幕的布局。编辑小组常常一起讨论，确定类似这样的跳越模式是否合理。

3. 交互式键入

662. 交互式键入可以在普查输入阶段使用，但也许更适合于调查，特别是在指定的项目可能影响调查结果的小型调查。交互式键入可能涉及手动或自动纠错，视可用来进行修改或纠正的信息而定。

663. 考虑小型调查的情况。对于小型调查来说，每一个答复都很重要。例如，如果一国进行 1% 的抽样调查，每一个答复便代表 100 个人、100 个住房单元或 100 项农业财产。在这些情况下，人口统计学家和其他社会科学家通常希望对过程进行大量控制。

664. 可以几种方式确立控制。人口统计学家和其他专家可以自己键入数据，在此过程中核查哪些答复错误、无效或不一致，并利用数据收集表中记录的信息。他们常常直接面对收集的信息，立即解决冲突、错码或其他不一致情况。有时，他们可能选择将不完整或无效的调查表发回现场。这类交互式键入能产生最佳效果，因为人口统计学家也充当数据输入操作员，但这样做耗费要大得多，没有几个国家能够负担得起。

665. 编辑小组可以制定非常详细的编辑规则，

以确定键入过程中出现特殊情况时数据输入操作员必须怎样做。他们可以决定，对于每个未解决的无效代码，数据输入操作员应键入什么代码。编辑小组可以解决详细规则没有涉及的情况，并可以修改规则（尽管键入的前一部分和后一部分可能不一致）。

666. 在埋头键入中发挥重要作用的跳越模式在此也很重要。因为采用埋头键入时，数据输入操作员必须注意并学会使用任何跳越模式。如上所述，跳越模式可以提高键入的速度，但常会造成一些质量方面的损失。对于交互式键入，一项来自经验的通常规则是跳越越少，质量越高。

B. 测试键入指令

667. 在拟定键入指令之后，国家普查 / 统计局在决定实际操作之前必须使实际数据输入操作员测试键入指令，不管是否使用埋头键入法。在测试键入指令时，可以改正系统的缺点，从而获得最佳的键入结果。

C. 验证

668. 国家普查 / 统计局还必须决定，应当采用哪一级别的验证。许多专家建议 100% 地进行验证。在这种情况下，所有项目都被重新键入，以确保收集

的数据正是输入机器供计算机处理的数据。然而，彻底验证并不实用，这或者是因为该国没有时间重新键入所有数据，或者是由于缺乏财政或人力资源。还有，如果测出的键入错误率很低，数据输入操作员犯的错误很少，彻底验证可能就不必要。

1. 依附式验证

669. 验证技术要么是依附式，要么是独立式。在依附式验证中，数据输入操作员在其他工作人员原先键入的数据之上键入数据。当击键不同时，软件包就告诉数据输入操作员，根据所用的程序，数据输入操作员要么取消以前的数据，要么注明不一致。由于键入的数据来自原始调查表，通常数据输入操作员自己便可以就原来的键入是否错误作出恰当的决定。

2. 独立式验证

670. 在独立式验证中，数据输入操作员从零开始重新键入数据；他们可以利用原始调查表建立一个完全独立的键入数据文件。然后使用一个计算机程序比较产生的两个文件，即原来键入的数据集和验证数据集，以测试不一致之处。可能需要一些手工操作来纠正无效和不一致的击键。

附件四

流程图样本

671. 编辑小组的任务之一是编拟编辑过程中所使用的变量的关系结构。流程图有助于确定变量之间的各种关系，从而补充了编写编辑规范的工作。这些相关关系既有助于主题专家和数据处理专家使编辑过程形象化，又促进两个群体之间的交流。

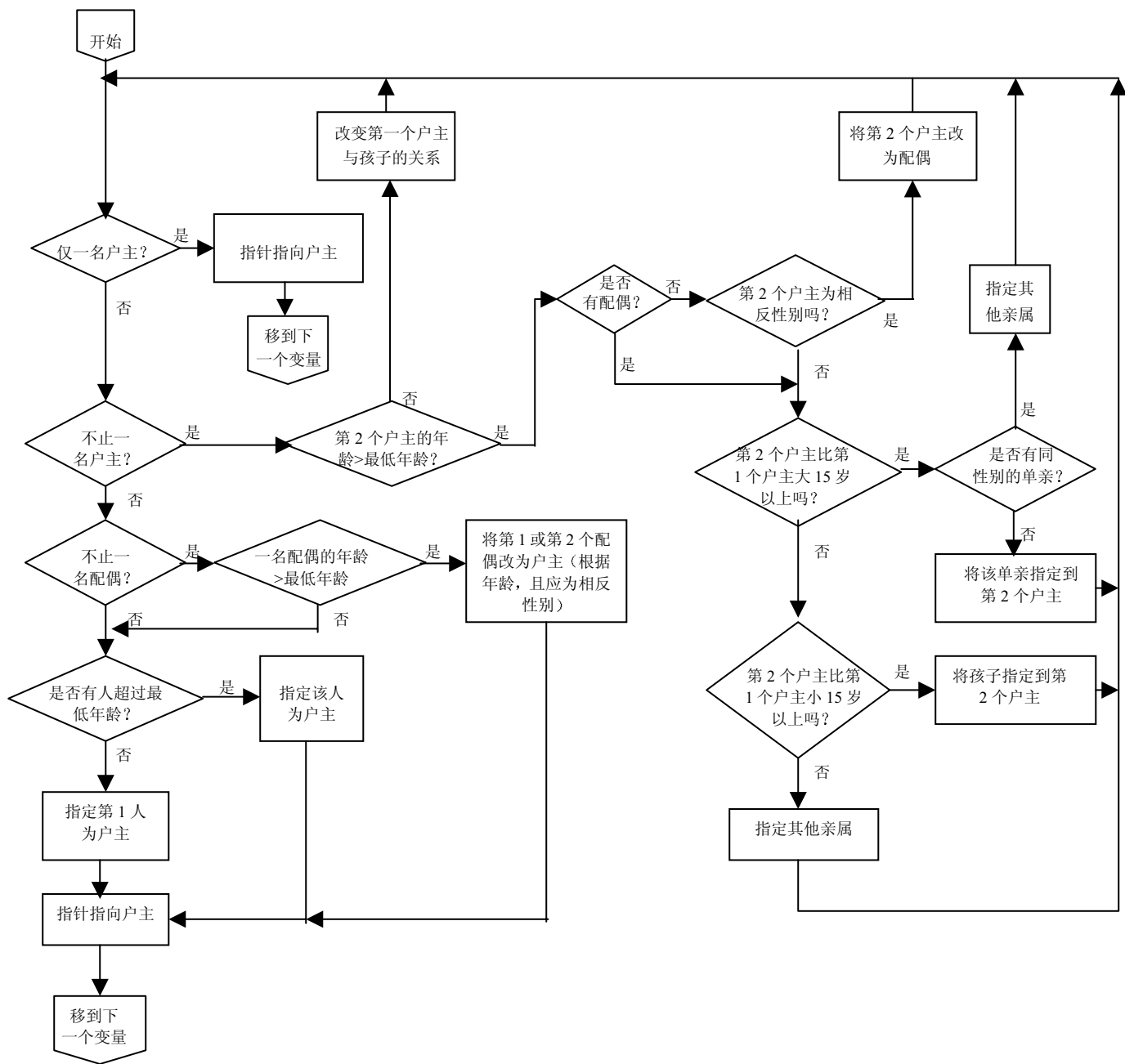
672. 下列各页给出三个流程图样本：

- (a) 用来确定户主的流程图；
- (b) 用来确定住户内是否有配偶的流程图；

(c) 用来编辑户主和配偶的性别变量的流程图。提供这些流程图仅为说明的目的，应当据此看待它们。编辑小组不妨根据国内的情况对流程图作必要的修改。

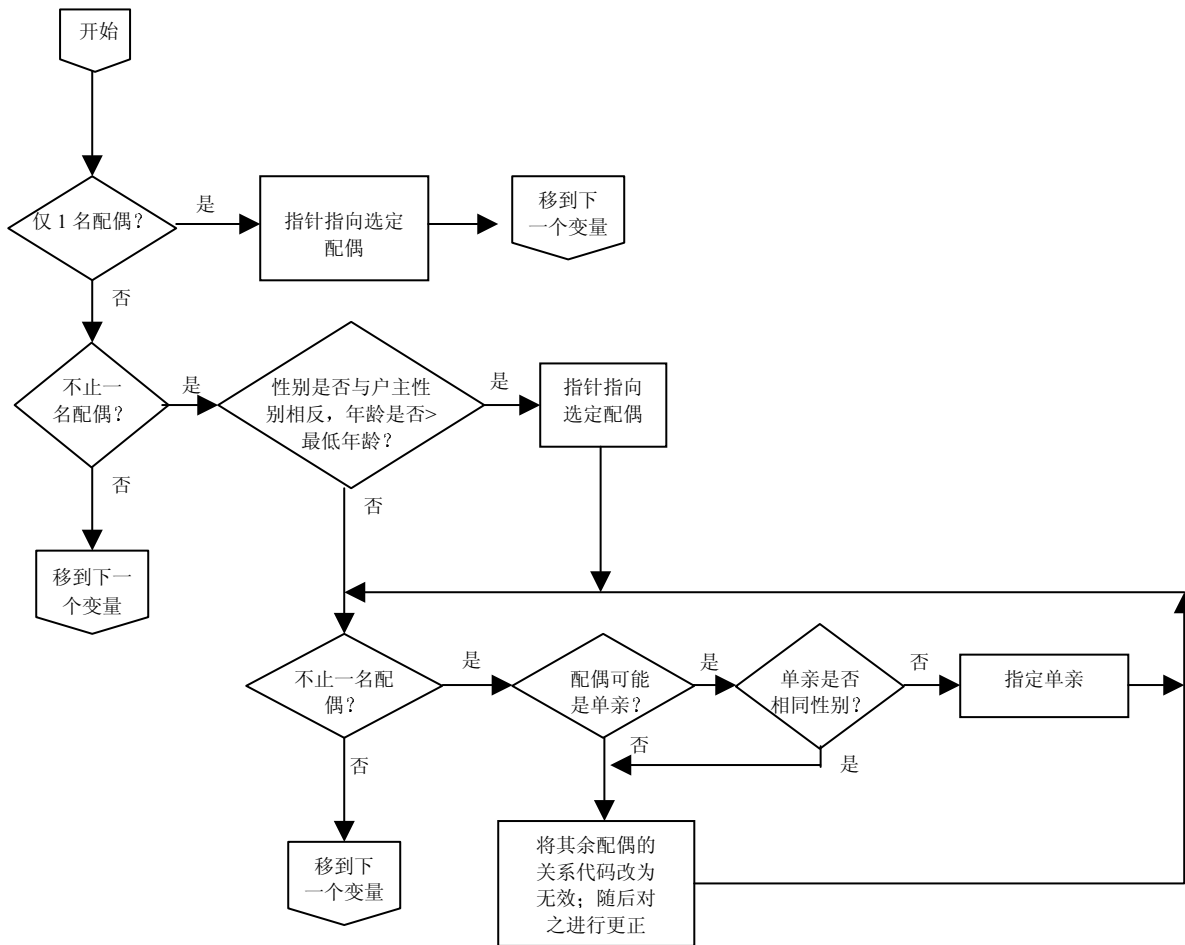
673. 应对普查中的各个变量编制编辑流程图。编辑小组应当共同努力编制流程图，数据处理专家应将它们与编辑规范一起用来开发编辑普查数据的计算机程序。流程图和编辑规范应精心制成文档，以便于将来普查和调查的数据处理。

图 A.四.1. 确定户主的流程图样本



*-由编辑组指定最低年龄

图 A.四.2. 确定住户是否存在配偶的流程图样本



注: HH=户主

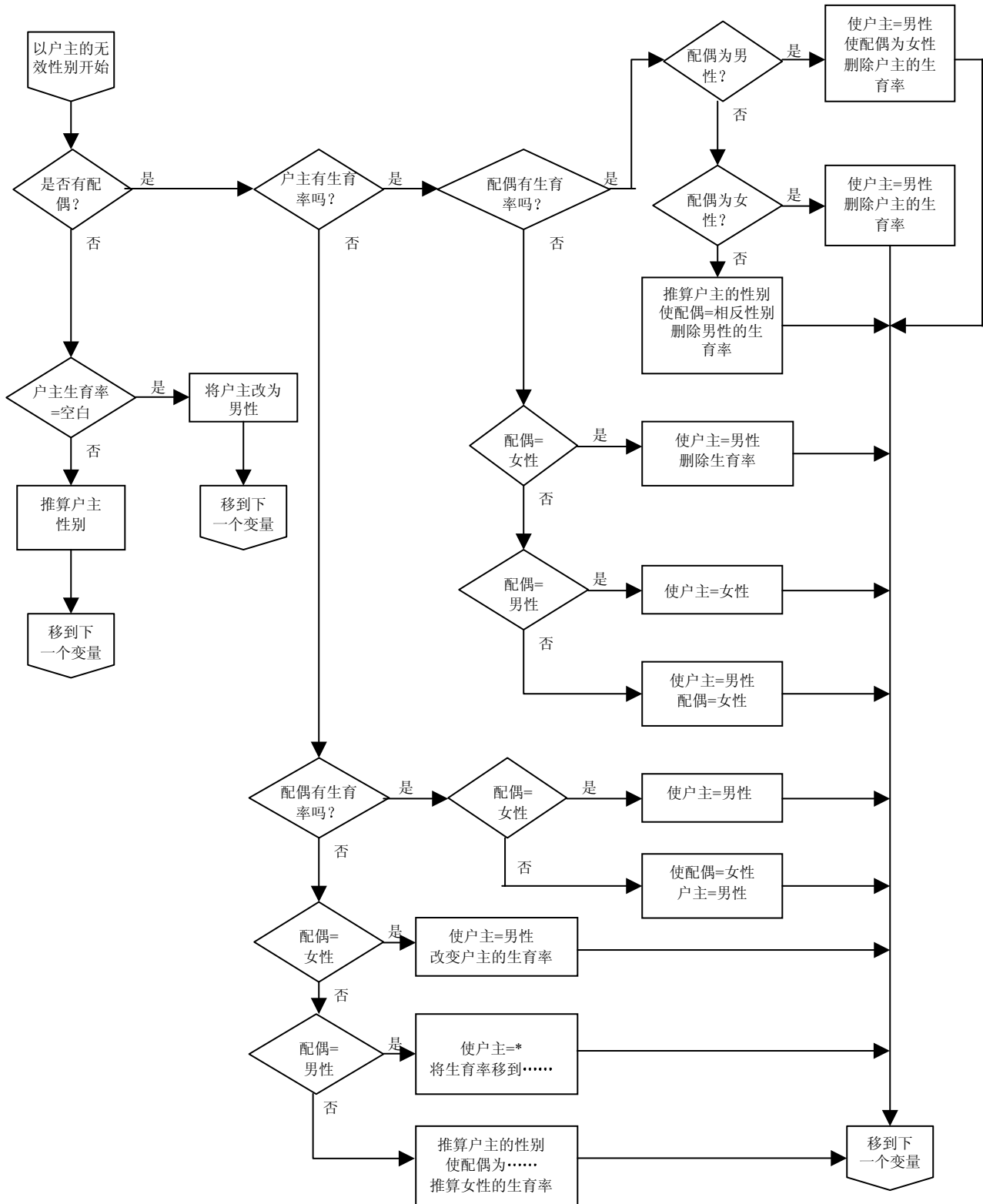
*-由编辑组指定最低年龄

图 A.四.3. 编辑户主和配偶性别变量的流程图样本



注: HH=户主

图 A.四.3. (续)



附件五

推算方法

674. 已制定了几种推算方法。下文所述的多数推算方法曾在 Kalton 和 Kasprzyk (1982 年, 1986 年)、Sande (1982 年) 和 Sarndal、Swensson 和 Wretman (1992 年) 的论文中作过论述。

675. 推算方法可以分为随机性或确定性, 取决于推算数据中的随意性程度。

676. **确定性推算方法**包括演绎性推算; 基于模型的推算方法, 如平均推算和回归推算; 酌情包括最近邻推算。

677. **演绎性推算方法**允许肯定地推导出缺失或不一致的值。这常常以对调查表其他项目的答复模式为依据。

678. 推算技术经常必须将不一定为真值的值替换掉。下述几段概述一些常用的推算程序。

679. 除了单一提供者动态推算算法之外, 下文所述方法涉及一次推算一个项目。这样, 在各个推算级别, 记录上的项目按顺序一个接一个被考虑。通常, 这项工作只需要只考虑明确涉及该项目或少数几个密切相关的变量的编辑项目。由于可能有些显式或隐式编辑项目, 把该项目与随后过程考虑的项目联系起来, 该程序便可以得出推算值, 同时通过目前考虑的编辑项目, 将造成该过程随后阶段考虑的其他编辑项目不能通过。只有当考虑完整的一组编辑项目包括所有隐式编辑项目时, 才能保证通过所有编辑项目。隐式编辑项目是可以合乎逻辑地合并两个或多个显式编辑项目推导出的编辑项目。

680. 在下列说明中, “通过编辑记录”指已通过有关所涉项目全部编辑的记录。“未通过编辑记录”指至少未能通过所涉项目一次编辑的编辑记录。

681. **总体平均推算**为所有未通过编辑记录的缺失或不一致项目指定已通过的编辑记录的项目平均数。这种方法产生合理的点估算, 但如果使用标准

的差异估计程序计算差异估计数, 这种方法就不那么吸引人了。除非推算比例非常低或使用经过修改以对推算作出解释说明的差异估计程序, 否则差异估计数可能被严重低估。

682. **分类平均推算**使用界定的推算类别, 以产生具有一定相似性的记录组。在每一类别内, 为所有未通过编辑记录的缺失或不一致项目推算已通过的编辑记录的项目平均数。这与总体平均推算非常类似, 但对分布和差异估计问题的影响可能没那么严重。

683. **回归推算**或更普遍的**基于模型的推算**使用通过编辑记录的数据来回归变量, 对该变量的推算是一组预测值变量所需要的。回归过程中的预测值可以是来自调查表或辅助变量的项目。然后使用回归方程式推算缺失或不一致项目值的值。这是基于模型的推算的特殊情况。这种方法常常用于商业调查应用中的连续变量, 在这些应用中, 使用以前情况的数据经常可以令人满意地预测当前情况的值。

684. **最近邻推算**或**距离函数匹配**从“最近”已通过编辑记录或未通过编辑记录分配一个项目值, “最近”按其他已知变量的距离函数来界定。可以在推算类别内使用这种方法。通常认为这种方法适合于连续变量, 但也可以适用于非数字变量。

685. **随机推算法**包括回归或任何其他确定性方法加上随机残差项, 以及流水式或非流水式方法。

686. 就每一种确定性方法而言, 有一个随机性对应方法。可以通过给源自确定性推算的推算值增加来自适当范围的随机残差项来实现这一点。这种程序将有助于更好地保留数据文件的频率结构。Kalton 和 Kasprzyk (1986 年) 审查了采用这种技术的一些方法。

687. **流水式和非流水式推算**试图产生比确定性方法所能够产生的更加现实的推算值可变性。流水

式推算程序用（随机）选定的值替代缺失或不一致的值，这些值源自已通过的当前调查或普查中的编辑记录。非流水式推算程序根据其他来源进行推算，这些来源常常是历史性数据，如以前的相同调查或普查的情况。有若干不同形式的流水式和非流水式方法。

688. **随机总体推算**是最简单的流水式推算形式。对于每个未通过的编辑记录，从所有已通过的编辑记录组中选择一个已通过的编辑记录，并推算出未通过编辑记录中该项目的报告值。

689. **类别内的随机推算**也使用推算类别，以便把提供者记录的随机选择局限于被认为与需要推算的记录具有某种相似性的记录组。

690. **顺序流水式推算**也使用推算类别，它的好处是一次通过数据文件就足以完成推算过程。该程序以每个推算类别的非流水式值开始，数据文件中的记录被依次考虑。当检测到一个已通过的编辑记录时，该项目的值将取代存储的推算类别的值。当检查到未通过的编辑记录时，其缺失或不一致的值被存储的值取代。推算类别数目不能过大，因为必须确保在每一个推算类别都有提供者。如果数据文件中记录的顺序为随机性的，这种办法将几乎等于类别内的随机推算。这种程序的一个缺点是，它常常导致多次使用提供者，从而可能对项目的分布和差异估计数造成不利影响。

691. **分层流水式推算**是对顺序流水式推算的强化，使用了大量推算类别。当初步推算类别中找不到提供者时，类别以分层方式瓦解，直到找到提供者为止。

692. **单一提供者流水式推算法**的目标是从一个单一提供者未通过的编辑记录推算数据。因此，它们允许共同推算被编辑认定为有问题的记录上的所有项目值。在实践中，目标常常是为记录中密切相关的各部分变量使用一个单一的提供者。这种办法具有很大的好处，它能够像上述流水式推算方法那样，不仅较好地保持边缘分布，而且能够保持共同的频率分布。单一提供者流水式推算法的另一个好处是它们减少推算值的问题，这些问题将使随后各部分变量中

考虑的编辑项目无法通过。在单一提供者流水式推算法中，已通过的编辑记录是已通过适用于该部分的所有编辑的记录。未通过的编辑记录是未能通过至少一次此类编辑的记录。

693. **Fellegi-Holt 编辑和推算法**(Fellegi 和 Holt, 1976 年)同时考虑所有编辑项目。Fellegi-Holt 编辑和推算法的一个特点是推算规则源于相应的编辑项目，而无明确的规范。对于每个未通过的编辑记录，首先执行一个找错步骤，确定推算的最小变量集以及可接受用来推算的值的范围，然后进行推算。在多数实施过程中，通过在编辑所涉但不要求推算的其他变量的基础上进行匹配，从已通过的编辑记录中选择单一的提供者。偶尔找不到合适的提供者，则必须使用缺省推算法。

694. **新推算法**(Bankier、Luc、Nadeau 和 Newcombe, 1996 年; Bankier, Lachance 和 Poirier, 1999 年)类似于 Fellegi-holti 法，因为它同时考虑所有编辑，并不明确指定推算行动并且从单一提供者进行推算。对于每个未通过的编辑记录，它依据可利用的潜在提供者确定变化最少的推算行动。这保证将有提供者可供利用。与 Fellegi-Holt 不同，新推算法先寻找提供者，然后确定变化最少的推算行动。新推算法通过利用编辑所涉的所有变量(包括可能被推算的变量)进行匹配来寻找提供者，附近的数字变量匹配项加上多数(但不一定是全部)其他变量的匹配项可以满足这种方法。确定基于各个潜在提供者的推算行动，并确认那些变化最少的推算行动。这种方法还考虑接近变化最少的推算行动；有时它们可以产生一些似是而非的推算记录。最后，随后随机选择其中一个变化最少和变化接近最少的推算行动，进行推算。

695. 虽然 Fellegi-Holt 法和新推算法在计算方面有着较高要求，但已有高效的算法，因此利用现代计算机实施和应用它们是可行的。对于新推算法来说尤其如此，新推算法与 Fellegi-Holt 法相比能够便利地处理更大型的编辑和推算问题。

696. 所有上述推算方法都对每个缺失或不一致值产生一个单一的推算值。所有这些方法都在某种程度上使该项目的通常分布失真，并在使用标准差异

估计程序时导致不适当的差异估计数。失真的程度依推算数量和所使用的方法而有很大的不同。

697. **多项推算**是 Rubin (1987 年) 建议使用的一种方法, 这种方法通过为每个需要推算的值进行几次推算来解决这一问题。然后, 从已完成的数据集中, 可以产生该项目的估计数。由此产生单个的合并估计数以及一批差异估计数, 后者将表明推算的值的确定性。多项推算法的一个缺点是它需要更多数据处理和计算估计数的工作。

698. 在多数推算系统中, 都组合使用一些推算方法; 在可能情况下使用演绎性推算, 然后使用一种或几种其他程序。多数国家统计局使用某种形式的动态推算法进行普查编辑和推算。顺序流水式推算和 Fellegi-Holt 法目前使用得最为普遍。在目前使用 Fellegi-Holt 法国家统计局中, 其中一个已转用新推算法, 另有一些国家正在考虑使用这种方法。然而, 考虑到预期的主要读者群, 本手册主要介绍一种顺序流水式推算形式。

附 件 六

计算机编辑软件包

699. 有了价格低廉的微机，各国应能够以彻底和有序的方式编辑普查和调查数据。直至最近，每个国家还不得以定制的方式自行编写编辑程序，需要代价高昂的调试和处理时间。但随着标准计算机编辑软件包的问世，就能十分便利地满足各国的编辑需要，而且并不需要很多的数据处理专门知识。

700. 利用计算机编辑软件包的一个有利条件是，如果正确利用，数据将保持一致和无错误，以便能够更及时地产生汇总表。可以利用许多计算机软件包如 SAS 和 SPSS 或其他高级语言来编写编辑程序。或者，一国可选择利用明确为编辑普查和调查数据编写的计算机软件包。就多数国家而言，一般编辑使用软件包，其速度比使用专门编写的程序来得快，因为与专门编写的程序相比，软件包将不需要同样多的数据处理知识。

701. 高质量的计算机编辑软件包将规定主题专家与程序设计人员之间进行交流。高质量的编辑软件包应允许注解语句或伪代码安排在编辑代码的旁边，除非编辑代码本身能使主题专家一目了然。人口学家和其他专家应能逐行审阅程序并准确理解程序在做些什么。

702. 一国可以考虑利用的任何现成的计算机编辑软件包必须能够执行和产生各种报告，以便进行编辑普查数据所需的各种检查、测试和推算。即使数据处理专家编写专用编辑程序时，这些要求也适用。

软件包必须满足下列要求：

(a) 有能力键入和（或）核查登录项数据。这种软件包应规定跳跃模式。例如，编辑小组可以决定遇到男性时生育率信息必须跳过；

(b) 进行结构编辑，这将使得能够确定应当存在的记录类型实际上是否存在，例如其中是否包括每个序列号的住房记录；

(c) 如果记录遗漏就产生记录（或）给现有记录加权；

(d) 确定每个变量有一个有效值；

(e) 存储已经编辑的全部或部分记录；

(f) 测试同一记录内和不同记录之间两个或多个特点是否一致。其中一个子集是测试住户内部的一致性，将各个答案与前面住户成员的答案进行对照检查。如果该国选择利用动态推算，就依靠流水技术推算各值；

(g) 利用一个记录或来自多个记录的数个值建立一个派生变量，并将派生变量插入适当的记录；

(h) 确定和消除重复的记录；

(i) 生成错误和变动日记，按小区域列示。

软件包或程序通常一次编辑一项记录，但现用的软件包也允许进行记录间检查，特别是住房单元内的记录间检查。

词 汇 表

阵列	一组数值。阵列有时叫做矩阵，可用来存储重复性的数值数据。
审计线索	跟踪字段中各值变化及每次变化的原因和来源的方法。审计线索一般在初始访问结束后开始。
自动更正	在无人力干预下由计算机更正数据错误。自动数据编辑的一个方面（Pierzchala, 1995年）。
记录间编辑	在涉及一份以上调查记录的字段上进行的编辑。统计编辑是记录间编辑的一个例子，因为分布产生于调查中所有记录各组字段之上（Pierzchala, 1995年）。
分类平均推算	利用界定的推算类别以建立具有某种程度相似性的记录组。
清洁记录	无遗漏值而且通过所有编辑的记录（Pierzchala, 1995年）。
代码表	一个数据项的所有允许（许可）值清单。
非流水式	初始静态矩阵；一种在更正开始前元素就给出而且在更正期间这些元素不改变的更正基础。例如，更正基础可以是上年的数据。业经修改的非流水式可按（可能聚集的）当前信息调整非流水值。
整套编辑	显式编辑和隐式编辑的结合。为产生可行的推算区域所需（如果有人希望推算符合编辑）（Pierzchala, 1995年）。
一致性编辑	检查确定性关系，如何部分相加等于总体或“收获的面积”始终小于“种植的面积”（Pierzchala, 1995年）。
数据捕获	使收集到的数据变为机器可读形式的过程。在捕获数据的软件的子模块中经常进行基本编辑检查。
演绎性推算	一种经常基于对调查表上其他项目所作答复的模型，能够肯定地推断遗漏或不一致数值的方法。
确定性编辑	一种如果违反则以概率 1 指明数据错误的编辑。例子：年龄=5 岁和地位=母亲。对比随机编辑（Pierzchala, 1995年）。
确定性推算	只有在一个字段的一个值促使记录符合所有编辑的条件下才出现这种情况。发生在某种情况下（例如一总体的各部分相加不等于总体）。在调查数据自动编辑和推算时要检查的第一个解法（Pierzchala, 1995年）。
距离函数	对于数值数据而言，根据接受者（候选者）和提供者双方记录的匹配变量界定并用来对相似性概念进行量化的函数。用于在流水式推算中找出匹配记录（Pierzchala, 1995年）。
距离函数匹配	从“最近”通过编辑的记录为未通过编辑记录分配一个项目值，其中“最近”利用其他已知变量方面的距离函数来界定。

提供者推算	将需要推算的每份记录，即接受者或候选者记录与来自例如流水式推算法中界定提供者群体的一份记录配对的方法（Pierzchala, 1995 年）。
编辑（定义 1）	对每个变量能够假定的值的逻辑限制（Pierzchala, 1995 年）。
编辑（定义 2）	检测禁止答案组合的规则（Pierzchala, 1995 年）。
编辑线索	见“审计线索”。
显式编辑	由主题专家明确编写的编辑（对比显式编辑与隐式编辑）（Pierzchala, 1995 年）。
未通过编辑记录	在编辑和推算中，至少未能通过有关所涉项目一次编辑的记录。
Fellegi-Holt 自动推算法	系这样一种更正法：尽量减少数据项变动的数目，而且利用 Fellegi-Holt 模型确定推算项目各组可接受的值或范围。可适用经由非流水法或流水法的顺序推算或同时推算。
Fellegi-Holt 系统	系指 Fellegi 和 Holt 在其 1976 年《美国统计协会杂志》论文中提出的假设和编辑与推算目标。Fellegi-Holt 模型的一个关键特点是它证明需要隐式编辑，以保证数据字段中不推算的一组值始终导致符合所有（情况）的最终（推算）记录。
标记	标记是一种变量，用来注明关于另一个或另一些变量的有用信息。例如，如果改变一个项目，从无效变至有效，就能利用标记来注明原始信息，或者注明该项目的值已经改变。
流程图	必须完成的各项职能的图解式说明。
手编辑	数据输入计算机前由人进行的编辑（见“人工编辑”）（Pierzchala, 1995 年）。
埋头数据输入	一种数据输入方式，指数据输入时数据输入机器不检测数据中的错误，允许操作员快速高效地录入数据。
抬头数据输入	一种数据输入方式，指数据输入时数据输入机器检测数据中的错误，允许操作员立即纠正错误。（Pierzchala, 1995 年）。见“交互式键入”。
流水式推算	一种推算方法，指提供者记录取自当前的抽样数据组（相形之下，非流水式指提供者记录来自过去调查数据的推算方法）（Pierzchala, 1995 年）。
隐式编辑	一种未加说明的编辑，按照逻辑派生于由主题专家编写的显式编辑（Pierzchala, 1995 年）。
推算	向某一字段分配一个值，或者作为未答复的值，或者取代被断定与一组编辑相矛盾的记录值（Pierzchala, 1995 年）。
交互式键入	一种数据输入方式，指在数据输入时数据输入机器检测数据中的错误，允许操作员立即更正错误。见“抬头数据输入”。
内部一致性	本用语有关给定抽样单位变量之间的关系，它是多数调查过程中进行编辑的原因（Ford, 1983 年；Pierzchala, 1995 年）。
宏观编辑	通过（1）检查聚合的数据，或（2）适用于记录总体的检查，检测个别错误。这种检查以估计为基础（Granquist, 1987 年；Pierzchala, 1995 年）。

人工编辑	在数据输入计算机前由人进行的编辑（见“手工编辑”）（Pierzchala, 1995年）。
匹配	在流水式推算过程中，使提供者记录与接受者（候选者）记录匹配的行为（Pierzchala, 1995年）。
匹配变量	用来找出接受者（候选者）记录与提供者记录之间匹配的变量（Pierzchala, 1995年）。
微观编辑	在记录一级数据上进行的传统编辑。宏观编辑的逻辑反义词（Pierzchala, 1995年）。
微宏观编辑	一种编辑过程，指详细的微观编辑由一种组合式的微观/宏观编辑与宏观/统计数据取而代之。组合过程中的微观编辑的详细度小于第一种。想法是“开发基于一种‘对估计产生影响’原理的调查编辑，而不是基于一种‘捕捉所有数据不一致’原理的调查编辑”（Granquist, 不同日期；Pierzchala, 1995年）。
最小集	需要推算的将保证所有编辑获得通过的最小集字段（Pierzchala, 1995年）。
基于模型的推算	利用来自通过编辑记录的数据以回归需要根据一组预测变量为其推算的变量。
多项推算	为需要推算的每个值推算数次并随后为该项目产生估计值。
多元编辑	一种编辑方式，指利用多元分布评估数据并找出范围以外值（Pierzchala, 1995年）。
最近邻推算	从“最近”通过编辑的记录为未通过编辑的记录分配一个项目值，其中“最近”要利用其他已知变量的距离函数界定。
新推算法	类似于 Fellegi-Holt 法，因该法同时考虑所有的编辑，不明确规定推算行动，而且从单一提供者推算。新推算法研究每项未通过编辑的记录，根据可加利用的潜在提供者确定最低限度变化的推算行动。
无答案	不完整的调查表或遗漏的调查表（Pierzchala, 1995年）。
范围以外值	位于按照某种范围确定规则所定的某个范围以外的项目值（Pierzchala, 1995年）。
总体平均推算	将通过编辑记录的项目平均值分配给所有未通过编辑的记录的遗漏或不一致的项目。
通过编辑记录	在编辑和推算期间，已通过有关所涉项目全部编辑的记录。
指示符	指示符为一变量，用来为一个项目或其他变量作标记以备日后查阅。例如，利用指示符来注明“户主”和“配偶”的行号，以便日后用来确保配偶双方性别相反而且双方都已婚。
生产运行	在从编辑或制表程序消除初始“错误”后处理大量数据的行为。
伪代码	书面编辑指令或说明。
质量错误	可能扭曲数据质量的错误：例如，导致偏差的系统错误（Granquist, 1984年；Pierzchala, 1995年）。
定量编辑	适用于按连续比例计量的字段的编辑（Pierzchala, 1995年）。

类别内的随机推算	利用推算类别将提供者记录的随机选择范围限制到被认为与需要推算的记录具有某种相似性的一集。
随机总体推算	就每项未通过编辑的记录而言，从一组完全通过编辑的记录随机选择一项通过编辑的记录，并为未通过编辑记录的有关项目推算其报告值。
记录	调查数据的磁存储和计算机可读的表示。通常每份调查表有一项记录，不过也有可能将一份调查表的数据分裂为一项以上的记录，例如人口和住房（Pierzchala, 1995 年）。
回归推算	利用通过编辑记录的数据，来回归需要依据一组预测变量推算的变量。
检索	在流水式推算过程中，检索提供者记录的行为（Pierzchala, 1995 年）。
顺序流水式推算	在一系列变量按顺序编辑时发生的推算，只将经编辑的值用作后续流水变量。
相似性	在数值数据中，根据规定的匹配变量两项记录接近的概念。利用距离函数按某些标准对此概念量化（Pierzchala, 1995 年）。
单一提供者流水式推算	从单一提供者未通过编辑的记录推算数据，允许对编辑确定有问题的记录的所有项目值进行联合推算。
统计编辑	根据调查对象数据的统计分析所作的一组检查：例如，两个字段的比率位于对假定有效报告人该项比率的统计分析确定的限度之间（Greenberg 和 Surdi, 1984 年；Pierzchala, 1995 年）。
统计推算	统计推算的一个例子将是利用某个回归模型，其中相关变量将进行推算，而独立变量的系数派生于假定有效的答案（Pierzchala, 1995 年）。
统计匹配（流水式推算中）	按照某些统计标准使提供者记录与接受者（候选者）记录匹配的行为，以便将提供者的数据转移至接受者（Pierzchala, 1995 年）。
随机编辑	一种如果违反则以概率小于 1 指明数据中错误的编辑（Pierzchala, 1995 年）。
结构编辑	基于两个或多个经编辑字段之间的逻辑关系的检查。例如，整体必须等于其各部分之和；或者由于调查表内固有的跳跃模式，位于不相交路径上的两个变量不能全为非零。结构编辑确保数据记录中保持调查表的结构（Pierzchala, 1995 年）。
结构推算	在数个变量之间结构关系成立时使用结构推算。例如，整体必须等于其各部分之和；因此，就某个母亲而言，所生的子女必须等于存活的子女加已死亡的子女（Pierzchala, 1995 年）。
验证编辑	在特定记录各字段之间所作的编辑检查。这包括检查每份记录的每个字段，以弄清它是否载有一个有效的登录项，并以某种预定的字段组合检查登录项，以确定各登录项是否彼此一致（Pierzchala, 1995 年）。
加权	在 Fellegi-Holt 编辑和推算学派内，根据可靠性将权数分派给字段。权数越大，字段就越有可能推算（所有其他条件相等）。也可将权数分配给编辑（Pierzchala, 1995 年）。
记录内的编辑	验证编辑的另一个名称（Pierzchala, 1995 年）。

参考文献

- Banister, J. (1980 年)。“普查编辑和推算的利用和滥用”。《亚洲及太平洋普查论坛》，第 6 卷，第 3 期，第 1–20 页。
- Bankier, M.、M. Lachance 和 P. Poirier (1999 年)。“新推算法的一般采用”。载于《调查研究方法科议事录》。弗吉尼亚州亚历山德里亚：美国统计学会（即将出版）。
- Bankier, M.、A. M. Houle 和 M. Luc (无日期)。“加拿大普查人口变量推算”。手稿。
- Bankier, M.、M. Luc, C. Nadeau 和 P. Newcombe (1996 年)。“同时推算数值和定性普查变量”。载于《调查研究方法科议事录》。弗吉尼亚州亚历山德里亚：美国统计学会，第 287–292 页。
- Boucher, L. (1991 年)。“制造厂商年薪微观编辑：所谓增值？”。载于《年度研究会议议事录》。华盛顿：美国统计局，第 765–781 页。
- Fellegi, I. P. 和 D. Holt (1976 年)。“自动编辑和推算的系统方法”。《美国统计学会杂志》，第 71 卷，第 353 (3 月) 期，第 17–35 页。
- Ford, Bany L. (1983 年)。“流水式过程概述”。载于《抽样调查不完整的数据》，第二卷，《理论与参考文献》。William G., Madin Ingram Olkin 和 Donald B. Rubin 编辑。
- Granquist, L. (1984 年)。“数据编辑及其对统计数据进一步处理的影响”。在统计计算讲习班上宣读的论文，布加勒斯特，1984 年 11 月 12–17 日。
- Granquist, L. (1987 年)。“瑞典统计局计算机支持编辑短期开发程序”。在数据编辑联合小组会议上所作的报告，马德里，1987 年 4 月 22–24 日。斯德哥尔摩：瑞典统计局。
- Granquist, L. (1997 年)。“编辑的新观点”。《国际统计评论》，第 65 卷，第 3 期，第 381–387 页。纽约：学术出版社。
- Granquist, L. 和 J. G. Kovar (1997 年)。“调查数据的编辑：多少才够？”载于《调查计量和过程质量》，L. Lyberg 等人合编。纽约：John Wiley and Sons 出版社，第 415–435 页。
- Greenberg, Brian 和 Rita Surdi (1984 年)。“比率编辑的灵活和交互式编辑与推算系统”。载于《美国统计学会调查研究方法科议事录》，弗吉尼亚州亚历山德里亚：美国统计学会，第 421–426 页。
- Kalton, G. 和 D. Kasprzyk (1982 年)。“遗漏调查答案的推算”。载于《调查研究方法科议事录》。美国统计学会，第 22–31 页。
- Kalton, G. 和 D. Kasprzyk (1986 年)。“遗漏调查数据的处理”。《调查方法》，第 12 卷，第 1–16 页。
- Naus, J. I. (1975 年)。《数据质量控制和编辑》。纽约：Marcel Dekker 出版社。
- Nordbotten, S. (1963 年)。“个人统计观察的自动编辑”。《欧洲统计学家会议统计标准和方法》，第 2 期。纽约：联合国。
- Pierzchala, M. (1995 年)。“编辑系统和软件”。载于《企业调查方法》，B. G. Cox 等人合编。纽约：John Wiley and Sons 出版社，第 411–425 页。

- Pullum, T. W.、T. Harpham 和 N. Ozsever (1986 年)。“大规模抽样调查的机器编辑：世界生育率调查的经验”。《国际统计评论》，第 54 卷，第 311–326 页。
- Rubin, D. B. (1987 年)。《调查中无答案的多项推算》。纽约：Wiley 出版社。
- Sande, I. G. (1982 年)。“调查中的推算：应对现实”。《美国统计学家》，第 36 卷，第 145–152 页。
- Sarndal, C. E.、B. Swensson 和 J. Wretman (1992 年)。《模型辅助调查抽样》。纽约：Springer-Verlag 出版社。
- 加拿大统计局 (1998 年)。《加拿大统计局质量指导方针》，第 3 版。渥太华：加拿大统计局。
- 联合国 (1992 年)。《人口和住房普查手册，第一部分：人口和住房普查的规划、组织与管理》。方法研究，F 辑，第 54 号，出售品编号：E.92.XVII.8。
- 联合国 (1992 年)。《人口和住房普查手册，第二部分：人口和社会特征》。方法研究，F 辑，第 54 号，出售品编号：E.91.XVII.9。
- 联合国 (1998 年)。《人口和住房普查的原则和建议，第一次修订本》，统计文件，M 辑，第 67/Rev.1 号。出售品编号：E.98.XVII.8。
- 联合国 (1999 年)。《统计使用的标准国家或地区代码》，统计文件，M 辑，第 49/Rev.4 号，出售品编号：M.98.XVII.9。
- 联合国统计委员会和欧洲经济委员会 (1994 年)。《统计数据编辑，第 1 卷：方法和技术》。欧洲统计学家会议，统计标准和研究系列，第 44 号，出售品编号：94.II.E.36。
- 联合国统计委员会和欧洲经济委员会 (1997 年)。《统计数据编辑，第 2 卷：方法和技术》。欧洲统计学家会议，统计标准和研究系列，第 48 号，出售品编号：96.II.E.30。

كيفية الحصول على منشورات الأمم المتحدة

يمكن الحصول على منشورات الأمم المتحدة من المكتبات ودور التوزيع في جميع أنحاء العالم . استعمل عنها من المكتبة التي تتعامل معها أو اكتب إلى : الأمم المتحدة ، قسم البيع في نيويورك أو في جنيف .

如何获取联合国出版物

联合国出版物在全世界各地的书店和经售处均有发售。请向书店询问或写信到纽约或日内瓦的联合国销售组。

HOW TO OBTAIN UNITED NATIONS PUBLICATIONS

United Nations publications may be obtained from bookstores and distributors throughout the world. Consult your bookstore or write to: United Nations, Sales Section, New York or Geneva.

COMMENT SE PROCURER LES PUBLICATIONS DES NATIONS UNIES

Les publications des Nations Unies sont en vente dans les librairies et les agences dépositaires du monde entier. Informez-vous auprès de votre libraire ou adressez-vous à : Nations Unies, Section des ventes, New York ou Genève.

КАК ПОЛУЧИТЬ ИЗДАНИЯ ОРГАНИЗАЦИИ ОБЪЕДИНЕННЫХ НАЦИЙ

Издания Организации Объединенных Наций можно купить в книжных магазинах и агентствах во всех районах мира. Наводите справки об изданиях в вашем книжном магазине или пишите по адресу: Организация Объединенных Наций, Секция по продаже изданий, Нью-Йорк или Женева.

COMO CONSEGUIR PUBLICACIONES DE LAS NACIONES UNIDAS

Las publicaciones de las Naciones Unidas están en venta en librerías y casas distribuidoras en todas partes del mundo. Consulte a su librero o diríjase a: Naciones Unidas, Sección de Ventas, Nueva York o Ginebra.
