

经济和社会事务部
统计司

方法研究

F辑 第96号

发展中国家和转型期国家 住户调查



联合国
2008年，纽约

联合国秘书处经济和社会事务部

是全球经济、社会和环境领域的政策与国家行动之间的一个重要纽带。该部在三个相互关联的主要领域开展工作：（一）汇编、产生和分析各种经济、社会和环境数据及信息，供联合国会员国在审查共同问题和评估各种政策选择时参考；（二）便于许多政府间机构的成员国就联合行动方针进行谈判以应对现有的或不断出现的全球挑战；（三）向有关政府就如何将联合国大会和首脑会议上拟定的政策框架转化为国家一级方案的方式方法提出建议，并通过技术援助协助各国进行能力建设。

说 明

联合国文件都用英文大写字母附加数字编号。提及这种编号系指联合国文件。

ST/ESA/STAT/SER.F/96

联合国出版物
出售品编号：C.05.XVII.6

ISBN 978-92-1-730042-4

联合国2005年版权所有©
翻印必究

序 言

住户调查是社会经济数据的一个重要来源。借以了解和监测发展政策的各项重要指标往往是从这些调查中得出的。在发展中国家，住户调查已成为收集数据的一种主要方式，它是其他数据收集方案和民事登记系统的补充，有时甚至会取代这些方案和系统。

本出版物介绍了发展中国家和转型期国家开展的住户调查在样本设计、调查实施、非抽样误差、调查成本和调查数据分析等若干重要方面的“最新水平”。本手册的主要目的是协助国家调查统计员以有效和可靠的方式设计住户调查，并使用户更充分地利用调查产生的数据。

本出版物共有25章，由全世界调查研究方法领域的知名专家撰写。他们中的大多数人都有协助发展中国家和转型期国家的国家统计局开展工作的实际经验。本出版物具有以下一些独特特点：

- 特别注重发展中国家和转型期国家的需要；
- 重点讨论可适用于不同国家和不同调查的标准和作业特点；
- 涉及调查成本（包括编制调查预算的经验实例）和对各细分部分调查成本的分析；
- 广泛涉及非抽样误差；
- 涉及住户调查数据分析的基本技术和高级技术，包括对用于分析复杂的调查数据的最新计算机软件包进行详细的经验比较；
- 介绍了一些发展中国家和转型期国家开展的住户调查的设计、实施和数据分析方面的实例；
- 介绍了一些关于发展中国家和转型期国家开展的大规模实际调查的案例研究，这些案例研究在设计类似的调查时可供借鉴。

本出版物是在联合国经济和社会事务部/统计司（经社部/统计司）先前为提高调查方法质量和增强国家统计系统的能力而发出的倡议基础上编制的。在过去二十年所提的倡议中，最为全面的倡议是国家住户调查能力方案（住户调查能力方案）。该方案的目的是协助发展中国家通过综合住户调查系统获取重要的人口和社会经济数据，从而支持发展规划、政策拟定和方案实施。该方案在很大程度上促进了许多发展中国家，特别是非洲发展中国家的统计发展，这些国家在1980年代完成了数量和种类都大为增多的调查并从中受益。此外，住户调查能力方案还为方法工作提供了支持，促成出版了若干技术性研究报告和手册。《住户调查手册》（修订本）¹概括介绍了设计和实施住户调查的相关

¹ 《方法研究》第31号，（联合国出版物，出售品编号：E.83.XVII.13）。

问题。继该手册之后，又发行了一系列讨论调查方法特定领域的问题和程序并涉及许多专题领域的出版物，其中包括：

- 《国家住户调查能力方案：综合住户调查方案的抽样基础和样本设计：初级版本》（DP/UN/INT-84-014/5E），纽约，1986年
- 《国家住户调查能力方案：住户调查中的抽样误差》（UNFPA/UN/INT-92-P80-15E），纽约，1993年
- 《国家住户调查能力方案：调查数据处理：对问题和程序的审查》（DP/UN/INT-81-041/1），纽约，1982年
- 《国家住户调查能力方案：住户调查中的非抽样误差：来源、评估和控制：初级版本》（DP/UN/INT-81-041/2），纽约，1982年
- 《国家住户调查能力方案：调查表的拟订和设计》（INT-84-014），纽约，1985年
- 《国家住户调查能力方案：住户收入和支出调查：技术性研究报告》（DP/UN/INT-88-X01/6E），纽约，1989年
- 《国家住户调查能力方案：住户健康调查指南》（INT/89/X06），纽约，1995年
- 《国家住户调查能力方案：稀有人口和易被遗忘人口的抽样调查》（INT-92-P80-16E），纽约，1993年

本出版物对上述出版物中详细论述的问题和程序的技术方面作了修订和补充，并特别着重说明它们如何适用于在发展中国家和转型期国家开展的调查。

联合国经济和社会事务部

统计司

司长

Paul Cheung

概 述

本出版物的编排方式如下。它由两大部分共25章组成。第一大部分21章，分成A至E五部分。以下是第一大部分每一部分的内容提要。

A部分： 调查设计和实施。本节包括三章内容。第二章概述发展中国家和转型期国家的住户调查设计方面的各种相关问题。第三章和第四章分别讨论发展中国家和转型期国家的调查问卷设计和调查实施方面的相关问题。

B部分： 样本设计。本节由介绍性说明和三章组成，讨论了样本设计的具体方面。第五章涉及主样本和主架构的设计。第六章讨论设计效果在样本设计和分析中的应用，第七章则对若干发展中国家开展的调查的设计效果进行了经验分析。

C部分： 非抽样误差。本节由介绍性说明和四章组成，论述发展中国家和转型期国家非抽样误差计量、评估和控制的各个方面。第八章讨论了非观察误差（不答复和未覆盖）。第九章讨论了计量误差。第十章介绍了适用于世界卫生组织的质量保证准则和程序，世界卫生组织调查是一项由世界卫生组织（卫生组织）赞助、在发展中国家开展调查的方案。第十一章描述了关于巴西开展的住户调查非抽样误差的计量、评估和校正的案例研究。

D部分： 调查成本。本节由介绍性说明和三章组成。第十二章介绍了分析发展中国家和转型期国家开展调查的调查成本的总体框架。第十三章利用经验数据，介绍了某个发展中国家开展收入和支出调查的成本模式。第十四章讨论了住户调查各个阶段和职能的预算编制的相关问题，包括用于比较和说明本章所讨论的重要预算编制问题的若干实例和案例研究。

E部分： 调查数据分析。本节由介绍性说明和七章内容组成，专门讨论了调查数据的分析。第十五章介绍了住户调查数据管理的详细准则。第十六章讨论了调查数据的基本表格分析，包括若干具体实例。第十七章讨论了利用多主题住户调查作为发展中国家减贫的一种手段。第十八章讨论的是如何采用多变量统计方法来创建住户调查数据索引。第十九章涉及调查数据的统计分析，着重于以模型为基础的基本分析技术，即多元线性回归、逻辑回归和多级方法。第二十章介绍了分析调查数据的高级方法，这些方法考虑了设计复杂性对分析的影响。最后，第二十一章讨论了评估调查数据的抽样误差所使用的各种方法，并介绍了一些实用的数据分析技术，对分析复杂的调查数据使用的若干计算机软件包做了比较。此外还强调了样本设计与数据分析之间的紧密联系。关于软件包的比较，包括各软件包的计算机输出的进一步详情，见本出版物所附光盘。

本出版物的第二部分由介绍性说明和四章内容组成，主要是案例研究，介绍了一些发展中国家和转型期国家开展调查的具体实例。这些章节详细而系统地介绍了由国际机构发起、由用户出资的调查和由国家编制预算开展的调查——作为国家统计制度经常调查方案的一部分。第二十二章介绍了人口与健康调查方案；第二十三章介绍了生活平衡量研究调查方案。对这两种调查的讨论包括如何计算若干主要特征估计数的设计效果。第二十四章以在老挝人民民主共和国开展的一项调查为例，讨论了住户预算调查的设计和实施。第二十五章讨论了转型期国家开展的调查在设计和实施方面的一般特点，并包含若干案例研究。

鸣 谢

编写一本内容如此广泛的出版物需要多方的合作。经社部/统计司得益于世界各地联合国共同系统内外许多顾问和组织的大力帮助。这些顾问都是具有丰富的设计、实施和分析复杂调查方面专门知识的专家，其中许多人还拥有在发展中国家和转型期国家开展调查的广泛经验。

本出版物的每一章都经历了非常严格的同行审查过程。首先，每一章都由两名审阅人进行审稿，他们是相关领域的知名专家。修订后的各章汇集在一起形成初稿。经社部/统计司于2002年10月在纽约召开专家组会议，对初稿进行了极其严格的审查。会后成立了编辑委员会，对出版物进行审查并就其结构和内容提出最后建议。经过审查过程的这一阶段，最终对整个出版物作了结构调整和简化，使其更加有条理、更完整、内容也更加连贯。根据专家组会议和编辑委员会的建议编写了新的章节，并对旧章节作了修订。在两名审阅人对修订的各章进行第三轮审查之后，才最后决定是否将其编入出版物。编辑人员小组然后对整个出版物进行最后的审查，确保所编材料在技术上具有可靠性、内容连贯，并且符合本出版物的主要目标。

经社部/统计司对Graham Kalton先生对本出版物做出的宝贵贡献深表感谢。Kalton先生既是专家组会议主席，也是编辑委员会主席，他审查了许多章节，在整个项目期间对经社部/统计司的工作人员提出了技术建议和知识指导。John Eltinge先生作为若干章节的审稿人和为制定本出版物框架而进行的某些背景研究工作的指导者兼合作人，在本出版物的最初构想阶段提供了大量指导，并继续在项目的各个方面发挥重要作用。James Lepkowski、Oladejo Ajayi、Hans Pettersson、Karol Krotki和Anthony Turner诸位先生对若干章节的编辑提供了重要帮助，并在项目的各个阶段提供了一般性指导和支持。

其他许多专家作为各章撰稿人、其他专家所撰章节的审稿人，或者同时作为撰稿人和审稿人，对本项目做出了贡献。另外一些专家通过出席专家组会议和对出版物初稿的各个方面进行建设性的审查，对本项目做出了贡献。参与本项目的专家的姓名及其所在单位见目录后的清单。

若没有经社部/统计司的一些工作人员在各个阶段所作的大量努力，本项目各项重大目标即使不是不可能实现，也是很难实现。Ibrahim Yansaneh先生提出了编写本出版物的提案，召集了其他的参与者，并对本项目的所有技术方面，包括编辑过程进行了协调。他还撰写了若干章节，并且是整个出版物的总编。经社部/统计司司长和副司长对项目的各个阶段予以鼓励，并提供了机构性支助。Stefan Schweinfest先生对项目的所有行政方面进行了管理，Sabine Warschburger女士设计并维护了项目网站，Denise Quiroga女士通过便利作者

和编辑人员之间许多文件的流动、对这些文件的不同格式和写作风格进行组织和协调并协助实施项目管理时间表，提供了一流的秘书协助。

目 录

	页 次
序言.....	iii
概述.....	v
鸣谢.....	vii
参与专家名单.....	xxvi
作者.....	xxviii
审稿人员.....	xxx
第一部分 调查设计、实施与分析	
第一章 导言.....	3
A. 发展中国家和转型期国家住户调查.....	3
B. 本出版物的目标.....	4
C. 上述目标的现实意义.....	5
A部分 调查设计与实施	
第二章 发展中国家和转型期国家住户调查样本设计问题概述.....	9
A. 导言.....	9
1. 发展中国家和转型期国家调查样本设计.....	9
2. 概述.....	10
B. 分层多级抽样.....	10
1. 明晰分层法.....	10
2. 隐含分层法.....	11
3. 初级抽样单位的样本选择.....	11
4. 概率与规模成正比的初级抽样单位采样.....	13
5. 住户样本选择.....	15
6. 每个初级抽样单位选定的住户数量.....	16
C. 抽样标架.....	18
1. 发展中国家和转型期国家调查中的抽样标架特征.....	18
2. 有关抽样标架的问题及可能的解决办法.....	18
3. 抽样标架的保持和评价.....	20
D. “域”的估算.....	20
1. “域”要素的必要性.....	20
2. 样本分配.....	21

E.	样本规模	21
1.	影响样本规模决策的诸多因素	21
2.	调查估计数的精确度	22
3.	数据质量	24
4.	成本和及时性	25
F.	调查分析	25
1.	抽样权数的拟定和调整	25
2.	住户调查数据分析	26
G.	结论	27
	附件 调查流程图	29
	第三章 发展中国家住户调查问卷设计概述	31
A.	导言	31
B.	大致结构	32
1.	调查目标	32
2.	制约因素	33
3.	一些实际忠告	34
C.	问卷设计的细节问题	34
1.	模块方法	35
2.	问卷格式化和前后一致性	36
3.	关于问卷设计细节的其他忠告	40
D.	问卷准备过程	40
1.	建立团队	40
2.	草拟问卷第一稿	41
3.	问卷的实地检验和定稿	42
E.	结论	43
	第四章 发展中国家住户调查的实施概述	45
A.	导言	45
B.	进入现场调查前的活动	46
1.	预算筹资	46
2.	工作计划	48
3.	提取住户样本	50
4.	编写培训手册	50
5.	外勤和数据录入工作人员的培训	51
6.	实地工作和数据录入计划	51
7.	进行试点检测	52
8.	发起宣传运动	52
C.	现场调查过程中开展的各项活动	52
1.	通讯和交通手段	53

2. 监督和质量保障	53
3. 数据管理	54
D. 实地工作、数据录入和数据处理之后需要开展的活动	54
1. 听取汇报	54
2. 最终数据集的制备和文件编制工作	55
3. 数据分析	55
E. 结论	55
B部分 样本设计	
引言	59
第五章 发展中国家住户调查主抽样标架和主样本的设计	63
A. 引言	63
B. 主抽样标架和主样本：概述	64
1. 主抽样标架	64
2. 主样本	65
3. 归纳和结论	67
C. 主抽样标架的设计	68
1. 数据和材料：质量评估	69
2. 关于主抽样标架覆盖面的决策	70
3. 关于基本标架单位的决策	70
4. 拟纳入抽样标架的标架单位信息	71
5. 主抽样标架的文件编制和保管	73
D. 主样本的设计	74
1. 主样本初级抽样单位的选择	74
2. 旨在减少初级抽样单位规模差异的地区合并与拆分	76
3. 初级抽样单位的分层和各层次的主样本分配	77
4. 初级抽样单位的抽样	78
5. 主样本的耐久性	79
6. 文件编制	80
7. 使用主样本调查公司企业	80
E. 结论	81
第六章 供样本设计使用的设计效果要素估算	83
A. 引言	83
B. 设计效果诸要素	86
1. 分层	86
2. 类集	90
3. 加权调整	94

C. 设计效果模型	96
D. 在样本设计中应用设计效果	100
E. 结论	103
第七章 发展中国家的调查设计效果分析	107
A. 导言	107
B. 调查	108
C. 设计效果	109
D. 同质性比率的计算	116
E. 讨论	118
附件 11项住户调查样本设计的介绍	120
C部分 非抽样误差	
导言	127
第八章 发展中国家住户调查中的非观察误差	129
A. 导言	129
B. 理解未覆盖和不答复误差的基本框架	130
C. 未覆盖误差	132
1. 未覆盖起因	132
2. 未覆盖误差	135
D. 不答复误差	138
1. 住户调查中不答复的起因	138
2. 不答复偏差	139
3. 计算不答复偏差	141
4. 减少和弥补住户调查中的单位不答复情况	142
5. 项目不答复与插补	144
第九章 住户调查测量误差：来源和测量	147
A. 导言	147
B. 测量误差的来源	149
1. 调查问卷的影响	149
2. 数据收集模式的影响	152
3. 访调员的影响	154
4. 受访者的影响	155
C. 量化测量误差的方法	157
1. 随机化实验	157
2. 认知研究方法	158
3. 重访研究	159

4. 记录检验研究.....	161
5. 访调员方差研究.....	163
6. 行为编码.....	164
D. 结束语：测量误差.....	164
第十章 调查中的质量保证：标准、准则和程序.....	171
A. 导言.....	172
B. 质量标准和保证程序.....	172
C. 质量保证准则的切实实施：世界卫生调查的实例.....	174
1. 选择调查机构.....	175
2. 抽样.....	176
3. 翻译.....	178
D. 培训.....	181
E. 调查的实施.....	183
F. 数据输入.....	187
G. 数据分析.....	189
H. 质量指标.....	191
1. 样本离差指数.....	191
2. 答复率.....	192
3. 漏失数据率.....	192
4. 检验 - 再检验访谈的可靠性系数.....	192
I. 国别报告.....	193
J. 实地访问.....	194
K. 结论.....	195
第十一章 巴西调查非抽样误差的报告和补救：目前的做法和未来的挑战.....	199
A. 导言.....	199
B. 巴西住户调查非抽样误差报告和抵消的现行做法.....	202
1. 覆盖误差.....	202
2. 不答复.....	205
3. 测量和处理误差.....	208
C. 挑战和展望.....	209
D. 有关进一步阅读的建议.....	211
D部分 调查成本	
导言.....	215
第十二章 发展中国家和转型国家调查成本问题的分析.....	217

A. 导言	217
1. 有效抽样设计的标准	217
2. 发展中国家和转型国家调查的成本结构组成部分	218
3. 本章概述	219
B. 调查成本的组成部分	219
C. 拥有广泛基础设施可加利用的调查的成本	220
1. 同筹备活动相关的因素	220
2. 同数据收集和处理相关的因素	220
D. 事先调查基础设施有限或没有此种设施的调查的成本	221
E. 同调查目标修改相关的因素	222
F. 关于调查成本报告的某些告诫	222
G. 摘要和结束语	223
附件 联合国儿童基金会（儿童基金会）多指标类集调查（MICS） 预算框架	226
第十三章 收支调查的成本模型	229
A. 导言	229
B. 成本模型和成本概算	230
C. 有效样本设计的成本模型	230
D. 案例研究：2002年老挝支出和消费调查	233
E. 老挝2002年支出和消费调查（LECS-3调查）实地工作成本 模型	234
F. 结束语	236
第十四章 为发展中国家住户调查制定预算编制框架	239
A. 导言	239
B. 初步考虑	240
1. 调查阶段	240
2. 调查的时间表	241
3. 调查类型	242
4. 预算与支出	243
5. 以前的调查	243
C. 预算框架内的主要会计科目分类	244
1. 人员	244
2. 交通	246
3. 设备	246
4. 消耗品	246
5. 其他成本	246
6. 按会计科目分类编制预算的例子	247
D. 预算框架内的主要调查活动	248

1. 为调查准备工作编制预算.....	249
2. 为实施调查编制预算.....	249
3. 为调查数据处理编制预算.....	249
4. 为调查报告编制预算.....	249
5. 为调查活动编制预算的例子.....	250
E. 汇总.....	251
F. 潜在预算限制和缺陷.....	252
G. 保存记录和摘要.....	253
H. 结论.....	254
附件 保留日记录和周记录的样表.....	255
E部分 调查数据分析	
导言.....	261
第十五章 住户调查数据管理指南.....	263
A. 导言.....	263
B. 数据管理和调查问卷设计.....	264
C. 数据录入和编辑的操作策略.....	265
D. 质量控制标准.....	268
E. 数据录入程序的开发.....	271
F. 调查数据集的整理和散发.....	273
G. 抽样过程中的数据管理.....	276
H. 建议摘要.....	285
第十六章 根据住户调查数据编制简单描述性统计资料.....	289
A. 导言.....	289
B. 变量与描述性统计资料.....	290
1. 变量类型.....	290
2. 简单描述性统计资料.....	291
3. 编制一个变量的描述性统计资料.....	293
4. 编制两个变量的描述性统计资料.....	295
5. 编制三个或更多变量的描述性统计资料.....	298
C. 编制描述性统计资料的一般建议.....	298
1. 数据准备.....	298
2. 编制结果.....	300
3. 构成一张好表格的元素.....	300
4. 权数的使用.....	303
D. 为住户调查编制综合报告（摘要）.....	303
1. 内容.....	303

2. 程序.....	304
E. 结论意见.....	304
第十七章 利用多主题住户调查改善发展中国家减贫政策.....	307
A. 导言.....	307
B. 描述性分析.....	308
1. 贫穷的定义.....	308
2. 描述贫穷概况.....	309
3. 利用贫穷概况进行基本政策分析.....	310
C. 住户调查数据的多元回归分析.....	312
1. 需求分析.....	312
2. 社会服务的使用.....	313
3. 政府具体方案所产生的影响.....	314
D. 小结.....	315
第十八章 构建指数的多变量法.....	317
A. 导言.....	317
B. 关于使用多变量法的一些限制.....	318
C. 多变量法概述.....	319
D. 图和汇总法.....	320
E. 类集分析.....	322
F. 主成分分析 (PCA).....	326
G. 多变量指数构建法.....	328
1. 建立消费支出模型, 构建收入代用指数.....	329
2. 用于构建“财富”指数的主成分分析 (PCA).....	330
H. 结论.....	332
第十九章 调查数据的统计分析.....	335
A. 导言.....	335
B. 描述统计学: 权数与方差估计.....	336
C. 分析统计.....	341
D. 关于回归建模的一般性评论.....	342
E. 线性回归模型.....	344
F. 逻辑回归模型.....	348
G. 多层次模型的应用.....	350
H. 旨在支持调查过程的建模.....	354
I. 结论.....	354
第二十章 较先进的调查数据分析方法.....	359
A. 导言.....	359

1. 样本设计与数据分析	359
2. 样本设计对分析有影响（和无影响）的实例	360
3. 基本概念	361
4. 设计效果及其在复杂样本数据分析中的作用	363
B. 复杂样本数据分析的基本方法	363
1. 模型设定是分析的基础	363
2. 模型与样本设计之间可能存在的关系：告知性设计 和非告知性设计	364
3. 使用适合复杂样本分析的标准分析软件包所涉及的问题	365
C. 回归分析和线性模型	366
1. 未纳入模型的设计变量效应和加权回归估计数	366
2. 检验设计对回归分析的影响	367
3. 采用告知性样本设计的多层次模型	369
D. 分类数据分析	370
1. 修改卡方检验法以检验拟合优度和独立性	370
2. 对数线性模型的归纳	372
E. 总结与结论	374
附件 正式定义和技术成果	375
第二十一章 调查数据的抽样误差估计	381
A. 调查样本设计	382
B. 复杂样本调查数据的数据分析问题	382
1. 加权分析	382
2. 方差估计概述	382
3. 没有替代抽样的有限总体修正（FPC）系数	383
4. 假层和假初级抽样单位	384
5. 描述许多复杂抽样计划的一种常见近似（WR）	384
6. 方差估计技术和调查设计变量	385
7. 复杂样本调查数据的分析	386
C. 方差估计方法	386
1. 采用泰勒级数线性化方法估计方差	386
2. 方差估计的重复方法	387
3. 平衡多次重复（BRR）	388
4. 刀切重复方法（JK）	389
5. 方差估计软件用户常犯的一些错误	390
D. 方差估计的软件包比较	390
E. 布隆迪样本调查数据集	393
1. 推断总体和总体参数	393
2. 抽样计划和数据收集	393
3. 方差估计的加权程序和设置	394

4. 调查数据分析的三个例子.....	394
F. 使用非样本调查程序分析样本调查数据.....	395
G. SAS 8.2中的样本调查程序.....	396
1. SURVEYMEANS和SURVEYREG 概述.....	396
2. SURVEYMEANS.....	397
3. SURVEYREG.....	397
4. 数值例子.....	398
5. 优点/缺点/费用.....	399
H. SUDAAN 8.0.....	399
1. SUDAAN概述.....	399
2. DESCRIPT.....	401
3. CROSSTAB.....	401
4. 数值例子.....	402
5. 优点/缺点/费用.....	403
I. STATA 7.0中抽样调查程序.....	403
1. STATA概述.....	403
2. SVYMEAN、SVYPROP、SVYTOTAL、SVYLC.....	404
3. SVYTAB.....	405
4. 数值例子.....	405
5. 优点/缺点/费用.....	406
J. Epi-Info 6.04d和Epi-Info 2002样本调查程序.....	406
1. Epi-Info概述.....	406
2. Epi-Info V6.04d (DOS)、CSAMPLE模块.....	407
3. Epi-Info 2002 (Windows).....	408
4. 数值例子.....	408
5. 优点/缺点/费用.....	409
K. WesVar 4.2.....	409
1. WesVar概述.....	409
2. 使用WesVar V4.2版.....	410
3. 数值例子.....	411
4. 优点/缺点/费用.....	412
L. PC-CARP.....	412
M. CENVAR.....	413
N. IVEware (Beta版).....	413
O. 结论和建议.....	414

第二部分 个案研究

导言.....	421
第二十二章 人口与健康调查.....	425

A. 导言	425
B. 历史	426
C. 内容	426
D. 抽样标架	427
E. 抽样阶段	428
F. 不答复情况的报告	430
G. 不答复率比较	431
H. 人口与健康调查的样本设计效果	432
I. 调查实施	435
J. 编制与翻译调查文件	435
K. 事先检验	436
L. 实地工作人员的征聘	437
M. 访调员培训	438
N. 实地工作	438
O. 数据处理	440
P. 分析及报告的编写	441
Q. 传播	441
R. 人口与健康调查数据的使用	441
S. 能力建设	442
T. 吸取的教训	442
附件：1990-2000年期间44个国家66项调查的住户答复率和妇女答复率 （部分地区）	446
第二十三章 生活水平衡量研究调查	449
A. 导言	449
B. 为什么进行生活水平衡量研究调查?	450
C. 生活水平衡量研究调查的主要特征	450
1. 内容及所采用的工具	451
2. 样本问题	452
3. 实地工作组织	454
4. 质量	454
5. 数据录入	457
6. 可持续性	457
D. 进行生活水平衡量研究调查的成本	458
E. 生活水平衡量研究设计对质量的效用如何?	460
1. 答复率	460
2. 项目不答复	460
3. 内部一致性检查	461
4. 样本设计效果	463
F. 生活水平衡量研究调查数据的使用	464

G. 结论	465
附件一 生活水平衡量研究调查清单.....	467
附件二 生活水平衡量研究调查的预算编制.....	469
附件三 生活水平衡量研究调查中样本设计对精确度和有效性的影响.....	470
第二十四章 住户预算调查的调查设计和样本设计.....	477
A. 导言	477
B. 调查设计	478
1. 住户预算调查中收集数据的方法.....	478
2. 计量问题.....	478
3. 标准期限.....	479
4. 访谈次数.....	480
5. 不答复.....	480
C. 样本设计	481
1. 分层、层的样本分配.....	481
2. 样本规模.....	481
3. 在不同时间进行抽样.....	482
D. 个案研究：1997/1998年老挝支出与消费调查.....	483
1. 调查工作的一般环境.....	483
2. 调查以及调查问卷的主题.....	483
3. 计量方法.....	483
4. 样本设计、实地工作.....	484
E. 吸取的经验教训	485
1. 计量方法、不答复.....	485
2. 样本设计、抽样误差.....	485
3. 使用时间利用日记的经验.....	486
4. 采用第二次老挝支出与消费调查估计国内生产总值....	486
F. 结论	487
第二十五章 转型国家住户调查.....	489
A. 对转型国家住户调查的综合评估	490
1. 导言.....	490
2. 转型前（1991-2000年）中东欧国家和苏联住户抽样 调查.....	490
3. 转型期间住户调查.....	492
4. 住户预算调查.....	492
5. 劳动力调查.....	493
6. 住户预算调查及劳动力调查中抽样设计及实施的共同 特征.....	494

7. 结论.....	502
B. 转型国家的住户抽样调查：个案研究.....	503
1. 爱沙尼亚住户抽样调查.....	503
2. 匈牙利住户预算调查和劳动力调查的设计与实施.....	506
3. 拉脱维亚住户调查的设计与实施.....	509
4. 立陶宛住户抽样调查.....	513
5. 波兰在转型期间的住户调查.....	515
6. 斯洛文尼亚的劳动力调查和住户预算调查.....	521

表

	页 次
表二.1. 选定的整群抽样规模与组内相关组合的设计效果	17
表二.2. 成本比率与组内相关选定组合的最佳子样本规模	18
表二.3. 根据各种样本规模在设计效果值为2.0的情况下 贫困率估计数的标准误差和置信区间.....	23
表二.4. 根据不同的样本规模在设计效果值为2.0的情况下 贫困率估计数的变异系数.....	24
表四.1. 3 000住户虚拟调查预算草案（货币单位：美元）	47
表六.1. 在两层示例中非比例抽样所产生的设计效果....	89
表六.2. 全国八个省份（用A - H表示）的人口分布和各省 样本分配的三种备选方案.....	101
表七.1. 11项住户调查研究的特点.....	109
表七.2. 非洲和东南亚七项调查的设计效果估计.....	110
表七.3. 按照全国和地区类型划分的部分住户估计数的估 计设计效果（1999年PNAD调查）.....	111
表七.4. 在全国范围内选择的个人特征及各分区的估计设 计效果（1999年PNAD调查）.....	112
表七.5. 1999年9月PME调查中若干估计的估计设计效果	113
表七.6. PPV调查中部分估计数的估计设计效果.....	114
表七.7. 各项调查设计效果的比较.....	115
表七.8. 总体设计效果被分为加权效果 ($d_w^2(\bar{y})$) 和类集 效果 ($d_c^2(\bar{y})$)	116
表七.9. 城市和农村地区的同质性比率.....	118
表十.1. 抽样质量一览表.....	179
表十.2. 翻译程序审查一览表.....	181
表十.3. 培训程序综合概述清单.....	184
表十.4. 调查实施情况审查一览表.....	186
表十.5. 数据输入过程一览表.....	190
表十一.1. 巴西主要住户抽样调查的部分特点.....	202

表十一.2.	从1991和2000年查点后调查获得的巴西普查遗漏率的估计 (%)	205
表十三.1.	村子实地工作估计时间.....	234
表十三.2.	LECS-3调查的估计成本 (美元每日津贴)	235
表十三.3.	村中最佳样本规模(mopt)和不同 ρ 值的实际设计(m = 15)的相对效率.....	236
表十四.1.	拟议的非正规经济部门调查时间表初稿.....	241
表十四.2.	会计科目分类与调查活动的汇总表.....	244
表十四.3.	已规划的工作人员时间 (天数) 与调查活动的汇总表.....	245
表十四.4.	按会计科目分类的成本占总预算的比例: 十年目标调查 (1999-2000年), 部分非洲国家.....	247
表十四.5.	分配给各会计科目分类的预算比例: 微型企业服务影响评估 (AIMS), 津巴布韦 (1999年)	248
表十四.6.	调查活动的成本占总体预算的比例: 十年目标调查 (1999-2000年), 部分非洲国家.....	250
表十四.7.	调查活动的成本占总体预算的比例: 津巴布韦微型企业服务影响评估调查 (1999年)	251
表十四.8.	按调查活动分列的会计科目分类成本计划占预算的比例: 津巴布韦微型企业服务影响评估调查 (1999年)	252
表十四.9.	按调查活动分列的会计科目分类成本在实施调查期间占预算的比例: 津巴布韦微型企业服务影响评估调查 (1999年)	252
表十五.1.	住户调查的数据被存储为简单的矩形文件.....	273
表十六.1.	2002年4月北马利亚纳群岛联邦赛班岛按年龄和性别划分的人口分布情况: 行百分比.....	291
表十六.2.	2002年4月北马利亚纳群岛联邦赛班岛按年龄和性别划分的人口分布情况: 列百分比.....	292
表十六.3.	1994年美属萨摩亚根据族群划分的住户收入主要统计指标.....	292
表十六.4.	1992-1993年越南住户照明的来源.....	293
表十六.5.	1992-1993年住户总支出简况: 越南.....	296
表十六.6.	1992-1993年越南城乡地区过去四周访问过某个保健设施的 (所有年龄组) 人口使用各种保健设施的情况.....	296
表十六.7.	1992-1993年越南按地区划分的住户总支出.....	298
表十八.1.	一些多变量技术及其目的.....	320
表十八.2.	表明有无各种农场特征的农场数据.....	324
表十八.3.	八个农场之间的相似性矩阵.....	325
表十八.4.	主成分分析的结果.....	326

表十八.5.	坦桑尼亚联合共和国乞力马扎罗地区消费支出预测指数构建中所使用的变量及其相应的权数....	330
表十八.6.	将人口分成五种财富五分位数的划分点.....	331
表十九.1.	住户调查设计的典型结构.....	336
表十九.2.	模型1的线性回归参数估计数的解释——设应变量为住户的工薪收入.....	345
表十九.3.	可估计的住户工薪收入（模型1）.....	345
表十九.4.	根据模型2得出的线性回归参数估计数的解释——设应变量为住户的工薪收入.....	346
表十九.5.	根据模型4得出的逻辑回归参数估计数的解释——设应变量为贫困线以下的住户指标.....	349
表二十.1.	采用各种调查设计，规模为3 850个农场的总体的普通最小二乘方估计量的偏差和平均数二乘方以及无偏估计量方差.....	367
表二十.2.	ANOVA加权及未加权回归比较表.....	368
表二十.3.	三种迭代卡方检验与简单随机抽样检验的比率..	370
表二十.4.	根据 X^2 和 X^2_c 对选自1971年大不列颠及北爱尔兰联合王国总住户调查的项目进行检验的估计渐近规模（名义规模为.05）.....	371
表二十.5.	根据 X^2 , $X^2/\hat{\sigma}^2$ 和 $X^2/\hat{\lambda}^2$ 对选自1971年大不列颠及北爱尔兰联合王国总住户调查的变量交叉分组进行检验的估计渐近规模（名义规模为.05）.	372
表二十.6.	X^2 和修正统计量 $X^2/\hat{\sigma}^2$, $X^2/\hat{\lambda}^2$, $X^2/\hat{\lambda}^2$: 2 x 5 x 4表的估计渐近显著性水平（SL）及名义显著性水平 $\alpha = 0.05$	373
表二十一.1.	5个软件包中PROCS的比较：血清反应为阳性的妇女估计百分比和人数，标准估计误差，刚刚生育过的妇女，布隆迪，1988-1989年.....	391
表二十一.2.	具有复杂样本调查数据方差估计能力的8种软件包的属性.....	392
表二十二.1.	48项人口与健康调查的平均 $d(\bar{y})$ 值和 $\hat{\sigma}$ 值，1984-1993年.....	434
表二十三.1.	越南住户调查问卷的内容，1997-1998年.....	452
表二十三.2.	新增单元示例.....	453
表二十三.3.	生活水平衡量研究调查中的质量控制.....	456
表二十三.4.	最近生活水平衡量研究调查的答复率.....	460
表二十三.5.	生活水平衡量研究和劳动力调查中收入数据缺失的频率.....	461
表二十三.6.	有完整消费总量指标的住户：最新的生活水平衡量研究调查提供的例子.....	462

表二十三.7.	数据的内部一致性：单元之间的成功联系.....	462
表二十三.8.	生活平衡量研究调查设计效果示例.....	463
表A三.1.	按变量划分的设计效果的差异，加纳，1987年..	471
表A三.2.	不同时间设计效果的差异，加纳，1987年和 1988年.....	472
表A三.3.	不同国家设计效果的差异.....	473
表A三.4.	分析变量的说明：个人一级.....	474
表A三.5.	分析变量的说明：住户一级.....	474
表二十四.1.	住户消费和耐用品拥有情况的设计效果.....	486
表二十四.2.	时间利用日记的样本中实际人数和预计人数之间的 比率.....	487
表二十五.1.	某些转型国家新的住户预算调查和劳动力调查， 1992-2000年：开始年份、周期、重新设计的最后 年份.....	493
表二十五.2.	2000年若干个转型国家中住户预算调查和劳动力 调查的样本规模、样本设计和估计方法.....	498
表二十五.3.	1992-2000年间某些转型国家住户预算调查的不答 复率.....	500
表二十五.4.	1992-2000年某些转型国家中劳动力调查的不答复 率.....	500
表二十五.5.	匈牙利2000年住户预算调查的成本结构.....	502
表二十五.6.	匈牙利2000年劳动力调查的成本结构.....	502

图

	页 次
图三.1. 问卷格式图解.....	37
图四.1. 制定和实施住户调查的工作计划.....	49
图十.1. 世界卫生组织调查质量保证程序.....	173
图十.2. 数据输入和质量监测过程.....	188
图十.3. 样本离差指数.....	192
图十五.1. 尼泊尔生活水平调查 II.....	275
图十五.2. 将电子表格作为初级抽样标架.....	277
图十五.3. 进行隐含分层.....	278
图十五.4. 抽选概率与规模成比例样本（第一步）.....	279
图十五.5. 抽选概率与规模成比例样本（第二步）.....	280
图十五.6. 抽选概率与规模成比例样本（第三步）.....	281
图十五.7. 抽选概率与规模成比例样本（第四步）.....	282
图十五.8. 被抽中的初级抽样单位的电子数据表.....	283
图十五.9. 计算第一阶段抽选概率.....	283
图十五.10. 编制住户清单的结果记录.....	284
图十五.11. 记录不答复.....	284
图十五.12. 计算第二阶段概率和抽样权数.....	285
图十六.1. 1992-1993年越南住户的照明来源（柱状图）...	294
图十六.2. 1992-1993年越南住户的照明来源（饼图） （百分比）.....	294
图十六.3. 2002年4月赛班人口的年龄分布情况（直方图）.	295
图十六.4. 1992-1993年越南城乡地区过去四周访问过某个 保健设施的（所有年龄组）人口使用各种保健 设施的情况（百分比）.....	297
图十八.1. 六种变量之间矩阵图示例.....	321
图十八.2. 由农场间相似性矩阵形成的树枝形结构关系图..	325
图十九.1. 权数的应用与统计估计数.....	337
图二十.1. 无选择.....	361
图二十.2. 依据X值的选择： $XL < X < XU$	361
图二十.3. 依据X值的选择： $X < XL$ ； $X > XU$	361
图二十.4. 依据Y值的选择： $YL < Y < YU$	361
图二十.5. 依据Y值的选择： $Y < YL$ ； $Y > YU$	361
图二十.6. 依据Y值的选择： $Y > YU$	361
图二十三.1. 生活水平衡量研究目的与调查工具之间的关系..	451
图二十三.2. 各小组的一个月活动计划.....	455
图二十三.3. 生活水平衡量研究调查的成本构成（各部分在 总成本中所占的份额）.....	459

参与专家名单

发展中国家和转型期国家住户调查作业特点专家组会议与会人员
(2002年10月8日至10日, 纽约)

Savitri Abeyasekera 雷丁大学 大不列颠及北爱尔兰联合王国雷丁市	Paul Glewwe 明尼苏达大学 美利坚合众国 明尼苏达州圣保罗市
Oladejo O. Ajayi 统计顾问 尼日利亚拉各斯伊科伊	Ivo Havinga 经社部/统计司 纽约州纽约市
Jeremiah Banda 经社部/统计司 纽约州纽约市	Rosaline Hirschowitz 南非统计局 南非比勒陀利亚
Grace Bediako 经社部/统计司 纽约州纽约市	Gareth Jones 联合国儿童基金会 纽约州纽约市
Donna Brogan 艾默里大学 美利坚合众国 佐治亚州亚特兰大市	Graham Kalton Westat公司 美利坚合众国 马里兰州罗克维尔市
Mary Chamie 经社部/统计司 纽约州纽约市	Hiroshi Kawamura 联合国 经社部/发展政策分析司 纽约州纽约市
James R. Chromy 研究三角研究所 研究三角园区 美利坚合众国北卡罗来纳州	Erica Keogh 津巴布韦大学 津巴布韦哈拉雷
Willem de Vries 经社部/统计司 纽约州纽约市	Jan Kordos 华沙经济学院 波兰华沙

James Lepkowski
 社会研究所
 美利坚合众国
 密歇根州安阿伯市

Anatoly Smyshlyaev
 联合国
 经社部/发展政策分析司
 纽约州纽约市

Gad Nathan
 希伯来大学
 以色列耶路撒冷

Pedro Silva
 巴西地理和统计数据局
 巴西里约热内卢

Frederico Neto
 联合国
 经社部/发展政策分析司
 纽约州纽约市

Diane Steele
 世界银行
 美利坚合众国
 哥伦比亚特区华盛顿

Colm O'Muircheartaigh
 芝加哥大学
 美利坚合众国
 伊利诺伊州芝加哥市

Sirageldin Suliman
 经社部/统计司
 纽约州纽约市

Hans Pettersson
 瑞典国家统计局
 瑞典斯德哥尔摩

T. Bedirhan Üstün
 世界卫生组织
 瑞士日内瓦

Hussein Sayed
 开罗大学
 埃及吉萨省奥尔曼市

Shyam Upadhyaya
 综合统计事务处
 尼泊尔加德满都

Michelle Schoch
 联合国人口基金
 纽约州纽约市

Martin Vaessen
 人口与健康调查方案
 ORC Macro*
 美利坚合众国
 马里兰州卡尔弗顿

* 一家民意调查公司。

Stefan Schweinfest
 经社部/统计司
 纽约州纽约市

Ibrahim Yansaneh
 国际公务员制度委员会
 [经社部/统计司]
 纽约州纽约市

作者

- | | |
|--|--|
| Savitri Abeyasekera
雷丁大学
大不列颠及北爱尔兰联合王国雷丁市 | Daniel Kasprzyk
数学政策研究公司
美利坚合众国
哥伦比亚特区华盛顿 |
| J. Michael Brick
Westat公司
美利坚合众国
马里兰州罗克维尔市 | Erica Keogh
津巴布韦大学
津巴布韦哈拉雷 |
| Donna Brogan
艾默里大学
美利坚合众国
佐治亚州亚特兰大市 | Jan Kordos
华沙经济学院
波兰华沙 |
| Somnath Chatterji
世界卫生组织
瑞士日内瓦 | Thanh Lê
Westat公司
美利坚合众国
马里兰州罗克维尔市 |
| James R. Chromy
研究三角研究所
研究三角园区
美利坚合众国北卡罗来纳州 | James Lepkowski
密歇根大学
美利坚合众国
密歇根州安阿伯市 |
| Paul Glewwe
明尼苏达大学
美利坚合众国
明尼苏达州圣保罗市 | Michael Levin
美国普查局
美利坚合众国
哥伦比亚特区华盛顿 |
| Hermann Habermann
美国普查局
美利坚合众国
马里兰州休特兰 | Abdelhay Mechbal
世界卫生组织
瑞士日内瓦 |
| Graham Kalton
Westat公司
美利坚合众国
马里兰州罗克维尔市 | Juan Muñoz
独立顾问
智利圣地亚哥 |

Christopher J.L. Murray
世界卫生组织
瑞士日内瓦

Gad Nathan
希伯来大学
以色列耶路撒冷

Hans Pettersson
瑞典国家统计局
瑞典斯德哥尔摩

Kinnon Scott
世界银行
美利坚合众国
哥伦比亚特区华盛顿

Pedro Silva
巴西地理和统计数据局
(巴西统计局)
巴西里约热内卢

Bounthavy Sisouphantong
国家统计局
老挝人民民主共和国万象

Diane Steele
世界银行
美利坚合众国
哥伦比亚特区华盛顿

Tilahun Temesgen
世界银行
美利坚合众国
哥伦比亚特区华盛顿

Mamadou Thiam
联合国教育、科学及文化组织
加拿大蒙特利尔

T. Bedirhan Üstun
世界卫生组织
瑞士日内瓦

Martin Vaessen
人口与健康调查方案
ORC Macro*
美利坚合众国
马里兰州卡尔弗顿

Vijay Verma
锡耶纳大学
意大利锡耶纳

Ibrahim Yansaneh
国际公务员制度委员会
[经社部/统计司]
纽约州纽约市

* 一家民意调查公司。

审稿人员

Oladejo Ajayi
统计顾问
尼日利亚拉各斯

Paul Biemer
研究三角研究所
研究三角园区
美利坚合众国北卡罗来纳州

Steven B. Cohen
保健研究和质量机构
美利坚合众国
马里兰州罗克维尔市

John Eltinge
美国劳工统计局
美利坚合众国
哥伦比亚特区华盛顿

Paul Glewwe
明尼苏达大学
美利坚合众国
明尼苏达州圣保罗市

Barry Graubard
国家癌症研究所
美利坚合众国
马里兰州贝塞斯达市

Stephen Haslett
梅西大学
新西兰
北帕默斯顿

Steven Heeringa
密歇根大学
美利坚合众国
密歇根州安阿伯市

Thomas B. Jabine
统计顾问
美利坚合众国
哥伦比亚特区华盛顿

Gareth Jones
联合国儿童基金会
纽约州纽约市

William D. Kalsbeek
北卡罗来纳大学
美利坚合众国
北卡罗来纳州查珀尔希尔市

Graham Kalton
Westat公司
美利坚合众国
马里兰州罗克维尔市

Ben Kiregyera
乌干达统计局
乌干达坎帕拉

Jan Kordos
华沙经济学院
波兰华沙

Phil Kott
 美国农业部
 全国农业统计处
 美利坚合众国
 弗吉尼亚州费尔法克斯市

Karol Krotki
 NuStats公司
 美利坚合众国
 得克萨斯州奥斯汀市

James Lepkowski
 密歇根大学
 美利坚合众国
 密歇根州安阿伯市

Dalisay Maligalig
 亚洲开发银行
 菲律宾马尼拉

David Marker
 Westat公司
 美利坚合众国
 马里兰州罗克维尔市

Juan Muñoz
 独立顾问
 智利圣地亚哥

Gad Nathan
 希伯来大学
 以色列耶路撒冷

Colm O' Muirheartaigh
 芝加哥大学
 美利坚合众国
 伊利诺伊州芝加哥市

Robert Pember
 国际劳工组织
 统计局
 瑞士日内瓦

Robert Santos
 NuStats公司
 美利坚合众国
 得克萨斯州奥斯汀市

Pedro Silva
 巴西地理和统计数据局
 (巴西统计局)
 巴西里约热内卢

Anthony G. Turner
 抽样顾问
 美利坚合众国
 新泽西州泽西城

Ibrahim Yansaneh
 国际公务员制度委员会
 [经社部/统计司]
 纽约州纽约市

第一部分

调查设计、实施与分析

第一章 导言

Ibrahim S. Yansaneh*

国际公务员制度委员会
联合国，纽约

* 经社部/统计司方法和分析股前任股长。

摘要

本章简要评述了发展中国家和转型期国家进行的住户调查，并概述了本出版物的广泛目标及其现实意义。

关键词：住户调查、运作特点、复杂调查设计、调查成本、调查误差。

A. 发展中国家和转型期国家住户调查

1. 以往数十年来，对发展中国家和转型期国家的家庭及个人人口统计和社会经济调查数据的需求与日俱增。这种数据已在各个层级的经济和社会政策分析、发展规划、计划管理及决策过程中变得不可或缺。为了满足这方面的需要，决策者和其他利益共享者经常求助于住户调查。于是，住户调查就成了发展中国家和转型期国家收集人口情报资料的最重要机制之一。现在这种调查已成为组织国家统计体系和制定政策的核心战略内容。当前，大多数国家都有了住户调查数据收集系统，但经验和基础结构的层次不尽相同。各国的国家统计局进行的调查一般都是为了多重目的，或者在性质和设计上实行一体化，提供有关各种人口的一系列人口调查和社会经济特征的可靠数据。发展中国家和转型期国家的住户调查也用于对中小企业和小型农业股份公司进行调查研究。

2. 在发展中国家和转型期国家，除了由国家经常预算拨付资金的全国性调查之外，还有各种国际机构主办的大量住户调查，其目的在于构建和监测这些机构所关注的各国特征和指标的估计数据，并对这些指标进行国际比较。大多数此类调查都是临时性的，但现如今愈来愈关注在联合国和世界银行等国际组织的技术援助下持续进行的多项目、多轮次综合调查方案的设计、实施、分析和推广的全过程。有关国际机构在发展中国家进行住户调查的突出实例有：ORC Macro为美国国际开发署执行的“人口与健康调查”（DHS）；在世界银行的技术援助下进行的“生活水平衡量研究”（LSMS）调查；以及由联合国儿童基金会进行的“多指标类集调查”（MICS）。这些调查方案分别在非洲、亚

洲、拉丁美洲和加勒比地区，以及中东地区的各种发展中国家实施。“人口与健康调查”和“生活平衡量研究”两项调查方案分别在第五章和第六章的案例研究中作了概述。另见世界银行提供的资料（2000年），其中详尽讨论了世界银行在发展中国家进行的包括“重点调查项目”和“核心福利指标问卷”（CWIQ）等在内的其他调查方案。关于“多指标类集调查”，见儿童基金会提供的资料（2000年）。“人口与健康调查”是早期调查方案“世界生育率调查”（WFS）的一个分支项目，此方案系由美国国际开发署和联合国人口基金会联合资助，在大不列颠及北爱尔兰联合王国、荷兰和日本政府的协助下进行的。关于“世界生育率调查”方案的详细情况，见Verma等（1980年）。

B. 本出版物的目标

3. 本出版物为在发展中国家进行调查提供了一个方法论框架。由于这些国家正在进行大量调查，所以，无论是住户调查数据的编制者还是用户，都对有关调查过程所有阶段的方法论研究及现行最佳方法的推广应用不断提出新的要求。这种方法论研究工作很多是在各国际组织以及经社部/统计司的赞助下通过该处的出版物和技术报告进行的。本出版物提供了这方面努力的最新成果。

4. 现在，发展中国家和转型期国家进行的大多数调查都以世界通用的方法和程序为基础。可是其中有不少调查是在有关国家的调查基础结构和技术能力相差十分悬殊、且预算非常短缺的环境中进行的。显而易见，不但需要改进根本性的调查方法，而且需要把这些方法传授给发展中国家和转型期国家。而达到这一目的的最佳途径，就是开展技术合作和加强统计能力建设。本出版物是作为这种能力建设中的一个工具而编写的，它为完善设计和实施住户调查并有效利用收集到的数据提供了所需要的中央技术资料及其他信息源。

5. 本出版物面向所有参与编制和使用调查数据的人，其中包括：

- 国家统计局的工作人员；
- 向各国提供技术援助的国际顾问；
- 参与住户调查数据分析的研究工作者和分析员；
- 与调查研究方法有关的教师和学生。

6. 本出版物提供了一个有关发展中国家和转型期国家的住户抽样调查设计、实施和分析等方面的综合数据与参考资料来源。读者在设计这种国家的新调查的时候可以使用本出版物第一部分中提供的一般方法论的信息及指导原则，同时参考第二部分的案例研究成果。具体地说，本出版物的目标是：

- (a) 提供一个涵盖发展中国家和转型期国家调查设计、实施与分析技术方面的中央数据和参考资料来源；
- (b) 协助调查实践者以较高的效率设计和实施住户调查；

- (c) 提供发展中国家和转型期国家已经或正在进行的各类调查案例研究，侧重于提供可归纳的一般特征，以帮助调查实践者为同一国家或其他国家设计和实施新的调查；
- (d) 较详细地考察调查工作的三个运作特征——即设计效果、成本和非抽样误差——的组成要素，并综合探讨这些特征或组成要素对不同调查和不同国家的可移植性；
- (e) 提供对复杂的抽样调查数据进行分析的实际指导原则，并对调查数据分析所用的计算机软件类型作详细比较。

C. 上述目标的现实意义

7. 在发展中国家和转型期国家进行的住户调查有许多共同特征。此外，各国、尤其同一地区的国家之间在基本人口关键特征方面也往往有许多相似之处。既然国与国之间在住户调查的样本设计和基础人口特点方面有相似之处，我们或许可以期望某些运作特点及其要素也比较相似或者可在国家之间进行移植。

8. 调查运作特点的可移植性有几方面的实用好处。首先，有关特定国家某一调查设计方面的信息可以为同一国家重复进行该项调查或者为该国或别的国家进行类似调查提供高效率的实际指导原则。第二，基本上没有或根本没有现行调查基础结构的国家，可以直接从具有较好调查基础结构和一般统计能力的其他国家的有关抽样设计和实施的经验数据中获益。第三，有可能从先进行的某项调查中“借用”与样本设计有关的代价高昂的信息，从而大量节约成本。再者，以往调查所取得的实践经验可以用于最大限度地提高拟议调查的设计效率。

9. 本出版物除了处理有关调查设计与实施的成本效益问题之外，还有一项重要的总目标，就是促进在发展中国家和转型期国家中进行高质量的住户调查工作。它以先前的联合国各项相关倡议为基础，譬如10多年前结束的《国家住户调查能力建设方案》。这里所载的各项案例研究在如下几方面提供了重要指导原则：在发展中国家和转型期国家行之有效的调查设计与实施；需要避免的易犯错误；以及在确保数据可靠性和减少调查总成本方面可以采取的改进措施。本出版物所述及的所有调查都是在发展中国家和转型期国家进行的，这一事实使其成为这些国家发展统计事业的一个针对性极强的有效工具。

10. 调查数据的分析与传播是发展中国家和转型期国家在能力建设方面亟待开发的“领域”之一。对许多调查数据所作的分析很少会超出基本比率和造册范围。调查数据的恰当分析和这种分析成果的及时传播，可以确保必要的信息能有效应用于制定政策和有关资源分配的决策。本出版物分别就如何进行较为精密的微观数据分析、如何阐释对数据分析进行设计的复杂性、如何将分析目标纳入设计阶段，以及如何使用特制软件包来分析复杂的调查数据等问题，提供了实际指导原则。

11. 总之，本出版物是全面了解发展中国家和转型期国家住户调查情况的综合参考资料来源。第一部分提供的技术材料配合第二部分提供的具体实例和个案分析，或许能在新的住户调查设计、实施与分析方面对全世界的调查实践者有所助益。

参考资料

联合国儿童基金会（儿童基金会）（2000年）。《十年终了多指标类集调查手册》。纽约：儿童基金会，2月。

Verma, V.、C. Scott和C. O'Muircheartaigh（1980年）。“世界生育率调查中的样本设计和抽样误差”。《皇家统计学会刊》，A辑，第143期，第431-473页。正在讨论中。

世界银行（2000年）。“非洲的贫穷：调查数据库”。请登陆<http://www4.worldbank.org/afr/poverty>。

A 部 分

调查设计与实施

第二章

发展中国家和转型期国家住户调查样本设计问题概述

Ibrahim S. Yansaneh*

国际公务员制度委员会
联合国，纽约

* 经社部/统计司方法和分析股前任股长。

摘 要

本章主要论述发展中国家住户调查所涉及的国家样本设计问题。论题包括抽样标架、样本大小、分层次多级抽样、“域”的估算和调查分析。此外，本章还全面介绍了调查过程的各个阶段，全书对这些阶段作了详尽处理，但本章着重强调每个阶段与样本设计过程的联系。

关键词：复杂样本设计、抽样标架、目标人口、分层、类集、初级抽样单位。

A. 引言

1. 发展中国家和转型期国家调查样本设计

1. 本章概述与发展中国家和转型期国家住户调查全国样本设计有关的问题。跟全书一样，本章的重点放在住户调查方面。没有明确包括商业和农业调查，但不少材料与这两方面有关。

2. 发展中国家和转型期国家的住户调查在样本设计方面有许多共同特征。大多数调查系以多级分层次的“域”概率样本设计为基础。为了节省成本，这些设计主要用于调查范围的开发和类集采访。通常在层次范围内进行样本选择（见B节）。在第一阶段选中的单位——在调查抽样文献中通常叫作“初级抽样单位”——往往是从前一次全国人口与住房普查中确定和使用的查点区中构建起来的。这种查点区或许是城市地区的一些市内行政区或农村地区的一些村庄。有些国家，候选的初级抽样单位包括普查监管区或行政区或行政分区。在每个初级抽样单位中选定的单位叫作“第二阶段单位”；在第三阶段选定的单位叫作“第三阶段单位”……如此类推。对于发展中国家和转型期国家的住户来说，第二阶段单位一般是居住单位或住户，而在第三阶段选定的单位通常是个人。一般来说，在一个多级设计中的最后阶段选定的单位叫作“最终抽样单位”。

3. 在发展中国家和转型期国家的调查中各国尽管有上述许多相似之处，但是就样本设计而言，国与国之间不尽相同，譬如讲，可能在目标人口、调查内容和目标、设计层次数、每一层次的抽样比率、初级抽样单位范围内的样本规模、以及选定层次的初级抽样单位数量等方面存在差异。此外，基础人口可能在特定人口特征的优势比率、各层次内部及各层之间的异质程度，以及各层内部和各层之间特定分组人口的分布等方面有所变化。

2. 概述

4. 本章由如下几部分构成：A节作概括性介绍。B节探讨分层次多级样本设计问题。首先描述概率与规模成正比的抽样工作。然后结合整群抽样介绍“设计效果”概念。接下来讨论初级抽样单位数量和初级抽样单位范围内的第二阶段单位（居住单位、住户、个人，等等）数量上的最佳选择。讨论中考虑的各种因素包括预定的调查估计数的精确度要求和有关实地组织工作的一些实际考虑。C节讨论抽样标架及相关问题，并就这些问题推荐一些可能的解决方案。D节讨论“域”的估计数问题，以及为满足相互竞争的需要以产生全国和分区层级估计数可能考虑的各种分配计划。E节讨论为达到预定的标准误差和估计数变异系数的精确度水准要求进行的抽样规模测定。F节讨论调查数据的分析，特别强调调查数据的适当分析必须考虑到产生数据的样本设计的特征。G节概括了发展中国家和转型期国家住户调查设计中的一些重要问题。附件中提供了一幅流程图，描绘典型调查过程中所涉及的各个重要步骤以及这些过程步骤之间的相互关系。

B. 分层多级抽样

5. 发展中国家和转型期国家的大多数调查都以分层次多级分组设计为基础。这么做有两个理由。首先，由于缺乏住户或住址清单，或者清单质量很差，所以有必要先选定一个地理单位样本，只在选定的单位内编制住户或住址清单。然后才可以从这些清单中选择住户样本。第二，采用多级设计可以控制收集数据的成本。我们在本节讨论一个典型多级设计中的统计和运作方面的问题。

1. 明晰分层法

6. 每个抽样阶段都要分层。不过，在采集初级抽样单位的时候划分阶段的好处尤为明显。因此在选择初级抽样单位以前对其进行完善的分层十分重要。

7. 分层的工作是将人口中的各种单位分割成相互独立而整体上包罗万象的一些小组或层次。然后再从每个层次选取不同的样本。分层的首要目的是提高调查估计数的精确度。在这种情况下，各层的构成方式应该是：就调查所关注的特征而言，同一层次的各单位要尽可能同质；不同层次的各单位要尽可能异质化。分层的其他好处包括：（一）管理方便灵活；（二）重要的“域”和特殊分组人口的代表性有保障。

8. 许多国家以往的样本设计和数据分析经验表明，在每个国家不同行政区之间和不同城乡地区之间人口特征分布差别悬殊（见第二十二章、第二十三章和第二十五章）。这就是为什么这些国家的调查中明晰的层次均以行政区和政区范围内的城乡地区为基础的原因之一。有些行政区——比如首都城市——可能没有农村成分，而有些政区则可能没有城市成分。最好在最终选定明晰抽样层次以前先检查一下住户和个人的抽样率分布情况。

9. 有些地区不但国家一级需要估计数，而且每个行政区或分区（诸如省、部或区）也需要估计数。可以通过根据上述各种关注“域”分层来控制样本的分布。例如，在第二十二章讨论的“人口与健康调查”（DHS）中，初步分层系以需要估计数的行政区为基础。在行政区内，再按照城市成分对农村成分或其他类的行政区划来进一步分层。在各“域”之间设定不均衡抽样率，以确保“域”的估计数足够准确。一般来说，如果有许多“域”要求可靠数据的话，就需要进行大规模的总体抽样。关于“域”的估算问题将在D节讨论。

2. 隐含分层法

10. 在每个明晰层次中，往往采用一种叫作“隐含分层法”的技巧来选择初级抽样单位。在选择样本之前，先要参照被认为与所关切的变量高度相关并可供该层次每个初级抽样单位所用的一、两个变量，对一个明晰层次中的各种初级抽样单位进行分类。然后选定一个成体系的初级抽样单位样本。采用隐含分层法可以确保该初级抽样单位的样本能推广运用到各种类别的分层变量。

11. 对于发展中国家和转型期国家的许多住户调查来说，隐含分层法系以明晰层次中各种单位的地理排序为基础。有时候使用隐含分层法来选择初级抽样单位，其变量包括：在城市地区有住宅区（按低收入、中等收入和高收入划分）、开支类别（通常按五分位数划分）、族群和居民区；在农村地区有耕作区、家禽或牲畜保有量、非农业工人所占的比例，等等。在社会经济调查中，隐含分层法的变量包括：属于贫困类住户所占的比例、受过中等以上教育的成年人所占的比例、距离大城市中心地区的远近，等等。用于隐含分层法的变量通常是从普查资料中获得的。

3. 初级抽样单位的样本选择

良好初级抽样单位的特征

12. 对于发展中国家和转型期国家的住户调查来说，初级抽样单位往往是同一层次中的一些小的地理区域。如果有普查资料可用，初级抽样单位可能是普查中确定和使用的查点区。有时也利用类似地区或地方人口登记造册的资料。农村地区可能把村庄列为初级抽样单位。在城市地区，初级抽样单位则可能以市内行政区或街区为基础。

13. 鉴于初级抽样单位会影响到此后调查过程所有各阶段的质量，所以一定要确保所指定的初级抽样单位质量良好并以较为有效的方式为调查所选定，这点十分重要。一般来说，初级抽样单位达到良好质量的标准是：

- (a) 有明晰可辨且长期稳定的界定范围；
- (b) 完全覆盖目标人口；
- (c) 有服务于抽样目的规模量数；
- (d) 有为分层所需的数据可用；
- (e) 要有足够的数量。

14. 选择样本之前需要对抽样标架进行质量评价。就查点区的范围而言，第一步是按照关注“域”逐一检查其普查计数。总的来说，应把相当大的注意力放在初级抽样单位的性质和住户及个人在整个人口和各个不同关注“域”中的分布方面。通过对这些分布情况的仔细审查，将为初级抽样单位的选择提供决策依据，并为达到良好初级抽样单位的质量标准而确定需要加以调整的单位。一般来说，如果住户和个人数量在不同的初级抽样单位和不同的时期变化幅度太大，就会给实地组织工作带来不利的影响。如果所选择的初级抽样单位概率均等，也会对调查估计数的精确度产生负面效应。

15. 自然选择的初级抽样单位往往不可用，因为它们缺乏一种或多种上述特征，在这个意义上是有缺陷的。这种初级抽样单位在使用前需要加以修改或调整。譬如讲，如果认为查点区的范围不是界定得很好，那么可以用范围大一些的或更明确界定的单位——比如行政区、村庄或社区——作初级抽样单位。另外，有时被认为过大的初级抽样单位可以拆分或者作为一个层次来单独处理，这种单位往往称之为确定型选择或“自我代表型”初级抽样单位（见 Kalton, 1983年）。较小的初级抽样单位通常与相邻单位组合在一起，以达到预定的每个初级抽样单位最低住户数量的要求。最好在选择样本之前对这些规模过小或过大的初级抽样单位加以调整。

16. 为了确保初级抽样单位范围内的住户均衡分布，有时候要将非常大的初级抽样单位分割成若干规模适中的小单位，需要从中随机抽选一个细分单位做进一步的实地工作，比如住户列表造册。这叫作“程序分块”或“分割”过程。请注意，对过大初级抽样单位的选择和分割意味着额外增加了一个抽样阶段，必须在加权过程中将其考虑在内。

17. 非常小的初级抽样单位也可以同初级抽样单位框架中的相邻初级抽样单位合并，以达到预定的初级抽样单位最低限度的规模要求。在选择初级抽样单位期间或之后进行这种组合可以大大减少初级抽样单位组合所牵涉的工作量。但这是一个冗长乏味的过程，需要坚持严格的规则并保存大量纪录。Kish（1965年）描述了在样本选择期间或之后合并初级抽样单位的程序。这种程序的一个不利方面是，它不能保证选定参与组合的初级抽样单位是相互毗邻的。因此，在规模不足的初级抽样单位数量很大的情况下，不主张采取这种做法。

规模量数不准确的问题及可能的解决办法

18. 在把调查用作初级抽样单位（这是在发展中国家和转型期国家常见的做法）的架构中最普遍的问题之一，就是规模量数可能极不准确。规模量数一般是基于前次普查的初级抽样单位中的个人或住户数量的计数。这些计数有可能早已过时，由于多种因素——比如人口迁移、战争和自然灾害等导致城市地区扩展和其他地区收缩——的缘故，致使有关的数据可能与当前实际情况有明显的差别。规模量数不准确就会导致第二阶段单位分布和子样本规模失控，而这有可能给日后的实地运作带来严重问题。解决规模量数不准确的办法之一，就是在选择住户之前先进行一次彻底的造册工作，以便在选定的初级抽样单位创建一个住户框架。另一个解决办法就是选择概率与估计规模成正比的初级抽样单位。下面第4节和第5节阐述了这两种做法。在使用普查查点区作为初级抽样单位方面还存在另外一些普遍问题，包括缺少质量较好的地图以及目标人口覆盖面不完整等，这是C节讨论的与抽样标架有关的问题之一。

4. 概率与规模成正比的初级抽样单位采样

19. 在选择样本之前，先使用B.1节和B.2节中列举的变量对初级抽样单位进行明晰分层和隐含分层。发展中国家和转型期国家的住户调查大多选择概率与规模量数成正比的初级抽样单位。在选择样本之前给每个初级抽样单位设定一个规模量数，这种规模量数通常以最近普查中所记录的或最近更新的住户或个人数量为基础。然后在每个明晰层次中单独选定一个其概率与指定规模量数成正比的样本。

20. 概率与规模成正比（PPS）的抽样是利用辅助数据大幅提升调查估计数精确度的一种方法，这种方法特别适合在规模量数比较准确、且所关注的变量与单位规模相互关联的情况下采用。它是大多数住户调查所选用的初级抽样单位采样方法。PPS抽样可以产生不均等的初级抽样单位选择概率。初级抽样单位的规模量数基本上决定了这种单位的选择概率。然而，在选定的初级抽样单位内部，当这种方法与某个适当的子抽样部分相结合用以选择住户的时候，它就可以导致产生一个总体上自我加权的住户样本，在这个样本中所有住户都有同样的选择概率，而不为其所在的初级抽样单位所左右。它的主要吸引力就在于促成每个初级抽样单位大体均衡的样本规模。

21. 对于住户调查来说，住户数量就是选择初级抽样单位所用的PPS规模变量的一个良好实例。诚然，一个初级抽样单位内的住户数量会随着时间的推移而不断变化，并且可能在选取样本的时候已经过时。不过正如第18段所讨论的，有几种方法可以处理这个问题。在农场调查中，经常采用的PPS规模量数就是农场的规模。这样选择的部分理由是：在农场调查中，各种典型的重要参数——诸如收入、作物产量、牲畜保有量和各项开支等——均与农场的规模相关联。在商业调查中，典型的PPS规模量数包括员工人数、经营场所的数量以及年销售额。跟住户数量一样，这些PPS规模量数也很可能依时而变，这一事实必须在样本设计过程中考虑在内。

22. 让我们来考虑从一个两阶段设计中取得一个住户样本：设 α 为第一阶段选定的初级抽样单位总数，在第二阶段选定住户样本。假定第 i^{th} 个初级抽样单位的规模量次（比如最后一次普查时的住户数量）为 M_i 。如果初级抽样单位是按概率与规模成正比（PPS）选择的，那么选定第 i^{th} 初级抽样单位的概率 P_i 可由下式给出：

$$P_i = \alpha \times \frac{M_i}{\sum_i M_i}$$

23. 现在，设 $P_{j|i}$ 代表在第 i^{th} 初级抽样单位中选定第 j^{th} 住户的有条件概率——即假定已在第一阶段选定了第 i^{th} 初级抽样单位。那么，根据该设计，在第 i^{th} 初级抽样单位中选定第 j^{th} 住户的无条件概率的选择方程式 P_{ij} 即为：

$$P_{ij} = P_i \times P_{j|i}$$

24. 若想在抽样份额为 $f = P_{ij}$ 的情况下取得一个同等概率的住户样本，那么必须按照与其所在的初级抽样单位选择概率成反比的适当的比率来选择住户，亦即：

$$P_{j|i} = \frac{f}{P_i}$$

25. 如果各初级抽样单位均为实际规模，且样本选择与数据收集之间没有规模量数上的任何变化，同时如果在每个取样的初级抽样单位中选定了 b 个住户的话，那么我们即可获得一个自我加权的住户样本，其选择概率由下式给出：

$$P_{ij} = \alpha \times \frac{M_i}{\sum_i M_i} \times \frac{b}{M_i} = \frac{\alpha \times b}{\sum_i M_i} = f$$

式中， f 为常数。

26. 这个程序的问题在于，实际当中很少能够知道准确的规模量数。不过往往有可能从最近的普查或其他一些可靠的来源获得一些不错的估计数，比如人口和住户计数。这样，我们就可以应用一种叫作“概率与估计规模成正比”（PPES）的抽样程序。在第二阶段选择住户的两段设计中，PPES抽样可以有两种选择，即：(a) 在每个初级抽样单位中按固定的比率选择住户；或者(b) 在每个提取的初级抽样单位中选取固定数量的住户。

27. 按固定比率的PPES住户抽样实施方法如下：设 N_i 代表实际规模量数实际值，并假定 M_i 值是 N_i 的正确估计数。这样，我们对第 i^{th} 初级抽样单位应用 b/M_i 的抽样率，即可得到如下式给出的样本规模：

$$b_i = \frac{b}{M_i} \times N_i$$

28. 请注意，在初级抽样单位范围内按照固定比率（与初级抽样单位规模数成反比）进行子抽样，需要确定每个初级抽样单位的取样比率，以便结合初级抽样单位的选择概率，获得一个同等概率的住户样本，而不管初级抽样单位的实际规模大小。可是这种程序不能控制第二阶段样本的规模，因此也就无法控制总体样本规模。在住户数量比预期数量多的情况下从初级抽样单位中抽样的住户就比较多，而在住户数量比预期数量少的情况下从初级抽样单位中抽样的住户就比较少。这对实地组织工作是有影响的。此外，如果早已过时的规模量数造成实际样本的极大差异，就需要改变抽样比率，以扩大样本规模，提高初级抽样单位间的同质性，而不惜在某种程度上偏离自我加权设计的要求。

29. 第二种程序，即每个初级抽样单位选择固定数量的住户，可以避免每个初级抽样单位样本规模变化无常的不利情况，但不会产生自我加权的样本。不过，如果在紧接在选择初级抽样单位样本之前更新规模量数的话，就有可能产生十分精确的近似值，进而形成大体上自我加权的住户样本。

30. 总之，即便把初级抽样单位范围内按固定比率进行的子抽样设计为产生自我加权样本，也会出现此种方法导致偏离自我加权住户样本要求的情况。另一方面，虽然在初级抽样单位范围内选取固定数量住户的做法往往不能产生自我加权的样本，但是会出现此种方法导致产生大体上自我加权住户样本的情况。一旦发生偏离自我加权设计的情况，必须运用权数来补偿不同初级抽样单位中所产生的微分选择概率。

5. 住户样本选择

31. 一俟完成初级抽样单位的样本选择，应立即执行旨在将每个选定的初级抽样单位中的所有住户或所有居住单位或住所列表造册的程序。有时候登记表所列的是居住单位。因此，如果是对某个居住单位进行采样的话，在选定居住单位中的所有住户都要包括在内。这一登记列表的步骤意在创建一个据以选择住户的最新抽样标架。有效完成这一步骤的重要意义不容低估。造册工作的质量是影响目标人口覆盖面的最重要因素之一。

32. 在对每个初级抽样单位进行样本选择之前，可以按照地理分布和据认为与重要调查变量高度相关的其他变量对列入清单的住户加以分类（见B.2节）。然后采用同等概率系统化抽样程序从按顺序排列的清单中采集住户样本。正如B.4节所述，可以按照旨在生成所有住户均等的整体选择概率的抽样比率，亦可按照旨在使每个初级抽样单位生成固定数量的抽样住户的程序，在初级抽样单位的取样范围内对住户进行选择。B.4节分别讨论了这些做法的优缺点。

33. 住户往往是最终抽样单位，并且围绕选定的住户和这些住户的所有成员收集信息。在有关收入和支出的特别单元方面，住户是分析单位，经常把其

中一位有见识的受访者选定为住户填报人。对于某些被认为对住户成员比较敏感的话题（比如家庭虐待），就在每个取样住户选定一个代表个人（一般为一人）的随机样本。

6. 每个初级抽样单位选定的住户数量

34. 初级抽样单位系由地理上群集的成套住户组成。因此跟一般住户相比，就调查特征而言同一住宅群里的住户都比较相似（比如收入、教育、职业，等等）。类集的住宅区可以大大节省收集数据的成本，但是与非类集住户调查设计相比，在同一住宅群各单位之间的相关性会使得调查估计数方差发生膨胀（或者说降低精确度）。由此可见，就类集住宅区而言，调查设计者面对着不得不在节省成本和相应的精确度损失之间做出适当权衡的挑战。

35. 因类集取样导致调查估计数方差膨胀，从而促成了所谓“设计效果”。“设计效果”是指势必导致依据同样规模的简单随机样本得出的估计数方差成倍增加的各种因素，就是说，需要考虑到由于分层法、类集和加权而致使实际样本设计中出现的各种错综复杂的情况。它被界定为依据复杂设计得出的估计数方差相对于依据同样规模简单随机样本的估计数方差之比。关于设计效果及其在抽样设计中应用的详细情况，见本出版物第六章和第七章及其中援引的参考资料。一项估计数[譬如一个估计平均数 (\bar{y})]（因类集而产生）的设计效果可以大致表示如下：

$$D^2(\bar{y}) = 1 + (b-1)\rho$$

式中， $D^2(\bar{y})$ 表示估计平均数(\bar{y})的设计效果； ρ 是组内相关；而 b 是拟从每个类集中选出的住户平均数，亦即平均类集样本规模。“组内相关”是衡量一个类集居住区范围内的各个单位（就重要变量而言）同质程度的一个尺度。由于在同一类集居住区内的各种单位一般彼此相似，所以组内相关几乎总是正相关。就人群总体而言，同类正相关也许是因为同一居民区的住户都属于同一收入阶层这个事实；也许都对日常生活问题抱持同样的态度；并且往往处于同样的环境状况（比如气候、传染病、自然灾害，等等）。

36. 如果在估算标准误差的时候不把设计效果考虑进去，就可能导致对调查结果的无效解释。应该指出的是， $D^2(\bar{y})$ 值的大小直接关系到整群抽样规模 b 值和组内相关系数(ρ)。对于一个固定的 ρ 值来说，设计效果与 b 值呈线性增长。因此，为了达到减小设计效果的目的，最好尽可能压缩整群抽样的规模。表二.1说明平均类集规模和组内相关如何影响设计效果。譬如讲，设平均整群抽样规模 b 为每个初级抽样单位有20个居住单位， $\rho = 0.05$ ，那么设计效果即为1.95。换言之，这个整群抽样设计所产生的估计数的方差相当于大约半数住户的非类集样本（即简单随机样本）估计数的方差。从表二.1右侧可以看出，对于较大的 ρ 值而言，精确度损失甚至更大。

表二.1

选定的整群抽样规模与组内相关组合的设计效果

类集 抽样规模(<i>b</i>)	组内相关(ρ)								
	0.005	0.01	0.02	0.03	0.04	0.05	0.10	0.20	0.30
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
10	1.05	1.09	1.18	1.27	1.36	1.45	1.90	2.80	3.70
15	1.07	1.14	1.28	1.42	1.56	1.70	2.40	3.80	5.20
20	1.10	1.19	1.38	1.57	1.76	1.95	2.90	4.80	6.70
30	1.15	1.29	1.58	1.87	2.16	2.45	3.90	6.80	9.70
50	1.25	1.49	1.98	2.47	2.96	3.45	5.90	10.80	15.70

37. 一般来说，每个初级抽样单位中所选定的最佳住户数量，取决于在初级抽样单位范围内收集数据的成本结构以及与调查变量相关的同质或类集程度。假定一个两阶段设计，初级抽样单位在第一阶段选定，住户在第二阶段选定。同时假定，与初级抽样单位和住户抽样有关的总成本的线性成本模式为：

$$C = aC_1 + abC_2$$

式中， C_1 和 C_2 分别是每增加一个初级抽样单位的成本和每增加一个住户的成本；而 a 和 b 分别表示选定的初级抽样单位数量和每个初级抽样单位选定住户的数量（Cochran, 1977年，第280页）。依照这一成本模式，旨在最大限度减少样本平均数方差的最佳 b 值的选择公式（见Kish, 1965年，第8.3节b）大致如下：

$$b_{opt} = \sqrt{\frac{C_1(1-\rho)}{C_2 \rho}}$$

38. 表二.2给出了各种成本比率 C_1/C_2 和组内相关的最佳第二阶段样本规模(b)。请注意，在所有其他方面均等的情况下，最佳样本规模随着组内相关增加而缩小（就是说，样本可以更广泛地扩散到各个类集），并且随着每增加一个住户的成本相对于每个初级抽样单位成本的增加而缩小。

39. 在推导最佳类集规模中使用的成本模式虽然过于简单，但对于一般指导来说这大概就足够了。由于大多数调查都具有多重目的性，涉及到各种不同的变量，并相应涉及到不同的 ρ 值，因而 b 值的选择也就常常涉及到几种不同的最佳条件在某种程度上的折衷。

40. 在缺少确切成本信息的情况下，可以使用表二.2，按照各种不同的成本比率和组内相关，来确定一个类集居住区住户数量的最佳选择。譬如讲，如果事前已知包含一个初级抽样单位所需的成本相当于包含一个住户所需成本的四倍，且一个重要变量的组内相关值为0.05，那么，最好在同类集居住区选择9个住户为宜。请注意，在一个类集内选择的最佳住户数量不取决于现有的调查预算总额。总预算只能决定所要选定的初级抽样单位的数量。

表二.2

成本比率与组内相关选定组合的最佳子样本规模

成本比率 (C_1/C_2)	组内相关				
	0.01	0.02	0.03	0.05	0.08
4	20	14	11	9	5
9	30	21	17	13	10
16	40	28	23	17	14
25	50	35	28	22	17

41. 一般来说，在初级抽样单位范围内确定各种初级抽样单位和住户样本之间的样本分配方面，需要考虑的事项包括（通过设计效果取得的）各项调查估计数的精确度、数据收集成本和现场组织工作。如果差旅费用很高——比如在农村地区，最好在每个初级抽样单位中选择少数几个初级抽样单位，而每个初级抽样单位多选一些住户。另一方面，如果——比如在城市地区——差旅费用比较低，那就最好多选一些初级抽样单位，而在每个初级抽样单位中少选一些住户。在农村地区，则每个初级抽样单位多选一些住户，也许效率会高。以上这些选择必须以有利于访调员和主管人员之间的有效分工为原则。

C. 抽样标架

1. 发展中国家和转型期国家调查中的抽样标架特征

42. 对大多数住户调查来说，其目标人口系由非固定的平民百姓所组成。为了从这种目标人口获取所需要的数据，往往在住户一级进行访问调查。一般来说，只有被认定为一个住户的永久居民的人才资格列为受访者。一个住户的永久居民当中暂时离家在外的人——比如在外地度假或临时住院的人，或者在外地就读的学生，如果他们的住户被选中的话，一般也包括在内。但如果在外地就读的学生在其就读居住地被取样的话，则不列为其住户受访者，因为这类学生的相关数据将从他们的常住地获取。通常被排除在发展中国家和转型期国家的住户调查以外的群体包括：生活在军营或私人家庭中的武装部队成员；服刑人员和住在医院、养老院或其他机构的人员；无家可归的人；以及游牧民。这些群体之所以大多被排除在调查以外，是因为通常从这些人收集资料的时候会遇到许多实际困难。不过，需要根据调查的具体目标来决定是否应该将某个群体纳入抽样标架。

2. 有关抽样标架的问题及可能的解决办法

43. 跟其他类型的调查一样，出自住户调查的数据质量在很大程度上取决于据以选取调查样本的抽样标架的质量。遗憾的是，有关抽样标架的问题是住户调查的一个不可避免的特征。本节讨论的是其中一些问题并推荐可能的解决办法。

44. Kish (1965年, 第2.7节) 提出了一个对与抽样标架有关的四个问题进行分类的有效方法及其可能的解决办法。这四个问题是: 未覆盖; 要素类集; 项目空白; 重复列表。现在, 我们结合发展中国家和转型期国家进行的多级调查设计来讨论这些误差。

45. “未覆盖”这个词语是指抽样标架没有覆盖所有目标人口, 以至于某些抽样单位没有可能被纳入样本。未覆盖是发展中国家和转型期国家进行住户调查当中值得关切的一个主要问题。在大多数发展中国家和转型期国家, 基于调查人口计数样本的估计数远远低于其他来源的人口估计数, 这一事实即可证明未覆盖问题所产生的影响。

46. 未覆盖分三个层级, 即: 初级抽样单位层级、住户层级和个人层级。对于发展中国家和转型期国家来说, 初级抽样单位的未覆盖问题不如住户及抽样住户中合格个人的未覆盖问题那么严重。譬如讲, 初级抽样单位未覆盖的问题多半是在一个国家的某些地区被故意排除在调查之外的情况下发生的, 而这些地区是因为发生战乱、自然灾害、或其他缘故而不宜进入才被排除在外的。另外, 一些住户或人口稀少的边远地区的居民有时也被排除在住户调查之外, 因为他们在整个人口中所占比例很小, 故而对人口数据影响甚微。但在住户和个人层级, 未覆盖则是一个比较严重的问题。由于存在涉及住户结构与成分的复杂定义和概念问题, 所以有些住户和个人可能被错误地排除在调查以外。有这样一种潜在可能性, 就是不同的访调员或负责住户及住户成员登记造册的人可能对上述定义和概念问题做出不一致的解释。因此有必要做出严格的操作指示, 以指导采访员哪些人应算作住户成员, 并使其正确把握住户或居住单位的确切定义。作为处理这方面问题的一个手段, 应该把住户及合格住户成员登记造册的质量问题列为在发展中国家和转型期国家开展有关方法论的工作与培训的一个关键“领域”。

47. 项目空白的问题是在抽样标架名册没有包含目标人口要素的情况下发生的。对于居住单位名册范围来说, 空白项目可能相当于一个空宅。在对特定分组人口进行抽样的时候也会出现这种问题, 比如上一年生孩子的妇女。有些被列入名单并抽样的住户不包括任何在上一年生产的妇女。如果可能的话可以在选择样本之前将空白项目从表册范围中删除。可是在实际执行的情况下这样做不划算。一种比较切合实际的解决办法是在选择样本之后鉴别和删除空白项目。不过, 删除空白项目意味着最终的样本缩小了, 而且规模变化无常。

48. 重复列表的问题是指目标人口单位不止一次地出现在抽样标架。譬如讲, 在对游牧民或一年中有部分时间在某一地点居住的人进行调查抽样时, 可能会发生这种问题。避免重复列表的一个办法, 就是预先规定一个惟一的列表作为实际列表, 而其他列表都算作空白表。只有在这个惟一列表被抽样的情况下, 该单位才被纳入样本。譬如讲, 游牧民不断迁移放牧牲畜的地点, 以寻找牧场和水源。他们可能在前往水源之地的时候被取样。根据牲畜的饮水周期(据报道马的饮水周期长一些), 有些牧民可能在调查收集资料期间要去不止一个水源地。为了避免重复列表, 可以在规定的日期以后, 以其首次到访一个

水源地为准，作为惟一的列表鉴别，尔后的各次到访均作为空白表处理。否则，抽样单位的权数需作调整，以将重复因素考虑在内。关于这方面的做法实例，见Yansaneh（2003年）。

49. 要素类集的问题是在一个单一的抽样标架列表实际上包含目标人口的多个单元的情况下发生的。譬如讲，一个住宅清单中可能包含一些内有不止一个住户的住所。在这种情况下，如果将与抽样住所有关联的所有住户都包括在内的话，就会产生这样一个样本：其中所有住户都有同样的概率被选作住所。请注意，在类集住所中随机选择单位之一的做法，可以自动导致选择的不均等概率，这种概率需要通过加权予以补偿。

3. 抽样标架的保持和评价

50. 良好抽样标架的构建和保持是一项既费财力又费时间的工作。发展中国家和转型期国家具有根据持续十年普查资料来源创建这种抽样标架的潜力。建议每个国家统计局都把创建和保持一个在前次普查中界定并使用的主要查点区抽样标架摆在高度优先地位。在完成普查后不久就应确立这样的抽样标架，因为离普查相隔时间越久，所牵涉的工作量越大。必须根据抽样标架对可能用作初级抽样单位的其他（可能较大的）地理区域设置相关标签。还应该包括可能对分层有用的数据，比如族群和种族的构成、中等开支或支出五分位数，等等。如果保持得当的话，这种主抽样标架可以用于包括反复调查在内的一体化调查系统。关于构建和保持主抽样标架的详细论述，见第五章。

D. “域”的估算

1. “域”要素的必要性

51. 近年来大多数国家不但在中央一级而且在国家所属各级行政区或辖区对可靠数据的需求与日俱增，主要因为大多数发展或干预规划都是在诸如行政地区或管辖区的地方各级实施的。就地方层级的规划实施或资源分配做出重要决策，需要有同一层级的准确数据作基础。

52. 为了本章讨论的目的，我们把“域”一词定义为：在调查设计中单独计划估计数的任何一个人口子集。一个“域”可以是一个层次，或多个层次的组合，或一个行政区，或者在该行政区范围内的城、镇（乡）或其他区划。例如，许多国家调查得出的估计数分别按行政区发表。然后这些行政区也可以按“域”来处理，每个政区“域”划分为两个层次（比如城市分组人口和农村分组人口）或更多的层次。“域”也可以是按年龄、种族和性别等特征界定的各种人口调查分组人口。可是，当“域”跨越层次界限的时候问题就复杂了，譬如一个由享有保健服务的住户组成的“域”。

53. 在一次特定调查中将所关注的“域”的数量控制在适中的水平上，这一点很重要。如果“域”的数量很多，且每一个“域”都要求提供可靠的估计

数，那么势必需要非常庞大的样本规模。与大规模样本有关的问题将在E节讨论。

2. 样本分配

54. 为所关注的“域”提供精确的估计数，就需要给各个“域”分配适当规模的样本。可是，如果对人口规模相差悬殊的不同的“域”要求同等的精确度的话，就会发生冲突。若想要所有“域”的估计数都达到同样水准的精确度，那么均衡的分配（亦即每个“域”中的样本规模均等）就是最有效的策略。然而，这种分配办法会严重降低国家级估计数的功效。对国家级的估计数来说，每个“域”采用均等抽样份额的按比例分配往往是最适当的分配办法。当“域”的规模相差很大、而且同时需要国家估计数和“域”的估计数的时候，就需要在均等分配和均等抽样份额之间作些妥协折衷。

55. Kish（1988年）推荐了一个按比例分配与均等分配之间的妥协办法，其依据是一种与 $n_i = n \frac{W_h}{H}$ 公式成比例的分配，式中， n 是总样本规模； W_h 是 h 层中的人口比率； H 是层数。对于非常小的层次来说，第二个条件支配着第一个条件，因而可以避免向规模过小的“域”作分配。

56. 一种可供选择的方法就是将较小“域”的样本规模扩大到能满足精确度要求的地步。如果一个“域”很小，按比例分配的结果是该“域”的样本规模可能太小，以至于不可能产生足够精确的估计数。那么补救办法就是对小“域”过度取样，或者按较高的比率取样。

57. 总之，发展中国家和转型期国家的调查设计者往往不得不在国家级的精确估计数与“域”级的精确估计数之间做出抉择。当所关注的各个“域”的规模相差悬殊的时候，这个问题就变得更加严重。规避这种两难境地的一个办法是把各个“域”界定成大致均等的规模，大概可以通过合并现有的“域”来实施。或者，可以保持各个“域”明显不同，并适当降低小“域”的估计数精确度水准，或者也许根本就不发表这些“域”的估计数。

E. 样本规模

1. 影响样本规模决策的诸多因素

58. 调查数据的编制者和用户都希望有较大的样本规模，因为据认为这是为了使样本更具“代表性”并最大限度地减少取样误差、从而增加调查估计数的可信度所必要的。这个论点的提出几乎全然不顾大规模的样本可能会增加非取样方面的误差这一事实。本节讨论在确定某项调查的适当样本规模时必须予以考虑的各种因素。

59. 在一项调查的适当样本规模方面驱动决策的有三个主要问题，就是：

- 调查估计数的精确度（或可信度）；
- 调查所收集数据的质量；

- 收集、处理和散发数据的时间成本和金钱成本。

现在让我们来逐一讨论这些要素。

2. 调查估计数的精确度

60. 在发展中国家和转型期国家，大多数调查的目标包括估计某个时间点上某一特征（比如贫困类住户比率）的水平，以及该水平在一定时期内的变化（比如在两个时间点之间的贫困率变化）。我们结合某个特征在某个时间点的估计水平来讨论调查估计数的精确度问题。在其余的讨论中，我们将使用贫困住户的百分比——我们称之为贫困率——作为所关注的特征。

61. 一项估计数的精确度系以其标准误差来衡量。估计某个特定“域”的某项估计贫困率 p （用 $se(p)$ 来表示）的标准误差的公式由下式给出：

$$se(p) = \sqrt{d^2(p) \times \left(1 - \frac{n}{N}\right) \times \frac{p(100-p)}{n}}$$

式中， n 代表所关注“域”的住户总数； N 代表有关“域”的住户总数；而 $d^2(p)$ 则代表与复杂调查设计有关的估计设计效果。² 该样本中的人口比率， n/N ，叫作抽样分式，而因子 $[1 - (n/N)]$ （样本中所不包括的人口比率）叫作有限总体矫正因素（ fpc ）。 fpc 代表对估计数的标准误差所作的调整，用以说明这样一个事实，即：该样本未加置换地取自一个限定人口。

² 虽然 n 在上述 $se(p)$ 的公式中实际上应为 $n-1$ ，但在大多数实际应用中， n 往往大到可以忽略 n 与 $n-1$ 之间差别的地步。

62. 让我们以越南的数据为例来加以说明。根据1999年普查，该国住户总数， N ，为16 661 366户。欲知基于1999年普查的住户分布的详细情况，见 Glewwe和Yansaneh（2000年）。请注意，面对如此之大的人口总量，有限总体矫正因素 fpc 无论在何种情况下都是可以忽略不计的。表二.3提供了在假定设计效果值为2.0的情况下，各种贫困率估计数的标准误差和95%的置信区间。一个95%的置信区间就是载有95%实值的概率。该表说明，对于一个特定的同等规模来说，标准误差随着贫困率的增长而增长，最高可达 $p=50\%$ 。与之相关的95%的置信区间也随着贫困率的增长而加宽，当贫困率达到50%的时候置信区间最宽。因此，一般而论，对于特定同等规模和设计效果来说，与贫困率接近50%的“域”相对而言，贫困率大大低于或高于50%的“域”可能获得较为精确的调查估计数。³ 这就是说，贫困率极低或极高的“域”将需要较小的样本规模才能达到与贫困率接近50%的“域”相同的标准误差。比如，让我们来考虑一个“域”中有50个住户的样本规模。如果这个“域”的估计贫困率只有5%，那么置信区间即为 $5 \pm 2.7\%$ ；如果这个“域”的估计贫困率为10%，其置信区间即为 $10 \pm 3.7\%$ ；如果这个“域”的估计贫困率为25%，置信区间即为 $25 \pm 5.4\%$ ；而如果这个“域”的估计贫困率为50%的话，那么置信区间即为 $50 \pm 6.2\%$ 。

³ 如果贫困率高于50%（即 $p > 50\%$ ），标准误差就相当于 $100 - p$ 贫困率的标准误差，因此可以根据表三.3来推断。譬如讲，估计贫困率75%的标准误差相当于估计贫困率25%的标准误差。

63. 当然，将样品规模扩大到500个住户以上即可缩小置信区间的宽度（换言之，样本估计数变得更精确）。然而置信宽度的缩小并不与样本规模的扩大成比例，而是与样本规模平方根的增加成比例，这里用 $\sqrt{n/500}$ 来表示，

表二.3

根据各种样本规模在设计效果值为2.0的情况下贫困率估计数的标准误差和置信区间

样本 规模	贫困率 (%)									
	5		10		25		40		50	
	标准 误差	置信 区间	标准 误差	置信 区间	标准 误差	置信 区间	标准 误差	置信 区间	标准 误差	置信 区间
250	1.95	(1.2, 8.8)	2.68	(4.7, 15.3)	3.87	(17.4, 32.6)	4.38	(31.4, 48.6)	4.47	(41.2, 58.8)
500	1.38	(2.3, 7.7)	1.90	(6.3, 13.7)	2.74	(19.6, 30.4)	3.10	(33.9, 46.1)	3.16	(43.8, 56.2)
750	1.13	(2.8, 7.2)	1.55	(7.0, 13.0)	2.24	(20.6, 29.4)	2.53	(35.0, 45.0)	2.58	(44.9, 55.1)
1 000	0.97	(3.1, 6.9)	1.34	(7.4, 12.6)	1.94	(21.2, 28.8)	2.19	(35.7, 44.3)	2.24	(45.6, 54.4)
1 500	0.80	(3.4, 6.6)	1.10	(7.9, 12.1)	1.58	(21.9, 28.1)	1.79	(36.5, 43.5)	1.83	(46.4, 53.6)
2 000	0.44	(4.1, 5.9)	0.95	(8.1, 11.9)	1.37	(22.3, 27.7)	1.55	(37.0, 43.0)	1.58	(46.9, 53.1)

其中， n 是新样本的规模。譬如讲，在一个贫困率为25%的“域”中，将样本规模从500翻番到1 000，即可将置信区间的宽度减小一个 $\sqrt{2}$ 商数，也就是说，从± 5.4%减小到± 3.8%。应根据由此而造成的调查业务管理的复杂程度、成本和非取样误差的增加情况，对置信区间宽度的减小予以仔细加权。

64. 调查估计数的精确度往往依据所关注估计数的变异系数来表示。跟以往一样，我们把注意力仅限于一个国家中被划为贫困类的住户百分比方面。一个贫困率估计数的变异系数用 $cv(p)$ 来表示，可由下式给出：

$$cv(p) = \frac{se(p)}{p} = \sqrt{d^2(p) \times (1 - \frac{n}{N}) \times \frac{(100-p)}{np}}$$

65. 表二.4提供了假定设计效果为2.0，一个估计的贫困率对各种不同样本规模的变异系数，其中 cv 以百分比来表示。该表说明，对于一个特定样本规模来说，估计贫困率的估计变异系数随着实际百分比的增加而稳步下降。对一个特定贫困率来说，变异系数也随着样本规模的缩小而减小。设样本规模为500：当 $p=5\%$ 的时候，其变异系数为28%；当 $p=10\%$ 的时候，变异系数为19%；当 $p=25\%$ 的时候，变异系数为11%；当 $p=40\%$ 的时候，变异系数为8%；当 $p=50\%$ 的时候，变异系数为6%；当 $p=60\%$ 的时候，变异系数为5%；当 $p=75\%$ 的时候，变异系数为4%；当 $p=90\%$ 的时候，变异系数为2%；当 $p=95\%$ 的时候，变异系数为1%。当样品规模扩大的时候，估计的变异系数便相应减小。请注意，与表二.3不同的是，表二.4中所展示的变异系数不是贫困率的对称函数。

表二.4.

根据不同的样本规模在设计效果值为2.0的情况下贫困率估计数的变异系数

样本规模	贫困率 (%)								
	5	10	25	40	50	60	75	90	95
250	39	27	15	11	9	7	5	3	2
500	28	19	11	8	6	5	4	2	1
750	23	15	9	6	5	4	3	2	1
1 000	19	13	8	5	4	4	3	1	1
1 500	16	11	6	4	4	3	2	1	1
2 000	14	9	5	4	3	3	2	1	1

3. 数据质量

66. 在确定一次调查的样本规模时，一个重要考虑方面就是即将收集的资料的质量问题。应尽可能保持数据的最高质量，以使人们对从中产生的估计数有信心，这一点十分重要。在调查的每个实施阶段检查数据质量是必不可少的。因此，重要的是要在合理的限度内保持样本规模，以便在保证时间效益和资金效益的双重意义上进行适当的检查和编辑工作。

67. 一个与样本规模有关、影响数据质量的因素就是从事这方面研究的人员数量。譬如讲，样本规模较小，所需调查采访人员就比较少，因而会对这些采访员进行更仔细的挑选。尤其在样本规模较小的情况下，所有采访员都应该从受过良好训练和有工作经验的阶层中挑选。此外，由于采访员人数少，使培训工作可以比较集中，并且可以把更多的调查资源配额投入到这方面来，因而采访员也就能够受到更好的培训。所需培训材料会少些，而采访员能在培训和实地工作期间得到更多的个别关注。所有这一切都会导致在收集资料阶段和后期数据编辑加工阶段少出问题。最终，可供分析的数据会有较高的质量，从而会让决策者对依据这些资料所作的决策有较大的信心。

68. 除了有关所收集数据质量的关注事项之外，大规模的样本也会使得最大限度减少拒绝响应调查的努力更加困难，而且成本会更高（见第八章）。尽可能把拒绝响应调查的现象限制在最低的水平上，以减少在调查估计数中出现重大偏差的可能性，此事至关重要（见F.1节）。如果我们不能确保相当一部分可能明显有别于业已纳入调查者的人口响应调查，就有可能出现重大偏差。譬如讲，农村地区有些相对高收入的人往往不大可能参加住户调查。如果不能把这类人口当中的很大一部分纳入调查的话，就会导致对诸如全国平均家庭收入、教育程度和识字率等人口特征估计不足。如果样本较小，对起初不愿参加调查的住户进行回访以说服他们参加就容易多了，而且成本效益较佳。由于说服起初不愿参加者、使其转变为参加者的工作可能既费金钱又费时间，所以给最优秀的采访员配备足够的资金和可用的时间以便促成不容推辞的有效交谈，这对确保调查数据的质量十分重要。

4. 成本和及时性

69. 调查样本的规模显然对调查成本有影响。一般来说，一项调查的总成本是固定间接成本和与选择及处理每个选样阶段的各个抽样单位有关各项成本的一个函数。因此，样本越大，实施调查的间接成本越高。第十二章对住户调查成本的相关组成部分进行了讨论。第十三章和第十四章提供了各项具体调查成本核算的经验实例。

70. 样本规模也对提供可供分析的数据所花费的时间有影响。及时提供可用的数据和估计数以便在适当更新的数据基础上做出政策抉择，这一点很重要。样本越大，为分析的目的而清理、编辑和权衡数据所花费的时间越多。

F. 调查分析

1. 抽样权数的拟定和调整

71. 为了弥补不够均衡的选择概率、因拒绝响应而造成的资料欠缺，以及样本与参考人口之间已知的差异，各种抽样权数是必不可少的。应该在估计所关注的人口特征的时候以及在估计所产生的调查估计数的标准误差的时候应用这些权数。

72. 可以认为，一个抽样单位的基权数就是为了估计的目的抽样单位所代表人口中的单位数。譬如讲，如果某一特定层次的抽样率是十分之一，那么从该层提取的任何抽样单位的基权数即为10；也就是说，被抽样的单位代表有关人口中的10个单位，其中包括该单位本身。

73. 抽样权数的展开通常从构建抽样单位的基权数入手，目的是纠正这些单位的不均衡选择概率。一般来说，一个抽样单位的基权数是其被选中纳入样本的概率之倒数。在多级设计的情况下，基权数必须体现每个阶段的选择概率。然后对抽样单位的基权数加以调整，以补偿未响应和未覆盖的情况，并使得经过加权的样本估计数与已知的人口总数相符。

74. 当最终调整的所有抽样单位的权数都一样的时候，该样本即可称之为“自我加权”的样本。而实际当中，由于如下几个方面的原因，样本并不是自我加权的。首先，抽样单位是按照不够均衡的选择概率选定的。的确，即便初级抽样单位经常是在概率与规模成正比的情况下选定的，而且住户是按某个适当比率从初级抽样单位中选出以达成自我加权设计，这种设计也会因为在每个住户选出一人接受采访而告无效。其次，选定的样本往往因为抽样标架有问题而出现一些纰漏（见C节），比如未响应或未覆盖。第三，由于各个“域”和特定分组人口都需要精确的估计数，所以往往不得不对这些“域”过度取样（见D节）。

75 如前所述，从所有抽样单位获得所有想要的信息这种情况十分罕见。譬如讲，有些住户也许一点儿数据都没有提供，有些住户也许只提供了部分数据——就是说，调查中有些问题获得了数据，但不是所有问题都获得了数据。

前一种不响应叫作单位不响应或根本不响应；后一种不响应叫作项目不响应。如果在答复者和不答复者之间有系统化差别的话，那么仅仅依据答复者得出的朴素估计数就会出偏差。为了减少这种偏差的潜在可能性，作为分析的一部分往往做一些调整，以补偿不响应的缺失。对项目不响应的标准补偿方法就是推算，这个问题不在本章讨论范围。关于推算方法及其在大型复杂调查中的应用的一般讨论，见Yansaneh, Wallace和Marker（1998年）及其中援引的参考资料。

76. 在单位不响应方面，有如下三种基本补偿做法：

- 对不响应的基权数调整；
- 选择超需要量的初级样本，以应对因不响应导致样本缩减的问题；
- 替换——即用没有被取样而具备类似于不响应住户关注特征的另外一个住户来取代不响应的住户。

77. 建议在住户调查中对不响应现象采取某种形式的补偿办法——要么调整响应住户的基权数，要么替换。替换的好处是它有助于将参与住户数量置于控制之下。但替换的办法也使得采访员摆脱了从原本抽样住户获取数据的压力。另外，试图替换不响应住户的做法需要花时间，而在替换过程中可能发生误差。譬如讲，可能用一个方便住户来替换不响应住户，而不是用专门指定的替补住户来替换。针对不响应问题调整样本权数的做法在世界各地的较大型调查中比较常用。基本上，这种调整把所有不响应的合格抽样单位都转换成相应单位。第八章较为详细地讨论了住户调查中的不响应和未覆盖的问题以及对这方面的缺失给予补偿的切实可行的办法（并见该章援引的各相关参考资料）。第十一章和第二部分的案例研究（第二十二章、二十三章和二十五章）还提供了具体调查的详情。

78. 可以在适当的时候对权数作进一步调整。譬如讲，如果有可靠的总人口资料，也可以采用分层后的调整办法，以便使某些变量的加权抽样分布符合已知的人口分布情况。关于如何分析采用分层后调整方法的调查数据，见Lehtonen和Pahkinen（1995年）提供的一些实际例证。

2. 住户调查数据分析

79. 为了对住户调查数据做出恰当的分析，必须具备如下几个条件：首先，有关的数据库必须载有反映样本选择过程的信息。特别是数据库要包含样本设计层次、初级抽样单位、二级抽样单位……等方面的适当标签。其次，要给数据文卷中的每个单位提供能反映每个抽样单位选择概率并对调查不响应及其他缺陷予以补偿的样本权数。第三，必须有足够的技术性文件资料对产生数据的调查进行样本设计。第四，数据文卷必须有适当的格式和结构以及有关在不同选样阶段各个抽样单位之间相互关联的必要信息。最后，必须有适当的计算机软件以及恰当使用这些软件的专门知识。

80. 需要有一个能体现实用样本设计复杂因素的、用以计算调查估计数标准误差估计值的专用软件程序。这些复杂因素包括分层法、类集和非均等概率抽样（加权）等。标准统计软件包一般不能用于标准误差估计，因为这种软件几乎总是假定数据是通过简单随机抽样获取的。一般来说，使用标准统计软件包会导致低估调查估计数的标准误差。现有几种可供复杂样本设计所获调查数据分析用的软件包。第二十一章对其中一些软件包进行了广泛评议和比较。

G. 结论

81. 我们重点围绕与发展中国家和转型期国家的住户调查设计有关的几个原则问题，提出如下结论：

- (a) 大多数住户调查的多重目的性：大多数发展中国家和转型期国家越来越重视建立与一次性特别调查相对应的多目的、多课题、多回合的、持续进行的一体化调查。调查的设计者从一开始就必须承认调查的多重目的性以及这种性质将对从中产生的数据提出的各种竞争要求。这些竞争要求通常对样本施加了往往难以满足的制约条件。因此，调查设计者的工作应包括与各方赞助者、决策者、国家统计局的数据处理者，以及国家各行政部門的用户进行广泛的讨论；这些初步讨论的目的是，在最终完成样本设计以前协调对调查设计提出的各种竞争要求并使之合理化；
- (b) 确定适当的样本规模：从一开始就应处理的主要问题之一，就是调查样本的适当规模问题。现在，不但在国家和地区级、而且在省一级甚至更低层次都对各项重要特征的精确估计数提出了日益增多的要求。这不同程度地导致要求扩大样本规模。花在通过大规模抽样减少取样误差以确保调查估计数可靠性方面的费用，远远超过了花在同样重要的、通过减少非取样误差确保数据质量问题方面的费用。建议调查设计者要对各种样本规模的选择和分配计划进行成本效益分析。部分成本效益分析应当包括讨论调查中的非取样误差问题及其对调查数据总体质量的影响。只能依据相关的成本效益来考虑对大规模样本的需求。正如D节所述，在样本分配当中应将考虑的重点放在所关注的各种“域”的方面，记住这一点很重要；
- (c) 调查设计与实施的文件编制工作：对许多调查来说，调查设计和实施过程都缺乏文件编制工作，或者这方面做得很不充分。要想使一个数据集对分析者和使用者有用，就绝对有必要在生成数据的设计全过程进行文件编制工作，其中包括样本选择、数据收集、数据文卷的编拟、抽样权数的构建（包括旨在弥补样本欠缺的任何调整），以及在可能的条件下详细规定标准误差的估计方法。假如没有这些文件编制工作，就不可能对数据做出恰当的分析。调查的文件编制工作对于与其他资料来源的沟通以及对于进行各种校验和补充分析也是必不可少的；

(d) 调查设计评价：设计完成后的分析评价工作是调查设计过程的一个重要方面。需要在规划阶段将这一重要实践列为整个预算编制过程的一个组成部分，为其指定划拨各种资源。对一项调查的现行设计进行评价，有助于改进未来调查的样本设计。这种评价可以揭示一些有益的信息，比如不均衡分配是否带来任何利得；以及现有规模量数与选择样本时的规模量数之间差距有多大（如果有的话）。然后可将此种信息用于为将来的调查制定更有成效的设计。

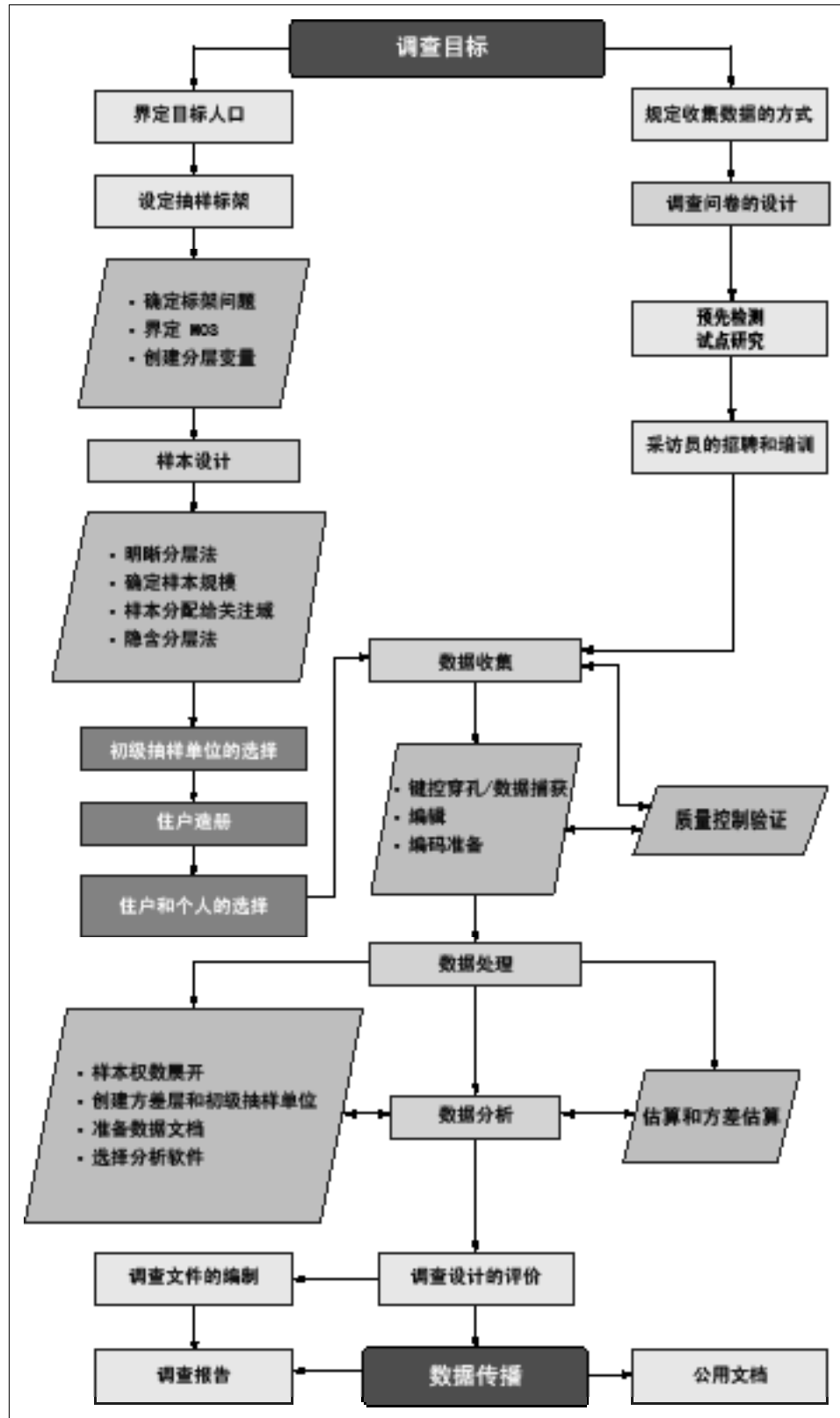
致 谢

笔者十分感激各位审稿人和编辑提出的建设性意见，尤其感谢Graham Kalton博士，多亏他的无数建议，本章初稿才有了相当大的改进。笔者在本章发表的意见纯属个人见解，未必反映联合国的政策。

参考资料

- Cochran, W. G. (1977年)。《抽样方法》，第三版。纽约：威利国际出版公司。
- Glewwe, P. 和I. Yansaneh (2000年)。《越南未来住户调查的发展》。《特派团向一般统计科提交的报告，越南》。
- Kalton, G. (1983年)。《抽样调查简介》。《社会科学汇编中的定量应用》，《赛奇大学论文，第35号》，加利福尼亚，比佛利山：赛奇出版社。
- Kish, L. (1965年)。《抽样调查》。纽约：威利国际出版公司。
- _____ (1976年)。“线性样本设计中的最佳和最接近效果”。《皇家统计学会志》，A辑，第139期，第80-95页。
- _____ (1988年)。“多种目的样本设计”。《调查方法》，第14期，第19-32页。
- _____ (1995年)。“统计效果的方法”。《官方统计杂志》，第11期，第55-77页。
- Lehtonen, R. 和E. J. Pahkinen (1995年)。《复杂调查设计与分析的实用方法》。纽约：威利国际出版公司。
- Lohr, Sharon (1999年)。《抽样：设计与分析》。加利福尼亚，太平洋丛林：达克斯伯利出版社。
- Yansaneh, I. S. (2000年)。《2000年土库曼斯坦小型人口普查的样本设计》。《特派团向国家统计和预测机构提交的报告，土库曼斯坦》。
- _____ (即将出版)。“样本权数的结构和运用”。《住户调查手册》。纽约：经社事务部/联合国统计司。正在筹备中。
- _____ L. Wallace和D. A. Marker (1998年)。《大型复杂数据集的填补方法：应用于国家雇主医疗保险调查》。《调查研究方法科的记录，美国统计学会》。维吉尼亚，亚历山德里亚：美国统计学会。第314-319页。

附件 调查流程图



第三章

发展中国家住户调查问卷设计概述

Paul Glewwe

明尼苏达州立大学应用经济学系
美利坚合众国明尼苏达州圣保罗

摘 要

本章评述有关适用于发展中国家住户调查问卷设计的若干基本问题。文章从问卷设计的第一步即拟订调查目标入手，然后结合基本制约因素的考虑对这些目标加以修订。讨论了上述广泛问题之后，就设计住户调查问卷的诸多方面提出了较为详细的忠告。本章还就问卷的实地检验和定稿提出建议。

关键词：问卷设计、调查目标、制约因素、试点检验、实地检验。

A. 引言

1. 住户调查可以提供有关各方面生活的大量信息。然而就问卷设计和现场实施而言，住户调查数据是否有用，在很大程度上取决于调查质量。设计调查问卷和实施住户调查的任务看似简单，可实际上住户调查的成功需要做艰苦的工作，付出大量的时间。

2. 本章对发展中国家住户调查问卷的设计过程作了基本概述。这里只是介绍性的陈述，因为问卷设计过程非常复杂，不可能在这样一个篇幅的章节作详尽描述。本章意在提出一些最重要的问题并逐一给予有益的忠告。凡是有意从事实地调查的读者都需要参考其他材料，以获得更详细的忠告。Grosh和Glewwe（2000年）开了一个很好的先河，他们就发展中国家的住户调查设计提供了极为详尽的信息。虽然在撰写此文章时作者心里想的是一个具体类型的调查——即世界银行主办的“生活水平衡量研究”（LSMS）项下的各次调查，但其中不乏几乎跟任何一类住户调查都有关系的忠告。读者还可以从Casley和Lury（1987年）、联合国（1985年）、Sudman和Bradburn（1982年）以及Converse和Presser（1986年）的文章中发现有关问卷设计的更加广泛——发表稍早——的论述。Husmanns、Merhan和Verma（1990年）详细讨论了如何设计一项劳动力调查。

⁴ 在一些调查中，抽样单位是住所而不是住户，但在这种情况下，可能被抽样住所的某些、亦可能所有住户成为调查的“报告单位”。另外，在所关注的人口中有某些人不能被纳入住户调查。这方面的实例有流浪儿童和游牧民。尽管如此，本章的大多数材料也适用于这些类别的人口。关于如何提取这些人口的样本，见联合国（1993年）。

3. 本章自始至终假定调查问卷将由对受访者进行家访的访调员实施，抽样单位是住户。⁴ 由于大多数住户调查收集的是住户每个成员的信息，所以调查不但以住户样本为基础，而且还以个人样本为基础。

4. 本章其余部分的编排如下：B节讨论“大致结构”，即调查的目标及其所面临的制约因素。C节就调查问卷的结构、格式和问卷设计的其他细节的组织提出意见。D节围绕从成立一个调查团队到问卷的实地检验和定稿的全过程提出建议。E节简要提出一些结论性意见。

B. 大致结构

5. 各种住户调查问卷在内涵和篇幅上差别很大。任何问卷的最后定稿都是上百、甚至上千决策过程的结果。必须要有个总体框架或“大致结构”才能确保拟订问卷的过程有序进行，并确保实现为调查规定的各项目标。为此，调查设计者必须就调查目标和开展调查所面临的诸多制约因素达成共识。本节解释如何确立总体框架，先从基本原则入手，然后提出一些实际意见。

1. 调查目标

⁵ 实施调查的组织希望得到答案的这些一般性问题，未必和调查问卷上请求住户成员回答的那些比较具体的问题一样。本节集中讨论前一类问题。

6. 政府机构和其他组织实施住户调查的目的是为了回答它们有关人口的一些问题。⁵ 因此，既然调查的目的是获取对这些问题的答案，那么调查问卷就应包含可以提供这些答案的数据。鉴于资源有限而且调查受访者的时间也有限，所以凡是无助于实现调查目标的数据就不应该收集。可见，住户调查设计的第一步就是要在调查目标上达成共识并将其写成文字。

7. 为了确立调查目标，调查设计者应从主办调查的组织希望得到答案的一套问题入手。有四类问题可以考虑。最简单的是关于当前人口基本特征的那一类问题。此类问题的实例有：

- 贫困人口的比率有多大？
- 失业率有多高？
- 营养不良的少年儿童占多大比例？
- 本国不同地区的农村住户分别种植哪些作物？

8. 第二类问题把住户特征与政府的政策和方案联系起来，以考察这些方案的覆盖面。此类问题的一个实例是：

- 参加某项特定方案的住户比率有多大？这些住户的特征与那些没有参加该方案的住户特征相比有何不同？

9. 第三类问题是关于住户特征在一定时期内的变化。政府机构和组织往往想要了解住户的生活状况是改善了还是恶化了。回答这类问题需要有相隔时间相当长的两次以上调查提供的数据，而每次调查都用相同的方式收集所关注的的数据。正如Deaton和Grosh（2000年）所解释的那样，收集数据的方式即便略有不同，也会导致数据不可比，从而造成误导。

10. 第四类、也是最后一类问题涉及到造成住户境况和特征的决定因素（或原因）。此类问题难以回答，因为它们不仅问正在发生什么事情，而且问为什么会发生这种事情。但这些问题往往是最重要的，因为提问者想要了解现行政策或方案、抑或假设的未来政策或方案对住户境况和特征有什么影响。如何来回答这些问题不一定能得到经济学家和其他社会科学家的认同，他们甚至有时会认为某个问题不可能得到答案。如果调查设计者认为此类问题很重要，那就需要进行周密的策划。但此种策划所涉及的问题超出了本章的讨论范围（关于对回答此类问题的要求，见Grosh和Glewwe（2000年）相关章节的详细讨论）。

11. 一旦就一套问题达成了共识，即可将问题表述为调查目标。譬如讲，关于当前失业率问题的存在就意味着，调查的目标之一是计量在经济活跃人口中的失业率。下一步是将这些问题按各自的重要程度依次排列起来。如果目标数量很多，就不大可能在一次调查中收集到为实现所有目标所需要的全部信息，这是预算水平低、调查能力有限和其他制约因素所决定的。如果发生了这种情况，就应该裁减一些（相对于实现目标所需要的收集信息的工作量而言）优先地位较低的目标。⁶ 在这个确定哪些调查目标可以达到的取舍过程中，有必要查看是否可以利用现有的其他一些资料来回答与某项目标有关的问题。任何可以利用其他来源的现有数据实现的目标都应该在新调查的目标清单中删除。这种成套合理目标的选择过程与其说是一门科学不如说是一门艺术，同时调查设计还必须考虑到以往收集与目标有关数据的经验和机构实施调查的整体能力等方面的因素。而一旦克服了上述种种挑战，这种方法就能帮助调查设计者们就一份住户调查所要实现的目标清单达成共识。

⁶ 裁减次要目标的一个替代方法，就是只从二级住户样本收集为实现目标所需的数据。这种做法牵涉的资源量会少一些，但也会降低估计数的精确度，并有可能增加实地调查的复杂程度。

12. 最后值得一提的是，有些调查设计者喜欢用一套表格来表现准备用调查数据来填写的成套问题或目标。这种方法通常叫作“列表计划”，对前三类问题最有用。较普遍的做法是利用在住户调查中收集的数据来回答问题（达到目标），这种方法可以称之为“数据分析计划”。这两种计划都可以非常详尽，随着住户调查的细节逐步搞定，方案也就成型了（这个问题在C节进一步讨论）。

2. 制约因素

13. 上述选择目标的过程必然要在“一整套”可行的制约条件下操作。调查设计者主要面临三大制约因素。首先最明显的是从事调查应拥有的财务资源。这个制约因素既限制了住户调查数量，也限制了采访特定住户的时间（反过来又限制了可以对每个住户提多少问题）。一般来说，样本规模（所调查的住户数量）和可从每个住户获取的信息量之间有各种不同的组合方式，而在既定的预算条件下，这两个调查特征之间存在某种权衡关系。尤其是，对于既定数额的财务资源来说，我们只能通过减少从每个住户收集的信息量来增加样本规模；反之亦然。⁷ 这对调查目标的数量和这些目标的精确度（亦即对基本问题回答的精确度）显然是有影响的：样本规模小，就可以从每户收集较多的数据，从而也就可以回答较多的关注问题，但由于样本规模缩小了，这些答案的

⁷ 通常，对于既定的预算来说，每户收集的信息量与采访户数之间的确切关系并不那么简单。譬如讲，事实上在既定的采访时间内，把问卷减掉一半并不能使样本规模翻一番，而且也不能通过缩短问卷来减少采访时间。

精确度也随之降低了。与此相关的一个问题是，就信息的准确性而言，数据的质量也受可用资源量的影响。比如说，如果可用的经费允许每户采访有更多的时间完成既定篇幅的问卷的话，所增加的时间即可用来回访住户，以纠正完成采访之后发现的误差或数据不一致的地方。

14. 调查设计者面临的第二个制约因素就是负责实施调查的组织的能力大小。庞大的样本规模和非常相近的住户问卷可能会超出执行组织在预期的质量水平上开展调查的能力。规模越大，需要雇用和培训的采访员及数据录入员就越多（假定要求完成调查的时限不能延长），这就是说，为了能雇用到必要数量的工作人员，该组织可能不得不降低对采访员和录入员的最低资格要求。同样地，住户问卷内容越广泛，就需要更多的培训工作和更多能胜任的工作人员，但在发展中国家，训练有素、工作能力较强的采访员和录入员往往供不应求。人们往往对这方面的制约因素认识不足，结果，在发展中国家进行的许多调查所产生的大量数据集质量有问题，因而其使用价值如何很难说。

15. 最后一个制约因素就是被采访住户提供所需信息的意愿和能力问题。首先，住户回答问题的意愿是有限的，因此，极为冗长的调查问卷所造成的响应负担很可能导致拒绝回答和/或数据不完整或不准确的比率很高。其次，即便受访者很愿意合作，但对于十分复杂的问题或者需要回想数月前甚至多少年前往事的问题，他们或许也很难回答。这与问卷的设计有直接的关系。譬如讲，如果提问少量的问题也许不能对住户的收入情况作出较为准确的估计，因此需要提出一长串问题，尤其对农村地区的务农住户情况更是如此：他们种了许多种作物，其中有部分收成留作自家消费，另一部分卖掉。

3. 一些实际忠告

16. 调查设计者需要在调查目标和面临的制约因素之间反复权衡和取舍，才能最终“涵盖”不但在上述制约条件下可行、而且在选定对调查单位最重要的目标这个意义上最理想的一套目标上。一旦弄清哪些事情是可行的，就有可能获得追加财务资源和为未来的采访员提供更多的培训，从而缓解这些制约因素。最近在同一国家完成的其他调查所取得的经验，应该可以在哪些事情可行、哪些事情不切实际的问题上提供有益的指导。如前所述，取得恰当的平衡与其说是一门科学不如说是一门艺术，不论本国经验还是外国经验都可以在取得这种平衡方面起到良好的指导作用。

C. 问卷设计的细节问题

17. 调查目标的“大致结构”一经确定，调查者就需要着手进行详尽而不可回避的、冗长乏味的问卷设计工作了。从一开始就应强调的总的要点是需要有个数据分析计划。此计划应详细解释需要哪些数据来实现既定的调查目标（或回答提出的问题）。调查设计者在编制调查问卷细节的时候必须经常参照这个计划。有的情况下，当问卷的详细设计工作对如何分析数据进行新的阐述的时候，必须对数据分析计划加以修改。凡在总体数据分析计划中没有用上的问题都要从问卷中删除。

18. 本章过于简单，从中难以详细了解如何将问卷设计工作与具体的目标及其相关的数据分析计划挂钩。若想了解有关不同类别调查的更加综合性的忠告，请参阅Grosh和Glewwe(2000年)。本节的其余部分将就如何完成制定住户调查细节的任务问题提供一些一般性的但是非常有用的忠告。

1. 模块方法

19. 住户调查问卷通常由几个部分组成，这些组成部分一般叫作“模块”。每个模块由一页或多页问题组成，分别就特定话题——比如住房、就业或保健等——收集相关信息。例如，第二十二章讨论的“人口与健康调查”系列就有关于避孕、生育取向和儿童防疫的模块。比较一般的做法是，几乎在任何一份就某个特定话题提几个问题的住户调查问卷中，都可以很方便地把这些问题集中到问卷的一页或多页，并把这一页或多页定名为该话题的模块；比如，上述教育问题就成了“教育模块”。这样一来，就可以把整个问卷看作一个模块集，少则3个模块，多则15或20个模块，依问卷所涉及的话题数量而定。每个模块包含若干问题，有时只有5-6个，但有时则多达50个甚至100多个。⁸ 非常大的模块（譬如含有50多个问题的模块）可以进一步细分为特定话题的子模块。譬如讲，一个关于就业问题的大型模块可以细分为如下子模块：第一职业；第二职业；就业史。无论如何，一份问卷的问题总量应保持在导致获取预期信息的最低限度。

⁸ 超过100个问题的模块可能会导致采访超时。关于问卷总体长度的问题，见D节。

20. 采用模块方法很方便，因为它可以让问卷的设计分两步走：第一步先决定需要哪些模块，也就是说，问卷将包括哪些话题以及模块的排列顺序。第二步是按问题逐一选择每个模块的设计。在实施两个步骤期间要不断参照调查目标和数据分析计划。

21. 依调查目标及其所面临的制约因素的不同，模块的选择和每个模块的细节会有很大的差异。不过，可以提出一些几乎适用于任何调查的一般性忠告。譬如讲，几乎所有调查都要收集有关家庭人口的信息以及有关这些人口的基本信息，比如年龄、性别和与户主的关系等。这些问题可以归纳成短短一页的“住户花名册”模块。该模块应该是问卷的首批模块之一，而且多数情况下应该是第一个模块。接下来许多住户调查问卷会问住户成员个人有关教育、就业、保健和迁移等话题的一些问题。任何这种提5个以上问题的话题大概都可以放入一个关于有关话题的专项模块。如果只问一个、两个或三个问题，也许把它们纳入住户花名册模块、抑或纳入另外一个提问家庭成员的模块会更方便些。

22. 一项住户调查的几乎所有模块都可以分为两大类：一类是上面所讲的提问住户成员个人的；另一类是提问住户的一般问题。关于前一类模块应该指出的是，向住户成员个人提出的问题不需要每个成员都一样。许多住户调查所提的问题只适用于某些类的家庭成员，比如五岁以下的儿童或育龄妇女。后一类问题的实例有，比如关于住户生活住所特征的问题，以及关于整个家庭在食品和非食品方面开支情况的问题。当然，任何这些模块的篇幅及其中问题的种类都取决于调查目标。

23. 最后，关于住户调查问卷中的模块顺序可以提几个一般性要点。首先，模块的次序应当与采访的运作次序相吻合，这样，采访员就可以从问卷第一页开始，接着一页一页地继续下去，直到问卷结尾，完成整个问卷的调查内容。有的情况下可能需要做例外处理，但一般来说按此顺序编排模块比较“自然”。

24. 第二个要点是：问卷中的第一批模式所包含的问题相对来说比较容易回答，其所属话题不太敏感。前面关于把住户花名册作为第一模块的建议与此处的建议是一致的，因为通常有关住户成员的基本信息都不是敏感话题。从不敏感话题的简单问题入手进行采访有助于采访员使受访的住户成员感到轻松自然，从而与之建立一种和谐亲善的沟通氛围。这就意味着，最敏感的模块应该放在问卷的末尾。这将使得采访员有尽可能多的时间取得住户成员信任，从而增加他们充分、如实地回答敏感问题的可能性。此外，倘若因为敏感问题终止了采访，起码已经取得了所有不敏感的信息。

25. 第三条原则就是将同一住户有可能回答的所有模块归拢到一起。比如说，有关食品和非食品开支的问题要放在一起，因为很可能一位家庭成员最能够回答这两类问题。这可以使这个家庭成员回答完他或她所能回答的这些模块中的所有问题，然后结束他或她的参与，而由其他家庭成员来回答剩余的模块。此处的一般要点是有效利用住户成员的时间，这会得到赞赏，从而增加他们的合作。这也有可能节省采访员的时间，因为每位受访者只需被召唤一次即可完成他或她对采访的贡献。

2. 问卷格式化和前后一致性

26. 选定了模块并排好了模块次序之后，接下来就开始了无可否认地冗长乏味的遴选具体问题并将其逐字逐句写成文字的工作。每当在特定国家进行这项工作的時候，最好先查看一下以往在该国——或许在邻近国家——就同样的话题进行的住户调查的成果。虽然最好的问题和措词取决于新调查的性质和目的，但还是可以提几点几乎对所有住户调查都适用的一般性建议。

27. 第一项建议就是几乎不论在任何情况下都要把问题写在问卷上，以便采访员通过读出问卷上的问题进行采访，这样就能保证对所有住户都提出同样的问题。另一种做法就是把调查问卷设计成措词尽量精简的形式，要求每个采访员用自己的话来提问题。这种做法不可取，因为它会导致许多差错。比如说，关于就业的模块有一个“问题”简单到只有“主要职业”几个字。这个问题就很不明确：它是指受访的当天或那一周从事的职业，还是以往12个月来的主要职业？对于一个从事两种职业的人来说，其“主要职业”是收入最高的工作，还是付出工时或工作日最多的工作？如果把问题详细写出来，就可以避免这种含混不清的提问，比如说：“过去七天来您做什么工作？如果您有不止一份工作，请告诉我在过去七天内您工作时间最长的工作。”图三.1提供了一页问卷的实例，是收集有关住房信息的（注意：所有问题都写成了完整的句子）。把所有问题都写出来的好处，在Scott等（1988年）的实验性研究成果中是显而易见的：没有写出来的问题所产生的差错比详细写出来的问题产生的差错多7到20倍。

图三.1
问卷格式图解

1. 这个住所是否属于您或某个家庭成员所有?

是..... 1

否..... 2

2. 您是如何取得这个住所?

继承..... 1

从某个家人购买..... 2

新建..... 3

合作安排..... 4

交易的结束..... 5

继承财产..... 6

其他..... 7

3. 您为这个单元支付了多少钱?

4. 您是否为这个住所分期付款?

是..... 1

否..... 2

5. 分期付款总价是多少钱?

金额 (货币单位)

时间单位

6. 您预计哪一年付清最后一笔分期付款?

年份

7. 您是否对土地拥有法定权利或在任何能证明所有权的文件?

是..... 1

否..... 2

8. 您是否对这个住所拥有法定权利或者任何能证明所有权的文件?

是..... 1

否..... 2

9. 是哪种类型的权利?

完全法定权利, 已注册..... 1

法定权利, 未注册..... 2

购买收据..... 3

其他..... 4

10. 如果您个人持有对这个住所的法定权利或文件?

请按照姓名写下您个人的身份证号码

第一身份证号码

第二身份证号码

11. 如果您购买的话, 您购买这个住所吗?

是..... 1

否..... 2

12. 如果您今天购买这个住所, 购买多少钱?

金额 (货币单位)

13. 请估计一下, 如果您让另一个人来住的话, 您能拿到多少租金?

金额 (货币单位)

时间单位

时间单位: 每天..... 3 每月..... 6 每年... 9

每周..... 4 每季度..... 7

每两周..... 5 每半年..... 8

200 下一页第20个问题

28. 第二项建议与第一项建议密切相关：问卷应载有对调查问卷中所使用的所有关键概念的精确定义，其基本目的是便于采访员在采访过程中遇到不寻常的情况时参考定义。此外，问卷还应包含一些对采访员的指导性备注，例如图3.1中的第10个问题就有这种备注。应在采访员手册中提供有关各种专用术语的更详尽阐述和解释。第四章讨论这种手册。

29. 第三项建议是尽可能坚持问题简明扼要，并使用普通的日常词语。另外，要仔细检查所有问题，以确保所提的问题不会“导致”或者有可能诱使受访者做出有倾向性的反应。如果问题很复杂，就把它拆分成两个或更多个独立的问题。让我们举例说明这个要点。假定需要了解一个人在过去七天内是雇佣劳动者还是自营职业者（或两者都是）。如果试图通过一个问题并使用多少令人难懂的技术术语来诱出答案的话，就会这样来提问：

- 过去七天来，您是挣工资或获取其他报酬的雇佣劳动者，还是一个家庭企业的自营职业者？您是同时参加这两类活动呢，还是两种活动都不参加？

这个问题可以用下面两个分开的问题来取代，同时少用技术术语：

1. 过去七天来您是否为某个不是这个家庭成员的人做有报酬的工作？
2. 过去七天来，您是否为自己而工作，比如作为一个农场主、或者一个货物或服务销售商？

图3.1中的问题8、9、10提供了这个要点的另外一个说明。调查设计者也许为了“缩短”问卷会试图把这几个问题合并成一个长的问题，比如：

- 您是否拥有这个住所的法定权利或文书？（如果有的话）这个住户中谁实际拥有这个法定权利？

但这个较长的问题会把许多受访者搞糊涂；果真如此的话，解释问题所花费的时间也许比三个问题分开来提问所花的时间还要长。

30. 第四，要把问卷设计成几乎所有问题的答案都可以用预先编码来表示。此类问题一般叫作“封闭式问题”。比如，受访者对答案为“是”或“否”的问题的回答，可在问卷中记录为“1”（代表“是”）和“2”（代表“否”）。这对采访者比较方便，因为他们只需写一个数字而不必写出完整的单词或短语。⁹更重要的是，这种方法越过了将带有采访员手写的（往往不合格）用一个或多个字作答的问卷交由办公室“编码员”对答案逐一写出数字编码的步骤。这一多余的步骤可能产生更多的差错，但是几乎在所有情况下均可避免。（不过，对于比较复杂的分类——比如职业和行业——编码来说，就需要有实地工作人员所不大可能有的技能和时间，因此建议这些类别应该根据采访员的手写答案由具有专门技能的办公室编码员来编码。）在图3.1中，对各种问题的所有可能答案都进行了预先编码，并且所有编码都在问题的同一页上给出（通常紧接在问题后面）。

⁹ 另一种可供选择的做法，就是允许采访员在预先编码答案旁边的方框中标记一个“X”或者一个钩形符号“√”。

31. 第五项建议与第三项建议有关。不同问题之间的编码设计应该一致。譬如讲，几乎在所有住户调查中都有许多答案为“是”或“否”的问题。问卷中所有这类问题的数字编码应该总是一样的，比如“1”代表“是”，“2”代表“否”。这个（以及其他）编码一经确立，就应该把它用在问卷上的所有“是”或“否”的问题上。这样，采访员就知道了对于所有“是”或“否”的问题来说，他或她应该用编码“1”代表“是”，用编码“2”代表“否”。也可以把这个办法推广到其他类别的答案。不少问卷都有用时间单位或距离来回答的问题。比如，“您上一次看病是什么时候？”或者“最近的公路离您家有多远？”时间单位可以按如下方式编码：1指分钟，2指小时，3指天数，4指周数，等等。于是，回答“10天”就可以用两个数字来记录：“10”和“3”，其中3是时间单位编码。同样，在距离的表述上，编码1可以指米数，编码2可以指公里数。确切的编码设计可以在各种调查之间有所不同；但要点在于，所有需要此种类型编码的问题都要尽可能使用相同的编码设计。¹⁰ 这项建议在图三.1中也有说明。请注意，在该页脚给出的时间单位编码，用于该页的两个问题，即问题5和13。

32. 这里关于编码设计的讨论提出了这样一个问题，就是采访员是应该告诉受访者有关问题的可能的答案呢，还是只读出问题而不读答案编码？一般来说，后一种方法比较好。受访者可能指定第一批答案之一，往往只是因为他们先听到了这些答案——即便后面的某个答案比较准确。另外，如果读出一大堆答案的话，受访者就可能在多种不同的可能答案当中选错了答案。

33. 第六项建议是，调查问卷应该包含一些“跳跃编码”，即指明，根据对前面问题的答案，哪些问题无需再向住户提问。譬如讲，调查中可能包括这样的问题：“过去七天来您找过工作吗？”如果回答说“是”，问卷可能接着问用什么方法找工作；但如果回答是“否”，这样的问题就不相干了。紧接在第一个问题之后应给出简要指示，比如“若回答否定，直接问第六个问题”，这样采访员就不会提问不相干的问题了。此类指示可以采用一些常规表示法，比如，上面的指示可以写成：“若否定，→问题6”。在图三.1中，可采取常规表示法的指示非常简单明了：即在相关答案编码之后用双引号中的数码来表示。比如说，在第一个问题中回答“否”的编码后面用符号“（»12）”来表示：如果对这个问题回答“否”，采访员应直接跳到问题12。

34. 关于格式化问题的最后一个要点是，提问题的方式应该允许受访者用自己的话来回答。这一要点最好通过实例来解释。在关于住房的调查中，可能会有一个关于住户的房租问题。依据不同的租房合同，有些受访者是按每周付一定数额的租金，其他一些受访者则可能按月交房租，甚至有些人按年交房租。这里的要点在于让受访者选择付租金的时间单位，因此提出的问题应该是“您为住所付多少租金？”，而不是“您每月为住所付多少租金？”，后面这种提问的问题在于强迫受访者按照月租回答问题。一位受访者可能清楚地知道他每周付\$50租金，但是如果要求用50美元乘以4.3他就可能算错，因而报告的答案可能就不是正确的租金数（每月217美元）。问卷的设计最好有助于采

¹⁰ 虽然对于时间和距离这样一些简单概念来说，同一国家各次调查的编码有所不同关系不大，但对于比较复杂的概念——比如职业类别或疾病种类——来说，就有充分的理由采用相同的编码设计，以确保不同时期不同调查的可比性。

访员用数字编码记下不同的时间单位，就像图三.1中的第5个问题所说明的那样，譬如讲，每周50美元即可在一个空格内记下50，然后在相邻的空格加上4（周的数字编码）。接下来进行数据分析，研究人员算错的可能性要比受访者小多了，他们可以很容易地将租金数额换算成普通的租金单位，譬如年付租金。

3. 关于问卷设计细节的其他忠告

35. 最后，还可以就问卷的设计提几点较为一般性的忠告。首先，对于某些非常重要的问题，比如住户人口或住户的各种收入来源，提一些“探询”性的问题，或许有助于受访者想起其可能忘记的什么事情。比如说，在取得一份住户成员名单之后，采访员可以提出下面的问题：

- 根据您提供的情况，这一户共有六口人。这个数对不对？家里是否还有别人，比如会不会有人在外面逗留几天或几星期？

36. 第二，问卷应该给每个住户及该住户的每个成员设计一个独特的编码，以便在问卷的所有部分识别这个人。这有助于数据分析员使同一住户和同一个人的信息相一致。几乎在所有情况下都应该是每户一份问卷；在个别情况下，如果一个住户使用两份以上的问卷就必须格外小心，一定要把同一住户的编码写在该户已填完的每份问卷（封面）上。

D. 问卷准备过程

37. 前面的讨论就如何设计住户调查问卷提供了忠告，但几乎只字未提那些亲身参与调查的人以及他们如何检查草拟出来的问卷。本节就问卷的起草、检验和定稿的过程提供一些建议。

1. 组建团队

38. 住户调查几乎总会导致产生大量的、一般来说比原本预想更加复杂的决策和行动。这就意味着，只靠一个人甚至一小组人简直不可能有足够的时间和技能来成功地设计一份住户调查。因此必须从过程一开始就组成一个“专家”团队，以确保调查工作的方方面面无一遗漏。这个团队应该有来自几个关键小组的代表组成。

39. 也许最重要的是团队要有一名或多名决策集团的成员，亦即代表计划利用调查所得信息进行决策的一个或多个集团利益的一位或多位个人。虽然这些人不是技术专家，但是需要由他们来告知（和提醒）团队其他成员调查的最终目标。有了这样一个小组，就能大大加强数据用户与数据生产者之间的沟通。

40. 另一个关键小组由研究人员和数据分析员组成，他们将利用数据中的信息来回答决策者感兴趣的问题。他们的作用是制定数据分析计划，以确保所收集的信息足以回答上述问题。有些情况下，回答决策者问题的工作很简单，但有些情况下又可能相当复杂。

41. 最后但并非最不重要的是数据收集员小组，其中包括采访员、监理员和录入员（含电脑技术员）。这些人通常是正式负责收集数据的工作的人员。他们以往收集住户调查数据的经验不可或缺。他们最清楚哪类问题住户能回答，哪类问题不能回答。这个小组里应该有一个在收集数据过程的数据录入阶段有经验的人。这个人提出的一些简单建议可能会大幅度提高所收集数据的精确度并缩短为分析工作准备数据所需要的时间。

2. 草拟问卷第一稿

42. 几乎任何住户调查问卷的第一稿都是在调查团队成员的一系列会议中产生的。跟任何类的初稿一样，设计的产品会不可避免地存在许多误差。本章提倡的模块法就意味着，第一草稿将由各种不同的模块汇总而成。当把不同的模块放入第一草稿的时候，有几个方面必须经过核查。

43. 首先，调查团队应当核查，这些模块作为一个组合体是否能够收集到所需要的全部信息。有可能某个模块的一个关键问题原本以为已被纳入另一模块了，而实际上并未纳入。有必要召开一次所有模块参与制订者的联席会议，以确保不要在问卷中遗漏一些重要信息。防止重复始终是一个要点。当把所有模块组合起来的时候可能会发现，有的问题在两个模块中问过两次。为了节省受访者和采访者的时间起见，通常这种重复现象应当清除。不应排除重复问题的惟一一种情况，就是当这种问题对某个极为重要的信息起到确认作用的时候，比如某人是否真的是某住户成员的问题。住户成员的年龄可以通过既提问现时年龄又提问出生日期的办法来核对；关于某人是否真的是住户成员的问题，核实的方法是提问：此人过去12个月内是否在别处住过；如果在别处住过，住了几个月（在如此提问之前，先问他或她在接受采访的这个住处住过几个月）。

44. 第二，要核查问卷的整体长度。在任何国家，受访者在回答住户调查的问题上愿意花多少时间都有一定的限度。同时，调查设计者一般都倾向于问大量的问题，因此使得最终的产品比预想的大多了。可以通过实地检验（随后论及）来回答采访一个典型住户需要花多长时间（以及受访者愿意花多少时间接受采访）的问题，但是有经验的采访员和监理员在检查了问卷之后可以告诉团队完成采访过程大约需要多长时间。取消收集“次等”信息的问题是令人遗憾的，但却是拟定任何住户调查问卷初稿的必要步骤。

45. 最后，应该检查问卷初稿在所有追溯期中的一致性。譬如讲，一次调查的一项目标可能是收集住户在过去数月或去年内各种来源的收入信息。需要检查问卷以确保所有收集收入数据的调查组成部分都用同样的追溯期。¹¹ 这一规则的主要例外情况是，偶尔会有前面讲的那种情况，即需要允许受访者在选择对其最容易回想的追溯期方面有一定的灵活性。

¹¹ 有些调查规定了时间参照点，比如在问到5年或10年前境况的时候。这些参照点有时涉及到具体日期、月份或年份，这方面也要检查以确保整个问卷的一致性。

3. 问卷的实地检验和定稿

46. 任何住户调查问卷——不论涉及范围多小、内容多么简单——定稿以前都必须先在少量住户中做试验，以检查问卷设计中存在的问题。几乎不论在任何情况下，一份新的住户问卷经过在有关人口的某些典型住户进行试验后，都会明显暴露出许多误差和欠缺。下面就此提出几点一般性规则；更详尽的处理意见见Grosh和Glewwe（2000年）以及Converse和Presser（1986年）。

47. 问卷草案的实地检验可以分为两个阶段。第一阶段一般叫作预先检测阶段，主要是在少数住户（譬如10-15户）对问卷的选定部分（模块）进行试验，以便对问卷草案的实际效果有个大概的认识。这种检测可以从问卷设计的早期阶段开始多次进行。第二阶段是对问卷草案进行综合性实地检验，一般称之为试点阶段。这是涉及到100-200户的较大型操作。试点住户应该不仅限于一个小地区，而应包括代表有关人口的好几个地区。对于针对城乡地区的调查来说，试点范围必须包括城乡两种地区。还应该在行将使用最终问卷的全国或地区的不同部分进行试点。最后，在住户的选择方面应该做到所有模块都要至少经过50个住户——但最好有50户以上——的检验。这就意味着，譬如讲，如果问卷中有一个收集有关小型家庭企业数据的模块的话，在试点阶段采访的住户中至少要有50户经营这种企业。

48. 大多数试点工作都需要有一两个星期对100-200个住户进行采访的阶段。调查团队的所有成员都要参加试点并尽量多地观摩采访。的确，对于任何在设计住户调查问卷方面经验不多的人来说，试点检测提供了极好的培训机会。试点检测所能提供信息的一个重要方面，就是对完成一户问卷填报所需时间的估计数。¹² 不过应该意识到，试点中所取得的估计数会过高估计（最高可达两成）实际调查中采访一个住户所需花费的时间，这一方面是试点访调员对问卷草案不熟悉，另一方面是因为问卷草案中存在的瑕疵和不完备之处也会影响他们的工作进度，而这方面的问题将在实际调查问卷中得到纠正。

¹² 在进行预先检测和试点检测的过程中，问卷草案应留有空档用以填写完成每个问卷模块的开始和结束的时间，每户采访中都要把这些时间段逐一记录下来。这样即可显示完成每个模块所需要的时间。

49. 另一关键点是，在使用多种语言的国家，要把问卷翻译成所有主要语言，并且用这些语言进行试点检测。这一点极为重要。特别是因为问卷所用的语言跟受访者用的语言不同，采访员不得不在采访过程中将一种语言翻译成另一种语言，要尽可能避免这种做法。研究表明（如Scott等，1988年），与事先将问卷翻译成受访者语言的做法相比，这种现场翻译的做法会使误差增加二至四成。为了检查一种翻译的准确性，一个人或除原译者之外的一组人应将业经翻译的问卷“反译”成原来的语言。然后拿这种反译文同原来语言的内容作比较，以确定该译文是否确切地传达了原问卷的内容；任何差异都说明可能有什么东西“在翻译中丢掉了”。Harkness、Van de Vijver和Mohler（2003年）是有关问卷翻译的一份有用的参考资料。

50. 关于试点检测的最后一个重要方面，就是它不但要检验问卷而且要检验整个实地工作计划，其中包括监理方法、数据录入以及采访手册之类的文字材料（这些都将在第四章中进一步讨论）。只有通过检验全过程，团队才可以

肯定调查实施工作已经准备就绪。一个有益的最后步骤是“快速分析”试点中收集的数据，以检查有可能被忽视的问题。

51. 完成试点检验之后，调查团队应立即召开几天会议，讨论试点成果并根据总结的经验教训修改问卷。前段所讲的试点检验数据的快速分析通常用一些简明扼要的表格来表述，应在会前准备好。有的情况下因为问题太多，可能需要进行第二次试点检验（规模未必同第一次试点一样大），以核实对问卷的较大修改在现场调查中效果如何。所有团队成员都要参加这些会议，并且还要吸收大部分或所有参加了试点检验工作的人参加。

52. 近年来围绕问卷设计问题开展了大量的研究工作，并为编制有效的问卷提供了各种新的重要方法。虽然这些方法尚未在发展中国家和转型期国家推广应用，但将来会明显扩大使用范围。本文对这些方法的介绍有限，但鼓励读者去查阅相关的参考资料。这些方法包括“焦点分组法”、“认知采访法”和“行为编码法”。Esposito和Rothgeb（1997年）及Biemer和Lyberg（2003年）对这些方法提供了很好的总体概述。另外，关于焦点分组，见Krueger和Casey（2000年）；关于认知采访，见Forsyth和Lessler（1991年）；关于行为编码，见Fowler和Cannell（1996年）。本出版物第九章还分别在C.2节和C.6节提供了关于焦点分组和行为编码的详细说明。

E. 结论

53. 本章就发展中国家住户问卷的设计问题提出了若干一般性建议。重点放在由访调员执行的问卷方面。有些住户调查还单独采用“社区问卷”来收集有关当地社区的数据。由于篇幅所限，这种问卷没有纳入本章的讨论。有关社区问卷设计的详细建议，见Frankenberg（2000年）。

54. 本章所涉及的话题确实不少，但每个话题只能作简要处理。凡正在策划此种调查者，务请参阅其他材料以获得更详尽的忠告。本章末提供的参考资料是个良好起点。

参考资料

Biemer, Paul P. 和 Lars E. Lyberg（2003年）。《调查质量简介》。纽约：威利国际出版公司。

Casley, Dennis 和 Denis Lury（1987年）。《发展中国家的数据收集》。联合国，牛津：克拉伦登出版社。

Converse, Jean M. 和 Stanley Presser（1986年）。《调查问题：编制标准的调查表》。加利福尼亚，比佛利山：赛奇出版社。

Deaton, Angus 和 Margaret Grosh（2000年）。“消费”。《设计发展中国家住户调查问卷：15年生活水平衡量研究的教训》，Margaret Grosh 和 Paul Glewwe 等人。纽约：牛津大学出版社（为世界银行）。

- Esposito, James L. 和 Jennifer M. Rothgeb (1997年)。“评估调查数据：从预备调查转变为质量评估”。《调查测量和程序质量》，Lars E. Lyberg 和其他人等。纽约：威利国际出版公司。
- Forsyth, Barbara H. 和 Judith T. Lessler (1991年)。“认知实验方法：分类法”。《调查中的测量误差》，Paul P. Biemer 和其他人等。纽约：威利国际出版公司。
- Fowler, F. J. 和 C. F. Cannell (1996年)。“利用行为守则确认调查问题中的认知错误”。《确认调查研究中的认知和交流过程的方法》。加利福尼亚，旧金山：乔西-巴斯出版社。
- Frankenberg, Elizabeth (2000年)。“社区和价格数据”。《设计发展中国家住户调查问卷：15年生活水平衡量研究的教训》，Margaret Grosh 和 Paul Glewwe 等人。纽约：牛津大学出版社（为世界银行）。
- Grosh, Margaret 和 Paul Glewwe 等 (2000年)。*《设计发展中国家住户调查问卷：15年生活水平衡量研究的教训》*。纽约：牛津大学出版社（为世界银行）。
- Harkness, Janet A.、Fons J.R. Van de Vijver 和 Peter Mohler (2003年)。*《跨文化调查方法》*。纽约：威利国际出版公司。
- Husmanns, R., F. Merhan 和 V. Verma (1990年)。*《从事经济活动人口、就业、失业和未充分就业情况的调查》*。《劳工局概念和方法手册》。日内瓦：国际劳工局。
- Krueger, Richard A. 和 Mary Anne Casey (2000年)。*《焦点小组：应用研究实用指南》*。加利福尼亚，千橡：赛奇出版社。
- Scott, Christopher 和其他人 (1988年)。“逐字问卷调查与实地翻译或表格：实验性研究”。《国际统计评论》，第56期，第3号，第259-278页。
- Sudman, Seymour 和 Norman M. Bradburn (1982年)。*《提问》*。《问卷设计实用指南》，加利福尼亚，旧金山：乔西-巴斯出版社。
- 联合国 (1985年)。*《联合国国家住户调查能力方案：调查问卷的编制和设计》* (INT-84-014)。纽约。
- 联合国 (1993年)。*《国家住户调查能力方案：稀少人口和难以定义人口的抽样》* (INT-92-P80-16E)。纽约。

第四章

发展中国家住户调查的实施概述

Paul Glewwe

明尼苏达州立大学应用经济学系
美利坚合众国明尼苏达州圣保罗

摘 要

本章评述有关发展中国家实施住户调查的基本问题，首先从现场调查前必须开展的各种活动入手，即：编制预算和工作计划；提取样本；训练调查工作人员和编写培训手册；以及拟定实地工作计划。此外还包括开展现场调查期间进行的各项活动：建立并保持完备的通讯和交通手段；订立监理议定书和旨在提高数据质量的其他活动；以及发展一个数据管理系统。本章末尾有一小节关于实地工作完成后的一些活动。最后是简要结论。

关键词：调查的实施、预算、工作计划、样本、培训、实地工作计划、通讯、交通、监理、数据管理。

A. 引言

1. 住户调查提供的信息价值高低在很大程度上取决于调查所收集数据的用途和精确度，而后者又反过来取决于调查是如何在现场实际执行的。本章就调查的实施问题提供一般性建议，其中除问卷设计外还包括执行住户调查全过程的几乎所有方面。

2. 人们可以把设计完好的住户调查问卷（及相关的数据分析计划）视为通往调查成功之路的中途点。必须通过有效实施调查才能抵达终点。有效实施调查的过程并非始于采访员着手访问指定给他们的住户，而是始于数月——且往往是在一两年——以前。本章B节讨论采访任何住户以前必须进行的一些活动；C节描述在现场调查的同时进行的一些活动；D节围绕实地工作结束后必须完成的工作提供了简短讨论；最后一节提出几点简要结论。虽然本章对此话题作了有益的介绍，但是过于简要，远不能详细提供所有必要的忠告。为了确保调查实现既定的目标，调查负责人应当参阅比本文详尽得多的各种相关论述。Grosh和Muñoz的文章（1996年）是个良好的起点：虽然它集中于世界银行主办的“生活水平衡量研究”（LSMS）相关调查，但其中大部分忠告适用于

几乎所有类别的住户调查。另外两份有用的参考资料是Casley和Lury的文章（1987年）和联合国资料（1984年）。

3. 本章从头至尾假定调查系由为此目的专门任命的一个组织健全的“核心”团队负责策划和实施。同时假定，调查问卷由走访住户家庭的采访员执行；抽样单位是住户。¹³最后请读者注意，本章的关注重点是发展中国家以及像中国和越南等低收入的转型经济体。尽管如此，大多数建议也适用于东欧和前苏联的较发达转型经济体。

¹³ 在有些调查中，抽样单位是住所而不是住户；但在此种情况下，被抽样住所中的某些或全部住户各自成为调查的“报告单位”。

B. 进入现场调查前的活动

4. 任何住户调查的第一项任务，都是组建一个全面管理调查工作的核心团队。第三章详细解释了该团队都包括哪些人。核心团队建成后，必须首先完成如下八项任务，才能着手住户采访工作：

- (a) 草拟临时预算，确保资金到位；
- (b) 制定所有其余活动的工作计划；
- (c) 提取拟采访住户的样本；
- (d) 编写培训手册；
- (e) 培训外勤和数据录入工作人员；
- (f) 编制实地工作和数据录入计划；
- (g) 进行试点检测；
- (h) 开展宣传运动。

以上任务清单大体上是按时间顺序排列的。现将各项任务分述如下。

1. 预算筹资

5. 对于几乎任何一项住户调查来说，财务资源都是哪些事情能办到的一个严重制约因素。这一制约因素所隐含的限制条件倒未必是显而易见的。几乎所有调查的首要任务都是编制一份预算草案，这项任务依据的假设是待抽样住户的数量是多少，以及工作人员调查一个典型住户需要多少时间。这个预算将是约略的，因为直到编制出详细问卷以前不可能知道某些成本细节。但在大多数情况下，预算草案应该与最终预算有一定的近似性（除非调查目标发生重大改变）。

6. 编制了预算草案之后，必须筹集到所需的资金。如果资金来源没有着落，恐怕调查的详细规划工作就得推迟进行，直到资金来源有了保障。这样可以避免一旦找不到资金来源的情况下浪费工时。

7. 尽管在没有掌握有关调查性质和类型的进一步信息的情况下很难在确定预算方面讲更多的话，但可以提几点一般性建议。首先应该对负责实施调查的组织能力进行评估。如果该组织缺乏某些技术能力——譬如讲，如果它基本上没有提取样本的技术专长，或者它的特点是缺乏利用新型信息技术的专长

长，那么也许就有必要外聘一些顾问。这会大幅增加调查成本，但无论如何这方面的额外费用还是值得的。第二，一个起步的好办法就是看一看该国或类似国家已经进行过的一些类似调查的预算资料。第三，为了避免意外费用带来的压力，应该明确补充一个大约相当于总预算额10%的“缓冲”预算，列为预算内的一个追加项目。这个项目通常叫作意外开支准备金。在成本极不确定的情况下，可能需要设定15%-20%的应急额度。

8. 为了使上述讨论更具体些，表四.1[Grosh和Muñoz所供资料（1996年）中的表8.2修改版本]提供了一个虚拟调查预算草案。在这个实例中，假定调查拟采访3 000个住户，数据收集为期一年。除了核心团队（见第三章）以外，还有四个外勤团队，每个团队有三名采访员、一名监理人和一名数据录入操作员。项目专用车辆的两名司机负责把这些团队送到工作地点。假定每位采访员在一年期内工作250天，（平均）每天采访一个住户。表四.1列出了所有人员的假定工资以及假设发给每位实地工作队员的每日“差旅津贴”。每个外勤团队配一台数据录入计算机，核心调查团队有三台数据分析计算机。另外还给出了（国内外）聘请顾问方面的假定成本。当然，该表仅为解说目的而提供；任何特定调查的成本都取决于样本大小、雇员数量、雇员薪酬、监理人与采访员的比率、每位采访员每天所能采访的户数、数据是在现场录入还是集中在一个地点录入，以及其他诸多因素。该表在这里权当一份“核对表”使用，以确保调查预算草案包含所有基本费用。

表四.1

3 000住户虚拟调查预算草案（货币单位：美元）

项 目	数 量	工作期限	单位成本	总 成 本
基本工资				
项目管理人	1	30个月	800/月	24 000
数据管理人	1	30个月	600/月	18 000
实地工作管理人	1	30个月	600/月	18 000
助理/会计	3	24个月	450/月	32 400
监理人	4	14个月	400/月	22 400
采访员	12	13个月	350/月	54 600
数据录入操作员	4	13个月	300/月	15 600
司机	2	13个月	300/月	7 800
小 计				192 800
差旅津贴				
项目管理人	1	90天	30/日	2 700
数据管理人	1	60天	30/日	1 800
实地工作管理人	1	90天	30/日	2 700
助理	2	60天	30/日	3 600
登记人员	10	60天	15/日	9 000
监理人	4	290天	15/日	17 400
采访员	12	270天	15/日	48 600

项 目	数 量	工作期限	单位成本	总 成 本
司机	2	270天	15/日	8 100
小 计				93 900
器材成本				
购车	2	-	20 000	40 000
燃料和维修	2	13个月	300/月	7 800
数据录入计算机	4	-	1 000	4 000
打印机、稳定剂, 等等	5	-	1 000	5 000
数据分析计算机	3	-	1 500	4 500
计算器/办公用品	-	30个月	350/月	10 500
复印机/传真	1	-	2 500	2 500
小 计				74 300
印刷成本				
问卷	3 500	-	2	7 000
培训手册	40	-	5	200
报告	500	-	5	2 500
小 计				9 700
顾问成本				
外国顾问	5	人-月	10 000/月	50 000
国际每日津贴	150	天数	150/日	22 500
国际差旅津贴	8	出差次数	2 000/次	16 000
本地顾问	5	人-月	3 000/月	15 000
小 计				103 500
应急准备金 (10%)				47 400
共 计				521 600

注：连字符(-)表示该项目不适用。

2. 工作计划

9. 资金来源落实后, 下一项工作就是拟定一项切实可行的工作计划, 这基本上就是一个从调查规划初期直到实地工作结束后的活动时间表。¹⁴ 工作计划包括如下每一种活动: 全盘管理(包括设备采购); 编拟问卷; 提取样本; 工作人员的指定、聘用和培训; 数据录入和数据管理; 现场调查活动; 以及数据分析、处理、编制文件和编写报告。上述每个具体领域都要有一个需要完成的任务清单和完成的日期(或曰期限)。要突出重要里程碑事件, 比如试点检测、实地工作启动日等。一般可用图表展示的这份任务清单。

10. 不用说, 上述诸多活动都是相互关联的, 因此必须加以协调。譬如讲, 在完成采购必要的设备、且指定(或聘用)并培训了执行任务的工作人员以前, 许多数据管理和数据分析活动就没办法开展。还必须牢记在心的是, 当发生了意外事件的时候, 就连最好的计划也不得不改变。回想起来, 多数计划

¹⁴ 这是一项总的工作计划, 包括实地工作开始以前(即采访任何住户以前)必须执行的许多工作任务。下文讲到, 还需要一份比较具体的“实地工作和数据录入计划”。

都过于理想化，以至于期限延后的情况屡见不鲜。各种活动时间表都要尽可能实事求是，并应包括某些“停工时间”，以便参与者能在发生不可避免的延误事件之后赶上进度。

11. 图四.1[根据Grosh和Muñoz（1996年）所供资料中的图8.1改编]是一个工作计划实例。该计划涵盖30个月的工作。标注星花(*)的地方表示有不同的活动在进行。从该图可以看出，在开始实地调查一年前就开始了准备工作。第八个月开始进行试点检测意味着，到这个月份以前问卷草案、人员培训和数据录入方案草稿均已准备就绪。实际实地工作定于第12个月份启动，预计

图四.1

制定和实施住户调查的工作计划

任 务	调查月数																													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
管理和后勤																														
任命核心调查团队	*																													
采购计算机		*					*	*	*	*																				
采购调查器材							*	*	*																					
宣传工作							*	*	*	*	*																			
采购/租赁车辆						*	*	*	*	*																				
问卷研拟																														
确定调查目标		*	*																											
筹备问卷草案			*	*	*																									
问卷起草会议							*	*																						
试点检测问卷草案定稿								*																						
试点检测									*																					
试点检测后的会议										*																				
印制问卷定稿											*																			
抽样																														
确定样本设计和范围		*	*																											
提取样本（初级抽样单位）					*																									
确定实地工作计划				*																										
初级抽样单位列表和地图标示				*	*	*	*																							
人员配置和培训																														
选择和培训试点检测工作人员							*	*																						
编写培训手册										*	*																			
采访员培训											*																			
数据管理																														
设计数据录入方案初稿							*	*	*																					
数据录入方案定稿										*	*																			
编写数据录入手册										*	*																			
培训数据录入员											*																			
实地工作												*	*	*	*	*	*	*	*	*	*	*	*	*	*					
分析和文件编制																														
分析计划草案																			*	*										
分析前一半数据																				*	*									
编写初步报告																						*								
创建首批完整数据集																								*	*					
初步数据分析																										*	*			
最终报告和文件编制																												*	*	*

为期一年。该工作计划还设想，当数据收集到一半的时候将着手起草报告。当然，任何特定调查的工作计划都会跟这份计划有所不同。这份计划草稿只是起到“核对表”的作用，用以说明各项不同的活动应在时间安排上相互协调。

3. 提取住户样本

12. 几乎在所有住户调查中都有一个关注人口，比如在调查中代表整个国家人口的住户。选择一套代表较大人口的住户的过程叫作抽样，而进行抽样的程序叫作抽样设计。在提取样本的时候需要考虑的问题很多——甚至多到像本章这样简短的概述所不可能尽数列举的地步。关于抽样的详细建议，见本书第二章、第五章和第六章。Kalton（1983年）介绍了抽样工作；另外可从Kish（1965年）、Cochran（1977年）和Lohr（1999年）的文章中找到更加全面的相关论述。

13. 本章有关抽样的讨论仅限于为调查团队提两点值得谨记在心的意见。首先，有时候将样本设计成适合进行为期12个月的住户采访是有益的。这种设计取调查现象的季节变异平均数，同时也使得数据可用于研究季节格局。第二，也许更重要的是，调查的策划者应当避免试图获取住户数量过大的样本。他们想要扩大样本规模——尤其是特别关注群体的样本规模——的愿望是很自然的，因为这样做可以减少调查中的取样误差。但是在许多情况下，随着样本规模的扩大，“漏查”（即“取样遗漏”）失误增多，这是由雇用的工作人员不太合格以及监理人与采访员的比率下降造成的。一种很有可能出现的情况是，因扩大样本规模所减少的抽样误差会被漏查失误所抵消。

4. 编写培训手册

14. 为所有参加培训的人——采访员、监理人和数据录入员——编写培训手册，大概是培训工作的一个最重要环节。分别为每一种人员准备手册是必要的，就是说，要有采访员手册、监理人手册和数据录入员手册。这些手册是培训的关键组成部分，必须在培训开始前准备好。更重要的是，在调查实施过程中，这些手册可以当作参考材料使用，故应包含各种不同类型的外勤和数据录入工作人员所需要的全部信息。¹⁵实际上，数据分析员也经常使用培训手册，以便更好地理解他们所分析的数据；这就意味着，所有手册都要额外增加一些印数，以满足这些分析员的需要。作为一条通用规则，每当发生疑问的时候，最好把相关的材料加入手册，而不要把它忽略掉。

¹⁵ 团队“外勤工作人员”系指采访员、监理人和其他工作人员，他们为了完成本职工作经常游走于采访住户的社区。如下文所讨论的，让数据录入员尽量在接近这些社区的地方工作有好处。在数据录入员跟随外勤工作人员旅行的调查中，也可以把他们叫作外勤工作人员，但在其他调查中他们不是外勤人员。本章使用“外勤和数据录入工作人员”这个短语是为了两种可能性都照顾到。

15. 培训手册要解释调查的目的和手册适用人员须执行的基本任务。还应提供处理异常情况所须遵循的程序，其中包括适用于处理意外问题的一般原则。手册还应解释如何填写作为份内工作必须填报的任何一种表格（这对监理人手册尤为重要）。鉴于即使准备最完备的手册也可能会有差错或遗漏，故应视需要准备一套或数套“补充教材”，以便在将手册发给外勤和数据录入工作人员之后对这些手册加以补充。

5. 外勤和数据录入工作人员的培训

16. 有些情况下，进行调查的组织会有为数众多的有经验的采访员、监理人和数据录入员。当新调查非常类似于该组织曾经做过的历次调查的时候，基本上不需要花多少时间进行新的培训，只需一两周的时间对新的问卷细节以及可能伴随新调查而发生的程序上的一些变化进行解释。可是有些情况下，新调查可能与该组织近期内所做过的任何调查都有相当大的差异，而且多数情况下各组织都需要至少雇用一些新的外勤和数据录入工作人员。在这些情况下，为了确保高质量的调查，必须进行通盘培训。譬如讲，新聘用的采访员和监理人必须经过一般培训才能接受有关调查的有针对性培训。一般来说，这些情况需要有两周以上的培训：通常需要三到四周才能保证采访员和监理人做好有效开展工作的准备。

17. 虽然培训的性质是由调查性质决定的，但是仍然可以就此提几点一般性意见。首先，培训要利用问卷包括大量采访实际住户的演练。第二，培训要强调理解调查的目标以及如何收集数据才有助于实现这些目标。集中掌握这方面的知识，而不是训练外勤和数据录入工作人员不容置疑地循规蹈矩，会有利于采访员和监理人恰当地应付没有预料到的问题和麻烦事。第三，最好比实际需要量多培训一些人并对受训者进行某种考核（包括笔试和实际采访部分）。可根据考核结果选拔那些考试成绩较高的人担任采访员和监理人。第四，要集中在一个地点进行培训，以保证所有外勤工作人员都能接受同样的培训，并且使培训本身达到最高质量。最后要认识到，培训质量可以对调查质量并最终对数据收集质量产生关键性的影响，这种认识很重要。整个调查团队必须充分重视培训工作，而不只是把它简单地委托一两成员去做。

6. 实地工作和数据录入计划

18. 外出到抽样地区对抽样住户进行采访的实际工作，一般称之为实地工作。鉴于实地工作和数据录入工作是密切配合的，所以本章将两者一起来讨论。实地工作应在培训完成后尽快启动（甚至在一周之内），以便最大限度地避免忘记培训中所学到的知识。在开始实地工作以前，务必（根据抽样计划）制定出选定的住户与行将从事这项工作的采访员、监理人和数据录入员相互匹配的非常详尽的工作计划。通常把调查工作人员组成几个团队，由一位监理人领导。每个团队被指定负责整个样本的一部分，并负责确保其指定部分的所有住户都采访到。

19. 在拟订实地工作计划的时候，有几条基本原则应谨记在心。首先，必须提供适当的交通运输手段，不仅为工作人员，也为物资供应。许多国家的住户调查经验表明，最常见的后勤问题是要保障外勤人员所用车辆的燃料、油料供应及完善的维修保养。第二，实地工作计划需要讲求实际，意思就是说，它应该以该国以往的住户调查经验为基础。如果要尝试某种新路子的话，所采用的方法应该作为试点检测的一部分先行验证（关于试点检测，见第三章）。第三，实地工作计划应有一个数据录入计划相配套，后者解释如何将填写完毕的

¹⁶ 这种短暂的停顿应选在一个“普通”时段，以免在调查遇到重要事件的时候中断数据收集工作。

问卷所产生的信息录入计算机并最终纳入中央办公室的主文档。第四，对于为期数月的现场调查，应在最初运转几个星期之后暂停一下，以评估实地工作和数据录入工作进展顺利与否。¹⁶ 很有可能根据头几个星期的经验，建议对实地工作和数据录入的程序作一些修改；这些改变应写成文字并提供给实地工作人员，像前面所讲的那样将其作为“增补内容”纳入他们的手册。第五，实地工作计划定稿之前应交给经验丰富的监理人和采访员查看，以征求他们的意见和建议。最后，要让采访员在每个初级抽样单位有足够的时间反复拜访被抽样的住户，以便从最有见识的受访者收集数据；假如不这么做，而由住户的另一位不太了解情况的成员“代答”，则如此收集的数据很可能会减损数据收集的准确性。

7. 进行试点检测

20. 所有住户调查都要进行一次对问卷设计、实地工作和数据录入计划以及所有其他相关方面的“检测”。这叫作试点检测。其中涉及到对调查所涵盖的全国各地100-200个住户进行采访。因为试点检测的主要目标之一是评价问卷的设计，所以这个问题在第三章进行了详细讨论。试点检测结束后开几天会，由核心调查团队和参加试点检测的人讨论试点期间发现的任何问题。然后与会者必须就问卷的最终定稿、最终工作计划和数据录入计划定稿、调查的任何其他方面达成一致。

8. 发起宣传运动

21. 住户调查应通过大众传媒公开宣布启动新一轮的住户调查，以提高公众对调查的认识，并有望鼓励被选中受访的住户积极合作。宣传运动的另一个好处是提高调查人员的士气。一般来说，不宜在一般宣传方面投入大量资金，因为得到相关信息的绝大多数住户并不是调查的受访者。不过，在不少情况下，通过与电视台和无线电广播电台、报纸和其他大众传媒组织联系，几乎无须任何费用即可达到宣传目的。报纸上的报道特别有用，因为采访员和监理人可将有关报道的复印件保留下来，给心存疑虑的住户看报纸上访调员是怎么说的。

22. 较有针对性的宣传也很有用，其中包括在选中为初级抽样单位的社区张贴传单，给选定受访的各住户发送公开信，等等。张贴的传单要色彩纷呈、引人注目，而不论传单还是公开信都要突出调查数据对改进政府政策的好处。信中还要强调数据是严格保密的；在许多国家可以援引特定法律条文作为机密保障。最后，要同地方社区领导接触，以解释调查的重要意义和利好。在说服地方领导人确信调查利好之后，他们就可以说服不情愿的住户参加调查。

C. 现场调查过程中开展的各项活动

23. 在所有这些准备活动完成以后，实际住户调查就开始了。每个国家实施住户调查的路子多少有所不同，但可以提出一些适用于所有国家的一般性忠告（直接参见下文）。本文假定实地工作是由巡回团队执行的。

1. 通讯和交通手段

24. 每个现场调查团队都需要有可靠的通讯线路通达中央调查管理机构，以便报告工作进展和问题，并尽快向中央办公室提供调查数据。发展中国家往往通讯能力薄弱，尤其是农村地区。但在大多数国家，电话业务已经改进到每个现场调查团队可在数小时、或最多一两天的路程内用上可靠的电话。实际上，移动电话现已成为许多发展中国家非常普通的通讯工具，尽管在农村地区还不是很普及。一个可供选择的简单办法，就是给那些在这种通讯技术覆盖地区工作的团队配置移动电话。对于在边远地区工作的团队，配置卫星电话设备或许是一项划得来的投资。

25. 对现场调查工作团队来说，运输问题也至关重要。各国在这方面的做法不尽相同，但至少每个团队都应该有可靠的运输保障，以便从一个工作地区转移到另一个地区。现场调查队员一旦得了重病需要立即救治，对于这种情况也必须有应急运输计划。不论正常运输还是紧急运输都要有某种后援系统的计划，以便在基本系统失灵的情况下派上用场。可靠的运输手段可在所有其他通讯手段都无效的情况下当作一种后备通讯方法来使用。

2. 监督和质量保障

26. 采访员的工作质量对任何住户调查来说都具有关键性的重要意义。有些采访员简直就没有工作能力，其他人在缺乏或根本没有奖励措施的情况也不可能全力以赴地投入工作。建立一个行之有效的实地工作监督制度是保证工作质量的关键所在。

27. 下面的建议会有助于监理人有效地监督和保证采访工作质量。首先，每位监理人应负责管理少数几位采访员：最多不超过五人，少到两三人即可。第二，每位监理人至少要拿出一半的时间用来检查采访员的工作质量。第三，要编制一份比较简短的检查对照表，监理人可用此表来检查填写完的问卷，以保证按照一些基本规则完成对每个受访住户的采访。每份调查问卷都要按照对照表上列出的条目逐项进行检查，并保留这些检查的文字记录。第四，监理人应在不预先通知的情况下巡视采访员的实地工作情况。这能确保采访员在预期的地方工作。此外，监理人要观察采访员如何进行住户采访，以查明采访员是否依照培训中教给他们的程序办事。第五，监理人要随机选择一些采访过的住户进行回访。需要另外准备一份较为详细的对照表，用来进行一种仅涉及一些关键要点的“微型采访”（比如某住户实际有几个常住人口），以确保采访员正确记录问卷上的最基本信息。第六，考虑到调查团队的流动性，实地工作计划在组织上应该使监理人能够伴随采访员从一个地方到另一个地方去完成他们的采访任务；如果监理人远离采访员，就很难开展监督工作。

28. 另外可就监督和数据保障提两点建议。首先应认真考虑使用笔记本电脑进行现场数据录入，并采用软件检查录入数据内部不一致的情况。这样，就可以在发现任何不一致的时候，让采访员返回有关住户那里收集正确信息。¹⁷

¹⁷ 现场使用笔记本电脑未必是件轻而易举的事。所涉及的问题包括缺少电源；计算机因灰尘、炎热和湿度大发生故障；当然，还有大量购置这种计算机的高成本问题。

第二，核心调查团队成员应在不事先通知的情况下巡视各现场调查工作队。这种巡视基本上是对监理人实施监督的手段，监理人的工作也需要督促检查。

3. 数据管理

29. 录入数据并将其处理成适合数据分析的形式，是任何调查的一项至关重要的工作。现在大多使用个人计算机和数据录入软件进行数据录入的工作。需要把软件设计成可用来检查数据的逻辑一致性。如果发现有不一致的地方，起码可以核查数据录入员的工作，以确定这是否只是单纯的数据录入差错问题。如果是在现场进行数据录入的话，就可能引进一种更好的制度——即采访员可以返回有关住户去纠正不一致的地方；但如果是在调查组织的中央总部进行数据录入的话，这几乎是不可能的事。

30. 数据管理系统的运行必须确保数据能尽快到达一个集中地点。这项原则的重要性有两个突出理由。首先，应该及时对第一周或第一个月的工作进行检查，以确保发送到中央办公室的数据没有严重问题。其次，几乎在任何情况下，信息送达分析员和决策人手中越快，信息价值就越高。

31. 在数据管理问题上还可以提几点具体建议。首先，要保有对所有抽样住户的完整调查结果报告，包括：受访者人数；不答复者人数；或不合格的单位。这种信息是为分析的目的对受访者数据记录加权所需要的。第二，数据录入软件程序在使用前必须经过彻底验证，而问卷试点检测期间就是验证这种软件的极好时机。第三，在向研究人员和数据分析员提供数据之前，要对数据集的各个部分逐一进行检查，以确保没有任何住户被错误地排除在外或被重复取样。第四，要向数据分析员提供一份“基本信息”文件，以保证他们了解如何使用相关数据。下面D节对此有进一步解释。

D. 实地工作、数据录入和数据处理之后需要开展的活动

32. 采访工作全部结束后还需要进一步开展几项活动才能圆满完成住户调查。这些活动通常都在收集数据的组织所在的中央总部进行。最突出的任务是数据分析，这方面的工作在本出版物的其他章节专门讨论，但是还需要进行其他一些后期工作。

1. 听取汇报

33. 所有监理人以及（可能的话）所有采访员和数据录入员要参加一次与核心调查团队一起召开的会议，来共同讨论调查期间遇到的问题，关于在未来的调查中避免类似问题的意见，并且——较普遍的做法是——就改进调查提供各种建议。这种会议应在完成调查后立即举行，以免现场和数据录入工作人员忘记有关的经验细节。要详细记录提出的各项建议，以便将其纳入下一次同类调查的计划工作。

2. 最终数据集的制备和文件编制工作

34. 几乎任何住户调查所收集的数据都可能管用好多年，不论收集数据的机构还是其他研究机构（或研究人员）都可能在日后的岁月就此提出各种报告和分析。为了避免混淆，应该编制一套最终的“正式”数据版本，并把它作为一切组织和个人进行一切分析的依据。最好能在收集完数据之后的一至三个月内拿出最终数据版本。因此，现场收集的数据必须经过严格核查和分析，以便找出任何差错或异常并加以定位和做出标记。当然，有些差错得经过数月乃至多年才会被发现，针对这种情况可以编制一个“修订版本”的数据集，供日后分析使用。

35. 任何数据分析都会有许多与数据相关的问题。这些问题可能包括从数据文档是如何建立的这种普通问题，具体到有关数据是如何收集的这样一些重要得多的问题。为了避免应接不暇的澄清要求占用工作人员的大量时间，收集数据的机构应该编制一份文件来解释数据是如何收集的以及数据文档是如何安排和格式化的。这种文件将载有对调查问卷中所未见的任何编码所作的说明，以及对任何偏离原定计划收集数据的情况的解释。该文件最好能说明最终样本和计划样本有何不同；换言之，有多少住户没有找到或拒绝参加调查，以及（如果适用的话）如何选定了新的住户来取代未接受采访的住户。除了这份文件外，供数据分析员使用的任何信息“包”还应包括一份问卷影印本和全套培训手册。

36. 在许多国家有关文件编制工作的最后一个问题，是将相关文件资料翻译成其他语言的问题。当前有不少研究人员所研究的国家，其语言他们看不懂，只能依靠翻译的问卷和其他文件。通常，与其让诸多不同的研究人员去做自己的、或许不够准确的翻译工作，不如把所有数据分析所需的材料翻译成某种通用的国际语言，最通用的就是英语（其他可能的语种是法语和西班牙语），这不失为一个好办法。虽然这或多或少是一项负担，但或许可以将这种翻译成本纳入初始调查预算，吁请捐款人特别为此目的提供资金。

3. 数据分析

37. 所有数据都是为分析目的而收集的，因此不言而喻，数据分析是数据收集之后的最终活动。这里需要指出的惟一要点，就是整体调查计划应对数据分析所需要的时间量进行实际估算，并将这方面的估计数纳入整个调查活动时间表。数据分析所花费的时间几乎总要比计划的时间长，但是，调查团队与数据分析人员协商越密切，依据这些数据进行分析所取得的结果可能就越精确，也就越有用。

E. 结论

38. 本章就发展中国家的住户调查提供了一般性建议。讨论涉及到许多议题，但每个议题只能作简要处理——从住户调查工作的复杂性来说，这是不可避免的。鉴于本章所提供的信息不能完全满足住户调查的需要，所以任何住户调查策划者都需要参考其他材料以获取更详尽的忠告。他们应该阅读本章导

言中列举的那些参考资料；另外，与进行过住户调查的有关个人和组织商讨本国以往相关调查的经验，无疑是个好做法。开展调查可能是一项单调乏味的工作，但如果精心工作，注意细微末节，并遵从本章忠告，就可以明显提高数据收集的质量和效用。

参考资料

Casley, Dennis和Denis Lury (1987年)。《发展中国家的数据收集》。联合王国，牛津：克来伦登出版社。

Cochran, William (1977年)。《抽样方法》。第三版。纽约：威利国际出版公司。

Grosh, Margaret和Juan Muñoz (1996年)。《规划和实施生活平衡量研究调查手册》。《生活平衡量研究工作文件，第126号》。哥伦比亚特区华盛顿：世界银行。

Kalton, Graham (1983年)。《抽样调查简介》。加利福尼亚，比佛利山：赛奇出版社。

Kish, Leslie (1965年)。《抽样调查》。纽约：威利国际出版公司。

Lohr, Sharon (1999年)。《抽样：设计与分析》。加利福尼亚，太平洋丛林：达克斯伯利出版社。

联合国 (1984年)。《住户调查手册 (修订版)》。方法研究，第31号。出售品编号：E. 83. XVII. 13。

B 部 分

样本设计

导 言

Vijay Verma

锡耶纳大学
意大利锡耶纳市

1. 本出版物第一部分综合介绍了有关住户调查设计与实施的主要技术问题。除了问卷设计之外，还概述了实际调查和样本设计的问题。本部分以较为具体的方式有选择地探讨了发展中国家和转型期国家的住户调查样本设计问题。其中包括三章：一章关于住户调查的主抽样标架和主样本的设计问题；另外两章关于设计效果的估量及其在样本设计中的应用问题。

2. 样本调查的目的，是根据对某个调查人口的有限数量单位（即一个样本）的观测结果，对该人口的总体适用性做出估量和推断。在这一过程中可能会出现各种原因造成的误差。通常对抽样误差和非抽样误差加以区分。可是从整个调查过程的观点来看，一种更基本的分类是区分“计量误差”和“估计误差”。计量误差是在对调查所包含的单位进行计量的结果偏离这些单位的实际（或真实）数值的情况下产生的，这种误差关系到调查所计数的个体单位层级的计量精确度，是围绕调查的实质内容发生的误差。此类误差有别于估计误差，后者是在从特定计数单位到需要估计数或推断的整个调研人口的外推过程中发生的。估计误差关系到从观测单位到目标人口的普遍化，是围绕样本设计与实施发生的误差。此类误差除了抽样的变异性之外还包括与样本选择和调查实施有关的各种偏差，比如覆盖面的误差和受访者不响应造成的误差。所有这些误差都是抽样统计人员所关切的基本问题。往往几次调查或几轮调查共用同样的抽样标架、主样本、样本设计、有时甚至同样的单位样本。在这种情况下，与抽样过程有关的误差在这些调查中带有共性，而较少取决于调查标的物。

3. 正是基于对计量与估计之分的了解才选择了本部分所讨论的问题。本出版物第二部分的各章讨论了关于估计的两个重要方面：一个方面是抽样标架，这个问题关系到所关注的人口的覆盖状况是否良好，并影响到所可能构建的抽样设计的成本效益；另一个方面是设计效果，这种效应提供了调查效率的定量计量，并有助于将设计结构与调查成本挂钩。当然，关于调查设计还有其他方面需要讨论，因此在探讨本部分章节的同时参考前一部分建立的框架是有益的，特别是第二章有关样本设计原则和方法的讨论。

4. 第五章非常务实而详尽地讨论了主样本和主抽样标架的概念。关于样本结果普遍化所及的人口定义，是调查规划与设计的一个基本方面。因此，行将调查的人口必须按照实际形态来表述，并据此选择所需要的样本类型。抽样标架就是这样一种表述方式。在最简单的情况下，抽样标架只不过是该人口中所有抽样单位的一个明晰清单；在比较复杂的设计中，抽样标架的表述也许是部分隐含的，但仍能说明所有抽样单位。实际上，所需的抽样标架是结合所需的样本结构及样本选择程序确定的。在多级抽样标架中（住户调查的抽样标架多以地区为基础），抽样标架的耐久性随着调查单位层次的降低而下滑。一端是初级抽样标架，它代表着有长远利用价值的重大投资；另一端是最终抽样单位清单（包括诸如地址、住户、特别是个人等），这种抽样单位则需要经常更新。

5. 初级抽样标架（叫作“初级抽样标架”）必须涵盖初级抽样单位（PSU）的整个人口。继第一阶段选择之后，任何较低层次的单位清单只须在前一阶段较高层次的选定单位范围之内即可。为了经济、方便起见，这项任务的一个或多个阶段可以合并或在许多调查之间共用。这些共用阶段所产生的样本叫作“主样本”。目标是提供最低到某一阶段的共用单位样本；从这一阶段开始，可以为个别调查进行进一步的抽样。主样本的利用包括如下目标：

- (a) 通过不同调查之间共享样本，节省开发并保持抽样标架和材料的成本；
- (b) 减少样本设计和选择成本；
- (c) 简化提取个体样本的技术过程；
- (d) 便利不同调查、特别是持续调查的连续轮次之间的实际运作上的联系；
- (e) 便利并在必要时限制和控制各种调查在同一抽样标架内提取多种样本。

6. 另外还必须认识到，就是在实际当中主样本也有其局限性：

- (a) 如果主样本的概念不能扩展到较低层次抽样阶段的话，节省的成本会很少，因为较低层次的单位比较少，而相应的抽样标架或清单需要经常更新；
- (b) 只有在主样本不止用于一项调查、最好是用于许多调查的情况下，才能适当节省成本；
- (c) 一个主样本的有效利用需要做长期规划，而在发展中国家的境况下这是难以做到的；
- (d) 如果在设计个别调查的时候缺乏适应共用主样本的灵活性，就可能成问题；
- (e) 可能在提取个别样本方面会增加技术上的复杂性；无论如何都需要详尽而准确地保有主样本的文件编制能力。

7. 将主样本的概念扩展到不包括初级抽样单位的样本而是包括初级抽样单位的整个人口，这也是有可能的。这就是第五章讨论的主抽样标架的概念。当现有的抽样标架不能充分涵盖关注人口而且/或者不包含有关高效而便捷地选择样本的信息时，在主抽样标架方面投资是值得的。使用主抽样标架也可以减少从限制性较大的主样本在选择样本的类型和规模上所受到的制约。

8. 第六章和第七章处理设计效果的重要概念。设计效果（或其平方根，有时叫作设计因子）是有关设计中的各种复杂因素对一个估计数方差影响的一种综合性概括度量法。对于某个特定统计量来说，设计效果是作为该统计量在实际设计条件下的方差与同样规模的简单随机样本（SRS）条件下的方差之比来计算。这样，它就提供了一种设计效率的度量方法。通过取实际方差与SRS方差之比作度量值，设计效果还排除了两者共同因素的影响，比如估计数大小和计量比例，人口变异和总体样本规模。这就使得此种度量法更具有从一种情况（调查、设计）转移到另一种情况的“便携性”。设计效果的这两个特点——即作为对设计效率的概括度量和便携式度量——使得这种度量法极具实用价值并在实际调查工作中广为采用。对许多统计量和多种多样的分组人口的估计数来说，设计效果的计算与分析对于评价当前的设计和新样本的设计是非常宝贵的。

9. 虽然这种度量法确实排除了前述抽样误差量值中的某些重要变异来源，但是设计效果量值依然取决于其他设计特征，比如在样本地区选择住户或个人的数量和方式。最重要的是务必牢记，设计效果是依变量和相关统计量而异的。绝不能仅靠一种设计效果来描述“该特定”设计的抽样效率。对于同一个设计来说，不同类别的变量和统计数据可能（而且往往）有极不相同的设计效果值；对不同分组人口来说，同一变量的不同估计数也有极不相同的设计效果值。第七章提供了涵盖6个国家10次调查的各种不同类别的变量，其一系列经验成果说明了各种调查之间和调查范围内的设计效果值的多样性。

第五章

发展中国家住户调查主抽样标架和主样本的设计

Hans Pettersson

瑞典国家统计局

瑞典斯德哥尔摩

摘 要

本章探讨有关主抽样标架和主样本的设计问题。导言后面分为如下几节：B节简要说明了展开和利用主抽样标架和主样本的理由；C节讨论了有关主抽样标架设计的主要问题；D节是关于主样本的，讨论了需要在设计阶段做出的若干重要决策（初级抽样单位的选择、抽样阶段的数量、分层、各层次间的样本分配，等等）。

关键词：主抽样标架、主样本、样本设计、多级样本。

A. 导言

1. 发展中国的国家统计局通常是国家级“官方”统计数据的主要提供者。国家统计局在履行这一职能的时候必须考虑到人口统计、社会和经济统计领域范围十分广大的信息需求。国家统计局使用各种不同的数据来源和方法收集数据。行政管理数据和注册资料在一定程度上是现成可用的，但样本调查永远是收集数据的一个重要方法。发展中国的大多数国家统计局每年进行几次调查。其中有些调查（比如生活水准计量调研、人口与健康调查、多指标类集调查等）在设计上相当规范，而其他一些调查则是为适应国家专项需要而“量身定制”的。基于对规划和协调不同调查工作的需要，正在加强住户调查方案项下各种调查活动的整合。现在，许多国家统计局都用长期调查计划取代了临时性的调查进程安排，根据长期计划持续进行或按照定期间隔开展涵盖各种不同议题的调查。联合国国家住户调查能力方案（NHSCP）在这一进程中发挥了重要作用。

2. 在住户调查方案中可以通过多种途径实现调查设计与运作一体化。可以在几项调查中采用相同的变量概念和定义。各项调查共用调查人员和设施，可以保障人员、设备的有效利用。一体化还可以包括在整个调查方案中所有调

查采用共同的抽样标架和样本。为各种调查制定一个主抽样标架（MSF）和一个主样本（MS），往往是一体化住户调查方案的一个重要组成部分。

3. 在第一抽样阶段采用地区单位的通用主抽样标架，可以提高住户调查方案中各项调查工作的成本效益。通常，扩展合理的抽样标架需要很高的成本。国家统计局通过确立一项持续进行的调查方案，就可以在多项调查中共同分担构建抽样标架的成本。

4. 如果各项调查从主抽样标架选定的一个通用主样本中选择各自的样本作为子样本的话，就可以将成本分担更推进一步。所有或大多数调查都采用一个主样本，即可降低每项调查在第二阶段及此后阶段中选择样本和准备抽样标架的成本。采用主抽样标架和主样本的上述成本优势也适用于在调查方案期间进行的一些预料外的临时调查；而且，在国家统计局没有正式调查方案的情况下的确也适用。

5. 本章将讨论有关主抽样标架和主样本的设计问题。联合国手册——《国家住户调查能力方案：一体化住户调查方案抽样标架和样本的设计》（联合国, 1986年）详尽描述了设计、准备和保有抽样标架和样本过程的各个步骤。该手册载有一个关于若干个案研究的附件。建议对此课题感兴趣的读者参阅该出版物中的详细论述。

B. 主抽样标架和主样本：概述

1. 主抽样标架

6. 正如第二章所述，发展中国家的住户样本通常是在几个抽样阶段中选取的。在第一阶段使用的抽样单位叫作“初级抽样单位”（PSUs）。这些单位是区域单位。它们可能是行政分区，比如辖区和地方选区，或者可能是为特别目的划定的地区，比如普查“查点区”（EAs）。第二阶段包括一个从选定的初级抽样单位中选出的“第二阶段抽样单位”（SSUs）样本。多级样本中的最后阶段抽样单位叫作“最终抽样单位”（USUs）。在一个多级样本中，每个选择阶段都需要有一个抽样标架——即从中选取样本的单位清单。第一阶段单位的抽样标架必须无遗漏、不重叠地涵盖整个调查人口，但是第二阶段抽样标架则只需保持在前一阶段选定的初级抽样单位范围内即可。

7. 如果初级抽样单位是行政单位，那么这些单位的清单可能是现成的，或者这样的清单可以轻而易举地根据行政案卷整理出来用作抽样标架。这样一个临时初级抽样单位清单可以在每次需要样本的时候搞出来。可是如果需要在某个时期开展一系列调查，最好准备并保有一个主抽样标架，以便随时用于各次调查。与每次调查临时准备一个专门抽样标架相比，节省的成本额是相当可观的。另外，这个抽样标架可为诸多调查所共用的事实，更容易证明扩展和保持主抽样标架在费用上的正当性，并促使主管机构在改进抽样标架的质量方面投入资源。

8. 主抽样标架基本上是一个涵盖全国的地区单位清单。每个单位可能有城/乡分类、较高层级单位识别（譬如该单位所属的政区和省份）、人口计数以及可能的其他方面的信息。每个地区单位必须保有关于单位界限的信息。以老挝人民民主共和国的住户调查为例，它的主抽样标架包括一份大约有11 000个村庄的清单。每个村都有关于住户数量、男性和女性人数、该村是城市还是农村（城市地区的行政分区也叫村）等方面的信息以及该村所属的政区和省份的信息。还有关于该村是否通公路的信息。

9. 主抽样标架的最普通类型是用普查查点区作基本抽样标架单位。通常，每个单位都有将该单位与较高层单位（行政分区）联系起来的信息。可以直接从这种类型的主抽样标架选择查点区的样本。还可以选择行政分区的样本并在选定的分区范围内选择查点区样本。

10. 一个有内在灵活性的最新主抽样标架，除了前面讨论的成本和质量方面的优势之外还有其他方面的好处。它便于为不同种类的调查迅速而简便地选择样本。另一个好处是，一个保持良好的主抽样标架会对下一次普查有使用价值。普查本身所需要的抽样标架类似于行将用于住户调查的抽样标架。如果在普查间歇期间有一个保持良好的主抽样标架在使用，那么开发普查所用的抽样标架的工作就容易多了。理想的情况是：在普查期内计划并构建一个新的主抽样标架，然后在下一次普查期间将其充分更新。

2. 主样本

11. 有可能在主抽样标架内为完全独立的不同调查选取样本。不过，在许多情况下，如果选择一个大样本（即主样本），然后为不同的（但有联系的）调查选择主样本下面的子样本，会得到实实在在的效益。不少国家调查局已决定开发一个主样本，以满足其住户调查的需要。

12. 所谓“主样本”就是一个可以从中选择子样本以满足不止一个调查或调查轮次需要的样本（联合国, 1986年）。它可以有多种形式。一种设计简便且相当普通的主样本就是由初级抽样单位组成的主样本，这里的初级抽样单位是指普查查点区。该样本用于两阶段样本选择，其中第二阶段抽样单位是居住单位或住户。

13. 可以采取多种方式进行子抽样。在（初级抽样单位的）初级层次上的子抽样将为每次调查给出一个独一无二的主样本初级抽样单位的子样本，也就是说，每次调查都有一个不同的查点区EAs样本。在二级层次上的子抽样将从每个主样本初级抽样单位当中给出居住单位的子样本；也就是说，每次调查都有同样的查点区样本，但在查点区内有不同的居住单位样本。子抽样可以独立进行，也可以采取某种受控选择过程，以确保样本间的重叠现象保持在预期的水平上。从主样本选择样本的另一种方法，就是从样本中选择独立的复制样本。可以选择一个或多个复制样本作为每次调查的子样本。这种构成要求主样本从一开始就是从一套完全独立的复制样本中建立起来的。

14. 各国的国家调查局可以通过使用主样本节省大量成本。选择主样本所付出的成本将由使用主样本的所有调查分担，这样就会降低每项调查选择样本的成本。由于主样本单位的选择基本上是办公室的操作（尤其在有完好主抽样标架的情况下），所以这一阶段节省的成本可能不是很多。当准备地图和主样本单位内居住单位子抽样标架的成本由各项调查共同分担的时候，节省下来的成本额就大多了。通常，建立子抽样标架所需要的实地工作量很大；而每次调查的这种实地工作成本的降低幅度几乎与使用同样子抽样标架的调查数量成正比。

15. 对于某些国家来说，由于实地巡回调查的难度之大、相关成本之高，一个经济实惠的做法就是在主样本初级抽样单位所在的地区或该地区附近招募采访员，并让他们在整个调查期内都驻扎在那里。在这种情况下可以使用相对较大的初级抽样单位。然后，通过使用这种初级抽样单位的一个固定主样本，即可获得明显效益，因为这就不用为每项调查选择一个新样本，也不必重新寻找或招募新的采访员了。

16. 使用共同的主样本可以缩短在本地区启动调查的时间。在许多发展中国家，采访员必须得到地区或地方当局的准许方可在该地区进行采访。例如，在老挝人民民主共和国和越南等国家，采访必须得到村长以上好几个行政层级的批准。当多项调查在同一地区进行的时候，就可以大大减少“开业”过程所需要的时间。

17. 多项调查共用主样本的初级抽样单位也可以减少采访员寻找住户所花费的时间。当具备了质量上乘的地图和子抽样标架的时候，采访员可以在该地区快速走访；有的情况下他们甚至可能在前次调查的时候在该地区工作过。可以实行永久性居住单位编号，以便利熟悉有关地区。有些主样本就是这么做的：Torene和Torene（1987年）就介绍了孟加拉国的主样本案例。

18. 使用主样本有可能在两个以上的调查中有重复取样的现象，这样可以通过将各次调查的住户数据联系起来，在低层次上进行数据整合。但是存在这样一种风险，即在多次使用样本单位的情况下可能会对调查成果的质量产生负面影响。参加了多项或几个轮次调查的住户，也许在后来的调查中就不愿意参与或者不太情愿给予准确的回答了。

19. 可见主样本对一个调查方案中的多次定期调查是有好处的（成本、统一和协调）。已经存在的主样本也可以让国家调查局有较好的准备来处理特别调查的抽样事宜：当特别调查需要的时候，可以从主样本中很快选取子样本。

20. 主样本的有利之处是显而易见的，但是也有一些不利之处或局限性。主样本的设计总要在方案的各项调查所产生的各种不同的设计要求中间作些妥协。主样本比较适合那些在有关地区的“域”估计数和目标人口分布方面设计要求适度兼容的调查。为主样本选定的设计量通常都能在相当程度上适合

调查方案中的大多数调查，但是没有任何一项调查完全适合。主样本的设计对各项调查规定了一些有时难以达到的约束和要求（比如在样本规模、类集、分层法等方面）。结果会使个别调查在效率上蒙受一定的损失。

21. 还有一些调查有主样本所不可能达到的设计要求，比如：

- 针对某些“小地区需要大样本”的行政区或地方辖区的调查（比如旨在评估某个地方辖区开发项目影响的调查）。
- 针对某些非均衡分布分组人口（比如族群）的调查。

22. 上面第一类调查的一个实例是在老挝人民民主共和国北部四省份定期进行的鸦片种植调查。其目的是评价老挝政府旨在减少鸦片种植的项目进展情况。在这个案例中，由于老挝的主样本不能满足对样本设计的要求，所以又为调查另外选择了一个样本。（本来有一种替代办法，就是在这四省份使用主样本的初级抽样单位，同时从主抽样标架中选取附加的初级抽样单位。）

23. 有的情况下也许不能完全达到利用主样本节省成本的目的。从一个主样本提取适合特定调查所具体需要的子样本，然后正确计算出选择概率，这得需要具备一定的技术能力。这种操作可能比单独选择一个样本还要复杂。在许多发展中国家的国家统计局，抽样统计员奇缺，这一事实可能会妨碍采用主样本，或者说实际上阻碍了主样本的开发。由于国家统计局缺乏抽样工作能力，对主样本利用不足的事例屡见不鲜。

3. 归纳和结论

24. 以上讨论的有利条件、不利条件和局限性可以归纳如下：

主抽样标架

- 具有成本效益；国家统计局可藉此将构建抽样标架的成本分摊到多项调查。
- 通常质量会比特设抽样标架好，因为它比较容易调动投资用于改进抽样标架的质量，从而延长范围使用期。
- 简化提取个别样本的技术程序；便于简易快捷地选择不同调查样本。
- 如果保持得好，它的使用价值可以延续到下一次普查。

主 样 本

- 节省成本：
 - 选择主样本单位的费用将由使用主样本的所有调查分担。
 - 准备住宅单位地图和子抽样标架的费用将由使用主样本的各项调查共同分担；但是子抽样标架需要定期更新，以补充新建住所和删除已拆毁的住宅单位。

- 在因为现场交通不便、成本高而采访员需要驻扎在或接近初级抽样单位地区的情况下，采用主样本有明显的效益。
- 提高工作效率：
 - 多项调查使用同一样本的初级抽样单位，可以缩短有关地区的调查启动时间和采访员寻找受访者的时间。
 - 主样本便于简易快捷地选择样本；在专项调查需要的时候可以很快地从主样本选取子样本。
- 数据整合：
 - 由于使用主样本有可能在两个以上调查中重复使用某些样本，所以可对这些调查所得的数据进行整合。
- 局限性和不利条件：
 - 主样本未必对所有调查都适用；有的时候，国家统计局可能会在调查方案期内面临主样本所满足不了的意外调查需要（其实这属于局限性而非不利条件）。
 - 在重复使用样本的情况下，尤其在住户一级，由于受到制约条件的影响和住户因累积受访的负担而不响应调查的现象增多，因而有产生种种偏差之虞。
 - 主样本的连续使用需要有必要的抽样技能，而这种技能国家统计局未必具备。

结 论

25. 主抽样标架和主样本显然有许多颇具吸引力的特征。对于国家统计局来说，可取的是保有一个维持完好的主抽样标架，以满足其住户调查的需要，而不论是在调查规划中进行的调查还是临时性的专项调查。许多国家统计局会发现，采取进一步措施为所有或大多数住户调查设计和使用一个主样本是有益的。

C. 主抽样标架的设计

26. 国家住户调查方案在诸如预计样本数量、人口覆盖面、分层法和样本规模等方面界定了主抽样标架和主样本的要求。如何在设计工作中达到这些要求，取决于各国构建抽样标架的条件。最重要的因素是可用于构建抽样标架的现有数据和其他材料的状况。我们将在下面第1小节简要讨论所需数据和材料的种类以及数据中可能存在的质量问题。

27. 国家统计局在评估了现有数据和材料之后，必须对与主抽样标架的如下几方面有关的关键特征做出决策：

- 主抽样标架的覆盖面（见第2小节）；

- 哪些地区单位应充当主抽样标架中的标架单位（见第3小节）；
- 主抽样标架中应包括有关标架单位的哪些信息（见第4小节）。

28. 有关抽样标架的完备的、经过妥善处理的文件编制和明晰的更新程序，对于主抽样标架的有效使用至关重要（见第5小节）。

1. 数据和材料：质量评估

29. 通常，最近的普查资料是最重要的数据和材料来源。在国家统计局打算用查点区作标架单位的情况下，这种资料来源的重要性是显而易见的；但即便使用其他（行政）单位，一般也需要有这些单位普查的人口和住户数据。带有户口和住户计数的查点区清单和查点区示意图是出自普查的基本材料。另外还有较大地区（辖区、地区）标有查点区的地图。通常，查点区使用表示城/乡分类及其所属行政区和分区的编码来识别。有时候编码还表示出查点区内是否有集体户人口（比如军事驻地的人口、学生宿舍的人口，等等）。

30. 各国之间在普查数据和材料的质量上差异悬殊，尤其是地图质量。有些国家（譬如南非）拥有储存在数据库中的数字化查点区地图，而其他一些国家（譬如老挝人民民主共和国）则根本没有像样的地图。有些国家，查点区地图往往非常粗糙，很难用于现场调查。由于查点区或许实际上由地点清单组成而不是由地区单位组成，所以分散在清单地点以外的人口就可能没有包括在抽样标架之内。一个与质量有关的、让抽样标架开发者感到有点儿头痛的特殊问题，就是检索普查材料、特别是检索地图非常困难。也许地图质量还不错，但如果难以检索，再好的地图也帮不了忙。普查过后查点区地图被“埋”在档案馆里，有时还杂乱无序，找起来很困难，这种情况至今仍然普遍存在。有些计数地图在档案馆遗失的事例也屡见不鲜。

31. 一般来说，普查资料的质量随着时间的推移而降低。在各查点区受到人口增长和迁移的不均衡影响的情况下，查点区的人口计数质量就肯定会降低。另外，行政单位的变化——比如边界变化或抽样单位的拆分或合并——也会导致普查信息过时。假如最近的一次普查是在七、八年前搞的话，普查信息肯定已经过时。

32. 主抽样标架设计的第一步必须是对可用于构建抽样标架的现有各种不同的资料进行鉴定和评估，其中不仅包括普查资料，而且还包括其他数据/材料；即便将普查当作主要资料来源，也可能还需要其他来源，以便更新或补充普查数据。可以提出如下问题：现有哪些数据/材料，其精确度如何？数据的现时性如何，多长时间更新一次？关于地图，应就其详尽程度和行政分区的边界标示到何种程度来评价。要努力估计出达到所要求的质量标准的查点区地图占多大比例。

33. 在现阶段的工作中，获得或编写一份对国家行政架构的确切而透彻的描述及其经过更新的行政区和分区的清单也很重要。

2. 关于主抽样标架覆盖面的决策

34. 需要尽早对主抽样标架的覆盖面做出决定。该不该把边缘地区某些人烟稀少的部分排除在抽样标架以外？大多数国家决定主抽样标架要充分覆盖全国，一般来说这是明智的决策，因为即使将某些人烟稀少的边缘地区排除在住户调查方案的定期调查以外，到时候可能还会出现需要对这些地区进行专项调查的情况。游牧民群体和高山部落属于特殊情况，对他们很难取样和进行实地采访。有些国家将这些群体排除在住户调查的目标群体之外。

35. 还必须对集体户人口的覆盖面问题做出决策。在有些国家，较大的集体户被界定为特殊计数领域（比如寄宿学校、大医院、军营、以及矿工集体宿舍等）。在这种情况下将这些领域排除在抽样标架以外是可能的。不过一般来说，最好还是把这些单位纳入抽样标架，以便为将来调查有关覆盖面的决策留有余地。

3. 关于基本标架单位的决策

36. “标架单位”，就是纳入主抽样标架的抽样单位。“基本标架单位”是主抽样标架内的最低层单位。一般来说，基本标架单位最好是小地区，以便在某项调查经费有限的情况下可以考虑将这些小区并入较大的抽样单位。

37. 普查查点区往往是基本标架单位的最好选择。查点区用作基本标架单位有几个好处。划分查点区的目的是产生人口规模大体相当的地区，这在某些抽样状况下是有利的。查点区标示在地图上；通常这种地图附有边界描述。一般可以得到在行政区划内标明各查点区位置的基本地图。普查中产生计算机化的查点区清单，这些清单可作为主抽样标架的起点。权衡利弊，将查点区用作标架单位有不少好处，但是前面第1小节中讨论的那几种质量问题也许会在某些情况下导致考虑别的解决办法。

38. 有些国家的行政分区小到可以充作基本标架单位；有的情况下把这些单位当作基本标架单位使用比查点区更好，譬如老挝人民民主共和国国家统计局保有的行政分区就介入了主抽样标架。该国曾经把查点区当作基本标架单位，但后来发现查点区的文件编制很难检索，而且普遍质量较差，以至于在现场调查中难以跟踪查点区的边界。在这种情况下决定利用“村”作为基本标架单位。老挝人民民主共和国的村是界定明确的行政单位。可是，在严格意义上“村”并不是地区单位。村与村之间的界线是模糊的，没有适当的地图可资使用，但是至于哪些住户属于某个特定的村，却一点儿也不含糊。

39. 把小于查点区的单位当作基本标架单位来使用的情况并不普遍，但是这种情况的确存在。泰国便是一例，在该国的市政辖区，查点区进一步划分成街区，按每个街区进行人口和住户普查计数。在主抽样标架MSF的市政部分，这些街区被当作基本标架单位使用。

40. 各种基本标架单位，不论是查点区还是其他类型的单位，就各该区的住户和人口而言在规模上不尽相同。即便意在创建人口规模变异不太大的查

点区，也会因为种种原因而发生偏离这种规则的情况（比如在人烟稀少、交通困难的地区可能建立人口较少的查点区）。其结果，在一些高低两级相差悬殊的极端情况下，通常查点区的大小会有相当大的差异。以越南为例：每个查点区平均有100个住户。在总共166 000个查点区当中，每个区的住户少到只有2户，最多的达到304户（Glewwe和Yansaneh, 2001年）。住户数在50户以下的查点区占1%。在老挝人民民主共和国，小查点区所占的比例甚至更大：6%的查点区户数都在25户以下。一般来说，这种作为基本标架单位的地区在人口规模上的差异不算个问题，但是太小的单位不宜用作抽样单位。极小的查点区在主抽样标架内是可以接受的；但是对于基于主抽样标架的样本来说，这些查点区需要跟邻近的查点区合并成适当的抽样单位。

4. 拟纳入抽样标架的标架单位信息

41. 可以使用基本标架单位的简单清单构建一个初期抽样标架，但是从这样一个范围提取有效样本的可能性是有限的。如果抽样标架包含一些可用于开发有效样本设计的标架单位的补充数据的话，这个范围的使用价值即可大幅提高。补充数据可分为三种类型：

- (a) 可将基本标架单位并入较大单位的信息。为了增加从该范围有效抽样的潜力，有一种方法就是设法从这个范围抽取各种不同类型的样本。因此，该范围最好包含有可能形成较大单位的信息，从而实现从该范围选择抽样单位的灵活性；
- (b) 有关单位规模的信息。如果每个标架单位包含规模量数的话，也可以提升该范围样本的效能。在单位规模差异很大的情况下，这一点特别重要；
- (c) 其他补充信息。可用于单位分层或作为估算阶段的辅助变量的信息，可以提高主抽样标架的样本效能。

将基本标架单位并入较大单位所需要的信息

42. 对于某些调查来说，初级抽样单位的最好替代物就是像查点区之类的小区。对其他一些调查来说，出于成本和抽样误差方面的考虑，可能会倾向于比查点区大得多的初级抽样单位。这些较大的初级抽样单位可由相邻查点区的群体来构建。另一种可能就是用地方选区或辖区之类的行政单位作初级抽样单位。在所有这些情况下，主抽样标架都有必要提供构建较大初级抽样单位的可能性。因此，主抽样标架中的标架单位记录要包含有关该标架单位所属较高层单位的信息，这一点很重要。

43. 一个已被许多国家采用的主抽样标架模型设计，是使用普查查点区作基本标架单位，其中的单位按地理排序编入一个等级式结构的较大（行政）单位。可以通过几种不同的方式从主抽样标架提取样本：(a) 通过查点区抽样；(b) 通过合并查点区以形成适当规模的初级抽样单位，并对这些初级抽样单位进行取样；(c) 在第一阶段对行政分区进行抽样，并在接下来的各阶段进行后

续抽样，直至查点区层级。在越南，主抽样标架的等级式结构包括如下几个层次：

- 省级
 - 地区级
 - 乡镇（农村），市内行政区（城市）
 - 村（农村），街区（城市）
 - 查点区

44. 如果所有标架单位（含基本标架单位和上级单位）都根据地域毗邻性所指定的标识符的话，就可以进一步增加选择抽样单位的灵活性。这样就可以用标架单位作“积木块”来构建毗邻标架单位所需规模的初级抽样单位。在前面小节所讲的越南和老挝人民民主共和国的情况下就需要这种运作。带有基于地域毗邻性的标识符的另一个好处是，可以利用按照地理排序抽样单位的系统抽样，从主抽样标架中选取地域分散的样本。

标架单位的规模量数

45. 如果标架单位的规模差异很大的话，包含规模量数特别重要。通常，规模量数是指标架单位内的人口、住户或住所的单位计数。有必要指出的是，规模量数不需要十分确切。实际上，它们差不多总是在一定程度上不够精确，因为规模量数所依据的是前一个时间点的数据，而事实上人口在不断变化，因而使得这种量数逐渐过时。规模量数的误差不会导致调查估计数出现偏差，但是的确会减损规模量数的使用效能，尤其在估算阶段使用规模量数的情况下。因此，应努力确保规模量数尽可能准确。

46. 在概率与规模成正比（PPS）的标架单位样本选择中，规模量数使用最普遍。规模量数的其他用途有：

- 确定各个层次样本的初级抽样单位分配；
- 形成按规模分类的单位层次；
- 作为比率或回归估计数的辅助变量；
- 形成可取规模的抽样单位。

标架单位所需的其他补充数据

47. 应该考虑在抽样标架中包含可用适度成本获取的有关标架单位的补充信息。有关标架单位中的人口密度、主流族群、主要经济活动和平均收入水平的信息往往是对分层有用的变量。

48. 在纳米比亚的主抽样标架中，首都温得和克的城市基本标架单位中包含了一个高、中、低收入水平概略分类法，从而得以在温得和克的城市子域形成两个收入水平层次。另一个实例是老挝的主抽样标架，其中农村标架单位载

有关于该单位距离公路远近的信息。采用主抽样标架的住户调查样本按照不通公路分层。

5. 主抽样标架的文件编制和保管

文件编制

49. 拥有一套保存完好、准确而便于查阅的主抽样标架文件资料是利用该抽样标架所必不可少的条件。如果这套文件资料的质量很糟糕，就必然不能充分发挥该抽样标架的效益。文件资料的核心是载有全部标架单位的数据库。标架单位的案卷内容应该包括：

- 初级数字标识符。它应该有一个能够识别所有行政区和标架单位所在分区的独一无二的编码。如果标架单位是按地域排序编号，那是一个优点。通常，查点区编码就有这些特征。完全数字标识比按名称和字母顺序编码好。在许多情况下，行政来源的和普查来源的现有地域编码系统适合用初级标识符。
- 二级标识符，标示标架单位所在的村名（或其他行政分区名称）。二级标识符用于在地图上和现场查找标架单位。
- 若干单位特征，比如（人口、住户的）规模量数、城市或农村、人口密度，等等。凡是可以用适度成本获取且质量尚可的所有数据都要包括进去。这些特征可用于分层、指定选择概率，以及作为估算用的辅助变量。
- 运作数据、单位变化的信息，以及样本用途标识等。

50. 抽样标架要易于查阅并可供分类、过滤和制作统计汇总表等各种操纵处理。如果将抽样标架存入计算机数据库的话，上述使用效果最佳。要避免采用只有专家才会用的格式。使用简单的Excel电子数据表往往就很管用。Excel使用起来很方便，不少人都会用，它并且具有在利用抽样标架制备样本时所需要的分类、过滤和聚合等功能。可以很容易地把电子数据表引入大多数其他软件包。

保持主抽样标架

51. 与主抽样标架的文件编制密切相关的是有关保持抽样标架的例行规定。在使用抽样标架期间，可能会发生一些对标架单位数量和界定都有影响的变化。保持一个主抽样标架所需要的工作量主要取决于标架单位的稳定性。标架单位可能发生两种变化，即：标架单位的界线变化和标架单位的特征变化。

52. 范围界线的改变首先影响到行政分区。由于政治和决策的缘故，改变行政分区界线的情况时有发生，尤其在较低层级。这些改变往往是伴随有关地区人口的实质性变化而发生的。由于原有单位的拆分或组合、或者对原单位做

了比较复杂的重新安排，产生了一些新单位。现有单位的界线改变也有可能不产生任何新单位。如果行政分区变动频繁，就需要动用大量资源才能保持抽样标架不断更新的精确状态。

53. 影响标架单位边界的变化必须记录在主抽样标架。需要建立一个收集有关行政变动信息的制度，以便跟踪这些变化。

54. 标架单位特征的改变不仅包括更名之类的变化，而且还包括较为实质性的改变，譬如规模量数（人口或住户/住所单位数量）的变化，或者城/乡分类变化。这些变化不一定非得反映在主抽样标架。但如前所述，有关规模量数的过时的信息会减损从该范围选取的样本效能。整个范围更新规模量数可能需要付出高昂的成本，一般来说不划算；但是特别对于迅速成长的城市周边地区来说，定期更新规模量数则不失为一个好主意。

55. 当人口发生大规模突变的时候——譬如在非法占用公地搭建住所的地区，当地方当局决定在这种地区拆除违章建筑的时候，就可能发生此种情况——规模量数的变化就成问题了。这种突变需要反映在抽样标架中。老挝人民民主共和国政府发起的边远山区农村移民计划提供了一个变化不太显著但仍然（对抽样标架）构成问题的实例。该国政府正在鼓励这些村子的老百姓转移到基本服务条件较好的村庄。这一进程的结果是，两年之内那里的村民大约减少了10%。显然，这些变化必须包括在抽样标架之中。

56. 当国家统计局的业务经费奇缺而又不得不竭力满足对统计成果需求的时候，主抽样标架的维持就有被忽视的危险。因此重要的是，国家统计局需要在早期阶段制定旨在更新抽样标架的计划和程序，并为此目的分配足够的资源。

D. 主样本的设计

57. “主样本”就是可以从中选取子样本以满足多项或多轮次调查需要的样本（联合国，1986年）。主要目标应该是为在分析“域”及其目标人口在本地区分布方面的设计要求有适当兼容性的住户调查提供样本。主样本是按照抽样阶段数量和作为最终抽样单位（USU）使用的单位类别界定的。在两个阶段中选中的以查点区为第二阶段单位的主样本，叫作两段查点区主样本。如果在第一阶段直接选出查点区，则叫作一段查点区主样本。这两种设计都是发展中国家常见的主样本设计。

58. 下面D.1节至D.4节讨论了开发主样本的几个重要步骤。D.5节和D.6节讨论有关主样本文件编制和保持的问题。最后，D.7节讨论在基本不是针对住户的调查中使用主样本的问题。

1. 主样本初级抽样单位的选择

59. 主抽样标架提供主样本的选择范围。主抽样标架中的基本标架单位有时候可以用作主样本的初级抽样单位。在别的情况下，我们可以决定构建比

主抽样标架中的基本标架单位大的初级抽样单位。在后一种情况下，通常把某种明确划定的行政单位（比如县、区等）当作初级抽样单位使用。但是也有用标架单位作“积木块”来构建初级抽样单位的情况；在这种情况下相邻的单位被并入适当规模的初级抽样单位。莱索托的主样本便是一例：其中的初级抽样单位是通过将相邻的普查查点区组合成包含300-400个住户的组别构建的。全国3 055个查点区被编入1 038个用作初级抽样单位的查点区组（Pettersson, 2001年）。

60. 在决定如何构建初级抽样单位的时候，要考虑关系到统计效率、成本和运作程序的若干因素。假定主抽样标架内的基本标架单位是查点区，那么，我们在哪些情况下宁愿使用比查点区大的单位作初级抽样单位呢？

- 如果已知很大一部分查点区的划界质量很糟糕，我们就可能使用较大的单位作初级抽样单位，因为较大区域的划界一般都比较稳定而明确。
- 在地区之间交通困难且/或差旅费很高的情况下，在提取初级抽样单位的地区或接近这种地区的地方招募采访员并且让他们在整个调查期内驻扎在那里，就交通困难及相关费用而言这是一个经济节俭的办法。
- 在广泛使用初级抽样单位作样本，以至于像查点区这样小的初级抽样单位会很快枯竭的情况下，解决这个问题的办法要么是采用较大的单位作初级抽样单位，要么坚持用查点区做初级抽样单位并轮换查点区的样本。当进入并在该地区开展调查的费用很高的时候，第一种选择比较可取。
- 从成本控制和抽样效率来考虑，当实行一段或多段抽样而涉及的单位大于基本标架单位已成为惯常做法的时候，如果——譬如讲——基本范围是查点区的话，我们就可能决定先采用较大单位（比如市内行政区）作初级抽样单位，然后在下一阶段选择初级抽样单位内的查点区或其他区域单位。
- 像有些调查采用的那种做法，把住户和个人的变量同社区变量联系起来。这种做法的一个实例是在保健调查中将个人保健变量同村里或社区的保健设施变量联系起来。另一个实例是在生活水准调查中将住户的变量同社区有关学校、道路、水源、公共卫生设施、当地物价等方面的变量联系起来。如果主样本要为多项同类调查所用的话，把社区（比如村、镇、市内行政区等）作为初级抽样单位会有多方面的好处。如果将社区用作初级抽样单位，我们就可以肯定第二阶段抽样单位的子样本将会很好地遍布全社区。

61. 大区域单位不适合用作初级抽样单位，因为这种区域太少了。从50-100个单位的人口中取样没有意义。特定人口的初级抽样单位最好在1 000个以

上，以便10%的样本可以为该样本产生100多个初级抽样单位。如果分数大大超过10%，就会降低抽样的成本效益。如果该样本中的初级抽样单位数量远低于100，就会增加差异。还应指出的是，在人口的不同部分采用不同类型的初级抽样单位可能效果较好，比如在城市地区采用查点区，在农村地区采用较大的单位。

2. 旨在减少初级抽样单位规模差异的地区合并与拆分

62. 在采取哪种类型的单位作初级抽样单位（在两区段情况下，用哪种单位作第二阶段抽样单位）的问题上作决定的时候，我们也许会发现存在一些较之可取规模过大或者过小的“非正常值”的抽样单位。

极小的抽样单位

63. 主样本中的初级抽样单位过小是个问题。多大的规模应视为可取的规模，这要取决于所期望的主样本工作负荷。南非国家统计局采用查点区作为主样本的初级抽样单位，它决定初级抽样单位的最小规模为100个住户。在准备主抽样单位期间，少于100个住户的查点区与相邻查点区合并。纳米比亚国家中央统计局的主样本采用的规则是初级抽样单位的规模须不少于80个住户。在普查中形成了2 162个查点区。在把相邻的小查点区合并以后，剩下了1 696个初级抽样单位。在这1 696个初级抽样单位中，405个是通过合并几个查点区组成的；其余的1 291个初级抽样单位个个都是只有一个查点区。

64. 如果小查点区数量很大，就非常有必要在选择前合并小查点区。可以用越南的案例来说明。为了调查的目的，越南中央统计局想要的区域样本规模是不少于70-75个住户。大约有5%的查点区（约计8 000个查点区）规模不足70个住户（Pettersson, 2001年）。将大约8 000个查点区并入相邻的查点区是个既乏味又耗时间的工作。

65. 有一个办法可以减少将小区单位合并成适当规模所需的工作量，就是只有当一个小区（初级抽样单位）碰巧被选入样本时，才将其并入别的区。Kish（1965年）设计了一个在选择期间或之后将小区并入相邻初级抽样单位的程序。

66. 减少合并小单位工作量的另一个办法，就是在拟定的第一阶段之前施加一个抽样阶段。有时我们可以不用拟定区域单位作初级抽样单位，而是用较大的区域作初级抽样单位。我们在选取的初级抽样单位中将小区单位（即我们原本拟定的初级抽样单位）并入相当规模的区域单位。合并小区单位的工作只在选定的第一阶段单位中进行，这与使用较小的区域作第一阶段单位的情况形成对比。上述替代办法在拟定的第一阶段之前增加了一个取样阶段，这也许会对设计效率有影响。可是，如果我们为每个选定的初级抽样单位只选取一个第二阶段抽样单位的话，该样本实际上相当于区域单位的第一阶段样本。在越南的案例中就采用了这个解决办法。该国决定使用较大的行政单位——即镇而不

是查点区——作为初级抽样单位。在选定的镇中规模不足的查点区与相邻的查点区合并，形成规模适当的单位。这样一来，合并相邻查点区的工作量就减少了。结果是没有合并8 000个查点区，而是将1 400个查点区并入了1 800个选定的镇。在选定的镇当中第二阶段选取了三个查点区（或者，在小查点区的情况下叫作“查点区组”）。

特大区域单位

67. 另一个极端是，可能有的情况下区域单位过大（不论就人口而言还是就地理区域而言），不合作初级抽样单位。不论哪种情况下，其造册成本都会大大高于普通区域单位（查点区或其他一些区域单位）。如果选取了某些特大初级抽样单位作主样本的话，这两种情况都会出问题。为了减少编制这些大单位中的住户范围清单的工作量，我们可以将这些大单位分别放进单独的层次，而按照降低的抽样率来选取初级抽样单位；我们可以通过增加初级抽样单位内的抽样率来维持总体抽样率。

68. 处理这种特大初级抽样单位问题的另一个办法，就是把初级抽样单位分成若干部分，并随机选取其中一部分。这个问题要比初级抽样单位过小的问题简单一些，主要是因为我们无须在选择主样本之前采取任何行动。只有当我们碰巧选取了一个特大初级抽样单位作主样本的时候，才有必要采取上述行动。

69. 另外一个问题是自上一次普查以来有些初级抽样单位显著增长或下滑的问题。经过一个时期之后人口总要发生变化，因而使得初级抽样单位的规模量数的精确度降低。一般的后果是增加了变异；但是尚未发生偏差。由于种种原因，譬如讲开拓市郊或者某些地区的大规模工程建设等，某些初级抽样单位发生显著变化，问题就变得严重了。必须设计处理这些变化的程序并将其列为维持主样本工作的一部分。联合国国家住户调查能力方案手册中讨论了两种策略，即：样本置换和样本修改（联合国，1986年）。

3. 初级抽样单位的分层和各层次的主样本分配

分 层

70. 主样本的初级抽样单位通常划分为国家主要行政区（省、地区等）层次；在这些行政区内又划分为城市部分和农村部分。其他通用分层因子是：城市化水平（大都会、城市、乡镇、村）以及社会经济和生态特征。在莱索托的主样本中，初级抽样单位分成10个行政区和4个农业经济区（低地、丘陵、山区和Senqu河流域），总共有23个层次，反映了农村地区的不同生活方式。

71. 在接近大城市的农村地区划定“城市边缘”层次是可能的。这样就可以照顾到在一定程度上依靠现代部门的农村住户。在大城市中，可以根据住房标准、收入水平或某些其他社会经济特征进行二级分层。

72. 在主层次范围内实现更深度分层的一个通用方法，就是根据分层标准对层次内部的初级抽样单位加以整理并系统化地选取样本（即所谓“隐含分层”）。隐含分层的一个好处就是无须划定层次的界限。

样本分配

73. 可以采取如下不同的形式来分配各层次的主样本初级抽样单位：

- 按照各层的人口比率进行分配；
- 各层次平均分配；
- 按照各层人口的平方根比率进行分配。

74. 许多主样本都是按照各层中的人口（人数或户数）比率分配给各层次。可是，按比率分配就意味着小层次的样本份额就小。当主要层次是需要单独调查估计数的国家行政区（比如省），而这些行政区的大小又相差悬殊的时候（实际情况往往就是如此），这种分配方法就可能成问题。各省的最高政府官员们（至少是小省份的官员们）会强烈要求在全省之间平均分配样本。在省与省的规模大小相差悬殊的情况下，平均分配就会导致各省之间在抽样比率上的实质性差异。在老挝1997年构建的主样本中，决定在19个省级层次中采取平均分配的办法，以实现省级估计数的精确度均等。结果是，最小的省份抽样比率竟比人口最多的省份高10倍之多。

75. 就城/乡不同的“域”而言，如果严格按比率分配，就会在一些城市人口少的国家出现城市样本小的现象。柬埔寨统计研究所编制的主样本是按比率分配给各省和城/乡地区的。这个由600个初级抽样单位组成的样本包括512个农村单位和88个城市单位。对某些调查来说，据认为城市样本太小了，因此要求追加城市初级抽样单位的取样。要是当初在主样本中或多或少额外增加一些城市域的取样的话，则较为明智。

76. 平方根分配法是“按比率分配”与“平均分配”的折衷产物，就是按照层次规模的平方根成比例地分配样本。越南和南非就采用了平方根分配法。Kish（1988年）推荐的代性折衷方案系依据与 $\sqrt{W_h/n}$ 成比例的分配，式中， n 是样本总体规模； W_h 是 h 层的相对规模；而 H 是层次数量。对于非常小的层次来说，第二项支配第一项；因此可以保证，对小层次的分配份额不会太小。

77. 另一种折衷办法就是搞一个适合于省级估计数的大型主样本，适用于作全国估计，并有一个从大型主样本选取的子样本。例如，菲律宾1996年的主样本是一个适用于省级估计数的扩大的样本，由3 416个初级抽样单位组成；同时带有一个包含2 247个初级抽样单位的子样本，在只需要地区级估计数的情况下充当核心主样本。

4. 初级抽样单位的抽样

78. 最普通的方法就是按照与规模成正比的概率（PPS）选取主样本的初级抽样单位。在这种情况下，选定一个初级抽样单位的概率与该初级抽样单位的人口成正比，因此初级抽样单位越大，其被纳入样本的几率越高。

79. 在初级抽样单位在规模上差异很大的情况下，这种方法有一些实际的好处。首先，它可以导致产生自动加权的样本。其次，它可以在初级抽样单位内部产生大体均等的样本，而这样的规模又反过来意味着大体均等的采访工作量，从现场调查的观点来看这种情况是可取的。前面第二章详细讨论了与规模成正比的概率PPS及其优点和局限性。

80. 选择PPS样本的方法有好多种。一种常用的方法是在层次内部进行系统化选择。如果初级抽样单位是按照层次内的某种地理排序列表的话，这将导致样本在主要层次中的良好地理分布（详见第二章）。莱索托、老挝人民民主共和国和越南的主样本都是采用系统化的与规模成正比的概率选择的，每个层次内部有一个随机选定的起点。

交叉子样本

81. 另一种选择样本的任择方法是选择一套交叉子样本。一个交叉子样本是这样一套子样本之一：其中的每个子样本自身构成目标人口的一个概率样本。

82. 在对主样本进行子抽样的时候使用交叉子样本有一些好处。子样本可以使样本规模比较灵活。供特定调查用的样本可以由一个或多个子样本构成。子样本还可以在多轮次的调查中用于样本置换。

83. 在主样本设计中，交叉子样本的使用不如对简单系统化选择方法的使用那么普遍。尼日利亚国家统计局提供了一个使用交叉子样本作主样本的实例（Ajayi, 2000年）。

5. 主样本的耐久性

84. 主样本的质量会随着时间的推移而退化。但如果在主抽样标架内所有单位的人口变化或多或少是均匀的，那么，随着人口变化而导致用以指定选择概率的规模量数过时的事实就不会成为问题。可是，通常情况并非如此。不同地区的人口增长率和迁移率是有差异的：有些农村地区往往人口增长率比较低，甚至呈下降趋势，而一些城市的郊区则人口增长率较高。当发生这种增长不平衡现象的时候，用于选择主样本的规模量数就不再反映调查人口的相对分布。这会导致增加主样本抽样估计数的误差。另外，政区边界的变更和分类（比如城/乡地区分类）标准的改变也会导致分层过时。

85. 正常情况下，每次普查（一般每10年一次）之后都要彻底修改主抽样标架。在普查间歇期内要定期更新抽样标架。如果有一个保存完好、定期更新的主抽样标架，就能从主抽样标架定期选择全新的主样本。于是问题就来了：一个主样本可以在不发生重大变化的情况下保持多长时间？主样本的耐久性在一定程度上取决于当地的诸多条件，比如国内迁移和行政单位的变化率。因此不可能提出一个适合所有情况的一般建议。一般情况下，一个主样本的效能经过三、四年之后会大幅度退化。在不加调整的情况下有关延长主样本使用期限的决策需要慎重考虑。

86. 关于处理主样本效能退化的问题，基本上有两种策略。一是定期选择全新的主样本，譬如在莱索托，主样本每三年更换一次。另一种策略是在较长时期内保留主样本，但定期加以调整，以补偿抽样标架和样本单位的变化效应。这些调整可以单独创建高增长率层次和制定处理影响到抽样单位或层次的行政区变更的规则。这种修正策略虽然在澳大利亚的主样本中应用过，但在发展中国家很少用。一个原因大概是从抽样观点来看这种策略太复杂，需要在设计和执行过程中多加小心并且要有较高的机能。

6. 文件编制

87. 如果主样本是从一个文件资料完备的主抽样标架选取的，到这个时候许多文件编制工作都已经完成。然而在发展中国家，文件编制工作往往是主样本的一个薄弱环节。有关信息是分散的，有的时候也很稀少，所以很难对样本选择实施跟踪和计算抽样概率。每个阶段的所有主样本选择程序和选择概率都必须充分编入文件。还必须保有关于哪个主样本已在特定调查样本中使用的记载。抽样单位必须采用标准的识别编码。

88. 主样本的文件编制还必须包括就重要估计数的抽样误差和设计效果而言主样本的性能量度。这些性能量度对根据主样本规划新调查的样本规模和样本分配十分有用。现在已有许多统计分析软件包程序可用来计算确切的变量和设计效果值（详见第二十一章）。

89. 文件编制工作还应包括主样本的辅助材料。如果为主样本的最终抽样单位准备了二级抽样标架，那么，这些抽样标架应当成为文件资料的一部分。二级抽样标架将包括街区或区段之类的区域单位或者像主样本最终抽样单位范围内住所之类的列表单位。

7. 使用主样本调查公司企业

90. 主样本的主要目的是为连续调查方案中的住户调查（以及适合主样本设计的任何特别调查）提供样本。因此，样本设计基本上是为一套基本住户调查服务。一般来说，它对其他类型单位的抽样是无效的。不过有些情况下，有可能将主样本用于与研究经济单位——比如家庭企业、自营企业和小型农业股份公司——特征有关的调查。

91. 在大多数发展中国家，在服务业、贸易和农业部门有很大一部分经济实体与私人住户有关。一般来说这些企业为数众多，规模很小，并广泛分布于整个人口当中。在这种企业和住户之间往往可能存在某种一对一的对应关系，并且可能是住户而不是这些小企业本身充当最终抽样单位。一个住户主样本可以用来对这些公司企业进行调查。这往往需要同自动加权设计分隔开来。Verma（2001年）讨论了提高经济单位调查样本设计效率的各种方法。

92. 不过，通常也有一些与住户没有联系的较大企业。这些企业一般数量很少，但是它们在许多估计总量（总产值、雇员总人数，等等）中占有很大比重。许多情况下它们在总人口中的分布也不平衡。由于区域主样本不能对这些

大单位进行有效的抽样，所以有必要单独为它们建立抽样标架。在许多情况下可以根据政府机构（譬如税务或特许机构）的案卷来构建这种抽样标架。除了根据主样本初级抽样单位选定的样本之外，还应该从上述清单中为调查选定所有特大单位和其余单位的一个样本。

93. 当住户调查与“社区调查”联系起来的时候，企业调查就会出现一种特殊情况。譬如在保健调查中，可以通过一项对覆盖每个原样本区（比如查点区）周围拓展地区卫生保健设施的调查来补充对个人/住户的调查。补充调查所获取的数据可用于两个目的：（a）可以跟住户数据联系起来，用以分析当地保健设施的质量和可利用的机会；（b）可用于产生有关保健设施数量与种类的全国估计数。关于第一个目的，住户和个人依然是分析单位；不牵涉任何新的抽样问题。第二个目的可能产生一些复杂因素。如果将原样本区周围的较大拓展区视为由与取样区并列的若干区域组成的较大区域（辖区、镇、普查监督区，等等），那么情况就简单了。由此产生的样本就相当于较大地区的一个样本，而较大区的选择概率相当于该较大区所包含的若干较小区选择概率的总合。可是，如果较大区域是根据“在距原样本区×公里范围内”的规则构建的话，选择概率的确定就比较复杂了。

E. 结论

94. 住户调查的设计和执行为所有国家统计局的一项重要任务。不少发展中国家的国家统计局每年进行好几项调查。出于对调查进行规划和协调的需要，促使各国正在努力整合住户调查方案中的各项调查。现在，许多国家统计局正在将一体化住户调查方案的理念变成现实。

95. 各种不同调查的样本设计是有关调查方案工作的一个重要组成部分。本章探讨了与主抽样标架和主样本的设计与发展有关的各种关键问题，阐释了完好保存主抽样标架的几点好处，并主张每个执行住户调查的国家统计局都应有一个保存完好的主抽样标架，以满足住户调查方案中的多项调查需要，同时也能满足在调查方案期间可能意外进行的某些专项调查的需要。此外，各国家统计局还可以更进一步为调查方案中的所有或大多数调查、或许也为预料之外的专项调查设计和使用一个主样本。

96. 本章综述了在开发主抽样标架和主样本的时候需要采取的重要步骤，并解说了一些发展中国家的主抽样标架和主样本。本章的格局不允许详尽处理所有与开发主抽样标架和主样本有关的重要问题。读者如欲了解更透彻的相关阐述，敬请参阅相关的联合国手册（联合国，1986年）。

参考资料

Ajayi, O. O. (2000年)。“尼日利亚农业抽样人口普查的调查方法与其他国家一些经验的对比”。于2000年9月19日至22日在北京举行的关于中国农业人口普查结果国际论坛上提交的论文。

- Glewwe, P. 和 I. Yansaneh (2001年)。《2002年至2010年期间多种目的住户调查的建议》。《特派团向一般统计科提交的报告, 越南》。
- Kish, L. (1965年)。《抽样调查》。纽约: 约翰·威利国际出版公司。
- _____ (1988年)。 “多种目的样本设计” 。《抽样方法》, 第14期, 第19-32页。
- Pettersson, H. (1994年)。《标准样本设计: 特派团向国家中央统计办公室提交的报告, 纳米比亚, 1994年5月》。瑞典统计局, 国际咨询办公室。
- _____ (2001年a) 《住户和商业调查的样本设计: 特派团向统计局提交的报告, 莱索托, 2001年5月21日至6月2日》。瑞典统计局, 国际咨询办公室。
- _____ (2001年b)。《关于总务干事住户调查标准样品设计的建议: 特派团向一般统计科提交的报告, 越南》。瑞典统计局, 国际咨询办公室。
- Rosen, B. (1997年)。《1997年老挝标准样品的设立: 特派团向国家统计局提交的报告, 老挝人民民主共和国》。瑞典统计局, 国际咨询办公室。
- Torene, R. 和 L. G. Torene (1987年)。 “运用标准样本的实践问题: 孟加拉国的经验” 。《国际统计学会通报: 第四十六次届会的记录, 东京, 1987年》, 第LII-2卷, 第493-511页。
- 联合国 (1986年)。《国家住户调查能力方案: 综合住户调查方案的抽样标架和样本设计 (暂定版本) 》。DP/UN/INT-84-014/5E, 纽约。
- Verma, V. (2001年)。 “国家调查的样本设计: 调查小型经济单位” 。《转型期统计学》, 第5期, 第3号 (2001年12月), 第367-382页。

第六章

供样本设计使用的设计效果要素估算

Graham Kalton
Westat

J. Michael Brick
Westat

Thanh Lê
Westat

美利坚合众国
马里兰州罗克维尔

美利坚合众国
马里兰州罗克维尔

美利坚合众国
马里兰州罗克维尔

摘 要

设计效果——即一个采用复杂样本设计的统计量方差与采用同等规模的简单随机样本或无约束样本的该统计量方差的比率——是抽样设计的一个极为有用的工具。可是，不应自动地采纳某项调查中发现的设计效果并将其用于另一项调查。一种设计效果代表着诸如分层、类集、不均等选择概率、以及对不响应和覆盖面遗漏的加权调整等多种要素的综合效应。不应当简单地从前一次调查引进某种总体设计效果，而应当仔细考虑其中所涉及的各项要素。本章逐一评述每种要素产生的设计效果，然后描述可将这些要素设计效果综合成为某种总体设计效果的各种模型。样本设计者可以根据这些要素为可供选择的样本设计方案构建总体设计效果估计数，然后用这些估计数来指导选择眼下正在策划的一项有效样本设计。

关键词：分层、类集、加权、组内相关系数。

A. 引言

1. 从本出版物的其他章节可以看出，发展中国家和转型期国家的住户调查运用了复杂的样本设计，其中包括多级抽样、分层和往往不均等的选择概率。采用复杂样本设计的一个后果是不能用标准统计教科书中提供的公式来计算调查估计数的抽样误差。那些公式都是以这样一种假设为依据，即：所观察的各种变量均为独立、恒等分布 (*iid*) 的随机变量。但这种假设对采用复杂样本设计获取的观测值不成立，因此需要一种不同寻常的估量调查估计数抽样误差的方法。

2. 基于复杂样本设计的调查估计数方差，可以通过某种形式的复制方法来估算，比如“刀切重复抽样法”或“平衡重复交叉抽样法”；亦可采用一种“泰勒级数线性化”方法估算[譬如见Wolter (1985年)；Rust (1985年)；

Verma (1993年); Lehtonen和Pahkinen (1994年); Rust和Rao (1996年)]。现有许多专用计算机程序可用来进行这种计算[其中不少已在Lepkowski和Bowles (1996年)的文章中做了评述; 另外还可以上网查阅: <http://www.fas.harvard.edu/-stats/survey-soft/iass.html>; 以及由美国统计协会调查研究方法处编写的“调查分析软件概要”, 该文件可以上网查阅: <http://www.fas.harvard.edu/-stats/survey-soft/survey-soft.html>]。如果方差是在考虑到复杂样本设计的情况下计算的, 那么所得到的方差估计数必然不同于采用“独立、恒等分布”(iid)变量的标准公式所得的结果。在许多情况下, 与复杂设计有关的方差要比用标准公式所得的方差为大——而且往往大得有确认价值。

3. 标准统计学教科书中所提供的方差公式适用于一种形式的样本设计, 那就是无约束抽样(又称简单随机带重置抽样)。采用这种设计, 调查人口中的单位是独立选择的, 且概率均等。这些单位用重置法进行取样, 就是说, 一个单位可能在样本中多次重复出现。假定一个 n 规模的无约束样本, 其变量 y 产生 y_1, y_2, \dots, y_n 个值。那么, 该样本平均数的方差 $\bar{v} = \sum y_i / n$ 即为:

$$V_u(\bar{y}) = \sigma^2 / n \quad (1)$$

式中, $\sigma^2 = \sum^N (Y_i - \bar{Y})^2 / N$ 是在人口(Y_1, Y_2, \dots, Y_N)中 N 个 y -值的要素方差, 而 $\bar{Y} = \sum Y_i / N$, 可用下式从样本中估算该方差:

$$v_u(\bar{y}) = s^2 / n \quad (2)$$

式中, $s^2 = \sum^* (y_i - \bar{y})^2 / (n-1)$ 。在标准的统计学教科书中可以找到同样一些公式。

4. 按照规则, 调查样本的选择是不带重置(而不是带重置)的, 因为, 如果只允许将有关单位纳入样本一次的话, 调查估计数就会比较精确(就是说, 他们的方差较低)。采用不带重置的简单随机抽样法——通常叫作“简单随机抽样”, 或称SRS, 系按照均等概率选择单位, 而所有可能从 N 个单位的人口中提取的 n 个成套不同单位都有同样的可能性构成该样本。若采用一个 n 规模的SRS, 则样本平均数 $\bar{y} = \sum y_i / n$ 的方差或方差估计数可由下面两式给出:

$$V_0(\bar{y}) = (1-f)S^2 / n \quad (3)$$

和

$$v_0(\bar{y}) = (1-f)s^2 / n \quad (4)$$

式中, $f = n/N$ 是抽样部分, $S^2 = \sum^N (Y_i - \bar{Y})^2 / (N-1)$ 而 $s^2 = \sum^* (y_i - \bar{y})^2 / (n-1)$ 。如果 N 值很大(一般调研情况均如此), σ^2 和 S^2 便大致相等。因此, 方程式

(1)中的无约束抽样平均数的方差和方程式(3)中的平均数方差之间的差异就是因子 $(1-f)$, 称之为“有限总体校正”(fpc)。在大多数实际情况中, 抽样因子 n/N 很小, 可按0处理。若如此处理, 方程式(3)和(4)中的fpc项约为1, 而带与不带重置的抽样差别可以忽略不计。

5. 上面给出的方差公式对复杂样本设计不适用,但它们确实可以充当基于复杂设计的估计数方差的比较基准。Kish (1965年) 杜撰了“设计效果”一词,用以代表从复杂设计所得的任何估计数,譬如 z 的方差与采用同样规模的SRS或无约束样本的估计数 z 之间的比率。¹⁸ 请注意,设计效果与一特定调查估计数 z 有关,而对于特定调查中的不同估计数来说则各不相同。另外请注意, z 可以是所关注的任何估计数,比如平均数、比率、总数或回归系数。

6. 设计效果既取决于所采用的复杂样本设计的形式也取决于所考虑的调查估计数。为了同时包含这两种特征,我们用 $D^2(z)$ 这个计数法来表示估计数的设计效果,式中,

$$D^2(z) = \frac{\text{复杂设计估计数 } z \text{ 的方差}}{\text{同样规模无约束样本估计数 } z \text{ 的方差}} = \frac{V_c(z)}{V_u(z)} \quad (5)$$

在这个计数法中使用了平方根,以便把 $D(z)$ 当作设计效果的平方根使用。这个的简单计数法很有用,因为它代表了行将应用于无约束样本设计中的标准误差的倍数,以给出它在复杂设计中(比如置信区间的计算)的标准误差。

7. 一个直接与设计效果有关的有用 n_{eff} 概念是“有效样本规模”,用来表示。所谓有效样本规模,就是一个无约束样本能够产生同复杂设计一样的那种调查估计数精确水准所应具有规模。因此,有效样本规模可由下式给出:

$$n_{eff} = n / D^2(z) \quad (6)$$

8. 上面给出的 $D^2(z)$ 的定义是为理论工作提供的,其中的真实方差 $V_c(z)$ 和 $V_u(z)$ 都是已知的。在实际应用当中,这些方差是根据样本估算的,然后,根据 $d^2(z)$ 来估算 $D^2(z)$ 。于是,

$$d^2(z) = \frac{v_c(z)}{v_u(z)} \quad (7)$$

式中, $v_c(z)$ 是用一种适合复杂设计的程序估算的;而 $v_u(z)$ 是用一个适合无约束抽样的公式估算的,其中未知的参数根据样本来估算。因此,比如说在样本平均数的情况中,

$$v_u(z) = s^2 / n \quad (8)$$

而在较大的样本中, s^2 可以用下式来估算:

$$\frac{\sum w_i (y_i - \bar{y})^2}{\sum w_i}$$

式中, y_i 和 w_i 分别是被抽样单位 i 的 y -值和权数;而 $\bar{y} = \sum w_i y_i / \sum w_i$ 则是人口平均数的加权估计数。在样本比率为 p 的情况中,对于较大的 n 值,

$$v_u(p) = \frac{p(1-p)}{n-1}$$

或者

¹⁸ 更确切地说, Kish (1982年)将 $Deff$ 定义为与一个以SRS方差为分母的比率,而将 $Deft^2$ 定义为以无约束样本方差为分母的比率。 $Deff$ 和 $Deft^2$ 之间的差异取决于是否包括fpc项的 $(1-f)$ 。鉴于该项在大多数国家住户调查中可以忽略不计,故 $Deff$ 和 $Deft^2$ 之间的差异几乎没有实际意义,因此在本章的其余篇幅中一律予以忽略。我们通篇假定fpc项可以忽略不计。另见Kish (1995年)。

$$v_a(p) = \frac{p(1-p)}{n}$$

式中， p 是人口比率的加权估计数。

9. 在界定设计效果和估计设计效果的时候，还有一个问题需要探讨。许多调查使用不均等选择概率的样本设计，在这种情况下，样本中的子群体的代表性可能就不成比例。譬如讲，在一个国家住户调查中，一个2 000住户的样本可能有50%选自城市地区，另外50%选自农村地区，而总人口当中只有30%的住户在城市地区。现在来考虑譬如讲城市住户估计平均数的设计效果。根据方程式(8)，分母是 s^2/n 。问题是如何计算 n 值。一种方法是使用实际的城市样本规模，这里是1 000。另一种可供选择的方法是使用 $n = 2 000$ 的简单随机抽样(SRS)的预期城市样本规模，这里就是 $0.3 \times 2 000 = 600$ 。以实际规模1 000为条件的第一种方法是最常用的方法，也是本章将采用的方法。不过，以第二种方法为基础计算设计效果的那种选择也可以用在某些估量方差的方案中。鉴于这两种方法所产生的数值可能差别很大，所以有必要了解二者的区别并做出适当的选择。

10. 现已证明设计效果概念是设计复杂样本的一个极为有用的工具。复杂的设计牵涉到多种设计要素的组合，比如分层、多级抽样、按不均等概率选择样本，等等。逐一分析每种要素的设计效果，最好要强调它们对调查估计数精确度的影响如何，这样才有利于指导有效样本设计的发展。我们将在B节评述各个要素的设计效果。在设计复杂样本的时候，一个有益的做法是构建多种要素的组合，以形成整个设计效果所需要的预测模型。我们将在C节简要评述这些模型。关于如何在样本设计中运用设计效果，我们将在D节通过一个虚拟的实例来加以说明。最后，E节总结了几点概括性的意见。

B. 设计效果诸要素

11. 本节考虑复杂样本设计的如下各种要素所产生的设计效果，即：按比例和非比例的分层；类集；不均等选择概率；对不响应的样本加权调整、对未覆盖的人口加权调整、以及为提高精确度的加权调整。本节分别对这些不同的要素进行考察；它们的合成效应在C节讨论。所考虑的主要统计量是人口平均数 \bar{Y} 的一个估计数（比如平均收入）。鉴于人口比率 P （比如贫困人口比率）实际上是算术平均数的一个特例，所以在处理上也包括一项比率。比率问题大概是在调查报告中用最广泛的统计项，所以将在适当时单独讨论。许多调查成果与总人口的子群有关，比如年龄在15-44岁的妇女，或者农村地区的人口。因此将讨论到加权和类集对子群设计效果的影响问题。

1. 分层

12. 我们首先来考虑在分层的单级样本平均数与层次范围内进行简单随机抽样的样本平均数有关的设计效果。分层样本平均数由下式给出：

$$\bar{y}_x = \sum_h \frac{N_h}{N} \sum_i \frac{y_{hi}}{n_h} = \sum_h W_h \bar{y}_h$$

式中, n_h 是从 h 层的 N_h 个单位中选取样本的规模; $N = \sum N_h$ 是人口规模; $W_h = N_h / N$ 是 h 层的人口比率; y_{hi} 是 h 层中被抽样单位 i 的数值; 而 $\bar{y}_h = \sum_i y_{hi} / n_h$ 是 h 层的样本平均数。在实践中, \bar{y}_x 是作为加权估计数计算的, 其中, 给每个抽样单位指定一个基权数, 即其选择概率的倒数 (暂时忽略样本和人口加权调整)。这里, h 层的每个单位都有一个选择概率 n_h / N_h , 因此基权数 $w_{hi} = w_h = N_h / n_h$ 。这样, \bar{y}_x 即可表示为:

$$\bar{y}_x = \frac{\sum_h \sum_i w_{hi} y_{hi}}{\sum_h \sum_i w_{hi}} = \frac{\sum_h \sum_i W_h y_{hi}}{\sum_h n_h W_h} \quad (9)$$

假定人口项可以忽略不计, 则分层平均数方差可由下式给出:

$$V(\bar{y}_x) = \sum_h \frac{w_h^2 S_h^2}{n_h} \quad (10)$$

式中, $S_h^2 = \sum_i (Y_{hi} - \bar{Y}_h)^2 / (N_h - 1)$ 是 h 层范围内的人口单位方差。

13. $V(\bar{y}_x)$ 数值的大小取决于样本在各层的分配方式。通常采取按比率分配方式, 因此, 一个层次中的样本规模与该层的人口规模成正比, 所有抽样单位的权数都一样。分层平均数降至简单未加权平均数 $\bar{y}_{pnp} = \sum y_{hi} / n$, 式中, $n = \sum n_h$ 是整个样本规模, 其方差降至

$$V(\bar{y}_{pnp}) = \frac{\sum W_h S_h^2}{n} = \frac{S_w^2}{n} \quad (11)$$

式中, S_w^2 表示层内单位平均方差。然后, 使用根据方程式 (3) 计算的一个简单随机样本的平均数方差, 得出按比例分层样本 \bar{y}_{pnp} 的设计效果, 其中忽略了 fpc 项, 并对方程式 (5) 表示的设计效果定义如下:

$$D^2(\bar{y}_{pnp}) = \frac{S_w^2}{S^2} \quad (12)$$

由于层内单位平均方差并不比整个样本方差大 (条件是 N_h 的值要大), 所以按比例样本平均数的设计效果不会大于 1。可见, 按比例分层不可能导致精确度降低, 一般来说反倒会在某种程度上提高精确度。如果各层的平均数 \bar{Y}_h 有差异, 就能提高精确度: 各层平均数差异越大, 提高精确度的幅度就越大。

14. 在许多调查中需要非比例分层的样本, 以便能够通过调查提供特定“域”的估计数。譬如讲, 调查的目标之一可能是生产一个国家各个地区的可靠估计数, 而各地区的人口可能不尽相同。为了实现这一目标, 可能就有必要给较小地区分配相对于按比例分层抽样来说规模上要大得多的样本。各不同层次之间在收集数据方面的成本相差悬殊, 这或许是不搞按比例分配的另一个

原因。在这种情况下，对数据收集成本较低的层次分配大于按比例的本规模，应该是一种比较理想的设计。

15. 按比例分层在数据精确度方面的所得未必适用于非比例的样本分配。为了简化在这个问题上的讨论，我们假定层内人口方差是固定不变的，换言之，所有层次一律按 $S_h^2 = S_c^2$ 处理。在国家住户调查中，当基于上述理由采取非比例分层的做法时，这一假定往往是合理的。按照这一假定，方程式（10）就简化成为

$$V(\bar{y}_{st}) = S_c^2 \sum_h \frac{W_h^2}{n_h} = \frac{S_c^2}{N} \sum_h W_h w_h \quad (13)$$

在这种情况下设计效果为：

$$D^2(\bar{y}_{st}) = \frac{S_c^2}{S^2} \frac{n}{N} \sum_h W_h w_h \quad (14)$$

16. 除了在推导方程式（14）中所使用的层内方差不变的假定之外，假定层内各种平均数大体均等也是有道理的，就是说，所有层次一律按 $\bar{Y}_h = \bar{Y}$ 处理。按照这进一步的假定， $S_c^2 = S^2$ 和设计效果便降至：

$$D^2(\bar{y}_{st}) = \frac{n}{N} \sum_h W_h w_h = n \sum_h \frac{W_h^2}{n_h} \quad (15)$$

¹⁹ 这份参考资料以非常有用的方式总结了大量相关成果。其中有许多关系已经众所周知，并且在几十年前就发表了。譬如，见Kish(1965年)和Kish(1976年)。

Kish (1992年)¹⁹将非比例分配所产生的设计效果表示为：

$$D^2(\bar{y}_{st}) = (\sum_h W_h w_h) (\sum_h W_h / w_h) \quad (16)$$

这个公式对样本设计很有用。但是不应该不加区分地应用这个公式，而无视其基本假定的合理性（见下文）。

17. 现在举一个简单例子来说明方程式（16）。设某国有两个地区：第一个地区人口占总人口的80%，第二个地区占总人口的20%（即 $W_1=4W_2$ ）。假定在进行调查中这两个地区的样本规模平均分配（ $n_1=n_2=1\ 000$ ）。上面提供的任何一种表述公式均可用于计算非比例分配的全国平均数估计数的设计效果（假定两地区的平均数和单位方差是一样的）。譬如讲，采用方程式（16）并注意到 $w_1=4w_2$ ，设计效果即为：

$$D^2(\bar{y}_{st}) = (4W_1 \cdot 4w_1 + W_2 \cdot w_2) \left(\frac{4W_2}{4w_2} + \frac{W_1}{w_1} \right) = 1.36$$

这是因为 $W_2 = 0.2$ 。为了达到两个地区大体相等的估计数精确度，采用了非比例分配，结果是：在有效样本规模为 $n_{eff} = 2\ 000 / 1.36 = 1\ 471$ 的情况下得出了全国估计平均数。

18. 表六.1展示了在只有两个层次的情况下，按照某些常见的过高抽样率进行非比例分配所产生的设计效果。每一栏开头的数字表示两个层次的权数比率，即相当于两个层次中抽样率之比的倒数。左边存根项是第一层内的各种人口比率。由于设计效果对称，约为0.50，故通过使用对应于 $(1-W_1)$ 的那

一行可以获得 $W_1 > 0.5$ 的数值。为了说明该表的用法，请考虑上面举的例子。在 $W_1 = 0.20$ 的那一行、过度抽样率为4的那一栏，给出的数值是 $D^2(\bar{y}_x) = 1.36$ 。从该表可以看出，设计效果伴随抽样率的增长而增长，层次内的人口比率接近50%。当各层的抽样率相差悬殊的时候，整个平均数的设计效果可能会很大，因而有效样本规模很小。在此案例中，非比例分配导致形成一个对于估计整个人口统计数非常无效的样本。

19. 许多国家的调查意在产生全国估计数以及国内各地区的估计数。通常，各地区的规模差别很大。在这种情况下，正如上述结果所显示的，在确定各地区的适当样本分配的时候就会发生冲突。按照各地区内部平均数均等和单位方差均等的假定，全国估计数的最佳分配是按比例分配；而对于地区估计数来说，最佳分配就是每个地区内的样本规模均等。为一种目的采用的最佳分配可能为别的目的产生很糟糕的样本。不过，有一种折衷的分配或许对两种目的都比较适合（见D节）。

表六.1
在两层示例中非比例抽样所产生的设计效果

W_1	W_1 对 W_2 的比率							
	1	2	3	4	5	8	10	20
0.05	1.00	1.02	1.06	1.11	1.15	1.29	1.38	1.86
0.10	1.00	1.05	1.12	1.20	1.29	1.55	1.73	2.62
0.15	1.00	1.06	1.17	1.29	1.41	1.78	2.03	3.30
0.20	1.00	1.08	1.21	1.36	1.51	1.98	2.30	3.89
0.25	1.00	1.09	1.25	1.42	1.60	2.15	2.52	4.38
0.35	1.00	1.11	1.30	1.51	1.73	2.39	2.84	5.11
0.50	1.00	1.13	1.33	1.56	1.80	2.53	3.03	5.51

20. 方程式（16）在样本设计中广泛用于评估对全国估计数采用非比例分配的效应。不过，用户在使用这个方程式的时候应注意如下假定条件，即：层内平均数均等和该层所依据的方差均等。首先考虑这样一种情况：平均数各异但方差均等。在这种情况下，非比例分层的设计效果可用方程式（14）来计算，并附加因子 $S_x^2 = S^2$ 。该因子小于1，因此设计效果不如用方程式（16）计算的那么大。但是这个设计效果代表了这种分层及这种非比例分配的整体效应。若想仅计量非比例分配的效应，则应在同等规模的非比例分层样本与按比例分层样本之间做适当的比较。根据方程式（11）和（13），且 $S_w^2 = S^2$ ，非比例设计的方差 \bar{y}_x 与按比例设计的方差 \bar{y}_{prop} 之间的比率是：

$$R = V(\bar{y}_x) / V(\bar{y}_{prop}) = (\sum_k W_k w_k) (\sum_k W_k / w_k)$$

因此，在这种情况下，可将方程式（16）解释为只是非比例分配的效应。

21. 关于层内单位方差均等的假定比较关键。上面的结果表明，在层内单位方差均等的情况下，非比例分配会导致总体估计数精确度的损失，但是在层内单位方差不均等的情况下却未必如此。的确，在层内方差不均等的

情况下，所采用的最佳抽样比率与层次标准方差成正比[譬如，见Cochran（1977年）]。这种类型的非比例分配广泛用于工商企业调查。在层内标准偏差明显不同的情况下，它可以导致按比例分配调查估计数的精确度明显提高。

22. 在住户调查中，假定层内方差均等或大致均等往往是合理的。一种层内方差可能不均等的估计数就是一个比率。一个比率是一种数值只取1和0、相当于有或没有特定特征的变量的平均数。这样一个变量的单位方差是 $\sigma^2 = P(1-P)$ ，式中， P 是有这种特征的人口比率。于是，在 h 层中有这种特征的比率 P_h 便是 $S_h^2 = P_h(1-P_h)$ 。如果 P_h 在各层之间有差异， S_h^2 也有差异。但是对0.2和0.8的比率来说， S_h^2 只是略有差异：高者对 $P_h = 0.5$ 来说差异为0.25；低者对 $P_h = 0.2$ 或0.8来说差异为0.16。

23. 为了说明层次比率差异性及其方差的效果，让我们回到前面举出的双层范例，其中 $W_1 = 0.8$ ， $W_2 = 0.2$ ，而 $n_1 = n_2$ ，并考虑 P_1 和 P_2 的两套不同数值。在案例1中，设 $P_1 = 0.5$ ， $P_2 = 0.8$ 。那么，通过用方程式（10）和（1）计算，总体设计效果即为 $D^2(\bar{y}_x) = 1.35$ ，而非比例和按比例两种设计方差的比率为 $R = 1.43$ 。在案例2中，设 $P_1 = 0.8$ ， $P_2 = 0.5$ 。则结果是 $D^2(\bar{y}_x) = 1.16$ ， $R = 1.26$ 。在这两个示例中所获得的 $D^2(\bar{y}_x)$ 和 R 的值可以同按照层内方差均等的假定获得的设计效果值1.36做比较。在这两个案例中，由于从分层取得精确度效益，各自的总体设计效果都小于1.36。在案例1中， R 值大于1.36，因为在抽样率较低的第一层，层内方差比较大。在案例2中情况恰好相反：因为第二层过度抽样，所以层内方差比较大。因此，这种过度抽样是符合提高精确度的要求的。实际上，在后一个案例中最佳分配应该是在第二层按照相当于第一层1.25倍的抽样率进行抽样。在这些例证中，即便层次比率相差很大，因而层内方差也有不小的出入，但是所得的 R 值——分别为1.26和1.43——却合理地接近1.36。这些计算结果说明了一个事实，即：根据方程式（16）计算得出的加权设计效果的近似值测定数对于大多数规划目的是恰当的——即使层内方差在一定程度上有出入。

24. 最后让我们考虑一个比较极端的例子：设 $P_1 = 0.05$ ， $P_2 = 0.5$ ；仍然设， $W_1 = 0.8$ ， $W_2 = 0.2$ ， $n_1 = n_2$ 。在此例中， $D^2(\bar{y}_x) = 0.67$ ， $R = 0.92$ 。这个例证说明，非比例分层可以产生精确度效益。可是，鉴于其所依据的假定条件，方程式（16）不能产生小于1的数值。由此可见，方程式（16）不应当在无视其基本假定条件的情况下不加区分地任意应用。

2. 类集

25. 现在我们来考虑大多数人口调查中总体设计效果的另外一个重要要素，亦即多级样本中的类集产生的设计效果。类集样本的目的是减少收集数据的成本，因为对稀疏散布在全国的或整个地区的住户进行列表和抽样是划不来的。一般的做法是采取两级或多级抽样方法，第一阶段初级抽样单位，即初级抽样单位明显界定了总体抽样的地理区域，抽样率与各区域所包含的估计住户或人口数量成正比。在选定的初级抽样单位范围内，可以进行一个或多个追加

区域抽样阶段，然后列出最终选取的次区域居住单位清单，并对清单中的住户进行抽样。对住户的调查，所收集的是被抽样住户的数据。对个人的调查，先编制一份选定住户的个人名单，从中选取符合调查条件的所有个人或个人样本。为了本节讨论的目的，我们假定住户调查只有两级抽样（即初级抽样单位抽样和住户抽样）。但是可以直接向多级拓展。

26. 在实际情况中初级抽样单位在规模上（即所包含的单位数量）是在不断变化的，因此，初级抽样单位是按照“与估计规模成比例的概率”

（PPES）进行抽样的。一般来说，从选定的初级抽样单位选取的样本彼此之间在规模上也是不尽相同的。但为了简便起见，我们先假定有关人口由初级抽样单位（ A ）（譬如普查查点区）组成，其中每个单位包含若干住户（ B ）。在每个选定的初级抽样单位中分别选取初级抽样单位 a 的一个简单随机样本和住户 $b \leq B$ 的一个简单随机样本（ $b=B$ 的特例代表单级类集样本）。我们假定一级有限总体校正因子可以忽略不计。针对选择住户的样本设计使用“等概率抽选法”（epsem），以使用简单不加权样本平均数 $\bar{y}_a = \sum_{\alpha} \sum_{\beta} y_{\alpha\beta} / n$ 来估算人口平均数，式中， $n = ab$ ，而下标 cl 表示类集。方差 \bar{y}_a 可以写作

$$V(\bar{y}_a) = \frac{S^2}{n} [1 + (b-1)\rho] \quad (17)$$

式中， S^2 是人口中的单位方差，而 ρ 是计量初级抽样单位中 y -变量同质性的组内相关系数。在实际当中，几乎对所有变量来说初级抽样单位范围内的各单位都或多或少彼此近似，尽管其近似程度通常很低。因此， ρ 值几乎总是正数，而且很小。

27. 在这种简单状况中，设计效果是：

$$D^2(\bar{y}_a) = 1 + (b-1)\rho \quad (18)$$

这一基本结果表明，在初级抽样单位中类集样本所产生的设计效果取决于两个因素：一是选定的初级抽样单位中子样本的规模（ b ）；二是组内相关（ ρ ）。因为 ρ 值一般为正数，所以类集所产生的设计效果值大于1。

28. 方程式（18）以及下面提供的其他方程式——的一个重要特征，就是该方程式取决于作为初级抽样单位中特定变量同质性测定值的 ρ 。²⁰对许多变量来说（比如年龄和性别）， ρ 值接近于零，量值虽小（譬如， $\rho = 0.03$ 至 0.05 ），但对其他变量来说却不容忽略；同时，对于某些变量来说它又可能很高（比如，在所有村民要么有机会、要么没有机会看病的情况下，到村里——作为一个初级抽样单位——的诊所看病的机会）。理论上 ρ 值为负数是可能的，但实际当中不大可能遇到（尽管 ρ 值的样本估计数往往是负的）。 ρ 值往往与初级抽样单位逆相关，因为一般情况下，类集量越大，就越趋多样化，尤其当初级抽样单位是地理区域的时候。在选择使用大而多样化的初级抽样单位的情况下，最佳调查设计就利用上述这些关系。在策划样本设计的时候，一些关键性的变量需要 ρ 值的估计数。通常，这些估计数系以从前历次调查的相

²⁰ 本节的讨论采用规模均等和非均等类集的层内同质性量数。

同或类似变量及初级抽样单位为基础，同时也基于对 ρ 值在各种类似变量和初级抽样单位之间可移植性的依赖。

29. 实际上，初级抽样单位并非规模均等，而且也不是采取简单随机抽样方法。在大多数国家住户调查的样本设计中，都采用PPES（与估计规模成比例的概率）抽样法来选择初级抽样单位的分层样本，因此不能直接应用方程式（18）。不过，只要在值的解释上稍做适当修改，对于多种多样的epsem（等概率抽选法）的样本设计来说，该方程式依然是计算类集设计效果的一个有用的模型。

30. 首先让我们考虑在已知确切规模量数的情况下，初级抽样单位中的PPS（概率与规模成正比）样本。在这种情况下， a 个初级抽样单位的一个PPS样本和 b 个住户的一个epsem（等概率抽选法）样本相结合，即可产生一个epsem总体设计。采用这种设计，方程式（18）依然成立，但现在 ρ 被解释为由这种子样本设计生成的最终类集范围内同质性的一个合成量数（Kalton, 1979年）。例如，对于一个通过系统抽样选择 b 个住户的子样本设计，和一个把每个抽样的初级抽样单位分成若干各包含 b 个住户的次区域并从中选取一个次区域的子样本设计来说，两者的 ρ 值是不同的（后者的 ρ 值可能比较大）。因此，这一拓展既涉及到PPS抽样，又涉及到子样本设计的各种可供选择的形式。

31. 现在来考虑初级抽样的分层问题。Kalton（1979年）说明，在一个总体epsem设计中——其中，选定 a 个初级抽样单位的一个分层样本，并在每个选定的初级抽样单位内按照均等概率对 b 个基本单位进行抽样——可以通过下式计算出由类集产生的设计效果的近似值：

$$D^2(\bar{y}_i) = 1 + (b-1)\bar{\rho} \quad (19)$$

式中， $\bar{\rho}$ 是层内同质性的平均量数，其假定条件是每个层次内的同质性量数大体差不多。在 b 值相当大的情况下，初级抽样单位的有效分层可以产生实质性的效益，因为方程式（18）中的总体同质性量数已被方程式（19）中较小的层内同质性量数所取代。换句话说，如果 b 值相当大的话，就可以大大减少因为初级抽样单位分层抽样所产生的 $(b-1)(\rho - \bar{\rho})$ 设计效果。

32. 以上，我们一直假定这么一个总体epsem的样本：在每个选定的初级抽样单位中样本规模均等，即 b 。这些条件只有在下面两种情况下才能得到满足，即：按照均等的概率选取规模均等的初级抽样单位样本；严格按照PPS（概率与规模成正比）的抽样方法选取同等规模的初级抽样单位样本。但在实际当中，这两种情况一个都不适用。实际情况是按照PPES（与估计规模成比例的概率）选取规模不均等的初级抽样单位样本，其中的估计规模量数在一定程度上是不准确的。在这种情况下，在抽取的初级抽样单位中应用子样本抽样率来做总体epsem设计，就会导致子样本规模的某些差异。只要子样本规模差异

不是太大，仍然可以用方程式（19）来算出近似值，而用子样本规模的平均数来取代 b ，亦即：

$$D^2(\bar{y}_d) = 1 + (\bar{b} - 1)\bar{\rho} \quad (20)$$

式中， $\bar{b} = \sum b_a / a$ ，而 b_a 是初级抽样单位 a 的基本单位数。方程式（20）业经证明对每个初级抽样单位中的抽样单位数相对稳定的情况极具实用价值。

33. 不过，如果每个初级抽样单位中的子样本规模差异很大的话，方程式（20）中的约计法就不适用了。Holt（1980年）将上述约计法加以拓展，通过用一个经过加权的平均子样本规模取代方程式（20）中的 \bar{b} 来处理不均等的子样本规模。不均等类集规模的类集所产生的设计效果可以写为：

$$D^2(\bar{y}_d) = 1 + (b' - 1)\bar{\rho} \quad (21)$$

式中， $b' = \sum b_a^2 / \sum b_a$ （ b' 的量值可视为加权平均数， $b' = \sum k_a b_a / \sum k_a$ ，其中 $k_a = b_a$ ）。如上所述，约计法适用于总体epsem（等概率抽选法）样本设计。

34. 举例来讲，假设有五个被抽样的初级抽样单位，其各自的规模分别为10、10、20、20和40个住户，同时假定 $\bar{\rho} = 0.05$ 。平均子样本规模为 $\bar{b} = 20$ ，而 $b' = 26$ 。这样，此例中采用约计法方程式（20）的类集所产生的设计效果即为1.95；相比之下，采用约计法方程式（21）所得的设计效果值为2.25。

35. Verma, Scott和O'Muircheartaigh（1980年）和Verma和Lê（1996年）提供了另外一种写出这种调整的方法，该法适用于不同“域”（比如城市域和农村域）的子样本规模相差悬殊的情况。假设两个域：设其中一个域在每个选定的初级抽样单位 a_1 中抽取 b_1 个住户样本，其中 $n_1 = a_1 b_1$ ；在另一个域中，从其余的 a_2 选定的初级抽样单位中抽取 b_2 个住户样本， $n_2 = a_2 b_2$ 。然后得出下式：

$$b' = (n_1 b_1 + n_2 b_2) / (n_1 + n_2)$$

36. 前面的讨论考虑了总人口平均数（和比率）的估计数类集所产生的设计效果。其中大多数处理方法同样适用于子样本估计数，但条件是必须仔细留意相关的假设条件。根据子样本在初级抽样单位间的分配情况，将其分成三种类型是有益的。一端是在初级抽样单位间平均分布的子群，叫作“交叉类”，比如年龄/性别子群一般属于交叉类。另一端是这样一些子群，其中每个子群集中于一个初级抽样单位子集，叫作“分割类”，城市子群和农村子群一般属于这一类。介于这两类之间的是基本上按初级抽样单位集中的，叫作“混合类”。

37. “交叉类”是按总样本在初级抽样单位间的分布。如果总样本在初级抽样单位间的分布相当均匀，那么方程式（20）可用于计算类集产生的设计效果的近似值，同时该方程式亦可用于交叉类。可是当它用于交叉类的时候有一个重要变化：现在 \bar{b} 代表每个初级抽样单位交叉类子样本的平均规模。由于这种变化，交叉类估计数的设计效果要比总样本估计数的设计效果小。

38. “分割类”构成整个样本中初级抽样单位的一个子集中的所有单位。因为一个分割类的子集样本规模跟初级抽样单位该子集总样本的规模一样，所以一般来说没有理由期望一个分割类估计数的设计效果会比总样本估计数的设计效果低。只有在该分割类中的每个初级抽样单位的平均子样本规模不同于总样本中平均子样本规模的情况下，或者，只有在同质性有差异的情况下（比如，其中包括由于在分割类中子样本设计不同致使合成 ρ 值有差异等），分割类的一个估计数的设计效果才会与总样本估计数的设计效果有所不同。如果总样本均匀分布在初级抽样单位中，仍然可以应用方程式（20），这时的 \bar{y} 和 ρ 均为分割类中全套初级抽样单位的数值。

39. “混合类”在初级抽样单位间的不均衡分布意味着方程式（20）在这种情况下不适用。为了估计混合类的一个估计数类集所产生的设计效果，可以使用方程式（21），其中 b_a 是初级抽样单位 a 中混合类抽样成员数。

3. 加权调整

40. 正如题为“分层”的B.1节所讨论的，采用非比例分层法的各层间不均衡选择概率需要在分析调查数据的时候使用权数。方程式（15）和（16）在假定各层的平均数和单位方差都均等的条件下，可以给出非比例分层及由此决定的不均等权数所产生的设计效果。现在我们来讨论这些比较适合在分析阶段确定权数效应的可供选择的公式形式。不过，我们首先提请注意在调查分析中使用各种不同权数的必要性[另见Kish（1992年）]。首先，正如我们在前面所特别指出的，之所以在分析中需要各种权数，是因为要补偿与非比例分层有关的不均衡选择概率。从更广泛的角度来说，它们是为了补偿因任何缘故导致的不均衡选择概率。补偿不均衡选择概率的权数是选择概率的倒数，通常叫作基权数。然后基权数往往经过调整后用于补偿不响应的情况，并使得加权的样本总数与已知的人口总数相一致。因此，最终的分析权数几乎总是在一定程度上有差异。

41. 即便对某些域过度抽样，通常样本设计也会偏离epsem，因为存在着各种范围问题。譬如讲，如果按均等的概率从一个住户范围选出了若干住户，然后在每个选定的住户中随机选择一位家庭成员，那么，家庭成员就是按照不均等的概率抽样的，因此就需要在分析时使用权数予以补偿。于是这些权数就产生了下面讨论的设计效果要素。顺便提一下，通过将所有选定住户成员纳入样本，这种加权效应是可以避免的。然而这种做法会开启另外一个类集阶段，从而增加了因住户成员的多种特征雷同而产生的类集效应。[关于在抽样住户中选择个人的这些备选方法所产生的设计效果，见Clark和Steel（2002年）。]

42. 因为范围问题而产生的另外一种常见的非epsem设计情况，就是采取两级样本设计，而按照与估计规模成比例的概率（PPES）对初级抽样单位进行抽样。如果规模量数比较准确，那么，在总体epsem设计中每个选定的初级抽样单位的样本规模对所有初级抽样单位都是大致相同的。可是，如果对某个选

定的初级抽样单位的规模严重估计不足的话，epsem设计就需要从该初级抽样单位抽取比平均单位数多得多的单位。鉴于收集如此大量的调查数据通常是不切实际的，因此可能抽取小一些的样本，这就导致了不均等的选择概率并需要带有补偿性的权数。

43. 几乎所有调查都会在一定程度上遇到不响应的情况。减少可能出现的不响应偏差的通常做法，就是可微分地调整受访者的基权数。具体做法包括鉴别样本中有不同响应率的各个子群，并且用相关子群的响应率倒数使每个子群的受访者权数膨胀（Brick和Kalton, 1996年）。这些加权调整的结果会使得权数偏离基权数，其效果往往是增强某个估计数的设计效果。

44. 如果有其他来源的有关人口的信息可用，即可对经过不响应调整的权数进一步加以调整，以便使加权样本估计数与人口信息相一致。譬如讲，如果有外部来源的较准确的区域人口估计数可用，可以使这些区域人口的样本估计数与外部估计数相吻合。这种人口加权调整往往通过一种分层后的调整来进行。它可以帮助补偿覆盖面遗漏并可以提高某些调查估计数的精确度。可是它也进一步增加了权数的多样性，这会对调整中所使用的与人口无关的调查估计数的精确度产生负面影响。

45. 在上述背景下，我们现在来考虑非比例分层设计效果的一般化，以评估各种权数的一般效应。Kish（1992年）提出了另外一种表示分层平均数的设计效果的方法，这对于在分析阶段计算非比例分层效应非常有用。下面的方程式只是对方程式（15）和（16）的一种不同的表述，因此基于“层次平均数均等”和“单位方差均等”的同样假定条件。因为它是根据样本计算的，所以设计效果指定为 $d^2(\bar{y}_x)$ ，并且

$$d^2(\bar{y}_x) = \frac{n \sum_h \sum_i w_{hi}^2}{(\sum_h \sum_i w_{hi})^2} = 1 + cv^2(w_{hi}) \quad (22)$$

式中， $cv(w_{hi})$ 是权数变异系数： $cv^2(w_{hi}) = \sum \sum (w_{hi} - \bar{w})^2 / n \bar{w}^2$ ；而 $\bar{w} = \sum \sum w_{hi} / n$ 是权数的平均数。

46. 上面方程式的一个更一般的形式是：

$$d^2(\bar{y}_x) = \frac{n \sum_j w_j^2}{(\sum_j w_j)^2} = 1 + cv^2(w_j) \quad (23)$$

式中，样本的 n 个单位个个都有自己的权数 w_j ($j=1, 2, \dots, n$)。方程式（23）给出的不均等加权所产生的设计效果取决于“这些权数与调查变量无关”的假定条件。如果该方程式的基本假定条件至少大体上成立的话，那么它就可以对不均等选择概率提供一个不均等加权效应的合理量数[关于在选择概率与调查变量相关的情况下一约计法设计的效应，见Spencer（2000年）]。

47. 不响应调整一般是在响应受访者和不响应受访者两方面都已知的辅助变量所界定的类别中进行的。为了有效减少不响应偏差，在调查中测得的各种

变量的确需要在这些加权类别之间有所差异。但一般来说差异不是很大——尤其在单位方差方面。因此，方程式（23）广泛用于检查不响应加权调整对调查估计数精确度的影响。这种检查可以通过只用基权数或者用不响应调整权数计算方程式（23）来进行。如果后者的计算得出的数值远远大于前者的计算结果，这就意味着不响应加权调整给调查估计数的精确度带来了实质性的损失。在这种情况下，最好通过推翻加权类别或者削减极大的权数来修改加权调整数，从而以减少精确度损失。

48. 虽然方程式（23）对于大多数不响应样本加权调整来说是合理的，但是它往往不能产生一个较好的有关人口加权调整效应的近似值。特别是，如果将权数后分层或校正到从外部来源已知的控制总数的话，那么，当 y 值的平均数设计效果与于一个或多个控制总数高度相关的时候，应用方程式（23）计算的 y 平均数设计效果近似值就很不可靠。譬如讲，假定将权数后分层到一个国家按性别划分的人口控制总数。现考虑将调查数据用于估计妇女在总人口中所占的比率的极端情况。在这个 y 变量与控制变量完全相关的情况中，估计的比率不受抽样误差的影响，因此方差为零。实际当中，不可能是完全相关，但对某些调查变量来说相关程度可能相当高。如果相关程度相当高的话，后分层或校正到已知的人口控制数就可以使调查估计数的精确度有相当可观的改进，但是通过使用方程式（23）是看不到这种改进的。相反，方程式（23）所显示的将是精确度方面的损失。

49. 以上讨论说明，不应该用方程式（23）去估算对基于与控制变量密切相关变量的估计数进行人口加权调整所产生的设计效果。不过，在大多数发展中国家的总人口调查中，很少有（如果有的话）可靠的控制变量可用，而且在任何可用的控制变量与调查变量之间也很少有紧密的关系。因此，使用方程式（23）过高估计加权所产生的设计效果的问题不会经常发生。然而，上面的讨论倒是提醒人们不要盲目地应用方程式（23）。

50. 我们在结束这部分关于加权的设计效果的讨论时，想就加权对子群估计数的影响问题谈点儿意见。本节和B.1节提出的所有结果都可以直接用于估算子群估计数方面的设计效果，只要将计算限制在子群成员即可。可是，若试图从完整样本的结果推导出子群估计数加权所产生的设计效果，则务须小心为好。要想使这种推导站得住脚，子群当中的权数分布必须跟完整样本的权数分布差不多才行。有的时候是这样，但并非总是这样。特别是，当使用非比例分层法估算某些域（即子群）的适当样本规模的时候，总样本估计数的设计效果值会超过1（在均等平均数和均等方差的假定条件下）。而域的估计数加权所产生的设计效果值则可能等于1，因为在域的范围使用了均等的选择概率。

C. 设计效果模型

51. 上节提出了分别与加权和类集有关的设计效果的结果，主要集中在平均数和比率方面的设计效果上。本节考虑加权与类集相结合的设计效果以及其他一些类型的估计数的设计效果，以拓展上述结果。

52. 为了拓展这些结果，使用了若干模型。这些模型在复杂样本设计的设计阶段和分析阶段都使用过（Kalton, 1977年；Wolter, 1985年）。从历史上来看，这些模型曾经在分析阶段发挥了重要作用。不过，它们在这方面的用途正在日趋减少。这些模型在未来的规划和新设计中的基本的——也是重要的——用途将是本文讨论的聚焦点。

53. 近年来，在计算复杂样本设计误差的计算能力和软件方面取得了重大进展。在取得这些进展以前，计算复杂样本估计数的有效抽样误差曾经是既费精力又费时间的工作。因此普遍的做法是仅限于数量相对较少的估计数直接计算抽样误差，并使用设计效果模型或其他模型来推导其他估计数的抽样误差。而今计算状况已有长足的进步，因此直接计算诸多估计数的抽样误差已经不再是个老大难问题。而且可以预料，今后还将在计算能力和软件方面取得进一步的进展。这样一来，为此目的使用设计效果模型的情况有望在很大程度上逐渐消失。

54. 在分析阶段使用抽样误差模型的另一个原因，是为调查报告简要汇总抽样误差提供一个手段，从而去除每个估计数都要附加抽样误差的必要性。有些情况下也可能有人争辩说，根据模型得出的抽样误差估计数或许要比直接抽样误差估计数更可取，因为这种比较准确。在某些情况下（譬如估算一个选定初级抽样单位数量很小的地区抽样误差），后面这个论据确实有一定的说服力。可是一般来说，不论出于哪种原因，使用模型报告抽样误差都是不可靠的。模型估计数的有效性取决于模型本身的有效性，而如果对直接抽样误差和基于模型的抽样误差进行比较的话，比较的结果往往会使人对模型的有效性产生严重疑问[譬如见Bye和Gallicchio（1989年）]。另外，虽然抽样误差模型可以为调查报告提供一个准确汇总抽样误差的手段，但是也给报告的使用者增加了不必要的负担——他们不得不根据模型去计算抽样误差。我们总的结论是，设计效果模型和其他抽样误差模型将在未来的调查分析中发挥有限的作用。

55. 相反，设计效果模型将继续在样本设计阶段发挥极其重要的作用。了解非比例样本分配的后果和类集做法对各类调查估计数精确度的影响，是有效进行样本设计的关键之所在。最明显不过的是，为关键估计数提供足够的精确度所需要的样本规模的确定，显然需要考虑到特定设计所产生的设计效果。另外，也可以通过考察不同设计模型所产生的效果来构建一个高效样本设计结构。请注意，为了在设计阶段应用模型，一些未知参数（比如 ρ ）的估计数是必不可少的。这就需要从以往的调查中产生这些参数的估计数（见下节的阐释）。

56. 我们先从描述用于推导“等概率抽选法”（epsem）样本中超出B.3节（题为“加权调整”）所考虑的平均数和比率范围的一系列统计数据的一类集效应模型入手。为了引进这些模型，让我们回到前面讨论过的子群平均数，其中对“交叉类”、“分割类”和“混合类”加以区分。交叉类，用 d 来表示，在

初级抽样单位中均衡分布，方程式（20）大致给出了交叉类平均数的设计效果，这里写为：

$$D^2(\bar{y}_{d,d}) = 1 + (\bar{b}_d - 1)\bar{\rho}_d \quad (24)$$

式中， \bar{b}_d 表示每个初级抽样单位的平均交叉类样本规模，而 $\bar{\rho}_d$ 是初级抽样单位中同质性 y 的综合量数。一个广泛采用的模型假定交叉类的同质性量数跟总人口的同质性一样，即： $\bar{\rho}_d = \bar{\rho}$ 。那么，交叉类平均数的设计效果可以用下式来估算：

$$d^2(\bar{y}_{d,d}) = 1 + (\bar{b}_d - 1)\hat{\rho} \quad (25)$$

式中， $\hat{\rho}$ 是由下式给出的全样本 $\bar{\rho}$ 的一个估计数：

$$\hat{\rho} = \frac{d^2(\bar{y}_{d,d}) - 1}{\bar{b}_d - 1} \quad (26)$$

57. 这种方法的一种普通拓展是计算涉及到各种相关变量的一套可比估计数的 $\hat{\rho}$ 值；如果 $\hat{\rho}$ 值相当雷同的话，就利用它们的某种形式的平均数来估算 $\bar{\rho}$ 值以及所有变量的子群估计数的 $\bar{\rho}_d$ 值。这种方法经常被用来为调查报告中汇总抽样误差提供设计效果模型。它也是为此目的而使用的一种形式的广义方差函数（GVF）的基础（Wolter, 1985年，第204页）。

58. 在某种类型变量中属于不同类别的各个子群比率——比如达到不同教育水平的子群比率或者属于不同行业类别的子群比率——的调查估计数，就是上述方法的一个特例。通常假定不同类别的 $\bar{\rho}$ 值是近似的，这样，只需估计其中一种类型的 $\bar{\rho}$ 值就行了；而一旦估算出 $\hat{\rho}$ 值，它就可以应用于所有其他类别。在只有两个类别的情况下（比如有电力供应的住户和无电力供应的住户），在数学上假设一个通用的 $\bar{\rho}$ 值是正确的，但在有两个以上类别的情况下它就未必适用了。比如，让我们考虑农业工人和矿业工人的比率估计数。在采矿业比较集中某些地区，几乎可以肯定农业工人的 $\bar{\rho}$ 值会远远低于矿工的 $\bar{\rho}$ 值。因此，笼而统之地为所有类别假设一个通用的 $\bar{\rho}$ 值就不妥当。

59. 当我们把根据方程式（25）推算出来的交叉类平均数方差与直接计算的平均数方差做比较的时候，就会发现前者一般估计偏低。这种比较的结果可能是这样一个事实决定的，即：有些子群虽然被划入交叉类，但是它们在初级抽样单位中的分布并不完全均衡。针对这个问题采用了一种补救办法，就是用下式计算的结果来修改方程式（25）：

$$d^2(\bar{y}_{d,d}) = 1 + k_d(\bar{b}_d - 1)\hat{\rho} \quad (27)$$

式中， $k_d > 1$ 。Kish（1995年）根据大量经验分析推荐了数值 $k_d = 1.2$ 或 1.3 ；Verma和Lê（1996年）则让 k_d 值随交叉类的规模而变（但 k_d 值永远大于1）。另一个可供选择的补救办法，就是根据方程式（21）用 $b'_d = \sum b_{d\alpha}^2 / \sum d_{d\alpha}$ 取代方程式（25）中的 \bar{b}_d 。

60. 现在让我们来简要考虑分析式统计的设计效果。最简单、也最广为采用的分析式统计形式，就是两个子群平均数或比例之间的差异。普遍发现，两个平均数间差异的设计效果值大于1，但是小于单独处理两个子群平均数所得的设计效果值（Kish和Frankel, 1974年；Kish, 1995年）。按方差表示如下：

$$V(\bar{y}_{d,d}) + V(\bar{y}_{d,d'}) < V(\bar{y}_{d,d} - \bar{y}_{d,d'}) < V(\bar{y}_{d,d} + \bar{y}_{d,d'}) \quad (28)$$

式中， d 和 d' 分别代表两个子群。当这两个子群都用同一个初级抽样单位表示的时候，平均数间差异的方差一般都低于上限。这一特征所产生的结果是，两个平均数间的协方差基本上总是正数，而这个正数协方差会减小差异方差。当子群分别属于不同组别初级抽样单位的切割类的时候，不会产生上述效应；而这种情况下适用上限。如果假定这两个子群的单位方差一样（亦即 $S_d^2 = S_{d'}^2$ ），那么这种不平等就会缩小到：

$$1 < D^2(\bar{y}_d - \bar{y}_{d'}) < \frac{n_d D^2(\bar{y}_d) + n_{d'} D^2(\bar{y}_{d'})}{n_d + n_{d'}}$$

61. 当两个不同的比率各自依据同一个多类别变量的时候，譬如讲，当请求两位受访者在几种备选方案当中做出选择，而分析者所感兴趣的是其中哪一种备选方案比另一种备选方案更受欢迎的时候，这两个比率之间的差异就构成一个特例。Kish等（1995年）考察了这种差异所产生的设计效果，并且凭经验发现，在这个特例中 $d^2(p_d - p_{d'}) = [d^2(\rho_d) + d^2(\rho_{d'})]/4$ 。

62. 上述研究的结论是，一般来说，平均数差异类集的设计效果要比整体平均数类集的设计效果小，这一发现可以推广到其他分析式统计。关于多类别回归系数的设计效果，曾提出一些明晰的经验证据和一些模拟建议，请参见Kish和Frankel（1974年）。回归系数方面的设计效果跟平均数差异的设计效果差不多。这符合人们的期望，并可以从如下事实得到证明，即：使用 $b = (\bar{y}_d - \bar{y}_{d'}) / (\bar{x}_d - \bar{x}_{d'})$ 这个公式，可以根据变量 x 相当有效地估算出简单线性回归 y 的斜率，其中，根据变量 x 分别计算出样本的第三变分上限（ u ）和第三变分下限（ l ）的 y 的平均数和 x 的平均数。关于回归和其他分析形式的设计效果，见Skinner, Holt和Smith（1989年）以及Lehtonen和Pahkinen（1994年）；关于复杂样本设计对调查数据分析精确度的影响，见Korn和Graubard（1999年）。

63. 在结束本节的时候，我们想围绕把整体设计效果分解成加权产生的设计效果和类集产生的设计效果两个要素的这一棘手问题谈几点意见。设计效果的计算公式 $d^2(\bar{y}) = v_c(\bar{y}) / v_a(\bar{y})$ 包含加权效应和类集效应的组合。可是在使用当前调查数据来策划未来调查的时候，需要把设计效果的这两个要素分开来。譬如讲，未来的调查可能是作为一项采用“等概率抽选法”（epsem）的调查设计的，而现行调查或许对某些“域”采取了过度抽样的做法。另外，即便未来的调查采用同样的初级抽样单位和分层法，仍有可能改变每个单位的子群规模。Kish（1995年）论述了这个问题，在这方面不存在单一而简便的解决办法。在这里，我们推荐一种仅可在权数随机而定或大体如此的情况下使用的方法。在这种情况下，我们可以用下述方程式将整体设计效果分解成加权和类集两种设计效果的一个产品：

$$d^2(\bar{y}) = d_w^2(\bar{y}) \cdot d_d^2(\bar{y}) \quad (29)$$

式中, $d_w^2(\bar{y})$ 是按照方程式 (23) 计算得出的权数进行加权所产生的设计效果, 而 $d_d^2(\bar{y})$ 是按照方程式 (20) 或 (21) 进行类集所产生的设计效果。在理论上方程式 (29) 没有什么正当的依据; 但是 Gabler, Haeder 和 Lahiri (1999年) 以方程式 (29) 作上限, 推导出设计效果。结合方程式 (20) 使用方程式 (29), 于是便通过下式估算出 $\bar{\rho}$ 值:

$$\hat{\rho} = \frac{[d^2(\bar{y})/d_w^2(\bar{y})]-1}{b-1} \quad (30)$$

从下文可以看出, 为了设计的目的, 参数 $\bar{\rho}$ 的估算要比类集设计效果的估算更为重要, 因为它更便于在不同设计当中移植。只要每个初级抽样单位的子样本规模维持不变, 一次调查中的类集设计效果即可应用于另一次的调查设计。

D. 在样本设计中应用设计效果

64. 本章前面讨论的各种设计效果模型可以成为策划新样本设计的有用工具。可是它们需要有经验数据、特别是同质性 $\bar{\rho}$ 的综合量值数据为依托。可以通过分析以往类似调查的设计效果获得这些数据。因此, 有关设计效果的数据积累是宝贵的。

65. 为“世界生育率调查”(WFS)和“人口与健康调查”(DHS)方案进行的广泛样本误差分析, 产生了大量可供使用的有关生育与保健人口统计调查设计效果的数据。世界生育率调查方案于1974至1982年总共在41个国家进行了42次调查。接着, 从1984年开始至今, 人口与健康调查方案先后在66个国家完成了120次调查, 大多数国家每隔3-5年重复进行一次该项调查。关于人口与健康调查抽样误差, 见 Verma 和 Lê (1996年); 关于世界生育率调查抽样误差的类似分析, 见 Kish, Groves 和 Krotki (1976年) 及 Verma, Scott 和 O'Muircheartaigh (1980年)。对这些方案抽样误差分析的一项重要发现是, 只要样本设计可比, 某项特定估计数的 $\bar{\rho}$ 值即可在各国之间有相当大的可移植性。因此, 在为一个国家设计新调查时, 可酌情利用邻国类似调查所提供的有关抽样误差的经验数据, 但须适当注意检查样本设计的可比性。

66. 下面提供的范例说明如何应用设计效果为一次虚构的国家调查进行样本设计。就此说明而言, 我们假定这是一个两阶段概率与规模成正比 (PPS) 的分层样本设计, 譬如以普查查点区为初级抽样单位, 以住户为第二阶段单位。我们假定所关注的关键统计数据是贫困住户的比率; 为了规划目的, 这个比率假定为25%, 并且在全国所有各省份差不多都是这个比率。初始设计说明中规定, 这个比率估计数的变异系数全国不得高于5%; 全国8个省份, 每省不得高于10%。此外, 样本应能有效地产生相当均匀地分布在8个省份的全国各个子群的精确统计估计数。如果采用样本随机抽样法, 那么变异系数即为:

$$CV = \sqrt{\frac{1-P}{nP}}$$

式中， P 是贫困住户比率（此例中 $P = 25\%$ ）。这个公式也可以用于复杂样本设计，但必须用有效样本规模 $n_{eff} = n/D^2(p)$ 来取代式中的 n 。

67. 需要解决的第一个问题就是如何在全国各省分配样本。表六.2展示了各省的人口分布状况 (w_h)，并分别列出各省之间样本按比例分配、样本规模均等分配、以及介于按比例分配和平均分配之间的折衷样本分配三种情况。在这个时间点上强制规定总样本规模为5 000个住户；必要的话，可以在以后改变此规模。

表六.2
全国八个省份（用A - H表示）的人口分布和各省样本分配的三种备选方案

	A	B	C	D	E	F	G	H	总计
W_h	0.33	0.24	0.20	0.10	0.05	0.04	0.02	0.02	1.00
按比例分配	1 650	1 200	1 000	500	250	200	100	100	5 000
均等样本规模分配	625	625	625	625	625	625	625	625	5 000
折衷样本分配	1 147	879	767	520	438	427	411	411	5 000

68. 在其他方面平等的情况下，按比例分配最适合产生全国估计数和子群估计数（后者的条件是，各省之间的子群分布均衡）。另一方面，均等样本规模分配最适合产生省级估计数。正如表六.2所示，由于 W_h 行所给出的各省人口规模迥异，这两种分配方式有很大的差别。按比例分配在小省份（E、F、G和H）产生的样本规模太小，以至于不能对其进行可靠的计算。而另一方面，均等样本规模分配则降低了全国估计数的精确度。这种精确度的损失可以根据方程式（15）计算出来，该方程式在此例中简化为 $H \sum W_h^2 = 1.77$ ，式中， H 系省份数量。因此，如果仅考虑非比例分配效应（亦即排除类集效应），对于全国估计数来说，5 000住户的样本规模可以减小到 $5 000/1.77 = 2 825$ 。

69. 由于采取平均分配方式而导致全国估计数（特别是子群估计数）精确度明显下降的情况是否可以接受，这要取决于全国估计数和省级估计数的相对重要性。往往全国估计数的重要性足以使得这种过大的损失不可接受。在这种情况下，介于按比例分配和平均分配之间的折衷分配方式有可能被认为既能满足全国估计数的需求又能满足省级估计数的需求。表六.2最末行中的折衷分配结果是根据Kish（1976年和1988年）针对全国和省级估计数同等重要的情况建议的一种分配方法计算出来的。由公式 $n_h \propto \sqrt{W_h^2 + H^{-2}}$ 给出的这种分配方法相对于按比例分配来说大大增加了小省份的样本规模，但是还没有大到像平均分配那么大的规模。这种分配的非均等加权设计效果值为1.22，而均等样本规模分配的设计效果值为1.77。我们将假定这次调查采用折衷的分配方式。

70. 需要解决的下一个问题是如何确定初级抽样单位的数量和每个初级抽样单位选择住户的可取数量。前面第二章讨论过，通过使用一个简单的成本模型，每个被抽样的初级抽样单位选择住户的最佳数量可以由下式给出：

$$b_{opt} = \sqrt{\frac{C^*(1-\rho)}{\rho}}$$

式中， C^* 是样本中增加一个初级抽样单位所需的成本与增加一个住户所需成本的比率。这个成本模型未免过分简化，而且 b_{opt} 公式也不该盲目地应用；但它毕竟提供了有益的指导。

71. 让我们假定，从现场调查工作的组织结构来看采用这个成本模型是合理的，并且通过分析成本结构得出 C^* 值约为16。另外让我们假定，前一次调查采用同样的初级抽样单位得出一个与贫困高度相关的特征的估计数为。将这些数字应用于上述公式，得出 $k_d = 17.4$ ；为简便起见，我们将其四舍五入到17。实际上，全国各地的成本比率 C^* 往往不是一成不变的；比如，城市地区的成本比率可能远远低于农村地区的成本比率。若果真如此，就可能在不同的地区采用不同的比值。这种复杂情况也不在本节的考虑范围之内。读者可以在本出版物中有关样本设计的好几个章节找到此类差异的实例。

72. 设 $\bar{p} = 0.05$ ， $b = 17$ ，类集的设计效果为：

$$D^2(p) = 1 + (b-1)\bar{p} = 1.80$$

在确定省级估计数的精确度时有必要将这种设计效果考虑进去。比如在H省，411个住户的有效样本规模是。因此，H省贫困住户比率的变异系数即为0.11。如果认为这一精确度水准还不够高，就需要增加H省（和G省）的样本规模。

73. 全国估计数的设计效果需要把类集设计效果和省际非比例分配设计效果结合起来。因此，对于整个国家的贫困住户比率来说，就可以根据方程式（29）得出估计的设计效果值为。因此，与实际样本规模5 000住户相对应的有效样本规模即为2 277户，而全国贫困住户比率估计的变异系数为0.036。往往有这种情况：总样本规模对于满足总人口估计数的精确度要求绰绰有余。比较令人关注的倒是人口子群估计数的精确度水准问题。在这种情况下，正如C节所述，在初级抽样单位当中均衡分布的交叉类的类集设计效果要小于总样本的设计效果。譬如，让我们来考虑一个占总人口三分之一的交叉类别。在这种情况下应用公式（27），设， $k_d = 1.2$ ， $\bar{b}_d = 17/3$ 那么类集设计效果值即为1.23。把类集设计效果和省际非比例分配设计效果结合起来，就得出交叉类估计数的总体设计效果值为；有效样本规模为 $5\ 000 / (3 \times 1.50) = 1\ 111$ 。这样，交叉类估计数的估计变异系数即为0.05。

74. 可以顺着上述思路进行计算，以评估关键调查估计数所可能达到的精确度，并可修改样本规模以达到预期的要求。在最终的样本估计数中，需要对不响应情况给予补偿。譬如讲，全国相当一致的响应率为90%，那么，上面计算的样本规模就需要增加11%。另外，由于做不响应调整、致使权数变异增加的缘故，最终可能导致设计效果或多或少有所增强。在计算拟用来生成预期样本规模的抽样份额的时候，也要对覆盖面遗漏情况给予补偿。90%的覆盖率，需要将抽样分数增加11%。

E. 结论

75. 了解设计效果及其各项要素，对于抽样设计的开发极为重要。譬如讲：

- 整个设计对关键调查估计数的影响力大小，可以用来确定所需要的样本规模。一个无约束样本的每一种关键估计数达到规定精确度标准所需要的样本规模，是可以计算出来的，然后用这个样本规模乘以特定估计数的设计效果值，即可得出在复杂样本设计中该估计数所需要的样本规模。然后通过审定每一种估计数所需的样本规模（大概是选取最大的样本规模），就可以选定最终样本规模。
- 当采用非比例分层样本设计法来提供既定精确度的域的估计数的时候，作为结果而发生的整个样本估计数的精确度损失和超越各种域的子群估计数的精确度损失，可以通过计算不同的可变加权指数所产生的设计效果值估算出来。如果发现损失过大，可以说明需要改变域的要求，从而减小可变加权指数。
- 如果对于某些关键估计数来说类集所产生的设计效果非常大，那就应该考虑是否有必要增加初级抽样单位的取样数量(a)同时缩小子样本规模(b)。

76. 虽然本章提供的公式对样本设计都很有用，但是不应该盲目应用。本文曾不止一次地提到，相关的公式是在诸多假定和简化条件下推导出来的。使用者务必对这些特征保持敏感，并结合他们的具体情况考虑有关公式能否提供适当的近似值。

77. 估算类集产生的设计效果需要掌握关键调查变量的 ρ 值估计数。这些估计数有欠完备，在所难免，但只要可用就行。片面地倾向于使用比预测值大的 ρ 值，就会导致规定需要较大的样本规模；所以说，这是一种保守的策略。

78. 最后应该指出的是，使用这些设计效果模型的目的在于产生行之有效的样本设计。如果公式不十分准确，就会带来一定的效益损失。然而使用不适当的模型来开发样本设计并不会影响调查估计数的有效性。若采用概率抽样法，调查估计数仍然是各项人口参数的有效估计数。

参考资料

- Brick, J. M. 和G. Kalton (1996年)。“处理调查研究中的丢失数据”。《医学研究的统计方法》，第5期，第215-238页。
- Bye, B. 和S. Gallicchio (1989年)。“来自收入和方案参与调查的社会保障方案参与者关于抽样方差估计值的说明”。《美国社会保障通报》，第5期，第10号，第4-21页。
- Clark, R. G. 和D. G. Steel (2002年)。“将住户作为抽样单位的影响”。《国际统计评论》，第70期，第289-314页。
- Cochran, W. G. (1977年)。《抽样方法》，第三版。纽约：威利国际出版公司。
- Gabler, S.、S. Haeder和P. Lahiri (1999年)。“以Kish加权和聚类设计效果公式为基础的模型”。《调查方法》，第25期，第105-106页。
- Holt, D. H. (1980年)。“Verma, V.、C. Scott和C. O'Muircheartaigh论文的讨论：世界生育率调查中的样本设计和抽样误差”。《皇家统计学会志》，A辑，第143期，第468-469页。
- Kalton, G. (1977年)。“估计调查抽样误差的实用方法”。《国际统计学会通报》，第47期，第3号，第495-514页。
- _____ (1979年)。“最终聚类抽样”。《皇家统计学会志》，A辑，第142期，第210-222页。
- Kish, L. (1965年)。《抽样调查》。纽约：威利国际出版公司。
- _____ (1976年)。“线性样本设计的最佳和最接近效果”。《皇家统计学会志》，A辑，第139期，第80-95页。
- _____ (1982年)。“设计效果”。《统计科学百科全书》，第2卷，S. Kotz和N. L. Johnson等，纽约：威利国际出版公司，第347-348页。
- _____ (1988年)。“多种目的样本设计”。《调查方法》，第14期，第19-32页。
- _____ (1992年)。“不等量的权数”。《官方统计杂志》，第8期，第182-200页。
- _____ (1995年)。“设计效果的方法”。《官方统计杂志》，第11期，第55-77页。
- _____和M. R. Frankel (1974年)。“复杂样本的推论”。《皇家统计学会志》，B辑，第36期，第1-37页。
- _____和其他人 (1995年)。“相关的设计效果”。《调查方法》，第21期，第117-124页。
- _____和其他人 (1976年)。“生育率调查中的抽样误差”。《世界生育率调查专题文件，第17号》。海牙：国际统计学会。

- Korn, E. L. 和 B. I. Graubard (1999年)。《健康调查分析》。纽约：威利国际出版公司。
- Lehtonen, R. 和 E. J. Pahkinen (1994年)。《复杂调查设计和分析的使用方法》，修订版。联合王国，奇切斯特：威利国际出版公司。
- Lepkowski, J. M. 和 J. Bowles (1996年)。“个人电脑的抽样误差软件”。《调查统计员》，第35期，第10-17页。
- Rust, K. F. (1985年)。“样本调查中复杂估值的方差估计”。《官方统计杂志》，第1期，第381-397页。
- _____和 J. N. K. Rao (1996年)。“复杂调查中使用重复方法的方差估计”。《医学研究的统计方法》，第5期，第283-310页。
- Skinner, C. J.、D. Holt 和 T. M. F. Smith 等 (1989年)。《复杂调查分析》。联合王国，奇切斯特：威利国际出版公司。
- Spencer, B. D. (2000年)。“概率抽选法校正测量时不平等权数的近似设计效果”。《调查方法》，第26期，第137-138页。
- 联合国 (1993年)。《国家住户调查能力方案：住户调查中的抽样误差》。UNFPA/UN/INT-92-P80-15E。纽约：联合国统计司，将由 Vijay Verma 出版。
- Verma, V. 和 T. Lê (1996年)。“人口与健康调查中抽样误差的分析”。《国际统计评论》，第64期，第265-294页。
- Verma, V.、C. Scott 和 C. O'Muircheartaigh (1980年)。“世界生育率调查的样本设计和抽样误差”。《皇家统计学会志》，A辑，第143期，第431-473页。
- Wolter, K. M. (1985年)。《方差估计简介》。纽约：施普林格出版社。

第七章

发展中国家的调查设计效果分析

Hans Pettersson

瑞典斯德哥尔摩
瑞典统计局

Pedro Luis do Nascimento Silva

巴西里约热内卢
国家科学统计局/巴西统计局

摘 要

本章将对来自7个国家11项住户调查的设计效果，以及设计非常类似的3项调查的设计效果进行比较；并针对住户消费和耐用品拥有情况方面的估计，就其设计效果和同质性比率（roh）进行比较。结尾部分将对各项调查估计概率进行讨论。

关键词：设计效果、效率、同质性比率、调查设计、抽样设计、类集。

A. 引言

1. 发展中国家住户调查中，将设计效果作为标准结果进行计算的做法还不常见。但一些标准化的调查除外，如：生活水平衡量研究（LSMS）调查、人口与健康调查（DHS）。对于这些调查来说，设计效果是跨国家进行计算和比较的（见第二十二章和第二十三章）。早些时候已经对世界生育率调查（WFS）方案（Verma, Scott和O'Muircheartaigh, 1980年）下的35项调查进行了深入的比较分析。

2. 本章将介绍来自7个国家11项调查的设计效果。对调查的选择具有主观性，主要以能否方便地获得相关资料为准。调查来自：巴西（3）、柬埔寨（1）、老挝人民民主共和国（1）、莱索托（1）、纳米比亚（2）、南非（2）和越南（1）。各项调查的性质各不相同，涉及不同的主题。在各种调查中，包括：多用途调查、劳动力调查、生活水平调查和人口调查。设计效果是针对一系列特征，尤其是调查计划目的而估算的。本章的主要目的是为了让读者对各种调查中的设计效果水平有一个大概的了解。

3. 通过对住户消费和耐用品使用方面几个变量的设计效果和同质性比率进行比较，对三项在设计方面很类似的调查进行了深入的分析。其目的是为了对不同人口群体大致相同的行为变量进行考察，并对调查结果的类似之处和可能的模型进行探讨。

B. 调查

4. 本章对各调查的设计效果进行了介绍，这些调查包括：

- 1997/1998年老挝支出与消费调查（LECS）
- 1999年柬埔寨社会经济调查（CSES）
- 1993/1994年纳米比亚家庭收入与支出调查（NHIES）
- 1995/1996年纳米比亚两次普查之间的人口调查（NIDS）
- 1999年越南多用途住户调查（VMPHS）
- 1997年莱索托劳动力调查（LFS）
- 1999年10月南非共和国住户调查（OHS）
- 2000年2月南非共和国劳动力调查
- 1999年巴西全国住户抽样调查（PNAD）
- 1999年9月巴西PME调查
- 1996/1997年巴西PPV调查

5. 表七.1对11项调查的主要设计特征进行了总结。所有调查都采取了标准的概率与规模成比例（PPS）两阶段调查法，但越南除外，越南采用的是三阶段调查法。PNAD对于小的非都会自治区也采用三阶段抽样，但这些只包括调查所覆盖人口的大约三分之一。大多数调查都将查点区（EA）作为初级抽样调查单位（PSU）（在有些情况下，对小查点区进行某些修改）。在这些情况下，初级抽样单位的数量平均为90-150个住户。有三项调查偏离了这一模式。莱索托两项调查的初级抽样单位数量更多：初级抽样单位为若干组查点区，平均数量为340-370个住户。另一方面，老挝调查中农村的初级抽样单位数量平均只有50个住户。

6. 在好几项调查中，初级抽样单位内的抽样规模（类集规模）大约为20个住户。在纳米比亚两次普查之间的人口调查中，每个初级抽样单位中的抽样规模高达50个住户。较低的是巴西的PPV调查，城市的每个初级抽样单位选了8个住户。南非的两项调查和柬埔寨调查则从每个初级抽样单位中选了10个住户。在大多数调查中，城市和农村的类集规模相同。

7. 大多数调查都是明确根据行政区划（省、地区）内的城市/农村地区进行分层。莱索托的劳动力调查对农业生态区进行了进一步分层，老挝的支出与消费调查则根据是否有公路做了进一步分层。巴西的PNAD和PME调查只是不明确地根据城市和农村进行分层，并在根据位置分类后，通过系统的概率与规模成正比抽样方法抽取初级抽样调查单位。

8. 所有调查中都采用了系统抽样方法在最终地区单位对住户进行抽样，但PPV调查除外，该调查通过简单随机抽样的办法对住户进行抽样。

表七.1

11项住户调查研究的特点

调 查	地区抽样阶段的数量	第一阶段样本: 选定样本初级抽样单位数量	初级抽样单位规模: 每个初级抽样单位的平均住户数量	类集规模: 每个初级抽样单位抽取的住户数量 (或两个地区抽样阶段的初级抽样单位)	抽样规模: 调查中住户数量	各层之间的样本分配
1997-1998年老挝支出与消费调查	1	R:348 U:102	R:51 U:87	R:20 U:20	R:6 960 U:2 040	不成比例
1999年柬埔寨社会经济调查	1	R:360 U:240	R:154 U:243	R:10 U:10	R:3 600 U:2 400	大约成比例
1993-1994年纳米比亚家庭收入与支出调查	1	R:123 U:96	R:152 U:148	R:20 U:20	R:2 685 U:1 712	大约成比例
1995-1996年纳米比亚两次普查之间的人口调查	1	R:120 U:82	R:152 U:148	R:50 U:50	R:5 600 U:3 900	大约成比例
1999年越南多用途住户调查	2	839个初级抽样单位, (每个初级抽样单位抽取2个第二阶段抽样单位)	R:1 417 U:2 579 第二阶段抽样单位: R:99 U:105	R:15 U:15	25 170	不成比例
1997年莱索托劳动力调查	1	R:80 U:40	R:370 U:341	R:33 (平均) U:25 (平均)	R:2 600 U:1 000	大约成比例
2000年南非共和国劳动力调查	1	R:426 U:1 148	R:最低100 ^a U:最低100 ^a	R:10 U:5	R:4 059 U:5 646	不成比例
1999年10月南非共和国住户调查	1	R:1 273 U:1 711	R:110-120 U:80-100	R:10 U:10	R:10 923 U:15 211	不成比例
1999年巴西PNAD调查	1或2	7 019	250	13	93 959	不成比例
1999年9月巴西PME调查	1	1 557	250	20	30 535	不成比例
1996-1997年巴西PPV调查	1	554	250	R:16 U:8	4 944	极其不成比例

注: R = 农村, U = 城市

^a 最低100。

9. 很多简单抽样设计的一个重要特征, 就是在各个省之间采用不成比例的样品分配, 目的是为了按省份取得足够精确的估计数。在有些情况下, 分析中各种用以补偿不成比例分配的权数差别很大。如, 在巴西的PPV调查中, 最大抽样权数和最小抽样权数之间的比率大约为40。调查抽样设计的详细情况将在附件中介绍。

C. 设计效果

10. 表七.2到表七.6显示了每项调查若干项估计的设计效果 ($d^2(\bar{y})$) (至于设计效果的计算, 见第六章)。设计效果是通过相关数据统计分析软件 (SUDAAN) 或StATA计算的。在有些情况下, 由国家统计署提供设计效果。²¹

²¹ 南非共和国劳动力调查和10月住户调查的设计效果由南非统计局David Stoker教授提供。越南多用途住户调查的设计效果由越南统计署社会与环境统计部门负责人Nguyen Phong先生提供。纳米比亚家庭收入与支出调查的设计效果由纳米比亚国家统计局的Alwis Weerasinghe先生计算。巴西调查的设计效果由巴西统计局Pedro Silva博士计算。其他调查的设计效果由Hans Pettersson博士根据各国家统计局机构提供的数据进行计算。

11. 各设计效果之间存在着很大的差异，但这种差异是难免的，因为样本设计存在着差异、调查中的变量存在着不同、各国人口状况也有别。有些效果值很高。表七.2至七.6中的结果显示，对于住户变量来说，设计效果值在6-10范围之间的情况并不少见。有些设计效果值在10-15之间。注意这些设计效果反应了复杂的分层类集样本设计以及各省之间不成比例分配（在适用的情况下）的效果。表七.2至七.6中显示的设计效果有助于说明发展中国家某些社会经济和人口住户调查的设计效果水平。

表七.2
非洲和东南亚七项调查的设计效果估计

		城 市	农 村	国 家
1997-1998年老挝支出 与消费调查	每个住户每月的总消费	3.8	7.8	5.4
	每个住户每月的食品消费	4.4	6.8	5.8
	拥有机动车的住户所占比率	1.3	3.3	2.1
	拥有电视机的住户所占比率	3.1	6.8	5.4
	拥有收音机的住户所占比率	2.7	4.8	4.5
	拥有录像机的住户所占比率	3.9	6.1	5.5
1999年柬埔寨社会经 济调查	每个住户每月的总消费	2.0	2.0	1.4
	每个住户每月的食品消费	3.1	3.2	3.2
	拥有电视机的住户所占比率	2.4	2.2	2.6
1993-1994年纳米比亚 家庭收入与支出调查	住户每年总消费	2.9	1.9	2.5
	家庭每年总收入	2.9	2.8	2.8
	拥有电视机的住户所占比率	6.0	4.6	4.1
	拥有收音机的住户所占比率	2.7	2.1	2.4
	拥有电话的住户所占比率	6.2	4.6	4.5
纳米比亚两次普查之间 人口调查	拥有电视机的住户所占比率	14.7	4.1	6.6
	用电照明的住户所占比率	4.4	3.9	4.2
	在过去12个月中遭遇住户成 员死亡的住户所占比率	2.1	4.3	2.3
1999年越南多用途住 户调查	贫困率	7.1
1997年莱索托劳动力 调查	就业率	5.6	3.1	6.6
	10岁以上尚未入学的人口所 占比率	4.6	5.9	5.5
	自给农民所占的比率	6.3	4.4	8.1
	单干工人所占的比率	3.0	1.4	2.4
1999年10月南非共和国 住户调查	就业率	4.0	3.6	3.8

		城 市	农 村	国 家
2000年南非共和国劳动 力调查	就业率	2.5	3.4	2.8

注： 两点 (..) 表明没有资料。

12. 表七.2显示的是，非洲和东南亚七项调查中国家一级、城市和农村分区的设计效果估计。大多数设计效果和住户社会经济变量有关。三项调查的设计效果主要与单独的劳动力变量有关。全国总体设计效果值平均为4.2。各种效果值的差异很大，从1.3到1.8不等，但大多数效果值在2.0-6.0之间。城市和农村分区的平均设计效果值分别为4.1和4.0。由于样本设计及变量存在着差异，所以很难初步找出表中各类变量（如，社会经济/劳动力）之间或地区（城市/农村）之间一般差异所导致的结果。表七.7力求对某些设计效果进行比较。

13. 表七.3介绍了巴西PNAD调查中某些住户估计数的设计效果估计。

表七.3

按照全国和地区类型划分的部分住户估计数的估计设计效果
(1999年PNAD调查)

变 量	全 国	大城市	大 自 治 区	其他地区
净普通水供应比率	9.80	6.60	6.74	10.73
来自水源的供水比率	9.24	4.04	4.19	9.43
拥有足够排水设施的比率	9.04	6.36	5.87	11.59
净普通自来水的比率	8.48	5.16	4.79	9.40
至少拥有一个浴室的比率	8.34	1.51	7.20	7.76
自有土地比率	8.10	11.53	4.49	7.09
拥有电的比率	7.92	1.03	4.43	7.27
拥有足够墙壁材料的比率	7.43	6.17	5.01	6.84
至少一个房间有自来水的比率	7.09	4.74	5.45	7.04
拥有足够屋顶材料的比率	5.68	2.91	2.41	5.65
每个住户拥有的平均房间数	5.32	6.26	4.50	5.09
拥有电话的比率	4.80	5.59	4.44	5.91
拥有电冰箱的比率	4.59	1.53	2.77	5.02
拥有洗衣机的比率	4.34	3.98	3.49	6.25
拥有彩色电视机的比率	4.31	1.77	2.76	4.88
拥有冷冻柜的比率	3.83	3.55	2.68	4.67
拥有滤水器的比率	3.39	2.50	2.07	4.37
拥有收音机的比率	3.01	1.46	1.62	3.29

变 量	全 国	大城市	大 自 治 区	其他地区
拥有黑白电视机的比率	2.79	1.50	1.30	2.93
平均租金	2.52	3.09	2.01	3.39
自有房屋的比率	2.46	3.18	1.74	2.30
租赁房屋的比率	2.32	2.71	1.78	2.51
房间用作寝室的平均数量	2.14	2.37	1.72	2.09

14. 至于国家一级的估计，其设计效果差异在2-10之间，平均值为5.5。有些变量的设计效果值要高一些——如，拥有净普通水供应的住户所占比率、来自水源的供水比例、拥有足够排水系统的比率。这种情况是预料中的，因为这些变量通常所表现的类集程度很高。有些“经济”变量的设计效果要低一些——如，平均租金、自有或租赁房屋的比率、房间用作寝室的平均数量。另外，正如所预料的那样，大城市和较大自治区的设计效果一般低于其他地区，大城市和较大自治区采用的是两阶段整群抽样，而其他地区，设计的类集程度更高（三阶段整群抽样）。

15. 表七.4显示的是一组个人变数计量的设计效果。

表七.4

在全国范围内选择的个人特征及各分区的估计设计效果（1999年PNAD调查）

变 量	国 家	大城市	大 自 治 区	其他地区
白色人种比率	15.97	11.97	8.14	19.97
黑色或有色人种比率	15.75	12.23	8.44	19.41
带薪工人比率	8.44	4.45	5.81	7.49
自谋职业比率	7.65	3.73	5.51	6.66
社会保障比率	6.59	2.93	3.28	8.45
文盲比率	6.33	3.67	4.37	7.10
主要职业平均收入	5.54	7.16	4.45	6.38
住房津贴比率	5.23	3.80	3.00	5.54
交通补贴比率	4.93	2.94	2.78	9.10
健康补贴比率	4.90	3.76	2.29	8.79
正在工作（10岁以上）的人的比率	4.79	1.97	1.67	7.08
食品津贴比率	3.35	2.60	2.08	4.60
正在工作（5-9岁）的孩子的比率	3.27	1.25	2.04	3.00
雇主比率	2.87	2.80	1.54	2.63
上学比率	1.88	1.75	1.57	1.94
教育津贴比率	1.87	1.85	1.74	2.22

16. 国家一级的估计设计效果值大约相差2至16。种族变量的设计效果值很高；与工作或收入有关的变量，其设计效果值也很高；诸如上学比率、教育津贴比率之类的变量，其设计效果值比较低。至于采用三阶段设计的其他地区，其设计效果值则更高。住户变量的设计效果值一般低于个人变量设计效果值，这是预料之中的事，因为人员的数量大于按初级抽样单位调查的住户数量。不同变量的设计效果差异很大，这也是预料之中的事，因为它们的类集程度不同。这些偏高的设计效果可以通过各层之间不成比例的样本分配得以解释，各层之间不成比例的分配导致了不同的权数。

17. 巴西PME调查每月公布的若干估计数设计效果见表七.5。该表选用了1999年9月的数值，因为它们的参考期和1999年PNAD调查相同。

表七.5

1999年9月PME调查中若干估计的估计设计效果

变 量	累西腓	萨尔瓦多	贝洛哈里桑塔	里约热内卢	圣保罗	Pôrto Alegre	全 部
主要职业平均收入	3.43	4.47	2.49	4.44	4.89	4.79	6.23
雇主比率	2.00	2.16	3.06	2.53	2.33	2.27	3.34
文盲比率	4.23	4.43	1.86	2.69	2.11	2.13	3.24
失业率	1.64	2.62	1.98	2.06	1.65	1.67	2.43
登记就业比率	1.61	1.87	1.66	1.50	1.40	1.75	2.02
从事经济活动的比率	1.59	1.99	1.78	1.61	1.31	1.40	1.96
带薪工人比率	1.51	1.67	1.43	1.37	1.34	1.55	1.88
自谋职业比率	1.53	2.26	1.60	1.47	1.19	1.14	1.78
上学比率	1.41	1.57	1.64	1.24	1.26	1.49	1.72

18. 至于该系列中其他月份同一估计的设计效果，尽管在这里没有介绍，但却做了计算，结果发现月和月之间的差异不大。10年中查点区的样本是固定的，短期内的样本规模也没有多少差异。主要职业平均收入的设计效果较大、文盲比率和雇主比率的设计效果为中等。这些和PNAD调查中大城市类似估计数的观察值一致，这不足为怪，因为除了PME调查中每个初级抽样单位抽取了更大的样本外，PME调查和PNAD调查基本上采用了相同的样本设计。其他变量的设计效果值在2.5以下。PME调查所估计的可比变量，其设计效果一般低于PNAD调查中的设计效果，这是因为相对于PNAD调查而言，PME调查中样本分配更均衡。

表七.6
PPV调查中部分估计数的估计设计效果

人口估计参数	估计差数
14岁以上的文盲人数	4.17
14岁以上的文盲比率	3.86
认为其健康状况“不佳”的人数	3.37
租赁房屋的比率	2.97
每个住户的平均人数	2.64
年龄在7-14岁之间的文盲人数	2.64
年龄在7-14岁之间的文盲比率	2.46
年龄在12-49岁之间、其孩子出生时死亡的妇女人数	2.03
年龄在12-49岁之间有孩子的妇女人数	2.02
年龄在12-49岁之间、其孩子出生时存活的妇女人数	2.02
依赖比率（年龄在0-14岁之间的人数加上年龄在65岁以上的人数除以年龄在15-64岁之间的人数）	1.99
年龄在12-49岁之间的每位妇女生育孩子的平均数	1.26

19. 表七.6显示了巴西PPV调查中一小部分估计数的估计设计效果。

20. 此处估计的设计效果值相差1.3至4.2。这些设计效果值相对较小，说明PPV调查中的类集程度较低，在PPV调查中，每个初级抽样单位只抽取了8个住户。这也说明问卷中人口和教育方面的大多数变量，加上住户一级的两个变量都予以了考虑。

21. 现在从表七.2至七.6中选出一组估计数，这些估计数在一次以上的调查中出现过。表七.7显示了这些调查的设计效果。设计效果被分为三组：(1) 住户消费与家庭收入；(2) 住户耐用品；(3) 就业与职业。在每一组内，我们将定义大致相同的估计数归为一类。

22. 国家一级估计的设计效果差异在1.4-6.6之间，中值为4.3。有些设计效果值很高。一个最突出的就是纳米比亚NIDS调查中拥有电视的城市住户比率，其设计效果值为14.7，这是因为其类集规模很大，为50个住户。如果类集规模和NHIES调查一样为20个住户，那么其设计效果值将为6.7，这和NHIES调查6.0的设计效果值接近。但该设计效果值仍然很高，在这种情况下，变量权数并没有起到多大的作用。每月食品消费调查中农村地区大多数估计的设计效果值也很高。在NHIES调查中，城市地区某些耐用品的设计效果值很高。

23. 除了南非两项调查和柬埔寨调查外，所有调查中城市和农村之间都存在明显的差异。老挝和巴西的调查中（见表七.2至七.6），城市的设计效果值

表七.7

各项调查设计效果比较

主题/特征	城市	农村	国家	说明
消费、家庭收入（住户变量）				
• 每月总消费（老挝人民民主共和国：LECS）	3.8	7.7	5.4	CSES调查中的类集规模是LECS调查和NHIES调查中类集规模的一半。
• 每月总消费（柬埔寨：CSES）	2.0	2.0	1.4	
• 国内住户总消费（纳米比亚：NHIES）	2.9	1.9	2.5	
• 每月食品消费（老挝人民民主共和国：LECS）	4.4	6.8	5.8	
• 每月食品消费（柬埔寨：CSES）	2.5	3.3	3.3	
住户耐用品（住户变量）				
• 拥有电视机的住户比率（老挝人民民主共和国：LECS）	3.1	6.8	5.4	NIDS调查中类集的规模是其他调查类集规模的一倍以上，这是城市地区设计效果值较大的原因（但不是农村地区设计效果值较低的原因）。
• 拥有电视机的住户比率（柬埔寨：CSES）	2.4	2.2	2.6	
• 拥有电视机的住户比率（纳米比亚：NHIES）	6.0	4.6	4.1	
• 拥有电视机的住户比率（纳米比亚：NIDS）	14.7	4.1	6.6	
• 拥有彩色电视机的住户比率（巴西：PNAD）	4.3	
• 拥有收音机的住户比率（老挝人民民主共和国：LECS）	2.7	4.8	4.5	
• 拥有收音机的住户比率（柬埔寨：CSES）	2.1	2.8	3.4	
• 拥有收音机的住户比率（纳米比亚：NHIES）	2.7	2.1	2.4	
• 拥有电话的住户比率（纳米比亚：NHIES）	6.2	4.6	4.5	
• 拥有电话的住户比率（巴西：PNAD）	-	-	4.8	
就业、职业（人变量）				
• 就业率（南非：OHS）	4.0	3.6	3.8	南非LFS调查和OHS调查城市地区设计效果的差异是因为LFS调查中城市地区的类集规模更小（LFS调查的类集规模为5个住户，OHS调查的类集规模为10个住户）。
• 就业率（南非：LFS）	2.5	3.4	2.8	
• 就业率（莱索托：LFS）	5.6	3.1	6.6	
• 就业率（巴西：PNAD）	-	-	4.8	

注：两点（..）说明没有资料。

（-）说明该项目不适用。

一般低于农村的设计效果值。纳米比亚和莱索托的调查中，城市的设计效果值要高于农村的设计效果值。（大多数调查中，城市和农村地区的类集规模相同的，所以其差异并不是由类集规模不同而引起的。）

24. 设计效果包括分层、不同加权、类集规模和类集同质性等方面的效果（有关效果的详细讨论见第六章）。表七.7中各项调查的样本设计可能大致相同，但在分层、类集规模和样本分配方面有着明显的差异。为了更具可比性，最好从设计效果中剔除类集规模和加权效果。

D. 同质性比率的计算

25. 可以采用LECS、CSES和NHIES调查中住户消费和耐用品拥有情况等估计，继续对更小一组的调查和变量进行分析，LECS、CSES和NHIES调查的样本设计类似。所有调查都采用了两阶段样本设计，将查点区作为初级抽样单位。省一级以及省内的城市/农村分区对初级抽样单位的分层方式大致相同。住户是通过在查点区内进行系统抽样选取的。但各层的样本分配不同。老挝调查中各省的分配均匀，而其他两项调查中，各省的样本分配接近均匀。分析的目的是为了考察复杂样本设计对不同人口群大致相同的估计指标精确度所带来的影响，并探讨各同质性比率的相同之处和可能的模式。

26. 第一步就是从设计效果值中剔除加权不同所带来的影响。表七.8中，由于加权和类集的原因，设计效果被分为若干部分。这些部分是按第六章中的第23个和第20个等式进行计算的。LECS调查中各省内的样本规模相同，这

表七.8

总体设计效果被分为加权效果 ($d_w^2(\bar{y})$) 和类集效果 ($d_d^2(\bar{y})$)

主题/特征	城 市			农 村		
	总 体	加 权	类 集	总 体	加 权	类 集
	$d^2(\bar{y})$	$d_w^2(\bar{y})$	$d_d^2(\bar{y})$	$d^2(\bar{y})$	$d_w^2(\bar{y})$	$d_d^2(\bar{y})$
住户消费、收入						
• 每月总消费 (LECS)	3.8	1.60	2.4	7.7	1.55	5.0
• 每月总消费 (CSES)	2.0	1.11	1.8	2.0	1.16	1.7
• 国内住户总消费 (NHIES)	2.9	1.20	2.4	1.9	1.23	1.5
• 每月食品消费 (LECS)	4.4	1.60	2.8	6.8	1.55	4.4
• 每月食品消费 (CSES)	2.5	1.11	2.3	3.3	1.16	2.8
• 住户总收入 (NHIES)	2.9	1.20	2.4	2.8	1.23	2.3
住户耐用品						
• 拥有电视机的住户比率 (LECS)	3.1	1.60	2.0	6.8	1.55	4.4
• 拥有电视机的住户比率 (CSES)	1.9	1.11	1.7	1.8	1.16	1.6
• 拥有电视机的住户比率 (NHIES)	6.0	1.20	5.0	4.6	1.23	3.7
• 拥有收音机的住户比率 (LECS)	2.7	1.60	1.7	4.8	1.55	3.1
• 拥有收音机的住户比率 (CSES)	2.1	1.11	1.9	2.3	1.16	2.9
• 拥有收音机的住户比率 (NHIES)	2.7	1.20	2.3	2.1	1.23	1.7
• 拥有录像机的住户比率 (LECS)	3.9	1.60	2.4	6.1	1.55	3.9
• 拥有电话的住户比率 (NHIES)	6.2	1.20	5.2	4.6	1.23	3.7

样就使得样本权数的差异很大。因此，LECS调查估计中，归因于加权的设计效果值很高。NHIES调查中，人口不太多的地区以及城市地区的抽样过多，从而使归因于加权的设计效果值在1.0以上，但这要比LECS调查的值低得多。CSES调查中城市地区的抽样也过多。

27. 三项调查所使用的设计都是从每个初级抽样单位中（通过系统抽样）抽取固定数量的住户。类集规模不变也是权数存在差异的原因之一，因为衡量初级抽样单位规模的尺度不完善，所以导致了总体抽样权数的差异。

28. 类集设计效果 $d_{ij}^2(\bar{y})$ 取决于类集样本规模。老挝和纳米比亚调查的类集样本规模为20个住户，而柬埔寨调查的每个类集为10个抽样住户。在对各调查结果进行比较时，为了剔除不同类集规模的影响，我们对表七.8中各项估计的同质性比率（*roh*）进行了计算（见第六章的第30个等式）。结果见表七.9。同质性比率用以衡量调查变量的初级抽样单位（查点区域）所具有的内部同质性。需要考察的问题就是各国同质性比率的水平和模式是否具有相同之处。

29. 由于类集的同质性可能在城市和农村类集之间存在着差异，所以这两个地区的同质性比率是分别计算的。结果见表七.9。该表中有些结果很突出：

- 三个国家中城市/农村同质性比率差异的模式不同。老挝调查中城市类集的同质性比率一直大大低于农村类集的同质性比率。城市/农村的平均比率为0.4。在纳米比亚的调查中，差异正好相反，城市类集的同质性比率平均起来大于农村类集的同质性比率——相差的系数为1.9。在柬埔寨的调查中，城市/农村的同质性比率没有明显的差异。
- LECS调查中农村类集的同质性比率很高（范围在0.110 - 0.209之间，中值为0.178）。城市类集的同质性比率要低得多（范围在0.036 - 0.092之间，中值为0.072）。
- 柬埔寨调查中农村地区每月食品消费的同质性比率很高（0.204）。该比率要比每月总消费的同质性比率高很多，也比住户耐用品估计指标的同质性比率高。

30. 老挝人民民主共和国调查中城市和农村同质性比率之间的差异很大，这主要是因为农村地区的同质性比率很高。这些结果和该国以前的LECS调查结果一致。由于农村的村庄很小，而且其社会经济条件非常相似，所以农村地区的同质性比率很高——这不是没有理由的。另外，城市地区收入水平间隔很小，使得社会经济条件很复杂。每月总消费和每月食品消费的季节性也对这些变量有影响。每个初级抽样单位的调查时间为1个月，对初级抽样单位的抽样时间在12个月内进行。因此，在地理区域类集之上有一个“季节类集”。有理由相信农村地区的这种季节性更强一些。

31. 纳米比亚调查中，商业农业地区中大部分农村初级抽样单位的差异很大，包括高收入的农场主住户和低收入的农场工人住户。另一方面，城市地区收入水平间隔很大，在分层中只考虑了其中的一部分。这些情况说明了城市地区住户消费和家庭收入的同质性比率较高的原因。

表七.9
城市和农村地区的同质性比率

主题/特征	城 市	农 村	城市/农村比率
住户消费、收入			
• 每月总消费 (LECS)	0.072	0.209	0.3
• 每月总消费 (CSES)	0.089	0.080	1.1
• 国内住户总消费 (NHIES)	0.071	0.025	2.9
• 每月食品消费 (LECS)	0.092	0.178	0.5
• 每月食品消费 (CSES)	0.139	0.204	0.7
• 总家庭收入 (NHIES)	0.071	0.058	1.2
住户耐用品			
• 拥有电视机 (LECS)	0.049	0.178	0.3
• 拥有电视机 (CSES)	0.079	0.061	1.3
• 拥有电视机 (NHIES)	0.200	0.125	1.6
• 拥有收音机 (LECS)	0.036	0.110	0.3
• 拥有收音机 (CSES)	0.100	0.109	0.9
• 拥有收音机 (NHIES)	0.063	0.032	1.9
• 拥有录像机的住户比率 (LECS)	0.076	0.154	0.5
• 拥有电话 (NHIES)	0.208	0.125	1.7

32. 除了上述两种原因外，还应该加上两点。一个原因是消费变量的设计效果（以及同质性比率）对高端的值很敏感。在有些情况下，剔除几个最高值将会大大改变设计效果。另一个原因是同质性比率的值不只是衡量类集同质性的尺度。当不同采访者，或几组采访者在不同的初级抽样单位进行采访时，同质性比率的值还反映了采访者差异效果。

E. 讨论

33. 在表七.9中，找出不同国家在同质性比率的水平或模式方面的相似之处可能性不大。在为一项调查设计样本时，如果抽样统计员希望将类似调查中的同质性比率用于另一国家，那么表七.9中的结果用处不大。在为此处研究的社会经济变量确定类集同质性程度时，国别人口条件似乎可以起到很重要的作用。诚然，本研究非常有限。惟一可以得出的一般结论的就是：在援引另一国家调查中的同质性比率时，应当谨慎。结果还说明有必要根据最近的调查来计算和证明设计效果与同质性比率，这样就可以将结果用于下次调查的设计。

34. 然而本研究的结果是不确定的，他们和通常的结果相反。对人口与健康调查进行的各项研究发现某一特定估计的同质性比率的估计值在各个国家之

间可以调用，但样本设计应该具有可比性（见第二十二章）。同样，对许多项世界生育率调查进行的研究显示，各国同质性比率的模式存在着类似之处。人口变量的同质性比率可能比社会经济变量的同质性比率“效用更高”，因而可能更具有可调用性。

附 件

11项住户调查样本设计的介绍

11项住户调查样本设计的简介如下：

1997/1998年老挝支出与消费调查（LECS）

以普查查点区（EAs）作为初级抽样单位（PSUs）。初级抽样单位根据18个省和城市/农村地区进行分层。农村查点区则根据“有公路”和“没有公路”进行了进一步分层。每个省则根据系统的PPS调查法抽出规模相等25个初级抽样单位样本（共450个初级抽样单位）（Rosen, 1997年）。每个初级抽样单位抽取了20个住户，这样共有9 000个住户样本。各省之间的样本分配相等，造成住户抽样权数的很大差异。

1999年柬埔寨的社会经济调查（CSES）

以村庄作为初级抽样单位。有几个公社和村庄没有包括进去，由于安全的原因不能前往这些地方调查。没有包括进去的地区占该国住户总数的3.4%。

根据生态区将村庄归为5层。金边被看作单独的一层，农村和城市地区各被视作单独的一层。这样四个地理区共有10层（金边、平原、洞里萨湖、沿海和高原/山区）。从每层抽出四个独立的村庄子样本。样本分配大致和各层成比例。

通过循环系统PPS抽样，抽取了600个村庄。每个村庄内抽取了10个住户（1999年柬埔寨王国国家统计局）。

1993/1994年纳米比亚的家庭收入与支出调查（NHIES）

初级抽样单位基本上为普查查点区。在抽取前，有些小的查点区和邻近的查点区合并在一起。初级抽样单位平均规模大约为150个住户。根据城市/农村分区和14个地区进行了第一阶段分层。第二阶段分层在城市地区进行，城市地区按“城市”和“小城市”（半城市）分层。样本的分配大致和各层成比例。但城市地区抽样略微偏多。通过系统PPS方法（Pettersson, 1994年）抽取了96个城市初级抽样单位样本和123个农村初级抽样单位样本。

1995/1996年纳米比亚两次普查之间的人口调查（NIDS）

设计和NHIES相同。抽取了82个城市初级抽样单位样本和120个农村初级抽样单位样本。在NIDS调查中，每个初级抽样单位中抽取的样本数量比较多，为50个住户，共有9 500个住户样本（Pettersson, 1997年）。

1999年越南的多用途住户调查（VMPHS）

农村地区以公社作为初级抽样单位。城市地区则以区为初级抽样单位。根据城市/农村和省（61个省）进行分层。通过PPS方法，抽取了839个公社。每个省的样本规模基本上相同，但大的省份被分配的样本规模稍微大一些。第二阶段抽样单位（第二阶段抽样单位）为公社内的村庄和区内的街区。又从被抽取的每个公社内抽取了两个第二阶段抽样单位。在每个第二阶段抽样单位内，抽取了15个住户。总共抽取了25 000个左右的住户（Phong, 2001年）。

1997年莱索托的劳动力调查

样本为两阶段样本。初级抽样单位是若干组查点区。初级抽样单位平均规模为370个住户。根据城市/农村分区、地区（10）和农业经济区（4）对初级抽样单位进行了分层，

共产生了33层。样本分配和各层成比例，但有两个例外：两个小层的抽样偏多。通过系统的PPS办法抽取了120个初级抽样单位。在初级抽样单位内，通过系统的随机抽样方法抽取了15-40个住户，共产生3 600个住户样本。所有合格的住户成员都被纳入调查中（Pettersson, 2001年）。

1999年10月南非共和国的住户调查（OHS）

以普查查点区（EAs）作为初级抽样单位。在抽样过程中，通过Kish（1965年）提议的方法将住户数量不足80个的查点区和列表上邻近的查点区合并在一起。城市初级抽样单位的平均规模为80-100个住户，而农村初级抽样单位的平均规模则为110-120个。根据9个省对初级抽样单位进行了分层。通过平方根分配将样本分配于各层。在每个省内，又进一步通过区委会（和大城市委员会）进行了分层。通过系统的PPS抽样方法，抽取了2 984个初级抽样单位样本，城市地区为1 711个，农村地区为1 273个。在每个初级抽样单位内，系统抽取了10个“调查点”样本（大致和住户相同）（Stoker, 2001年）。

2000年2月南非共和国的劳动力调查

2000年2月劳动力调查第一次采用了一个新的标准样本，该样本是根据1996年普查数据库于1999年末建立的。该样本包括2 000个初级抽样单位（当年稍后的时候，该样本被扩大为3 000个初级抽样单位）。以普查查点区作为初级抽样单位，不足100个住户的查点区和邻近的查点区合并在一起。根据9个省对初级抽样单位进行了分层。通过平方根分配将样本分配于各层。在每个初级抽样单位内，形成了规模为10个调查点的类集，每个类集分布在整个初级抽样单位中。抽取了一组类集用于将来的劳动力调查。

由于预算问题，决定将劳动力调查按比例减少为10 000个调查点。其影响如下：在所有的城市初级抽样单位中，从所确定的类集中只抽取了5个调查点。就农村样本而言，从农村初级抽样单位集中抽取了PPS系统子样本，该子样本包括了50%的农村初级抽样单位。在抽取的初级抽样单位中，所确定的包括10个调查点的整个类集成为样本的一部分（Stoker, 2001年）。

1999年巴西全国入户抽样调查（PNAD）

PNAD调查的年度样本大约为115 000个住户，代表了除北部（亚马逊）农村地区以外的巴西所有地区。根据地理情况进行分层，共分为36层。在36层中，有18个州是以每个州作为一层，另有9个州是将每个州再细分为两层。位于州首府周围大城市地区的初级抽样单位形成一层，该州其余的初级抽样单位形成一层。在大城市地区形成的层中，设计为两段整群抽样，在这类抽样中，初级抽样单位为普查查点区，通过系统PPS抽样方法抽取，规模大小相当于最近一次普查中所获得的私人经济数量。在抽取初级抽样单位之前，是根据地理代码对它们进行分类的，从而使根据自治区和城市-农村状况进行的分层不明显。

在不属于大城市地区的层中，初级抽样单位为自治区。这些是根据大小和地理情况进行分层的，形成了人口规模大致相同的层（采用最新普查数据）。在每一层中，通过系统的PPS抽样方法抽取了两个自治区（这些层中的初级抽样单位），以总人口作为衡量规模的尺度。在进行系统抽样之前，有些自治区因为人口很多而被宣布为“确定”的初级抽样单位，并因此被纳入具有“确定”性的自治区样本中。在抽取的每个自治区内，通过系统的PPS抽样方法抽取了查点区，其规模大小相当于最近一次普查中私人经济的数量。在抽样的最后一个阶段，通过从每年更新的名单中进行系统抽样，在查点区内抽取住户。抽取住户的每个成员都被纳入调查中。原本应该从每个查点区中抽取13个住户的目标样本。但为了减少由于衡量规模的尺度过时而引起的权数差异，在每个查点区内使用了不变的抽样分数而非不变的抽样规模，这样就产生了不同的类集。

样本在各层的分配不成比例。最大权数和最小权数的比大致为8。

巴西1999年9月的PME调查

PME调查为劳动力调查，包括巴西六个最大城市大约40 000个住户的月样本，最新的主要劳动力指标就是得自这种调查。在大城市地区一层，样本设计和PNAD相同，但目标类集抽样除外——PME为20，而PNAD为13。

1996/1997年巴西的PPV调查

PPV调查旨在对生活水平进行衡量，采用了住户生活水平衡量研究（LSMS）方法——生活水平衡量研究由世界银行倡议并在各个国家进行（Grosh和Muñoz, 1996年）。于1996-1997年进行的巴西调查对大量的人口、社会和经济特征进行了调查，采用了4 944个住户样本，这些样本是从巴西东北地区和东南地区的554个查点区中抽取的。样本设计为两阶段分层类集样本。分层包括两个步骤。首先，形成了10个地理区域层，以确定6个大城市地区（福塔雷萨、累西腓、萨尔瓦多、贝洛哈里桑塔、里约热内卢和圣保罗），另外还要加上其他四个层，这四个层覆盖东北地区和东南地区的其余部分，并进一步细分为城市和农村查点区。在这10个地理区域层中的每一层内，又根据1991年普查中记录的每人平均家庭收入，将查点区进一步细分为3层。因此，共形成了30个层。

总样本规模固定在554个查点区，东北地区278个，东南地区276个。各层内查点区的分配和每层内查点区的数量成比例。通过概率与规模成正比的置换取样法抽取了查点区，将每个查点区的私人经济数量作为衡量规模的尺度。在抽取的每个城市查点区内，通过不置换简单随机抽样法抽取了8个固定住户。为了节省成本，每个农村查点区抽取的数量为16个住户。

尽管同PNAD和PME相比，PPV调查的抽样规模较小，但该调查提供了有关设计效果方面的有用信息，因为它对查点区直接进行了收入分层，而且每个查点区的取样规模也比其他调查规模小。另一个重要特征就是估计仅使用标准的反向抽样概率权数，没有使用人口预测校准法。PPV调查中样本权数差异很大，最大权数是最小权数的40多倍。

参考资料

- Grosh, M. 和 Muñoz, J. (1996年)。《规划和实施生活水平衡量研究手册》。
《生活水平衡量研究工作文件, 第 1 2 6 号》。哥伦比亚特区华盛顿: 世界
银行。
- Kish, L. (1965年)。《抽样调查》。纽约: 威利国际出版公司。
- 柬埔寨王国, 国家统计局 (1999年)。《1999年柬埔寨社会经济调查: 关于调
查设计和实施的技术报告》。金边。
- Pettersson, H. (1994年)。《标准样本设计: 特派团向国家中央统计办公室
提交的报告, 纳米比亚, 1994年5月》。国家咨询办公室, 瑞士统计局。
- _____ (1997年)。《1992-96年标准样本业绩的评估: 特派团向国家中央
统计办公室提交的报告, 纳米比亚, 1997年5月》。瑞典统计局, 国际咨
询办公室。
- _____ (2001年)。《住户和商业调查的样本设计: 特派团向统计局提交
的报告, 莱索托, 2001年5月21日至6月2日》。瑞典统计局, 国际咨询办
公室。
- Phong, N. (2001年)。关于1999年越南多种目的住户调查样本设计的个人信
函。
- Rosen, B. (1997年)。《1997年老挝标准样本的设立。特派团向国家统计中心
提交的报告, 老挝人民民主共和国》。瑞典统计局, 国际咨询办公室。
- Stoker, D. (2001年)。关于南非共和国10月住户和劳动力调查样本设计的个
人信函。
- Verma, V.、C. Scott和C. O'Muircheartaigh (1980年)。“世界生育率调查的样
本设计和抽样误差”。《皇家统计学会志》, A辑, 第143期, 第4部分,
第431-473页。

C 部 分

非抽样误差

导 言

James Lepkowski

密歇根大学
美国密歇根州安阿伯

1. 本出版物之前的各部分及章节已经较全面介绍了从某个人口群体抽取代表性概率样本时所出现的抽样误差。本部分将介绍住户调查中出现的其他误差。和抽样误差一样，这些误差中的某些误差会随着可能抽取的样本发生变化或随着测量过程的重复而发生变化，另一些误差则是不变的或是系统的，不会因样本的不同而不同。

2. 在样本设计框架中，变量误差通常又被称作抽样方差。有些属于固定抽样误差（其中有一些已经介绍过），固定的抽样误差又被称作偏差。如，有意将人口群体的某个子群排除在外，就会使该人口子群不在调查范围之内，这样不管抽取哪个样本，都将会出现一个大小相同的误差。

3. 非抽样误差涉及非观察误差或测量误差，非观察误差是在不能从某个抽样单位或某个变量中获取资料时发生，测量误差是在收集调查变量值时发生。非观察误差通常是固定不变的，通常需要对调查估计值的偏差进行考虑。测量误差有时是固定的，但也可以是变动的。

4. 在非观察误差中，有两个重要的误差来源：未覆盖误差和不答复误差。在概率抽样中，必须对人口因素有一个完善的定义，每个人口因素有一个非零的选择机会。当人口群体某个因素实际上没有选择机会的时候，就会出现未覆盖的问题。该因素无法进入抽取的样本中。不答复是指某个要素已经被选入样本，但却没有收集到有关该要素的反应数据。这是有可能发生的——因为某个家庭或某个人可能不愿合作，或者因为存在着语言障碍、健康限制或调查期间家中无人。

5. 计量误差的来源有很多，包括应答者、采访者、监督者甚至数据处理系统。如果一个应答者忘记了所需要的信息而做出了不正确的回答，或在回答敏感性问题时扭曲信息，那么就可能出现应答计量误差。这些应答误差很可能构成偏差，因为不管什么时候提问题，应答者都常常会忘记或扭曲某些信息。这些误差也可能是变动的。有些应答者可能在某个时候忘记了某些信息，又在某些时候想起某个信息。

6. 有关这几种误差，有四个方面需要调查设计者考虑。第一，需要对误差进行认真界定，并检查调查过程中误差的来源，包括调查过程中可能是哪一部分出现了这种误差。第二，应该考虑如何计量误差的大小——这是一个很棘手的问题。第三，应该制定各种办法来降低误差的程度，当然这时常需要补充调查资源。最后，每项调查都会出现非抽样误差，调查设计者应该在调查结果中努力弥补这些误差。

7. 本部分第八章和第九章从概念的角度分别考察了非观察误差和计量误差，对各种不同种类的这些误差进行了说明。第十章和第十一章更详细地介绍了这些误差的处理办法，第十章分析了这些误差对调查结果的总体影响，第十一章以一个国家——巴西为例，对这些种类的误差进行了个案研究。

第八章

发展中国家住户调查中的非观察误差

James Lepkowski

密歇根大学

美国密歇根州安阿伯

摘 要

调查中采取非观察做法是因为不能对某些目标人口群体或样本进行计量。非观察可能是整体的，在这种情况下，不对某个单元（如某个住户或某个人）进行任何计量；非观察也可能是部分的，在这种情况下，只对某个单元进行某些（但不是全部）计量。本章将讨论非观察的两个来源：未覆盖和不答复。未覆盖是指相关人口单元没有机会被调查抽取。不答复是指被调查选中的住户或个人不参与调查，或虽然参加了但却不提供完整信息。本章将考察非观察的原因和结果，以及对非观察误差进行补救的措施。当那些没有被考虑的人口或样本不同于所观察的人口或样本时，未覆盖和不答复可能会使调查估计值出现偏差。由于这些偏差可能很严重，所以本章将对未覆盖和不答复的若干补救措施和调整办法进行讨论。

关键词：不答复、未覆盖、偏差、目标人口、抽样标架、反应率。

A. 引言

1. 调查研究中的非观察是由于没有对调查目标人口的某个部分进行计量所致。非观测可能是整体的，这种情况是指不进行任何计量；非观测也可能是局部的，在这种情况下，只进行某些（而非全部）计量。

2. 非观察的一个显著来源就是抽样过程。由于普查是一种旨在对人口所有因素进行计量的调查，所以只有在普查中，才不会因为抽取样本而出现非观察。源于抽样的非观察会导致抽样误差，这个问题将在本出版物第六章和第七章进行讨论。本章将不讨论这种来源的非观察。

3. 本章将讨论非观察的其他两个来源，也就是未覆盖和不答复。正如稍后将要进行详细介绍的那样，未覆盖是指相关人口单元没有机会被调查抽取；不答复是指某个被抽取的单元完全不或部分不参与调查。本章将对这些非观察的原因、结果、减少这些情况的措施、降低调查估计值偏差的方法进行讨论。

未覆盖和不答复的结果包括：调查结果可能出现偏差。如果没有被考虑的人口不同于被观察的人口，那么调查结果和人口实际情况之间就会出现差异。这种差异就属于非观察偏差，非观察偏差可能会很严重。

4. 当然，即使不对部分人口进行计量，也有可能不会出现非观察偏差。尽管记录非观察的情况比较简单，但发现无观察偏差却很困难。正因为如此，才会使得非观察偏差很少成为研究的主题。有时非观察对整个调查或大多数调查问题一点影响都没有，这方面的例子是有的。有时非观察会导致某个问题的调查估计出现很大偏差，或导致一组问题的估计出现很大偏差，从而导致调查的所有结果都不可信，这方面的例子也是有的。

5. 有关非观察的研究有很多。本章只能对住户调查中未覆盖和不答复误差的性质进行介绍。至于更详细的情况，读者可以参看所提供的参考资料。下一部分将介绍区分未覆盖和不答复的基本框架，其后各部分将分别对每种误差的来源进行讨论。

B. 理解未覆盖和不答复误差的基本框架

6. 了解未覆盖和不答复之间的差异就需要了解人口群的性质和抽样的框架。目标人口是由许多元素组成的，调查设计人员希望就这些元素进行调查估计。如，可能会要求一个调查设计人员设计一项调查来研究某个国家中年龄在15岁或15岁以上的劳动力参与情况。人口群显然有界定明晰的地理限制（国家的边界）以及单位特征方面的限制，如年龄限制。

7. 目标人口的定义还有其他不太明显的方面。如，居住在一个国家中人的含义。很多调查采用居住的定义，根据该定义，一个人必须在过去的绝大部分时间居住在某个国家，或者在刚迁移到一个国家的情况下，必须打算在这个国家永久居住。人口的某些部分可能不在某些调查主题的范畴内。如，根据定义监狱里的人员或诸如军事之类的机构可能不在某些经济条件调查的范围之内。这些机构可能被排除在外，因为它们包括了不在计量概念范畴之列的人员。目标人口定义本身也有时间界限。调查所感兴趣的很可能是当前的劳动力参与情况而非个人的历史情况。如果这样，那么调查是对某个时间点的人口特征进行估计。

8. 目标人口是推断出的人口。调查结果将最终被视为特定人口的调查。调查通常是为了对特定国家的人口特征进行衡量。不管抽样程序是否包括一个国家的某些人，调查的最终报告都有可能不加限制地做出有关该国整个人口的陈述。如，尽管调查将某个机构的人员排除在外，但最终报告可能会说调查结果适用于居住在这个国家的所有人口，不知情的读者可能会由此认为有关结果也能代表该机构的人员，但抽样过程并不包括这些人员。因此，在对调查进行介绍时，就有必要在有关调查的出版物中对目标和调查人口进行认真而全面的描述。

9. 目标人群通常不同于另一种被称为抽样标架的人口群体，也就是样本实际上据以抽取样本的一组元素。抽样标架是一组用以抽取样本的材料，可

能不会和目标人口完全一致。如，在有些国家，公共安全部门（如公安部门）编制的地址登记簿被用作抽样标架。人口群中的某些家庭有可能不在这些抽样标架中，这样抽样标架就不同于目标人口群。

10. 在其他情况下，由于结构或故意作为的缘故，抽样标架会不同于目标人口。由于行政管理或成本方面的考虑，有一部分人口可能不在框架范围内。如，一个国家的某一个地区、若干个行政区或某一个省目前可能出现内乱，公安部门可能对该地区的出入进行限制。这样尽管有在该地区进行抽样的材料，调查设计人员还是有可能故意将该地区排除在框架之外。

11. 将某一部分人口排除在外，也可能是基于成本方面的考虑。在很多国家，那些居住在边远且人口稀薄地区的人群被排除在抽样标架之外，因为如果对他们进行抽样，调查成本会很高。另外在一些有很多本土语言的国家，翻译和招聘能说所有语言的采访者需要很高的费用，调查设计者可能会和主办单位一道，专门将那些不能说一种该国主要语言的人员排除在外。在这种情况下，只有在确定一个家庭以及该家庭成员的语言能力之后，才有可能将某个人排除在外。主要通过家庭中的筛选来决定取舍。

12. 另一方面，调查设计者可能会将这种问题视作不答复，也就是由于语言被排除在外的缘故被看作未覆盖，由于无法交流的缘故则被看作不答复。哪些语言应该被排除在外的决定在某种程度上取决于问题的大小。如，某一个国家的调查可能限于那些能说一种官方认可语言的人群，这种决定可能会将很多不能说官方认可语言的人群排除在外。相反，在其他国家，几乎所有人都能说一种官方语言，少数人说的是非官方语言，因而没有这种语言版本的调查问卷，可能会和这一小部分人进行接触，但不会采访他们。在前一种情况下，最好进行认真记录，并将其语言被排除在外的人群列为未覆盖。在后一种情况下，最好将没有采访的人群列为不答复。

13. 当目标人口中有一些因素和抽样标架中的列表不吻合时，会发生未覆盖问题。在住户调查中，最典型的未覆盖问题是在以下几种情况下发生：现场准备的列表没有将住宅单位包括在内，使用了过时的或不准确的行政管理住户清单，或当居民住户清单没有将某个住户内的成员包括进去。

14. 未覆盖是指人口群体某个要素没有机会被选作调查的样本，不答复是指没有能够从一个被抽样的合格单位——目标人口中的一个单元中收集到信息。未覆盖问题是在被用作抽取样本的框架存在误差或问题时出现的；而不答复问题是在框架已被确定、从框架中抽取样本元素之后发生。如，假设在一个被抽样的住户中，因为一项临时性的工作，该住户男居民需要在本村外呆一周，所以在采访时他不在家。如果在首次采访期间，提供住户信息的人把他忘了，因而该居民没有被列入住户花名册，那么就会发生未覆盖问题。另一方面，如果一个居民被列入花名册，可在采访村子期间，他不在家，那么通过调查得到的只是该居民自报的数据，因不是从他那里收集到的数据，这种情况就属于不答复。

15. 未覆盖通常会涉及整个单元，如家庭或个人。不答复可能涉及整个单元或个别数据项目。如，由于某个住户住在一家零售店的楼上，所以村子的花名册没有将该住户包括进去，在这种情况下，整个单元不在抽样标架之列，因而属于未覆盖。在以下情况下可能发生不答复问题：被列入框架的住户不愿参与调查；家庭的某些成员配合调查并提供了数据，但另一些成员不在家或拒绝对所有调查做出回答。这两种形式的单元或全部不答复（家庭或个人）和另一种情况相反，在另一种情况下，家庭的某个成员除了对个别项目没有做出回答外，对其他所有调查问题都做出了回答。如，一个家庭成员可能担心官方会对未报告的收入采取行政措施，因而不愿提供其在非正式经济领域的收入数据。后一种不答复就是所谓的项目不答复。请注意，在这种情况下，不答复的类型还取决于所分析的单元是个人还是住户：如果分析单元是住户，那么个人不答复就属于项目不答复，但如果分析单元是个人，那么个人不答复就属于单元不答复。

16. 权衡未覆盖和不答复之间的关系也很重要。尽管在某个调查中，可以通过认真研究发现未覆盖或不答复的很多来源，但人们可能希望降低这些问题所引起的误差——而这又需要支出有限而稀缺的资源。为降低这两个来源所引起的误差，可能存在着对稀缺资源的竞争。

17. 如，某个国家有40种主要语言或方言，调查工具被翻译成5种语言，占人口80%的家庭说这5种语言。第六种最常用语言的人群占整个人口的3%。同时假设调查活动规定：为了确定有关人员是否在家，将在两天内对一个住户走访两次，而且被走访两次的家庭中，有10%的家庭为不答复，因为在调查采访的两天期间，没有人在家。在这种情况下，调查设计者需要在资源方面做出选择：可能会花费更多的资金将调查工具翻译成第六种语言，以便将占人口3%的这第六种语言的人群包括进去；也可能会花费更多的资金让采访者在每个村庄走访三天或四天以便提高在家住户成员的比率。

18. 额外调查资源究竟是用于翻译，还是用于延长走访家庭的次数，有关这方面的决定将取决于预计偏差的大小以及所涉及的成本和资源。偏差既取决于未覆盖或不答复的程度，又取决于已覆盖人口和未覆盖人口之间或有反应和不答复抽样人群之间的差异。

19. 在调查设计中，时常需要进行这种成本与误差之间的权衡。至于用以进行这种权衡所需要的数据以及如何进行权衡，其详细情况不在本章讨论的范围之列。大多数调查中，进行这种权衡的信息是有限的，所做的权衡也是非正式的。

C. 未覆盖误差

1. 未覆盖起因

20. 在住户调查中，未覆盖起因在于用以抽取样本的框架材料。由于发展中国家以及某些转型期国家的很多住户调查都采用区域抽样法，所以在现有讨论中，框架和未覆盖问题将限于根据区域抽样法进行的住户调查。

21. 区域抽样通常要结合进行多阶段抽样。初级抽样以及有些时候的第二阶段抽样涉及可能被视作住户类集的地理区域。在某些后期的抽样阶段中，必须为一组相对较小的地理区域获取或编制一张住户清单。在抽样的最后阶段，为每个被抽样的区域编制了一份住户成员或居民清单。因此，在考察这种调查的未覆盖问题时，需要考虑三类单元：地理单元、家庭和个人。正如稍后将要讨论的那样，这些单元也可能成为住户调查中不答复的来源。

22. 抽样标架缺陷引起的地理单元未覆盖问题并不多见，因为作为大多数区域框架基础的普查材料覆盖了某个人口群的整个地理区域。地理区域未覆盖的问题确实会发生，但正如以上所述，这种情况并不多见。一项调查可以通过设计来对一个国家的整个人口或一个国家的地区进行推断，最终报告中所提到的人口确实可以包括居住在整个人口的人口，但可以不从整个国家中抽取样本。

23. 如，在调查设计中，调查设计人员可以确定一些区域，这些区域的人口在整个人口中所占的份额很有限，因此将其纳入调查的成本非常高。他们可能会有意将这些地理区域排除在框架之外。然而在对调查结果进行报告时，不会提到或只是简单提到这些被忽略的地区。报告的读者可能会以为，或报告本身可能会让读者觉得调查结果适用于整个国家或地区——可实际上却有一部分人群没有被包括进去。实际上，在这种情况下，未覆盖误差一般很小，因而通常忽略不计。

24. 应该记住的是——预期的目标人口（也就是居住在一个国家整个地理区域的人口）和居住在所覆盖地理区域内的有限“调查人口”之间存在着区别。如果记录不完整，那么数据的使用者就有可能以为调查样本覆盖了整个人口，可事实上却并非如此。

25. 未覆盖问题的一个更重要的起因就是家庭。大多数调查都将家庭看作通常居住在一个住宅单位内的一群人。因此有两个因素很重要：常住居民的定义和住宅单位的定义。

26. 住宅单位的定义很复杂，因为需要考虑某个建筑物是否被用作住处；居住在该建筑物内的人员是否吃住彼此分开（诸如公寓楼之类的多单元楼房就属于这种情况）。分开居住意味着居民可以直接从建筑物的外面或从公用的大厅或走廊进入居住区，“分开吃”通常意味着有一个地方做饭，或者居民完全可以自由选择所吃的东西。

27. 很难将这种广义的定义用于各国或各国不同地区的众多生活情况。大多数住宅单位都很容易确定，如单一或独立家庭住宅、有不同入口但却共墙的联式住宅、多单元建筑内的单元住宅等。但有很多住宅单位却很难分类或发现。如，在城市贫民区，当人们居住在由回收材料或废旧材料搭建的建筑物内时，可能很难确定独立的住宅单位。另外还有些住宅单位，如果不从街道、小巷或小道入口处细心观察，将无法发现这些地方。

28. 在农村地区，可能很容易确定用作住宅的建筑物，但如果一个建筑物内的社会安排很复杂，那么就很难确定独立的住宅单位。如，在部落人群中，

被用作住宅的是一个长条的房屋，其中只有一个入口，房屋有独立的隔间供各家睡觉用，但集体或个人家庭做饭的地方却是共用的。也就是说独立的隔间本身并不是居住单元，因为它们没有独立的入口或做饭区域。在这种情况下，住户的概念——通常居住在一个特定住宅单位内的一组人——就很难适用。因为根据这一概念，不知道是应该将整个建筑作为一个住宅单位呢，还是应该将每个隔间作为一个住宅单位。实际上，应该将整个长条的房屋作为一个住宅单位或居住区，如果对其进行抽样的话，那么在现场列出住户清单时，所确定的所有住户都被纳入调查中。

29. 也有一些住处不被视作住宅单位。由他人监管而由个人居住的机构住宅，如孤儿院、监狱或医院不被视作住宅单位。学生宿舍、男修道院、女修道院、无家可归人员的庇护所属于特殊类型的住处，这一类住处不一定需要提供与机构住宅有关的监管。过渡性或季节性住处也是一个问题，如农村地区可能有单独的住宅单位供季节性劳动人员居住，每年只有一个季节或几个季节占用这些住宅。季节性居民通常有可能居住在其他地方，因此不应将其包括在季节性单元的住户中。

30. 在发展中国家中，多阶段区域抽样要求在调查过程的某个点为较小的地理区域（如城市的某个街区或农村地区的某个查点区）编制住宅单位清单。当兼职调查人员被派往现场编制住宅单位清单时，如果遇到上述有着复杂安排的住宅，那么常会出现未覆盖问题。大多数住宅单位的确定都很简单，但还是会时常发生遗漏住宅单位的情况，因为在将有着几个组成部分的定义用于有着复杂安排的住宅时，兼职调查人员的经验还很有限。

31. 由于列出住宅单位清单的行为属于暂时性的行为，所以未覆盖问题的解决就变得更加困难。在列出住宅单位清单时，某个住宅单位可能没有被居住，或者正在兴建当中。如果在未来某个点进行调查，那么可能需要将这类单元列入清单。如果在多轮单一固定样本调查或若干个不同调查中使用住宅单位清单，那么通常会努力将那些没有被居住或正在兴建中的建筑单元包括进去。

32. 转型期国家的调查中，也有可能使用由某个行政管理机构编制的清单。但在将其用于住户调查中时，需要认真考虑其质量。以上列出的、在编制调查清单时可能出现的问题也会在行政机构编制的清单中出现。

33. 因此，在列出住宅单位清单的过程中，某些类型的住户可能没有被纳入进去。如果不额外投入大量的调查资源，可能很难发现这种未覆盖情况。

34. 最后，在一个被抽样的住宅单位内，列出常住居民的清单也是列出住户清单过程的一部分。需要有操作规则来指导采访者，以便让他们了解应该将哪些人作为常住居民纳入住宅单位中。正如住宅单位的情况一样，有关这方面的很多决定都是很简单。在联络的时候，所遇到的大多数人都在住宅单位内，因为这是他们惟一的住处。当然，也会有一些人尽管只有一个住处，但在联系时却不在住宅单位内。

35. 但对于有些人来说，该住宅单位是其数处住宅当中的一处。在这种情况下，兼职调查人员就应在现场做出决定，以确定被抽样的住宅单位是否属于此人的常住地。也很难让提供住户信息的人准确报告某些居民的居住安排情况。代理另一居民所报告的这种信息可能不会完全准确。

36. 被调查的人也有可能出于个人的原因而故意将其知道属于常住居民的人员排除在外。如，某个住户正在领取政府津贴，但某个人住在其住宅单位内就有可能使其没有资格接受政府津贴。另外，由于财务问题（如债务）或法律问题（如犯罪活动）的缘故，被调查的人还有可能隐瞒那些不希望被公共机构或私人机构发现的居民。

37. 被调查人员也有可能出于文化或认识方面的缘故而隐瞒住户中的某个成员。被调查人员可能不会报告一个不足一周岁的婴儿，因为其文化认为这些婴儿还没有达到成为人的年龄。如果他们认为调查机构对儿童数据不感兴趣，那么也有可能将婴儿排除在外。有些人纯粹是忘了将某人包括进去，不管是婴儿还是年纪更大一些的人员。

38. 这样，住户调查中的未覆盖问题可能是由各种定义和调查环境引起的。如果这一问题会导致调查结果出现误差，就应予以重视。

2. 未覆盖误差

39. 假设某人口群人数为 N ，该人口群的某个特征为 Y ，该人群中有 N_{nc} 人不在调查抽样标架内，需要估计 Y 的平均值。假设人数为 N 的人口群的平均值为 \bar{Y} ，在抽样标架内的人口平均值为 \bar{Y}_c ，不在抽样标架内的人口平均值为 \bar{Y}_{nc} 。与未覆盖有关的误差又被称作样本平均值的未覆盖偏差 \bar{y}_c ——这只以包括在样本中的人口为基础，它实际上是估计数 \bar{Y}_c 而非 \bar{Y} 。

40. 样本平均值的偏差 \bar{y}_c 取决于两个因素：不在抽样标架内的人口所占的比率，以及已覆盖人口平均值和未覆盖人口平均值之间的差。其公式如下：

$$B(\bar{y}_c) = (N_{nc}/N)(\bar{Y}_c - \bar{Y}_{nc})$$

41. 这种未覆盖偏差公式有助于了解调查设计人员是如何处理未覆盖问题的。为了减少未覆盖方面的误差，或减轻未覆盖误差的影响，调查人员要么使已覆盖人员平均值和未覆盖人员平均值之差变小，要么使不在抽样标架内的人口所占的比率变小。

42. 该公式的一个问题就是，在大多数调查中都难以知道 $(\bar{Y}_c - \bar{Y}_{nc})$ 之差或 (N_{nc}/N) 之比率。另外，未覆盖率 (N_{nc}/N) 还有可能随着子集的不同而各异。差也会随着变量和人员子集（如，根据某个人口特征——如年龄——定义的地区或子群）的不同而各异。这样，未覆盖误差不是调查的特性而是个别特征的特性，它是一种估计的统计值。

43. 在很多政府调查机构中，通常需要估计一个总值。与总值有关的未覆盖偏差不仅取决于所调查特征的已覆盖单位和未覆盖单位之间的差，还取决

于未覆盖的数量（而非比率），也就是说，如果回答者的估计总数为 $\hat{Y}_y = N\bar{y}_y$ ，那么偏差为： $B(\hat{Y}_y) = N_{nc}(\bar{Y}_y - \bar{Y}_{nc})$ 。

未覆盖误差的减少、衡量与报告

44. 在家庭调查中，处理未覆盖误差的方法大概有四种：

- 通过改进现场调查办法减少未覆盖的情况。
- 制定计量未覆盖误差大小的办法，并在调查中包括误差的水平。
- 努力通过统计调整来弥补未覆盖误差。
- 在调查报告中，尽可能全面地报告调查的未覆盖特征。

45. 在家庭调查中，通常通过使用多个抽样标架，或改进调查中编制清单的办法来减少未覆盖误差。多个抽样标架更多地被用于住宅单位而非人员调查，它们需要有单独的住宅单位清单——这为现场编制清单带来了特殊的问题。

46. 如，假设某个特定国家中，很难在现场正确列出农业工人季节性住宅单位的清单。同时，还假设在农场或企业聘用了季节工人并为其提供住处的情况下，负责农业生产、教育或社会福利的某个机构拥有有关这些农场或企业中季节性住宅单位的数量或类型方面的清单。在这种情况下可以将另一个来源的季节性住宅单位清单作为一个单独的抽样标架。对于编制住宅单位清单的现场采访人员来说，如果机构清单中已经包括了他们将要编制清单的领域，那么就应该为他们提供一张这些机构所拥有的农场或企业清单，并告诉他们不要编制这方面的季节性住宅单位清单。这样，用于调查的住宅单位样本将从采访者编制的住宅单位清单中以及政府机构保留的清单中抽取。毫无疑问，两种清单都有可能存在未覆盖问题，也有可能存在“覆盖过度”的问题。但使用两种抽样标架可以减少未覆盖的情况，也可以减少由此引起的误差。

47. 还有必要考虑改进编制清单过程的办法。如果某个行政管理机构可以提供住宅单位清单，那么在抽样前，应该通过现场矫正对这些清单进行检查。在将采访者派往地理区域时，可以让他们带一张行政管理机构提供的住宅单位清单，并告诉他们在检查有关地区时如何根据所提供的清单对住宅单位进行核查、增添或删除。

48. 还可以对采访者进行培训，让他们学会在现场使用“半开区间”的办法，以便发现行政管理机构清单或现场清单中漏登的住宅单位。根据半开区间的办法，需要从一个地址清单中选择一个住宅单位，由一位采访者对被抽样单位进行走访，另外还需要有一张明示或暗示的排序清单。到了被抽样的单位，应该要求采访者调查在被抽取的住宅单位和清单上的下一个住宅之间是否还有其他的住宅单位。

49. 清单中下一个单位是通过地理区域中事先确定的某种路线来确定的。如，在城市的街区，要求编制清单的采访者从某个特定的拐角开始，然后围绕街区按顺时针方向进行。住宅单位清单就是按照这种顺时针方向编制的。

50. 如果采访者发现某个住宅单位不在清单范围内，但在被抽取的住宅单位和清单中的下一个单位之间，那么应该要求采访者将遗漏的住宅单位加入样本中，并进行采访。如果有若干个遗漏的单位，那么采访者需要征求调查中心办公室的意见，以避免引起现场调查活动的混乱。

51. 在住户内，改进编制清单办法可能需要有调查问题方面的顺序表，以便让采访者向提供住宅单位信息的人了解遗漏的人。如，可以要求调查采访者调查是否有任何婴儿不在常住居民清单之列。要求采访者选择合适的被调查人，或要求他们向被调查者重念人员清单上的名字以确保没有人被遗漏，也可以改进编制住户清单的办法。

52. 评估未覆盖偏差也是应该考虑的一个重要因素，当然这是一个难以解决的问题。调查机构如何发现那些没有纳入其清单中的单位呢？计量未覆盖通常是一项费用昂贵的调查工作，所以不经常进行。

53. 评估未覆盖偏差的一个常用办法，是对那些可以进行比较的变量的调查结果进行比较，调查结果可以从外部或独立机构获得。为了评估未覆盖的大小情况，调查可以将其抽样人员的年龄和性别分布情况与最近一次普查或行政记录中的分布情况进行比较，两者之间的差异将能说明未覆盖问题的情况。为了计量某个变量的未覆盖误差，可以将有关统计值与某个独立机构的统计值进行比较。如，调查中所报告的有关整个样本和主要子群的总工资收入可以和行政机构报告的工资收入进行比较。在传统的研究中，Kish和Hess（1950年）将一项调查的住宅单位分布情况与最近一次普查中有关街区住宅单位分布的数据进行了比较。这种比较有助于对调查数据收集过程中未覆盖问题的性质做进一步的了解。

54. 可以通过双系统计量办法或相关个案比较法来对未覆盖误差进行计量，但这种方法的费用会更高。普查采用双系统方法来对普查活动的覆盖情况进行计量[如，可以参见Marks（1978年）]。在普查中，将一项独立的调查和普查结果进行比较以发现未覆盖问题。未覆盖大小的评估需要将调查样本和普查元素逐一进行比较，以发现哪些样本元素没有在普查中出现。这些办法和动物群落环境研究中使用的“捕捉-再捕捉抽样”法密切相关。

55. 由于住户调查普遍受未覆盖误差的影响，所以在统计过程中，很多调查采用事后分层或人口控制调整的统计办法来对调查结果进行调整，以便弥补未覆盖误差。这些调整办法和以上用以计量未覆盖误差大小的方法非常接近。如，根据年龄和性别划分的样本分配情况可以和外部机构（如最近的普查或人口预测）根据年龄和性别划分的样本分配情况进行比较。如果某个年龄-性别组的样本分配偏低（或偏高），那么应该对年龄-性别组的所有抽样人员数据

进行加权，以便增加（减少）其在调查结果中所占的份额。在分析中，为正确处理加权问题，需要有经过加权的估计值。

56. 对于未覆盖问题应该考虑的最后一个因素就是一个良好的报告，这对于任何一个统计机构都很重要。分析报告应该对目标人口，包括除外项目，做出明确的定义。抽样标架的描述应该足够详细，以便读者能够看出未覆盖问题可能会怎样产生，甚至对潜在误差的大小做出非正式的估计。将有关抽样标架的质量评估情况纳入参考资料或附件中也很有帮助，如住宅单位清单或行政管理机构清单的质量检查情况、住宅单位人员的原有清单与为进行质量控制评估而进行的再采访时所编制的清单进行比较的情况。

57. 一个更困难的问题就是对人口或人口子群的覆盖率或未覆盖偏差进行报告。只有当所开展的调查已在某个时间对非未覆盖问题进行了估计时，才有可能进行这类估计，对一次性横断面调查进行这种估计十分困难，或根本不可能。

58. 最后，如果进行了事后分层或人口控制调整，那么调查文件必须对人口重要子群的调整程序以及调整大小进行描述。

D. 不答复误差

59. 不答复误差的定义、计量、减少、弥补和报告与未覆盖误差有很多相同之处。因此，现有部分的安排和第三部分的安排很类似。但必须明确的是不答复和未覆盖是完全不同的两个问题，它们各自的来源不同，各自的解决方案在某些情况下也有所不同。在未覆盖的情况下，除了未覆盖人口部分的位置和一般特征外，调查设计者几乎从不知道其他方面的情况，在不答复的情况下，他们至少知道不答复者的框架信息。一般认为住户调查中不答复问题更普遍，因此其对调查估计偏差的影响也可能更大。

60. 如上所述，在住户调查中通常会发现两类不答复问题。这两类不答复问题对调查结果的影响有很大的不同。用以计量、减少、报告和弥补它们的方法也在某些方面有所不同。尽管每一类不答复问题都可以用一整部分的篇幅进行分析，但本部分将把这两类问题放在一起讨论。

1. 住户调查中不答复的起因

61. 在住户调查中，若干不同种类的单位都有可能发生单位不答复问题。正如未覆盖问题一样，初级抽样单位和第二阶段抽样单位都有可能发生不答复问题。如，初级抽样单位可能包括一个国家的某个区或分区。由于天气条件或自然灾害的原因，可能无法在第一抽样阶段或第二抽样阶段已经抽取的某个地区或分区中进行调查活动。因此虽然调查中包括该单位，但在调查期间却无法从该单位中的任何住户收集资料。

62. 住户调查更有可能发生不答复问题。名单上已经被选作样本的住宅单位可能有人住，而且也尝试进行过采访。但当采访者走访住宅单位时，可能会

有几个不利的事件会阻止采访者收集资料。住户成员可能不愿以个人的身份或作为整个单位的代表来参与调查。

63. 尽管住宅单位有人住，但住户可能在整个调查期间都不在家。在有些发展中国家，这类问题很多：住宅单位显然有人住，但在收集的整个期间却一直锁着。

64. 在很多国家，尽管在收集数据期间住宅单位有人在家，但语言却成了一个问题。调查问卷可能没有被翻译成该住户所使用的语言，或采访者可能不会说当地的语言。为了避免不答复问题，可以就地聘请翻译人员陪同采访者走访家庭并进行翻译。其他调查拒绝采用这种办法，因为不知道翻译是否正确，也不知道在各户的翻译是否一致。然而，可以将那些由于语言障碍而不能提供回答的住户列为不答复单位。作为一种变通的办法，如果住户不能讲一门已经翻译过的语言，那么有些调查机构将把他们排除在调查之外。这些住户就成为未覆盖住户而非不答复住户。不管是将这类住户视作未覆盖，还是视作不答复，调查机构都必须在调查文件中对此进行明确说明。

65. 个人不答复也有可能发生。在调查允许代理回答调查问题的情况下，如果某个住户有人在采访期间不在家，那么可以从该住户的其他成员那里收集资料。在调查要求被调查者对某些问题或所有问题都亲自回答的情况下，如果由于一个人在调查期间不在家、拒绝回答或有其他障碍（如语言）而使调查无法进行时，那么就会出现不答复问题。健康状况，不管是长期性的，如听力障碍或失明，还是暂时性的，如一段严重的急性疾病，都有可能让一个人无法做出回答。

66. 至于有语言问题的住户，有些调查机构将有语言障碍或永久性健康问题的人员纳入未覆盖范畴，而将那些有临时性健康问题的人员纳入不答复范畴（Seligson和Jutkowitz，1994年）。至于如何进行这方面的分类，并没有一个被广泛认可的规则。对于收入或支出调查，有临时性健康问题的人员数量不多，因此调查机构可以不将他们纳入调查。然而对于健康状况调查，这些人的回答可能有很大的不同，因而如果将他们排除在调查之外，将可能引起令人担忧的问题。在这种情况下，可以将他们列为不答复范畴。由于缺乏一个被广泛认可的做法，所以调查机构应该在调查报告中准确报告某个调查是如何处理这种问题的。

2. 不答复偏差

67. 在住户调查中，对不答复问题的研究要比未覆盖问题的研究多得多[如，见Groves和Couper（1998年），以及Lessler和Kalsbeek（1992年）的评论]。研究的这种侧重点与若干个因素有关。

68. 在某种程度上说，未覆盖问题不如不答复那么明显。可以说根本没有未覆盖住户或个人方面的资料可用以研究。但不答复单位却是可以观察和计算的，还有可能说服他们参与调查。

69. 发达国家有一种看法认为未覆盖问题不像不答复问题那么重要，因为未覆盖率要低于不答复率。发展中国家的情况可能正好相反，在发展中国家，不答复率较低，而未覆盖率远远高于发达国家。请回忆一下：样本平均值的未覆盖偏差取决于两个因素——未覆盖率的大小，以及已覆盖人员平均值和未覆盖人员平均值之间的差。同样，对于不答复来说，样本平均值的不答复偏差可能取决于不答复人口所占的比率，以及已反应人群平均值和没有反应人群平均值之间的差。

70. 和未覆盖的情况一样，假设人口群的某个特征为 Y ，人口平均值 \bar{Y} 包括做出反应人群的平均值 \bar{Y}_r 和没有做出反应人群的平均值 \bar{Y}_{nr} ，需要估计 Y 的平均值。假设被抽取但没有做出回答的人数为 N_{nr} 。那么回答者的样本平均值偏差 \bar{y}_r 的公式如下： $B(\bar{y}_r) = (N_{nr} / N)(\bar{Y}_r - \bar{Y}_{nr})$ 。至于未覆盖问题，调查设计者要么必须使不答复率很低，要么使有反应住户、个人和不答复住户、个人之间的差异很小。这种一般性的框架可以用来理解调查中项目一级的不答复问题。然而项目不答复偏差的问题更复杂，因为通常会将各个项目结合在一起考虑，所以项目不答复是若干个项目不答复的结合。

71. 在未覆盖的情况下，未覆盖的差异或比率都不知道；但在不答复的情况下，如果调查设计周全的话，那么可以对未覆盖率做出比较准确的估计。设计周密的调查可以保留被选作研究的每个样本单位（不管是住户、人员还是个别数据项目）的详细部署记录。因此可以直接从调查数据对不答复率进行估计。如果重要子群之间，尤其是住户地理子群之间的反应率不同，那么也有数据可供观察。

72. 估计反应人员和不答复人员之间的差异需要收集和衡量更广泛的数据。在收集调查数据期间通常无法对不答复人员的有关特征进行衡量。但在调查过程中，可以进行特别的研究以便从不答复单位中推理出有关反应。

73. 相对于一次性或横断面调查中的不答复而言，在随后的波动式固定样本调查中的不答复提供了更多的数据，可以根据这些数据对潜在的不答复偏差效应进行研究和调整。固定样本调查是指对同一单位进行追踪，并在一段时间内重复从追踪单位中收取数据的调查。调查过程中可能有一部分跟踪的单位流失了，这导致了固定样本不答复或样本损耗不答复。但可以根据以前收集的固定样本波动数据对固定样本不答复进行调查，以进一步了解反应者和不答复者之间的差异，并以此作为以下调整的基础。弥补固定样本不答复的办法见Lepkowski（1988年）。

74. 由于有关不答复者的信息要略多于未覆盖者的信息，同时由于可以采用行为模式对不答复问题进行研究和补偿，所以不答复问题方面的研究要多于未覆盖方面的研究。如果对所有样本单位而不仅仅是反应样本进行了详细记录，就可以直接根据样本数据对反应者和不答复者进行比较。另外，部分不答复是由住户或个人的行为引起的，是一种自我选择的行为。调查设计者可以参照社会学、心理学和社会心理学方面的广泛文献，研究个人和群体是如何对参

与各项活动做出决定的。如果可以提供有关不答复者的某些数据，就可以对行为模式进行分析，以了解调查中不答复情况的决定性因素。

3. 计算不答复偏差

75. 计量不答复偏差需要计量不答复率以及调查变量的反应者和不答复者之间的差异。为样本数据的住户或人员计算不答复率又需要对所有抽样的可能结果做出界定，然后确定如何将这结果用于不答复率的计算。如，完全采访和部分采访（那些具有足够数据，能为关键研究概念提供相关信息的采访）通常归为一类。

76. 符合条件的未采访情况是指那些在人口群内被调查确定为受访者，但却没有从中收集到数据的人群。如，如果一项调查限于年龄在15岁或15岁以上的人群，那么合格的未采访人群是指那些年龄在15岁或15岁以上，但又未从中收集到数据的人群。通常情况下，未采访情况至少有三个来源：拒绝（Ref）——已被联系过，但不参与研究的人员或住户；未联系（NC）——在收集数据期间，没有被联系的人或住户；其他（Oth）——由于其他原因引起的未采访情况，如语言障碍或健康问题。最后还有一些不合格的情况（Inelig）——如年龄在15岁以下的人群，以及是否合格还不清楚的情况（Unk）。

77. 根据以上简化后的结果，反应率可以通过几种方式计算。以下是一种被普遍认可的反应率计算方法（其中Int表示一项调查中完全采访和部分采访的数量）：

$$\bar{R} = \frac{Int}{Int + Ref + NC + Oth + \epsilon \times Unk}$$

其中，在是否合格还不清楚的情况中，估计某个比率（ ϵ ）是合格的。通常情况下，估计的这一部分合格情况可以根据现有数据进行计算——可以使用所有已经被确定为合格的已知合格率（结果为Int、Ref、NC和Oth的情况）来计算，其公式如下：

$$\tilde{\epsilon} = \frac{Int + Ref + NC + Oth}{Int + Ref + NC + Oth + Inelig}$$

78. 至于重复对同一住户进行采访的调查，或重复对从一个住户样本中抽取的一组人进行采访的住户调查，不答复情况还需要考虑到其他因素，这些因素对反应率的计算有影响。这种纵向的跟踪调查在首次主要采访中会出现单位不答复情况，这和横断面调查一样。另外，在后几次主要的采访中可能无法从某些被追踪样本成员那里获取数据。反应率的计算必须考虑到由于首次及其后几次主要数据收集过程中的不答复情况所引起的缺失。固定样本调查中反应率的计算不在本出版物的研究范围之内，有关这方面的更多情况可访问美国公共舆论研究协会网站（<http://www.aapor.org>。路径：Survey Methods）。

79. 已反应人员和不答复人员平均值或其他统计指标之差很难计量。可以将某些变量的调查结果和外部机构的调查结果进行比较，以便了解调查估计值和外部机构的数据之间是否存在很大差异。但可能很难使用这种方法，因为调

查所使用的定义和方法可能和外部机构所使用的定义和方法不同，这样在解释所观察到的差异时，情况会变得更复杂。换句话说，调查估计值和外部机构估计值之间的差异可能不是由不答复引起的。

80. 衡量已反应人员和不答复人员的差异需要很高的费用。从原则上讲，在有足够资源的情况下，有时假设反应情况可以根据不答复情况获得。但很少有足够的资源可用来从每个不答复情况中获取数据。另一种方法就是，可以从不答复人员中抽取第二阶段样本或双重样本，将其余的所有调查资源都用作从这种子样品中收集数据。

81. 从统计上讲，有不少文献介绍了若干个设计特征方面不答复情况的两阶段抽样（如可以参见Cochran, 1977年, 第13.6部分）。在费用受到限制的情况下，如果从两阶段不答复样本中获取了完全反应情况，那么在第二阶段中可以确定一个最优抽样分数——这个抽样分数可以将两阶段估计平均值的抽样方差最小化。

4. 减少和弥补住户调查中的单位不答复情况

82. 用以降低不答复率的各种特别方法似乎是可行的，在很多情况下，可以通过这些特别方法来减少单位不答复情况。最近，已提出了各种以社会学和心理学原则为基础的综合理论[见Groves和Couper（1998年）]，可以通过更全面地了解住户调查中不答复情况的发生方式，从这些理论中得出减少不答复情况的方法。这些更加全面的理论框架不属于本章的讨论范围，本章将对试验研究中已经被证明有效的减少不答复情况的几种方法进行介绍。

83. 在大多数样本调查中，重复走访或“回访”是比较常用的方法。调查采访者对某个住户或合格人员的联系不仅仅只有一次，他们会对该住户或合格人员进行“回访”，以便充分获取采访情况。回访的次数、回访日程安排、用以说服不愿参与者或难以联络者的各种采访技巧都属于现场研究的主题。然而，这些调查特征却没有一个可行的标准。各个国家在反应率、公众对调查的接受度以及人口流动方面存在着差异，所以无法在回访方面确立一种统一的理论。由于存在着公众对不同主题的接受度问题，所以即使在一个国家的不同调查之间，也难以确定一个回访标准。但建议通过最好的采访者来说服那些拒绝参与者。

84. 在住户调查中，并没有经验证明某种方法（包括回访）可以带来较高的反应率。通常情况下，会结合使用几种方法。经验证明，在由采访者进行的住户调查中，以下方法都有助于提高反应率：通过提前打电话或写信的方式预先给予通知、使用个人化的信函、提供调查主办单位的信息、向潜在的反应人员说明数据是如何使用的。在发展中国家和转型期国家的调查中，采用激励性的措施是有争议的，很多国家都不建议采用这种方法。但在发达国家，这些方法正变得越来越普遍[有关调查方法的研究文献，见Kulka（1995年）]。

85. 重视采访者的采访技巧也可以提高反应率。对采访者进行培训，以便让采访者可以针对反应者的不同反应采取不同的技巧，这种办法也能提高一定的反应率。根据超过调查目标的调查成果和工作质量，对采访者给予奖励措施，这种办法也对提高调查的反应率有一定的积极作用。

86. 每一项住户调查都肯定会存在单位不答复问题。调查设计通常对单位不答复的样本规模进行调整，并计算出弥补权数，以对估计和分析进行调整。

87. 对不答复的样本规模进行调整需要估计，这种估计应该在收集预计的单位不答复率数据之前进行，它通常是一项调查的特别估计，主要根据过去相关人口、调查主题或其他因素的调查数据进行。在一次性横断面调查中，这种估计通常需要假设其他调查经验将会在未来调查中重现采用。在重复进行的横断面调查中，会定期或不定期地对同一人口群进行抽样，因此可以提供有关数据对预计的反应率进行估计。在固定样本调查中，需要在一定时间内对样本单位进行跟踪，所以在估计时，不仅需要对初次的单位不答复情况进行预计，还需要对之后的样本损耗不答复情况进行预计，在样本损耗不答复中，调查后期无法对那些在初期调查中曾配合进行调查的主体进行采访（主要由于被拒绝、不能找到该主体或其他原因）。

88. 由于费用或精确度方面的原因，样本规模的调整会使所需要的样本规模扩大，其目的是为了有足够的样本单位，从而获得满意的结果。如，要求完整住户采访的最后样本规模为1 000，预计的不答复率为20%。那么为了使完整住户采访的最后样本为1 000，调查就应该抽取 $1\ 000 / (1 - 0.2) = 1\ 250$ 个样本。如果预计反应率是准确的，那么最后样本规模将会和所要求的最后规模大致相同。应该分配一些单位让采访者进行采访，并要求他们尽可能多地从这些单位获得回应。不得采取替代法。

89. 处理单位不答复的另一个方法就是替代法。这种方法是让采访者自己决定是否接触某个单位，也就是说通过采访者的主观判断，而非客观的随机抽样来确定所要接触的抽样单位。采用替代方法处理不答复问题可以准确获得所要求的样本规模。但大量证据表明替代方法也有可能使样本与已知的人口分配不一致[如，见Stephan和McCarthy（1958年），他们介绍了一种与配额抽样法密切相关的非概率办法]。

90. 可以对最后调查数据进行统计调整，以便在某种程度上对可能发生的不答复偏差进行弥补。最常见的弥补方法需要确定不答复调整的权数。

91. 不答复调整权数的确定要求所有反应者和所有不答复者可以得到同样的信息。由于对不答复者的了解不多，所以在大多数住户调查中，可供这类调整使用的变量很有限。在大多数情况下，已知的有关不答复者的初步信息就是地理位置，也就是说住户的所在地。

92. 如，假设某个住户调查使用区域抽样法，根据这种方法，在抽样的第一阶段抽取普查查点区。收集数据的过程中，在某个查点区内，并非所有被调

查抽取的住户都会提供数据。通过简单的不答复权数调整方案，可以将增加的权数分配到某查点区内所有做出反应的住户中，以便对该地区不答复的住户进行弥补。如果某个查点区内有90%的住户做出了反应，那么该地区做出反应的住户权数将按 $1/0.9 = 1.11$ 的系数增加。如果另一个区域内，有80%的住户做出了反应，那么系数就是： $1/0.8 = 1.25$ ，该查点区内所有做出反应的住户权数都按这个系数增加。在最后的样本中，所有不答复住户都被剔除出去，实际上按照零的系数对他们进行了加权。

93. 在有些情况下，将行政机构的数据和调查反应数据进行比较也可用以确定加权的调整幅度。如，可能已经将行政机构的数据用以抽取样本。那么可以对样本反应者分配权数——这种权数的分配应该能够使某些重要变量经过加权后的反应者分配情况和行政机构数据中所报告的分配情况一致。

94. 不答复调整的幅度也可以根据一个模型进行确定。如果某项调查中被抽样住户的反应情况属于单纯的反应或不答复状况，而且还有关于反应和不答复住户的数据，那么反应状况可以根据现有变量进行回归。然后可以使用逻辑回归系数对每个住户反应的概率进行预测。如上所述，预测概率的倒数可用来计算权数，有时还被称为反应倾向权数。另外，直接按预测概率计算的权数往往变化很大，所以通常将预计的概率分类，然后使用中点、中值或平均预测概率的倒数，或将分类的加权反应率作为权数，给每类分配一个权数。

5. 项目不答复与插补

95. 最近活跃的一个研究领域就是项目不答复[如，见由Groves等人所做的评论（2002年）]。对于项目不答复，每个不答复情况都有大量的资料。这些资料有助于对项目不答复情况以及根据更复杂的统计模型计量、减少和弥补项目不答复的可能性进行更全面的了解。

96. 如，假设一项有关健康和保健服务情况的住户调查中，有90%的回答者回答了所有问题，有10%的回答者除了没有报告其前一个月的工资收入外，回答了所有其余的问题。那么90%回答者所提供的信息可用来确定一个统计模型，来了解健康、健康保健和工资收入之间的关系。而这些模型又可以被用以制定各种方法来减少工资收入方面的不答复情况，以此来弥补或预测缺失的工资收入值。

97. 替代缺失的项目值又称为插补，调查中使用这种方法已经有几十年了。住户以及其他调查中所使用的插补办法，见Kalton和Kasprzyk（1986年）以及Brick和Kalton（1996年）。几十年以来，插补法一直被各项调查用以弥补缺失的项目值。其基本思想就是用—个值来替代缺失的项目值，可以采用有关该受访者（如，住户或人）或调查中其他受访者的信息来预测这个值。

98. 如，可以通过回归模型来确定插补值。假设—项调查的变量为 Y ，可以为 Y 确定—个模型，采用调查的 p 个其他变量 X_1, \dots, X_p 对 Y 进行预测。该模型为：

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi} + \varepsilon_i$$

将变量Y以及“预测”变量 X_1, \dots, X_p 中没有缺失的项目填入该模型，之后，可以根据上述模型得出的估计参数来预测缺失情况的Y值。对于单位 i^{th} 来说，其变量Y的预测值为：

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \dots + \hat{\beta}_p X_{pi}$$

99. 插补的这种回归模型可以有几种形式。回归预测可以包括一项预测“残值”，该残值将被加入预测值。对于需要插补的情况，可以采用一种称为序列热卡插补的方法，通过回归插补方式，将从另一种情况下“借”来的残值有效地加入到有着类似值的数据文件中去。

100. 插补方法的最新发展中考虑了这样一个问题：插补会使那些根据插补值得出的估计数更具不确定性。可以通过诸如“刀切式”方差估计之类的方差估计法、插补模型、多重插补法来解决不确定性问题。在多重插补法中，多次重复插补，插补值的不确定性被纳入方差估计中。

101. 在调查中，有几种方法可以减少项目不答复情况，如对调查的采访者进行培训，使他们能够对调查问卷中那些无法解释或不完整的答复进行探究。调查设计者确实会为某些特定的项目增加一些跟踪的问题，这样在遇到诸如“我不知道”或“我不会回答这个问题”之类的反应时，就可以根据这些跟踪问题做进一步的探究。如，有关收入问题的不答复率往往要高于其他项目。所以有关收入方面的调查有时会在某些收入项目下增加一系列问题，这些问题可以“揭示”一个范围，在此基础上报告收入情况。如果回答者不愿意回答或不知道收入是多少，那么可以通过以下提问来进行探究：收入会超过XXX个单位吗？是介于YYY和XXX个单位之间吗？等等。通过这些问题就可以确定一个收入的范围，并据此对收入进行报告。

102. 进行住户调查的机构应该定期对各调查项目的不答复频率进行检查，以便确定项目不答复问题在调查中的重要性。除几个重要项目外，项目不答复率很少公布。所以用户通常需要决定项目不答复问题是否会对其分析造成影响。应该将重要项目的不答复率以及无反映率很高的项目包括在调查文件内。

鸣 谢

密歇根大学调查方法专业在职研究生Kenneth Coleman，在我考察拉丁美洲和南美的调查方法时为我提供了极其宝贵的帮助，在此，我向他表示感谢。

参考资料

- Brick, J. M. 和G. Kalton (1996年)。“处理调查研究中的丢失数据”。《医学研究的统计方法》，第5期，第215-238页。
- Cochran, W. G. (1977年)。《抽样方法》，第三版。纽约：约翰威利国际出版公司。
- Groves, R. M. (1989年)。《调查误差与抽样成本》。纽约：约翰·威利国际出版公司。
- _____和M. P. Couper (1998年)。《住户访谈调查中的不答复情况》。纽约：约翰·威利国际出版公司。
- Groves, R. M. 和其他人 (2002年)。《不答复调查》。纽约：约翰·威利国际出版公司。
- Kalton, G. 和D. Kasprzyk (1986年)。“如何对待丢失的调查数据”。《调查方法》，第12期，第1-16页。
- Kish, L. 和I. Hess (1950年)。“样本住所的未覆盖情况”。《美国统计学会刊》，第53期，第509-524页。
- Kulka, R. (1995年)。“运用激励机制调查‘很难普及’的答复者：实验法研究和目前研究做法的简单评论”。《关于统计方法新趋向的讨论会》。《统计政策工作文件，第23号》。哥伦比亚特区华盛顿：美国管理和预算办公室，第256-299页。
- Lessler, J. 和W. Kalsbeek (1992年)。《调查中的不答复误差》。纽约：约翰·威利国际出版公司。
- Lepkowski, James M. (1988年)。“如何对待小组调查中的间歇式不答复情况”。《小组调查设计和分析》，D. Kasprzyk、G. Duncan和M. P. Singh等。纽约：约翰·威利国际出版公司。
- Marks, E. S. (1978年)。“双重制度估计在人口普查评估中的作用”。《人口规模和增长双重制度估计的发展态势》。K. J. Krotki等。艾伯塔，埃德蒙顿，艾伯塔大学出版社。
- Seligson, M. A. 和J. Jutkowitz (1994年)。《危地马拉民主发展的价值与前景》。维吉尼亚，阿灵顿：发展协会。

第九章

住户调查测量误差：来源和测量

Daniel Kasprzyk

数学方法研究

美国哥伦比亚特区华盛顿

摘 要

本章描述了在样本调查中所发现的测量误差的主要来源以及用于量化测量误差的典型方法，讨论了测量误差的四个来源——调查问卷、数据收集模式、访调员和受访者，并通过误差的这四个来源描述了测量误差是如何在样本调查中产生的。此外，本章还描述了用于量化测量误差的方法，例如随机化实验、认知研究、重复测量研究以及记录检验研究，此外，还提供了例证以证明这些方法的应用。

关键词：测量误差、测量误差来源、测量误差量化方法。

A. 导言

1. 住户调查数据的收集采用了多种方法。数据收集过程中的一个固有假设就是被衡量的特性和概念可以精确定义，能够通过一套清晰的程序获得，并具有独立于调查的真值。测量误差就是受访者提供的特性值与这些特性的真值（但是未知的值）之间的差异。因此，测量误差与调查数据收集过程中的变量观测有关，所以，测量误差有时又被称为“观测误差”（Groves, 1989年）。

2. 本章以一份工作文件中有关测量误差的一章为基础，该工作文件由一个评量并报告美国联邦统计方法委员会调查数据质量的小组委员会提供（2001年）。因此，许多参考资料和例子涉及到在美国和其他发达国家的调查。不过，本章的论述适用于所有的调查，无论这些调查是在何处进行的。因此，本章也同样适用于在发展中国家和转型期国家开展的调查。

3. 在样本调查测量误差方面有一份重要文献 [见Beimer等（1991年）和Lyberg等（1997年）]，用于回顾重要的测量误差问题。在对调查的重复试验进行调查估计期间，测量误差可能产生偏倚和可变误差（方差）。在受访者对问题的答复与真值的差异中，测量偏误或答复偏误是作为一种系统的模式或方式而发生的。例如，受访者可能容易忘记报告其第二或第三职业的收入，导致

某些受访者所报告的收入低于其实际收入。如果对单位（住户、人员、访调员和调查问卷）多次提出的同一问题所获得的报告值不同，则这种差异也是误差的来源。简单答复方差反映了某一受访者对某一重复调查问题的随机变化（也就是说，如果向受访者数次提出相同的问题，受访者可能会提供不同的答案）。访调员对受访者答案的不同感受也是可变误差的一个来源，被称为访调员方差。访调员方差是与答复方差相关的一种形式，这种方差的出现是因为对相同的访调员访问的样本单位而言，答复误差是相互关联的。

4. 在文献中很容易找到用于研究测量误差的若干一般方法。一种方法是将调查结果与另一来源可能更为准确的数据进行比较。数据可以处于单独的样本单位层面，像在“记录检验研究”中的一样。举个简单的例子，如果受访者被问及年龄，其答复可以与其出生记录进行核实。不过，我们应当认识到，即使在这种简单的情况下，谁也不能肯定出生记录没有错误。尽管如此，在样本调查中研究测量误差的一个方法就是将调查答复与来自其他独立有效来源的数据进行对比。利用其他来源的数据对测量误差进行评估的另一种方法是在总体水平上进行分析，也就是说，将以调查为基础的估计与其他来源的人口估计进行对比。第二种方法包括对某些样本单位进行重复测量。典型的方法是调查重访程序，包括将第一次访谈获得的答复与紧随第一次访谈之后进行的第二次访谈获得的结果进行对比。研究测量误差的第三种方法必须从全部调查样本中选定随机的子样本并区别对待，例如使用备选调查问卷或问题或不同的数据收集模式。最后，测量误差也可以进行定性评估。方法包括焦点小组和受控的实验室环境，例如认知研究实验室。

5. 本章说明在样本调查及其测量过程中发现的测量误差的主要来源。设定程序对测量误差进行量化的成本昂贵并且往往难以实施。出于这个原因，同时由于这是一种良好做法，调查管理人员更强调通过良好的规划和良好的调查实践努力控制测量误差的来源。这种实践包括检验调查材料、调查问卷和程序，制定并检验定义清晰、操作可行的调查概念，特别着重强调难以接近的子群体的数据收集问题，招募合格的外地工作人员时执行高标准，制定并实施详细严格的培训计划，并为实地工作人员提供详细而清晰的书面指示。具体来说，控制非抽样误差和测量误差本身就需要进行广泛的讨论。例如，联合国（1982年）公布的报告中包括一份控制住户调查中非抽样误差的清单。本章并不涉及这个问题，而是将重点集中在描述样本调查中测量误差的主要来源以及量化测量误差的典型方法上。

6. 与Biemer等（1991年）一样，本章将讨论误差的四个来源：调查问卷、数据收集模式、访调员和受访者。本章使用大量篇幅描述在样本调查的过程中，这些误差来源是如何导致测量误差的。接着讨论了量化测量误差的一些方法。这些方法包括随机化实验、认知研究、重复测量研究以及记录检验研究。量化测量误差总是要求在调查前、调查中和调查后采取额外的措施。启动量化的测量误差来源研究经常遇到的困难是进行研究的时间和费用要求。不过，对测量误差的研究对量化目前调查的误差水平，以及说明在以后的调查

中有待做哪些改进这两个方面均具有极为重要的价值。这些研究对重复调查计划特别有用。

B. 测量误差的来源

7. Biemer等（1991年）确定了测量误差的四个主要来源：

- **调查问卷：**调查问卷的设计、视觉布局、所覆盖的主题以及问题的措辞等方面的影响。
- **数据收集模式：**包括调查问卷如何提供给受访者（例如：邮寄、人员上门或日记）的影响。受访者当着访调员的面亲口回答或使用日记回答时，对问题的答复可能不同。
- **访调员：**访调员会对问题的答复产生影响。在处理不恰当的答复时，访调员可能会通过有意不阅读答复中的有些内容或者进行不适当的追问，或者通过添加可能迷惑或误导受访者的其他信息，从而导致调查结果出现误差。
- **受访者：**因为受访者的经历、知识和态度存在差异，所以受访者对调查问卷的问题会做出不同的解释，从而对调查结果产生影响。

8. 在进行样本调查时，这四个来源是至关重要的。调查问卷是向受访者正式提出问题以获取信息的方法。数据收集模式是指调查问卷是以何种方式交付或提供（自行提供还是人员上门）给受访者的。访调员在人员上门的模式中是调查问卷的交付者。受访者是问卷的接受者。这四者中每一个都会导致测量过程中出现误差。大部分调查即使考虑到这些误差来源，也都是孤立看待。但是，这些误差来源之间相互作用，例如，访调员和受访者的特点可能相互作用，从而导致误差的出现，却无法明白这种误差是由哪一个单独来源造成的。测量误差如何在这四个误差来源的范围内产生，这一点将在下文讨论。

1. 调查问卷的影响

9. 调查问卷是数据收集者从调查受访者处获得信息的工具。近20年来，调查问卷设计的基本原理曾一度被视为艺术化程度大于科学化程度，已经成为大量文献关注的主题（Sirken等, 1999年；Schwarz, 1997年；Sudman、Bradburn和Schwarz, 1996年；Bradburn和Sudman, 1991年）。调查问卷或调查问卷的特点，即问题表述的方式或者调查问卷的格式都可能会影响到个体在调查中的答复。在本节中，我们描述在数据收集过程中，调查问卷是如何导致误差出现的。

规范问题

10. 在进行调查规划时，往往因为研究目标不明确、概念定义不清以及调查问卷中收集的信息不一致而产生问题。调查问卷中的问题可能无法获得满足研究目标所需要的信息。因为调查问卷调查说明表述不清、定义不清或所需

的概念难以衡量而产生数据规范问题。例如，一项调查可以询问“怀孕期间接受的孕妇保健”，但是没有界定哪一次怀孕或该问题涉及的时间段。即使是基本的问题也会产生含糊的答复，例如：你有几份工作？如果没有界定工作的性质——临时或固定工作和/或全职或兼职工作，答案势必含糊。复合分析概念，例如一个人的总收入，如果没有对受访者确定和界定收入的各个部分，则可能得不到完整的报告。

问题的表述

11. 如果想让受访者按照设计者的意图理解调查问卷中的问题，则这些问题的表述必须准确而清晰。因为调查问卷是数据收集者和受访者之间的一种交流形式，因此有许多的潜在误差来源。首先，调查问卷的设计者可能没有清晰地阐明他/她试图衡量的概念。其次，即使概念得到了清晰的阐述，这个概念也可能没有恰当地在问题或一系列问题中得到表述；即使概念清楚，并且得到了忠实的表述，受访者的理解也可能与调查问卷设计者的意图不一致。调查问卷设计者和受访者之间的语言和文化差异或经历和背景差异可能会导致对问题的误解。这些差异在有数个不同民族的发展中国家和转型期国家中尤为重要。Vaessen等（1987年）讨论了在使用多种语言的国家进行调查时出现的语言问题。

12. 对于在样本调查中提出的问题，至少有两个层次的理解。第一个层次是对问题的字面意思进行简单的理解。受访者熟悉问题中所有单词吗？受访者能够回忆起与他/她对这些单词的理解相匹配的信息并提供有意义的答复吗？但是，要答复一个问题，受访者必须推测调查问卷的意图；也就是说，要答复问题，受访者必须确定问题的实际意义（Schwarz、Groves 和Schuman, 1995年）。正是这第二个因素使问题的表述成为比仅仅编写需要低级阅读水平的问题项更困难、更复杂的工作。为了制作设计周密的工具，需要受访者的投入，即需要他们对问题的解释和理解。认知研究方法为获得这种投入提供了有用的方法（见第C.2节）。

问题的长度

13. 常识和良好的书写习惯告诉我们，使问题简短能够获得清楚的解释结果。但是，研究发现，较长的问题可能比较短的问题更能从受访者那里获得更为准确的细节，至少在涉及症状和看病（Marquis和Cannell, 1971年）以及酒精和麻醉品使用（Bradburn、Sudman 和Associates, 1979年）方面的报告行为如此。较长的问题可以提供更多信息或提示来帮助受访者回忆，并且可以使受访者有更多时间对询问的信息进行思考。

调查问卷的长度

14. 研究人员和分析人员总是想问尽可能多的问题，但是调查方法学家却认为，如果调查问卷太长，可能会导致误差的出现。受访者会变得不专心或疲

劳，这取决于他/她的特点（例如，年龄或健康状况）、主题的特性、与访调员的关系、调查问卷的设计和访谈的模式等。

问题的次序

15. 研究人员发现，问题的次序会影响答复（Schuman和Presser, 1981年），特别是在对态度和观点的调查中。已经观察到同化现象——后面的答复被引导到与前面的问题相同的方向，或者相反，后面的答复被引导到与前面的问题相反的方向。受访者可能还会利用从前面与术语的含义有关的项目中获得的信息帮助自己答复后面的问题。

答复的类别

16. 问题的答复类别可能会通过向受访者提示调查问卷设计者的哪些想法很重要从而影响对问题的答复。受访者推测，调查问卷的设计者认为某一项所包含的类别是最重要的。如果受访者认为答复的类别对受访者不合适，则会导致问题意图的混乱。各类别的顺序也会对答复产生影响。受访者可能会在访谈中自鸣得意，并且会在答复范围内有计划地回答同一处的问题，答复的是靠前的选项而不是靠后的选项，或者选择提供的靠后答复。

17. 答复类别的次序产生的影响也可能会受到访谈模式的影响。如果问题项是自行完成的，则表上位置靠前的答复类别更容易被想起并得到认可（首因效应），因为受访者有更多时间处理这些问题。如果问题项是访调员提供的，则靠后的类别更容易被回想起来（近因效应）。

开放式问题和闭合式问题

18. 为受访者提供了一系列指定答案选项的问题形式（闭合式）所得到的不同答复比没有为受访者提供这种选项的问题形式（开放式）要多（Bishop等, 1988年）。一个给定的答案在开放式问题中得到答复的可能性，与它被列为闭合式选项相比要小（Bradburn, 1983年）。闭合的形式可能会帮助受访者想起一些他们可能忘记的东西。答复的选项还可能会向受访者说明适当的答复层次或类型 [例如，见Schwarz、Groves和Schuman（1995年）以及Schwarz和Hippler（1991年）]。

调查问卷的格式

19. 自填式调查问卷的实际“外表”，也就是说，调查问卷的格式和版面，可能会有助于或妨碍受访者做出准确的答复。格式混乱的调查问卷设计可能会让受访者感到迷惑，并会导致对跳转模式的误解，或者导致对问题和提示的曲解。Jenkins和Dillman（1997年）为自填式美国人口调查问卷提供了若干设计原则。如果没有考虑到另一国的文化和语言因素，在将这些原则转用于该国时应审慎行事。

2. 数据收集模式的影响

20. 确定数据收集的最佳模式需要进行包括各种调查方法问题在内的决策。财力往往在这种决策中扮演着重要的角色；然而，调查问卷的内容、目标人口、目标答复率、数据收集期的长度以及预期的测量误差都是在决定最恰当的数据收集模式过程中需要重点考虑的问题。虽然技术的发展使电话在数据收集中的使用日益增多，数据收集的其他方式在进行调查的过程中也提供了许多不同的选择。Lyberg和Kasprzyk（1991年）对不同的数据收集方法以及伴随这些方法的测量误差源进行了概述。下文对这一概述进行了汇总。

面对面访谈

21. 面对面访谈是在发展中国家和转型期国家收集数据的主要方法。在大多数情况下，访调员向受访者提供一份有客观标准的调查问卷，并将受访者的答复填入纸质调查问卷。这种利用纸和笔的个人访谈（PAPI）方法历史悠久。新出现的便携式电脑使得面对面访谈可以通过计算机辅助个人访谈（CAPI）进行。访调员访问受访者的住所，利用便携式电脑而不是纸质调查问卷进行访谈。参见Couper等（1998年）对计算机辅助个人访谈相关事务的讨论。计算机辅助个人访谈方法最大的优势体现在质量控制和减少答复误差方面。访调员将答复输入电脑文件。访谈软件确保正确遵循了调查问卷的跳转模式，而且在访谈时，答复即被输入并进行了合理编辑；因此，节省了调查数据处理阶段的时间和资源。

22. 在面对面访谈中可能进行复杂的访谈，视觉辅助手段可能用来帮助受访者答复问题，训练有素的老练访调员能够与被访者建立密切的关系，并能得到更完整、更准确的答复。但是，访调员可能会影响受访者对问题的答复，并由此在调查评估中产生偏误，或者产生C.3节中所讨论的访调员方差影响。访调员能够通过个性与行为特性的组合来影响答复。特别值得关注的是不合社会要求的特性或行为。受访者可能很不愿意向访调员报告这些特性或行为。De-Maio（1984年）指出，社会愿望因素似乎包含两方面的因素——某事物好以及某事物不好的观念，以及受访者想要表现得“好”并通过答复问题表现出来。

23. 与住户调查面对面访谈有关的测量误差的另一个可能来源是访谈过程中另一个住户成员有可能会出现。住户成员可能会影响受访者的答复，尤其是当访谈的问题较为敏感时更是如此。例如，当另一名住户成员在场时，很难让受访者真实地答复有关非法药物使用的问题。当另一名住户成员在场时，即使表面上看来无伤大雅的问题也被视为敏感问题（例如，配偶在场的情况下，询问有关婚姻或生育史的问题）。

自行完成的调查

24. 自行完成的调查问卷中的测量误差来源与面对面访谈中出现的测量误差来源不同。自助式调查显然不存在访调员的影响，受到“社会愿望”影响的风险也较小。这种调查还为在不让受访者感到为难的情况下就敏感或有危险的

话题提出问题提供了一种方法。这种调查方式的另一个好处是，如果有必要，可以同时向一个住户中多个受访者展开调查（Dillman, 1983年）。可另一方面，如果目标人口中个体的受教育程度很低或没有受过教育，或者个体在阅读方面存在困难，则自行完成的调查可能会产生系统性偏误。这种偏误可以在对“开放式”问题的答复中观察到，在自行完成的调查中，对这种问题的答复不如访调员调查时所得到的答复全面、详细。这种数据收集方法在识字率低的国家不太理想，不过，即使目标人口的受教育程度相当高，受访者依然可能读错并曲解问题和提示。通常，在自行完成的调查中，问题项的答复率较低，但是，一旦对问题做出了答复，数据的质量一般都较高。自行完成的调查，也许比其他的数据收集方式更受益于良好的调查问卷设计和格式以及表述清晰的调查问卷项目。自行完成的调查中有一种特别类型的调查，那就是自行完成的邮件调查，在这种调查中，调查问卷的发放和收回全部通过邮寄完成，要求受访者自行填写调查问卷（Dillman, 1978年；1991年；2000年）。

日记调查

25. 日记调查是一种自助式调查形式，用于需要在一段时间内对某种行为进行详细报告的主题（例如，支出、时间利用和看电视）。为了尽量减少或避免回忆错误，鼓励受访者使用日记并在事情发生后立即记录对有关事件或主题的答复。日记模式的成功之处在于受访者在记录信息以及填写典型的“繁琐”表格时能够扮演积极的角色。这一模式还要求目标人口能够阅读并理解日记问题，这一条件在识字率低的国家不用。资料收集程序通常要求访调员联系受访者以提供日记，得到受访者的合作并解释数据记录的程序。经过预先约定的一段时间后，访调员收回日记，如果日记尚未完成，访调员帮助受访者完成日记。

26. Lyberg和Kasprzyk（1991年）为这一模式确定了许多测量误差来源。例如，极少注意或根本不注意记录事件的受访者在记忆犹新时，可能没有记录这些事件。日记本身的版式和格式以及问题项的复杂性，可能让受访者在实践中感到非常困难。此外，使用日记可能会导致受访者改变他们的行为方式；例如，必须在支出日记中列明购买行为可能会导致受访者改变他/她的购买行为。对支出调查中测量误差的讨论，尤其是调查中有关日记问题的讨论见Neter（1970年）和Kantorowitz（1992年）。对面对面访谈以及日记调查所获数据的对比见Silberstein和Scott（1991年）。

直接观察

27. 直接观察作为一种数据收集方法，要求访调员利用自己的感觉（视觉、听觉、触觉、试验）或物理测量设备进行数据收集。多门学科都使用这种方法，例如，在农业调查中估算农作物产量时（“目测”）以及在住户调查中评估受访者的住宅质量时。观测者导致测量误差的方式与访调员导致测量误差

的方式相似；例如，观测者可能会误解概念，并误解需要记录的信息，还可能因为自满或厌倦而在事后改变记录信息的方式。

3. 访调员的影响

28. 在许多样本调查中，访调员都扮演着极为重要的角色。作为数据收集过程的一个基本部分，他/她的表现能够影响调查数据的质量。不过，访调员是数据收集过程中调查研究者/调查管理者能够努力控制的因素之一；因此，使用策略——通过挑选和雇用、培训以及工作表现监督——以尽可能缩小与访调员角色相伴的误差（Fowler, 1991年）。因为存在着个体差异，每一名访调员都以不同的方式处理调查事务；例如，个别访调员可能不会严格地按照书面材料提出问题，不会正确地遵循跳转模式或者以恰当的方式寻求答案。他们可能不会严格地遵循指示，这既可能出于特定的目的，也可能是因为指示不太清楚。访调员可能会无意识地改变他们的音调，或者在个人的特殊习惯中表现出其他变化。

29. 每一名访调员都可能犯多报或少报的错误。如果所有访调员的多报和少报能够近似抵消，结果会出现小的总体访调员偏误。然而，单个访调员的误差也可能很大，而且处于相同的方向，从而导致这些访调员的偏误较大。单个访调员的偏误变化被称为访调员方差，这种访调员方差会对调查评估的精确度产生重要影响。

相互关联的访调员方差

30. 20世纪60年代初，Kish（1962年）使用访调员内部关联系数，他用 p 来表示这个系数，找到了一种评价访调员方差对调查评估影响的方法。 p 是访调员方差分量与调查变量总方差的比值，可以通过对方差的简单分析来估算这个值。

31. 在进展顺利的面对面调查中，对大部分变量来说， p 的典型值约为0.02。尽管很小，但它对评估精确度的影响却可能很大。样本平均方差乘以 $1 + p(n - 1)$ ，式中， n 是访调员工作量的平均样本量。如果 p 的值为0.02，10名访调员的工作量将使方差增加18%，如果工作量为25，则方差会增大48%。因此，即使 p 的值很小，也会显著降低调查统计的精确度。从实践和经济利益的角度出发，访调员的工作量通常都很大。因此，一名出现系统性偏误的访调员会影响到从大量受访者那里获得结果，而且对方差的影响可能会很大。

访调员特点

32. 研究文献无助于确定优秀访调员的特点。在大不列颠及北爱尔兰联合王国，Collins（1980年）发现，没有理由建议将访调员的重点招募对象集中于妇女而不是男子，或者集中于中产阶级，以及中年人而不是青年或老人。Weiss（1968年）对纽约市领受福利救济的母亲样本进行了研究，对几个问题的准确性进行了验证，发现访调员和受访者在年龄、教育、社会经济地位方

面的相似性并没有导致更好的报告结果。Sudman等（1977年）对访调员获得敏感信息的预期困难进行了研究，发现预期和实际的访谈困难之间的关系产生的影响不大。Groves（1989年）对许多研究进行了审查，得出的结论是，如果测量与人口统计学特点有关，则通常会产生人口统计影响，但是其他方面则并非如此；例如，如果问题与种族有关，则可能会在访调员种族的基础上产生某种影响。

控制访调员误差的方法

33. 在一定程度上，调查管理者可以通过访调员培训、管理或监督以及工作量处理来控制访调员误差。一个长度足以覆盖访谈技巧和技术以及提供有关特别调查的信息培训程序有助于对访谈过程标准化（Fowler, 1991年）进行衡量。许多人相信，标准化的访谈程序能够减少访调员的影响。

34. 监督和绩效监测的目标是通过观察和绩效统计对绩效进行监测，并确定存在问题的调查问题，并且构成了访调员质量控制系统中的另一个组成部分。重访程序和实地观察用来评估单个访调员的表现。实地观察使用大量编码表或详细的观察者指南，以便监督者检查访调员是否正确地遵循了程序。例如，观察内容可以包括访调员的外表和行为，对他/她自己以及调查的介绍，提出问题和记录答案的方法，样品卡的使用以及客观的调查，访调员手册的正确使用。再比如，可以使用磁带（录像或录音），并对访调员的行为进行编码和分析（Lyberg和Kasprzyk, 1991年）。

35. 缩小访调员方差的另一个方法是降低平均工作量；不过，这样做的前提是可以使用素质相同的额外访调员。Groves和Magilavy（1986年）将访调员的最佳工作量作为访调员雇用与培训成本、访谈成本、访调员内部相关性的函数来进行讨论。由于在相同的调查统计中，访调员内部相关性存在变化，因此很难确定是什么构成了最佳工作量。

36. 可以通过避免调查问卷设计问题，提供清晰而明确的说明和定义，培训访调员遵循说明，尽可能减少在获得答复时对访调员可变技巧的依赖来减少访调员的影响。

4. 受访者的影响

37. 受访者不提供准确的答复可能会促进测量误差的产生。Groves（1989年）注意到传统的访谈过程模式（Kahn和Cannell, 1957年）以及调查答复的认知科学方面。Hastie和Carlston（1980年）确定了调查受访者形成并提供答案的五个连续阶段：

- 信息编码，涉及形成记忆或记住知识的过程。
- 理解调查问题，涉及有关调查问卷单词和短语的知识以及受访者对调查目的的印象、问题的背景和形式、以及访调员提问时的行为。

- 从记忆中检索信息，涉及受访者努力在她/他的记忆中搜索有关信息。
- 对正确答案的判断，涉及受访者根据追忆的信息对问题的备选答复进行选择。
- 答复的传递，涉及对受访者检索到相关信息后的正确答复的影响以及受访者清晰表述答复的能力。

38. 调查过程的许多方面都会影响受访者在这五个阶段所做答复的质量。以下举例说明影响答复效果的因素。

受访者规则

39. 受访者规则确定了用于确定回答调查问卷人员的合格标准，在答复过程中扮演着重要角色。如果一项调查收集有关住户的信息，住户中不同的合格受访者对问题的答复也可能不同。收集样本住户中个人信息的调查可以使用自行报告或代理人报告的方式进行。自行报告与代理人报告之间的差别在于主题（例如，自行报告更适用于意见调查）。联合国（1982年）描述了在土耳其人口调查中有关代理人答复影响的小规模试验结果。Blair、Menon和Bickart（1991年）对自行报告与代理人报告的研究进行了文献论述。

问 题

40. 问题的表述和复杂性以及调查问卷的设计可能都会影响到受访者如何以及是否理解了问题（详情见B.1节）。受访者是否愿意积极提供正确的答案受到所提问题的类型的影响，还受到确定答案这一任务难度的影响以及受访者对答复的社会愿望的看法。

访 调 员

41. 访调员的视觉信息（例如，年龄、性别、衣着、面部表情）和声音提示（例如，语调、语速、音调）都可能会影响受访者对问题的理解。

回 忆 期

42. 时间通常会削弱回想事实或事件的能力。记忆力减退，会导致受访者很难回忆起距离调查时间很久以前的活动。例如，在世界生育率调查中，有些国家对近期出生情况的日期记录比以前出生情况的日期记录似乎更为准确（Singh, 1987年）。调查设计者可以按照抽样误差和可能的偏误寻找使总均方误差最小的回忆期；例如，Huang（1993年）发现，通过扩大样本规模提高精确度并将基准期从四个月增加到六个月并不能补偿因为回想损失而增大的偏误。Eisenhower、Mathiowetz和Morganstein（1991年）讨论了利用记忆辅助手段（例如，日历、地图、日记）来减少回想偏误的问题。Mathiowetz（2000年）报告了一项元分析的结果，这一结果检验了回溯报告的质量是回忆期长度的函数这一假设。

套 叠

43. 套叠发生在受访者将一个事件作为发生在基准期内的事件进行报告，而该事件实际上发生在基准期之外的情况下。利用限制技术（例如，为建立一个基准日而专门进行初次访谈，或者利用一个重要日期或活动作为基准期的起始日期）能够减少套叠的影响（Neter和Waksberg，1964年）。

固定样本调查/纵向调查

44. 与受访者有关的其他因素促使固定样本调查或纵向调查发生调查误差。首先，当受访者对出现在两个不同地方的相同或相似问题做出不同答复时，可能出现虚假的变化度量，做出这种答复的原因是在回答同一问题时出现的随机变异，而不是真正的变化。Kalton、McMillen和Kasprzyk（1986年）提供了在连续进行某项纵向调查时出现测量误差的例证。他们将年龄、种族、性别以及行业和职业作为变量，并在“美国收入和计划参与调查”中发现了测量误差。“美国收入和计划参与调查质量概述”对这一问题以及在调查中确定的其他测量误差进行了讨论（美国普查局，1998年）。通过将以前访谈的答复用于当前的访谈，相关访谈技术能够降低虚假变化的发生率。Hill（1994年）发现，相关访谈的结果是职业和就业行业方面的变化量度完全改进，但是也会错过对真实变化的报告，因此在使用时有必要对其进行选择。Mathiowetz和McGonagle（2000年）审查了当前在计算机辅助访谈环境中的做法以及相关访谈对数据质量的影响的经验证据。

45. 在固定样本调查中，固定样本调整或“抽样时间”偏误是误差的另一个潜在来源。调整是指在受访者已经进行过一次或多次事先访谈后答复发生的变化。Woltman和Bushery（1977年）调查了“美国国家犯罪受害调查”中的抽样时间偏误，将个人的受害报告与在同一个月内已经进行过的各等级固定样本经验（即以前访谈的数量）进行比较。他们发现，在事先访谈数量增加的情况下，受害报告比率通常会不断下降。Kalton、Kasprzyk和McMillen（1989年）也对这一误差来源进行了讨论。

C. 量化测量误差的方法

46. 有几种常规方法可以量化测量误差。为了研究测量偏误，不同的处理方法（例如选择性调查问卷以及问题或不同的数据收集模式）都能有助于在全部调查样本中随机选择子样本。测量误差能够在定性的情况下进行研究，例如焦点小组或认知研究实验室。另一种方法涉及对样本单位的重复测量，例如在调查重访程序中就使用这种方法。最后，还有“记录检验研究”，这一方法将调查答复与从其他来源得到的更为准确的数据进行对比，从而估计测量误差。这些方法都将在下文中进行讨论。

1. 随机化实验

47. 随机化实验是评估测量误差时经常使用的方法。调查研究者在提到这一方法时使用了许多名称，例如，互渗式样本、分样本实验、分固定样本实

验、随机半样实验以及分抽签实验。针对同一设计的随机子样本对正在测量的有关特定误差采取不同的处理措施。为了研究可变误差，将许多被认为是误差来源的不同实体包括在内并进行比较（例如，在访调员方差估计中涉及许多不同的访调员）。为了研究偏误，通常对两到三种处理措施进行比较（例如，两种不同的数据收集模式），其中的一种方法是首选方法。调查进行前的实地试验通常包括随机化实验，以评估其他方法、程序和调查问卷。

48. 例如，随机化实验可以用来检测调查问卷长度的影响。样本单位被随机分配给两组中的一组，这两组人员中，一组使用“简短”问题，另一组使用“长”问题。假设能够获得一个独立的数据来源，每一组的答复就可以与来自假定准确可靠的数据源的估计值进行比较。同样，问题顺序的影响也可以通过改变给随机样本提供的调查问卷中的问题的顺序进行评估。这一方法被用于在多米尼加共和国进行的一项调查中，该调查是在世界范围内开展的人口与健康调查方案的一部分；三分之二的样本使用核心调查问卷，三分之一的样本使用的是实验调查问卷。这样做的目标是确定因使用两套问题而产生的答复差异（Westoff、Goldman和Moreno，1990年）。

2. 认知研究方法

49. 过去20年来，认知研究方法在减少测量误差方面的应用得到了迅速增长。这些方法最初被用来洞察受访者的思维过程，现在被越来越多地用于观察传统的实地试验（Schwarz和Sudman，1996年；Sudman、Bradburn和Schwarz，1996年）。受访者向调查问卷设计者提供有关他们如何理解调查问卷中问题项的信息。这一方法是劳动密集型的，而且在每名受访者身上的花费都很高；因此，认知试验都在小规模样本中进行。认知访谈的一个缺点是只能在小规模的非随机样本中进行。调查问卷的设计者应当意识到，试验结果揭示了潜在问题，可是并不一定代表潜在的调查受访者。

50. 应用最为广泛的方法依靠的是口头协议（Willis、Royston和Bercini，1991年）。受访者被要求填写调查问卷草稿并描述他们如何理解调查问卷中的各个项目。访调员将对她或她想从受访者那里获得有关特定的字词、定义、跳转模式或调查问卷中的其他要素的特定反馈进行调查。要求受访者确定他们所有不清楚的事情。还可能要求受访者在填写调查问卷时（“同时边想边说”）或在事后汇报讨论（“回顾式边想边说”）中这样做。设计者可能会增加问题以调查后继访谈调查问卷中不同项目或要素的明确性。这项技术的优点在于不会出现因访调员而引起的偏误。其缺点在于，如果受访者不自在或者不善于描述自己的思想，这种方法的效果就不好（Willis，1994年）。

51. 相关技术涉及在受访者完成某一题项之后，访调员立即就这一问题的某些特点向受访者提出问题（Nolin和Chandler，1996年）。这种方法对受访者是否自在及其思维表达能力的依赖性较小，可是把研究限制在调查设计者认为他能够询问的项目上。这一方法还可能产生访调员偏误，因为调

查依赖于访调员。由于追问方法与访谈不同，因此有人认为这种方法不自然（Willis, 1994年）。

52. 其他方法能够使受访者利用在焦点小组进行的提问填写调查问卷。焦点小组的优势是组员之间能够互动，这样，就可以探讨一对一访谈所没有涉及到的领域。

53. 召集专家小组对调查问卷提出批评，是发现问卷是否存在问题的有效方式（Czaja和Blair, 1996年）。调查设计专业人员和/或调查项目专业人员在与调查问卷设计者举行会议之前的几天获得调查问卷。在小组会上，大家分别对调查问卷上的问题一一进行评议并提出意见。

54. 认知研究方法现在被广泛用于发达国家调查中的调查问卷设计以及减少测量误差方面。Sudman、Bradburn和Schwarz（1996年）总结了他们在调查方法学方面的主要成果。Tucker（1997年）讨论了在调查研究中应用认知心理学的方法问题。

3. 重访研究

55. 重访——在访谈调查中对同一单位进行重复测量——是指询问以前访谈问题（或其中的部分问题）的访谈。重访通常利用调查样本单位中的小规模子样本（通常约为5%）进行。进行重访出于以下一个或多个目的：

- 查明伪造数据的访调员；
- 确定误解程序并需要接受补救性培训的访调员；
- 估计简单的答复方差；
- 估计答复偏误。

56. 前两个目的提供信息，说明因访调员的影响而造成的测量误差。后两个目的提供信息，说明因所有四个来源（即访调员、调查问卷、受访者和数据收集模式）共同影响而产生的测量误差。

57. 下文讨论了四种重访类型中每一种的特别设计要求 [见Forsman和Schreiner（1991年）]。此外，还介绍了一些分析重访数据的方法以及结果的局限性。

访调员造假重访

58. 访调员可能采用若干种方法伪造调查结果：例如，访调员可以为部分或全部问题拼凑答案，或者访调员可以故意不遵守调查程序。为了发现造假情况，需要抽取重访样本，并且通常由监督人员进行重访。造假率是指通过造假重访发现的在访谈中造假的访调员的比例，这一比例可以计算出来。Schreiner、Pennie和Newbrough（1988年）报告说，美国现有人口调查中这一比率为0.4%，美国国家犯罪受害调查中这一比率为0.4%，纽约市住房及闲置率调查中这一比率为6.5%，以上这些调查都是由美国普查局完成的。

访调员评价重访

59. 重访程序中把确定未能以可接受的水平完成工作的访调员的重访定义为访调员评价重访。这种重访的目的是确定误解调查程序的访调员，并确定需要进行进一步培训的访调员。这种类型重访的大部分设计特点都与造假重访中的设计特点相同。以统计质量控制理论为基础的容差表可以用来确定调整后的访谈差异数量是否超过了可以容许的限度。美国普查局的重访程序中采用的可接受质量容差标准为6%至10% (Forsman和Schreiner, 1991年)。

简单答复方差重访

60. 简单答复方差重访是对原始访谈程序的独立复制。在重访中尽可能重复原始访谈中所有的准则、程序和过程。重访样本是原始样本设计中的代表性子样本。重访中使用原始访谈中的访调员、数据收集模式、受访者规则和调查问卷。在实践中并不总是遵循这些假设；例如，如果原始的调查问卷很长，则使用精简的原始调查问卷。不对原始访谈和重访之间的差异进行调和。

61. 从简单答复方差重访中得到的统计估计就是总差异率 (GDR)，就是原始访谈与重访答复之间的平均平方差。总差异率除以2就是简单答复方差 (SRV) 的无偏估计。对特性来说会有两个可能的结果，总差异率与在原始访谈和重访中存在不同答复情况的百分比相等。Brick、Rizzo和Wernimont (1997年) 为解释采用总差异率测量的答复方差提供了一般规则。

62. 另一个统计量是不一致性指数 (IOI)，计量总体方差占简单答复方差的比例。因此，

$$IOI = \frac{GDR}{s_1^2 + s_2^2}$$

式中， s_1^2 是原始访谈的样本方差， s_2^2 是重访的样本方差。

63. 对不一致性指数的值一般解释如下：

- 小于20的不一致性指数是低相对答复方差；
- 20至50之间的不一致性指数是中等相对答复方差；
- 50以上的不一致性指数是高相对答复方差。

64. 答复方差的测量标准——总差异率和 inconsistency 指数——向数据用户提供有关调查问题的可靠性和答复一致性的信息。在联合国 (1982年) 住户调查非抽样误差中，可以找到将总差异率和 inconsistency 指数用于秘鲁生育力调查的部分变量的例证。作为人口与健康调查方案第二阶段的一部分，在巴基斯坦对在主调查中访谈的妇女子样本使用了重访程序以评估全国范围内答复的一致性 (Curtis和Arnold, 1994年)。Westoff、Goldman和Moreno (1990年) 对一项重访研究进行了描述，该研究是多米尼加共和国人口与健康调查方案的一部分，之所以引人注意是因为它在若干方面需要进行折衷，例如将访谈限制在少

数地区和目标人口的子集中。印度进行重访调查的目标是搞清答复方差，美国普查局（1985年）对这项调查进行了说明，并审查了普查评估程序。

65. 一项定期调查利用重访程序不断改进调查问卷，Feindt、Schreiner和Bushery（1997年）介绍了该项调查为此做出的努力。如果重访调查发现问题的差异率较高，则可以开始利用认知研究方法对调查问卷进行改进。这些方法可以确定问题产生的原因，并提出可能的解决方案。在下一轮调查访谈中，可以对修改过的问题进行重访，以确定是否对可靠性进行了改进。然后对其余难题重复这一过程。

答复偏误重访

66. 测量答复偏误的重访旨在获得原始样本设计中代表性子样本的真实或正确答复。为了获得真实的答案，使用了最有经验的访调员和监督员。此外，重访中的受访者是具有见识的受访者或者是为他们自己答复问题的住户成员。在重访中使用原始的访谈问题，并且就两次答复的差别与受访者进行核对以确定“真实性”。另一种方法使用一系列的探究性问题取代原始问题，以获得准确的答复，然后与受访者核对两次答复间的差异。关于为获得答复有偏估计数而进行的重访调查，其讨论见美国普查局（1985年）发布的描述普查评估程序的报告。

67. 为确定真实答案进行的核对有一定的限制条件。受访者可能故意报告虚假信息，并在原始访谈和重访中一直报告这一信息，以致核对的重访不会得出“真实的”估计。在对美国现有人口调查重访数据进行的一项研究中，Biemer和Forsman（1992年）发现，原始访谈错误有多达50%在经核对的重访中没有被发现。

68. 对答复偏误的估计通过计算净差异率（NDR）来估算，假定原始访谈答复和经核对的重访答复的平均差额就是“真实的”答案。在这种情况下：

$$NDR = \frac{1}{n} \sum_{i=1}^n (y_{0i} - y_{1i})$$

式中， n 是重访样本规模； y_0 是原始访谈的答复； y_1 是经核对后的重访答复，假定为真实的答复。

69. 净差异率提供了有关调查问题准确度的信息，并确定导致出现偏误的问题。在分析数据和解释结果时，这一偏误的存在需要加以考虑。Brick等（1996年）利用集中性重访来更好地了解受访者的视角及其回答问题的理由，最终对答复偏误做出估计。尽管工作中使用的样本规模不大，但是作者得出的结论是这一方法在观察和测量偏误方面具有潜力。偏误校正估计得到了发展，这说明了在有办法测量偏误时，对估计的潜在影响。

4. 记录检验研究

70. 记录检验研究将个别样本案例的调查答复与从外部来源（该来源通常被假定为包含调查变量的真值）获得的值进行比较。这种研究被用来评估因为

所有四种测量误差来源（访调员、调查问卷、受访者和数据收集模式）综合影响所造成的答复偏误。

71. Groves（1989年）描述了三种记录检验研究的设计：

- 反向记录检验；
- 正向记录检验；
- 全面设计记录检验。

72. 在反向记录检验研究中，选取调查样本的来源拥有关于重要研究特点的准确数据。答复偏误估算以调查答复与准确数据来源之间的比较为基础。

73. 记录来源通常是具有特定特点的单位清单（住户或人员），例如那些接受特定形式政府帮助的住户或人员。在这种情况下，反向记录检验研究不测量多报误差（即报告他们不具备的特点的单位）。这些研究只能测量正确报告或没有正确报告特点的样本来源记录的比例。例如，美国法律执行援助管理局（1972年）曾开展过反向记录检验研究以评估受害报告中的误差。在调查访谈中，受害者报告的犯罪占警察部门记录的已知犯罪的74%。

74. 在正向记录检验研究中，获得调查答复后，需要寻找包括有关调查受访者准确信息的外部记录系统。答复误差估算以调查答复与记录系统值的比较为基础。正向记录检验研究提供了测量多报的机会。这种研究的困难之一是要联系记录保存机构，并在获取这些信息时征得受访者同意。如果调查答复显示该单位没有特定的特点，可能就难以从记录系统中搜寻该单位。因此，正向记录检验研究的能力限于测量少报。Chaney（1994年）描述了一项正向记录检验研究，该研究将教师对自己学术资格的评价与大学记录进行比较。数据显示，对所获学位的类型和年份以及专业方向的自我评价大部分是正确的；但是，对课程和学分的报告不太准确。

75. 全面设计记录检验研究同时综合了反向和正向记录检验设计的特点。从覆盖全部人口的标架中选抽样本，并找到了与该样本有关的所有来源的记录。因此，比较调查答复与调查答复人的所有记录（即来自样本标架以及外部来源的记录），就可以测量与多报和少报相关的误差。尽管这种类型的记录检验研究避免了反向和正向记录检验研究的缺点，但它确实需要有一个覆盖人口中所有单位以及这些单位的所有相关事件的数据库。Marquis和Moore（1990年）详细说明了为了估计美国收入和计划参与调查中的测量误差而进行的一项全面记录检验研究的设计和分析。在这项研究中，有四个州实施八个联邦和州福利方案，关于方案福利金领受数额的调查数据与同一方案的行政记录不符。调查质量概述（美国普查局，1998年）对所述设计和分析进行了总结。

76. 记录检验研究的三种类型共有的局限性与以下三个假设有关，这三个假设在实践中不切实际并且永远无法验证：首先，记录系统中不存在覆盖范围、无答复或丢失数据的误差；第二，这些系统中的单个记录是完整的、准确

的并且没有测量误差；第三，不存在匹配误差（在匹配受访者调查记录过程中出现的部分误差）或这一误差最小。

77. 具有特定特点的答复偏差能够根据以下公式，通过调查答复与该特点的记录检验值的平均差异进行估计：

$$\text{Response Bias} = \frac{1}{n} \sum_{i=1}^n (Y_i - X_i)$$

式中： n 是记录检验研究样本规模； $Y_i = i^{\text{th}}$ 样本人员的调查答复； $X_i = i^{\text{th}}$ 样本人员的记录检验值。

78. 从记录检验研究中得到的答复偏误量度提供了有关调查问题准确性以及确定产生有偏估计的问题的信息。这些量度也能被用于评估各种备选的调查设计特点，如调查问卷设计、回忆期、数据收集模式以及限制技术。例如，Cash和Moss（1972年）提供了在北卡罗来纳州三个县就机动车辆事故报告进行的一项反向记录检验研究的结果。访谈在包含样本人员的住户中进行，这些样本人员被确定为在访谈前12个月内曾经卷入机动车事故的人员。研究表明，访谈之前3个月内发生的事故只有3.4%没有报告，而访谈之前9至12个月内发生的事故则有27%以上没有报告。

5. 访调员方差研究

79. 为了研究访调员方差，访调员的分配必须是随机的，以便不同的访调员获得的结果的差异是访调员自身的影响所致。

80. 访调员方差通过把每名访调员分配给不同但是相似的受访者来进行估算，也就是说，这些受访者在调查变量方面具有相同的属性。在实践中，这种均等是通过随机化来得到保证的。样本被分成随机的子集，每一个都代表相同的人口，每一名访调员针对样本的不同子集开展工作。通过这种设计，每名访调员利用大规模调查中除其规模以外的所有基本属性，进行小规模调查。O'Muirheartaigh（1982年）介绍了在世界生育力调查中为了测量因访调员导致的答复方差而采用的方法，并提供了在秘鲁（1984年a）和莱索托（1984年b）进行的调查的答复方差估计数。

81. 在面对面访谈设计中，按地理位置相互分派访调员，以避免高额的差旅费用。所指定区域的规模足以满足一名访调员的工作量。确定成对的指定区域，并确定成对的访调员。在每个指定的区域中，样本住宅单位中随机选定的一半被分配给一对访调员中的每名访调员。这样，每名访调员都在两个指定的区域内完成访谈，而且每个指定的区域内都由两名不同的访调员负责。设计中包含着一项实验（对两名访调员在两个指定区域内的每个区域所得的结论进行比较），有多少成对的访调员，这个实验就被重复多少次。Bailey、Moore和Bailar（1978年）提供了在八个城市中举行的美国国家犯罪受害调查中个人相互交叉访谈的例子。

6. 行为编码

82. 访调员在培训和工作中的表现可以通过行为编码进行评估。经培训的观察员对一个访谈样本进行观察，并录制有关访谈或访谈样本的代码，然后根据磁带进行编码。将代码分配以记录各访调员的主要口头活动和行为，如提问、追问以及答复的整理。例如，代码可以对以下方面进行分类，即访调员如何阅读问题，提问是否正确和完整，提出的问题是否稍有变化和省略，以及访调员是否对问题进行了充分的复述或没有完成问题。编码系统包括追问是否引导受访者做出特定的答复，进一步确定问题或没有指示，对答复是否进行了正确或不正确的总结以及各种其他的行为是否恰当或不恰当。编码后的结果反映了访调员在何种程度上采用了他/她受训时学到的方法，即访调员在培训中学会应该避免的一种行为被定义为“不正确的”或“不恰当的”行为。为了保持每一次编码访谈高水平的编码可靠性，另外一名编码员应该独立为访谈子样本进行编码。

83. 行为编码系统能够让新的访调员明白他们的哪些访谈技术是可以令人接受的，哪些是不能被接受的。访调员和监督者可以以此编码系统为基础对实地工作进行检查并讨论通过编码发现的问题。此外，该系统提供一种评估访调员表现的手段，可以将访调员的表现与其他访调员的表现进行比较，或者与其他编码访谈中的个人表现进行比较（Cannell、Lawson和Hauser，1975年）。

84. Oksenberg、Cannell和Blixt（1996年）描述了美国卫生保健研究和质量局于1987年开展的国家医疗开支调查中进行的一项研究，在该研究中，对访调员的行为进行了录音记录、编码和分析，其目的在于确定访调员和受访者的问题。该研究试图了解访谈行为是否偏离了访调员培训中所包含的原则和技术。作者报告说，访调员往往不按照已有措词提问，有时候，他们以能够影响答复的方式提出问题。访调员没有进行足够多的必要追问；而他们追问时，则往往带有指示性或不恰当性。

D. 结束语：测量误差

85. 测量误差出现在数据收集过程。在这一过程中确定了四个主要来源：调查问卷、数据收集方法或模式、访调员和受访者。要想量化特定类型的测量误差的存在及其数量，就应当预先计划和深思熟虑。除非进行小规模（即有限样本）的研究，否则都需要进行需要随机子样本、重访和记录检验的特殊研究。这些研究的费用通常都很高，并需要一名统计员对数据进行分析。不过，如果对这个问题在调查准备阶段可能无法完全得到解决给予充分的关注，或者如果调查中的误差来源特别明显，调查管理者应采取措施，设计特殊的研究，量化误差的主要来源或问题来源。

86. 在一项调查中，不能过于强调为了了解和量化测量误差而进行研究的重要性。如果被衡量的调查概念是新概念，并且很复杂，则这一点尤为重要。使用者进行的分析依赖于他们拥有高质量的数据并了解数据的特点和局限性。测量误差研究要求对调查方案做出明确承诺，因为这些研究费用昂贵并且

非常耗时。但是，这种承诺并不随着研究的实施和进行而结束。必须对这些研究进行分析，并报告结果，这样分析人员就能对测量误差对结果的影响做出自己的评价。方法和技术报告中一般会报告侧重于试验和实验分析以及数据质量评估的特别研究 [例如，见人口与健康调查方案编制的方法和技术报告 (Stanton, Abderrahim和Hill, 1997年; Institute for Resource Development/Macro Systems Inc., 1990年; Curtis, 1995年)]。最后，测量误差研究的结论对改进下一次调查的投入非常重要。测量的重大改进在很大程度上依赖于知识和以前调查的结果。今后改进调查数据的质量要求调查研究专业人员做出承诺。

参考资料

- Bailey, L.、T. F. Moore和B. A. Bailar (1978年)。“国家犯罪调查城市样本中八个目标城市的访谈者方差研究”。《美国统计学会志》，第73期，第16-23页。
- Biemer, P. P. 和G. Forsman (1992年)。“将重新访谈数据的质量应用于目前的人口调查”。《美国统计学会志》，第87期，第915-923页。
- Biemer, P. P. 和其他人等 (1991年)。《调查中的测量误差》。纽约：约翰·威利国际出版公司。
- Bishop, G. F. 和其他人 (1988年)。《自填和电话调查中答复效果的对比》。《电话调查方法》，R. M. Groves和其他人等。纽约：约翰·威利国际出版公司，第321-340页。
- Blair, J.、G. Menon和B. Bickart (1991年)。“调查问题本人和代替答复的测量效果：从信息处理角度看”。《调查中的测量误差》，P. Biemer和其他人等。纽约：约翰·威利国际出版公司，第145-166页。
- Bradburn, N. M. (1983年)。“答复效果”。《调查研究手册》，P. H. Rossi、J. D. Wright和A. B. Anderson等。纽约：学术出版社，第289-328页。
- _____和S. Sudman (1991年)。“问卷设计的目前情况”。《调查中的测量误差》，P. Biemer和其他人等。纽约：约翰·威利国际出版公司，第29-40页。
- _____和Associates (1979年)。《改善访谈方法和问卷设计：调查研究中危险问题的答复效果》。加利福尼亚，旧金山：乔西-巴斯出版社。
- Brick, J. M.、L. Rizzo和J. Wernimont (1997年)。《学校安全和惩戒及学校戒备内容的访谈结果》。哥伦比亚特区华盛顿：美国教育部，国家教育统计中心。NCES97 - 339。
- Brick, J. M. 和其他人 (1996年)。《国家住户教育调查方案中答复偏差的评估：95个成人教育调查》。《工作文件，第96-13号》。哥伦比亚特区华盛顿，美国教育部，国家教育统计中心。
- Cannell, C. F.、S. A. Lawson和D. L. Hauser (1975年)。《评估访谈者业绩的技巧》。密歇根，安阿伯：密歇根大学，调查研究中心。

- Cash, W. S. 和A. J. Moss (1972年)。“机动车辆事故受害报告人的最佳回忆时期”。《生命和健康统计学》，第2期，第50号。哥伦比亚特区华盛顿：公共卫生局。
- Chaney, B. (1994年)。《教师本人关于其大专教育报告的准确性：教师学习成绩报告单、学校和支援调查》。《工作文件，第94-04号》，哥伦比亚特区华盛顿：美国教育部，国家教育统计中心。
- Collins, M. (1980年)。“访谈者差异：关于问题的审查”。《市场调研学会杂志》，第22期，第2号，第77-95页。
- Couper, M. P. 和其他人等 (1998年)。《计算机协助的调查资料收集》。纽约：约翰·威利国际出版公司。
- Curtis, S. L. (1995年)。《人口与健康调查-II中婴儿和儿童死亡率直接估计所使用数据质量的评估》。《专题文件，第3号》。马里兰，格鲁曼卡沃顿：Macro国际公司。
- _____和F. Arnold (1994年)。《根据重新访谈调查对巴基斯坦人口与健康调查进行的评估》。《专题文件，第1号》。马里兰，格鲁曼卡沃顿：Macro国际公司。
- Czaja R. 和J. Blair (1996年)。《设计调查：决定和程序指南》。加利福尼亚，千橡：Pine Forge出版社（属于赛奇出版社）。
- DeMaio, T. J. (1984年)。“社会愿望和调查测量：评论”。《调查主观现象》，C. F. Turner和E. Martin等。纽约：拉塞尔·塞奇出版社，第257-282页。
- Dillman, D. A. (1978年)。“邮件和电话调查：总体设计方法”。纽约：约翰·威利国际出版公司。
- _____ (1983年)。“邮件和其他自填问卷”。《调查研究手册》，P. Rossi、R. A. Wright和B. A. Anderson等。纽约：学术出版社，第359-377页。
- _____ (1991年)。“邮件调查的设计和管理”。《社会学年评》，第17期，第225-249页。
- _____ (2000年)。“邮件和网络调查：专用设计方法”。纽约：约翰·威利国际出版公司。
- Eisenhower, D.、N. A. Mathiowetz和D. Morganstein (1991年)。“回忆误差：根源和偏差削减方法”。《调查中的测量误差》，P. Biemer和其他人等。纽约：约翰·威利国际出版公司。第127-144页。
- Feindt, P.、I. Schreiner和J. Bushery (1997年)。《中心访谈：调查质量管理的工具》。《调查研究方法科的记录》。维吉尼亚，亚历山德里亚：美国统计学会，第105-110页。
- Forsman, G. 和I. Schreiner (1991年)。“重新访谈的设计和分析：概览”。《调查中的测量误差》，P. Biemer和其他人等。纽约：约翰·威利国际出版公司，第279-302页。

- Fowler, F. J. (1991年)。“通过访谈者培训、监督和其他方法减少与访谈者相关的误差”。《调查中的衡量误差》，P. Biemer和其他人等。纽约：约翰·威利国际出版公司，第259-275页。
- Groves, R. M. (1989年)。《调查误差与调查成本》。纽约：约翰·威利国际出版公司。
- _____和L. J. Magilavy (1986年)。“测量和解释访谈者效果”。《公共舆论季刊》，第50期，第251-256页。
- Hastie, R和D. Carlston (1980年)。“个人记忆的理论性问题”。《个人记忆：社会知觉中的认知偏差》，R. Hastie和其他人等。新泽西，希尔斯岱尔：劳伦斯·埃尔伯姆出版社，第1-53页。
- Hill, D. H. (1994年)。“小组调查中非独立和独立数据收集的相对经验有效性”。《官方统计杂志》，第10期，第4号，第359-380页。
- Huang, H. (1993年)。《关于收入和计划参与回忆长度研究的报告》。美国人口普查局的内部报告，哥伦比亚特区华盛顿。
- 资源开发机构/宏观制度公司 (1990年)。《人口与健康调查-I数据的评估》。《人口与健康调查方法报告，第1号》。马里兰，哥伦比亚：资源开发机构/Macro系统国际公司。
- Jenkins, C. 和D. Dillman (1997年)。“关于自填问卷设计的理论”。《调查测量和程序质量》，L. Lyberg和其他人等。纽约：约翰·威利国际出版公司，第165-196页。
- Kahn, R. L. 和C. F. Cannell (1957年)。《访谈的动态性》。纽约：约翰·威利国际出版公司。
- Kalton, G.、D. Kasprzyk和D. B. McMillen (1989年)。“小组调查中的非抽样误差”。《小组调查》，D. Kasprzyk和其他人等。纽约：约翰·威利国际出版公司，第249-270页。
- Kalton, G.、D. McMillen和D. Kazprzyk (1986年)。“收入和方案参与中的非抽样误差问题”。《人口普查局第二次年度研究会议的记录》。哥伦比亚特区华盛顿，第147-164页。
- Kantorowitz, M. (1992年)。《家庭支出调查中的方法问题》，Vitoria-Gasters, autonomous community of Euskadi: Euskal Estatistika-Erakundea, Instituto Vasco de Estadística。
- Kish, L. (1962年)。“因态度差异造成的访谈者方差的研究”。《美国统计学会刊》，第57期，第92-115页。
- Lyberg, L. 和D. Kasprzyk (1991年)。“数据收集方法和衡量误差：概览”。《调查中的测量误差》，P. Biemer和其他人等。纽约：约翰·威利国际出版公司，第237-258页。
- _____ P. Biemer、M. Collins、E. D. DeLeeuw、C. Dippo、N. Schwartz和D. Trewin (1997年)。《调查测量和程序质量》。纽约：约翰·威利国际出版公司。

- Marquis, K. H. 和 C. F. Cannell (1971年)。“健康访谈中一些经验方法对报告的影响”。《生命和健康统计学》，哥伦比亚特区华盛顿：公共卫生局，第2辑（数据评估和方法研究），第41号。
- _____和 J. C. Moore (1990年)。“收入和方案参与调查方案报告中的测量误差”。《人口普查局1990年度研究会议的记录》。哥伦比亚特区华盛顿，第721-745页。
- Mathiowetz, N. (2000年)。“回忆长度对调查数据质量的影响”。《第四次官方统计方法问题国际会议的记录》。斯德哥尔摩：瑞典统计局。请登陆：http://www.scb.se/Grupp/Omscb/_Dokument/Mathiowetz.pdf (2004年6月3日起可访问)。
- _____和 K. McGonagle (2000年)。“关于住户调查中独立访谈目前情况的评估”。《官方统计杂志》，第16期，第401-418页。
- Neter, J. (1970年)。“消费者支出报告中的测量误差”。《市场调研杂志》，第七期，第11-25页。
- _____和 J. Waksberg (1964年)。“住户访谈支出数据中答复误差的研究”。《美国统计学会志》，第59期，第8-55页。
- Nolin, M. J. 和 K. Chandler (1996年)。“国家住户教育调查中认知实验和记录访谈的使用”。哥伦比亚特区华盛顿：美国教育部，国家教育统计中心。NCES96 - 332。
- Oksenberg, L.、C. Cannell 和 S. Blixt (1996年)。“住户调查中访谈者和答复者行为的分析”。《国家医药支出调查方法7》。马里兰，罗克维尔：公共卫生局，保健和政策研究机构。
- O'Muircheartaigh, C. (1982年)。“答复误差项目的方法”。《世界生育率调查报告，第28号》。荷兰，福尔堡：国际统计学会。
- _____ (1984年a)。“莱索托生育率调查中答复方差的范围和模式”。《世界生育率调查科学报告，第70号》。荷兰，福尔堡：国际统计学会。
- _____ (1984年b)。“秘鲁生育率调查中答复偏差的范围和模式”。《世界生育率调查科学报告，第45号》。荷兰，福尔堡：国际统计学会。
- Schreiner, I.、K. Pennie 和 J. Newbrough (1988年)。“人口普查局调查中访谈者伪造的情况”。《调查研究方法科的记录》。维吉尼亚，亚历山德里亚：美国统计学会，第491-496页。
- Schuman, H. 和 S. Presser (1981年)。“态度调查的问题与答案”。纽约：学术出版社。
- Schwarz, N. (1997年)。“问卷设计：从概念到答案障碍重重”。《调查测量和程序质量》，L. Lyberg 和其他人等。纽约：约翰·威利国际出版公司，第29-46页。
- _____ R. M. Groves 和 H. Schuman (1995年)。“调查方法”。《调查方法方案工作文件汇编》，密歇根，安阿伯，密歇根大学，调查研究机构。

- _____和H. Hippler (1991年)。“替代备选方法：其选择和陈述顺序的影响”。《调查中的测量误差》，P. Biemer和其他人等。纽约：约翰·威利国际出版公司，第41-56页。
- _____和S. Sudman (1996年)。《回答问题：调查研究中确定认知和交流过程的方法》。加利福尼亚，旧金山：乔西-巴斯出版社。
- Silberstein, A. 和S. Scott (1991年)。“支出日记式记录调查及其相关的误差”。《调查中的测量误差》，P. Biemer和其他人等。纽约：约翰·威利国际出版公司，第303-326页。
- Singh, S. (1987年)。“数据质量的评估”。《世界生育率调查：评估》，J. Cleland和C. Scott等。纽约：牛津大学出版社，第618-643页。
- Sirken, M. 和其他人 (1999年)。《认知和调查研究》。纽约：约翰·威利国际出版公司。
- Stanton, C.、N. Abderrahim和K. Hill (1997年)。《人口与健康调查孕产妇死亡率指标：关于数据质量和数据使用意义的评估》。《人口与健康调查分析报告，第4号》。马里兰，格鲁曼卡沃顿：Macro国际公司。
- Sudman, S.、N. Bradburn和N. Schwarz (1996年)。《思考答案：将认知过程应用为调查方法》。加利福尼亚，旧金山：乔西-巴斯出版社。
- _____和其他人 (1977年)。“适中的期望：访谈者预先期望答复的效果”。《社会学方法和研究》，第6期，第2号，171-182页。
- Tucker, C. (1997年)。“关于调查研究中心理学应用的方法问题”。《社会学方法杂志》，第55期，第67-92页。
- 联合国 (1982年)。《国家住户调查能力方案：住户调查中的非抽样误差：根源、评估和控制：暂定版本》，DP/UN/INT-81-041/2。纽约：联合国技术合作发展部和统计办公室。
- 美国人口普查局 (1985年)。《评估人口和住户普查》。《统计培训文件》。哥伦比亚特区华盛顿ISP-TR-5。
- _____ (1998年)。《收入和方案参与调查质量简介》，第三版。哥伦比亚特区华盛顿：美国商务部。
- 美国联邦统计方法委员会 (2001年)。《测量和报告调查中的误差来源》，《统计政策工作文件，第31号》。哥伦比亚特区华盛顿：美国管理和预算办公室。请登陆：<http://www.fcs.gov> (2004年5月14日起可访问)。
- 美国执法援助管理局 (1972年)。《圣何塞已知犯罪受害者的方法测试》。《第一号统计技术报告》。哥伦比亚特区华盛顿。
- Vaessen, M. 和其他人 (1987年)。“将问卷翻译为当地语言”。《世界生育率调查：评估》，J. Cleland和C. Scott等。纽约：牛津大学出版社，第173-191页。
- Weiss, C. (1968年)。“福利母亲访谈应对的有效性”。《公共舆论季刊》，第32期，第622-633页。

- Westoff, C.、N. Goldman和L. Moreno (1990年)。《多米尼加共和国实验性研究：关于生育率和儿童健康信息的评估》。新泽西，普林斯顿：人口研究办公室；以及马里兰，哥伦比亚：资源开发机构/Macro系统国际公司。
- Willis, G. B. (1994年)。《认知访谈和问卷设计；培训手册》。《认知方法工作人员工作文件，第7号》。马里兰，海厄特维尔：国家健康统计中心。
- _____ P. Royston和D. Bercini (1991年)。“编制和测试调查问题时口头报告方法的使用”。《应用认知心理》，第5期，第251-267页。
- Woltman, H. F. 和J. B. Bushery (1977年)。《国家犯罪调查小组方差研究的最新报告》。美国人口普查局的内部报告，哥伦比亚特区华盛顿。

第十章

调查中的质量保证：标准、准则和程序

T. Bedirhan Üstun、Somnath Chatterji、Abdelhay Mechbal
和Christopher J. L.Murray

代表世界卫生调查（WHS）合作者*
世界卫生组织
政策依据和信息部门
瑞士日内瓦

* 世界卫生调查合作者的详细名单列于世界卫生调查的网站，网址是：<http://www.who.int/whs/>。

摘 要

调查质量对获得精确、可靠、有效的结果至关重要。调查小组应执行系统的质量保证程序，以防做出不切实际的调查，并把数据收集的误差降至最小。制定提高质量的有效策略将有助于及时收集高质量的数据，并使结果有效。“质量保证”也可被看作是实施预先设定的关于调查结构、过程和结果的业务标准的组织工具。调查小组应该坚持明确的质量标准，遵循规定的程序达到上述标准。各种程序应该是透明的，受到系统地监测，并作为调查实施和结果总文件的一部分审慎地提交报告。在切合实际的范围内，用可计量的参数对调查进行衡量和总结也非常重要。

本章概要论述获得质量保证量度的方法，不仅仅论述简单的控制机制。一项大型国际调查（由多个调查机构在71个不同国家开展的世界卫生调查）用来说明总质量保证方案应用的详细情况。该调查旨在收集可比数据，用于评估使用国内代表性样本的各参与国卫生制度的不同方面。根据世界卫生调查结果的重要性，出台了有力的质量保证程序，聘请国际专家组成外部同行审查小组，支持各国取得有关以下问题的经一致同意的可行质量标准：样本抽选方法、合格答复率的获取、漏失数据的处理、可靠性量度的计算、人口子群和各国的数据可比性检查。

关键词：质量保证、质量指标、世界卫生调查、漏失数据、答复率、抽样、可靠性、人口间可比性、国际比较。

A. 引言

1. 调查设计和实施的基本特征之一是调查的“质量”（Lyberg等，1997年）。在每次数据收集活动中，结果取决于投入；常言道：种瓜得瓜，种豆得豆。除了调查工具和分析技术的质量外，调查结果的质量主要取决于调查的实施，包括合理的抽样方法和调查问卷的正确管理。

2. 为获得最好的质量，每个调查小组都应当遵守一套标准的调查实施准则。这些准则包括：

- (a) 调查的每一步需要遵循的质量标准；
- (b) 质量保证（QA）程序，确定了在实际情况下监测调查实施的具体措施；
- (c) 对质量保证过程的评估，该过程用于衡量质量保证标准对调查结果和程序的影响，以提高整个质量保证过程的关联性及有效性（Biemer等，1991年）。

3. 准则的整体目标是提供支持，提高质量，而不是审计调查的实施。由于任何调查都是一项大的投资，涉及到很多参与方，所以其结果极为重要，会影响到一个国家的政策，因此，质量是严肃的业务工作重点，这一点极为重要。质量保证被看作是整个调查中的一个持续过程，贯穿通过收集和分析数据进行准备和抽样一直到编写报告的整个调查过程。准则还旨在保证各使用者更好地理解调查的设计。制定标准程序的目的是协助保证：

- 数据收集对国家的需要是相关的、有意义的；
- 数据在国家内部和国家之间进行对比，找出不同人口的异同；
- 调查的实际实施应遵守达成的协议；
- 数据收集中的误差降到最小；
- 数据收集的能力随时间不断提高。

B. 质量标准和保证程序

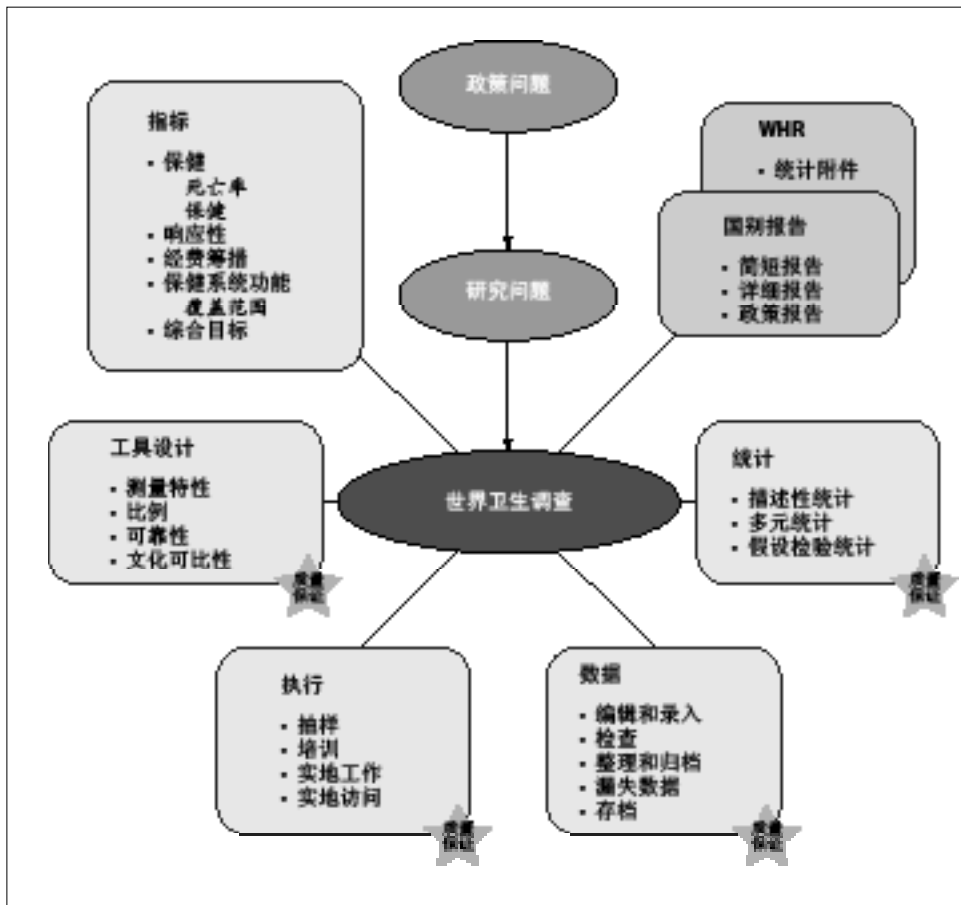
4. 质量保证（加拿大统计局，1998年）被定义为保持和提高调查数据的可靠性和有效性而对数据进行收集、处理和分析的方法或程序。质量保证可以被理解为具有类似但不同的意义。在本章，我们使用总体质量管理模式，检查每一步的调查过程，并试图总结出一种方法，这种方法既能减少抽样和非抽样误差，又能提高调查的关联性和可行性，还能提高国家实施调查的能力。为实现这一目标，并能保持切实可行，本章将采用世界卫生调查质量标准和保证程序（世界卫生组织，2002年），所有步骤包括：

- 选择调查机构；
- 抽样；

- 翻译；
- 培训；
- 调查实施；
- 数据输入/数据采集；
- 数据分析；
- 质量指标；
- 国别报告；
- 实地访问。

5. 图十.1描述了整个世界卫生调查的周期，并说明实施调查的各个阶段的上述步骤。由世界卫生调查大批参与者和国际专家起草的质量保证准则，旨在确定为了完成和监测高质量调查，可以采取什么样的最佳做法。调查实施的每一步都涉及到一定的质量检查。例如，重要的一点是，调查工具具有很好的测量特性，抽样能够代表目标人口，数据清楚而完整。

图十.1
世界卫生调查质量保证程序



6. 此套程序只不过是一个例子，说明针对作为一个过程的调查设计和实施，以及改善调查在关联性、准确性、一致性和可比性方面的产出的“质量保证方法”。任何设计和实施调查小组都可使用类似的方法，同时铭记其调查的具体目标以及本章提出的质量保证标准的可行性。最重要的是，应该重视质量，并应在业务工作范围内指导和监测质量。质量保证过程的结果应该使用定量术语进行报告，在可能进行测量的情况下使用适当的指标（例如，抽样比率、答复率、漏失数据、应用的试验 - 重新试验可靠性）。同时，应该使用定性术语，总结调查的结构、过程和结果。

C. 质量保证准则的切实实施：世界卫生调查的实例

7. 上述总体质量保证战略已经在世界卫生调查内实施，以提高调查的质量，包括亚洲和撒哈拉以南的若干发展中国家。本部分旨在使用质量保证标准、程序和报告作为具体指南。只要能符合其目的，其他调查小组可以使用这个例子。据我们所知，这是国际调查质量保证程序的首次系统应用，各实施机构和合作者已经发现这些程序在组织和各自的工作中非常实用。初始数据表明，可以尽早发现误差并防止误差，提高结果的完整性、准确性和有效性。

8. 世界卫生组织（WHO）已经开始开展世界卫生调查，把该调查作为现实的数据收集平台，用于不断收集有关不同人口和卫生制度的卫生信息（Üstün等, 2003年a, 2003年b; Valentine、de Silva和Murray, 2000年; 世界卫生组织, 2000年）。世界卫生调查满足各国对详细和持续的卫生信息系统的需求，并通过调查收集数据，用来衡量人口卫生的基本参数，同时，还为普通人口调查提供了程序和工具，为世界卫生组织的成员国提供可比数据。这些方法和工具在结构上可作为样本，并且通过对文献进行科学审查、与国际专家广泛协商和在63个以上国家用40种语言进行大规模试点试验后进行修改（Üstün等, 2003年a, 2003年c; 2001年）。世界卫生调查可以在实施过程中通过合作者的不断投入而不断发展，这些合作者包括政策制定者、调查机构、科学家和其他有关各方。各国和世界卫生组织共同拥有数据，有义务长期收集数据，提高本地能力，使用调查结果指导卫生政策的制定和实施。

9. 本章系统评述了调查过程的每一步骤，调查问卷调查和测试除外，因为这两部分另作评述（参见Üstün等, 2003年b）。本章还介绍了世界卫生调查在每一领域的质量保证标准。这些标准都是可取的，通过这些标准可以提高效率，防止不可取的做法。目前比过去更要注意质量，因为世界卫生调查的数据对世界卫生组织成员国日益重要，并且对卫生政策起着越来越重要的影响。世界卫生调查因此制定了实施调查的通用准则，旨在通过减少可预防的误差加强世界卫生调查的可靠性和有效性。已经采取的质量保证准则将成为世界卫生调查的主要组织工具，同时，还用于调查工作、准备工作和计划实施的组织过程中。因此，本章为需要特别注意的关键问题提供总体指导，以保证收集到高质量的数据。

10. 这些准则还将用作调查管理人和质量保证顾问（这是一批具有广泛调查经验、担任整个过程同行审查人的国际专家）的评价模板。这些人员将对各国进行实地访问，支持这些国家实施世界卫生调查，并对过程进行有序的详细评估，这将有助于各国对质量进行系统评估，以及确定调查活动中需要改进的地方。

1. 选择调查机构

11. 开展国内调查需要广泛的知识、技能、资源和专长。这些要求使得调查活动的组织根据不同国家和部门的不同风格和传统进行。为保证在既定的国家内选择能胜任的调查小组进行世界卫生调查，有必要确定优秀的调查机构，制定约聘条件的标准。世界卫生调查通常的做法是，通过与各国卫生部、区域办事处和世界卫生组织国家的代表或联络官确定这些机构。鉴于调查的规模和复杂性，世界卫生组织的章程要求，可行性应该通过合同投标程序进行论证。该过程开始时，要求胜任的调查机构根据抽样、访谈和数据收集的技术规定向世界卫生调查投标。[世界卫生调查的技术规定可在世界卫生调查的网站查询，网址：www.who.int/whs。] 根据一系列标准对这些投标书进行比较，然后做出最终决定。

12. 评价潜在机构的绩效标准包括：

- 这些机构的以往业绩记录（即这些机构至少在最近五年进行大型全国调查的经验，抽样规模应在3 000人或以上）；
- 这些机构开展全面调查的能力（即抽样、培训、数据收集和分析）；
- 这些机构在不同数据收集方式方面的经验，包括面对面访谈（以及其他方式，如电话、信函、电脑等方式的访谈）；
- 以前所做调查的文献情况（包括样本表示法的调查度量、全国人口的覆盖面、访谈的质量、费用和培训方式、质量保证及其他调查程序）；
- 对调查日程的一般时限记录以及这些机构在规定时间内完成调查的能力的记录；
- 这些机构与卫生部、国家统计机构和其他机构密切合作，开发和利用良好的卫生系统基础设施的潜力。

13. 合同投标程序对确定质量和成本方面的最佳报价是有用的，有助于对一个国家内的所有报价者做出比较评估。这样，世界卫生组织和卫生部可以确定最佳调查机构，以期建设进一步的调查能力，把世界卫生调查的数据纳入卫生信息系统。约聘的过程还有利于对未能提供结果以及未能保证质量的行为进行处罚。鼓励联合投标，以保证相关各合伙人（例如卫生部和国家统计局）共同努力，确保获得最佳抽样标架。

14. 利用上述标准对提交的不同方案概述进行仔细审查。对这种对比性分析进行记录。

15. 总之，重要的是不仅应确定符合有关国家预期调查技术规范的优秀机构，还应为该机构提供必要的技术支持，以获得预期结果。对于大规模全国性调查，往往有必要在该国建立不同小组、机构和人员之间的伙伴关系，这些小组、机构和人员应具备设计、培训、实施、数据处理、分析和编写报告方面的必备知识。

2. 抽样

16. 好的抽样等于好的调查。如果抽样设计/或实施或两者都有误，则很难弥补样本设计中有限的代表性或补充缺失的信息。调查结果会莫名其妙地出现偏误，而且偏误的数量往往也不可量化。

17. 由于需要在实地进行大量应用，世界卫生组织和一批国际技术专家确定了一套准则，以确保在世界卫生调查中获得好的样本 [为参与国制订的世界卫生调查抽样准则见世界卫生组织的网站 (www.who.int/whs)]。科学抽样的标准以概率选择方法为基础，该标准广为人知并得到认可 (Üstün等, 2001年; Kish, 1995年a)。但是，由于操作欠佳，抽样程序在实地的实施缺乏监督，以及/或在特定情况下实施成本过高，这些标准常常没有得到充分遵守。

18. 世界卫生组织的准则强调，调查抽样的科学原则是明确的标准，同时还提供妥善的抽样计划范例，并确定各国需遵守的质量保证标准。必要时，世界卫生组织和技术顾问将向各国提供技术支持。以下是世界卫生调查抽样中的重要方面：

- (a) 世界卫生调查的抽样目标应是实际人口（即住在该国的所有人员，包括外来工人、移民和难民）而不是法律上的人口（只包括该国公民）。样本要具有良好的代表性，要成为该国全部人口的“缩影”，这一点非常重要。为此，样本必须代表居住在该国的所有人并覆盖该国的所有地域；
- (b) 根据调查目的不同，样本规模必须足以提供对国家级或国以下各级利益数量的良好（可靠）估算；与此同时，调查管理者必须平衡对更大样本规模的需求，以便根据相应增加的调查成本做出更好的估算。大的样本规模并不能弥补低劣的质量。出于各种目的，可能要求有足够的少数民族代表（例如种族或其他小团体），这可能需多抽样（即提供更高的选择概率）。如果因为任何科学研究问题而需要对分组人口多抽样，则必须清晰而详细地对这种行为进行说明。如果要多抽样，在数据分析阶段需要进行微分加权，以修正因为多抽样而造成的失真；
- (c) 在世界卫生调查中，一个覆盖全部有关关键利益子群体90%的抽样标架（即一份地理区域、住户或个人的清单，从中选抽样本，例如一份

计算机处理过的人口清单、一次最近的普查、选民名册等)被认为是可以接受的。参与国应使用能够获得的最近的抽样标架。如果抽样标架已过时两年或两年以上,通常有必要在样本选择的倒数第二个阶段查点或列出所有住户以更新该标架。在进行选择之前,可以使用快速清点方法以更新主要抽样单位大小的衡量方法;这样的方法包括采用由于绘图过时或其他原因而缺乏最新标架的选定路径进行清点。在选择出的抽样领域除采用快速清点方法外,还可使用其他来源更新标架,这些方法如当地邮局提供的邮政地址、水电费用收取公司的名单等。重要的是,人口要根据最新的人口调查进行科学权衡;

- (d) 世界卫生组织的样本目标是一般人口中18岁及以上的一切成年人。²²在大多数情况下,它以最新的普查资料作为其抽样标架的基础。住户是根据多级分层整群抽样程序进行选择。每个住户的单个成员通过随机选择程序进行选择[例如,Kish表法(Kish,1995年a)或其他方法,如最后的出生日期法和Trohdahl/Carter/Bryant法(Bryant,1975年)]。只要详细记录下选择的号码,这一阶段就可以使用随机编好的表格。不论使用何种选择技术,都应尽力在这一领域实际的实施阶段减少选择偏误。各国应该设法设计能够实现调查的测量目标的最简单的抽样计划。如果设计过于复杂,则很难实施,并且不能控制误差的出现。保证质量的关键在于可行性以及拥有监测抽样设计的数据轨迹;
- (e) 世界卫生组织使用了联合国关于住户的定义;²³但是,该定义可能因各地情况不同而有差异。应该在国别报告中详细说明这些住户定义的差别可能对抽样造成的影响。如果各国使用了住户抽样标架,建议这些国家在调查中使用与原来标架中住户定义相同的定义;
- (f) 世界卫生组织采用科学的抽样策略,它包括调查中所有人的已知非零选择概率。在抽样的每一个阶段使用严格的概率方法非常关键,这样,可以使抽样数据推断整个人口。否则,调查结果将不具代表性和有效性;
- (g) 很难在普通人口调查中纳入专门机构的人口,因为需要制订不同的标架。在机构(如医院、疗养院、军营和监狱)中进行访谈还存在很多道德问题。由于不同国家有不同的机构收容方式,所以不可能找到单一的解决方案。如果在调查期间能访谈这些人,世界卫生组织试图把因健康状况而被送入专门机构的人纳入到调查范围之内。然后,我们使用普查所得的机构人口率检查调查中所得比率的一致性。当然,这只是世界卫生组织的一个特殊考虑,因为和不生活在这些专门机构中的人相比,生活在特殊机构的人如疗养院中的病人和长期住院的病人等,其身体状况往往要差的多,所以生活在专门机构的人需要被纳入抽样标架,以减少低估健康状况的可能性;

²² 目前,世界卫生组织的受访者仅为成年人。以后的工作旨在开展一项也将儿童纳入受访者的调查。

²³ 联合国把住户定义为居住在同一住所、共同使用炊具和餐具(即用餐的来源相同)的一群人。对世界卫生组织,个人如果因健康问题目前正在某一机构治疗则通常被看作是住户的一部分。这种在某个机构内的人口应包括在住户的成员名单之内。

- (h) 世界卫生组织抽样准则明确解释了单位不答复的含义以及目标样本和已获得样本的不答复率计算方式。世界卫生组织的抽样策略不允许用另外一个住户或个人代替不答复；
- (i) 抽样调查的结果应该报告重要调查变量的标准误差，以使用户可以明白统计用语中的测量误差；
- (j) 使用地理信息系统（GIS）可以通过核证抽样计划的实地执行而提高结果的质量；换句话说，访谈实际上已经在某个地点开展，而不是所谓的街头巷尾或虚构的访谈（De Lepper、Scholten和Stern, 1995年）。在调查中，地理信息系统可以通过链接到保健设施、水和其他环境资源的距离等信息和测定的卫生参数（如健康状况、疾病和危险因素），增加数据的价值。还可以在地图上标示任何参数的分布性质，从而显示健康的不平衡性。为此，世界卫生组织一直使用全球定位系统（GPS）装置和数字化的地图按一定的准则对数据进行地理编码（请查询<http://www3.who.int/whosis/gis>）。由于地理编码信息可能会违反数据保护标准，因此采取了一定的法律措施保持个人信息的机密性。

抽样的评价

19. 在调查开始之前应该对抽样策略进行评价，以便评估分层的正确性、人口的代表性是否充分以及所选群体的规模和分布。报告应该详细记录在实地采用的确切程序，同时注意到设计的偏差，以使用户可以更好地了解调查结果的质量。

20. 在数据收集阶段，为准确起见（比如使用Kish表和住户名册的准确性），实地和/或办事处监督人员应细心监测对住户和个人的选择。

21. 收集数据后，使用数据分析量度（详见下文讨论）评估数据的质量，评估所采用的手段如下：

- 总体统计，称作“抽样离差指数”（SDI）；
- 试验——重复试验的可靠性，以表示不同访调员所使用工具的“稳定性”；
- 关于不答复和漏失数据的程度的信息。

22. 这些程序在有关数据分析的部分进行更为详细的描述。表十.1是关于抽样质量的详细一览表。

3. 翻译

23. 要对跨文化的数据进行有意义的比较，需要衡量不同国家相同结构的相关工具。对现有调查工具进行科学的审查，向大量专家广泛咨询，并在多国调查研究中进行系统的实地试验后，开发出了世界卫生组织调查工具（Üstün等，2003年a）。我们已经介绍过调查工具的特点、相关性以及在其他

地方的文化适用性（Üstün等，2003年b）。对其他的调查而言，设计者必须努力获得最佳的工具和措施，并确保他们的工具适合他们的目的，具有良好的测量性，并已经通过彻底的小规模试验以确保其可行性和稳定性。

表十.1

抽样质量一览表

- 人口组成概览（城市/农村、少数民族、语言、抽样过多的群体）
- 抽样标架和抽样阶段数：
 - 抽样标架是否涵盖所有目标人口？
 - 抽样标架有多新？
- 抽样标架内分层
- 每个阶段的抽样单位：已知的选择概率
- 每个阶段的抽样单位规模：保证所有抽样单位有一个超过预先设定的最小规模的量度
- 检查抽样单位的“实地”规模，以及在每个选定的地址有一个还是多个住户和如何在这些住户中进行选择等问题
- 所选样本的规模
- 住户的概率权数
- 受访者的概率权数
- 有关使用和正确实施Kish表（或其他）的培训
- 检查在住户中选择受访者的程序
- 有关实际实施中的抽样、离差、权数和标准误差的报告

24. 一旦你拥有了良好的调查工具，翻译就是保证不同语言中各种问题对等性的重要手段之一。由于我们生活在一个多元社会，所以，用以度量调查中的相同概念的翻译，质量上乘至关重要。

25. 在一个国家，文件通常根据该国的不同语言群体的大小翻译成多种语言。建议对任何一个超过总人口5%的语言群体，都应用其语言对他们进行访谈。如果某个受访者受访的语言没有经过正式翻译，重点将放在对主要概念的理解上。访调员应该利用该国现有的一种译文，采用总体准则，使用没有翻译的语言提出问题。大型的多国调查面临的另一个挑战是，非洲和亚洲国家的许多语言都不是书面的，因此没有文字材料。在这种情况下，建议继续准备与准则一致的标准翻译，并利用该国另一种熟悉语言的文字翻译拟定书面译文。

26. 世界卫生组织调查文的翻译准则来自世界卫生组织与众多合作伙伴和语言专家开展和实施的大量国际研究。世界卫生组织的翻译准则可以在世界卫生组织

查的网站（www.who.int/whs）查询。该准则强调保持概念对等的重要性，并保证制定一个程序，查明可能存在的缺陷，避免曲解含义：

- 翻译的目的应当是编制当地能够理解的调查问卷。
- 翻译问题的原意时，应尽可能使用当地语言中对等的术语。
- 逐个问题说明的目的应该是传递问题以及事先编码的答复选项的原有含义。
- 调查问卷首先应该由对主题内容关键概念有基本了解的卫生和调查专家翻译。由语言学专家对一系列经过挑选的关键术语以及在第一次直接翻译中被证明有问题的术语进行回译，并对这些术语可能的翻译和建议的翻译提出意见。在该国调查主任监督下的一个编辑小组对翻译和回译进行审查，并将翻译质量上报给世界卫生组织。
- 应该利用焦点小组和定性语言学方法（例如编制当地表达方式的详细目录，并与其他语言中的表达方式进行比较）来提高质量。世界卫生组织已经对特定语言中的翻译和认知访谈进行了系统的研究，并将这些研究的结果纳入世界卫生调查问卷的现有文本。对于使用翻译的调查问卷进行的“认知访谈”（即对何种主题被理解为问题的含义这一问题进行的进一步探索研究），仍建议利用当地的主题进行这种访谈。应当将世界卫生调查的所有文件（即世界卫生调查问卷、逐个问题的说明、调查手册和培训手册）翻译成当地语言。数据输入程序仍然使用英文。但是，如果该国已经根据世界卫生组织的规范利用电子媒介对世界卫生调查的问卷进行了翻译，则数据输入程序能够自动以其他语言生成。
- 每一个世界卫生调查参与国都应该在试点阶段接近尾声时提交一份有关翻译工作质量的报告。如果发现有些术语特别难译，应提交特别语言评价表，对难译的特性加以说明。
- 该国质量保证顾问应该特别关注翻译过程的实施步骤，并应与该国的调查主任一起检查关键术语清单。
- 在许多方言和/或语言没有文字形式的国家，应该与世界卫生组织讨论具体翻译协议。

翻译评价

27. 世界卫生调查试点访谈开始之前，应该向世界卫生组织提交一份完全翻译的调查问卷。译文应该由特定语言的相关专家进行检查，如有必要，可向提交国提出意见。

28. 应审查回译的关键术语的清单以及关于翻译过程和相应产生的问题的报告。语言评价表（Üstün等，2001年）应该先由国别调查管理者进行系统的

检查，然后由世界卫生组织进行检查，以便找出问题特别大的项目，并且在可行时，能够在各种语言之间找到一种共同的解决方案。

29. 如果对某个术语、短语或问题不理解，则应当和访调员讨论如何理解在实地采用的程序这一问题。这些讨论应审查访调员在何种程度上“解释”和“说明”对受访者提出的问题。

表十.2

翻译程序审查一览表

<ul style="list-style-type: none"> • 该国所讲的语言；主要语言群体的覆盖程度； • 参与翻译过程的有哪些人？ • 所有所需材料是否都已经翻译？ <ul style="list-style-type: none"> • 调查问卷 • 附录 • 管理指南（仅适用于访调员不懂英语的情况） • 调查手册（仅适用于访调员不懂英语的情况） • 结果编码 • 翻译中出现何种问题？ • 协议内容是什么（比如，是把译件全文还是主要项目清单交给世界卫生组织）？ • 是否填写完语言评价表？
--

D. 培训

30. 调查小组的培训是质量的关键。培训是一个持续的过程，应该在数据收集开始或在数据收集过程中进行，最后，随着实地工作时间的结束，写出详细的报告。

31. 参与调查的各级调查小组都应该接受培训，从访调员到培训人员到监督人员，以及在全国监督过程的中心小组等，均应接受培训。这样将保证所有参与人员在保证数据的高质量方面明确自己的地位和作用。

32. 整个培训的目的如下：

- 保证调查材料的统一应用；
- 说明研究和研究协议的理由；
- 动员访调员；
- 提供切实可行的建议；
- 提高数据的总体质量。

33. 为完成部分培训目的，世界卫生组织已经为来自所有参与国的主要调查人员举办了世界卫生调查区域培训讲习班，并编写了各种培训教材，包括涉及所有培训问题的培训录像带和教育光盘。

访调员的挑选

34. 使用经验丰富的访谈人员以及熟悉调查主题的其他人员都是非常重要的。

35. 访调员应该至少在其所在国家接受过完整的教育，并且熟练掌握该国的主要语言。各国应在挑选之前决定需要何种更高教育水平以及采取何种正式评价方式。

36. 各国自行决定是否要求访调员是卫生工作者。使用单独的数据库记录访调员的个人特点（包括年龄、性别、受教育水平、职业培训、工作状况、过去的调查经验等等）。然后，这些信息可以和负责每份调查问卷的访调员的标识号相连，并对每个访调员的业绩进行分析。

培训时间、方法和内容

37. 培训的时间长短应该以保证访调员既熟悉成功访谈的技巧又熟悉所要使用的调查问卷内容为准。对于有经验的访调员，培训时间可以比那些无经验访调员的培训时间短一些。

38. 世界卫生调查建议培训时间为3—5天，有经验的访调员如果只需接受调查问卷内容方面的培训，则只需要3天。对于其他所有访调员，建议延长培训时间。

39. 所有培训应该尽可能由同一培训小组承担，以保证同一期的访调员和不同时间及地点的不同访调员均获得标准的培训。为降低成本、支持地区培训，培训可以分散、分级进行。但是，这种成本上的利得与分散或变化不定的培训的弊端相比，得不偿失。

40. 如果在数据收集期间能够进行，则强烈建议开展辅助培训。建议最好在世界卫生调查中期的某个时候进行。辅助培训可以审查数据收集的各个方面，重点放在复杂而有难度的工作上或访调员未能充分遵循的准则上。这样的辅助培训有利于对已经完成的工作量以及积极的方面进行反馈，包括监督人员和中心调查小组向访调员进行反馈，以及访调员向监督人员和调查小组进行反馈。

41. 培训方法应该包括尽可能多的访谈角色扮演（每个访调员应该至少扮演一次）。这种方法有助于更加有效地模仿访谈技巧。为使角色扮演更加有效，在培训之前要准备好不同的书面材料，这样，就可以说明访谈的不同分支结构、所允许的解釋的特性以及在对难对付的受访者进行访谈期间预计会出现的问题。

42. 除角色扮演外，在开始真正收集数据之前，至少应该有一次机会对访谈小组以外的真正受访者进行访谈。在培训期间，练习性访谈应该尽可能录音或录像以便进行复习和反馈讨论。如有可能，鼓励世界卫生调查参与国制定与世界卫生组织的录像类似的标准培训录像。每次角色扮演或练习性访谈之后均予以反馈。

43. 应该为所有访调员提供培训材料，用作参考资料。提供的所有材料在培训期间均进行全面温习，并将相关的部分翻译成该国使用的各种语言。

44. 培训的内容应该包括以下方面：

- 管理问题；
- 实地工作计划；
- 温习提供的所有材料；
- 接触程序、同意表和机密性。

访谈应包括：

- 实地访谈程序；
- 实地监督和报告程序；
- 调查小组的组成以及小组成员的职责。

培训的评价

45. 应该在各级进行培训评价。对访调员进行评价以便确定这些访调员是否具有有效访谈的能力以及他们需要何种支持。反过来，访调员可以对培训以及培训工作者进行评价。在数据收集开始时和在实地工作结束时，应该进行持续评价。

46. 中心调查小组必须对监督人员进行类似的评价。值得一提的是，培训的性质必须保持与监督人员应该完成的任务相一致，例如把拒答转换为愿意回答、对所选的访谈进行反复核实以及对访谈内容进行编辑。在培训期间应对这些程序起草详细的协议，并做出清楚的说明。

47. 培训结束时可以对访调员进行正式评价，为合格的访调员颁发某种形式的证书。该项事宜由各国自行决定并实施。

E. 调查的实施

48. 无论是从后勤管理还是从其他方面来看，计划和管理调查的实施都是一项复杂的任务。需要做好充分的准备、安排以及在实地的各种力量的调度，从而获得理想的样本。从策略上看，调查实施是确定调查数据质量优劣的关键因素。因此，注意实际调查的实施质量以及对此进行监测是非常重要的，这样，在实施的过程中就可以解决问题。

表十.3

培训程序综合概述清单

- 培训期数
- 培训天数
- 哪些人提供培训，他们在培训和卫生调查领域拥有哪些专门知识？
- 使用了何种文献？
- 实践部分：实际场景中角色扮演的情況
- 培训中遇到的问题
- 对培训的评价

49. 在整个过程中，在实地展开调查的方式是确定质量的一个步骤。各国对调查进行周密而有力的统一组织将有助于保证质量。每一步（即印刷调查问卷、制作样本清单、登记主题、派遣访谈小组、在实地展开日常监督、编辑调查问卷等等）均应该仔细计划并加以审查，以确保质量。具体的做法是：

- (a) 每一个调查小组都应该制定中心调查实施计划以及任务日程表，明确列出调查后勤工作的细节。该计划应确认对某一特定地区的某一部分确定样本进行一定数量的电话访谈（包括回复电话）需要多少访调员及其成功率。同时，应该考虑到预计的非答复率和不完整的访谈以及调查小组在当地的存。
- (b) 每一个调查小组应该拥有一名监督人，负责监督和协调访调员的工作，并提供现场培训和支持。世界卫生调查的监督员和访调员间的理想比率视国家和地点的不同，变化范围在1:5到1:10之间。
- (c) 监督人应该在工作日开始时与访调员一道展开日常工作，一天结束的时候检查工作结果。在检查中，访调员将向监督人介绍其访谈过程和结果。监督人必须检查完成的访谈，确保访调员对各住户家的受访者做出正确选择，并确保调查问卷既完整，编码又准确。
- (d) 世界卫生调查每个参与国的调查中心都应该记工作日志以监测调查工作的进展。日志应记录以下内容：
 - 能联系上的受访者人数、已完成的访谈和未完成的访谈；
 - 答复、拒答和无联系率；
 - 回访数和访谈结果。

必须掌握每个访调员的信息，以确保其工作得到监督人的持续监测。可以使用访调员数据库以便提供个别反馈，同时对未来的雇用人选做出决定。

- (e) 各国应在世界卫生调查开始阶段展开试验调查。该试验调查应该持续一到两周的时间，作为正式调查的排练。试验样本的50%须由另一位

访调员重新访谈，以证明访谈应用的稳定性。试验阶段应该得到批评性评价，并与世界卫生组织进行讨论。试验所得数据应迅速进行分析，以发现任何特别的实施问题。由于调查中所使用的工具在试点之前本已经过广泛的事先试验，试点试验的目的在于找出语言和可行性方面存在的小问题，使真正的调查阶段得到更好的规划。同时在调查中还要使用跳跃方式发现明显的特别错误。从试点得到的反馈将纠正这些错误，并进行微调。与世界卫生组织协商后，应开始正式研究。

- (f) 应该意识到印刷和调查问卷实际整理的有用性（例如，旋转装置的彩色编码和答复卡片的分层）。所有参与国均应向世界卫生组织寄送一份印制的文件。
- (g) 根据世界卫生组织的合约规定，监督人员或其他小组应该对10%的受访者进行随机检查，这种检查可以通过电话进行，也可以当面进行，以确保初始访谈恰当正确。复查访谈应涉及基本的人口统计学信息以及初始访谈中没有收集的信息。
- (h) 根据世界卫生组织的合约规定，随机抽查10%的全部受访者样本，由另一访调员在第一次访谈后的7天内进行第二次访谈，这样即可评估调查问卷的可靠性（再检验的受访者不应是上述第（8）条所述的复查的受访者）。
- (i) 应对答复率进行持续的监测，每个中心应该采取各种策略提高调查的参与程度，减少非答复率。例如，可以通过电视、广播、报纸或地方媒体渠道发布公告，也可以寄信或卡片给参加者，向地方卫生工作者谋求帮助，实施鼓励参与的机制，与地方传统当局或其他公认的当局举行谈判等，这些都是可用来最大限度提高答复率的公关手段。个别中心也可使用特别的方法。
- (j) 每项调查均旨在获得最高答复率。世界卫生组织合约规定要求总体答复率至少是75%，但这并不意味着在调查实施中停留在75%的水平上，这只是指根据许多国家过去的调查，世界卫生组织合作者一致同意的最低可接受标准。在许多情况下，世界卫生组织调查答复率一直很高，答复率可随国家的不同而不同，可以和该国其他调查的答复率进行比较。在计算答复率时，所有国家均应使用完整访谈的相同定义。在数据整理程序中使用了一种算法，以确定根据一组关键变量进行的访谈是否完整。
- (k) 回访：根据世界卫生组织的规定，调查小组应该进行10次回访（包括打电话，留下便条或卡片说明访调员访问过）。回访的平均数取决于答复率，每个中心应该检查每次追加回访的收获，与世界卫生组织就该国的回访数是否充分进行协商。
- (l) 调查实施严重依赖手头的资源。每项调查都应当在该国范围内进行评估。非常有必要比较各项调查同该国的其他可比调查。在评价中应考虑到地方风俗和传统。在有关对质量的影响方面，用较少的访调员在

较长的研究期限内做更多的访谈，还是用较多的访调员在较短的研究期限内做较少的访谈，应在两者之间进行权衡。

表十.4
调查实施情况审查一览表

试验调查

- 试验在何地展开？
- 为试验提供了何种培训？
- 数据输入中是否存在数据问题？
- 数据分析：查看结果；遇到了何种问题？
- 试验中是否发生了方法上的变化？
- 在翻译方面是否因试验发生了变化？

真正的调查

- 访调员、监督人员和中心协调员的人数：
 - 如何进行监督？反馈
- 后勤安排：
 - 差旅：到达住户的方便程度如何？乘坐何种交通工具？
 - 小组的组织
- 联系程序：
 - 联系受访者的方便程度如何？
 - 访问过几次？
 - 拒答率如何？拒绝接受访谈的主要原因是什么？
- 访调员的报酬
- 同意书的签字和记录（作为调查问卷的一部分或单独一页）
- 监督者的实地检查程序
- 集中检查的程序
- 把调查问卷送回中心办事处和安全性
- 对调查问卷的最终检查和纠错程序
- 检查程序和监督

每周调查情况报告

- 评估访谈过程
- 审查答复、拒答和不联系的比例：保证答复率
- 监测结果并保证收集了数据

核查记录

- 联系（已经取得的联系/试图进行的联系）的次数是否有详细记录？
 - 每个访调员所做的访谈中是否有至少10%得到了核查，以保证部分答案保持不变（年龄、学历和住户组成），并保证确实进行过访谈？
 - 检查已经做过的访谈数和访谈日程安排的规划
 - 核实完成的访谈和拒绝访谈的最终结果编码已经合理分配
 - 检查知情同意书是否签署

从调查问卷和数据输入程序分离出的所有标识信息
对待采取的行动提出建议的报告草案

F. 数据输入

50. 调查的持续产出就是数据，重要的是要准确而及时地获取数据。世界卫生组织调查数据输入过程是经过规划的，这样，就可以立即输入地方数据并进行集中协调。在收集到数据以后尽快将它转移到计算机媒体非常必要。任何错误一经发现，都可以在实地调查过程中予以处理。

51. 下面的图十.2描述了世界卫生组织调查中的数据流以及与该数据流有关的质量保证步骤。右方列示的是在国家一级完成的任务，左方列示的是在世界卫生组织完成的任务。

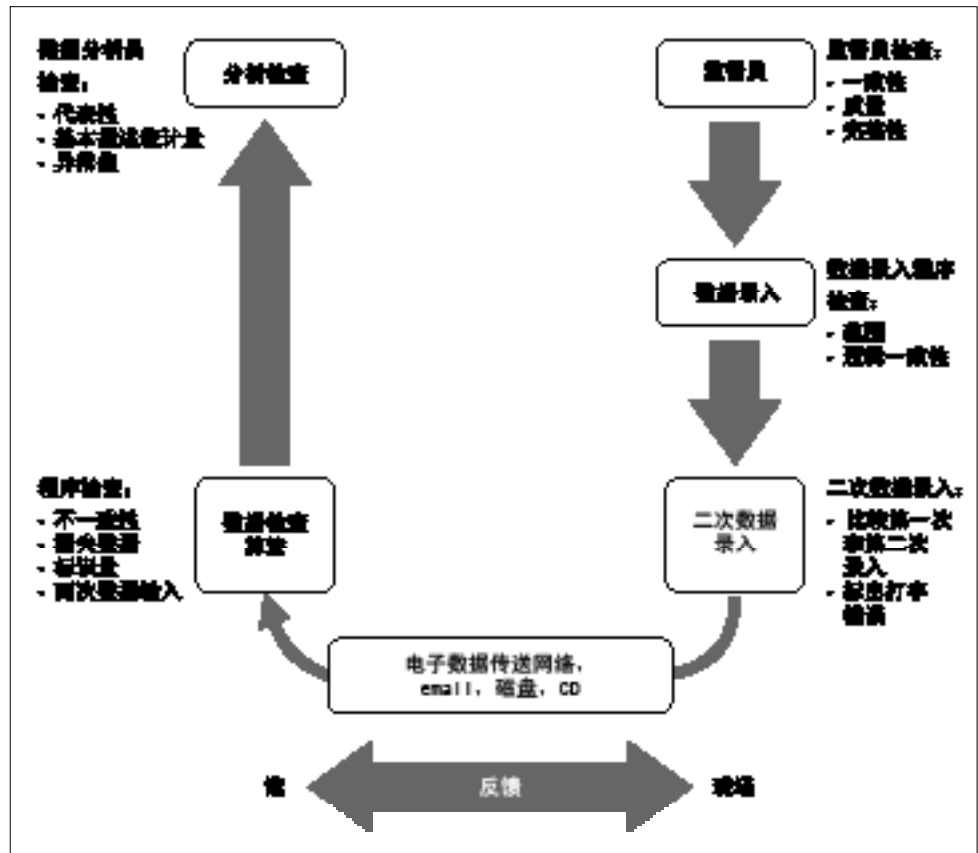
52. 访谈结束后，采取以下步骤：

- 数据输入开始前，监督人检查调查表。
- 使用世界卫生组织数据输入程序进行数据输入（或者数据采集/登记）。该程序检查范围（例如允许的答复变量范围），保证相关代码的逻辑一致性（例如，疾病持续的时间不得超过一个人的年龄，男子不可能患有妇科疾病等）。
- 进行第二次数据输入，目的在于发现录入错误以及偶尔漏掉的问题。
- 使用email、CD-ROM或磁盘等方式，将数据分批发给世界卫生组织。
- 一旦世界卫生组织收到数据，各种程序便开始检查是否有不一致的地方、遗漏值、标识号问题或试验/二次试验的情况。这些程序生成的报告将返还各国。另外，从现场国家获得的纠正信息应当应用于数据。
- 数据分析者检查数据的代表性、基本描述统计量和异常值。代表性通过比较实际样本的年龄 - 性别分布与预期人口分布进行检查。基本描述

统计量用来确定答复的分布并确定偏态分布、怪异的结果和异常值。

- 世界卫生组织把反馈发给各国，如有必要，各国将根据反馈提交校正的信息和/或解释。

图十.2
数据输入和质量监测过程



53. 有关数据输入的重要质量问题:

- 数据输入应使用数据输入程序进行，该程序具有质量检查特征。因此，使用不含这些特征的其他程序可能有不利之处。
- 填好的访谈表应该由监督人检查，然后开始数据输入工作。
- 除负责的小组成员外，其他人员不得使用数据输入程序，这对数据保密十分必要。
- 需要两次数据输入，以避免数据的录入或编辑错误。在第二次输入完成后，数据输入程序确认数据输入了两次。
- 各国在输入标识号（ID）时应该非常仔细。有效的标识清单将送交各

国。程序拥有检查和的数字，以确保标识编码的正确输入。使用正确的标识对再检验情形尤其重要，因为标识用来匹配检验个案与再检验个案。

- 数据应定期提交世界卫生组织，譬如，可以每天或每周提交一次。
- 一旦世界卫生组织从各国获得数据，随着数据收集的继续进行，对数据进行检查，并向各国提供反馈。
- 使用一定的规则保持有关数据的完整和准确，例如，对数据进行检查从而确定同一位受访者是否被利用了两次，并确定漏失数据的数量。

54. 确认信息与调查问卷分开，一旦输入信息，数据输入程序将保密信息保存在一个单独的文档。各国有责任对信息保密。如果通过互联网传送数据，要通过加密保证数据的安全。

数据输入的评价

55. 应该对以下各个方面进行认真监测和审查（见表十.5）：

- 数据输入人员的人数及其培训；
- 每人每天输入的表格数，包括错误率；
- 数据输入的检查程序和监督；
- 实地访谈完成和数据输入之间的时间间隔；
- 已完成的访谈提交世界卫生组织的数量和频率以及在发送数据方面遇到的问题。

56. 在计算机辅助的访谈中，由于在访谈的过程中就可以输入数据，所以数据输入方面的一些问题就会减到最少，但是，这些计算机程序要求安装检查工具，以便确保利用跳转和转移规则正确实施访谈，同时保证输入规定范围内的相应数据。

G. 数据分析

57. 在对世界卫生调查数据进行实质性数据分析前，应对数据质量进行大量的系统检查。这些检查的汇编被称为“世界卫生调查量度”，提供了数据质量的综合指数。

58. 调查量度的要素：

- 完整性，包括答复率（考虑到某些住户的合格性可能是未知的，在这种情况下，须对合格的住户比例做出估算。如果把这样的住户从答复率的计算中剔除，则需对这些住户没有合格受访者这一假设做出合理清楚的说明）和不完整的调查问卷或未答题项。计算出全体受访者每一题的漏失数据频率以及每个受访者在全部题项层面的漏失数据频率，这有助于发现调查实施的问题，尤其是调查问卷中有问题的题项。

表十.5
数据输入过程一览表

- 数据输入人员是哪些人？
- 数据输入人员的完成率和错误率如何？是否有数据输入人员需要接受再培训？
- 观察数据输入过程。记录分配给每一个访调员的调查问卷数所使用的是何种系统？
- 讨论数据分析和数据质量矩阵的计算以及对更多支助的需求
- **调查问卷**

从每个访调员处选择若干完成的调查问卷并检查：

- 调查问卷删除了姓名
- 调查问卷没有封面
- 住户名册是随机的，而且按要求完成
- 字迹清晰整洁
- 正确地记录了选项（例如，圈出了选项，而不是打勾号，加下划线或打叉号）
- 应当回答的开放式问题已经回答
- 逐字记录了开放式问题
- 问题正确地跳过
- 应由妇女回答的问题只由妇女回答
- 两次输入数据。
- 使用数据输入程序：
 - 核实数据的保密性和安全性
 - 数据是否两次输入？
 - 根据硬拷贝检查数据库中的编码
 - 检查范围、一致性、路径及其他错误
 - 检查漏失数据的数量

- 样本离差指数（SDI），用来衡量样本在代表性方面与目标人口的偏差程度。如果该测量显示偏差很大，那么，需分层进行分析。样本离差指数可以使用卡方统计进行正式评估。如果对某些关键子群体的抽样故意高出既定标准，则应该考虑到这一问题，以便按照故意过多抽样系数调整样本离差指数。

- 可靠性，用来表示对不同时间、不同访调员的同一受访者使用相同测量工具所得结果的重复性。该分析使用了在50%的试验访谈和10%全体样本中采用的检验/再检验协议。
- 与外部有效数据进行比较，与其他调查结果比较，比如与普查、调查和服务数据以及私营部门和公共部门的数据进行比较。

59. 这些量度将在下一部分另作详述。数据处理在国家一级进行，原因是国家可提供必要的能力，并且在世界卫生组织总部进行。

60. 进一步的国家级数据分析对于保证结果的有效利用十分必要。世界卫生组织总部和区域办公室将确定在数据的全面分析方面需要支持的国家，并制定提供这种支持的机制。

数据分析评价

61. 此方面的评价需要讨论该国是否具备进行分析的技能，以及需要多少支助或该国可以向其他国家提供多少支助。

H. 质量指标

62. 利用指标概括质量保证是有用的。这些指标以后可以用来评价影响调查质量的其他关联因素，这样就完成了质量循环。据我们了解，在综合措施中，还没有提出一套系统的指标用来监测和报告调查的质量。世界卫生组织使用下文所述的一些可计量指标以及在同行审查过程分阶段进行的定性评估（即质量保证报告）。

63. 总体看来，任何住户调查容易出现两种误差：抽样误差和非抽样误差。抽样误差的出现是因为调查是在整体样本的基础上进行的，而不是在全部整体的基础上进行的。抽样误差受到了样本规模、总体中很多方面发生的变异以及样本设计其他方面如分层和类集效应的影响。另一方面，非抽样误差受到了诸如调查项目概念的特性、抽样标架的准确性和完整性、实地实际选择程序与预定样本设计之间的一致程度以及调查实施误差等方面的影响。上述最后一个因素产生的问题有：调查问卷设计不当，访调员在提问时出错以及答复错误或回答时报告错误、数据输入错误和其他处理错误、不答复或估算方法不正确。下文中论述了一些对测量和量化有益的非抽样误差。

64. 在监测调查数据的最终结果时，目前使用了以下标准指标用于监测世界卫生组织调查的数据质量。

1. 样本离差指数

65. 样本离差指数（SDI）²⁴显示了样本中年龄和性别各层在独立来源的人口数据中的比例，而独立来源的人口被假定为是标准的。世界卫生组织使用了联合国人口数据库作为独立的来源，但是也可以使用其他更新的可靠人口数据来源。样本离差指数是样本数据质量在代表性方面的一个指标（即该样本多

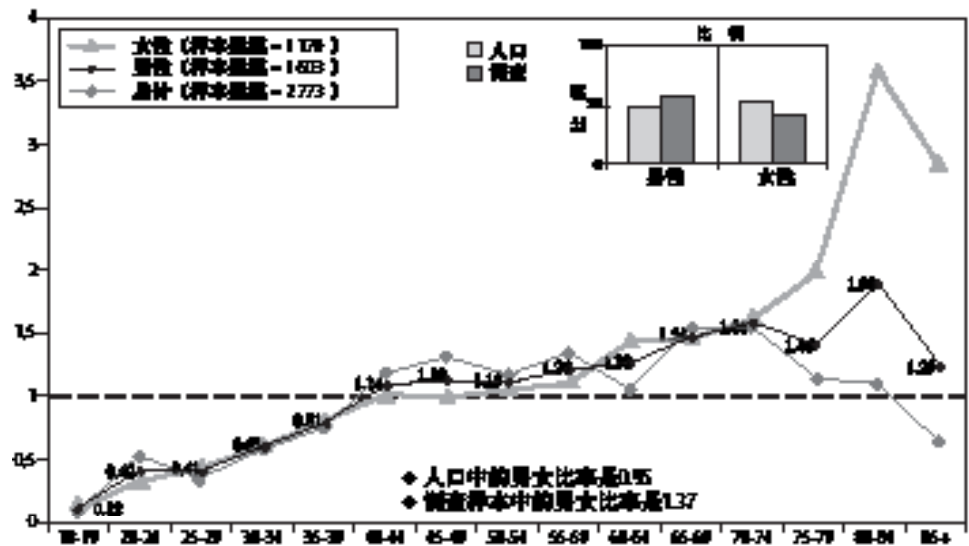
$$^{24} \text{ 在 } SDI = \sum_{a=1}^a |1 - index_a|$$

中， a = 年龄组，指数是年龄组中的样本与年龄组中的人口比率，其来源是联合国人口数据库或其他最新的来源，如该国的普查。该指数说明样本在年龄或性别分布方面代表的人口数量，可以用卡方或pi-star同质性检验进行检验。

么恰当地代表了整个人口)。比率1显示调查样本符合某一年龄或性别群体的特征,而大于1或小于1则表示该年龄或性别群体的抽样过多或过少。

66. 由于抽样误差,在调查中很难得到预期值1(最理想的代表性)。图十.3列示其中一项调查的样本离差指数,说明代表年轻群体的样本少而代表年长群体的样本过多,特别是代表年长男子的样本过多。

图十.3
样本离差指数



2. 答复率

67. 答复率说明所选样本中的访谈完成率,即在有资格纳入调查的个人或住户中完成访谈的数量(如果选定的“住户”最后被发现是空房子,则该“住户”被视为不合格)。该指标显示了调查在获得理想的100%答复率方面的表现有多好。最低可接受的答复率为60%,而世界卫生组织调查要求答复率起码为75%。

3. 漏失数据率

68. 漏失数据率被定义为在访谈受访者时漏失题项的比例。世界卫生组织调查通过衡量对题项的答复未达到最低规定的人的比例(例如,住户面对面访谈为10%)来确定访谈的质量,从而也确定有资格的受访者不予答复率居高(5%以上)的问题项。

4. 检验 - 再检验访谈的可靠性系数

69. 检验 - 再检验访谈的可靠性系数说明访谈在两种情况下的响应变异性方面的稳定性。这种系数被计算为随机纠正的一致率(即持续变量的分类和组内相关系数卡帕统计)。该项指标指的是调查中某个给定的项/问题在重复访谈时产生相同结果的概率如何。一般说来,大于0.4的分值被看

作是可接受的；大于0.6的分值被视为良好，而大于0.8的分值被视为极佳（Cohen, 1960年；Fleiss, 1981年）。

70. 在抽样过程中产生的数据误差方面，调查质量的主要指标对调查的每个关键统计量来说都是估算的标准误差。它显示抽样误差围绕某个给定估算值的估计范围（例如，加减3%）。如有可能，就计算多级类集样本的有关量度和设计效果系数，该系数是实际样本的方差与相同规模假定简单随机样本方差的比率。由于大型调查因费用（包括交通费）问题，真正简单随机样本不切实际，所以，常规的做法是为了与随机样本进行比较，计算抽样方差（标准误差的平方）（Kish, 1995年b）。对于世界卫生组织的有关指标而言，设计效果介于1 - 6之间一般被视为可接受的数值。

I. 国别报告

71. 质量保证的一个重要特点与报告数据方面的最终产出有关，原因是调查在下述方面产生了影响，即它对我们的知识基础和进一步的政策方针的提供具有附加值。正确的报告显然与世界卫生组织与国家需求的关联性密切相关。世界卫生组织调查结果将在许多不同类型的报告中公布，即：

(a) 世界卫生组织每个参与国的国别报告：

(一) 为决策人士和公众提供执行摘要；

(二) 为研究人员或其他科学部门用户提供详细报告；

(b) 有关特定问题的区域和国际报告。

72. 国别报告 [上文第71段第(a)项] 的初始模板包括：

- 导言，包括（例如，驱动政策制定的信息和有关卫生系统的信息）；
- 调查实施的讨论（例如，包括调查描述、抽样方法、培训、数据收集和处理、质量保证程序、调查量度描述）；
- 调查结果和政策影响概述（例如，应包括卫生系统的投入、人口和住户特征、卫生干预的覆盖范围、人口健康、卫生系统的响应力；卫生支出）；
- 结论：针对卫生政策和监测该国千年发展目标的特别建议。

73. 将与各国、区域办事处和其他相关各方相互合作，进一步完善所述模板。

74. 应通过媒体、研讨会和其他活动明确制定国别报告的宣传策略。有必要在政策辩论中让不同的利益相关者使用从调查中所获得的信息。

75. 各国应该为编制国别报告负主要责任。世界卫生组织将帮助提供基本的数据和技术支持及工具，与编制小组一起编写和讨论这些国别报告。

76. 世界卫生调查在获得有关人口卫生和卫生系统的信息方面起重要作用。这些因素包括卫生系统工作评价框架的许多要素。另外，调查提供有关其他方面的详细信息，如特别的危险因素、卫生系统的功能、特殊疾病流行病学和卫生服务。因此，从世界卫生调查数据中获取最好的信息价值是非常重要的。

77. 一些国家可能也想使用世界卫生调查数据进行国内分区分析。在大多数情况下，这可能需更大的样本规模。换句话说，世界卫生调查数据可以和其他数据来源，如普查和其他调查一起使用。

78. 从长远来看，世界卫生调查的模块结构预计能将有关卫生和卫生系统的各种调查合并到某个单项调查中。

国别报告的评价

79. 数据分析和国别报告的起草是调查工作的高潮阶段。报告的质量和讨论结果的方式将决定以后几轮调查的实施方式，并决定结果对该国政策制定和监测的影响。

J. 实地访问

80. 世界卫生调查参与国事先清楚他们在实施世界卫生调查和质量保证程序方面应该做哪些工作。记录这方面的实地工作是非常重要的。为实现这一目标，世界卫生组织将聘请独立的质量保证顾问，这些顾问将到每个国家进行实地访问。这些实地访问实际上是调查实施过程中的一项外部同行审查，将单独记录对质量保证标准的遵行情况。实地访问还将提供机会发现调查过程中问题，尽早解决问题。国家小组和质量保证顾问将按照世界卫生组织的准则一道对总体调查质量做出分阶段评估。

81. 质量保证是一个过程，不可能只被视为实地访问的单一活动。质量保证顾问和国家调查小组之间的关系可以被看作是由三个阶段组成的长期过程：实地访问之前、之中和之后。

82. 实地访问之前，各国和质量保证顾问应编写访问文件，文件涉及本份文件概要介绍的世界卫生组织质量保证准则的基本格式，并包括实地访问清单的所有方面。这份文件还将包含有关实地、调查机构、抽样设计、本地专门知识、本地使用的工具和培训包以及世界卫生调查国别报告的模板等所有背景信息。没有提供的资料将在实地访问期间提供。

83. 世界卫生组织总部的国家官员以及质量保证顾问将与国内的主要调查员或主要调查官员直接联系，使得质量保证过程成为调查实施过程的有机组成部分。这将有助于在调查中建立质量保证文化。质量保证过程的目的在于审计或管制，而在于通过提供帮助和支持保证世界卫生调查的质量。

84. 为使实地访问取得最好的结果，实地访问应安排在培训结束和数据收集开始时。实地访问将重点放在调查过程的方方面面，也就是说，发现问题，提出解决办法，熟悉地方情况，提供支持，建立长期的关系。

85. 质量保证顾问（QAA）在访问各国时，其职责是发现问题并调查实施中的实力。他们的主要任务是审查该国使用的世界卫生调查实施过程，并发现同预期质量保证标准有无偏差之处。顾问们对这种偏差是否重要以及如何纠正这些偏差做出的判断非常重要。质量保证顾问还应该通过与世界卫生组织总部进行讨论直接提供支持，或安排其他实体提供相关支助。

86. 质量保证顾问将根据结构清楚的清单进行评价，该清单将涵盖按主次顺序排列的各个步骤。评价应该包括对“调查量度”（只要在实地访问时已经输入了一些数据）进行的分析，其中包括数据质量指标。

87. 将与各国调查小组和世界卫生组织一起进行质量保证评价。各国事先应该明确他们在质量保证程序方面应该做的工作。

88. 实地访问报告编写完成之后，应编制世界卫生调查国别报告，该报告是实地访问和国家支持的最终产物。实地访问时就应该着手起草国别报告，并探讨特别的编制策略，包括如何使用政策拟定方面的结果。

K. 结论

89. 质量保证是调查实施中的核心问题。在调查的每个阶段都有必要而且有可能确定质量保证机制。如果这些机制在操作层面上进行了界定，那么，这些机制就可以得到度量，也就可以监测整个调查的质量。

90. 质量保证要求实施调查的人员转变观念，因为每一阶段的审查和评估都是硬性规定。

91. 在整个调查过程中持续进行质量指标评估很有必要。这个过程不应仅被视为后续过程，还应用于通过发现问题和正确解决问题进行必要的中期纠错。生产过程中的这种重要的持续质量改善或总体质量管理应纳入所有调查。

92. 由于现在可以使用电脑工具，所以就可以开发调查管理和跟踪系统，对调查过程进行持续跟踪，这有助于提高数据的信度。

93. 在定性报告和定量指标（即样本离差指数、答复率、漏失数据比例和检验 - 再检验的可靠性）方面以系统的方式记录关键问题（如关于调查实施和培训等问题）是非常重要的，以便为数据用户提供有关调查质量的重要信息。

94. 质量保证过程的预期结果是产生能够获得优质数据的调查。这样，调查结果可以记录为有效、可靠和可比。

95. 这些质量保证程序的持续实施将为容许的国际数据采集工作确定标准，监测这些标准的方法将继续不断发展。

鸣 谢

来自不同国家和机构的下述调查专家参与了世界卫生调查质量保证准则的编写，谨对他们表示感谢：

Farid Abolhassani博士，伊朗伊斯兰共和国
 Sergio Aguilar-Gaxiola博士，美国
 Atalay Alem博士，埃塞俄比亚
 Lorna Bailie博士，加拿大
 Russell Blamey博士，澳大利亚
 Carlos Gomez-Restrepo博士，哥伦比亚
 Oye Gureje博士，尼日利亚
 Holub Jiri博士，捷克共和国
 Mark Isserow先生，南非
 Feng Jiang博士，中国
 Jean-Louis Lanoe先生，法国
 Howard Meltzer教授，大不列颠及北爱尔兰联合王国
 Steve Motlatla先生，南非
 Lipika Nanda女士，印度
 Kültegin Ögel博士，土耳其
 Gustavo Olaiz Fernandez博士，墨西哥
 Mhamed Ouakrim博士，摩洛哥
 Jorun Ramm博士，挪威
 Wafa Salloum博士，阿拉伯叙利亚共和国
 Shen Mingming博士，中国
 Benjamin Vicente博士，智利

抽样顾问

Steve Heeringa教授，美国密歇根大学社会研究学院
 Nanjamma Chinnappa教授，印度，国际调查统计员协会前任主席

世界卫生组织区域顾问

M. Mohale M.女士，世界卫生组织非洲区域办事处区域顾问
 Siddiqi Sameen博士，世界卫生组织东地中海区域办事处区域顾问
 Amina Elghamry博士，世界卫生组织东地中海区域办事处区域顾问
 Lars Moller博士，世界卫生组织欧洲区域办事处区域顾问
 Myint Htwe博士，世界卫生组织东南亚区域办事处区域顾问
 Soe Nyunt-U博士，世界卫生组织西太平洋区域办事处区域顾问

参考资料

- Biemer, P. P. 和其他人等 (1991年)。《调查中的衡量误差》。纽约：威利国际出版公司。
- Bryant, B. E. (1975年)。“改变住户构成时答复者的选择”。《市场调研杂志》，第12期，第129-135页。
- Cohen, J. (1960年)。“名义尺度协议的折算率”。《教育和心理测量》，第20期，第37-46页。
- DeLepper, M. H.、H. Scholten和R. Stern等 (1995年)。《公共和环境健康地理信息系统的增加值》。荷兰，多德雷赫特：克鲁维尔学术出版社。
- Fleiss, J. L. (1981年)。《比率和比例的统计方法》，第二版。纽约：约翰·威利国际出版公司。
- Kish, L. (1995年a)。《抽样调查》。纽约：约翰·威利国际出版公司。
- _____ (1995年b)。“设计效果的方法”。《官方统计杂志》，第11期，第55-77页。
- Lyberg, L. E. 和其他人等 (1997年)。《调查测量和程序质量》。纽约：威利国际出版公司。
- 加拿大统计局 (1998年)。《质量指标》，第三版。渥太华。
- Üstün, T. B. 和其他人 (2001年)。《残疾与文化：普遍性与多样性》。德国，哥廷根：霍格林夫-哈伯出版社。
- _____ (2003年a)。“2000-2001年卫生组织健康与应对多国调查研究”。《保健系统业绩评估：辩论、方法和经验》(C. J. L. Murray和D. B. Evans等)。日内瓦：卫生组织。
- _____ (2003年b)。“世界卫生调查”。《保健系统业绩评估：辩论、方法和经验》(C. J. L. Murray和D. B. Evans等)。日内瓦：卫生组织。
- _____ (2003年c)。《世界卫生组织残疾评定量表二：发展和心理学测试》。日内瓦：卫生组织。卫生组织/国家健康联合项目合作者研究所合作完成。
- Valentine, N. B.、A. de Silva和C. J. L. Murray (2000年)。《估计应对程度和191个国家的分布情况：方法与结果》。《全球证据讨论方案文件汇编，第22号》。日内瓦，卫生组织。
- 世界卫生组织 (2000年)。《世界卫生报告》。日内瓦：卫生组织。
- _____ (2002年)。《世界卫生调查：质量保证和指标：国家调查小组实施质量保证的程序和质量保证建议》。日内瓦：卫生组织。

第十一章

巴西调查非抽样误差的报告和补救：目前的做法和未来的挑战

Pedro Luis do Nascimento Silva

国家统计学学校/巴西地理和统计研究所（统计学校/地理统计所）

巴西，里约热内卢

摘 要

本章讨论巴西非抽样误差当前一些报告和补救的做法，其中考虑三类误差：覆盖误差、不答复和测量与处理误差。它还确定致使难以集中精力注意非抽样误差测量和控制的一些因素。此外，它还确定最近一些可能有助于使情况改观的倡议。

关键词：调查过程、覆盖、不答复、测量误差、调查报告、数据质量。

A. 引言

1. 适用于某种未知目标数量（或参数）统计或估计的误差概念必须界定。它指的是估计数（如 \hat{Y} ）与所有误差源消除情况下将会得出或报告的理论上的“真实参数值”（如 Y ）之差。或许正如有人认为是的那样，一个更好的术语将是偏差[见Platek和Särndal著作中的讨论（2001年，第5节）]。不过，术语误差已经非常根深蒂固，因此我们不应尝试避开它。这里我们关注的是调查误差，也就是说基于调查数据的估计的误差。根据Lyberg等人（1997年，第13页），“可将调查误差分解为两大类：抽样和非抽样误差”。使用现代术语讨论调查误差是数据质量更广泛的讨论的组成部分。

2. 为说明该概念，假设某项调查中报告的某个人口群的平均月收入的估计数是900美元，而且从全数调查获得而无报告和处理误差的这一人口群成员的实际月平均收入是850美元。那么在本例中，估计误差将是950美元。一般来说，调查误差是不被人觉察的，因为真实的参数值是不被人觉察的（或无法觉察的）。至少统计估计数抽样误差可被观察到的一个例子是计算机记录的抽样提供的误差，其中估计数与利用完整数据集计算的值之间的差在需要的情况下可以计算。普查记录的公开使用样本提供实际应用的一个例子。在巴西，自1970年以来，这种类型的样本是从普查记录选出的。不过，像这样的情况属于例外而不是通例。

3. 抽样误差指基于样本调查的估计数与普查利用相同测量方法进行时会得到的相应人口值之差，并“由通过观察样本而不是整个人口群所造成”（Särndal、Swensson和Wretman，1992年，第16页）。“非抽样误差包括影响调查的所有其他误差”（同上）。在包括普查在内的所有种类的调查中，非抽样误差都能发生而且确实发生。在普查和采用大型样本的调查中，非抽样误差是人们必须关注的主要误差源。

4. 调查估计可能出现两类误差：偏倚误差和变量误差。偏倚指影响调查估计数预期值并且使它偏离目标参数实际值的误差。变量误差影响调查估计数在调查过程潜在重复中分布的分散程度。关于抽样误差，利用适当的抽样程序、样本容量和估计方法，通常可避免偏倚或者可使偏倚忽略不计。因此，散布是抽样误差分布的主要方面，应加以考虑。说明这种散布的一个关键参数是标准误差，即抽样误差分布的标准偏差。

5. 非抽样误差包括两大类误差（Särndal、Swensson和Wretman，1992年，第16页）：“不观察引起的误差”和“观察中的误差”。之所以产生不观察引起的误差，原因是不能从目标人口各组成部分获得所需的数据（覆盖误差）或不能从所选样本各部分获取所需的数据（不答复误差）。覆盖或框架误差指调查框架中调查单位的错误包含、遗漏和重复，导致目标人口的覆盖过度或不足。发生不答复误差的原因是没有从为调查选定的单位获取数据。观察误差可分为三类：设定误差、测量误差和处理误差。Biemer和Fecso（1995年，第15章）将设定误差界定为下述三种情况下发生的误差：“（1）调查概念不可计量或界定不明确；（2）调查目标未适当具体规定；或（3）收集的数据与规定的概念或目标变量不相对应”。测量误差涉及在数据收集后所观察到的调查问题和变量的值不同于利用理想或金本位测量方法将会获得的对应的实际值。处理误差是在处理所收集的数据的过程中发生的误差，即在调查数据的编码、键入、编辑、加权和制表过程中发生的误差。所有这些类型的误差在B节的各分节中论述，但设定误差除外。我们不讨论设定误差并不是说它们不重要，只是因为这些误差的讨论和处理在巴西还没有完全确立。

6. 对非抽样误差分类的其他办法在联合国手册（见联合国，1982年）中讨论。在有些情况下，不答复、覆盖面与测量误差之间没有明确的分界线，就如在一次多阶段住户样本调查中在一个接受调查的住户中遗漏一个住户成员时那样：这是测量误差、不答复误差还是覆盖问题？

7. 非抽样误差还能划分为非抽样方差和非抽样偏倚。如果使同一样本在相同的基本条件下经受调查过程的假设性重复（联合国，1982年，第20页），非抽样方差计量调查估计数中的差异。非抽样偏倚指产生于调查过程和调查条件的误差，而且，这种偏倚产生的调查估计数的预期值不同于实际参数值。作为非抽样偏倚的一个例子，假定人口群中的个人倾向于将其收入平均低报30%。然后，不管抽样设计和估计程序，在无任何外部信息的情况下，平均收入的调查估计数将比人口群成员平均收入实际值平均小30%。本章的多数讨论论述非抽样偏倚的避免或抵消问题。

8. 近年来，样本调查中的数据质量问题日益受到注意，若干倡议和出版物论述该专题，包括若干次国际会议（见D节）。遗憾的是，讨论仍主要限于发达国家，发展中国家和转型国家很少参与和做贡献。这是在审查在这些不同会议和倡议后发表的议事录和出版物后得出的主要结论。不过，在杂志《转型中的统计学》上对于转型国家的调查最近也就这一专题发表了几篇论文（Kordos, 2002年），但是这份杂志在发展中世界的图书馆订阅范围似乎并不广。

9. 关于抽样误差，存在着一种统一的测量和估计理论[例如见Särndal、Swensson和Wretman（1992年）]，它得到作为调查实践方面抽样标准的概率抽样方法和技术的广泛传播的支持（Kalton, 2002年），而且还得到使得这一理论能实际应用于实际调查的标准通用软件的支持。如果恰当地抽取和收集样本，调查估计的抽样可变性的估计数就相对易于计算。发展中国家和转型国家的许多调查已经在这样做了，尽管这种做法仍远未成为一种强制性的标准。

10. 不过，此种变异度量的传播和分析仍远远落在后面。在许多调查中，抽样误差估计数既不计算也不公布，或只针对少量选定的变量/估计数做出计算/公布。一般来说，它们不能用于大多数的调查估计，因为它涉及大量计算业务。虽然这可能使得外部用户难以评估所关心的某种特定变量的抽样变异程度，但有可能通过将它与估计了标准误差的一个类似变量作比较而估定它的数量级。关于调查估计的评注往往忽略估计的可变程度。例如，1980年开始的巴西月劳动力调查（巴西地理和统计研究所，2002年b）计算和公布根据调查估计的主要指标的变异系数的月估计数。不过，不为连续月份或隔年月份间此类指标之差计算标准误差的估计数。然而，每月连同估计数公布的多数调查评注是关于变动情况（月指标的变动）。只是到了最近，才为内部分析计算变化估计数标准误差的估计数[见Correa、Silva和Freitas（2002年）]，而且这些估计数尚未经常提供给调查结果的外部用户使用。在估计数“复杂”时也是这种情况，按季调整的一系列劳动力市场指标就属于这种情况。

11. 如果有关抽样误差的情况远不理想，其中理论和软件都能广泛利用，而且抽样文化已广泛传播，发展中国家住户调查和其他调查的非抽样误差的处理则更加不发达。缺乏广泛接受的统一理论[见Lyberg等人（1997年，第十三页）；Platek和Särndal（2001年）；以及随后的讨论]，缺乏用来汇编关于非抽样误差部分的信息和估计它们参数的标准方法，以及缺乏承认测量、评估和报告这些误差重要性的文化，都意味着非抽样误差及其测量和评估在发展中国家或转型国家进行的调查中受到的关注较少。这不是说发展中国家或转型国家进行的调查多数质量低下，而是强调我们对于它们的质量水平知之很少。

12. 借助于关于发展中国家和转型国家进行的调查的非抽样误差测量和控制状况的这种背景信息，我们进而讨论关于巴西经验的当前做法的状况（B节）。虽然局限于在一国（巴西）发现的情况，我们相信，鉴于关于这一主题文献稀缺，本次讨论对其他发展中国家的统计人员是相关的。然后，我们再次从巴西调查实践的角度出发，表明发展中国家和转型国家今后改进调查做法面临的挑战（C节）。

B. 巴西住户调查非抽样误差报告和抵消的现行做法

13. 在巴西，覆盖面广的主要定期住户抽样调查由巴西地理和统计研究所（地理统计所）这一巴西中央统计研究所进行。为帮助读者了解这些调查的参考资料，我们在表十一.1中列示它们的主要特点、覆盖面和时期。

表十一.1

巴西主要住户抽样调查的部分特点

调查名称	时 期	人口覆盖面	专题/主题
普查	每10年（最近一次在2000年）	我国私人和集体住户的居民	住户项目、婚姻状况、生育率、死亡率、宗教、种族、教育、劳动力，收入
全国住户抽样调查（住户抽样调查）	年度，普查年除外	我国私人和集体住户中的居民，北部区域农村地区除外	住户项目、宗教、种族、教育、劳动力、收入和各种专题的特殊补充项目
月劳动力调查（月劳力调查）	每月	6大都市区私人住户中的居民	教育、劳动力、收入
住户支出调查（支出调查）	1974-1975、1986-1987、1995-1996、2002-2003	2002-2003年度版全国；前两版11个大都市区；1974-1975年度版全国	住户项目、家庭收支
生活水平衡量调查（水平调查）	1996-1997	东北部和东南部区域私人住户中的居民	广泛覆盖有关生活水平衡量的专题
城市非正规经济调查（非规经调）	1997	城市地区私人住户中介入非正规经济的居民	非正规经济领域的劳动力、收入和经营特点

1. 覆盖误差

14. 覆盖误差指调查人口单位覆盖不足或过度。目标人口中的单位从框架遗漏并因而不能接受调查时，发生覆盖不足。不属于目标人口的单位列入框架而且在抽样前无法将它们与合格单位分开以及在框架包括合格单位的重复时，发生覆盖过度。覆盖误差也可能指不准确或过时的框架信息引起的调查单位在阶层中的错误分类（例如，在一个住户因未占住而被排除在抽样过程外时，当时在进行调查时住户实际上已被占住）。有关调查估计数的覆盖不足的损害性通常比覆盖过度的大。我们没有办法重新找出遗漏的单位，但全域以外的单位经常能在实地工作或数据处理期间认定并适当加以更正或调整；不过，全域以外的单位造成每合格单位的调查成本增加。

15. 进行普查与进行抽样调查相比，覆盖问题经常被认为更加重要，因为在普查中，不用担心出现抽样误差。不过，这是一种错误观念。在有些抽样调查中，覆盖问题有时和抽样误差一样可能是个大问题，如果不是更大的话。例如，抽样调查有时可能将某些难以到达地区或难以检查类别的单位排除在抽样过程之外（从而给它们以零包括概率）。发生的原因可能有访问者安全（例如，调查工作将涉及冲突或暴力高发地区）或调查成本（例如，前往部分地区

访问费用高昂得难以负担或花时间太长)。如果目标人口群的定义不精确说明此类排除情况,造成的调查将导致覆盖不足问题。此类问题可能使估计数发生偏倚,因为被排除在调查人口以外的单位将往往有别于被包括在内的单位。在调查打算覆盖此类难以覆盖的人口时,需要做出特殊规划来确保扩大覆盖范围,以便将这些群体包括在目标人口中,或打算做出推测的人口口。

16. 在电话覆盖率低和文盲率可能高的国家进行的某些重复调查,出现一个相关的问题,其中数据收集必须依赖于面对面的访问。当这些调查的访问期短时,它们的覆盖面往往可能限于易于到达的地区。例如在巴西,月劳动力调查(月劳力调查)只在6个都市地区进行(巴西地理和统计研究所,2002年b)。它的目标人口的限制性定义是这种调查的相关性的主要批评源之一:由于目标人口对于许多使用而言限制性太大,它并不提供关于该国其他地方就业和失业演变的信息。虽然调查正确地报告它同生活在6个都市区的“调查人口”相关的数字,但许多用户错误地解释这6个区之和的数字,好像它们同巴西的整个人口相关。计划重新设计调查,以便在2003-2004年度处理这个问题。类似的问题也出现在其他调查中,例如像1987-1988年度和1995-1996年度巴西收支调查(覆盖面限于11个都市区)和1996-1997年度巴西生活平衡量研究调查(覆盖面仅限于东北和东南区域)。在巴西进行的主要的“国家”年度住户抽样调查在较小的程度上也属于此种情况(巴西地理和统计研究所,2002年a)。由于进入费用高得负担不起,这项调查不覆盖巴西北部区域的农村地区。Bianchini和Albieri(1998年)更详细地讨论了在巴西进行的各种住户调查的方法和覆盖面。

17. 其他发展中国家和转型国家的许多调查遇到了类似的问题,其中,国家一部分难以到达地区的覆盖费用通常太昂贵。关于这个问题遵循的一个重要规则是,基于调查的任何出版物应当清楚说明该项调查有效覆盖的人口情况,并视情况说明排除在外的潜在相关分组。

18. 覆盖误差的度量不经常与调查估计数一起公布,以便允许外部用户在它们的分析中独立评估覆盖问题的影响。这些度量可能只在每隔10年左右公布普查数字时才提供,而且即使在这种情况下,也不将它们与前十年中进行的住户调查的覆盖问题直接联系。

19. 在巴西,进行较全面覆盖面分析的惟一“调查”是普查。这通常通过将查点后抽样调查与人口分析结合起来完成。查点后抽样调查(PES)的主要目的是评估普查或类似调查的覆盖情况,尽管在许多国家应用中,查点后抽样调查也经常用来评价调查内容。在巴西,2000年普查后的查点后抽样调查抽取了大约1 000个查点区的样本,并且由独立的查点员小组检查抽取的样本,这些查点员须与正规普查员遵循相同的程序。在收集查点后抽样调查数据后,进行匹配以对正规普查数据中的对应单位定位。然后利用这种匹配活动的结果适用双重制度估计法[例如见Marks(1973年)],它产生覆盖不足的估计数见下文表十一.2报告的数字。还可利用基于出生和死亡的行政记录的人口存量 and 流量的人口分析来检查普查的人口计数并评估其覆盖程度。在巴西,这种做法只

在南部和东南部区域的某些州富有成效，这里的出生和死亡记录非常准确，足以为此目的提供有用的信息。

20. 将查点后抽样调查普遍适用于普查覆盖面估计和分析的一个严重障碍是它们的高成本。要使这些调查的结果变得可靠，必须仔细规划和执行调查。此外，重要的是它们还提供在某种程度上进行分列的结果，不然，它们的有用性非常有限。在有些情况下，不能获取进行此种调查所需的资源，而在另一些情况下，普查规划者可能认为，这些资源用于改进普查作业本身将更合适。不过，如不测量和检测关键问题在什么地方，进行改进如果不是不可能的话也是困难的。查点后抽样调查帮助确定覆盖问题的关键源头，并能提供关于在今后的普查中数据收集需要加以改进的方面的信息，以及可用来抵消损失的覆盖面的覆盖面不足的估计数。因此，我们强烈建议在编制普查预算和规划期间，为普查数据收集作业后立即进行的合理规模的查点后抽样调查留出所需的资源。覆盖面的人口分析评估一般比查点后抽样调查的费用低廉，但它要求能同时利用外部数据源和人口组成方法的知识。还有，在可能的情况下，还应为进行这种分析编制预算和为其留出时间，作为主要普查评价作业的组成部分。

21. 在不论发达还是不发达的大多数国家，不为覆盖不足调整普查数字。其原因可能是没有广泛接受的理论或方法可用来更正覆盖误差，或者来自查点后抽样调查的覆盖不足估计数的可靠性不够，抑或政治因素阻止变更普查估计数，抑或原因可能是这些因素和其他因素兼而有之。因此，来自普查数据公布的人口估计的覆盖不足基本上未得到弥补。在有些情况下，关于普查覆盖不足的信息（如有）可视为“机密信息”，而且可能不提供给一般用户查阅，其原因是这样一种认识，即如果解释不当，此类信息可能损害普查结果的可信性。我们建议不应采取这种做法，而是应公布查点后抽样调查的结果，或者提供给相关的普查用户。

22. 以上讨论与调查人口的广泛覆盖有关。适当的覆盖评价这一问题对于特别关注的分组人口而言甚至更为严重，例如少数民族或其他少数派，因为查点后抽样调查所需的样本规模一般超出可加利用的预算资源。对于发展中国家普查和其他住户调查中此类分组人口的覆盖率知之甚少。在巴西，自1970年以来进行的每次普查的查点后调查都未能提供民族群体或可能令人关心的其他分组人口的估计数。它们的估计数按大的地域（州）细分，限于住户和个人的总体少计。2000年普查的覆盖不足估计的结果最近才出现（Oliveira等人，2003年）。这里我们只介绍国家一级的结果，包括1991年和2000年普查的住户和个人的遗漏率估计。覆盖不足率1991和2000年相似，2000年的总体比率略低。对于改进巴西普查范围内进行的查点后抽样调查的建议一向是扩大覆盖不足估计的范围，以包括相关的分组人口，如按民族或年龄组界定的那些分组人口。

表十一.2

从1991和2000年查点后调查获得的巴西普查遗漏率的估计 (%)

覆盖类别	1991年普查	2000年普查
私人占住的住户	4.5	4.4
生活在私人占住未遗漏住户中的人	4.0	2.6
私人占住住户遗漏的总人数	8.3	7.9

资料来源：Oliveira等人（2003年）。

23. 表十一.2中的数字高于某些发达国家类似普查所报告的数字。遗漏率显示了一个不能忽略的覆盖不足量。到目前为止，如绝大多数国家一样，巴西公布的普查结果未对估计的覆盖不足做出任何调整。不过，后来对普查后公布的人口预测作了此种调整。需要结合有关要使查点后抽样调查的估计数可用于这一目的时所需的可靠性的讨论、规划和决定进行研究，以评估调整普查覆盖不足估计数的潜在影响。

2. 不答复

24. 术语“不答复”指某些调查单位遗漏（单位不答复）、某些调查单位在一轮或多轮集体或重复调查中遗漏（被动不答复）或甚至就调查单位内某些变量遗漏（项目不答复）的数据。不答复影响每次调查，不论是普查还是抽样。它还可能影响行政来源的用于统计生产的数据。多数调查采用某些作业程序避免或减少不答复的发生。在调查的答复不是“随机答复”（重要的分组人口组中的差别不答复）和答复率低时，不答复就更成问题。如果不答复是随机答复，它的主要影响是样本规模减少引起的调查估计数的方差增大。然而，如果调查参与（答复）取决于答复者和（或）访问者的某些特性和特点，那么偏倚就是需要担心的主要问题，特别是对于不答复率较高的情况而言。

25. Särndal、Swensson和Wretman（1992年，第575页）说：“处理不答复的主要技术是加权调整和估算。加权调整意指增加在估计中适用于答复者y值的权数以补偿由于不答复而损失的值……。估算意指用‘好的’人为值取代遗漏值。”

26. 在三类不答复中，单位不答复是最难补救的类别，因为在调查框架和记录中可用于该目的的信息通常极少。用来抵消单位不答复负面影响最常用的补救方法是加权调整，其中答复单位将它们的权数增加以说明不答复引起的抽样单位的损失；但是甚至这种非常简单类型的补救也不始终适用。波动和项目不答复的补救经常通过估算进行，因为在此类情况下不答复的单位提供的一些信息可用来指导估算并因此减少偏倚（见Kalton，1983年；1986年）。

27. 不答复有各种原因。一种原因可能是未联系上选定的调查单位，联系不上的因素是需要及时进行调查，住户难以查点和答复者不在家。一种原因也可能是拒绝合作以及没有能力答复或参加调查。在发展中国家进行的住户调查

中，因拒绝而造成的不答复率往往低下，这就像通过教育赋予公民权力的工作开展不够一样，主要是因为潜在的答复者不太愿意和能够拒绝与调查合作；而且较高的文盲率意味着多数数据收集仍利用面对面的访问进行，而不是采用电话访问或邮寄调查问卷。这两个因素都有助于降低拒绝或不合作的比率，而且这两个因素还都可能导致调查内部的差别不答复，受教育较多和较富的人更倾向于变成不答复者。与此同时，答复或参与调查并不一定导致报告的准确性提高；在许多情况下，较高的答复率实际上可能掩盖某些种类数据的故意误报，特别是与月收入或财富相关的变量，原因是政府对官员不信任。

28. 发展中国家的普查受不答复影响。在巴西，普查使用两类调查问卷：一种是短表，关于人口项目只有几个问题（性别、年龄、同户主关系和识字情况），另一种是较大和较详尽的表，列有社会经济项目（种族、宗教、教育、劳动力、收入、生育率、死亡率等），并且还包括短表上的所有问题。长表用于通过每个查点区的住户概率抽样选择的住户。小城市抽样率较高（五分之一）而普查年估计人口达到或超过1.5万的城市较低（十分之一）。普查中总的单位不答复率非常低（巴西2000年普查约为0.8%）。不过，对于短表上的变量（要求所有参加住户答复的那些变量，叫作全域组），未为不答复做出补救。这有三个原因：首先，不答复率被认为相当低；其次，关于不答复住户的信息非常少因而不能使补救方法变得有效；第三，在普查背景下没有天然的框架可用于进行加权调整。由于头两个原因，以及还由于对在像这样的情况下进行的估算所增大的偏见，通过某种类型的供体法估算遗漏的普查表的替代办法也不非常受人欢迎，对于从普查内的样本得到的估计数，进行基于校准法的加权调整，以便部分补救单位不答复。

29. 在有些抽样调查中采用了一种类似的办法。巴西的主要住户调查中的两个——年度全国住户抽样调查（住户抽样调查）和月劳动力调查（月劳力调查）——使用不具体的不答复补救法（见Bianchini和Albieri, 1998年）。对答复单位权数的惟一调整通过与大都市地区或国家级总人口的校准来进行，因此它们不能补救例如按性别和年龄界定的人口组内的差别不答复。其原因大都同作业考虑相关，例如很早以前开发的用于估计的专用软件的维护和所感受的忽略不答复的简单性。两项调查都记录它们的不答复水平，但在载有主要调查结果的出版物中未发布关于这个问题的信息。不过，提供了可以导出不答复估计数的微数据文件，因为来自不答复单位的记录也包括在此类文件中，并有适当的代码确认不答复的原因。月劳力调查最近作了重新设计（巴西地理和统计研究所，2002年b），并开始至少利用一种简单的再加权法以补救发现的单位不答复。进一步的发展可能包括引入校准估计量，它们将尝试更正关于年龄和性别的差别不答复。不过，相关的研究尚处于早期阶段，其动机是这样的观察，即不答复或许是失业率月估计数中轮流组偏倚的原因之一（Pfeffermann、Silva和Freitas, 2000年）。

30. 利用较先进的不答复调整法的一项巴西调查是住户支出调查（支出调查）（1995-1996年度最后一轮，2002-2003年度一轮正在进行）。本

项调查利用再加权法和估算法的组合来补救不答复（Bianchini和Albieri，1998年）。进行了权数调整以补救单位不答复，而供体估算法被用来填补在数据收集和编辑处理后答案仍遗漏的变量或变量块。更加注意处理不答复问题的动机是本项调查中发现的不答复率，较之一般住房调查的为大。鉴于调查类型造成的答复负担大得多（访问住户至少两次，而且它们被要求在两周期间对费用做出详细的记录），预期不答复率较大。调查方法的报告包括对不答复的分析，但介绍主要结果的出版物未这样做。

31. 然而在巴西进行的另一项调查即生活水平衡量调查（水平调查）——是世界银行生活水平衡量调查方案的组成部分——利用住户的替代以补救单位不答复。在巴西，这种做法很少使用，其他主要住户调查都没有采用这种做法。

32. 在对同一国内进行的各种调查进行审查后，出现的格局是没有一种标准办法补救和报告单位不答复。不答复的方法和处理在各次调查间互不相同，它成为所经历的不答复程度、调查坚持国际建议的情况和实施补救方法和程序的所感受的需要和能力的一个函数。可用来改进这种情况的一个做法是定期编写住户调查的“质量概况”报告。与试图将有关使用的方法和数据局限性的所有可加利用的信息都包括在基本普查或调查出版物中相比，这种做法通常可能更为切实可行和有用。

33. 关于项目不答复，情况并无很大的不同。在从1980年开始的巴西普查中，估算法被用来填补空白，并且还用来取代利用主题专家规定的编辑规则检测到的不一致的值。在1991和2000年，在像DIA（数据自动检测和估算）（Garcia Rubio和Criado，1990年）和NIM（新估算法）（Poirier、Bankier和Lachance，2001年）等软件中执行的供体法与Fellegi-Holt法的组合，被用来执行普查长短表的综合编辑和估算。2000年，除了分类变量的估算外，还借助于回归树法执行收入变量的估算，该方法用来发现供体记录，然后再利用从中发现的收入值填补不完整记录中遗漏的收入项。巴西普查中，在处理结束时微数据文件中所有普查记录都没有遗漏的值，这是第一次。尽管关于编辑和估算量的大部分信息只在专门报告中提供，但普查的编辑和估算战略有详尽记录。一项建议提出，使这些报告更易于查阅的办法是利用互联网传播它们。

34. 其他住户调查中遗漏和可疑数据的处理进展得不够好。在住户抽样调查和月劳力调查中，都利用了计算机程序检测误差，但仍有大量的“人工编辑”，而且极少利用计算机辅助估算法补救项目不答复。如果在编辑阶段结束时项目遗漏，则将它们编码为“不明”。近年来取得的进步专注于整合编辑步骤与数据输入，以便减少处理费用和时间。由于更便宜和更便携的计算机的出现，使得地理统计所能够继续进行进一步的整合。2000年这个十年的修订的月劳力调查于2001年10月开始收集一个平行的样本，其含量与经常调查中使用的相同，其中数据利用计算机辅助（掌上）的面对面访问获取。关于掌上电脑的成效尚无最后报告，但在头几个月后，据报告数据收集工作进展顺利。这种技术使得调查管理者能够专注于源头的质量改进，其途径是将所有的转移指令和

有效性检验嵌入数据收集仪器，从而在源头避免键入错误和其他错误。如在普查中那样，收入的不答复将利用回归树法找到供体来补救。不过，这种新调查的结果只是到最近才提供，而且数据收集与老系列并行了一整年，然后它们才被发布而且新系列取代老系列。目前正在对新的数据收集和处理办法进行更广泛和更详细的评估。

35. 在月劳力调查中，将每个住户在样本中保持两个时期，每期待个月，中间相隔8个月。因此，在原则上，每当一个住户或住户成员在第一回合后的任何调查回合中遗漏，都可利用来自先前完整访问的数据补救波动不答复。数据的这种利用在老系列中未出现，也不打算用于新系列，尽管它代表着可由调查管理者加以考虑的一种改进。

36. 从项目不答复和不一致或可疑数据编辑和估算做法的跨调查分析出现的格局是一种非标准化格局，不同的调查沿着不同的方法路径进行。在普查中显然大规模地应用自动编辑和估算法，而较小的调查并不如此经常地采用类似的方法。从下述意义上说，也许存在着一种调查规模效应，即投资于开发和应用可接受的方法和程序进行自动估算对于普查来说是有正当理由的，但对于较小的调查就不是这样，后者出结果的时间也较短。对于像巴西月劳力调查这样的重复调查，虽然出结果的时间短，但对数据编辑和估算的方法作较大的投资也许可以获得收益，因为有可能在许多连续的调查回合中可以利用这种投资。

3. 测量和处理误差

37. 测量和处理误差需要调查问题和变量在数据收集和处理后的观察值，它们不同于如果使用理想或金本位的测量和处理方法将会得到的对应真值。

38. 从发展中国家和转型国家进行的住户调查和测量、补救和报告的角度看，这个专题或许是受注意最少的专题。若干现代发展可被视为朝着改进调查做法迈进以减少测量误差。首先，数据收集的计算机和辅助方法的使用有助于减少转录误差，因为答复者的答案被直接输入计算机并可立即来编辑和分析。还有，问题的流动受计算机控制并可使之依赖于答案，防止访问者引入错误。答复可根据预计的范围检查和甚至根据同一答复者先前的回答进行检查。可以标出可疑或令人意外的数据并要求访问者探查答复者。因此，原则上，可以获得质量较高和少出测量误差的数据。不过，除了减少项目遗漏值率和超范围值率外，没有多少证据表明计算机辅助访问比纸笔访问有任何质量优势。

39. 另一类进步涉及开发和应用通用软件进行数据编辑和估算（Criado和Cabria，1990年）。如B节中已经提及的那样，普查采用了自动编辑和估算软件来检测和补救测量误差和某些类型的处理误差（例如，编码和键入错误），同时还有项目不答复。某些抽样调查也这样做了。不过，在这种办法范围内适用的补救类型只能处理所谓的随机误差。利用标准编辑软件很少检测到系统误差。

40. 可能导致调查中减少处理误差的另一类发展是开发了计算机辅助编辑软件，以及数据捕获设备和软件。

41. 虽然测量和处理误差的预防可能取得了一些进步，但测量、最终补救和报告测量误差的方法的应用却并不如此。关于测量误差的实践大多专注于预防，而且在这方面做了被认为重要的事情后，并不很多地注意评估调查规划和执行工作取得了多大的成功。由于缺乏标准的测量指导理论，使得制定质量目标和评估此类目标实现成为一项艰巨的任务。例如，虽然我们确实看到调查抽样计划中界定了样本规模，其目标是使某些关键估计数的变动系数（相对标准误差）低于事先提出的规定值，但我们很少见到调查收集和处理计划的目标是使项目估算水平低于规定的水平，或者旨在使观察到的度量保持在高概率的对应“真值”的订定容差（即最大偏差）范围之内。期望所有类型非抽样误差的现实的定量目标都能事先制定也许是不切实际的；不过，我们主张调查组织至少应当做出努力，以测量非抽样误差并利用此种度量确定今后改进的目标和监测这些目标的实现情况。

C. 挑战和展望

42. （抽样）调查作为社会科学的一个关键观察工具经过50多年的广泛传播后，抽样误差及其控制、测量和解释的概念已达到了某种成熟程度，尽管我们已经指出，世界各地许多调查的结果在公布时未包括任何抽样误差估计数。非抽样误差方面取得的进展少得多，至少就在发展中国家进行的调查而言是这样。这种情况并非偶然。调查中非抽样误差是一个难题。一方面，它们来自调查中的许多源头。抵消一类误差的努力经常造成另一种误差的增大。预防方法不仅取决于技术，而且依赖于文化和环境，使得成功的经验难以普及和推广。补救方法通常很复杂，而且恰当执行的费用很昂贵。在预算非常有限时进行调查的背景下，测量和评估难以进行，而且公布结果的限期定得越来越紧，以满足信息饥渴社会日益增长的需求。在与此相似背景下，正确的做法始终应当是优先预防而不是测量和补救，但是这给评估预防努力取得多大成功没有留下多少空间，因而使未来改进的前景变得暗淡。

43. 对统计事务缺乏了解的有些用户可能误解关于调查中非抽样误差的报告。因此，在有些调查环境下，此类报告的公布有时候被视为是不可取，主要原因是统计知识和文化不发达，在识字和识数较普遍缺乏的人口中，它们的发展可能特别艰巨，许多发展中国家的情况就是这样。同样经常出现的情况是在制作机构中也缺乏统计专长，导致难以认识问题和采取主动行动抵消它们，以及测量此种行动取得多大成功。无论如何，我们鼓励编写和公布此类报告，统计机构应力争使报告尽量清楚明了，供识字成人查阅。

44. 即使情况不是很好，但有些新的发展令人鼓舞。最近，若干主要统计机构、统计和调查学术协会以及多边政府组织给予数据质量课题以关注，这是一个令人高兴的发展。我们在这里应当提及的主要倡议是国际货币基金组织（货币基金组织）的通用数据传播系统（GDSS系统）和专用数据传播标准

(SDDS标准)，它们正试图借助于各国自愿坚持其中任一倡议以促进关于统计数据质量的报告的标准化。据货币基金组织(2001年)说：“GDDS系统是一个结构化过程，基金组织成员国通过它长期自愿致力于改进它们的统计系统生产和传播的数据的质量，以满足宏观经济分析的需要。”另据货币基金组织说：“GDDS系统提供经济、财政和社会人口统计资料汇编和传播两方面的合理统计做法。它确定对社会及人口发展的经济分析和监测特别相关的数据集，并制定有关它们发展、生产和传播的目标和建议。特别注意用户的需要，通过有关数据的质量和完整性准则，以及公众对数据的利用来满足这种需要”(同上)。

45. 这些倡议的主要贡献是向各国提供：(a) 数据质量框架(见<http://dsbb.imf.org/dqrsindex.htm>)，可协助确定关键问题领域和改进数据质量的目标；(b) 在广泛的调查和统计产出的范围内考虑改进数据质量的经济奖励(采用的形式是延长或获得准入国际资本市场的机会)；(c) 具有共同兴趣的社区，社区由此能够推进数据质量问题的讨论而不必担心受到误解；和(d) 在需要时对评价和改进方案的技术支持。这不是一个普遍的倡议，因为不是每个国家都是货币基金组织的成员国。不过，就此事同131个国家进行了接触，而且截至目前已有46个国家决定坚持使用GDDS系统，而且另有50个国家已取得SDDS标准的较高认购地位，并符合关于其统计产出质量评估的一套较为严格的控制手段和标准。

46. 对货币基金组织或其他组织倡导的数据质量标准进行的详细讨论不属于本章的范围，但鼓励读者使用这里列示的参考资料探究这个问题。发展中国家应当参加讨论目前出台的标准，决定是否尝试坚持上述任一倡议，并在相关的情况下为标准的界定和修订做出贡献。最重要的是，发展中国家的统计机构可将这些标准作为出发点(如果本地没有可供利用的类似规定)，以促进它们的成员和工作人员以及在它们的用户界提高质量意识。

47. 我们应在这里提到的另一项倡议，特别是因为它对巴西和其他拉丁美洲国家产生了影响，是欧洲联盟(欧盟)和南方共同市场的统计合作项目。²⁵根据项目目标：“欧洲联盟和南方共同市场国家签署了一项关于‘与南方共同市场国家开展统计合作’的协定，其主要目的是统计方法的修好，²⁶以便能够使用基于共同接受的条件的各种统计数据，特别是有关贸易货物和服务以及一般须进行统计测量的任何领域的那些数据。”该项目“预期同时实现南方共同市场国家内部及它们与欧洲联盟之间统计方法的标准化”。(要了解更多的详情，访问网址：<http://www.ibge.gov.br/mercosur/english/index.html>。)这一项目已经促进了若干课程和培训研讨会，并在这样做的过程中，为改进调查做法和提高对调查误差及其测量的认识做出贡献。

48. 这些倡议对于支持发展中国家的统计机构改善它们的状况十分必要：它们的统计资料可能质量很好，但它们往往不了解好到何种程度。发达国家提供给发展中国家以及发展中国家相互之间的国际合作，对于逐步更好地测量和报告非抽样调查误差和调查数据质量的其他方面至关重要。

²⁵ 南方共同市场是南部共同市场，即共同签订了一项自由贸易协定的一组国家，包括巴西、阿根廷、巴拉圭和乌拉圭。

²⁶ 该术语在这里的意思是协调统一。

D. 有关进一步阅读的建议

49. 建议作为进一步阅读的主题的会议包括：

- 调查测量误差国际会议，1990年在亚利桑那州塔克森举行（见Biemer等人，1991年）；
- 调查测量和过程质量国际会议，1995年在联合王国布里斯托尔举行（见Lyberg等人，1997年）；
- 调查不答复问题国际会议，1999年在俄勒冈州波特兰举行（见Groves等人，2001年）；
- 官方统计质量问题国际会议，2001年在瑞典斯德哥尔摩举行（访问<http://www.q2001.scb.se/>）；
- 加拿大统计局2001年专题讨论会，在加拿大渥太华举行，重点是从方法角度实现统计机构的数据质量（访问<http://www.statcan.ca/english/conferences/symposium2001/session21/s21c.pdf>）；
- 国际统计学会（统计学会）第五十三届会议，2001年在大韩民国汉城举行，本届会议举行了一次关于“统计机构质量方案”的应邀论文会议，论述国家和国际统计局处理数据质量的方法（访问：<http://www.nso.go.kr/isi2001>）；
- 2000年统计质量研讨会，由货币基金组织赞助，2000年在大韩民国济州举行（访问<http://www.nso.go.kr/sqs2000/sqs12.htm>）；
- 改进调查国际会议，2002年在丹麦哥本哈根举行（访问：<http://www.icis.dk/>）。

参考资料

- Bianchini, Z.M. 和 S. Albieri (1998年)。“针对巴西使用的主要住户样本的审查”。《经济与社会发展统计国际会议的记录》。墨西哥，阿瓜斯卡连特斯，1998年：Instituto Nacional de Estadística, Geografía e Informática (INEGI)。
- Biemer, P. P. 和 R. S. Fecso (1995年)。《评估和控制商业调查中的测量误差》，Cox和其他人等。《商业调查方法》，纽约：约翰·威利国际出版公司。
- Biemer, P. P. 和其他人 (1991年)。《调查中的测量误差》。纽约：约翰·威利国际出版公司。
- Correa, S. T.、P. L. do Nascimento Silva 和 M. P. S. Freitas (2002年)。Estimação de variância para o estimador da diferença entre duas taxas na pesquisa mensal de emprego. In 15° Simpósio Nacional de Probabilidade e Estatística. Aguas de Lindóia, Brazil, São Paulo, Brazil: Associação Brasileira de Estatística。

- Criado, I. V. 和 M. S. B. Cabria (1990年)。Procedimiento de depuración de datos estadísticos, cuaderno 20. Vitoria-Gasteiz, Spain: EUSTAT Instituto Vasco de Estadística。
- GarciaRubio, E. 和 I. V. Criado (1990年)。“DIA系统: 质量数据的自动填补软件”。《美国人口普查局第六次年度研究会议的记录》(维吉尼亚, 阿灵顿)。哥伦比亚特区华盛顿: 美国人口普查局。
- Groves, R. M. 和其他人 (2001年)。《不答复调查》。纽约: 约翰·威利国际出版公司。
- 巴西地理统计局 (2002年a)。http://www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad99/metodologia99.shtm。
- _____ (2002年b)。http://www.ibge.net/home/estatistica/indicadores/trabalhoerendimento/pme/default.shtm。
- 国际货币基金组织 (2001年)。“数据公布通用系统指南”。哥伦比亚特区华盛顿: 基金组织统计学部。请登陆: http://dsbb/imf/org/applications/web/gdds/gddsguidelangs)。
- Kalton, G. (1983年)。《弥补丢失的调查数据》。《研究报告汇编》。密歇根, 安阿伯: 密歇根大学, 社会学研究所。
- _____ (1986年)。“处理小组调查中的间歇式不答复情况”。《官方统计杂志》, 第2期, 第3号, 第303-314页。
- _____ (2002年)。“抽样调查实践中的模型(修订)”。《官方统计杂志》, 第18期, 第2号, 第129-154页。
- Kordos, J. (2002年)。《个人通讯》。
- Lyberg, L. 和其他人等 (1997年)。《调查测量和程序质量》。纽约: 约翰·威利国际出版公司。
- Marks, E. S. (1973年)。“双重制度估计在人口普查评估中的作用”。《内部报告》。哥伦比亚特区华盛顿: 美国人口普查局。
- Oliveira, L. C. 和其他人 (2003年)。Censo Demográfico 2000: Resultados da Pesquisa de Avaliação da Cobertura da Coleta. Textos para Discussão, No. 9. Rio de Janeiro: IBGE, Directoria de Pesquisas。
- Pfeffermann, D.、P. L. Nascimento de Silva 和 M. P. S. Freitas (2000年)。“巴西劳动力轮调计划对公布的估计值质量的影响”。《内部报告》。里约热内卢: 巴西地理统计局, 方法部。
- Platek, R. 和 C. E. Särndal (2001年)。“是否能提供一名统计员?” 《官方统计杂志》, 第17期, 第1号, 第1-20页。
- Poirier, P.、M. Bankier 和 M. Lachance (2001年)。“加拿大人口普查编辑和修正系统内的有效方法”。统计联席会议上提交的文件, 美国统计学会。
- Särndal, C. E.、B. Swensson 和 J. Wretman (1992年)。《调查取样的模型帮助》。纽约: 施普林格出版社。
- 联合国 (1982年)。《国家住户调查能力方案: 住户调查中的不答复误差: 根源、评估和控制: 暂定版本》。DP/UN/INT-81-041/2。纽约: 技术合作发展部和统计办公室。

D 部 分

调查成本

导 言

James Lepkowski

密歇根大学
美国密歇根州安阿伯

1. 前面几节中审议了住户调查中产生的抽样和非抽样误差，以便更好地了解调查估计数的质量。在几乎所有类型的此类误差中，都能利用各种方法缩小误差的规模。不过，这些方法的执行经常需要增加成本。由于调查确定了支付费用的预算，投放追加的资源减少一种误差源意味着将资源从一个领域转到另一个程序。调查设计牵涉不断地权衡成本和调查误差。

2. 例如，假定在一次特定的住户调查中，有一个分组人口说调查问卷没有译本的语言。调查设计人员起初可能决定将此群体排除在调查范围之外，从而产生覆盖问题。或者，他们可能决定缩小样本规模以减少调查费用，然后利用节约的成本将调查问卷译成一种新的语言，雇用讲这种语言的访问者，将这些住户重新纳入调查范围。

3. 鉴于调查设计经常是一系列的此种权衡，为了做出稳妥的决定，应提供好的信息，说明产生于不同来源的误差的性质和规模（例如前例中的抽样方差和未覆盖偏倚）以及与不同调查程序有关的成本。前几节审议了误差源和误差的大小。本节将审议调查成本的性质。

4. 调查中的成本因素产生于三个层面。第一个在调查的规划阶段，此时必须事先估计成本。规划或“预算”阶段的成本概算是难以获得的，除非有事先的经验可资利用。持续进行调查作业能够为规划新一轮调查提供相关的成本数据，尽管下一层面的成本考虑因素——调查成本的监测——经常起干扰作用。

5. 调查组织或甚至偶而进行调查的其他组织，很少拥有完善的系统，以便采取能够使成本数据可用于规划的方式跟踪成本。成本集合在一个核算系统中，但是这些系统并不将成本分为调查设计者需要用于规则目的的那种类别。在尝试进行此种成本监测的情况下，如果必须将新的系统加到作业上，可能增加调查本身的成本。

6. 如果在进行着的作业中监测成本，就有可能在数据收集期间更系统地考虑调查设计的变化。可将成本信息用于预测一次作业中节约的量以及将资源再分配到另一领域的影响可能有多大。

7. 调查规划中资源的再分配通过对成本水平与分布在多个误差源中的误差之间的权衡进行考虑来加以确定。样本设计发展是一个能够和正式做出这些权衡以便为解决资源分配问题找到最佳方案的领域。

8. 例如，如在第二章中所讨论的那样，基于在一个来自广泛分布的人口地区概率样本中提取的群集的调查，必须考虑限制群集的数量，以降低数据收集成本。不过，限制群集数目意味着在每个样本群集中所作观察的数目必须增加，以便保持总的样本规模。然而，每个群集中子样本规模的增大提高样本估计数的可变性。换句话说，由于通过提取较少的群集而成本下降，抽样方差增大。所需的是指导选择多少群集，以便在实现规定的精确水平的前提下能够将成本降到最低限度，或在成本给定前提下尽量缩小抽样方差。在样本设计中，这个问题有一种数学解法。

9. 成本误差间的权衡还产生于调查设计的其他方面。例如，在住户调查中减少住户不答复的一个方法是对靠单次访问未得到答复的住户访问一次以上。可指示访问者在调查数据收集期间对住户访问多至四五次以获取答案。对某些样本住户反复访问减少可包括在样本中的样本住户的数目。反复访问以减少住户不答复的成本可限制样本规模。因此，为减少不答复偏倚，加大了在减少不答复方面所做的努力，这方面的成本可增大抽样方差。在一个领域为降低成本所做的努力再次要求重新分配资源，并且引起调查设计另一领域误差增大的潜在可能性。

10. 本节各章审议围绕调查设计中成本的规划、监测和再分配出现的若干问题。它们使用发展中国家和转型国家住户调查的数据来说明调查数据收集发生的成本类型和在某种程度上成本的大小。由于各国之间的调查工作差异甚大，而且各洲之间差异甚至更大，所提供的具体的成本信息对某一特定国家的调查规划不一定有用。不过，希望下列各章介绍的成本源和成本水平有助于不同背景下的调查设计人员在各自的调查当中更加充分地理解调查成本和成本误差权衡。

第十二章

发展中国家和转型国家调查成本问题的分析

Ibrahim S. Yansaneh*

国际公务员制度委员会
联合国，纽约

* 联合国统计司方法和分析股前任股长。

摘 要

本章概括讨论同发展中国家和转型国家设计和实施住户调查的成本相关的关键问题。将一次调查的总成本分解为同它的设计和实施的各方面相联系的较为详尽的各个组成部分。对于拥有广泛的调查基础设施的国家和很少拥有或没有调查基础设施的国家，将分别考虑成本因素。另外还审议各国之间的成本可比性问题。

关键词：调查基础设施、每次访问的增量成本、效率、成本可比性、成本因素。

A. 导言

1. 有效抽样设计标准

1. 一般来说，有效的抽样设计必须满足两项标准之一：它必须在固定预算的制约下提供合理精确的估计数，或尽量减少规定精确水平的执行成本。本章重点放在第一项标准，涉及的任务是制订最有效的设计方案，并且能够以与可获预算相一致的成本实施该方案以及对资源进行合理有效的利用。在发展中国家和转型国家，拟订关于设计和实施的重要决定时，最大的一个制约因素是调查成本。与在发达国家一样，发展中国家和转型国家设计一项调查时涉及调查估计数精确性与实施成本间通常的权衡。所选人口数量估算量的方差是关心的主要方面，精确性一般从这个方面加以度量。精确性的其他相关度量包括均方误差或总调查误差，它还纳入误差的偏倚组成部分。

2. 在典型的情况下，精确性与成本间权衡的形式数学展开涉及优化受到相对简单制约因素制约的良性方差或成本函数。不过，由于提供的成本和方差信息有限，通常应当将这种优化做法视为只对优先设计方案或只为在实施中真正实现的精确性和成本值提供粗略近似值。针对在发达国家开展的调

查，现已对这些问题进行了深入审议。例如见Andersen、Kasper和Frankel（1979年）、Cochran（1977年）、Groves（1989年）、Kish（1965年；1976年）以及Linacre和Trewin（1993年）的著作和其中引用的参考资料。此外，关于成本和精确性作为评价国家统计局系统许多标准中的两项标准，更广泛的讨论见de Vries的著作（1999年，第70页）及其中引证的参考资料。关于发展中国家和转型国家所选调查成本的经验分析，以及发展中国家和转型国家调查设计中成本/误差权衡的更为详尽的讨论，见第十三和十四章及D节导言（调查成本）。

3. 发展中国家和转型国家调查设计的一个主要制约因素是缺乏关于同调查实施各方面相联系的成本的信息或此种信息不充分。尽管存在上述制约因素，人们经常发现不同调查的成本中有一定数量的共同结构可用于新调查的设计。在有些情况下，这种共同结构限于若干成本组成或来源的相对规模的定性指标。在另一些情况下，提供的实际成本可被视为在一组国家，特别是人口分布和调查基础设施水平类似的国家具有相当的同质性。

4. 本章在发展中国家和转型国家的调查范围内分析成本问题，并调查一个国家的调查成本或相关组成部分可用来改进另一国类似调查设计的程度。换句话说，本章试图论述调查成本在各国间可否移植的问题。此种分析的用途有两个：首先，它有潜力提供部分解决办法，用来解决发展中国家和转型国家调查成本信息稀缺这一问题。其次，在各国之间就样本设计、调查基础设施和人口分布具有类似性的程度上，可以预料这些国家至少有部分调查成本具有类似性。此种成本信息社会可从一国的某项调查析取，并用来在不同的国家设计新的调查，或提高同一国同一项调查的设计效率。在这样做的过程中，调查设计人员必须认识到各国间调查成本结构广泛的可变性。一般情况下，可变成本组成部分是特定国家特有的，而有些固定成本在各国之间可加比较。

2. 发展中国家和转型国家调查的成本结构组成部分

5. 在本章，我们的侧重点放在有效调查设计的第一项标准上，也即就给定预算拨款产生合理精确的调查估计数的设计。发展中国家和转型国家进行的许多调查由国际金融和发展机构委托进行，这些机构需要数据用于关于发展援助项目的决策或者为决策者或受益国的决策者提供支助。发展中国家的调查的3个突出例子是人口与健康调查（DSH），由ORC Macro代为美国国际开发署进行；生活平衡量研究（生活平衡量研究）调查，由世界银行进行；以及多指标类集调查（MICS），由联合国儿童基金会（儿童基金会）进行。此外，许多其他的调查由国家统计局和国家统计系统内的其他机构定期进行。还有大量的较小规模的调查由捐助方委托并由当地小型组织（例如非政府组织）进行。不用说，在这些调查的设计工作方面，成本问题至关重要。

6. 在处理成本问题时，重要的是应当认识到，发展中国家的调查设计具有许多共同特点。例如，多数调查基于多阶段分层区域概率设计。初级抽样单位（PSUs）经常根据前述国家普查中确定和使用的查点区域构建。次级抽样单

位一般是居住单位或住户，而且最终的抽样单位通常是住户或个人。分层域和分析域一般由行政区域和这些区域的城市/农村子域的交集形成。由于这些类似点以及按照上文第2段提及的文献，重要的是应研究在发展中国家调查组内人们能够在多大程度上确定共同的成本结构。关于在发展中国家和转型国家进行的调查，其设计和实施的某些一般背景见本出版物第一部分A节（设计节）和第2部分的案例研究。关于发展中国家某项具体调查成本组成部分更详尽的处理，见第十三章。所选发展中国家和转型国家进行的调查的成本组成部分的经验比较见第十四章。

7. 在本章中，我们应将注意力限于国家统计局或国家统计系统内其他政府机构进行的主要国家住户调查。这些包括住户预算调查、收支调查以及人口与健康调查。即使市场调查和各种组织临时进行的其他较小规模的调查提供有用的信息源并用于国家政策决定和发展计划，也不将它们列入本次讨论的范围。不过，讨论中提出的关键问题也适用于这些类型的调查。大多数例子基于人口与健康调查以及生活水平衡量研究调查，但关键问题可广泛适用于所有的住户调查。

3. 本章概述

8. 本章的结构如下：B节讨论一项调查的总成本典型地分解为较详细的组成部分的问题。后面的3节定性地说明影响发展中国家和转型国家进行的调查的总成本的某些因素。C节审查对大量调查基础设施已到位的情况可能具有重要性的成本因素。D节考虑事先的调查基础设施有限或没有此类设施的情况。E节讨论成本结构可能因调查目标修改而发生的变化。F节提供关于报告的调查成本的解释的某些相关告诫评论。G节提供一些结论性评论和讨论中未充分展开的某些要点。如Ajayi提供（2002年）的那样，附件给出的例子是一个框架用于在发展中国家和转型国家进行的儿童基金会多指标类集调查（MICS）的预算编制。

B. 调查成本的组成部分

9. 调查成本的数学基础一般假定总成本C为所选初级抽样单位和选定要素数目的一个线性函数。此种函数的一个例子是

$$C = c_0 + \sum_{h=1}^L n_h c_h + \sum_{h=1}^L \sum_{i=1}^{n_h} n_{hi} c_{hi} \quad (1)$$

其中， c_0 代表发起调查的固定成本； c_h 等于从在分层h内的一个追加初级抽样单位（PSU）收集信息的增量成本； n_h 是抽取的初级抽样单位数； c_{hi} 等于访问分层h内初级抽样单位i内部一个追加住户的增量成本；以及 n_{hi} 是初级抽样单位i中被抽中住户的数量。例如见Cochran（1977年，第5.5和11.13-11.14节）和Groves（1989年，第2章）。一般来说，成本系数 c_0 、 c_h 和 c_{hi} 将取决于在各国间和在国家内部不同调查间可能互不相同的大量因素。这些因素将在下文各节进行详细讨论。

10. 请注意表达式(1)是可加考虑的许多可能的成本函数之一。例如Cochran (1977年, 第313页) 讨论将一个同列举次级抽样单位(作为子抽样住户访问前的一个中间阶段)相联系的单列成本组成部分包括在选定初级单位内, 其中, 该组成部分取决于每个初级单位内的次级单位数。还有, 对于一个三阶段设计, 即个人从住户内随机被选为访问对象的设计而言, 上文(1)中将有一个额外的项, 表示与访问选定住户内追加人数相联系的增量成本。

11. 此外, 一个更为现实的成本函数常常是一个逐步函数而不是线性函数。例如, 如果10次访问能在一天内进行, 那么额外的第十一次访问就要求额外工作一天并因此需要大量的成本, 而增加的第十二次访问可能只使总成本增加少许。另外, 重要的是应指出, 关于诸如样本初级抽样单位等问题的决定有时受实际考虑因素而不是成本和精确性考虑因素影响。例如, 可能的情况是有人想花一整周在一个初级抽样单位内访问调查。在这种情况下, 工作量不到一周将不可行, 尽管相当于两周工作的双倍工作量也许是可能的。因此, 在此种情况下, 样本初级抽样单位将不由成本和设计效果的考虑因素直接决定, 而是由实施方面的实际制约因素决定。

12. 我们在下节讨论取决于有关国家调查基础设施水平的调查成本。该节的核心信息是, 在拥有大量调查基础设施的国家与基础设施很少或者没有基础设施的国家之间, 调查的总成本差幅很大。不过, 必须记住, 在发展中国家和转型国家, 人们将不得不在调查的规划阶段评估基础设施的程度, 而不是依赖于历史记录。在某个时间拥有一流调查基础设施的国家, 随着时间的推移, 它的基础设施逐步减少到使该国从第一组国家(在C节中考虑)转到第二组国家(在D节中考虑)的程度并非不常见。

C. 拥有广泛基础设施可加利用的调查的成本

1. 同筹备活动相关的因素

13. 一次性调查的大量成本是为筹备活动供资发生的成本[例如见Grosh和Muñoz (1996年, 第199页)], 因此, 此类活动的资金在调查过程的初期支出。成本相对固定的筹备活动包括多个政府机构对调查规划的协调、框架发展、样本设计、调查问卷设计、调查问卷和其他调查材料的印刷以及针对潜在答复者的宣传。取决于样本规模(初级单位或住户一级)的筹备活动成本包括征聘和培训实地工作人员(例如, 制表人、访问调查者、监管员和译员)。

14. 筹备活动的成本取决于各种当地因素, 如调查人员的多少和报酬率、设备的类型和数量、各种项目如文具和其他用品的价格以及交通和通信方式等。此外, 成本还受下述因素的严重影响, 即调查是首次进行的断面研究(单位成本相对较高), 还是一项连续调查的组成部分(单位成本较低)。

2. 同数据收集和处理相关的因素

15. 数据收集和处理的成本也涉及固定和可变的两个组成部分; 但就多数而言, 数据收集的成本是可变的, 也就是说, 取决于选定的初级抽样单位和住

户数。这些成本包括列出选定初级单位内的住户或列出选定住户内的个人、访问和实地监督等的费用。数据收集成本还包括初级抽样单位之间及其内部的差旅费。这些数据收集成本取决于访问作业的组织安排、调查问卷的长度、是否使用口译人员和接受访问的单位数。

16. 减少差旅费的一个选项是建立由监督员和访问员组成的国家调查组，并在各区域间调用该小组，而不是建立区域小组。重要的是应指出这个选项还提高了数据的质量。这种做法也能用于这样的情况：数据收集滚动进行，或调查作业牵涉使用昂贵的设备。多个调查工作组的模型已被用于发展中国家和转型国家的许多调查中，例如生活水平衡量研究系列（Grosh和Muñoz，1996年，第5章）。在区域间语言有变化的发展中国家和转型国家，更有效的做法可能是根据对每个区域所讲语言的熟练程度建立调查组。

17. 数据收集和处理费用的很大部分同实地活动的协调和调查材料的费用相关。在集中的数据收集和处理系统中，与检索填好的调查问卷并将它们发往总部联系在一起的费用可能是巨大的。此外，预算还必须考虑到与监测调查活动和结果，例如在实地开展的列表和子抽样程序、参照预定水平而言对关键关心领域的答复率等有关的潜在高昂费用。此类活动的有效监测使得调查执行人员能在数据收集期间视需要采取纠正措施，而不是等到数据收集后再发现不足，到那时再作补救所需的费用高得可能承担不起。

18. 作为数据处理的组成部分，数据录入、编辑和估算工作可能涉及固定和可变成本混在一起，这取决于用于这一过程的自动化程度。经证明，数据处理的其他主要成本是固定的，而且包括计算设备和软件的成本，以及发展权数和方差估算量及其他数据分析工作。例如，权数的计算与抽样的初级抽样单位或住户数无关；而且在加权程序开发和编程后，计算一个追加住户的权数的增量成本可忽略不计。

19. 数据处理的成本取决于预算包括多少层面的分析。对于某些调查而言，仅以表的形式对收集的数据进行初步分析。对于像人口与健康调查以及生活水平衡量研究调查这样的其他调查，进行了更详尽的统计分析，作为对受益国政府和捐助机构的政策建议的基础。例如，人口与健康调查和生活水平衡量研究调查对它们的调查微观数据进行各种类型的详尽分析，并在一系列分析和方法报告（在人口与健康调查情况下）以及工作文件（在生活水平衡量研究调查情况下）中公布各自的调查结论。有些例子包括在下述部分参考资料中。在编制报告和传播结果，以及向其他分析人员提供各种服务方面，发生的费用也很高，向其他分析人员提供的各种服务可能包括编制元数据和组织培训讲习班。

D. 事先调查基础设施有限或没有此种设施的调查的成本

20. 在事先调查基础设施相对极少的国家，赞助机构可能需要投放大量的资源，进行拥有大量调查基础设施的国家不需要进行的能力建设（Grosh和

Muñoz, 1996年, 第8章)。由于缺乏基础设施, 筹备活动、实地作业和数据处理的成本可能都将大幅增加。

21. 能力建设一般牵涉对人员进行广泛的初始培训。在事先调查基础设施有限或没有此等设施的国家, 与基础设施十分发达的国家相比, 发展调查需要对外部专门知识加以利用所涉的成本通常很大。此外, 随着调查组织取得经验, 实地工作人员对时间的利用效率往往随之提高。另外, 在拥有大量先前调查经验的国家, 差旅的必要性大大降低, 因为此类国家的统计机构可能拥有有经验的区域数据收集小组, 或为调查实地工作人员提供交通工具。这些优势可节省交通、培训成本和其他人事费用, 无先前调查史的国家通常在调查预算中列入车辆, 而且这个项目可能占调查总成本的一大部分 (Grosn和Muñoz, 1996年, 第8章)。某些调查基础设施或先前调查史的存在具有重大影响的预算项目的其他例子是计算机设备和用于确定住户的地图。

E. 同调查目标修改相关的因素

22. 上文指出, 许多成本因素同调查设计的特点挂钩, 包括样本规模, 调查问卷的长度; 单元数; 用于样本选择和列表的具体方法、试点检验及调查问卷的设计和翻译。就特定的设计而言, 其中所产生的某些成本在各国间大致是不变的。

23. 不过, 发展中国家和转型国家的调查设计经常不得不加以修改, 以便适应受益国政府或其他利益相关方特定的规格。例如, 一国政府可能决定扩大调查的目标, 以便包括其他的国家优先事项。这反过来可能导致: (a) 在调查问卷中列入追加的单元; 或(b) 如果国家以下组别的关键变量的估计数希望达到国家级估计数的相同精确度, 增加报告域的数目。

24. 这些修改会以若干方式影响成本与数据质量间的权衡。首先, 它们可能直接导致访问者收集数据所需时间总量大幅增加, 原因是由于列入追加的调查问卷单元平均访问时间增加[第23(a)段], 或者由于报告域数目增加访问次数以数量级增加[第23(b)段]。其次, 如果调查组织拥有的训练有素的访问者和实地监督员数目相对固定, 那么修改可能导致成本增加, 因为需要培训增加的访问者以及每分钟访问时间所需的监督时间要增加。或者, 可将训练有素的实地工作人员的人数保持不变, 但带来数据收集期延长并因而增加成本的双重后果。第三, 上述增加可能导致非抽样误差相对于抽样误差的规模增大。例如, 将额外单元列入调查问卷可能由于问题检验不够或答复者厌倦而扩大非抽样误差。访问次数或平均访问时间长度增加需要使用更多相对缺乏经验的访问者, 这也可能增加非抽样误差。

F. 关于调查成本报告的某些告诫

25. 应对若干因素加以考虑, 以确保在合理的共同基础上进行调查间和国家间的成本比较。首先, 发展中国家和转型国家的调查由若干不同的组织赞

助，这些组织的政策和核算程序通常不同。例如，对于某些赞助机构，重要的可能是区分赞助机构的成本与实施调查的总成本。

26. 其次，可能重要的是采用比较方式说明以实物提供的调查支助，例如实地人员的运输工具。在有些情况下，实物支助可由国家统计局提供，例如将它的长期实地工作人员分配给国际赞助的调查。虽然此种成本可作为实物考虑而且不计入分项列出的预算，但它们代表一种机会成本，因为调查举措是一项追加的活动，它占去可由国家统计局从事的其他潜在工作的时间。

27. 类似的评论适用于外部技术援助的提供。在没有调查基础设施或者没有进行调查的历史的国家，这一项可能尤其重要。就许多调查而言，此类技术援助由进行或赞助调查的国际机构以实物方式提供，因此不直接包括在调查预算中。不过，有时此种技术援助分包出去，并因而列入预算中。例如，1998年土库曼斯坦的生活水平衡量研究调查即按照与世界银行签署的合同依靠三角研究所的技术援助进行。

28. 第三，由于B节中给出的分级成本结构（表达式1），重要的是应区分一项调查的总成本与每次完成的访问的成本。例如，由于可以获得较多的资源和在较大程度关心国家以下级报告的可靠的估计数，较大的发展中国家和转型国家往往在它们的调查中使用较大的样本规模（联合国儿童基金会，2000年，第4章）。由于调查人员为数众多发生的交通运输和薪金成本高，大国的调查与小国的调查相比，总成本往往较高。不过，总成本较高的大国每次完成访问的成本有时可能较低，原因是规模经济和固定成本分布在较大的样本中。

29. 第四，由于样本设计的特殊性，可能使总成本和每次访问成本的评价变得复杂化。例如，由于使用过度抽样或使用筛选样本确保为小量或难以从框架信息确定的某些分组人口（例如，有5岁以下子女的住户）实现精确目标，成本可能增大。最后，对于住户规模变化大的人口的调查而言，可能也很重要是应区分每个联系的住户的成本与每次完成的访问的成本。

G. 摘要和结束语

30. 发展中国家和转型国家的大多数调查，在预算限制严格和甚至批准的预算的兑换都不确定的环境下进行。因此，影响调查成本的因素的分析是发展中国家和转型国家调查设计和规划过程最重要的方面之一。本章介绍一个进行此种分析的框架，而且还审议调查成本及相关组成部分在多大程度上可在调查设计和住户人口分布及其他因素相似的国家间移植。

31. 大规模的国家调查被用来说明关键的问题，但讨论也可适用于发展中国家和转型国家的国家统计局内部进行的许多其他类型的较小规模的调查。在人们能够确定这些调查中共同成本结构的限度内，人们可以利用关于一个国家一次调查的成本组成部分的信息，为另一国类似的调查的设计提供有益的指导原则，或提高同一国某项新调查的设计效率。已经指出过，在审议调查时拥有大量调查基础设施的国家与没有或极少拥有基础设施的国家之间，调查成本

差异很大。另外还强调了一些告诫，在比较各国间调查总成本时应当加以考虑。

32. 我们最后重申同关系到发展中国家和转型国家调查成本的一些重要问题相关联的要点，即：

- (a) 即使对成本组成部分进行仔细的分析可以揭示各国家组间或调查组间共同的成本结构，也应当认识到调查预算通常不只是分国别，而且还有明确的时间性。因此，重要的是汇编成本数据和编写一份行政报告，以文件形式记录每次住户调查的调查过程每个阶段的成本的各个组成部分。应为方差及其组成部分将同类信息记录成文件。关于成本和方差的此类信息可以两种方式加以利用：首先，用于做出重要的预算和管理决定，其次，用于证明各种样本设计决定如何受不同的成本和方差组成部分影响。一般来说，对于调查过程的每个阶段，成本和方差及其组成部分的文件编写应是发展中国家和转型国家国家统计局标准作业程序不可分割的组成部分；
- (b) 即使调查总成本纳入固定和可变两种成本，正是调查预算中的可变成本，在设计调查的过程中应审慎地加以控制和操纵。有些固定成本，例如用于多个政府机构的调查规划的协调，以及用于针对潜在答复者的宣传的那些固定成本，往往超出调查设计者控制的范围，而且无论如何对于审议中的国家、时间和调查过于具体；
- (c) 如在第十四章中所讨论的那样，在用户付费的调查与国家预算的调查之间，编制预算的考虑因素是不同的。前者的设计精心而且实施起来较为顺利，所有的紧要组成部分都事先支付，而后者通常受预算和国家拨款的限制。对于这类调查，规划的预算与实际的预算间往往存在着巨大的差异，它不由精确性考虑因素所决定，而是由相对于国家其他预算优先事项的调查资金的可获性决定；
- (d) 由于发展中国家和转型国家开展多数调查的预算环境非常严格，对于调查设计者来说，重要的是探索调查预算编制或在无预算情况下调查的执行方面的非货币方法。例如，也许可以与现有的调查合用基础设施；利用已为另一项调查选定的单位的子样本；或安排一名访问者收集多项调查的数据。还应从所需的时间量的角度考虑一项调查的某些方面的预算编制；
- (e) 我们在前面论证过，如果一个国家缺乏调查基础设施和一般的统计能力，调查成本就会大幅增加。因此，建设和加强调查基础设施是一项值得的投资，它能导致发展中国家和转型国家从长远降低调查预算。建设此类调查基础设施和促进一般统计发展的最有效的办法之一是通过发展中国家和转型国家的国家统计局与较发达统计系统之间的技术合作，并与国际统计和供资机构及其他利益相关方协作。不过，为了给受益国带来积极的成果，此类技术合作努力必须精心设计和精

心实施。统计方面技术合作良好做法的实践指导原则由联合国概述（1998年，附件），并得到联合国统计委员会1999年3月4日第十三届会议核可。

鸣 谢

我感谢三位审阅人和2002年10月在纽约联合国总部举行的发展中国家和转型国家调查作业特点分析专家组会议的与会者，因为他们提出的意见非常具有建设性，这些意见使本章第一稿大有改进。不过，其中发表的意见是我个人的意见，并不一定反映联合国的政策。

参考资料

- Ajayi, O. O. (2002年)。“调查的预算框架”。《个人通讯》。
- Andersen, R、J. Kasper和M. R. Frankel (1979年)。《整体性调查误差》。加利福尼亚，旧金山：乔西 - 巴斯出版社。
- Cochran, W. G. (1977年)。《抽样方法》，第三版。纽约：威利国际出版公司。
- deVries, W. (1999年)。“我们是否符合标准？关于国家统计系统业绩的问题”。《国际统计评论》，第67期，第63-77页。
- Grosh, M. E. 和J. Muñoz (1996年)。《规划和实施生活平衡量研究调查手册》。《生活平衡量研究工作文件，第126号》。哥伦比亚特区华盛顿：国际复兴开发银行，世界银行。
- Groves, R. M. (1989年)。《调查误差与调查成本》。纽约：威利国际出版公司。
- Kish, L. (1965年)。《抽样调查》。纽约：威利国际出版公司。
- _____ (1976年)。“线性样本设计的最佳和最接近效果”。《皇家统计学会志》，A辑，第139期，第80-95页。
- Linacre, S. J. 和D. J. Trewin (1993年)。“整体性调查设计：应用于建筑业收集”。《官方统计杂志》，第9期，第611-621页。
- 联合国 (1998年)。“统计学技术合作良好做法指导原则：秘书处的说明”。E/CN. 3/1999/19。10月15日。
- 联合国儿童基金会 (2000年)。《十年终了多指标类集调查手册》。纽约：联合国儿童基金会。
- Yansaneh, I. S. 和J. L. Eltinge (2000年)。“发展中国家调查的设计效果和成本问题”。《调查研究方法科的记录》。维吉尼亚，亚历山德里亚：美国统计学会，第770-775页。

附件

联合国儿童基金会（儿童基金会）多指标类集调查（MICS）预算框架

成本类别	总成本	活动类别								
		筹备/ 敏感化	试点 调查	调查设计和 样本准备	培 训	主要调查 实施	数据输入	数据处理 和分析	报告撰写	传播和进一 步分析
人事										
每日津贴										
交通										
消耗品										
设备										
其他成本										
总成本										
执行机构（名称）										

补充细节

1. 样本规模：住户数 _____ 类集数 _____
2. 查点持续时间：天数 _____
3. 查点员培训期限：天数 _____
4. 实地查点员/监督员人数：查点员 _____ 监督员 _____
5. 数据录入：每份调查问卷的击键数：数目 _____
6. 儿童基金会捐款：美元 _____

成本计算框架

列入成本和活动类别的项目

成本类别	活动类别
人事（薪金）	筹备/敏感化
顾问费	调查问卷编制
实地监督员	虚表的制作
访问者/查点员	翻译和倒译
司机	调查问卷的预测试
译员	查点前后的宣传
当地向导	
数据录入员	试点调查
计算机程序员	培训
加班费	数据收集
奖金	数据分析
协调委员会	试点调查报告
	调查设计和样本编制
每日津贴（膳宿）	规划
实地监督员	样本编制
访问员/查点员	
司机	培训
译员	培训材料的编制
当地向导（伙食补贴）	译成培训语言
顾问/监测员	培训的执行
	主要调查的执行
交通费	执行
租车费	监测和监督
公交补贴	数据检索
燃料	
保养费	数据输入
顾问视察	数据录入
	错误检查
消耗品	数据处理和分析
文具（纸、笔等）	数据处理
身份证	数据清除
文档封袋	指标制作
计算：用品（纸张、软盘、色带、盒式磁带）	分析表
	报告编写
设备	传播和进一步分析
人体测量设备（磅秤、长度计等）	报告印刷
	分发
其他成本	反馈会议
印刷（调查问卷等）	进一步分析
地图、列表和说明手册复印件	
设备维护	
通信（电话、传真、邮件等）	
合同（数据处理、报告编写）	

第十三章

收支调查的成本模型

Hans Pettersson

瑞典统计局

瑞典斯德哥尔摩

Bounthavy Sisouphanthong

国家统计局中心

老挝人民民主共和国万象

摘 要

本章说明在老挝人民民主共和国建立支出和消费调查成本模型的工作。它首先简要讨论成本模型和估计模型中各组成部分的问题，然后描述2002年老挝支出和消费调查的设计。基于调查概算发展而成的成本模型，被用于计算关于类集中同质率的不同假设下的最优类集规模。本章最后分析相对于最优条件下的效率的所选样本设计的效率。

关键词：调查设计、调查成本、效率、成本模型、最优样本规模。

A. 引言

1. 多阶段类集样本的设计涉及若干项决定。要做出的一个重要的决定是如何以尽量最好的方法在各样本阶段中分配样本。类集样本一般对成本和方差具有相反的影响：它减少成本并增大方差。多阶段样本的经济设计要求抽样统计人员估计和平衡这些影响。对于这项任务，他（或她）需要关于可归属于不同抽样阶段的方差的信息，以及关于依赖于每个阶段样本规模的可变成本的信息。

2. 为许多共同的多阶段设计开发了方差模型，但在统计人员中成本模型的开发受到较少的注意。当今，至少为发展中国家许多调查的最重要的估计汇编了方差和设计效果。成本模型用来设计样本不太常见。部分问题是许多国家统计研究所缺乏关于调查成本的详细信息，这使得难以编制调查的准确预算和建立现实的成本模型。

3. 在本章中，我们简要地讨论成本模型并说明如何将成本模型与方差模型合起来用于在两阶段设计中如何找到初级抽样单位（PSU）内的最佳样本规模。我们为老挝人民民主共和国的一项支出和消费调查制作一个成本模型，并用该模型在初级抽样单位内计算最佳样本规模。

B. 成本模型和成本概算

成本模型

4. 两阶段样本的简单成本模型可表示如下：

$$C = C_0 + C_1 \cdot n + C_2 \cdot n \cdot m \quad (1)$$

其中 n = 样本中初级抽样单位 (PSU) 的数目； m = 每个初级抽样单位的样本中次级抽样单位 (SSU) (例如住户) 的数目； C_0 = 进行调查的固定成本，它独立于样本初级抽样单位的数目和每个初级抽样单位的次级抽样单位的数目，包括调查规划成本、发展调查设计的成本、筹备工作成本、调查管理成本及数据处理、结果分析和表示的成本 (其中数据处理的部分成本取决于样本规模并因此不是固定的成本，但这里将它忽略不计)； C_1 = 给样本增加一个初级抽样单位的平均成本，由访问员和监督员往返于初级抽样单位与居住地之间或初级抽样单位之间的差旅费 (燃料成本、司机薪金) 和访问者的薪金组成，包括获取初级抽样单位地图和其他材料的成本，在当地确立调查的成本，例如需要与地方当局会晤和获得其许可的成本，以及初级抽样单位内部居住单位/住户列示和抽样的成本； C_2 = 将一个额外住户包括在样本中的平均成本，包括住户定位、与它联系和对它访问的成本，其中成本由访问员和监督员的薪金和每日津贴组成，以及访问员和监督员在初级抽样单位中的差旅费。

5. 与已经开发的较为复杂的成本模型相比，这一成本模型是简单的，Hansen、Hurwitz和Madow (1953年) 发展了将初级抽样单位之间的差旅费孤立起来的一个模型，其中

$$C = C_0 + C_1 \cdot n + C_2 \cdot n \cdot m + C_3 \cdot \sqrt{n} \quad (2)$$

增加一个初级抽样单位的成本 (C_1) 包括就位差旅费 (数据收集旅途中从访问者家中到第一个访问的初级抽样单位再从最后受访的初级抽样单位回到家中的差旅费)，但不包括由数据项 $C_3 \cdot \sqrt{n}$ 覆盖的初级抽样单位之间的差旅费。将初级抽样单位间差旅费和就位差旅费孤立出来的模型也被提了出来 (Kalsbeek、Mendoza和Budescu, 1983年)。Groves (1989年) 较为广泛地讨论了成本模型，包括各种复杂的形式，例如非线性、不连续的阶跃函数成本表达式。不过，成本模型数学形式的复杂性经常使最优性的检索变得更难。此外，准确数据的缺乏也经常妨碍复杂模型的使用。在本章中，将使用简单模型 (1) 而且假定第二阶段的单位是住户。

成本概算

6. 调查管理者根据先前类似性质调查的信息对于具体调查作业所需的时间往往有很好的主意。经常可将以前调查 (或试点调查) 的经验用于合理估计确定和访问住户时每户所需的时间。在这些情况下，可以汇编 C_2 的合理估计数。通常更成问题的是 C_0 的估计，它涉及间接费用和若干项目/活动中工作的工作人员的费用的分配。往往难以对行政、专业和监督人员所需的时间做出估

计。通常，先前的调查没有可靠的成本记录显示此类工作人员的费用。另外，许多调查利用外国捐助方提供的技术援助（技援）。在许多情况下，可能难以分出技援顾问在一项具体调查上所花的时间。

7. 计划 C_1 的合理估计数通常很困难，因为在样本增加一个初级抽样单位时，这样的计算涉及确定访问者增加差旅费的影响。差旅费取决于所覆盖地区的大小，分配给每个访问者的初级抽样单位数目以及访问者的旅行方式。差旅包括数据收集出行期间初级抽样单位之间的差旅和进驻差旅。

8. 设有捷径可以克服做出可靠成本估计固有的困难。先前调查或试点调查准确而颇为详尽的成本计算非常宝贵。除了以前的经验和试点外，人们还可以通过在进行中的调查中建立特殊的成本监测能力获取所需的成本数据。例如，美利坚合众国的全国卫生访问调查就采取了这种做法（Kalsbeek、Botman和Massey，1994年）。

C. 有效样本设计的成本模型

9. 成本建模可用于两个目的：

- 为了预算目的，根据成本模型中的单位成本和不同阶段规划的样本规模编制一项调查预算；
- 通过合并成本模型与抽样误差模型找到有效的样本设计方案。

10. 在本章中，我们关心的主要是利用成本模型找到一种有效的设计方案。我们假定一个两阶段的设计，而且住户在第2阶段从初级抽样单位中选出。问题可以这样一种方法说明：鉴于成本模型中表示的成本结构，应如何在两个抽样阶段分配样本。通常为城市和农村阶层以及在某些情况下为其他阶层做单列的成本模型。在这种情况下，问题还包括样本在城市和农村（及其他）阶层的分配。

11. 在尝试制定有效的设计方案时，我们不必考虑固定成本（ C_0 ）；重要的部分是实地工作成本： $C_1 \cdot n + C_2 \cdot n \cdot m$ 。用实地总成本除以访问次数（ $n \cdot m$ ），可求出每次访问的估计实地工作成本（ C_f ），已知

$$C_f = C_2 + C_1/m \quad (3)$$

设计的方差可表示为

$$\text{Var} = V \cdot (1 + \rho h (m - 1)) \quad (4)$$

其中 V 是住户简单随机抽样下的方差； ρ 是同质率（Kish，1965年）；另见上文第六章；而 m 是初级抽样单位中的样本规模。

从（3）可以看出，通过使 m 尽量变大可以最大限度地降低每次访问的实地工作成本（ C_f ）。从（4）同样可清楚地看出，方差随 m 增大而增大（而且通过 $m = 1$ 尽量缩小方差）。住户的最佳数目 m_{opt} 是最大限度缩小 $\text{Var} \cdot C_f$ 的 m 的值，其中

$$\text{Var} \cdot C_f = V \cdot (1 + \text{roh} (m - 1)) \cdot (C_2 + C_f/m) \quad (5)$$

现已证明 (Kish, 1965年), 最佳样本规模可用下述公式求出

$$m_{opt} = \sqrt{\frac{C_1 (1 - \rho)}{C_2 \rho}} \quad (6)$$

12. 等式 (6) 中的第一个因子 C_1/C_2 是第一和第二阶段单位成本之间的成本比。将一个新的初级抽样单位包括在样本中的成本 (C_1) 将始终大于将一个新的住户包括在选定的初级抽样单位中的成本 (C_2), 因此成本比将始终远高于 1.0。成本比越高, 选定一个新的初级抽样单位与在选定的初级抽样单位中选定更多的住户相比更费钱; 因此, 我们应在已经选定的初级抽样单位中选择更多的住户。

13. 数量 ρ 测量初级抽样单位的内部同质性。在内部同质性高时, 在初级抽样单位中取一个大的住户样本是不可取的, 因为从样本中每个新住户所获的信息量将很小 (因为住户非常类似)。这反映在 (6) 中的第二个因子中。当 ρ 高时, 这个因子以及 m_{opt} 变小 (对于给定的成本比)。

14. ρ 值通常是根据先前调查估计的设计效果求出的。对于多数人口变量而言, ρ 值往往很小——通常小于 0.01。对于许多社会经济变量而言, ρ 值可能高于 0.1, 在有些情况下高至 0.2 或 0.3。

15. 成本比还必须根据先前调查的经验求出。应当指出, 不必从成本角度表示比率。通常将时间 (从访问者所需天数的角度) 用作单位而不是成本: 数学运算将大致一样 (有些差旅费可忽略不计)。成本比的水平取决于实地工作的设计。对于花在访问上的时间非常短的调查, 成本比可能为 20 - 50。例如, 如果独立于住户访问外的每个初级抽样单位所需的时间是 3 天, 而且访问者能够每天覆盖 10 户, 成本比 (计算为时间比 T_1/T_2) 将是 30 ($T_1 = 3$ 天和 $T_2 = 0.1$ 天)。在访问时间很长的调查中, 成本比可能低于 10。

16. 计算中采用的数学可能留下这样的印象, 即从每个初级抽样单位选出多少住户的问题可以得出精确和清楚的答案。不过情况几乎都不是这样, 这有几个原因, 即:

- 成本模型是现实的极其粗略的近似。需要简化以便使成本模型可加管理 (如 B 节中讨论)。
- 成本和 ρ 值的估计具有不确定性。
- 最佳效果适用于许多调查变量中的一个。如果调查中重要的调查变量有不同水平的 ρ , 那么将不存在单一的最佳类集规模, 而是存在若干不同的规模。

17. 计算将颇为粗略地显示对于不同的 ρ 值最佳样本规模是多少。可将这信息用来决定初级抽样单位中合理地适合于有重要调查变量的一个样本规模。关于最后决定, 可能还有其他因素需加考虑, 这些因素往往同实地工作的实际制约因素有关。

D. 案例研究：2002年老挝支出和消费调查

18. 老挝人民民主共和国国家统计局（国统中心）在过去十年进行了两次支出和消费调查。老挝第一次支出和消费调查（LECS-1调查）于1992-1993年度进行；第二次（LECS-2调查）于1997-1998年度进行；而第三次（LECS-3调查）于2002-2003年度进行。本节说明LECS-3调查。

19. 调查的数据用于多个目的，最重要的是为国民账户产生住户消费和生产的国家估计数。这包括估计住户农业活动和商业活动中的生产。

LECS-3的样本设计

20. 样本由通过两阶段样本设计方案选出的8 100户组成。村庄是初级抽样单位（初级抽样单位）。村庄根据18个省和在省内根据城市/农村部门分层。农村的村庄再根据“通路”村和“不通路”村进一步分层。第一阶段的总样本由540个村组成。将样本与普查得到的人口数的平方根成比例地分配到各省。在每个省用一个与规模成正比的系统概率程序选出初级抽样单位。

21. 所选村庄的住户在调查前先列表。在每个村采用系统抽样选出15户，提供一个8 100户的样本。每村选出15户的决定主要基于实际考虑。在E节中，我们比较15户样本的效率与关于同质率的不同假设条件下的最佳样本规模。

LECS-3中的数据收集

22. 数据通过下述手段收集：(a) 住户调查问卷；(b) 村调查问卷；以及(c) 价格收集表。最后两份调查问卷主要充当用以收集住户调查补充信息的工具。

23. 住户调查问卷的一大部分与先前调查的一样，只是对先前调查中编写不当的问题进行了一些修改。根据所有交易的每日记录收集了一整月的支出和消费数据。在月底，向住户询问前12个月内购买耐用品的情况。在该月期间，住户的每个成员应记录24小时期间的的时间利用情况。为某个“昨天”测量住户每个成员的大米消费量，以便更为精确地度量每人每顿的摄入量。

24. 村调查问卷发给村长填写，覆盖各种项目如道路和交通、水、电、卫生设施、当地市场、学校等。价格收集表由访问者用来收集121种商品当地价格的数据。

实地工作

25. 通过由住户所记的日记测量每天消费情况，不仅给住户也给实地访问者造成沉重的负担。许多住户，尤其在农村地区，记日记的任务经常需要支助。为了确保数据的质量能令人接受，必要的做法是让访问员在村里住一个整月，而不是来回到村里进行反复的访问和随访。这项决定也得到下述情况的支持：许多村，特别是山村，出入很困难（有些村需要步行数天才能到达）。

26. 在先前的调查中，每村由两名访问员组成的工作组开展实地工作。LECS-3调查考虑了单个访问员设计方案。不过，在最终分析中，同访问员安全和福利有关的因素倾向于村里设两名访问员。在4周期间，访问员数次访问选定的住户。访问员还与村里领导人合作填写村调查问卷并更新村登记册。当月期间，访问员还收集了当地市场的价格数据。

27. 实地工作人员包括180名访问员，他们被安排在90个两人工作组内。省统计局的36名监督员和总局的10名中央监督员监督各工作组。

E. 老挝2002年支出和消费调查（LECS-3调查）实地工作成本模型

成本概算

28. LECS-3调查很大程度上类似于前两次的LECS调查。因此，前两次调查中的实地工作所需的时间方面的经验被用来估计LECS-3调查的实地工作成本。

29. 表十三.1载有LECS-3调查各村实地工作所需时间的估计数。为城市和农村地区分开做出了估计。

表十三.1
村子实地工作估计时间

	实地旅行 (天数/村)	介绍调查、村里住户列表和选定、收集村信息 (天数/村)	住户访问工作 (天数/村)
城市地区（100个村）			
省监督员	1.5	0.5	3
访问员（2人组）	3	7	47
农村地区（440个村）			
省监督员	3	0.5	3
访问员（2人组）	6	7	47

30. 表十三.2载有根据表十三.1中的时间估计数计算的实地工作估计成本。成本包括差旅费（通常乘坐轿车或公共汽车）和实地工作时间的实地补贴（每日津贴）。从事调查的工作人员无例外地是国统中心在编人员，被分配从事调查工作，作为其正常职责的组成部分。因此，成本项目不包括正常的薪金。

成本模型

31. 表十三.2中的纵列A和B列示同调查的村子的选择和准备有关的成本。这些纵列各项目之和除以村子数，得出将一个村子列入调查范围和美元平均成本（ C_i ）：城市地区： $C_i = (1\ 540 + 2\ 490 + 450 + 5\ 060) / 100 = 95$ ；而农村地区： $C_i = (15\ 850 + 25\ 560 + 1\ 990 + 22\ 260) / 440 = 149$ 。所有差旅费差均被视为村子间的差旅；因此所有的差旅费都包括在 C_i 中。

表十三.2

LECS-3调查的估计成本（美元每日津贴）

	实地旅行差费（旅行期间的每日津贴和估计的差旅费）	介绍调查，村里住户列表和选定，收集村信息	住户访问工作
	A	B	C
城市地区（100个村）			
省监督员	1 540	450	2 710
访问员（2人组）	2 490	5 060	33 970
农村地区（440个村）			
省监督员	15 850	1 990	11 950
访问员（2人组）	25 560	22 260	149 460
合 计	45 440	29 760	198 090

32. 表十三.2中纵列C列示同住户访问有关的调查成本。主要项目是访问者的时间。本纵列各项之和除以住户数，得出将一住户列入调查范围的美元平均成本（ C_2 ）：城市地区： $C_2 = (2\,710 + 33\,970)/(100.15) = 24$ ；而农村地区： $C_2 = (11\,950 + 149\,460)/(440.15) = 24$ 。将 C_1 和 C_2 的估计值插入，成本函数变为

$$\text{城市: } C_{\text{fieldwork}} = 95 \cdot n + 24 \cdot n \cdot m \quad (7)$$

$$\text{农村: } C_{\text{fieldwork}} = 149 \cdot n + 24 \cdot n \cdot m \quad (8)$$

33. 人事费不包括在编人员薪金，其结果是低估了 C_1 和 C_2 并因而低估了 $C_{\text{fieldwork}}$ 。不过，对于优化设计最重要的是成本比 C_1/C_2 。我们可以预计成本比仅略受省略薪金的影响，因为省略对 C_1 和 C_2 的影响非常相似。

34. 第一阶段与第二阶段样本间的成本比是：城市地区 $C_1/C_2 = 95/24 = 3.9$ ，而农村地区为 $149/24 = 6.1$ 。这些成本比颇低，并反映这样一种情况，即在使用访问者支助的日记法的当月，每户访问和随访需要相当长的时间。LECS-3调查在这方面是一项非同寻常的调查。

村子中的最佳样本规模

35. 在先前的LECS调查中，两名访问员的工作量为每村20户。对于LECS-3调查，样本规模缩小至15户。工作量从20户减至15户的原因是：与先前的调查相比，LECS-3调查中的住户访问大大延长。另外，LECS-3调查包含未列入先前调查的价格调查问卷。

36. 一村两名访问员覆盖15户样本的设计方案效率如何？可将成本模型连同方差模型一起用于评估15户样本的相对效率。

37. 在表十三.3中, 为不同的 ρ 值列示 m 的最优值。我们的设计的相对效率列示于行3和4。将它计算为 $\text{Var}.C_f$ (见(5)) 的最低值与给定 ρ 值和样本规模15的 $\text{Var}.C_f$ 实际值之间的比率。 ρ 值高至0.10效率较高; ρ 值等于0.2和以上, 效率非常低, 并且往往越来越低。

表十三.3

村中最佳样本规模 (m_{opt}) 和不同 ρ 值的实际设计 ($m=15$) 的相对效率

	$\rho=0.01$	$\rho=0.05$	$\rho=0.10$	$\rho=0.15$	$\rho=0.2$	$\rho=0.25$
m_{opt} , 城市地区	20	9	6	5	4	4
m_{opt} , 农村地区	24	11	8	6	5	4
城市相对效率 (百分比)	99	94	82	73	66	61
农村相对效率 (百分比)	96	98	89	81	75	70

38. 先前LECS调查中 ρ 的计算表明, 对于重要的LECS变量, ρ 值存在着明显的城乡差异。城市地区的 ρ 值大大低于农村地区的 ρ 值。对于LECS调查中的许多城市估计数, 我们可以预计 ρ 值的范围在0.04-0.08之间, 在此种情况下, 8至9户的样本将是最佳样本。我们每个初级抽样单位15户样本的设计将达到85-95%的相对效率。农村地区的 ρ 值在0.11 - 0.20的范围内, 在这种情况下, 5至7户的样本将最佳。我们的样本将达到75-88%的相对效率。存在着某种不确定性, 特别是关于我们在LECS-3调查中的重要变量方面可以预计的 ρ 值。我们仍然能够把握地得出结论认为, 我们的15户样本超过了最佳水平。

39. 这些结果对于今后的LECS调查具有哪些实际影响? 在城市地区的效率损失小; 因此, 我们可以决定坚持15户的替代方案。在农村地区, 我们希望缩小每个初级抽样单位的样本。不过, 目前的实地工作安排即访问员在初级抽样单位中不得不呆一整月, 使得难以大幅度缩减工作量。这就是说, 访问员在这个月内的工作量将不充实。也许可以给访问员分配其他任务以充实工作时间的访问量, 例如, 当月在该地区进行社区调查。这是否是一种可行的选择方案, 尚有待讨论。

F. 结束语

40. LECS-3调查中的实地工作的成本模型已作了开发和分析。它表明, 调查的成本比C1/C2相当低。主要原因是LECS-3调查采用了由访问者支助的耗时日记法, 其中, 访问员在村里呆一整月, 并向住户提供日记所需的一切协助。在这方面, 与其他住户收支调查相比, LECS-3调查是一项异乎寻常的调查, 在那些调查中, 每户的访问时间通常较短。

41. 初级抽样单位内最佳样本规模的计算表明, 目前的15户样本规模在最佳规模之上, 尤其是在农村地区。不过, 实际的制约因素可能使得难以缩小样本规模。

42. 应当指出，成本模型只是现实的粗略近似；任何简单的模型都不能完全把握整个复杂性。可以建立较复杂的模型，例如包括各种阶跃函数成本表达式。不过，成本模型数学形式的复杂性常常将使最优性的确定工作越发难以进行。

参考资料

Groves, R. M. (1989年)。《调查误差与调查成本》。纽约：约翰·威利国际出版公司。

Hansen, M. H.、W. N. Hurwitz和W. G. Madow (1953年)。《样本调查方法与理论》，第一期。纽约：约翰·威利国际出版公司。

Kalsbeek, W.、O. M. Mendoza和D. V. Budescu(1983年)。“多级抽样中的最有利分配成本模式”。《调查方法》，第9期，第2号，第154-177页。

Kalsbeek W. D.、S. L. Botman和J. T. Massey (1994年)。“国家健康访谈调查中的成本效率和许可呼叫尝试的数量”。《官方统计杂志》，第10期，第2号，第133-153页。

Kish, L. (1965年)。《抽样调查》。纽约：约翰·威利国际出版公司。

第十四章

为发展中国家住户调查制定预算编制框架

Erica Keogh

津巴布韦大学统计系

津巴布韦哈拉雷

摘 要

本章的目的是为调查工作提出关于编制精细与合理预算的建议。向读者说明，有两种方式可以审查这种预算，一种是按照会计科目分类，另一种是按照调查活动，因此，鼓励利用在每一个调查活动内详细列明会计科目分类的做法来编制预算。最终成品是一个成本汇总表，也可以用于整个调查工作中，用以记录实际支出。极其鼓励为实际调查成本提供证明和进行讨论，以便为今后的调查工作提供投入资料。本章自始至终一直在强调抽样调查的设计和预算编制之间的关键相互作用。

关键词：调查设计、调查预算、调查实施。

A. 导言

1. 无论是在时间方面，还是在财力方面，调查都是一项成本很高的工作；因此，务必为调查工作一开始直到结束期间可能发生的支出制定详细计划。另外，还必须为意外事项、紧急情况和意料不到的经济变化制定计划，并且必须保证拟议预算会涉及到这些不可预见的事件。为意外事项制定计划的一种方式是在调查过程中培养调整调查工作范围的能力，包括抽样规模，从而使入更有能力灵活处理没有预料到而且可能对调查的实施工作产生影响的经济变化。调查预算应被视为是一种动态的调查过程，随着调查的实施期间实际需求的变化而变化。监测支出的工具要与预算一起制定，并且要不断更新，以反映实际预算进度。

2. 由于预算的规模及其在调查工作范围内对各组成部分的预算分配会对调查结果的质量产生直接影响，因此不能过多强调详细计划和预算编制的重要性。第七章详细讨论了住户调查设计期间的成本问题。联合国（1984年）强调了权衡成本和的重要性，内容如下：“理想地说，应该在分析利用稀缺资源各种替代方法的成本和效益的基础上，确定优先事项”（第1.5段）。通常，

调查预算是固定的预算，样本设计者的任务是在本预算范围内进行一种误差水平能够令人接受的设计。

3. 为拟议的调查编制详细预算通常是一件很烦琐的工作，因为它需要做出极小的规划和准备。另外，调查规划者还会在规划时陷入进退两难的境地，因为在出台最终调查计划之前，无法对预算做出正确估计，而且预算编制工作还必须在最终调查规划/设计之前进行。因此，在以前调查中的预算编制和成本计算方面的经验就起到非常重要的作用。另外还有必要记住，如果不考虑成本，就无法考虑最佳抽样分配：例如，如果使用分层抽样法，那么就能在使精确度固定的成本降至最小或固定成本的精确度达到最佳之间做出选择（Scheaffer, Mendenhall和Ott, 1990年）。但是，成本模型经常与实际不符，不允许在调查期间有可能出现的情况变化，通常只考虑一种变量的误差。因此，保留详细的预算编制和最终支出记录是一项重要工作，其目的是支持不断加强如下宣传，即鼓励调查从业人员采集成本信息以便为今后的调查规划提供帮助。

4. 传统上，规划和/或决策必须使用调查数据，因此，应尽快获取调查结果。通常，调查工作必须在严格的时间框架内进行，出资机构对调查各个阶段的最后截止日期都作了明确规定。但是，必须记住，多用一点时间就可能获得质量更好的数据；因此，调查从业人员应该在调查工作的预算编制阶段就做好提倡这种做法的准备。例如，正如通常的情况那样，如果分配给数据管理和分析的时间和/或预算不够充裕，那么调查结果的质量就可能有危险。因此，必须在预算编制阶段“巧妙地处理”时间、成本和误差，以便制定能够完成调查的最适当框架。

5. 本章的目的是重点阐述：

- 如何着手编制预算；
- 预计将在实施调查时出现的缺陷；
- 制定用以管理和报告调查财务的工具；

特别提到发展中国家的个人访调式住户调查。

B. 初步考虑

1. 调查阶段

6. 作为开始，在对住户调查预算的主要组成部分进行相当详细的审查之前，明智的做法是提醒自己记住调查的主要阶段，因为每个调查阶段的成本都必须做出规划，并且必须尽量遵守。调查阶段可以总结如下：

- 调查设计和准备；
- 调查的实施；
- 调查汇报。

需要每天分配大约24次访调。目前，调查问卷的长度、每天的访调数量以及被访人之间的距离都决定着所需实地工作人员的数量。

10. 上图说明：

- 有多少个调查阶段重叠，例如，数据录入设计将与调查问卷定稿同时进行，数据录入本身会在收到第一批调查问卷之后马上开始，数量整理工作可以在所有数据完成录入之前就能够进行。
- 有多少项任务会在整个调查期内继续进行，例如，编写报告是调查协调人的持续任务，因为每一调查阶段都必须报告。
- 在有些情况下，各个阶段只能逐个完成，例如，调查问卷的最终印制工作不能在试点完成之前进行，而且印制期限很短，与主要培训工作同时进行（请记住，始终建议你们在培训之后尽快开始访调）。

3. 调查类型

11. 预算编制可能取决于所要开展的调查的类型。就预算编制而言，在此要考虑的主要类型的调查有两种，即编入国家预算的调查和用户付费的调查。

编入国家预算的调查

12. 每个国家都有具体的（政府）部门负责执行定期调查，例如，健康和营养状况调查、住房人口调查、收入调查、消费调查、支出调查以及农业和牲畜调查。在这些调查中，大部分都可能：

- 拥有某种已经到位并且在此种性质的调查中再三使用的共同基础设施，换句话说，它是“综合”方案的一部分；
- 已被编入中央政府的预算范围之内，尽管也可能会要求捐助人提供补充资金；
- 有永久编制人员参加调查；
- 拥有信息技术设备和交通设施等等。

换句话说，这些调查是与公共部门某些单位的日常生活不可分割的一部分，同样，将在很大程度上侧重于先前对本次研究的预算编制工作投入的研究。这些调查通常利用全国有代表性的样本进行，并且通常具有一定灵活度的时间表，截止期限是用月来表示，而不会具体到天。本章其余部分介绍的一些预算编制项目可能不适用于这种调查。

用户付费调查

13. 用户付费调查与任何中央政府方案都没有关系，而是在各种国内和国际非政府组织和捐助人的资助下由私人组织进行。这些调查可能是“一次性”的，用户可能马上需要高质量的调查结果。另一方面，这种调查可能用于

方案监测，有的时候，为了给今后活动的规划建立模型，可能需要扩展数据分析。进行这种调查的机构可能拥有：

- 调查进程可依赖的有限基础设施；
- 一批可供这类研究选用的工作人员；
- 有限的信息技术设备和交通设施；
- 有限的固定资源。

或者它们拥有完善的组织机构，在最近期间执行多次这种调查。固定资源和间接费用必须编入预算，而且，如果该组织是一个私人组织，那么利润因素也必须加以考虑。这种调查的抽样规模通常不会太大，通常这种调查只会集中在国家内部的少数地理区域。严格的时间表和截止期限通常是这些调查的特征之一，但遗憾的是，由于实际规划不充分，故数据的质量常常受到损失。

4. 预算与支出

14. 早在开始实施调查之前，就应为调查编制预算，而且，在开始实际规划之前，必须拟定预算编制框架并提交给供资组织。因此，关于调查设计的某些基本假设必须在编制预算的时候做出。另外，实际调查支出反映研究期间发生的真实情况。调查实施人员需要了解这种特性，并且认识到预算必须顾及最终成本。支出在很大程度上取决于同通货膨胀、汇率等变化有关的时间，当然，各国的情况不同，有时候差别相当大。建议按照人工作日、旅行距离等，以及按照预计成本（使用国际货币）执行预算编制，以便更好地应对国内急剧增长的通货膨胀和宏观经济状况发生的类似意外变化。如前所述，调查预算是一种动态预算，实施期间一旦发生实际支出，调查预算就会不断变化。

5. 以前的调查

15. “学习以往的经验”是我们大家都熟悉的格言。但是，就调查预算编制而言，这比人们预料的要困难得多。从全世界范围来看，存在一种相当恶劣地/不完全地报告调查成本的趋势，这就意味着要为下一次调查规划找到有用的信息是一件很难的事。作者在向最近进行过调查的一些组织索要成本信息时发现，在大多数情况下，只能提供原始预算信息，可是据报告说，由于许多如通货膨胀等外部因素的影响，致使实际成本分配与预算分配存在差异。所有地方似乎都没有报告实际成本，所有各方似乎都已经承认这是一种正常和可以接受的现象，只要调查工作在预算范围完成就行。对过去调查预算进行审查的另一个问题是缺少有关隐性成本的报告，例如，免费使用车辆、主管人员的工资等。事实上，这些费用经常被作为间接费用而未编入调查预算，这一现象会在今后对研究人员产生误导。

16. 希望通过阅读本章内容，鼓励调查实施人员记录日常费用，并且为这些费用提供充分的支出凭证，以便今后的研究人员能够从过去的经验中学到东

西。为每次访调的成本提供充分的支出凭证对于那些希望在今后为类似调查编制预算的人来说极其宝贵。每次访调成本涉及到所有方面的实际调查成本，包括设计、实地调查、数据处理和报告，并且提供极其全面的实际成本汇总。

C. 预算框架内的主要会计科目分类

17. 有两种方式可以用来审查调查预算和调查支出，即根据调查活动或共同会计程序审查。有人建议，在制定预算框架时，最好在每一个调查活动范围内单独考虑会计科目分类。然后可以利用从每一个活动采集到的信息，对全部会计科目分类进行总结，并且将它们汇总以便提交给供资机构。同时，向供资机构说明每一个调查活动的详细预算编制，以便强调每一个活动的特殊需要也会有帮助。下文的表十四.2介绍了这种做法的一个例子，利用汇总表从两个角度来说明编制预算的需求。

表十四.2

会计科目分类与调查活动的汇总表

	咨 询	设 计	抽 样	实地调查	数据处理	报 告	总 计
人员							
交通							
设备							
消耗品							
其他							
总 计							

18. 通过将它与表十四.1中的甘特图提供的时间表进行对比，可以看出表十四.2的目的在于突出表十四.1列示的相同调查活动。尽管由于空间的原因，已在上文表十四.2的各单元格内进行了分组，但还是有必要对这些成本计算的具体计算方式作进一步详细说明。

19. 本节将集中说明与调查预算编制有关的会计科目分类，而D节将集中介绍调查活动的预算编制问题，E节将进行“归总”。下文所述种类并不详尽，其他具体调查的种类可能需要（重新）定义。

1. 人员

20. 所有工作人员的工资和薪酬都必须精心计算，并列入预算。这里需要考虑的额外成本包括在调查工作长期延续时可能产生的费用：例如，通货膨胀不断加剧势必要求工资上涨。还必须对健康状况不佳和工作人员调动做出规划。

21. 付给工作人员的薪酬应该符合当地的实际情况，但应该记住，由于调查人员长时间工作，包括晚上和周末，并且经常是属于合同工作，因此，在给予报酬时应该考虑到这一点。可能需要附加福利，并且必须列入预算编制程

序。记住，如果工人认为他们没有得到应有的报酬，那么他们的工作往往会出错，从而增加非抽样误差。根据调查的时间长短，建议按天、按周或按月支付工作人员的费用。资金必须从调查工作的一开始就要到位，从而能够及时和足额支付工作人员的薪酬和工资。如果调查员和组长远离家庭到外地工作，那么就需要支付外地补贴。有些调查实施人员习惯在“每次完成访调”的基础上支付实地调查人员的薪酬。但是，这种做法可能导致很多偏见，建议不要采取这种做法。

22. 所有类别的工作人员，从最低级到最高级，都应该予以说明，包括那些只在项目中做兼职的工作人员在内。调查时间表将指导人们对每个潜在工作人员的工作时间进行评估。

23. 绘制工作人员的工作量图是编制薪酬和工作预算的一种方式。它再次利用汇总表做法，总体介绍每个调查小组的每个成员可能使用的时间。下表十四.3列示了一个例子。同上，在规划过程期间将会需要每一个单元格内的其他详细资料。

表十四.3
已规划的工作人员时间（天数）与调查活动的汇总表

	人员数量	每项活动中的人工作日天数						
		咨询	设计	抽样	实地调查	资料处理	报告	总天数
管理人员								
主管								
组长								
调查员								
资料员								
分析员								
秘书人员								
司机								
其他								
总天数								

24. 应该向实地调查员提供每日补贴（每日津贴），用以支付他们在工作时的餐饮和其他基本需要。这种补贴的数额多少应该在当地的限制范围之内，但也许比平常稍多一点，这样才能应付缺少食物的局势和保证这种资金可以用于紧急情况。

25. 离开家庭到外地工作的所有工作人员的住宿费用也必须编入预算，并且必须及时发放。在许多情况下，当工作人员从一个地方来到另一地方时，他们往往更喜欢自己找住处；但在另外一些情况下，集中安排住处可能会更好。

2. 交通

26. 如果一个人知道被访人的住址，那么交通成本很好估算，也就是说，在基本抽样设计已经确定之后。根据不同情况，可以建议调查员自己找交通工具，做好交通费用的记录，以便事后报销，或者可以选择向每一组实地调查员提供交通工具。最好使用后一项选择，因为调查小组将要作为一个“团队”工作，并且这样更加便于组长随时了解访调进度。不能预见的额外成本包括燃料价格的上涨、意料不到的天气状况致使某些道路无法通行等等，这些不测事件应该包含在应急费用之中。

27. 组长与调查管理人员之间定期举行会议的交通费用也应该编入预算，再次以坚持一贯数据采集方法为目的。

28. 可能需要为实地调查人员购买或租用车辆/摩托车/自行车，而且在通货膨胀不断加剧的情况下，要将这些项目编入预算可能会有困难。

3. 设备

29. 通常，在调查工作进行之前提前对可能发生的设备支出做出比较准确的估计不是一件难事。这些方面的预算可能会出现的问题通常集中在不断上涨的价格和能否搞到所需物品。如果情况可能是这样，建议提前采购所需物品，并且采购足够多的物品以满足整个调查工作之需。信息技术、通信、复印和印刷设备也必须在此予以考虑。

4. 消耗品

30. 在这一部分预算中需要考虑的项目包括各类文具、软件及实地调查需求如袋子、地图、识别文件和书写板、其他办公设施等等。印刷和复印消耗品将会成为这一部分调查预算的重要组成部分，因为在整个调查期间，全天24小时都应当能够用到复印设备。

5. 其他成本

31. 在调查工作进行期间，总是会发生少量宣传和信息费用。这些活动的多少完全取决调查的性质和规模，并且可能在整个调查期间的各种时候进行。这种活动的例子包括：与包括社区领导人和最终用户在内的所有有关当事人的会谈或研讨会、提前与被访人联系、作广告等。宣传活动应该在整个调查期间一直持续进行，因为要向有关当事人反馈信息，以便为最终公布结果做准备。

32. 在有些调查阶段，需要雇用大量的工作人员。必须要有足够的空间用以组织长时间的会议（例如，在培训期间）、存放调查问卷和供数据录入人员及主管在舒适的环境中工作等。有时候，需要租用备用场所，例如，与实地调查区更近的场所，而在另外一些时候，这些场所应随时可以利用。

33. 除非做好充分准备，否则培训费用可能会大幅度增加。培训费用包括培训设施的接待费和培训性访调的交通费，以及所有参与人员的每日津贴费用。所有这些费用都需要予以考虑。

34. 在进行调查时必需的通信方式容易被忘掉。这些将包括使用电话、电子邮件、传真和邮政。通常难以对这些项目进行预算，因为人们无法知道所需的数量。一般来讲，只设定一个总数字，通常将总体预算中的一部分比例用作通信经费，希望能够支付实际费用。必须与实地调查工作队保持持续通信，从而保证迅速处理无法预料的事件，保证坚持一贯的数据采集方法。在移动电话网络比较可靠的国家，这些工具提供了极其有用的即时通信手段。

35. “隐性”成本涉及到已经“到位”的项目/基础设施的预算编制，如计算机或办公场所等。其他可能不太明显的隐性预算包括被雇用执行不止一个项目的任务的人员的业务费用、交通费用以及许多不同项目都要使用的消耗品的费用，每一种费用都有自己的预算。通常，最好尽量在调查规划期间对将要花费/使用的实际时间/数量做出估计，尽管有的时候只能将这些额外间接成本粗略估计为占总体预算中的一部分百分比。重要的是要查明所有这些隐性成本，并且说明其用途，从而在今后调查的规划中，能够了解它们并做出相应的规划，尽管届时的情形可能已经发生变化。

6. 按会计科目分类编制预算的例子

36. 如前所述，极难获得有关实际调查成本的信息。下面的第一个例子由Ajayi（2002年）提供，并且提到从许多非洲国家收集的关于在联合国要求提供儿童、产妇保健和福利指标之前进行的十年目标（EDG）调查的成本计算信息。

表十四.4

按会计科目分类的成本占总预算的比例：十年目标调查（1999–2000年），部分非洲国家（百分比）

国 家	人 员 ^a	交 通	设 备	消 耗 品	其 他	抽 样 规 模
安哥拉	62.7	22.2	9.6	1.3	4.2	6 000
博茨瓦纳	79.2	0 ^b	10.1	3.5	7.2	7 000
厄立特里亚	64.0	0 ^b	28.0	4.8	3.2	4 000
肯尼亚	62.3	22.8	3.3	4.7	6.9	7 000
莱索托	75.1	5.2	5.8	2.3	11.6	7 500
马达加斯加	31.2	6.5	33.3	12.8	16.1	6 500
马拉维	32.0	17.3	23.9	21.6	5.2	6 000
索马里	43.8	17.7	5.0	1.0	32.5	2 200
南非	69.3	24.0	1.5	3.7	1.5	30 000
斯威士兰	29.8	4.3	1.9	1.0	63.0	4 500
坦桑尼亚联合共和国	77.9	12.8	1.6	1.2	6.5	3 000
赞比亚	81.8	5.2	2.0	5.6	5.4	8 000
总 计	62.9	14.9	7.4	6.3	8.5	7 054

资料来源：Ajayi（2002年）。

^a 包括每日津贴。

^b 表明不可能单独抽取此信息。

例 2

37. 12个国家提供了按会计科目分类的调查成本信息。下文表十四.4介绍了在编入国家预算的调查中所使用的会计科目分类例子，说明用于每项成本的预算占总预算的比例。

38. 从表十四.4中明确看出，在按会计科目分类编制预算时，不同国家的类似调查存在相当大的差异。我们希望在样本规模不断扩大的同时，分配给人事成本的预算也会有一定比例的增加；但实际情况并非如此，例如，拿南非和坦桑尼亚联合共和国进行比较，情况就并非如此。然而，预计大部分调查的人事成本可能真的会占到人事成本总预算的三分之二，包括实地调查期间的每日津贴。对于任何国家调查而言，下一个成本最高的项目可能就是交通，当然，交通成本会随着需要覆盖的区域的不同而各不相同，并且可能会占到总体预算的百分之十五到二十。这些调查的经费是由联合国儿童基金会(儿童基金会)和有关政府提供，儿童基金会提供的资金因国家不同数额差距很大。

例 3

39. 本例提到了在1999年进行的一次住户调查的预算编制问题，这次调查是作为调查津巴布韦小额供资业务情况的“微型企业服务影响评估(AIMS)”研究(Barnes和Keogh, 1999年; Barnes, 2001年)的一部分，并因此提到了一次用户付费调查[由国际管理系统通过美国国际开发署(美援署)出资]。

40. 表十四.5说明，大部分(75%)预算被分配给人事成本，包括每日津贴。出现这种情况的部分原因在于调查设计，而调查设计是在1997年进行的一项基线调查后进行的，需要确定相同被访人的住址和/或身份，这是一项极为费时的工作。

表十四.5

分配给各会计科目分类的预算比例：微型企业服务影响评估(AIMS)，津巴布韦(1999年)(百分比)

人 事	交 通	消 耗 品	其 他	样本规模 ^a
75	8	912	5	691

* 由于各种原因，1 997名被访人中有92人的地址没有确定，因此，最终样本规模为599人。

D. 预算框架内的主要调查活动

41. 只有知道了需要编制预算的调查的所有方面，才能对将要使用的会计科目分类进行定义和做出安排。下一步就是考虑调查阶段，并且利用已经定义的会计科目分类，为每一个阶段单独拟定完整的预算。这将导致利用C节所述的汇总表表示法拟订预算框架。

42. 如果把未来的成本资料编制工作放在心上，那么随着实际调查工作逐步深入，实际成本就会一目了然，按同样方式编制预算将使比较更加容易，并且使人一眼就能看出预算和成本之间的显著差异。

43. 另外，这种办法还有助于了解数据质量、调查时间表和预算之间的密切联系。

1. 为调查准备工作编制预算

44. 在调查的这一阶段，人们会遇到为实施调查所需的一切准备工作编制预算这一问题。应该依次考虑所有会计科目分类，并且确切估计每个种类都需要什么。如果工作所在环境的通货膨胀率很高，那么最好提前为消耗品、文具、设备、车辆使用等下订单。与准备和最终确定抽样设计、调查问卷及其随附手册以及早期准备数据录入和管理一样，人员招聘和宣传都是重要活动。

45. 调查设计过程的主要部分是准备抽样标架。调查类型将决定抽样标架的性质，但有时候更新现有标架或形成新的标架会需要相当长时间或大量旅行，有的时候是两方面兼而有之。这包括需要确定清单，无论是家庭、村庄，还是一些更高级别抽样单位的清单，并且这种清单需要单独的预算拨款。

46. 其他可能花费大量时间的活动是编制调查问卷以及培训和实地调查手册。

2. 为实施调查编制预算

47. 因为实施调查可能是调查中成本最高的一个方面，所以在每一个会计科目分类范围内，对每一种可能的情况进行精心的预算编制是一件极其重要的工作。分配给最终印制调查问卷的时间和预算必须仔细考虑透彻，并且提前做出可靠的资源规划。重要的是要记住，在实地调查工作开始进行的同时，中心办公活动应该促进数据录入工作的开展。

48. 正如在前面所强调的那样，不得为了不超预算而对分配给实地调查工作的时间进行调整，因为这样做的结果可能是由于增加非抽样误差而损害数据质量。

3. 为调查数据处理编制预算

49. 编制数据录入、验证、整理和分析的预算计划时，应考虑到一切可能的情况，以便保证这些活动不会因仓促完成而导致报告质量不好和不完整。本阶段的印刷量很大，在文具方面敷衍了事也会有损于结果的总体质量。还将需要有充分的信息技术设施，包括已录入数据的备份设施（光盘、磁盘）。

4. 为调查报告编制预算

50. 一旦完成实地调查工作并且全面展开数据录入工作，调查工作将进入下一个预算编制阶段，即报告和定稿阶段。调查设计在此会再次起到作用，因

为它将决定数据分析的范围和所需报告的级别。最好持续保留整个调查活动期间的所有资料，因为有关各种活动、决定、问题和成本的每日记录将会成为报告的说明部分的好材料。应该对各种会计科目分类予以仔细考虑，并且为最后调查阶段的每一个会计科目分类拨足经费。

5. 为调查活动编制预算的例子

51. 在此从调查活动角度说明上文第C.6节所举例子中的信息。

例 4

52. 回过头再来谈谈例2（十年目标调查），在本例中，提供了关于10个国家特定调查活动成本计算的资料。下文表十四.6是一个汇总表。

表十四.6

调查活动的成本占总体预算的比例：十年目标调查（1999-2000年），部分非洲国家（百分比）

国 家	准 备	实 施 ^a	数据处理 ^b	报 告 ^c	样本规模
安哥拉	0 ^d	83.6	6.1	10.3	6 000
博茨瓦纳	10.4 ^d	59.1	21.7	8.8	7 000
肯尼亚	0 ^d	93.9	2.6	3.5	7 000
莱索托	0 ^d	73.2	18.6	8.8	7 500
马达加斯加	0.3	78.6	3.0	18.1	6 500
马拉维	5.0	62.7	16.4	15.9	6 000
南非	1.3	93.1	2.9	2.7	30 000
斯威士兰	63.0	23.4	7.5	6.1	4 500
坦桑尼亚联合共和国	22.7	72.4	3.6	1.3	3 000
赞比亚	0.4	92.0	6.4	1.2	8 000
总 计	7.0	81.0	6.0	6.0	7 054

资料来源：Ajayi（2002年）。

^a 包括培训、设计、试点和数据采集。

^b 包括数据录入，管理和分析。

^c 包括报告编写和分发。

^d 表明不能单独抽取此种信息。

53. 除斯威士兰外，所有国家都表明，必须分配给调查实施的成本所占的比例很大：可能比较合理的估计是，这一调查阶段的预算占总体预算中的70-90%。由于马拉维的数字（正如表十四.4所显示的那样）表明它的设备成本非常高，因此，这可以用来解释为什么表十四.6中的数据处理和报告成本在总体预算中所占比例较高的原因。但是，博茨瓦纳和莱索托的数据处理成本相对较高却无法解释。在这种情况下，已经要求各国提供一个关于成本问题的“汇总表”，说明各调查活动范围内的会计科目分类；遗憾的是，只有坦桑尼亚联合共和国和厄立特里亚提供了这种汇总表。

例 5

54. 再来看例3，下文表十四.7介绍了有关津巴布韦1999年微型企业服务影响评估调查（AIMS）按调查活动分列的成本信息。

表十四.7

调查活动的成本占总体预算的比例：津巴布韦微型企业服务影响评估调查（1999年）（百分比）

准 备	实 施 ^a	数据 ^b 处理 ^b	报 告 ^c	样本规模
4	85	8	3	599

^a 包括被访人的位置、设计、培训、试点和数据采集。

^b 包括录入、管理和整理。

^c 仅指截止制作纯净数据集的本地化报告。详细的数据分析和最终报告的编制将根据单独的合同进行。

55. 本用户付费例子在上文说明，调查实施成本之所以在总预算中占有相当高的比例，可能是因为这次微型企业服务影响评估调查的样本由上次（1997年）调查随访的691名被访人组成，确定他们的处所耗费的成本相当高（占总预算的22%）。

E. 汇总

56. 一旦各种会计科目分类准备好每一类调查活动的成本，就可以按调查活动来列示各种会计科目分类的汇总表，以期为调查预算的最终审议提供方便。制作这种汇总表有助于调查规划人员对全球级别上的调查工作进行审查，消除不一致和重叠之处，突出预期的主要成本；并且有助于出资机构比较各种调查的成本，从而有利于更好地评估拟议预算的有效性。

57. 如在上文例4中所述，在21个参与十年目标调查的国家中，实际上只有两个国家按照要求提供了按调查活动的会计科目分类列示的成本汇总表。因此，对于本例而言，我们无法按调查活动编辑会计科目分类汇总表。

58. 但是，下文表十四.8提供了有关微型企业服务影响评估调查以及表十四.5和表十四.7中的交叉分类的信息。

59. 上文表十四.8所示的这种汇总表明确地表达了调查工作的预算要求，将会有利于鼓励出资机构考虑出资申请。另外，如果提供这些详细资料，就可以更加容易调整预算，以满足在通货膨胀加剧时出现的意外需要。最后，对整个调查过程期间必须发生的支出进行持续记录也很容易适应类似的实际成本汇总表。显然，还需要有与上文汇总表类似但也包含实际美元金额的汇总表。

60. 出资机构在提交拟议预算时希望看到的最终汇总表是按照住房或其他抽样单元进行的成本估计。通过与类似调查进行比较，这种数字还可以作为实际审议预算的界限数值。

表十四.8

按调查活动分列的会计科目分类成本计划占预算的比例：津巴布韦微型企业服务影响评估调查（1999年）（百分比）

	准 备	实 施	数 据	报 告	总 计
人员	3	65	5	2	75
交通	0	8	0	0	8
消耗品	0.9	9	2	0.1	12
其他	0.1	3	1	0.9	5
总 计	4.0	85	8	3	100

61. 这种汇总表有助于顺应调查实施期间的动态变化，因为它提供了一种全局观点，从而能够让人看到如何在一个领域减少支出，而在另一个更加急需经费的领域增加支出。调查设计、筹到的资金和实际执行情况中的变化都可以用这种方式进行调节。在实际实施微型企业服务影响评估调查（1999年）时，必须对拟议预算进行修改，主要是在人事成本领域，原因是通货膨胀出人意料地不断加剧。调查实施人员能够从实地调查（执行）活动的消耗品、交通和其他类别资金中转拨资金，以补充有正当理由的额外人事成本。下文表十四.9是这次调查的实际支出汇总表。

表十四.9

按调查活动分列的会计科目分类成本在实施调查期间占预算的比例：津巴布韦微型企业服务影响评估调查（1999年）（百分比）

	准 备	实 施	数 据	报 告	总 计
人员	3.3	69.3	5.6	2.5	80.7
交通	0	6.6	0	0	6.6
消耗品	0.6	7.1	2.1	0.1	9.9
其他	0.1	2.5	0	0.2	2.8
总 计	4.0	85.5	7.7	2.8	100

F. 潜在预算限制和缺陷

62. 无论对调查工作做出多么精心的规划，实地调查的具体情况绝不会与预期相符。事先了解这一点非常重要，这样就可以在最终预算申请中列入所谓的应急成本。这一类成本通常与前几节中所建议的方面合在一起在总成本中占一定比例：通常可以接受的应急成本为5-10%。

63. 除了列入一定比例的应急成本外，在对调查进行规划时还应充分了解当地状况，特别是在国内政治和/或经济局势不稳定的时候。应该在提交预

算建议时让出资机构了解这种可能性，并且通过在调查进行期间与出资机构保持不断联系，以便能够在发生预算超支的事件时，迅速向出资机构发出警告。这种事件既包括人为问题，也包括环境问题；并且地方政治、经济、天气状况、移民活动等问题也必须成为与出资和/或委托调查的机构保持持续联系的内容。

64. 例如，在津巴布韦1999年微型企业服务影响评估调查中，通货膨胀在数月内稳步上升，而且调查协调人员认为他们在编制调查预算时已经考虑到这一问题。但是，就在实地调查工作就要开始时，当局将美元汇率冻结在与实际不符的低水平上，因而与不断提高的通货膨胀率不相符；计划的成本变得完全不现实。幸运的是，国际管理系统对此表示体谅，并且允许增加成本以完成调查工作。

65. 如果出现上文所说的情况，调查实施人员可能有必要裁减人员，只保留效率最高的人员，或者以其他方式削减成本，例如，通过使用较低成本的文具、使用公共交通代替租用交通、合并业务以降低间接成本等方式。换句话说，如果调查的出资机构允许，最好在调查进程报告书中列入一项关于调查范围可能会因为意料不到的情况而发生变动的说明，例如，允许改变样本规模以便考虑到成本的不断增加。

G. 保存记录和摘要

66. 前文提过，如果想要了解已经做出的所有决定和在做出这些决定时所考虑过的备选方案，那么就很有必要对调查工作进行期间发生的事件做持续的日常记录，其中包括记录支出。

67. 调查协调人员应该在调查的准备阶段设计出一系列的表格供所有职员详尽记录日常活动和支出。这种表格应该包括工作的时间、完成的任务、访问细节、交通细节等，这些详细资料可以每周进行一次汇总。用这种方式，既能够保持密切注意预算，又能够在早期阶段查明可能出现的问题。另外，应该尽量建立只凭有效收据付款的支付制度。

68. 监测和报告实际的日常调查活动及其相应成本是一项重要的调查管理职责。不同的记录表格适合于不同的调查阶段。

调查设计

69. 在本阶段期间，调查管理人员将会与各种活动保持密切联系，从而使监测成为一项非常直接的任务。每日日记是记录谁做了什么的一种有用的方式，并且这项工作可以在周报中进行总结。有关交通、消耗品、住宿等实际成本的平行记录可以保留，并辅以每周汇总表，从而提供每周成本报告。本章附件提供了有关保存日记录和周记录的样表。

调查的实施

70. 在调查实施期间，调查管理人员在很大程度上需要依赖其实地调查组的组长向其提供有关他们的日常记日记活动以及记录的实际成本和收据方面的情况。管理人员还应每周进行一次汇总，详细说明所有成本及调查组成员的工作天数，以便于核对已经使用的预算比例。

调查报告

71. 调查管理人员还将与本阶段的各种活动保持更加密切的联系，对日常活动及成本记日记的制度将有助于保持每周汇总。附件中提供了将要使用的表格。

按预算跟踪支出

72. 对于负有根据预算对支出进行持续跟踪职责的人来说，最好应该每周提供一次最新支出概况及预算分配情况（例子见附件）。如果从调查期一开始时就出台这种机制，预测将会发生的问题是一件非常简单的事，而且如果有必要，可以申请对预算进行重新分配。调查从业人员应该认识到，一旦调查工作开始，增加预算就是一件不同寻常的事，因此，在拿出最终产品过程中，调整是取得成功的关键。

H. 结论

73. 本章的目的是通过详细考虑调查的各个组成部分，提供对调查预算规划有用的线索和建议。本章建议的一种动态方法包括从两个角度出发编制预算，并且举例进行了说明。

74. 仍要强调的是，要想成功实施一项透明、可靠和高质量的调查工作，这种详细规划至关重要。同样重要的是必须保留一切活动和行动的日常工作记录，因为这些记录可以顺利输入会计过程，并将作为一种可靠的记录予以保留，供今后调查规划使用。

附 件

保留日记录和周记录的样表

人员日常活动记录			
姓 名			
日 期	活 动	地 点	所用时间
总天数			

每日访调记录				
调查员姓名				
日 期	地 点	访调代号	所用时间	访调结果/评论

人员活动记录每周汇总			
组长姓名		报告日期	
人员姓名	活动总结	地 点	总 天 数
总天数			

人员日常支出记录

姓 名					
日 期	地 点	活 动	支 出 细 目	金 额 (美 元)	收 据 编 号
总金额(美元)					

每周支出记录

组长姓名			报告日期		
人员姓名	地 点	活 动	支 出 细 目	金 额 (美 元)	收 据 编 号
总金额(美元)					

每周支出汇总

姓 名						
项 目	第 1 周	第 2 周	第 3 周	第 4 周	第 5 周	第 6 周
人员						
工资/薪酬						
住所						
餐饮						
其他						
交通						
消耗品						
其他						

每周支出汇总*								
项 目	预 算	累积支出	第 1 周	第 2 周	第 3 周	第 4 周	第 5 周	第 6 周
人员								
工资/薪酬								
住所								
餐饮								
其他								
交通								
燃料								
车辆								
公交								
其他								
设备								
消耗品								
其他								
迄今总额								

* 可另建一电子数据表（例如，用EXCEL）。

参考资料

- Ajayi, D. (2002年)。《个人通讯》。
- Barnes, C. (2001年)。《津巴布韦Zambuko信托评估》。对微型企业服务影响力的评估项目。哥伦比亚特区华盛顿：国际管理系统。
- _____和E. Keogh (1999年)。《对津巴布韦小额供资方案影响力的评估：基线结果》。《对微型企业服务影响力的评估文件》，哥伦比亚特区华盛顿：国际管理系统。
- Greenfield, T. (1996年)。《研究方法：研究生指导》。纽约：约翰·威利国际出版公司，阿诺德，第306页。
- Groves, R. M. (1989年)。《调查误差与调查成本》。纽约：约翰·威利国际出版公司。
- _____和J. M. Lepkowski (1985年)。“双重框架混合模式调查设计”。《官方统计杂志》，第1期，第3号，第263-286页。
- Scheaffer, R. L.、W. Mendenhall和L. Ott (1990年)。《基本抽样调查》(第四版)。加利福尼亚，贝尔蒙特：华兹伍斯出版公司，第97页。
- 联合国 (1984年)。《住户调查手册 (修订版)》。“方法研究”，第31号，出售品编号：E. 83. XVII. 13。

E 部 分

调查数据分析

导 言

Graham Kalton

Westat

美国马里兰州罗克维尔

1. 在收集调查数据后，需要对调查数据进行准备以进行分析。数据准备包括三个重要的组成部分。首先，正如第十五章将要讨论的那样，在考虑到现有的计算工具和分析软件的情况下，需要确定如何安排数据才能最有效地进行分析。调查分析通常涉及两个或两个以上不同单位的分析：尤其是，在很多住户调查中，住户和人员属于各自独立的分析单位。因此数据文件应该能够有效处理分层结构。如，需要考虑到这样一个事实：人员归属于住户，而各住户的人数存在着差异。

2. 数据准备的第二个组成部分是数据整理或编辑。当然，调查反应肯定会存在着各种可以发现的误差。如，反应和其他反应不一致，或没有可能性。在进行分析前需要解决这些误差问题（数据整理的详细情况见第十五章）

3. 数据整理的一项重要工作是最后确定每个被抽样单位的分析状况。被选作样本的所有单位应纳入以下类别之一：受访者、合格的不答复者、不合格单位或是否合格还不清楚的不答复单位（见第八章）。一般来说，对诸如受访者方面所进行的分类，只有用于被抽样单位的调查问卷是不够的。通常情况下，需要为进行此种分类的单位收集最低数量的可接受数据。这样，确定反应状况就有必要对调查问卷进行分析。但应该注意的是，即使在分析中将某个单位纳入受访者的范畴，也很有可能存在着一些没有获得满意答复的项目。为了解决这一问题，可以采用某种形式的估算方法来为缺失的反应指定值。

4. 数据准备的最后一步要求确定所有被抽样单位的分析状况：计算调查的权数。需要计算出每个分析单位的调查权数。由于确定权数首先需要确定所有被抽样单位的抽选概率，所以很重要的一点就是需要在抽样过程中，认真记录抽选概率。被抽样单位的初始权数或基权数被计算为单位抽选概率的倒数。然后对受访者的基权数进行调整以补偿合格的不答复者以及一部分还不清楚是否合格的不答复者。通常情况下，还需要做进一步的调整，以便使某些重要变量在经过加权调整后的样本分布情况和来自外部来源的这些变量的已知分布情况一致。第十五章和第十九章对权数的形成进行了介绍。

5. 数据准备的一项重要工作就是要将分析所需要的抽样信息记录在每个受访者数据记录中。在分析期间，每个做出答复的单位都需要调查权数，以便

对调查人口的参数做出有效的估计。还需要有关每个做出答复的单位的初级抽样单位和层的信息，以便为调查估计准确计算抽样误差（见第二十一章）。

6. 有两个因素需要考虑，这两个因素使调查数据分析不同于标准统计文献所介绍的分析。一个就是需要在调查分析中使用调查权数以便补偿不均衡抽样概率、不答复和未覆盖情况。如果在分析中不使用权数，那么很可能使总体值的估计数失真。

7. 调查分析的第二个区别因素是需要计算调查估计的抽样误差，采用的方式是考虑调查的复杂样本设计。现行标准统计文献中所介绍的理论假设抽样为非限制性抽样，而大多数住户调查则采用分层的多级抽样。一般来说，根据分层多级样本进行估计的抽样误差要大于根据同等规模无限制样本进行估计的抽样误差，因此采用标准统计文献中的公式将会夸大估计的精确度（见第六章、第七章和第二十一章）。这就是说，根据标准统计软件包对调查估计所得出的标准误差估计是无效的。值得庆幸的是，如今有很多调查分析软件包可用来根据复杂样本设计中获得的调查数据做出合理的抽样误差估计。第二十一章对这些软件包进行了介绍。

8. 与政府调查一起进行的很多分析都属于描述性分析。通常情况下都以表格形式对结果进行报告，表格单元中包括平均值、百分数和总数。有时，这些数值采用图示法显示。从狭义的统计术语来说，有关估计通常都很简单，惟一的问题就是需要确保已经使用了调查权数。但还应该考虑定义和说明方面的重要问题。需要对将要度量的概念进行周密的界定（如，贫穷：见第十七章），并且根据现有统计目的以合适的方式对需要度量的一组单位进行详细说明。另外，对结果的说明应该能够明确说明度量的对象是什么，并说明是为哪一组单位进行度量的。至于对简单描述性估计进行说明的指南，见第十六章。

9. 就调查反应而言，对所要度量的概念进行合理界定通常都比较简单。但有时概念会更复杂一些，因而可能需要采用多元统计方法确定一个指数来对有关概念进行度量，如：类集分析和主成分分析。第十八章介绍了几个例子，如其中一个例子是根据一些变量的有关信息——如，住户是否有电、每个卧室的人数以及饮用水主要类型——来决定“财富”指数。

10. 最后，应该指出的是，尽管进行描述性估计仍然是调查分析的主要形式，但已越来越多地使用各种分析技巧来对调查数据进行分析。这些技巧通常用来研究各变量之间的关系，探讨可能存在的因果关系。这类分析的最常见形式就是：建立一个统计模型以便根据一组自变量（或预测变量）对某个因变量进行最佳的预测。如果属于连续性因变量（如，家庭收入），那么可以采用多元线性回归法。如果属于二元反应的分变量（如，住户有自来水或没有自来水），那么可以使用逻辑回归方法。第十九章和第二十章对这些方法以及复杂样本设计对这些方法的影响进行了介绍。第十九章还介绍了多层次建模在调查中的使用，第二十章讨论了复杂样本设计对各分类变量之间关联性的标准卡方检验产生的影响。

第十五章

住户调查数据管理指南

Juan Muñoz

Sistemas Integrales

智利圣地亚哥

摘 要

本章对数据管理在设计和实施全国住户调查中的作用进行了介绍。首先，对数据管理和调查问卷设计之间的关系进行了讨论。其次，对过去、现在和未来录入和编辑调查数据的各种选择及其对调查管理的总体影响进行了探讨。后面的部分为质量控制标准的定义、复杂全国住户调查数据录入程序的编制以及调查数据集的散发提供了准则。最后部分讨论了数据管理在支持调查样本设计方面所起的作用。

关键词：一致性检验、数据整理、数据编辑、数据管理、住户调查、质量控制标准。

A. 导言

1. 尽管一再强调数据管理在住户调查中的重要作用，但通常情况下还是会将数据管理视作与调查制表阶段有关的一组任务，换句话说，数据管理是调查项目接近尾声时所进行的活动，这些活动使用调查总部数据整理办公室的计算机，一般由数据分析师和计算机程序设计人员进行控制。

2. 对调查数据管理的这种局限性看法正在改变。过去二十多年的经验显示数据管理能够而且应该在调查工作的最初阶段起到关键作用。另外显而易见的是，数据管理并不随着初次统计报告的公布而结束。

3. 世界银行进行的生活水平衡量研究以及成功地将基于计算机的质量控制和调查实地工作结合起来的调查，显然证明了分析阶段前进行数据管理的成效。即使数据录入不是实地工作的一部分，数据管理员也应该参与调查问卷的设计，以确保调查所观测的统计单位能够被正确识别；访调员所使用的跳过指令既明确又正确；特意做出的备份最终被纳入可在以后用来有效控制一致性的调查问卷。

4. 有人认为在调查项目结束时，调查产生的最终产品是印刷出版物和一系列统计表，如今，这种观念已经被数据库观念取代，不仅统计机构可以采用这种数据库来编制初始统计表，而且研究人员、决策人员和普通公众也可以使用这种数据库。有关调查结果的描述性简要报告不再被视为最后一个步骤，而是各种分析工作的入手点，这种工作在项目正式结束和调查小组解散后还可能持续多年。

5. 本章将首先讨论调查数据管理和调查问卷设计之间的关系，然后对过去、现在和未来录入和编辑调查数据的各种选择以及各种选择对调查管理的总体影响进行探讨。以后各部分将为质量控制标准的定义、复杂全国住户调查数据录入程序的编制以及调查数据集的散发提供指导。最后部分将讨论数据管理在支持调查样本设计方面所起的作用。

B. 数据管理和调查问卷设计

6. 调查数据管理与调查问卷设计同时开始，并对后者有着很大的影响。在起草每份重要调查问卷草案时，都应征求数据管理员的意见，因为数据管理员尤其能够敏锐地发现观测单位的定义、跳过模式等方面存在的缺陷。在这方面，调查问卷一些格式应予以注意，本节将对此进行探讨。

7. 被观测统计单位的性质与标识。每个住户调查不仅要收集主要统计单位——住户——的信息，而且要收集住户内各种从属单位——人员、预算项目、土地、农作物等——的信息。调查问卷应该明确清晰地说明这些单位具体是什么，并确保被观测的每个单位都正确地有一个独特的标识。

8. 有关住户的信息一般放在调查问卷的封面。住户信息有时包括一系列冗长的数字和字母，这些数字和字母代表地理位置和用来选取住户的抽样程序。虽然这似乎是显而易见的，但将这些代码作为住户标识时，应该进行认真评估，因为这会很烦琐、容易产生错误并且费用很高（仅仅对几百家住户样本进行标识通常就需要20或20位以上的数字）；有时，这种做法甚至不能保证每个单元的独特标识，如当封面上的地理代码识别出某一住所，但未能将某一住所内多家住户的情况考虑在内。另一种简单安全的方法就是通过手写或印在调查问卷封面上的、甚至由印刷所事先印制的简单序号来识别住户。这样，封面上的地理位置、城市/农村状况、抽样代码以及其余数据就成为住户的重要属性，这些属性必须包括在调查数据集内，但不一定是为了识别之目的。处于这两个极端（列出所有详细的抽样代码和简单的住户序号）之间的一个不错的折中办法是：调查中的初级抽样单位（PSUs）采用三位或四位序号，每个初级抽样单位内的住户采用两位数字号。

9. 从属统计单位的性质通常是显而易见的（如，住户成员为个人），但如果统计单位看似属于一个单一的单位，但事实上却是另一种具有多重属性的单位，那么其性质就不明确。如，在要求一个人对其工作的主要活动进行报告时，如果他同时从事着多种同等重要的活动，或者在某个特定的基准期拥有一份以上的工作，那么就可能出现这种情况。同样，在要求一位妇女报告其最后

一个孩子的性别或体重时，如果她最后生育的孩子是体重不同的一男一女双胞胎，那么也有可能出现含糊不清的情况。当然，可以通过周密的调查问卷设计和试用来防止这种情况的出现，但这种情况的出现通常不明显，正是在这一方面，有经验的数据管理员的重要远见可为调查项目的专家提供宝贵协助，协助他们发现这些情况。

10. 不管性质如何，都应该对住户内的从属单位进行独特的标识。可以通过访调员分配数字代码的方式来实现这一目的，但一般来说，如果可能，最好事先将这些识别码印在调查问卷上。

11. 固有的重复。设计调查问卷时可考虑列入有意重复的内容，旨在检查访调员的错误或数据录入错误。最常见的例子有：

- 在包含货币金额的栏下，为“总计”加一条底线。通常由访调员来进行这一类的加总工作，但即使并非由访调员负责加总工作，列入这些“总数”也是方便的，因为它是发现数据录入错误或遗漏的有效方法（通常也是惟一的方法）。事实上，可以在任何一个数据栏的底端加一栏“总计”以进行质量控制，即使数据的加总不能对某个量进行有意义的度量也是如此（如，对于所购买的各种食品，可以在其数量栏（而不是货币金额栏）的底端加一栏“总计”，即使所加总的是不同类别的数据也是如此，如，若干公斤的面包，若干公斤的土豆，甚至若干升的牛奶）。以下讨论印刷检查时将对此一点作进一步的阐述。
- 对某些重要变量（如某个人的职业或活动，或消费项目的性质）的代码加一个校验码。校验码是可以根据代码其余数位计算出来的数字或字母，可以在数据录入时通过算术的方法计算。常见的校验码运算方法如下：将代码最后一个数位乘以2，倒数第二个数位乘以3，依此类推（如果代码为六位数以上，那么可以按顺序重复以2、3、4、5、6、7作为乘数进行计算），然后将结果相加。校验码就是这一总数与11的第二小倍数之差（10通常由字母K表示）。确定校验码运算方法的目的是为了让较常见的编码错误（如，数位颠倒或数位遗漏）将会得出错误的校验码。

C. 数据录入和编辑的操作策略

12. 很多住户调查仍然将数据录入和编辑视为继实地调查之后由中心机构进行的活动，而其他调查则已经将数据录入工作纳入实地活动中。在不久的将来，可能会朝着计算机辅助访谈的方向发展。本部分将讨论各种策略所具有的组织意义，以及每种选择下数据录入和编辑软件的共同特征和独特属性。

13. 集中式数据录入。在微型计算机出现之前，集中式数据录入是惟为大家所熟知的方法，如今很多调查仍然使用这种方法。这种方法将数据录入视为一个工业生产过程，数据录入将在访谈结束之后在数据录入部门集中进行。其目的是为了将原材料（纸质调查问卷上的信息）转化为中间产品（可由机器

阅读的文件），应（使用编辑程序和文书工作）进一步改进中间产品，从而获得最终产品，也就是所谓的纯净数据库。

14. 在首次数据录入阶段，首要的问题就是要确保速度，确保文件上的信息能够完全反应调查问卷中所收集的信息。数据录入人员不应“思考”其工作内容，而只需如实录入提供给他们的数据。有时，调查问卷采用双盲数据录入方式，以确保正确录入。

15. 截止1970年代中期，数据录入由性能非常有限的特殊机器进行。目前这一过程几乎都是由微型计算机进行，尽管可以在计算机上安装质量控制检验程序，但实际上却很少这么做。目前最普遍的观点就是：很少在数据录入过程中进行质量控制检验，因为操作人员并没有接受过这方面的培训，因而无法在发现错误时，做出决定来采取相应的措施。另外，发现错误并找出解决方案会使数据录入过程放慢。持有这种观点的学派认为应该将质量控制检验单独纳入编辑过程。

16. 实地数据录入。从1980年代开始，将基于计算机的质量控制纳入实地工作已经被看作改善住户调查质量和及时性的关键之一。这种思路最初由世界银行生活水平衡量研究（LSMS）调查提出，之后在其他各种复杂的住户调查中得以运用。根据这一策略，数据录入和一致性控制是逐户进行的，并以此作为实地工作的一部分，这样就可以通过最后回访住户来解决各种误差和不一致问题。

17. 将基于计算机的质量控制纳入实地工作，其最重要和最直接的好处就是可以大大提高信息的质量。因为各种错误和不一致问题可以在访调员仍然在实地的时候加以解决，而不是等到数据整理办公室“整理”数据时才得以解决。另外，由于办公室进行数据整理不仅费时而且程序冗长，所以其整理的数据库至多只是保证数据内在的一致性，但不一定能够反应实地所观测到的事实。这种不确定性是由种种决定造成的（一般情况下，这些决定是没有记录的），因为这些决定是在远离数据收集所在地的情况下做出的，而且是在收集数据很长一段时间后才做出的。

18. 将基于计算机的质量控制纳入实地工作所产生的数据库可以及时用作制表和分析，一般情况下，实地工作结束后的几个星期内就可以使用有关数据库。事实上，甚至可以在进行调查的时候准备数据库，这样调查管理人员就可以对实地工作进行有效的监管。

19. 将基于计算机的质量控制纳入实地工作的另一个间接好处是：它可以促进所有访调员在收集资料的整个过程中采用统一的标准，在采用这种方法前，实际上很难做到这一点。计算机确实是调查监督人员的好助手，它不仅廉洁而且不知疲倦。

20. 将基于计算机的质量控制纳入实地工作还对调查的组织工作有着各种意义，其中最重要的是：它要求实地工作人员组成各种小组。实地工作小组通常由一位监督人员负责，除了2-4名访调员外，还包括一位数据录入员。

21. 实地工作的组织取决于现有的技术。两个最常用的配置包括台式计算机和笔记本电脑，并需要采取以下步骤：

- 让数据录入人员在固定的地方（一般是统计机构的区域办公室）用台式计算机工作，对实地工作进行组织，以便让小组的其余人员对每个调查地点（一般为初级抽样单位）至少访问两次，从而使数据录入人员能够有时间在两次访问期间录入数据并对数据的一致性进行检验。在第二次及以后的访问期间，访调员将针对数据录入程序发现的错误、缺漏或不一致问题，重新提出问题。
- 让数据录入人员用笔记本电脑工作，并让该人员和小组其余人员一道访问调查地。整个小组人员将一直呆在调查场所，直到所有数据都被录入并且所有数据都经过数据录入程序证实为完整和正确时为止。

22. 上述两种选择都需要具备各种外围条件，调查策划人员和管理人员应认真考虑这些条件。其中一个条件就是需要确保计算机拥有不间断的电源，在供电状况差的地区，这可能是一个问题。如果在固定地点使用台式电脑，那么可能需要安装发电机，这样就需要一直备有发电机燃料。如果使用可以移动的笔记本电脑，那么可能需要使用便携式太阳能电池板。

23. 两种策略一个显著的重要差别就是：如果将基于计算机的质量控制纳入实地工作，那么需要在调查开始前编制和调试数据录入和编辑程序。至于集中式数据录入，这么做也可以带来方便（其目的是为了让数据录入能够与实地工作同步进行），但却不是绝对必要的。

24. 无纸访谈。使用手提式电脑而非纸调查问卷非常具有吸引力，因为它的优点是：访谈的某些部分可以自动进行，如跳过指令。尽管这种技术已经使用了近20年，但发展中国家却很少采取措施来将这种策略用于复杂的住户调查。事实上，到目前为止，即使在最先进的国家统计机构中，无纸调查也只是限于一些相对简单的工作，如就业调查、收集用于消费价格指数的价格资料。

25. 其中一个可能的原因是，尽管无纸提问有助于遵循有始有终的线性流程的访谈，但发展中国家和转型国家进行的很多住户调查却有可能要求对每个住户进行多次访问，对住户的每个成员单独进行访谈，或需要一些结构不太严格的其他程序。

26. 尽管缺乏真正的经验，但可以通过一些观测来了解无纸调查问卷的设计和应该考虑哪些问题：

- 在有些情况下，数据录入程序界面可以包括一系列逐一出现在计算机屏幕上的问题，但在其他情况下，可能需要复制纸调查问卷的结构和直观格式，以便在同一时间显示出很多数据录入领域。在有关支出和消费的模块中，这一点似乎尤其重要，因为访调员需要同时“看到”很多消费项目。在有疑问的情况下，界面还必须能够使有关人员得以标出有关问题，并且也应当使返回住户进行第二次访谈而无需重复所有问题成为可能。

- 调查问卷的设计一般需要数月的时间，并需要不同领域众多人员的参加（调查项目专家、从事调查的人员等）。对于纸质调查问卷来说，需要对各“代”调查问卷进行准备、分类、讨论和试验，直至确定最后版本的调查问卷为止。对于某些实际上永远不会出现在纸质调查问卷上的内容，也应确定对其采取的相应步骤。
- 必须围绕新技术，对访调员的培训进行重新设计。我们知道如何对访调员培训（通过理论研讨会、情景模拟、模拟访谈和培训手册等），从而让他们能够根据纸质调查问卷进行调查。但我们却很少花精力开发无纸调查技术。
- 最后，必须开发各种有效的监管方法。经过半个多世纪的演变，已经有很多方法（对调查问卷进行目视检查，对访谈进行核查等）可以对实地访调员的工作进行核实。但所有这些方法都是围绕纸质调查问卷的概念进行的，因此需要对其进行重新设计，以便将其用于无纸访谈。新技术将能为有效的监管提供各种截然不同的，甚至更为有力的选择方案，这是极有可能的。如，除了记录数据文件外，很多手提式电脑还具备语音录制功能，这种功能可自动记录访谈的随机部分。通过添加全球定位系统（GPS）性能，还可以自动记录访谈的时间和方位。同样，需要对各种细节进行说明并对其进行实地检查，并将有关细节纳入调查实地工作的一般计划中。

D. 质量控制标准

27. 不管是选择哪种策略进行质量控制，调查问卷上的数据需要接受五种检验：范围检验、根据参照数据检验、跳过检验、一致性检验和印刷检验。本部分将回顾这些检验的性质，以及在各种操作配置下实施这些检验的方式。

28. 范围检验旨在确保调查的每个变量所包括的数据处于有效值的有限范围内。在调查问卷事先为分类变量确定的各种值中，分类变量只能拥有其中的一个值（如，性别代码只能是“1”代表男性，或“2”代表女性）；按年代顺序排列的变量应该包括有效日期；数值变量应该在所规定的最小值和最大值范围内（如年龄在0-95之间）。

29. 范围检验的一个特殊情况是：根据外部参考表对来自两个或两个以上密切相关领域的数据进行检验。以下为一些共同的情况：

- 人体测量数据的一致性。在这种情况下，将根据世界卫生组织的标准参考表对记录的身高、体重和年龄值进行检验。如果标准指标的值（年龄标准身高、年龄标准体重及身高标准体重）偏离常态三个以上的标准差，那么就有可能存在着误差，应该对该值做出标记，以便重新进行测量。
- 食品消费数据的一致性。在这种情况下，将根据具体项目的单价表对所记录的食品代码值、购买的数量以及支出的金额进行检验。

30. 一般来说，即使在中心机构录入数据，也最好在数据录入的初始阶段对范围错误进行检验和纠正，而不是将这种质量控制一直推迟到编辑阶段，因为范围错误通常是由于数据录入本身造成的而不是访调员的错误。在录入超出范围的值时，可能会启动错误标记（如，发出“嘟嘟”声或屏幕上出现闪光区域）。如果仅是打印错误，那么数据录入人员可以立刻进行纠正。但如果录入的值与调查问卷上的值一致，那么就可以不考虑该错误标记。在这种情况下，应该对错误进行报告，以便事后由文书人员对调查问卷进行检验，并纠正有关错误（如果是在实地录入数据，那么可以由访调员在第二次访谈期间纠正有关错误）。同时，有疑问的数据项目可以用一种可以记录疑问状况的特殊格式进行存储。

31. 跳过检验。这种检验是为了核实跳过模式是否被正确执行。如，可以通过简单的检验来核实：如果一名儿童对第一个问题——“是否入学”的回答是“没有”，那么针对这名儿童，一些只向学童提问的问题是否没有被记录下来。还可以通过更复杂的检验来核实是否为每个受访者填写了调查问卷的适当单元。根据受访者的年龄和性别不同，住户的每个成员应该就调查问卷中的具体部分进行回答（或跳过不回答）。如，年龄在5岁以下的儿童所回答的问题应属于人体测量部分，不应向他们询问职业方面的问题。年龄在15-49岁之间的女性可以纳入生育问题部分，但却不能将男人包括在这一部分。

32. 未来的某个时候，发展中国家将会在调查中普遍采用计算机辅助（无纸）访谈，到那个时候，至少在某些情况下，可以通过数据录入程序本身来控制跳过方案。但按照其他操作配置（数据录入中心机构，实地数据录入），数据录入程序实际上不应该自动进行跳过。如，如果对于“你入学了吗？”这个问题所录入的答案是“没有入学”，但一些需要就所在学校类别和年级等录入数据的部分仍然应提供给数据录入人员。在这种情况下，如果答案实际上已记录在调查问卷上，就可以录入答案，程序将标出“跳过不正确”的记号。监管人员或访调员（或集中编辑的文书人员）可以在稍后的时间确定错误的性质是什么。但，答案“没有入学”也很有可能是“已经入学”。如果数据录入程序自动跳过以后的字段，就无法发现和纠正错误。

33. 一致性检验。这些检验的目的是为了核实一个问题的值和另一个问题的值是否保持一致。如果各种值都来自同一统计单位，那么只需要进行简单的检验，如某个人的出生日期和年龄。更复杂一些的一致性检验涉及对来自两个或两个以上不同观测单位的信息进行比较。

34. 一致性检验的数量是多少，并没有自然限度。在复杂的住户调查中，一个好的数据录入程序可能会有几百个检验项目。一般来说，检验的项目越多，最终数据集的质量也就越高。但由于编写数据录入和编辑程序的时间总是很有限（通常情况下在两个月左右），所以需要专业的知识和良好的判断来决定应该对哪些项目进行检验。某些一致性检验项目适用于几乎所有住户调查，经验证明，这些一致性检验特别有效，因而差不多成了实际的标准。这些项目如下：

- 住户的人口一致性。对所有住户成员年龄和性别方面的一致性进行检验以明确亲属关系。如，父母应比孩子至少年长（假定）15岁，配偶的性别应该不一样等。
- 职业一致性。某些内容的有无应该和住户成员自报的职业一致。如，如果某些成员在劳动力一栏所报告的职业为农民，那么就应该有耕作部分的内容，而且也只有在这种情况下才应该有耕作部分的内容。
- 年龄和其他个人特征之间的一致性。可以就每个人年龄和个人特征之间是否一致进行检验，个人特征如：婚姻状况、与户主的关系、目前正在上学的孩子所在的年级（用于目前正在校的孩子）或最后一次上学时所在的年级（用于已经辍学的成员）等。如，8岁的孩子所在的年级不应在三年级以上。
- 支出。在这种情况下，可以进行若干不同项目的一致性检验。在一个住户中，只有当一个或一个以上的个别记录显示出某个孩子在上学的情况时，才能在诸如教科书和学费方面的住户消费记录中出现正数。同样，只有当住户拥有与电有关的服务设施时，才能报告与电有关的支出。
- 控制合计数。如前文所述，在可以添加一系列数值的情况下，添加控制合计数是进行调查问卷设计的一个有益原则。数据录入程序应该校验控制合计数是否等于个别数值的总和。

35. 印刷检验。在早些年的调查数据处理中，打印错误检查几乎是数据录入时所进行的惟一质量控制手段。为此，一般由两位不同的操作人员对每份调查问卷录入两次。如今则很少使用这些所谓的双盲数据录入办法，因为如今可以采用其他一致性控制措施，这些措施的采用使得印刷检验没有必要。但在某些情况下，这可能是一种美好的愿望，而非可靠的假设。

36. 典型的印刷错误是在数字录入时出现数位颠倒的问题（如录入“14”而非“41”）。在对婚姻状况或家庭关系进行一致性检验时可以发现类似这样的年龄错误。如，如果某个已婚或寡居人士的真实年龄为41岁，但由于录入错误，出现在调查问卷中的年龄却为14岁，那么在根据婚姻状况检验年龄时就会出现一个显示该错误的标记。但如果在每月的肉类支出方面出现同样的错误，那么很可能不容易发现，因为14美元和41美元都是有效的金额。

37. 这说明将数据管理观念纳入调查问卷设计阶段是很重要的。如，控制合计数可以大大减少印刷错误，因为要求访调员用计算器对数据加总的方法和双盲数据录入办法类似。对于某些重要的变量，可以采用校验码来实现这一目的。调查问卷某些部分的数据录入也可以采取真正的双盲数据录入法，但整个调查问卷都采取这种办法是没有必要的，也是不切实际的——除了其他原因外，这是因为现代数据录入一般是由一个数据录入员进行，而非两个不同的数据录入员进行。

E. 数据录入程序的开发

38. 编写一个好的调查数据录入和编辑程序既是一门技术，又是一种技能。本部分讨论目前可用来促进这一进程的技术层面的某些开发平台，并探讨一些较微妙的问题，这些问题与数据录入人员以及调查数据集未来用户的界面设计有关。

39. 开发平台。市场上有很多数据录入和编辑程序开发平台，但适合复杂住户调查中数据管理要求的开发平台却很少。世界银行于1990年代中期进行的一次审查发现当时适用两种基于DOS的平台：世界银行内部开发的生活水平衡量研究（LSMS）软件包及美国普查局微型计算机集成处理系统（IMPS）程序。自此次审查以来，为了适应不断变化的硬件和操作系统环境，两种平台都进行了改进。微型计算机集成处理系统已经由普查与调查处理系统（CSPRO）取代。普查与调查处理系统是一种基于Windows的应用程序，这种应用程序除了充当数据录入和编辑程序的开发环境这一基本角色外，还具备某些制表功能。生活水平衡量研究软件包已经朝着LSD-2000演变，这是一种基于Excel的应用程序，其目的是为了同时开发调查问卷和数据录入程序。

40. 经验证明，在很多国家中，CSPRO和LSD-2000（或其前身）能够为许多国家的复杂全国住户调查的有效数据录入和编辑程序的开发提供支持。这些平台不仅容易获得而且容易使用。几乎任何一个程序设计人员——事实上，几乎任何基本了解计算机的人——都可以在几个星期内掌握初步开发工作数据录入程序所需要的技能。

41. 设计原则。遗憾的是，对于需要开发什么样的数据录入程序，开发平台不能在这方面为程序设计人员提供建议。甚至可以说，平台的用户友好性面临的风险是极易开发出不适用的数据录入程序。将工具的掌握和善于利用工具的技巧混淆起来是错误的，调查管理人员应该避免这种错误。为避免这种错误，调查管理人员应让有经验的程序设计人员和调查项目专家共同参与调查数据录入和编辑程序的开发。在这一方面，提供一些切实可行的准则是有益的：

- 数据录入屏的设计。数据录入屏应该尽可能和调查问卷中相应的页面相同。譬如，如果调查问卷以矩阵的方式提出个人问题（矩阵的“行”为问题，“列”为住户成员，或“行”为住户成员，“列”为问题），那么最好为每个人准备一个单独的数据录入屏，而不是将纸格子复制在计算机屏幕上。不将整个格子复制在计算机屏幕上的一个原因之一是受访者的数量是可变的，另一个更重要的原因就是所观测的统计单位是人而非住户。
- 区分不可能和未必可能的情况。如果从逻辑上讲或在自然情况下，某种情况是不可能的（如，女儿比其母亲年长），那么数据录入程序应该将其标记为错误。如果某种情况在自然情况下不是不可能的，但可能性却非常小（如女儿比其母亲小15岁以下），那么数据录入程序也应

该对这种情况做出反应。数据录入程序最好能够像监管人员一样，能够通过调查问卷的目视检验，对错误的严重程度做出评估，然后根据错误的严重程度做出不同反应。在将数据录入纳入实地工作时，这种“智能”程序尤其重要。遗憾的是，有些程序设计人员在这方面没有投入足够的精力。一个明显的迹象就是总是将定量变量的上限定义为“999……”（数据录入区有多长，就应该有多少个9）。这种做法的反作用是显而易见的：数据录入区当然应该足够长，以便能够录入尽可能大的值，但上限应该足够小，以便将未必可能的值标记为异常值。

- 报告错误所使用的语言。数据录入程序中的某些质量控制标准可以报告较简单方法所发现的错误，这些简单方法要么是不言自明的，要么是稍加培训就能够理解的。如，生活水平衡量研究数据录入程序通过在有疑问的值旁边显示向上“↑↑↑”或向下“↓↓↓”的闪光箭头来报告范围检验错误，箭头向上还是向下取决于问题值是偏高还是偏低。如，在对住户人口的一致性进行检验后，可能会出现这样一句话：“警告：露西（身份代码05，女，9岁）不太可能是玛丽的女儿（身份代码02，女，21岁）”，这些文字最好能够打印出来，而不仅仅是在计算机屏幕上出现。相对于一些看似较简单的办法（如采用错误代码）而言，这种“具有写作能力的智能”程序可能花费更长的时间编写，但这种程序可以节省很多实地工作时间和实地人员培训时间，还可以免去程序设计人员编写错误代码本的麻烦。
- 变量代码。一项复杂的住户调查通常包括数以百计的变量。负责数据录入程序的程序设计人员需要根据所用开发平台的具体习惯，通过代码来查阅这些变量。为此目的，应该在数据录入程序开发之初选择一个简单合理的编码系统，因为这将有助于开发小组成员之间进行交流，还有助于在编制和散发调查数据集的稍后阶段节省时间。但寻找一个好的编码系统可能要比想象的更难。这一过程的开始阶段可能较容易，比如，可以为头几个变量确定诸如“AGE”、“GENDER”之类的代码，但这一过程很快就会变得难以管理，因为寻找足够的助忆码越来越难了。一种不错的选择是干脆参照调查问卷上的有关部分和问题编号来进行编码，而不管代码是否一目了然。（例如，如果“年龄”和“性别”这两个变量分别是第1部分第4个问题和第5个问题的变量，那么其编码可分别为“S1Q4”和“S1Q5”。）
- 数据录入工作量。在将数据录入纳入实地工作时，数据录入最自然简单的工作单位就是住户。这是因为在这些条件下，数据录入人员所处理的总是一份或几份调查问卷，此外还因为一致性控制措施以及错误的报告是逐户进行的。在数据录入中心，工作量可以为若干区组，每个区组为10-20个住户（如调查地或初级抽样单位）。这么做的原因是：

(1) 每一区组应该由一个数据录入人员单独在一个计算机上录入，录入至多需要几天时间；(2) 随时都应当轻松地存储和检索调查问卷的相应内容。

F. 调查数据集的整理和散发

42. 调查数据集的结构必须能够反应调查中被观测统计单位的性质。换句话说，来自复杂住户调查的数据不能按照下表十五.1所显示的形式进行存储，也就是说，不能存储为简单的矩形文件，在其中调查问卷上，每个住户占一行，每个变量占一列。

表十五.1

住户调查的数据被存储为简单的矩形文件

	变量 1	变量 2	...	变量 j	...	变量 m
住户 1			
住户 2			
...
住户 i			...	数据 i, j	...	
...
住户 n			

43. 如果有关住户的所有问题都属于一个统计单位，那么这种结构（又称为“平面文件”）是可行的。但正如以前所讨论的那样，情况并非如此。有些问题涉及从属统计单位，每个住户内的这些从属统计单位数量是可变的，如人员、农作物和消费项目等。将每个住户成员的年龄和性别存储为不同住户级别的变量不仅浪费（因为所要求的变量数目将取决于最大住户的规模，而非平均住户规模），而且在分析阶段也非常麻烦（因为即使简单的工作，如获得年龄-性别分布情况，也需要费力地对每个住户内数量可变的年龄-性别组进行查找）。

44. 处理众多不同的统计单位很复杂，CSPRO和LSD-2000两个平台所使用的文件结构都能很好地处理这种复杂的问题，此外还能尽量减少存储要求，在分析阶段也能和统计软件很好地结合。

45. 数据结构应该使所观测的每个统计单位和计算机文件记录之间保持一对一的关系，为此，每种统计单位需要采用不同的记录类型。如，为了对住户花名册上的数据进行管理，应该为花名册上的变量确定一个记录类型，和个别项目相应的数据应该单独存储在该类别下的记录中。同样，在食品消费单元

中，应该有一个记录类型和食品项目相对应，与个别项目相应的数据应该存储在该类型下的单独记录中。

46. 每个记录类型下的记录数量可以不同。这样，由于无需将每个记录保存成最大文件，从而可以节省所需的存储空间。

47. 从原则上讲，每个统计单位只需要一个记录类型，但有时由于实际情况的需要，可能会为同一单位确定一个以上的记录类型。如，有关教育和健康的问题可能需要存储在两个不同的记录类型下，尽管这两种情况下的统计单位都是个人。

48. 每个记录都标有一个独特的代码，代码分三个或三个以上部分。第一部分是“记录类型”，记录类型出现在每个记录的开始，其作用是为了说明信息的来源，如说明信息是来自封面、健康单元还是食品支出。在所有记录中，记录类型的后面为住户号码。在大多数的记录类型中，还需要有第三部分识别符，第三部分的识别符用作区分住户内同一类别的各统计单位。如，人员识别码或支出项目代码。在某些情况下，所观测的单位只有一个，在这种情况下，第三部分的识别符就没有必要。如，住房特征通常只按照一家只有一个住处的情况进行收集。在有些情况下，可能还需要增加第四部分代码。如，第三个识别码可能代表住户企业，第四个代码可能代表每个企业所拥有的每种设备。

49. 紧跟识别码之后的是调查中为每个具体单位所记录的实际数据，实际数据按调查问卷中有关问题的顺序记录在长度固定的区段中。所有数据都按美国信息交换标准代码（ASCII）的标准格式存储。

50. 调查数据集只能按照独立的平面文件进行组织（每个记录类型一个文件）以方便散发，因为原有结构中长度固定的字段格式足以将数据导入标准的数据库管理系统（DBMSs）做进一步的处理，或导入标准的统计软件进行制表和分析。将数据导入DBMS很容易，因为原有结构几乎可以直接将被其接受为个别表格的录入信息转换为标准的数据库格式（DBF）（在这种情况下，记录识别符将充当各表之间的自然关系联系纽带）。

51. 数据集的散发还要求每个记录类型的结构被正确记录在所谓的调查代码本中，代码本应提供给任何对数据集感兴趣的用户，它应该明确说明记录中每个变量的位置和长度。对于分类变量来说，也应该对编码作具体说明。下图十五.1所显示的是尼泊尔生活水平调查代码本（某些变量的编码被省略）。

52. CPro和LSD-2000这两个平台都可以编制调查代码本，作为数据录入程序开发过程的副产品。LSD-2000还可以提供界面将数据录入文件转换为DBF文件，并将数据导入最常用的统计软件中（Ariel、CPro、SAS、SPSS和Stata）。这说明有必要在数据录入程序开发阶段认真确定一个变量编码系统：如果做好了这项工作，那么只要有数据集，调查分析员就能够立即使用调查数据。

G. 抽样过程中的数据管理

53. 本节将讨论数据管理在住户调查样本的设计和实施工中所起的作用，并为抽样标架的计算机化以及初级抽样的进行提供了建议，包括适用于隐含分层以及初级抽样单位的概率与规模成比例（PPS）抽样的实用方法。至于开发一个将倒数第二个抽样单位作为其前一个抽样阶段副产品的数据库，本章对其进行了讨论，并强调它在实地调查中作为管理工具的作用，以及其内容如何可以通过实地生成的信息（如，列出的住户清单以及不答复数据等）进行更新，以便得出用于分析阶段的抽样权数。

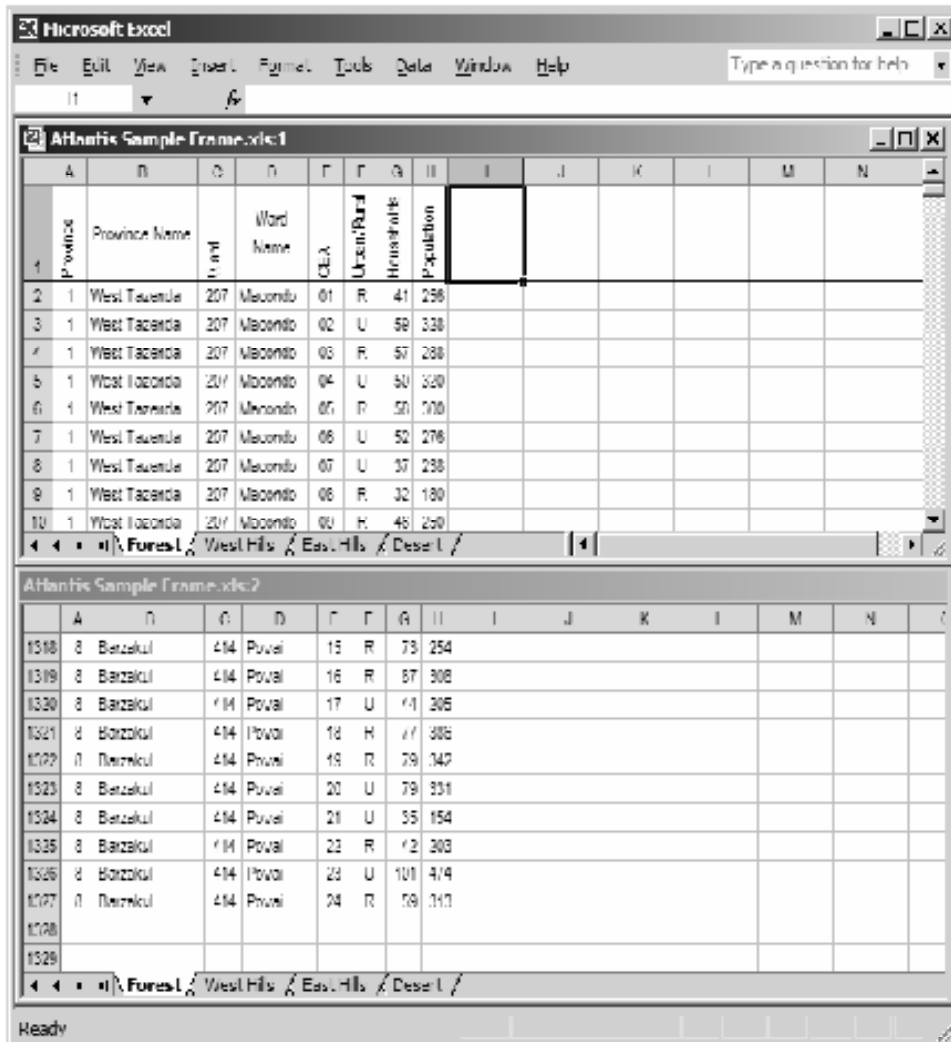
54. 初级抽样标架的组织。很多住户调查的初级抽样单位是大多数全国普查确定的查点区（CEA）。为确定初级抽样标架，一个简单而有效的方法是创建一个包括全国所有查点区清单的计算机文件。除非一个国家有大量查点区（如孟加拉国家有80 000个以上的查点区），否则实现上述目的的最好方法就是采用诸如Excel之类的电子表格程序，每个查点区占一行，需要的所有信息放在“列”中。电子表格程序应该包括有关查点区的全部地理识别信息以及度量查点区大小的指标（如，人口、住户数量或住所数量）。一般来说，如果能为每个样本层创建一个不同的工作表，那么会更方便一些。下图十五.2显示初级抽样标架在某个假设国家“森林”层中的样子（Excel屏被分为两个窗口以同时显示第一个和最后一个查点区）。

55. 在本例中，“森林”这一层中的1 326个查点区是通过地理代码和国家行政区划（省和行政区）的名称以及每个行政区内的序号识别的。抽样标架还包括普查时每个查点区的住户数量和人口，并显出查点区是城市还是农村。

56. 在进行抽样以后步骤之前，很关键的一点是要核实抽样标架是否完整和准确，采用的办法是根据统计机构公布的普查总数核对人口数字。另外，重要的是核实查点区的规模是否大得足以用作初级抽样单位。如，如果样本设计要求倒数第二阶段的每个类集为25家住户，那么如果查点区的住户不足25家，就无法满足该要求。在这种情况下，小的查点区应该和邻近地区的查点区结合在一起以构成初级抽样单位。如果必须通过人工、通过不断参阅普查图来搜寻邻近的查点区，那么这一过程会冗长而缓慢。但由于统计机构通常根据某种地理标准来为查点区分配序号（所谓蛇形或“螺旋”形排序），所以电子数据表上属于邻近的查点区也就是该地区邻近的查点区，一般可以在电子数据表上自动对查点区进行组合。在我们的例子中，每个查点区有30多家住户，所以没有必要对查点区进行组合。但应该注意的是，根据上图对查点区进行组合是不切实际的，因为在数字清单中城市和农村查点区是混在一起的，在实际的国家中，这种情况是不太可能遇到的。换句话说，如果清单中城市和农村查点区是分散的而不是归在一起的，那么就无法通过计算机将邻近的查点区归在一起。

图十五.2

将电子表格作为初级抽样标架



57. 第一个抽样阶段前的另一个步骤是确定是否需要根据某种设计标准来对抽样标架进行分类，以便在每个明确的层内对样本进行隐含分层，行政区划几乎总是被用作这一目的。但在某些情况下，另一种标准，也就是城市/农村分层，可能会更重要一些。假设在我们的例子中，城市/农村分类就属于这种情况，那么首先需要根据城市/农村这一标准来对抽样标架进行分类，然后依次根据省份、行政区进行分类，最后根据查点区的序号进行分类。根据图十五.3中电子制表程序所提供的“分类”指令很容易做到这一点。

58. 抽选概率与规模成比例的初级抽样单位。绝大多数住户调查都是采用概率与规模成比例（PPS）的方法抽选初级抽样单位。如果抽样标架中有现成的数据，那么查点区的住户数量一般可用作度量其规模的尺度，但在某些情况下，可以使用人口或住处数量作为度量其规模的尺度。以下将说明概率与规模成比例的程序，假定设计要求在“森林”层（见图十五.4）抽选概率与住户数量（工作表的G列）成比例的88个查点区。

图十五.3
进行隐含分层

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Province Name	Ward Name	Word Name	CLA	Urban/Rural	No. of Inhabitants	Population							
2	1	West Tawenda	207	Macondo	02	U	58	528						
3	1	West Tawenda	207	Mhnanah	04	U	50	370						
4	1	West Iazezide	207	Macondo	06	U	52	270						
5	1	West Tawenda	207	Macondo	07	U	37	238						
6	1	West Iazezide	207	Mhnanah	11	U	68	367						
7	1	West Tawenda	207	Macondo	12	U	40	238						
8	1	West Tawenda	207	Mhnanah	17	U	53	312						
9	1	West Iazezide	207	Macondo	19	U	70	271						
10	1	West Tawenda	211	Dakuli	02	U	50	290						

	A	U	C	D	L	I	42	11	I	J	R	L	M	N
1318	8	Banzakul	414	Powai	08	R	189	531						
1319	8	Pozmitul	414	Powai	09	R	107	367						
1320	0	Danzakul	414	Powai	12	R	136	410						
1321	8	Banzakul	414	Powai	14	R	115	411						
1322	U	Lanzakul	414	Powai	15	R	73	254						
1323	8	Banzakul	414	Powai	18	R	87	308						
1324	8	Pozmitul	414	Powai	18	R	77	366						
1325	8	Banzakul	414	Powai	19	R	79	342						
1326	8	Pozmitul	414	Powai	22	R	42	203						
1327	U	Lanzakul	414	Powai	24	R	59	211						
1328														
1329														

59. 首先，在电子数据表中创建新的一列，这一列为查点区的累积规模。在I2单元格里录入公式= I1 + G2，然后将该公式沿着I列一直复制到最后一行。请注意，I列中的最后一行将为“森林”层住户的总数（110 388）。

60. 其次，创建另一列，该列为查点区按比例缩放后的累积规模，用I列中的值乘以缩放比例系数88/110 388（这样做的目的是为了一个从零增长到查点区数目的列被选取，并且与查点区的规模成比例，见图十五.5）。在单元格J2中录入公式= I2 * 88/110 388，然后将该公式沿着J列一直复制到最后一行：

61. 第三，在新一列的最顶端，录入一个值在0-1之间均匀分布的随机数，然后将该数和J列中所有行的数相加，从而创建一个新列，新列是通过随机数转换后按比例缩放的累积规模（见图十五.6）。可以在电子数据表内自动选择随机数，但最好在外部选择这个随机转换数（如，使用一张随机数表格），以防止系统在重新计算工作簿时所选样本与原来选取的样本不同。如，在单元格K1中录入随机数0.73，然后在单元格K2中录入公式= J2 + K\$1，然后将该公式沿着K列一直复制到最后一行。

图十五.4

抽选概率与规模成比例样本（第一步）

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Province	Province Name	MA	Ward Name	CL	Urban/Rural	Households	Population						
2	1	West Tazenda	207	Macondo	02	U	58	268	0.97					
3	1	West Tazenda	207	Macondo	07	U	50	330	1.09					
4	1	West Tazenda	207	Macondo	08	U	57	245	1.11					
5	1	West Tazenda	207	Macondo	07	U	37	238	1.98					
6	1	West Tazenda	207	Macondo	11	U	65	357	2.65					
7	1	West Tazenda	207	Macondo	12	U	40	235	2.05					
8	1	West Tazenda	207	Macondo	17	U	53	312	3.59					
9	1	West Tazenda	207	Macondo	19	U	70	231	4.29					
10	1	West Tazenda	211	Deluli	02	U	50	290	1.79					

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1318	B	Banzakai	414	Puwai	08	R	183	531	109,813					
1319	B	Banzakai	414	Puwai	08	R	107	382	109,720					
1320	U	Banzakai	414	Puwai	12	R	126	419	109,056					
1321	B	Banzakai	414	Puwai	14	R	115	411	109,871					
1322	R	Banzakai	414	Puwai	15	R	74	254	110,044					
1323	B	Banzakai	414	Puwai	18	R	87	308	110,131					
1324	R	Banzakai	414	Puwai	18	R	77	385	110,208					
1325	U	Banzakai	414	Puwai	19	R	79	342	110,287					
1326	B	Banzakai	414	Puwai	22	R	72	303	110,323					
1327	U	Banzakai	414	Puwai	24	R	18	151	110,381					
1328														
1329														

62. 至于通过随机数转换后按比例缩放的累积规模，如果其整数部分发生变化，那么样本将由整数部分发生变化的那一行决定。本例中，在1号省份（West Tazenda）207号行政区（Macondo）17号查点区这一列，通过随机数转换后按比例缩放的累积规模从0.97变化为1.02，说明这是样本中选取的第一个查点区。同一省份的226号行政区（Balayan）1号查点区这一列，通过随机数转换后按比例缩放的累积规模从1.99变化为2.09，整数部分再一次发生变化，说明这是样本中选取的第二个查点区。在单元格L2中录入公式= INT（K2）- INT（K1），然后沿着L列一直复制到最后一行就可以自动标出被选中的样本。L列中非零值所在的那一行就是要选取的样本（见图十五.7）。

63. 第一阶段抽选的所有抽样单位清单应该转到另一张工作表上，该工作表将成为调查管理的重要工具。如，调查管理人员可以增加一列来记录每个初级抽样单位所有重要活动的详细情况（实地工作和数据录入的预期日期和实际日期，有关主管小组的信息等）。

图十五.5

抽选概率与规模成比例样本（第二步）

The image shows two Excel worksheets. The top worksheet, 'Atlantic Sample Frame.xls1', has columns A through N. Row 2 is highlighted, showing data for 'West Tazemba' with a value of 0.05 in column J. The bottom worksheet, 'Atlantic Sample Frame.xls2', has columns A through O. It contains a list of UICs and their corresponding values in columns G through O.

64. 工作表将特别用以计算从样本中获取无偏估计值所需的抽选概率和相关的增大因子（或权数）。无需根据层对这种汇总工作表进行划分。最好将所有被抽选的初级抽样单位列入一张独特的工作表，在其中一列列明层。在本例中，被抽选的88个查点区中，头19个查点区的“样本”工作表见图十五.8。

65. 抽选概率和抽样权数。第一阶段抽选概率 $P(1)$ 可很简单地在“样本”工作表中进行计算：将样本初级抽样单位中的住户数量乘以每层中所抽选的初级抽样单位数量（下图十五.9中的G列和K列），然后用结果除以该层的住户总数（J列）。在单元格L2中录入公式 $= K2 * G2/J2$ ，然后沿着L列一直往下复制。

66. 其后各阶段的抽选概率当然取决于抽样设计的具体情况。我们将对这样一个两阶段抽样设计的抽选概率计算进行说明：两阶段抽样设计的第二阶段，每个初级抽样单位根据相同的概率来选择固定数量的住户。实际上，这种样本设计是实践中最常用的样本设计之一。第二阶段每个初级抽样单位抽选的

住户数量因层的不同而不同。但在本例假设的国家中，我们假设所有层中每个查点区的住户数量为12家。

图十五.6
抽选概率与规模成比例样本（第三步）

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Province	Province Name	UIC	Ward Name	CFA	UIC	Households	Population			0.78			
2	1	West Tazenda	207	Masundu	02	U	59 328	59	0.05	0.78				
3	1	West Tazenda	207	Mweneni	04	U	50 320	50	0.05	0.67				
4	1	West Tazenda	207	Macondo	06	U	52 210	52	0.11	0.06				
5	1	West Tazenda	207	Macondo	07	U	37 338	37	0.16	0.89				
6	1	West Tazenda	207	Macondo	11	U	50 250	50	0.21	0.94				
7	1	West Tazenda	207	Masundu	12	U	40 338	40	0.21	0.67				
8	1	West Tazenda	207	Mweneni	17	U	53 310	53	0.21	1.07				
9	1	West Tazenda	207	Macondo	19	U	70 231	70	0.24	1.07				
10	1	West Tazenda	211	Dobuli	02	U	50 290	50	0.38	1.11				

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1318	8	Banzukul	414	Puwai	06	R	188 521	109,613	87 38	88 11				
1319	8	Henritul	414	Puwai	09	R	110 387	101,120	87 47	88 20				
1320	8	Banzukul	414	Puwai	12	R	136 418	109,856	87 58	88 31				
1321	8	Henritul	414	Puwai	14	R	115 411	101,971	87 67	88 40				
1322	U	Uanzukul	414	Puwai	15	U	70 254	110,044	07 73	08 46				
1323	8	Banzukul	414	Puwai	16	R	87 308	110,131	87 80	88 53				
1324	U	Henritul	414	Puwai	18	U	70 389	110,238	07 85	08 59				
1325	8	Banzukul	414	Puwai	19	R	79 312	110,287	87 92	88 65				
1326	8	Henritul	414	Puwai	22	R	42 203	110,329	87 05	88 68				
1327	U	Uanzukul	414	Puwai	24	U	29 313	110,308	08 00	08 73				
1328														
1329														

67. 在这种抽样阶段中，一般要求对于每个被抽选的初级抽样单位都应编制住户清单。住户清单不需要通过计算机处理，因为可以根据纸清单通过手工选取调查所要访问的住户。但将清单录入计算机文件有很多优点（如，如果第一阶段抽选的初级抽样单位是所谓的标准样本，而这种标准的样本将用于各种调查，或用于多轮的固定样本调查，那么计算机文件就能显示出其优势）。

68. 编制清单时在每个被抽选为样本的初级抽样单位中实际发现的住户数量一般不同于初级抽样标架中通过普查所记录的原始“住户数量”。应该在“样本”工作表中添加一列以记录所列出的住户数量。如果清单是用计算机进行处理的，那么可以通过程序填写这一列（如，采用Excel macros）。否则，在编制住户清单时填写该列应成为调查管理人员的高度优先工作。在图十五.10中，“住户数量”和“列入清单的住户数量”框架分别显示在G列和M列。

图十五.7

抽选概率与规模成比例样本（第四步）

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Province	Province Name	District	Ward Name	CEK	Urban/Rural	No. households	Population			0.75			
2	1	West Tazenda	207	Macondo	02	U	59	338	59	0.05	0.75	0		
3	1	West Tazenda	207	Macondo	04	U	50	370	109	0.02	0.82	0		
4	1	West Tazenda	207	Macondo	06	U	52	246	101	0.10	0.89	0		
5	1	West Tazenda	207	Macondo	07	U	37	238	198	0.16	0.88	0		
6	1	West Tazenda	207	Macondo	11	U	60	377	206	0.21	0.94	0		
7	1	West Tazenda	207	Macondo	12	U	40	298	308	0.21	0.97	0		
8	1	West Tazenda	207	Macondo	17	U	53	312	389	0.20	1.00	0		
9	1	West Tazenda	207	Macondo	19	U	70	321	429	0.34	1.07	0		
10	1	West Tazenda	211	Debuli	02	U	50	200	479	0.38	1.11	0		
11	1	West Tazenda	211	Debuli	03	U	81	370	509	0.45	1.10	0		
12	1	West Tazenda	211	Debuli	06	U	77	258	637	0.51	1.21	0		
13	1	West Tazenda	211	Debuli	07	U	60	262	607	0.66	1.20	0		
14	1	West Tazenda	211	Debuli	08	U	45	312	740	0.59	1.32	0		
15	1	West Tazenda	211	Debuli	10	U	75	303	815	0.65	1.38	0		
16	1	West Tazenda	211	Debuli	11	U	62	311	977	0.70	1.40	0		
17	1	West Tazenda	211	Debuli	12	U	52	291	929	0.71	1.47	0		
18	1	West Tazenda	211	Debuli	13	U	63	297	1027	0.61	1.52	0		
19	1	West Tazenda	211	Debuli	15	U	82	345	1,094	0.81	1.57	0		
20	1	West Tazenda	211	Debuli	17	U	53	289	1,107	0.66	1.61	0		
21	1	West Tazenda	211	Debuli	19	U	60	308	1,167	0.62	1.66	0		
22	1	West Tazenda	211	Debuli	18	U	57	325	1,224	0.68	1.71	0		
23	1	West Tazenda	211	Debuli	20	U	54	311	1,213	1.02	1.71	0		
24	1	West Tazenda	211	Debuli	21	U	57	319	1,310	1.07	1.80	0		
25	1	West Tazenda	211	Debuli	24	U	50	347	1,329	1.12	1.85	0		
26	1	West Tazenda	211	Debuli	25	U	51	321	1,420	1.16	1.89	0		
27	1	West Tazenda	211	Debuli	28	U	58	338	1,508	1.20	1.93	0		
28	1	West Tazenda	211	Debuli	30	U	70	304	1,506	1.26	1.99	0		
29	1	West Tazenda	228	Balayan	01	U	125	183	1,711	1.36	2.06	0		
30	1	West Tazenda	228	Balayan	05	U	41	247	1,752	1.40	2.13	0		
31	1	West Tazenda	228	Balayan	06	U	75	289	1,827	1.46	2.19	0		

69. 至于其有用信息被实际记录在调查数据集中的住户，以及由于各种原因没有相关信息的住户，应该在实地工作和数据管理工作完成之后，在“样本”工作表中添加几列，然后按照每个初级抽样单位记录所述住户数量。添加一个“无效问卷”列的标准不响应理由在本出版物其他部分详细讨论（如，见第八章和第二十二章第F部分有关拒绝回答、住房闲置等方面的内容）。另外，当调查不能将基于计算机的质量控制纳入实地工作时，可能需要添加一个“无效问卷”列。遗憾的是，集中数据录入技术的常见结果就是不能将基于计算机的质量控制纳入实地工作。

70. 我们将继续以下图十五.11中的例子为例，并对有关情况进行简化，假设在“样本”工作表中添加了两列，一列为“数据集内的住户数量”，另一列为“不答复”总数。

图十五.8

被抽中的初级抽样单位的电子数据表

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Province	District Name	Ward	Ward Name	CEA	Urban/Rural	Household No. in the U	Population	Region					
2	1	West Tazenda	207	Mosarvin	17	U	53	312	Forest					
3	1	West Tazenda	226	Udeyay	01	U	125	403	Forest					
4	1	West Tazenda	226	Boloyan	53	U	69	394	Forest					
5	1	West Tazenda	226	Koloyan	90	U	43	102	Forest					
6	1	West Tazenda	242	Haliyal	52	U	48	279	Forest					
7	1	West Tazenda	255	Grasnu	15	U	52	333	Forest					
8	1	West Tazenda	259	Haza	04	U	79	295	Forest					
9	1	West Tazenda	401	Tuloo	21	U	84	361	Forest					
10	1	West Tazenda	401	Toiho	38	U	131	413	Forest					
11	2	East Tazenda	267	Kanadu	06	U	125	511	Forest					
12	2	East Tazenda	267	Kanadu	25	U	105	474	Forest					
13	2	East Tazenda	270	Queta	02	U	166	500	Forest					
14	2	East Tazenda	270	Queta	21	U	138	407	Forest					
15	2	East Tazenda	270	Queta	45	U	174	485	Forest					
16	2	East Tazenda	275	Mosken	10	U	150	368	Forest					
17	2	East Tazenda	275	Mosken	30	U	106	351	Forest					
18	2	East Tazenda	280	Ludza	06	U	180	505	Forest					
19	2	East Tazenda	280	Ludza	18	U	261	555	Forest					
20	2	East Tazenda	280	Ludza	30	U	132	473	Forest					

图十五.9

计算第一阶段抽选概率

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Province	District Name	Ward	Ward Name	CEA	Urban/Rural	Household No. in the U	Population	Region	Number of HHs in the Ustam	Number of HHs in the Ustam (%)			
2	1	West Tazenda	207	Mosarvin	17	U	53	312	Forest	110,338	88	0.01225		
3	1	West Tazenda	226	Udeyay	01	U	125	403	Forest	110,338	88	0.02965		
4	1	West Tazenda	226	Boloyan	53	U	69	394	Forest	110,338	88	0.05201		
5	1	West Tazenda	226	Koloyan	90	U	43	102	Forest	110,338	88	0.03425		
6	1	West Tazenda	242	Haliyal	52	U	48	279	Forest	110,338	88	0.02424		
7	1	West Tazenda	255	Grasnu	15	U	52	333	Forest	110,338	88	0.01145		
8	1	West Tazenda	259	Haza	04	U	79	305	Forest	110,338	88	0.06296		
9	1	West Tazenda	401	Tuloo	21	U	84	361	Forest	110,338	88	0.06896		
10	1	West Tazenda	401	Toiho	38	U	131	453	Forest	110,338	88	0.10363		
11	2	East Tazenda	267	Kanadu	06	U	125	511	Forest	110,338	88	0.09950		
12	2	East Tazenda	267	Kanadu	25	U	105	474	Forest	110,338	88	0.08370		
13	2	East Tazenda	270	Queta	02	U	166	500	Forest	110,338	88	0.13733		
14	2	East Tazenda	270	Queta	21	U	138	407	Forest	110,338	88	0.11001		
15	2	East Tazenda	270	Queta	45	U	177	485	Forest	110,338	88	0.14110		
16	2	East Tazenda	275	Mosken	10	U	150	368	Forest	110,338	88	0.11950		
17	2	East Tazenda	275	Mosken	30	U	106	351	Forest	110,338	88	0.08150		
18	2	East Tazenda	280	Ludza	06	U	180	505	Forest	110,338	88	0.14171		
19	2	East Tazenda	280	Ludza	18	U	261	555	Forest	110,338	88	0.20307		

图十五. 10
编制住户清单的结果记录

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Province	Province Name	Dist	Dist Name	UIC	UIC Code	UIC Population	UIC Status	UIC Type	Number of HHs in the Stratum	Number of PSUs in the Stratum	H-11	Number of HHs Listed	
2	1	West Tazenda	207	Mnenedi	17	U 53	317	Forest		110,388	88	0.04726	58	
3	1	West Tazenda	220	Delayan	04	U 125	403	Forest		110,388	88	0.09965	150	
4	1	West Tazenda	226	Boloyan	53	U 60	304	Forest		110,388	88	0.05901	60	
5	1	West Tazenda	220	Delayan	90	U 40	192	Forest		110,388	88	0.03420	44	
6	1	West Tazenda	242	Haliyal	52	U 48	279	Forest		110,388	88	0.03827	62	
7	1	West Tazenda	255	Omasu	15	U 52	333	Forest		110,388	88	0.04345	45	
8	1	West Tazenda	259	Pasat	04	U 79	395	Forest		110,388	88	0.06298	74	
9	1	West Tazenda	401	Tolbo	21	U 84	361	Forest		110,388	88	0.06696	95	
10	1	West Tazenda	401	Tolbo	30	U 130	423	Forest		110,388	88	0.06350	90	
11	2	East Tazenda	287	Kamalu	08	U 125	511	Forest		110,388	88	0.08985	117	
12	2	East Tazenda	267	Kanadu	21	U 101	494	Forest		110,388	88	0.08171	101	
13	2	East Tazenda	270	Quella	02	U 188	590	Forest		110,388	88	0.13233	174	
14	2	East Tazenda	270	Quella	21	U 138	477	Forest		110,388	88	0.11001	138	
15	2	East Tazenda	270	Quella	40	U 171	425	Forest		110,388	88	0.04110	102	
16	2	East Tazenda	275	Mookan	10	U 150	368	Forest		110,388	88	0.11958	150	
17	2	East Tazenda	275	Mookan	30	U 100	271	Forest		110,388	88	0.08450	132	
18	2	East Tazenda	280	Luliza	06	U 188	685	Forest		110,388	88	0.14328	191	
19	2	East Tazenda	280	Luliza	15	U 261	555	Forest		110,388	88	0.20907	285	

图十五. 11
记录不答复

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Province	Province Name	Dist	Dist Name	UIC	UIC Code	UIC Population	UIC Status	UIC Type	Number of HHs in the Stratum	Number of PSUs in the Stratum	H-11	Number of HHs Listed	Number of HHs in the Stratum	Non-response			
2	1	West Tazenda	207	Mnenedi	17	U 53	317	Forest		110,388	88	0.04726	58	11	1			
3	1	West Tazenda	220	Delayan	04	U 125	403	Forest		110,388	88	0.09965	150	12	0			
4	1	West Tazenda	226	Boloyan	53	U 60	304	Forest		110,388	88	0.05901	60	12	0			
5	1	West Tazenda	220	Delayan	90	U 40	192	Forest		110,388	88	0.03420	44	11	1			
6	1	West Tazenda	242	Haliyal	52	U 48	279	Forest		110,388	88	0.03827	62	8	4			
7	1	West Tazenda	255	Omasu	15	U 52	333	Forest		110,388	88	0.04345	46	11	1			
8	1	West Tazenda	259	Pasat	04	U 79	395	Forest		110,388	88	0.06298	74	11	1			
9	1	West Tazenda	401	Tolbo	21	U 84	361	Forest		110,388	88	0.06696	95	12	0			
10	1	West Tazenda	401	Tolbo	30	U 130	423	Forest		110,388	88	0.06350	90	8	4			
11	2	East Tazenda	287	Kamalu	08	U 125	511	Forest		110,388	88	0.08985	117	12	0			
12	2	East Tazenda	267	Kanadu	21	U 101	494	Forest		110,388	88	0.08171	101	12	0			
13	2	East Tazenda	270	Quella	02	U 188	590	Forest		110,388	88	0.13233	174	11	1			
14	2	East Tazenda	270	Quella	21	U 138	477	Forest		110,388	88	0.11001	138	10	2			
15	2	East Tazenda	270	Quella	40	U 171	425	Forest		110,388	88	0.04110	102	12	0			
16	2	East Tazenda	275	Mookan	10	U 150	368	Forest		110,388	88	0.11958	150	12	0			
17	2	East Tazenda	275	Mookan	30	U 100	271	Forest		110,388	88	0.08450	132	11	1			
18	2	East Tazenda	280	Luliza	06	U 188	685	Forest		110,388	88	0.14328	191	11	1			
19	2	East Tazenda	280	Luliza	15	U 261	555	Forest		110,388	88	0.20907	285	11	1			

71. 尽管没有普遍接受的模型可用于不答复情况，但一种最常见的假设是：最终数据集内“有用的”住户实际上是住户各自的初级抽样单位内所列全

部住户的等概率样本（详细讨论见第二章和第八章）。根据这种假设，只需要将有用住户的数量除以所列住户数量就可以计算出第二阶段抽选每个有用住户的概率 $P(2)$ 。初级抽样单位中每个住户的总抽选概率为 $P(1)*P(2)$ ，抽样权数为该概率的倒数。

72. 在电子数据表中可以很容易根据这些公式进行计算（见图十五.12）。在单元格P2中录入公式= N2/M2，单元格Q2中录入公式= L2*P2，单元格R2中录入公式= 1/Q2，然后沿着P、Q和R列一直往下复制。

73. 这种根据概率计算的权数适用于各初级抽样单位的所有住户。有些调查从业人员可能采用“事后分层”的方法来对这些权数做进一步的调整，以确保调查估计和某些已知的人口分布（如，从样本调查外部来源获得的年龄和性别分布，或总消费数据）相称。这些调整是通过特殊的软件直接在调查数据集中进行的，而不是在抽样电子数据表中进行，一般是在每个住户或每个人的基础上而非每个初级抽样单位的基础上进行。

图十五.12
计算第二阶段概率和抽样权数

Household No	Household Name	Ward Name	U.P. No	U.P. Name	Block No	Block Name	Number of Households in the Stratum	Number of Households in the Sample	W ₁	Number of Households in the Sample	Number of Households in the Stratum	W ₂	W ₂ *W ₁	Weight		
1	West Tanzania	201	Mwanza	11	11	212	Forest	110,000	66	0.0006	66	11	1	0.0006	0.0001	124.00
2	West Tanzania	206	Kilimanjaro	23	13	485	Forest	110,000	66	0.0006	125	12	11	0.0006	0.0001	124.00
3	West Tanzania	206	Kilimanjaro	23	13	485	Forest	110,000	66	0.0006	81	12	11	0.0006	0.0001	124.00
4	West Tanzania	206	Kilimanjaro	23	13	485	Forest	110,000	66	0.0006	47	11	1	0.0006	0.0001	124.00
5	West Tanzania	207	Kilimanjaro	24	14	219	Forest	110,000	66	0.0006	87	8	7	0.0006	0.0001	207.50
6	West Tanzania	251	Mwanza	25	17	101	Forest	410,000	181	0.0015	61	15	1	0.0015	0.0001	101.00
7	West Tanzania	258	Pwani	28	18	785	Forest	410,000	66	0.0006	79	14	1	0.0006	0.0001	101.00
8	West Tanzania	421	Tabora	28	18	785	Forest	410,000	66	0.0006	87	13	6	0.0006	0.0001	101.00
9	West Tanzania	421	Tabora	28	18	785	Forest	410,000	66	0.0006	80	8	7	0.0006	0.0001	101.00
10	East Tanzania	267	Mwanza	27	17	761	Forest	410,000	66	0.0006	117	13	6	0.0006	0.0001	87.80
11	East Tanzania	267	Mwanza	27	17	761	Forest	410,000	66	0.0006	121	13	6	0.0006	0.0001	101.00
12	East Tanzania	270	Geita	27	17	761	Forest	410,000	66	0.0006	121	14	1	0.0006	0.0001	101.00
13	East Tanzania	270	Geita	27	17	761	Forest	410,000	66	0.0006	121	13	7	0.0006	0.0001	101.00
14	East Tanzania	270	Geita	27	17	761	Forest	410,000	66	0.0006	121	13	7	0.0006	0.0001	101.00
15	East Tanzania	270	Geita	27	17	761	Forest	410,000	66	0.0006	121	12	0	0.0006	0.0001	101.00
16	East Tanzania	270	Geita	27	17	761	Forest	410,000	66	0.0006	121	12	0	0.0006	0.0001	101.00
17	East Tanzania	270	Geita	27	17	761	Forest	410,000	66	0.0006	121	11	1	0.0006	0.0001	101.00
18	East Tanzania	280	Lushoto	26	16	555	Forest	110,000	66	0.0006	131	11	1	0.0006	0.0001	117.10
19	East Tanzania	280	Lushoto	26	16	555	Forest	110,000	66	0.0006	285	11	1	0.0006	0.0001	124.50

H. 建议摘要

74. 本章旨在阐明在调查的各个阶段采用数据管理标准的重要性，而不是将数据管理视作仅属于最后分析阶段的部分工作。在这方面的一个最明显的例子是有关生活水平衡量研究的调查，在该项调查中，调查问卷的设计、实地工作的计划和实施、数据的录入和处理都采取了有助于对数据进行正确管理的方式，甚至在收集任何数据前也是如此。指导这项工作的各项原则是本章的核心内容。尽管这些原则会因特定国家的具体应用而有不同的特征，但仍然可以将这些原则进行如下的归纳和总结：

- (a) 调查数据管理的入手点是调查问卷设计，在这一阶段需要：
- (一) 对统计单位进行正确的标识。建议调查的初级抽样单位采用简单的或改进的三位或四位数字号，而每个初级抽样单位内的住户采用两位数字号，另外还应对住户内的每个从属单位进行合理的序列标识；
 - (二) 固有的重复。在设计调查问卷时应列入有意重复的内容，这是为了发现访调员的错误或数据录入错误。这方面的例子是为总计数加一行底线，或者对重要变量的代码加一个校验位数。
- (b) 在实地工作中，应该考虑以下方面：
- (一) 数据录入和编辑的操作策略。建议有关国家对实地录入所有数据的选择方案进行认真考虑。可以由一位在固定场所工作的数据录入人员而不是被调查住户的数据录入人员进行有关工作；由一位操作人员和调查小组其余人员一道在访谈每个住户时直接将数据录入便携式计算机；或使用掌上电脑通过至今尚未加以适当研究的无纸访谈进行（但还需要对无纸访谈做进一步的研究）。在实地录入所有数据以及集中录入数据将大大有助于保证质量和一致性；
 - (二) 质量控制标准。调查问卷上的数据需要由五种不同的质量控制机制加以控制：范围检验、根据参照表检验、跳过检验、一致性检验和印刷检验；
 - (三) 数据录入技术。根据世界银行于1995年进行的审查，有两个可靠的数据录入和编辑平台适用于复杂的住户调查，这两个平台是：世界银行内部开发的生活水平衡量研究软件包和美国普查局的微型计算机集成处理系统程序。其各自的更新版本分别为LSD-2000和CSPro。在设计数据录入和编辑工具时，除了需要考虑那些对各国自身条件有影响作用的现有专业经验以及其他因素以外，还应该考虑一些基本准则：除了一些例外情况外，计算机屏幕应该和其相应的调查问卷页面相同；数据录入程序应该能够辨别不可能的情况和未必可能的情况，并为每种情况做特别的标记；对错误进行报告的语言和表述应该通俗易懂；
 - (四) 组织和散发调查数据集。在这一方面，不适合采用平面数据文件，因为平面数据文件不能正确处理住户内的从属统计单位（人员、农作物和消费品等）。所采用的文件结构最好是每种统计单位采用不同的记录类型。
- (c) 最后，数据管理在执行抽样协议的主要阶段提供指导，对协议的执行起到了一定的帮助作用，所述阶段有：初级抽样标架的组织（通常根据现有的最新查点区创建）；抽选概率与规模成比例的初级抽样单位（规模由住户、住所数量或人口数量计量）；抽选概率及相关抽样权数的计算。

参考资料

- Ainsworth, M. 和 J. Muñoz (1986年)。《科特迪瓦生活水平调查：设计与实施》。《生活水平衡量研究工作文件，第26号》。哥伦比亚特区华盛顿：世界银行。
- Blaizeau, D. (1998年)。“西非经济和货币联盟中的七次支出调查”。《国际调查统计员协会/国际官方统计协会经济与社会发展联席会议的记录》。墨西哥，阿瓜斯卡连特斯：国际统计学会。
- _____和 J. L. Dubois (1990年)。 *Connaitre les Conditions de Vie des Ménages dans les Pays en Développement*. Paris: Documentation française。
- Blaizeau, D.和 J. Muñoz (1998年)。 *LSD-2000. Logiciel de Saisie des Données: Pour Saisir les Données d'une Enquête Complexe*. Paris: Institut national de la statistique et des études économiques。
- Grosh, M. 和 J. Muñoz (1996年)。《规划和实施生活水平衡量研究调查手册》。《生活水平衡量研究工作文件，第126号》。哥伦比亚特区华盛顿：世界银行。
- Muñoz, J. (1989年)。“复杂社会经济调查的数据管理：从问卷设计到数据分析”。《国际统计学会第四十七次届会的记录》。巴黎：国际统计学会。
- _____ (1996年)。Cómo mejorar la calidad de la información: opciones para mejorar la organización del trabajo de campo, el sistema de entrada de datos, el análisis de consistencia y el manejo de la base de datos. In *Reunión de Iniciación del Programa para el Mejoramiento de las Encuestas de Condiciones de Vida en América Latina y El Caribe*. Asunción: 美洲开发银行。
- _____ (1998年)。“预算-消费调查：新挑战与展望”。《国际调查统计员协会/国际官方统计协会经济与社会发展联席会议的记录》。墨西哥，阿瓜斯卡连特斯：国际统计学会。
- 美国人口普查局。CSPRO人口普查与调查处理系统，请登陆：<http://www.census.gov/ipc/www/cspro/>。

第十六章

根据住户调查数据编制简单描述性统计资料

Paul Glewwe

明尼苏达大学应用经济学系
美国明尼苏达州圣保罗

Michael Levin

美国普查局
美国华盛顿

摘 要

本章提供计算和列示住户调查数据基本描述性统计资料的一般准则。本章进行的分析是基本分析，包括编制易于被广大读者理解的相对简单的图表。此外，本章还就如何在广泛散发的综合报告中如何使用图表提出了建议。

关键词：描述性统计资料、表、图、统计摘要、散发。

A. 导言

1. 只有对数据进行分析才能实现住户调查数据的真正价值。数据分析的类别很多：从包含非常简单的主要统计指标的分析到极其复杂的多变量分析不等。本章是以后四章的导论，正因为如此，本章重点介绍的都是一些基本问题和相对简单的方法。其后各章将对更复杂的内容进行介绍。

2. 可以通过各种方式运用大多数住户调查数据来说明属于调查主要焦点的现象。从某种意义上来说，数据分析的入手点是基本的描述性统计，如有关主要变量的均值和频数表。然而，数据分析的最根本入手点在于需要通过收集数据来回答的问题。因此，几乎在任何一项住户调查中，首要任务就是要确定调查目标和设计调查问卷，以便所收集的数据能够实现这些目标。这说明应该在收集数据前，同时进行调查设计和数据分析规划。第三章对此进行了详细的解释。本章将重点介绍数据分析的实际问题，假设已经根据第三章的建议为数据分析制定了合理的策略。

3. 本章的安排如下：B节介绍变量和简单描述性统计资料的类型；C节提出一般性建议，说明如何根据住户调查数据编制和提供基本描述性统计资料；D节提出建议，说明如何编制一般性报告（通常称为统计摘要），将住户调查的基本成果传播给广大读者。最后一节简要提出了一些结论性意见。

B. 变量与描述性统计资料

4. 大多数住户调查都是就一个主题收集数据，而其他住户调查则就各种主题收集数据。在两种情况下，所收集的数据都可以被视作各种变量的集合，其中有些变量具有独立的意义，而另有一些变量是在和其他变量进行比较时才具有一定的意义。很多变量将在住户一级发生变化，如住处类型，而其他变量将在个人一级发生变化，如年龄和婚姻状况。有些调查可能收集只在社区一级发生变化的数据，如当地市场出售的各种商品的价格。²⁷

²⁷ 在大多数住户调查中，住户被定义为一群人，这群人：(a) 居住在同一住处；(b) 每天至少在一起吃一顿饭；(c) 共享用以购买商品和其他服务的收入和其他资源。有些住户调查根据当地情况对这一定义进行了修改，但这一问题不在本章的讨论范围之内。“社区”则更难界定，但为本章之目的，可以将其定义为居住在同一村庄、城镇或城市某个地区的一群住户。至于“社区”定义的详细介绍讨论，请参见Frankenberg（2000年）。

5. 任何数据分析的第一步都是为了产生一个包括所有相关变量的数据集。数据分析员可以对基本的描述性统计资料进行计算，让变量“自己说话”。这方面的方法相对少一些。本节将就如何使用这些方法进行介绍。首先，将对不同变量和描述性统计资料进行简要讨论，然后就单个变量、两个变量、三个或更多变量的数据编制方法进行讨论。

1. 变量类型

6. 住户调查就两类别型的变量收集数据，“分类”变量和“数值”变量。分类变量的特征是本身不是数字，而是类别或类型，如，住处特征（地面材料、墙面材料、洗手间类型等）和个体特征（种族、婚姻状况和职业）等。实际上，可以为这些特征分配代码，某个种族为“代码1”，另一个种族为“代码2”等，但这是一种任意的做法。相反，数值变量本身就是数字，如一个住处的房间数量，拥有土地的数量，或某个家庭成员的收入。在本章中，分类变量可能产生的各种不同结果将称之为“类别”，数值变量可能产生的各种不同结果将称之为“值”。

7. 在编制任何一类变量数据时，有必要就变量类别或值的数量进行区分。如果类别/值的数量很少（如，不足10），那么编制有关变量分布的全部信息不仅方便而且资料丰富。但是，如果值/类别的数量很多，如10个以上，那么一般最好编制有关变量分布的总体或主要统计指标。举例即可明确说明这一点。如，在某个国家，人口所包括的种族团体数量可能很少，也许只有四个，对于这样一个国家，可以较容易地通过一个简单的表或图显示每个族群中被抽样住户的百分比。然而，在另一个国家，可能有数以百计的族群。显示如400个不同族群的每个族群中被抽样住户的百分比将是一件非常乏味的工作。在大多数情况下，如果将许多不同的族群汇总成少数大类，然后显示每个大类中住户的百分比，那么不仅会简单一些，而且也能提供足够的信息。

8. 以上所举的例子采用了族群这个分类变量，这个例子也适用于数值变量。有些数值变量（如在过去一周中，一个人生病的天数）只有为数不多的值，那么可以通过简单的表或图显示全部分布情况。然而对于另一个数值变量来说（如农场拥有动物的数量），则可能有很多的值，这样最好只编制分布的某些主要统计指标。处理分类变量和数值变量的主要区别在于：在值/类别的数量有很多时，如何进行归类。对于分类变量来说，一旦决定不显示全部分布情况，那么就必须将分类变量分为若干个大类。对于数值变量来说，可以将

有关变量分为几个大类，但可以显示主要统计指标，如平均值、标准差、或者最小值和最大值。以下分节将对最常见的描述性统计资料进行简单介绍。

2. 简单描述性统计资料

9. 通过简单描述性统计资料，图和表可以提供有关变量的基本信息。这些统计资料包括，但不限于百分比分布、中位数、平均值和标准差。本分节研究这些简单的统计资料，并采用北马里亚纳群岛联邦赛班岛和美属萨摩亚的住户调查数据举例说明。

10. 百分比分布。住户调查很少收集100、1 000或10 000个人或住户的数据。假设拥有某个分类变量类别方面的数据，如某个人口中男性的数量和女性的数量，或拥有某个数值变量方面的数据，如，同一人口群中各成员按年计算的年龄。通常情况下，显示每个类别中被观测对象的数量不如显示每个类别中被观测对象的百分比。可以从表十六.1中的头三列数字看出这一点。如果没有给出百分数分布情况，那么大多数用户都难以理解这些结果，但一个人可以根据表十六.1中最后三列的数据较容易地了解不同年龄组男性的比率和女性的比率。当然，一个人可能对列中的百分比感兴趣，也就是不同年龄组男性的百分比和女性的百分比，有关这方面的情况见表十六.2。（第三种可能性是在表中显示所有年龄按性别加起来为100%的比率，但这种做法较少引人关注。）两个表都说明，可以列示分类变量和数值变量的百分比分布情况。

表十六.1

2002年4月北马里亚纳群岛联邦赛班岛按年龄和性别划分的人口分布情况：
行百分比

大组距年龄组,按年计算	数 量			行百分比		
	总 计	男	女	总 计	男	女
总人数	67 011	29 668	37 343	100.0	44.3	55.7
15岁以下	16 915	8 703	8 212	100.0	51.5	48.5
15 - 29岁	18 950	5 765	13 184	100.0	30.4	69.6
30 - 44岁	20 803	9 654	11 149	100.0	46.4	53.6
45 - 59岁	8 105	4 458	3 648	100.0	55.0	45.0
60岁或以上	2 239	1 088	1 150	100.0	48.6	51.4

11. 表十六.1显示各年龄组的性别分布情况不同。这也反应出了一些在表十六.1和十六.2中无法看到的信息，也就是赛班移民工人——赛班服装厂内有很多移民工，尤其是女工。尽管赛班最年轻一组的人群中，男性人数略多于女性；但下一个年龄组，也就是15-29岁之间的年龄组中，每70个女性中只有30个男性；在30-44岁之间的年龄组中，女性也多于男性，由于赛班服装厂的工人大多数为年龄在20-40之间的女性，所以以上情况和这一情况相吻合；再下一组，也就是45-59岁之间的年龄组中，男性多于女性。表十六.2中的列百分比说明男性的最大年龄组为30-44岁，女性的最大年龄组为15-29岁，这一年龄组的女性在服装厂工作的可能性最大。

资料来源：北马里亚纳群岛联邦当前劳动力调查第10轮。

说明：数据来源是从住户和所有集体居住人员中抽选的10%的随机样本。

表十六.2

2002年4月北马里亚纳群岛联邦赛班岛按年龄和性别划分的人口分布情况：
列百分比

大组距年龄组,按年计算	数 量			列百分比		
	总 计	男	女	总 计	男	女
总人数	67 011	29 668	37 343	100.0	100.0	100.0
15岁以下	16 915	8 703	8 212	25.2	29.3	22.0
15 - 29岁	18 950	5 765	13 184	28.3	19.4	35.3
30 - 44岁	20 803	9 654	11 149	31.0	32.5	29.9
45 - 59岁	8 105	4 458	3 648	12.1	15.0	9.8
60岁或以上	2 239	1 088	1 150	3.3	3.7	3.1

资料来源：北马里亚纳群岛联邦当前劳动力调查第10轮。

说明：数据来源是从住户和所有集体居住人员中抽选的10%的随机样本。

12. 中位数。度量数值变量最常用的统计量度是均值和中位数。（根据定义，分类变量不是数字，因此无法计算此类变量的均值和中位数。）中位数是分布的中点，均值是值的算术平均数。中位数常用于诸如年龄和收入之类的变量，因为它对异常值不太敏感。举一个极端的例子：假设一项调查中有99个人，每年收入在8 000美元到12 000美元之间，围绕10 000美元呈对称性的分布。这样均值和中位数都为10 000美元。现假设该年期间增加一个年收入为500 000美元的人，这样平均数将为15 000美元左右，而中位数仍然为10 000美元左右。至于很多收入类的变量，公布的报告通常显示均值和中位数。

13. 再次回到赛班的调查数据，2002年4月，赛班人口年龄的中位数为28.5岁，也就是说，有半数人口的年龄在28.5岁以上，半数人口的年龄在28.5岁以下。女性的年龄中位数低于男性（分别为27.6和30.5），这是因为有很多年轻女移民工在服装厂工作。

14. 均值和标准差。如上所述，均值是数值变量的算术平均数。通常计算（妇女）生育孩子的数量、收入以及其他数值变量的均值。标准差度量数值变量和该变量平均值之间的平均距离，用以度量数值变量分布的离散程度。

15. 表十六.3显示1995年美属萨摩亚住户调查中年收入的中位数和均值。该调查从该地区所有住户中抽选20%的随机样本。住户收入的均值高于中位数并不令人吃惊，因为有些住户的工资非常高，并且从其他来源获得了较高的收入。来自汤加的移民比较贫穷，其收入的均值和中位数都偏低即可以看出这一点；“其他族群”收入的均值和中位数都很高，这说明他们相当富裕。

表十六.3

1994年美属萨摩亚根据族群划分的住户收入主要统计指标

年 收 入	总 计	萨 摩 亚	汤 加	其他族裔群体
被调查住户的数量	8 367	7 332	244	790
中位数（美元）	15 715	15 786	7 215	23 072
平均值（美元）	20 670	20 582	8 547	25 260

资料来源：1995年美属萨摩亚住户调查。

说明：数据是从所有住户中抽选的20%的未加权随机样本。

3. 编制一个变量的描述性统计资料

16. 在只有一个变量的情况下，编制住户调查的描述性统计资料最简单。本分节说明如何编制分类变量和数值变量的描述性统计资料。

17. 显示整个分布情况。对于分类变量或数值变量来说，如果其类别或值的数量很少，如10个或10个以下，那么可以极其容易地显示其分布情况。可以用一个表通过显示变量每个类别或数值的频数来显示变量的整个（百分比）分布情况。这方面的例子见表十六.4，该表为越南住户照明主要来源的样本频数（未加权）以及百分比分布情况。很多住户调查需要采用权数对人口某个变量的分布情况进行估计，在这种情况下，显示原始样本频数可能会产生混乱，因此是不可取的。C节将对权数的运用进行讨论。（越南调查是以自行加权的样本为基础的，所以无需权数。）最后，报告估计的百分比频数的标准误差也很有用（这一问题的详细讨论见第二十一章，调查样本设计的其他特征以及权数的使用都会使这一问题复杂化）。

18. 在有些情况下，某个变量类别或值的数量可能很多，但只需要几个类别或值就可以说明分布的主要情况。在这种情况下，可能没有必要显示每个类别或值的频数。为了防止表格读者不至于被信息量搞得焦头烂额，一个办法就是将罕见的情况归入“其他”类中。如，对于任何一个类别或值，如果其频数不足1%，那么就可以将其归入“其他”类中。表十六.4也正是这么做的，“其他”类中包括的是一些罕见的情况，如火把和手电筒。在有些情况下，可能还存在其他自然组。譬如，在很多国家中，族群和宗教团体可以划分为众多界限分明的类别，但可以将这些精确分类的团体分为数量不多的几个大类。在很多情况下，显示更具有普遍性团体的数据就足够了。这项原则的主要例外情况是那些尽管很罕见，但却可能具有特定意义的类别。一般来说，可以单独报告这些“罕见但却具有特别的意义”的类别，但在这种情况下应该显示出标准误差，这一点很重要，因为稀少类别估计值的精确度较低。

19. 在很多情况下，如果用图而不是表来编制数据，那么会更直观，也更能引起人们的兴趣。对于一个只有少数几个类别或值的变量，用图显示数据的一个常见方法就是采用柱状图或直方图，在图中，柱的高度代表每个类别或值的相对频数。有关这方面的例子见图十六.1，该图采用了表十六.4中的数据。显示某个变量类别或值的相对频数还有一种常见的方法——采用饼图，饼图是一个圆圈，可以通过饼“块”的大小来显示相对频数。这方面的例子见图十六.2，该图也显示了表十六.4中的信息。至于如何设计有效的图表，详细建议见Tuftte（1983年）以及Wild和Seber（2000年）。

表十六.4

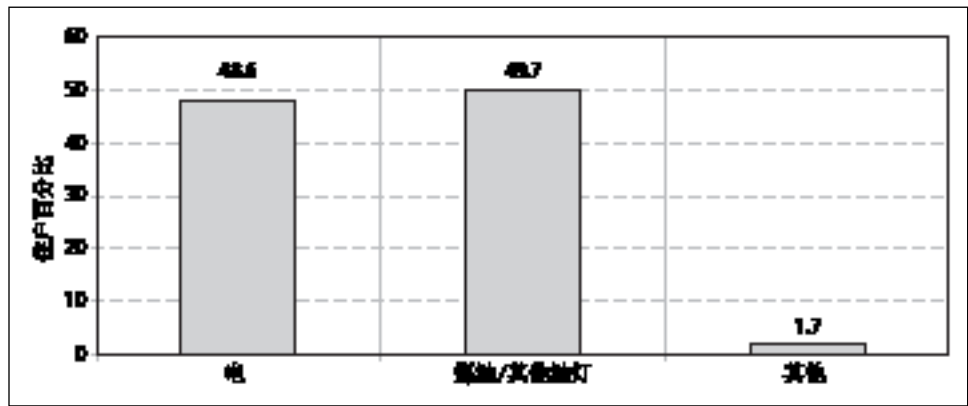
1992-1993年越南住户照明的来源

照明方法	住户数量	住户百分比（标准误差）
电	2 333	48.6 (0.7)
煤油/其他油灯	2 386	49.7 (0.7)
其他	81	1.7 (0.2)
样本的总住户数量	4 800	100.0

资料来源：1992-1993年越南生活水平调查。
说明：数据未加权。

图十六.1

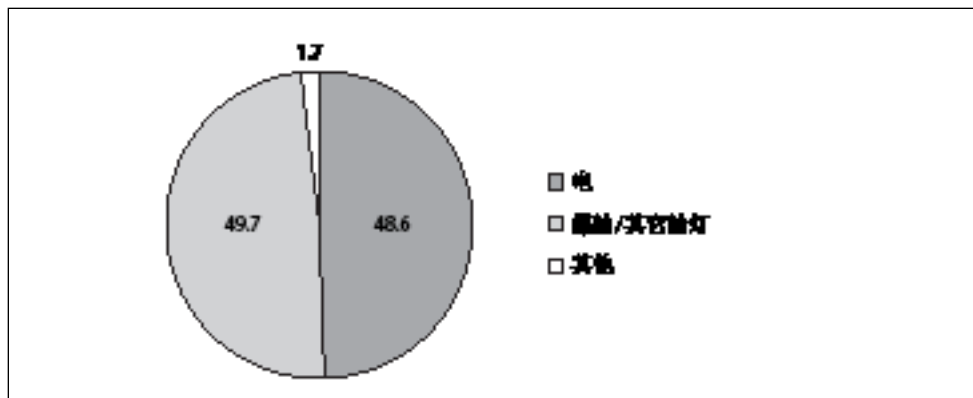
1992-1993年越南住户的照明来源 (柱状图)



资料来源：1992-1993年越南生活水平调查。
说明：样本规模：4 800个住户。

图十六.2

1992-1993年越南住户的照明来源 (饼图) (百分比)



资料来源：1992-1993年越南生活水平调查。
说明：样本规模：4 800个住户。

20. 显示有很多类别或值的变量。分类变量和数值变量可能经常有很多类别或值。对于分类变量来说，为避免编制过于详细的表或图，惟一的方法就是如上所述，将众多类别归为若干个大类和/或将所有罕见的值归入“其他”类中。对于数值变量来说，则有两个截然不同的选择。

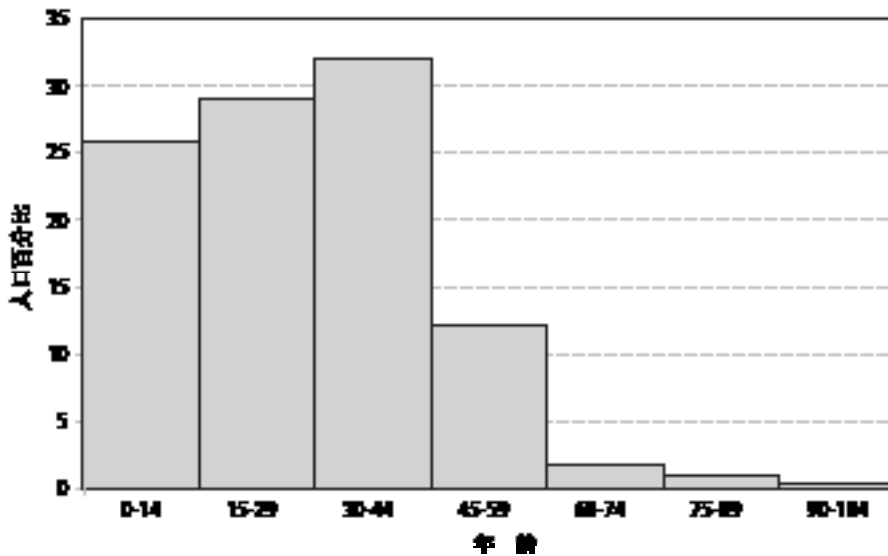
21. 首先，如果一个数值变量有很多值，那么可以将该变量的域分为数量不多的几个区间，然后根据以上针对变量只有少数类别或值的情况而采取的任何一种方式来显示信息。譬如，对于表十六.1和十六.2中的年龄变量，就采取了这种做法。这种方法还可以用在图中：如果一个数值变量有很多值，那么有关该变量分布的信息可以采用显示频数的图来显示，根据频数，变量被分为数量不多的类别。这种图的一个例子就是直方图，直方图显示的是有关变量的大致密度函数。直方图将数值变量的域划分为数量相对少的“子域”，“子域”通常被称作“箱”。每个箱由一个圆柱代表，圆柱的面积和相应子域内样本的百分比成比例。图十六.3就是按照这种方法来显示表十六.2中的年龄数据的。第一个箱代表0-14之间的子域，第二个箱代表15-29之间的子域……。²⁸请

²⁸ 该直方图将年龄在60-99之间的人口分为三组（60-74、75-89、90-104），其中每组的年龄跨度都相同，都为15年，这和与60岁以下各组的年龄跨度相同。这么做是为了确保直方图中每个柱的面积和每个年龄组中人口的百分比成比例。

注意，和图十七.1中柱状图不同的是，直方图中各“柱”之间没有间隔。因为直方图水平轴所显示的是变量的域，变量的域之间通常没有“间隔”。

图十六.3

2002年4月赛班人口的年龄分布情况（直方图）



资料来源：北马里亚纳群岛联邦当前劳动力调查第10轮。

22. 如果一个数值变量有很多值，那么显示这种数值变量的也许是最常见的第二种方法是编制其分布情况的某些主要统计指标，如均值、中位数和标准差。只能通过表中显示这些统计指标来编制指标，无法在一个图中显示单一数值变量的主要统计指标。除了均值、中位数和标准差外，编制最低值和最高值、上四分位数和下四分位数²⁹以及偏态量度也很有用。这方面的例子见表十六.5。

4. 编制两个变量的描述性统计资料

23. 相对于独立考察单一的变量而言，考察两个或更多变量之间的关系通常能够更深入地了解有关主题。然而，显示数据的可能性也同时增加了一个数量级。本分节将介绍各种常见的方法，区分拥有少数类别或值的变量和拥有很多值的变量。

24. 拥有少数类别或值的两个变量。显示两个变量相互关系的最简单情况就是两个变量都只有少数的类别或值。在简单的双向表中，可以将一个变量的类别或值作为“列”，将另一个变量的类别或值作为“行”。这方面的例子见表十六.6，该表说明越南城乡地区所使用的不同类型的健康服务供应商。在本例中，“列”总计为100%。如上所述，另一种选择就是让“行”总计为100%。在根据越南调查所举的例子中，各行百分比数字（总计为100%）将说明每类保健设施在越南城乡地区的分布情况。第三种选择就是：该表的每个“单元

²⁹ 某一分布的下四分位值四分位数比被观察观测对象的第一四分位值四分位数（25%）大，比第三四分位值四分位数（75%）小，而上四分位值四分位数比观察观测的第三四分位值四分位数（75%）大，比第一四分位值四分位数（25%）小。

格”给出某个特定地理区域（城市或农村）某个人访问保健设施的（联合）概率的频数（以百分比表示），在这种情况下，所有“行”和“列”百分比的总计将为100%。但是，这种方法很少使用，因为条件分布通常更令人感兴趣。无论如何，好的做法是报告充分的数据，以便读者都可以根据表中提供的数据推导出所有三种类型的的频数。

表十六.5

1992-1993年住户总支出简况：越南（千盾/年）

均值	6 531
标准差	5 375
中位数	5 088
下四分位数	3 364
上四分位数	7 900
最小值	235
最大值	100 478

资料来源：1992-1993年越南生活水平调查。

说明：样本规模：4 799个住户。

表十六.6

1992-1993年越南城乡地区过去四周访问过某个保健设施的（所有年龄组）人口使用各种保健设施的情况

看病地点	城市地区		农村地区	
	频 数	百分比（标准误差）	频 数	百分比（标准误差）
医院或诊所	251	45.0(2.1)	430	25.0(1.0)
乡镇保健中心	30	5.4(1.0)	318	18.5(0.9)
健康服务提供者家中	213	38.2(2.1)	595	34.6(1.1)
病人家中	50	9.0(1.2)	376	20.1(1.0)
其他	14	2.5(0.7)	29	1.7(0.3)
总 计	558	100.0	1 718	100.0

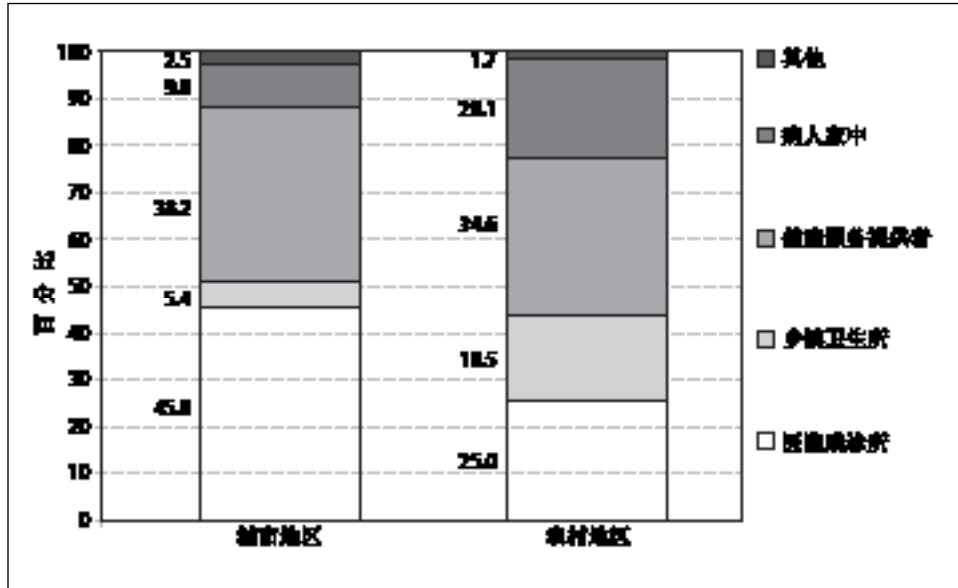
资料来源：1992-1993年越南生活水平调查。

25. 如果两个变量只有少数的值，那么可以采取几种方法用图来显示这两个变量关系的信息。在显示列百分比或行百分比的时候，一个简单的方法就是让若干个纵向列的百分比总计为100%。每一列代表一个变量具体的值，其他变量分布的频数在每个列中以阴影部分显示。图十六.4就是采取了这种方法来显示越南的保健设施数据。电子数据表软件包可以列示许多其他方法，供人们使用。

26. 拥有少数的类别/值的单变量和拥有很多值的数值变量。另一种常见的情况就是有两个变量。一个是有少数类别或值的变量（也许是经过归类数量变少了），另一个是有很多值的数值变量。在这方面，一个最常见方法就是以拥有少数类别和值的每个变量的值为条件，按照数值变量的均值来显示数据。另外，还可以增加其他方面的信息，如中位数和标准差。这方面的例子见表十六.7，该表显示的是1992-1993年越南平均住户总支出水平，其住户是根据该

图十六.4

1992-1993年越南城乡地区过去四周访问过某个保健设施的（所有年龄组）人口使用各种保健设施的情况（百分比）



资料来源：1992-1993年越南生活水平调查。
说明：样本规模：2 276。

国的七个地区分类的。这可以用一个“截面图”柱状图来显示，在该图中，每个柱（X轴）代表一个地区，柱的长度（Y轴）和每个地区的平均收入成比例。

27. 另一个办法就是将变量的域分为少数类别，从而将连续变量转为离散变量。譬如，有时可以根据住户收入或支出将住户分为20%的最贫穷住户，20%的次贫穷住户。在进行这种划分后，可以采用同样的方法按照上述方法显示两个离散变量的数据。一个具体的例子就是通过修改图十六.4来显示5个柱，每个柱代表每项收入的五分位数。

28. 两个拥有很多值的数值变量。统计员通常根据两个数值变量的相关系数（两个变量的协方差除以方差积的平方根）来提供这两个变量的摘要信息。但一般读者通常不熟悉这种统计方法。另一种办法就是用图的形式在一个散点图中显示数据，在散点图中，每个点代表一个观测值。譬如，这种图可以采用两种不同调查中同一住户的观测值（每个调查期间一个）来显示住户收入在两个调查期间的相关程度。

29. 使用散点图的一个问题是：当样本规模很大时，图将会变得过于“拥挤”以至于无法对其进行解释。为避免这一问题，可以从观测对象中抽选随机子样本（如，从观测对象中抽选十分之一的样本）。散点图的另一个问题就是如何对其进行调整以便考虑将抽样的权数考虑进去。一个简单的方法就是创建重复观测，每个观测的重复数为抽样权数。这无疑会使散点图变得过于拥挤，因此在创建重复观测后，只能将观测对象的随机子样本包括在散点图中。

表十六.7

1992-1993年越南按地区划分的住户总支出（千盾/年）

地 区	平均总支出 (圆括号内为标准误差)
北部高原	4 792(95.5)
红河三角洲	5 306(110.4)
中北部	4 708(107.7)
中部海岸	7 280(234.8)
中部山地	6 173(373.7)
东南部	10 786(398.5)
湄公河三角洲	7 801(167.4)
越南全国	6 531(77.6)

资料来源：1992-1993年越南生活水平调查。

说明：样本规模：4 799个住户。

5. 编制三个或更多变量的描述性统计资料

30. 从原则上讲，可以通过表和图的形式来显示三个或三个以上变量之间的关系。然而实际上却很少这么做，因为变量的增加不仅使变量之间的关系更加复杂和难以理解，而且通过简单表或图显示其关系的方法也变得更加复杂。在实践中，有时可以显示三个变量之间的描述性关系，但却几乎无法显示四个或四个以上变量之间的描述性关系。

31. 对于三个变量来说，最简单的办法就是将某个变量确定为“条件”变量。这个变量要么有少数离散的值，或者如果属于连续的值，那么应该通过计算其在域内少数几个区间中的分布情况，来对其进行“离散”。之后，就可以为该条件变量的每个类别或值建立单独的表或图。譬如，假设某人希望显示三个变量之间的关系：户主的教育情况、住户的收入水平和儿童营养不良的发生率。可以通过一个单独的表或图显示每种教育水平下收入和某个儿童营养状况指标（如发育迟缓发生率）之间的关系。譬如，这可以显示出，对于户主文化程度较高的住户，其收入和儿童营养之间的关系较弱。

C. 编制描述性统计资料的一般建议

1. 数据准备

32. 在生成有关数据，以便将这些数据纳入表中或图中之前，必须准备用以分析的数据。数据准备涉及三项不同的工作：对数据进行检验，以便删除那些很不准确的观测值；生成复杂的（衍生）变量；全面记录用于所有分析的“正式”数据集。在所有三项工作中，如果能在开始阶段格外努力并注意到细节情况，那么可以在将来的工作中节省很多时间和资源。本分节将对这些工作进行简单的介绍。更详尽的情况，请查阅第十五章。

33. 实际上，不管如何精心策划和实施每项住户调查，都会有一些变量存在着一些不太可信的观测值。这方面的问题涉及很多方面，从项目不答复

（见第十一章）或其他明显的误差（如，将三岁的孩子确定为户主）一直到不太明显的错误（如，某个住户收入很高，但支出却属中等）。在很多情况下，误差是由于纸调查问卷中数据录入不准确造成的，所以首先应该检验纸调查问卷。这种数据录入错误可以很容易解决。如果调查问卷上出现了异常的数据，那么可以有几种选择。首先，可以将有关变量的值视作“缺失”的值。如果这种情况不多，那么在根据该变量计算任何一个图或表的数据时，可以不考虑这些观测值。³⁰ 如果这种情况很多，那么可以将“缺失”值作为某个分类变量的一个独特类别进行计算，可以将其标为“未报告”或“未声明”。其次，如果大多数异常情况集中在少数几个住户，那么可以放弃这几个住户。第三，如果很多住户的某些变量有很多观测值存在着问题，那么需要决定是否应该报告有关变量的结果。

³⁰ 这种方法的缺点是：每个表的样本规模可能略有不同。尽管这会引起混乱，但在每个表的底端加一个注释，说明“有些观察观测值被略去了”，这种做法就足以能够说明情况明确了。

34. 解决缺失数据的方法有很多，其中之一就是“估算”缺失的值。估算法为未知情况、“未报告”情况以及值不真实的情况分配各种值，包括热卡估算法和最近邻法，根据这种方法，可以为不答复情况做“最准确的猜测”。采用这种方法的原因很简单：在某些特征上具有类似之处的住户或人员很可能在其他特征上也具有类似之处。如，某个村庄房子的墙壁和屋顶可能和其他农村地区房子的墙壁和屋顶类似，而与城市地区的房子可能不同。同样，某个住户内的大多数成员所信仰的宗教和所归属的种族将会相同。调查小组必须根据有关国家的人口、社会、经济和住房条件确定一个可以遵照执行的具体规则。

35. 尽管估算法非常有用，但可能存在着一些严重的问题。负责数据分析的小组成员必须决定：是根据情况逐一更换缺失的值还是采用某种估算法更换缺失的值。应该考虑到有关做法对最终表格的影响。估算1%或2%的情况对最终结果的影响可能很小，或没有影响。如果大约有5%的情况缺失或与其他项目不一致，那么还有可能采用估算法。但如果需要估算更大比例的值，如10%或10%以上的值，那么有关变量很可能不宜用于数据显示和分析，在这种情况下不应该显示该变量的结果。关于估算法和缺失值处理的详细建议，读者可查阅第八章、第十一章以及这些章节的参考资料。

36. 数据准备的另一方面就是对复杂的（衍生）变量进行计算。在很多住户调查中，总住户收入和/或总住户支出是根据很多变量的值计算的。如，总支出的计算通常需要将100个或100个以上的食品和非食品项目支出加总。从理论上说，这些变量的计算很简单，但在实践当中，可能会出现很多问题。譬如，在计算农村住户的农业收入和支出时，有时农业利润会出现负值。如果某个住户的数据出现异常情况，那么可以查看总计中的各个分项，这样可能会发现导致问题的一两项原因。继续以农业利润为例，购买的某些农用品，其价格可能非常高。在这种情况下，可以运用平均价格来重新计算利润。

37. 遗憾的是，在出现问题时准备数据集与其说是一门科学还不如说是一门艺术。在无法明确哪种选择更好时，需要对此做出决定。最后，应该对所做出的选择进行记录，一般的情况下，应该对“原始数据”被转化为表和图的

整个过程进行记录，这一点很重要。在进行记录时，应该对有关过程和用以操作、转化数据的所有计算机程序作简短的叙述。

2. 编制结果

38. 编制基本统计结果的最佳方法因调查类型和读者的不同而不同。但有一些一般性的建议几乎可以适用于所有的情况。

39. 最重要的一般性建议就是编制的结果要明确，这又包括几个更加具体的建议。首先，所有变量的定义都必须准确清楚。譬如，在以图表显示住户“收入”时，收入变量要么为“人均收入”，要么为“总住户收入”，但绝对不应只是“收入”。一些诸如收入和支出之类的复杂变量可以在文本中和图表的注脚中予以明确的定义。收入是税前收入呢还是税后收入呢？它包括业主占住的房子的价值吗？收入是指每周收入、每月收入还是每年收入呢？所有这些都必须非常明确。对于很多变量来说，在文中列出住户调查问卷中用以推导变量的字句很有用。如，对于有关成人文化水平的数据来说，应该非常明确地说明这一变量是如何定义的。可以根据有关人员上学的年数、签名的能力、受访者阅读报纸的能力对该变量进行定义，或者可以根据对受访者进行的某种测试进行定义。不同的定义可能会得出截然不同的结果。

40. 第二个具体的建议是：离散变量的百分比分布应该非常明确：是有关住户的百分比呢还是有关人员的百分比（也就是人口的百分比）。在很多情况下，这会导致不同的结果。在很多国家，文化水平较高的人员，其家庭规模也较小。这意味着：户主文化水平较高的住户人口比率小于户主文化水平较高的住户比率。第三个建议是：图应该显示出用以绘制图形的数据，如图十六.1中的柱状图显示出越南住户三个照明来源各自的百分比，图十六.2中的饼图也是如此。

41. 最后有其他方面的几个建议。首先，报告中不应该有太多的表格，每个表格中不应有太多的数据。有时统计机构提供了包括各种细枝末节的数百个表格，大多数读者都不太可能对这些细枝末节感兴趣。每个表格中的细节也常常如此。编制报告的人员应该对编制各种表格的目的进行讨论，如果某个表格或某个表格中的详细信息用处不大，那么应该将一些无关紧要的信息排除在外。第二，对于调查中某些最重要的变量，应该就其抽样误差估计编制报告。另外，显示重要变量或指标的置信区间也非常有用。这一点是不言而喻的，但却常常被疏忽。这说明有必要向读者传达住户调查信息的精确度。第三，每个表格应该列明样本的规模。

3. 构成一张好表格的元素

42. 本分节将就编制各种显示住户调查信息的表格提出一些具体的建议。在编制各种表和图时，可以遵循下述一般原则：表中的信息应足以让用户无需参照报告的正文就能够对它们进行准确的解释。这一点非常重要，因为很多用户不会参考报告的有关内容，他们只是拷贝报告中的表格以备将来使用。

43. 以下只是一些一般性的建议。对于任何一项调查来说，调查小组必须决定哪些做法最合适。一旦选定了有关做法，那么就应该严格遵照执行。但在某些情况下，为了说明某一点或显示统计分析的具体类型，可能有必要偏离某种做法。最后，有关表格的所有准则几乎都适用于各种图。

44. 构成一张好表格的各元素见表十六.6。每个表应该包括：一个明确的标题、地理标志（如有必要）、列标题、行标题、数据来源以及有关的注释。

45. 标题。标题应该对表格进行简要的描述。描述内容包括：(a) 表的编号；(b) 被考虑的人口或其他全域（包括分析单位，如住户或个人）；(c) 说明行中的内容；(d) 说明列中的内容；(e) 调查所涉国家或地区；(f) 调查的年限。

46. 至于表格编号，大多数报告的表格都采用连续的编号，如从表十六.1开始一直到最后。有时，有些国家采用字母和数字对不同组别的表进行编号，如住房表采用H01、H02……等编号，人口表格采用P01、P02……等编号。这种办法的优点是简单明了，但缺点是报告受编号的限制，增、减编号非常麻烦。

47. 全域是表涉及的人口或住房基数。如果所有人口都包括在表中，那么标题中可以省略全域：所假设的是总人口。相反，如果表格包括子总体，如劳动力人口，而且潜在的劳动力被定义为10岁或10岁以上的人员，那么标题就应该包括如下的措辞：“年龄在10岁或10岁以上的人口”。

48. 表十六.6的标题还显示了表格的行和列所包括的内容。尤其是，该标题显示：该表显示了各类保健设施类型的使用信息（行），而且有关信息是按城乡地区分别列出的（列）。将国家或地区包括在标题中将使地理全域一目了然。如果研究人员希望对不同国家的调查结果进行比较，这种做法就非常重要。显然，收集数据的国家统计局当然知道自己国家的名称，但不同国家的人员在使用有关表时可能需要这方面的信息来对各国进行区分。

49. 最后，标题还应该包括调查的年份，以明示调查期间。有时，一个国家的统计机构可能希望在同一表中显示两个或两个以上不同调查的数据。那么标题中可能会出现两个日期，如，“1990年和2000年”或“1980年至2000年”。调查小组应该决定是否要写出一系列的日期（如“1980年、1990年和2000年”，而不是简单但不够全面的日期，如“1980年至2000年”）；一旦做出决定，国家就应始终执行。

50. 地理标志。如果对较低一级的地理区域重复使用相同的表格，那么每个表格的标题中应该包括一个地理标志以明确哪一种表格适用于哪一个地理区域。如，如果对越南的七个地区重复使用表十六.6，那么应该在紧挨着表格标题的下一行用括号标明地区的名字。也可以使用“非地理标志”。譬如，可以对主要族群或民族重复使用相同的表格。

51. 列标题。表格的每一列都应有一个“标题”。列标题可以有一个以上的“级别”。如表十六.6中，头两列的标题为“城市地区”，最后两列的标题为“农村地区”，在“城市地区”和“农村地区”标题下又分别列有观测值“频数”和“百分比”分布的标题。列标题的另一个内容就是“总计”或“总数”，如表十六.3的第一列。至于这些列应该放在什么位置，调查小组应该做出决定。传统的做法是将总计放在最后，所有特征首先在各列显示。但，如果表格有很多页，而且有很多列的信息，那么调查小组可能倾向于将总计放在各列中的前面（左方）。如果将总计放在前面，那么读者无需翻阅表格的全部内容就可以立即知道各列的总计。

52. 在确定列标题及其相应数据列的间隔时，应尽可能减少页面的空白处。应该考虑到列中最大数的位数、属性名称的字母数量以及所用字体允许的总“间隔”数。使用的字体很重要，因而应该在制表的早期选好字体。

53. 行标题。调查小组必须就行标题的具体做法做出决定。行标题应该在左边，每行只能列一个变量。行标题应该包括行所显示的变量名称，可以包括子类别（嵌套变量）。如，一个行的“组”标题下可以分设两行，一行为男性，一行为女性。需要确定一些习惯的做法，以便区分不同的组标题。通常的做法是：对于不同“级别”的变量，采用不同的缩进方式。

54. 数字的精确度。很多表格提供的重要数字过多。在显示百分比时，基本上在小数点以后保留一位就足够了。在小数点以后保留两个或两个以上的位数很少能提供有用的信息，而且还有三个缺点：分散读者的注意力、浪费空间、使人对精确度产生错误的感觉。四到五位数的数字很少需要小数点。在显示大的数字时，应该以“千”或“百万”表示，这样做的目的是为了显示的数字不超过四位数或五位数。

55. 资料来源。数据来源应该为调查的全名，通常在表格的底端（见表十六.6）。但有时表格显示的是一个国家中一项以上的调查，或一个以上国家的调查。在这种情况下，资料来源中的信息就更加重要。调查名称中还应包括日期。如果资料来源属于公布的报告，那么有必要区分报告的公布日期和收集资料的日期。譬如，一个国家收集资料的时间可能为1990年，但数据公布的日期为1992年。因此资料来源应该为“1990年生育调查，1992年”，1992年为公布日期。

56. 说明。说明提供的是直接的信息，该信息对表中的结果进行合理的解释。譬如，表十六.1和十六.2中的说明表明被抽样人口包括独居或集体居住的所有人员。除了表格底端的说明外，和表格一起的正文中还应该包括一系列的定义和解释。正文应该包括各种特征的定义，如正文可用来说明出生地是指母亲去医院生产前的居住地，而不是医院所在地。正文还可能就如何获得和使用数据做出解释。譬如，如果收集的是生日和年龄方面的数据，若生日和年龄不一致，应该以生日为准，这一信息有助于某些用户，如人口统计学家，对解释数据的最佳方法进行评估。

4. 权数的使用

57. 本分节简述在根据住户调查数据编制图表时如何使用权数。更详细的内容，见第二、六、十九、二十和二十一章及其相应的参考资料。

58. 至于调查加权，最简单的住户调查样本设计就是“自行加权”的样本设计。在这种情况下，分析中实际不需要使用权数，因为人口中每个住户在样本中被抽选的概率是相同的。本章在几个例子中所使用的1992年-1993年越南生活水平调查就是这一类的调查。然而不同住户的回答率存在着差异，这说明应该计算权数来纠正这种差异。更重要的是，大多数住户调查并不是自行加权的调查，因为这些调查为特别关心的某些总体抽选的样本非常大。对于这些调查来说，必须使用权数来反应抽选概率的不同，以便正确计算调查中有关特征的无偏估计值。

59. 准确的权数必须包括三个部分。第一个部分包括“基权数”或“设计权数”。这些权数解释调查的初始样本设计规定的不同住户群之间的抽选概率的变差（也就是在样本不是自行加权样本的情况下）。第二部分是对不答复率的变差进行调整。譬如，在许多发展中国家，相对于中低收入住户而言，较富有的住户更不太愿意接受访谈。基权数需要根据所有住户群答复率的倒数来“增大”。最后，在某些情况下，可能会进行“事后分层调整”。事后分层的基本原理是：某个独立的数据来源，如普查，可能会按年龄、性别和族裔群体提供较为精确的人口分布估计值。这样，如果有关这些分布的调查估计值和这些独立来源的估计值不一致，那么可能需要对调查数据重新加权，以便使两种分布估计值保持一致。第二部分和第三部分的详细情况见Lundström和Särndal（1999年）。

D. 为住户调查编制综合报告（摘要）

60. 大多数住户调查首先通过出版一份综合报告的形式来公布其调查结果，综合报告较为详尽地提供了调查所收集的全部信息。相对于充分利用某些数据的专门报告而言，这种报告的散发范围更广。这些综合报告有时又被称作“统计摘要”。本节根据Grosh和Muñoz（1996年），就这些报告的编制提供一些具体的建议。

1. 内容

61. 综合报告的主要材料是大量的图表。这些图表应该能够反映调查收集的所有重要信息。范围较窄的主题，应由侧重点更强的专题报告进行深入分析。有关表格应该附有少量文本，文本内容应该足以说明有关表格中的信息类型。尽管可以做出的解释对未来的研究有作用，但没有必要得出特定的政策性结论。

62. 可以根据地理区域、性别或年龄对最重要的基本信息进行细分。如果调查包括收入或支出数据，可以根据收入或支出群体进行细分。在有些国家，这些不同群体之间存在着很大的差异，可以在另外的表中对这些差异的性质做

进一步的探讨。在其他国家，这些差异可能不会很大，这样就没有必要对其进行更为详尽的探讨。

63. 除了包括住户调查数据的结果外，综合报告还应用数页的篇幅对调查本身进行介绍，包括样本规模、样本设计、调查的起始日期和中止日期以及收集数据的方式。所用的调查问卷应该包括在主要报告的附件中。

2. 程序

64. 一份好的综合统计报告是由一组人员编制的，其中最好包括一些以前有过编制报告经历的人。一些小组成员应该集中负责编制图表的技术性工作，另一些人员则负责表格的内容和表格所附的文本。技术型小组成员可以选择其最为熟悉的统计软件，因为大多数统计软件都能够生成图表所需要的数据。但标准误的估计可能需要专用于此目的的软件，因为住户调查样本设计实际上总是非常复杂，标准统计软件包无法对其进行适当处理（有关这些问题的讨论，见第二十一章）。

65. 负责图表内容的小组成员应该就报告所载主题咨询政府机构相关专家的意见，这将能够确保有关图表能够以最适合这些机构使用的方式来编制数据。向国际援助机构征求意见也可能很有用，因为这些机构在进行项目策划时也有可能用到这些数据（至于如何组建有效的调查小组，详细的讨论见第三章）。

E. 结论意见

66. 本章简单介绍了如何根据住户调查数据编制简单描述性统计资料，所涉及的只是最基本的一般性做法。由于所介绍的大多数内容只不过是一些常识，所以数据分析师如果在进行调查分析时遇到特殊问题，就应该运用自己的常识。还可以运用更为复杂的方法对住户调查数据进行分析，其中一些方法将在以后的章节中进行讨论。总之，任何一个住户调查的数据分析应该适合调查的主要主题和目标，研究人员应该查看专业书籍和杂志，以便获得指导性意见来处理与这些主题有关的问题。

参考资料

- Frankenberg, Elizabeth (2000年)。“社区与价格数据”。《设计发展中国家住户调查问卷：15年生活水平衡量研究的教训》，M. Grosh和P. Glewwe等人。纽约：牛津大学出版社，世界银行。
- Grosh, Margaret和Juan Muñoz (1996年)。《规划和实施生活水平衡量研究调查手册》。《生活水平衡量研究工作文件，第126号》。哥伦比亚特区华盛顿：世界银行。
- Lundström, S. 和C. E. Särndal (1999年)。“将校准作为样本调查中对待不答复情况的标准方法”，《官方统计杂志》，第13期，第2号，第305-327页。
- Tufte, Edward (1983年)。《数量信息的显示》。康涅狄格，柴郡：制图出版社。
- Wild, C. J. 和G. A. F. Seber (2000年)。《抓住机会：数据分析与推理入门》。纽约：威利国际出版公司。

第十七章

利用多主题住户调查改善发展中国家减贫政策

Paul Glewwe

明尼苏达大学

应用经济学系

美国明尼苏达州圣保罗

摘 要

本章阐述发展中国家研究人员和政府官员如何利用住户调查来制订减贫政策。本章从相对简单的描述性分析入手，突出强调住户调查数据发挥的关键作用：它们所提供的信息显示哪些人是贫穷人口，以及这些贫穷人口所具有的特征。然后，本章利用多元回归技术，讨论更加复杂的多变量分析。对于每种分析，都有例子说明是如何利用住户调查数据来制订减贫政策的。

关键词：贫穷、政策制订、描述性分析、多变量分析。

A. 导言

1. 几乎所有的发展中国家都接受这样的事实：经济与社会发展的首要目标之一就是减少并最终消除贫穷。虽然所有的政府都可能抱有同样的目标，但他们所实施的减贫政策却不尽相同。贫穷的性质以及贫穷人口的特征因国家而异，因此，减贫政策也应当因地制宜。

2. 本章内容过于概括，无法对发展中国家政府政策如何能够影响贫穷的诸多方法进行详细探讨。关于最新处理方法的详细情况，见Lipton与Ravalion（1995年）及世界银行（2001年）。尽管如此，本章还是可以对此进行概述，为方便起见，本章将政府政策分为4大类。第一类包括宏观经济政策，即影响经济增长与稳定的全面经济政策。最重要的宏观经济政策包括税收与政府支出的总体水平、（影响利率和通货膨胀率的）货币政策、（影响汇率、外贸以及外国资本流动的）国际经济政策、以及有关银行和其他金融机构的政策。第二类包括那些对价格产生影响的政策，如对具体货物与服务的税收与补贴。第三类政府政策是政府如何提供公共服务和基础设施，如保健中心、学校以及交通和通讯网络。最后一类包括旨在向贫穷人员提供直接援助的政府计划。这些政策的例子是墨西哥的教育、卫生和粮食方案（教卫粮方案），该方案为子

女正常上学的贫穷家庭提供现金资助。此外还有牙买加的食品券计划，政府向贫穷家庭提供一种食品券，贫穷家庭可以凭票在当地商店购买食品。所有这四类政策对贫穷都会起到重要作用。

3. 这几类政策的任何一种对于贫穷产生的影响都取决于贫穷人口的特点及其行为，有时，还取决于非贫穷人口的特点和行为。例如，政府对特定食品的补贴会降低这些食品的价格，这种政府补贴对贫穷的影响取决于贫穷人口购买这些食品的程度。这就是说为选择最有效的减贫政策，各国政府需要有关其国内贫穷人口的特点和行为的信息。住户调查提供这方面的重要信息。

4. 几乎所有发展中国家，甚至最贫穷国家，都会进行某些形式的住户调查，如收入和支出调查、劳动力调查、人口与健康调查。这些调查提供大量信息，可用来更好地了解贫穷性质，分析政府政策可能对贫穷人口发挥的作用。本章阐述如何将发展中国家住户调查用于制订减贫政策。B节首先说明在根据调查数据计算出的简单描述性统计指标中，我们可以了解到什么。C节讨论基于多变量分析的更复杂方法，其后是简短的结论部分。

B. 描述性分析

5. 为确保政府减贫政策和计划行之有效，政府需要了解政策是否能够真正落实到贫穷人口以及这些政策的效果。遗憾的是，发展中国家通常缺少这方面的信息。如，加快经济发展速度的某些政策措施可能会使某些职业的收入增幅大于其他职业。这就产生了一个问题，哪些职业中的贫穷人口比例最高。在价格政策中也存在同样的问题。如果政府计划对诸如石油之类的产品提高税收，这种措施对贫穷人口的影响将取决于贫穷家庭是否大量使用这些产品。对于是否应在某些地区开办新的学校和诊所存在着同样的问题：即这些地区的贫穷家庭是否相对集中。最后，至于向贫穷人口提供直接福利的任何计划，无论以服务、实物还是以现金形式提供，计划管理人员都想知道该计划的受益人有多少人是贫穷人口，又有多少贫穷人口能够从该计划中受益。

6. 遗憾的是，许多发展中国家对于贫穷人口的分布情况及其特征知之甚少，因而他们对于贫穷人口从政府政策和计划中受益的情况，或受害程度了解不多。住户调查可以弥补这方面的许多信息差距。本节通过许多发展中国家的实例，讨论如何可以做到这一点。虽然利用住户调查数据来了解贫穷状况都非常简单，只是生成一些简单的图表，但此类信息通常比通过更复杂分析得到的信息更具价值。

1. 贫穷的定义

7. 在调查政府政策对贫穷人口产生的影响之前，人们必须清楚哪些人属于贫穷人口，这就要求对贫穷进行定义。人们对于贫穷总是看法不一。但是，如果个人和家庭有机会过上丰衣足食的生活，对于它们原则上应当达到的最低“体面”生活水准，人们的看法大体一致。大多数有关贫穷的讨论都将重点放在物质需求上，而不是放在政治自由、人权以及心理健康上面，本章也是如

此。意见最为一致的最明显物质需求包括：(a) 足够的饮食；(b) 基本的住所/住房；和(c) 饮用水以及卫生的废物处置方式。大多数观察家还将基本的受教育机会和简单的预防性保健列入其中。有些人则主张将更多的实物和服务，如文化或娱乐活动，包括在内。但对于应包括哪些实物和服务，人们意见不一，甚至是否应该将这些东西或服务包括在内也没有统一的意见。

8. 哲学家、经济学家以及社会科学家可能会花大量的时间，而且通常已经花费了大量的时间讨论个人和家庭为了不被视为贫穷而对一揽子货物和服务的最低需求。一旦将最低限度的一揽子货物和服务需求确定下来，那么只要消费中缺少该一揽子需求中的项目，就可以将所缺少的项目视作贫穷的“指标”。被许多经济学家广为采纳的另一种更为实用的方法认为：几乎所有的一揽子货物和服务项目都需要花钱，所以真正的问题并不在于一揽子货物和服务的具体内容，而在于其货币成本。这种方法以一定数量的货币作为“贫穷线”，那些收入或支出低于此金额的个人和家庭将被列为贫穷人口或家庭。事实上，发展中国家中，很多货币形式的贫穷线都必须首先确定满足人们基本需求的一揽子货物和服务。例如，在一揽子需求中，有一项就是能够满足最低营养需求的一系列食品品种，而这也反映了国家的食品消费结构。其次就是要计算一揽子需求的价格。本章的其余部分假定采取这种方法；至于如何划分贫穷线，详情见Ravallion（1998年）。

2. 描述贫穷概况

9. 一旦根据住户收入或住户支出对贫穷做出了可行的定义，便可利用住户调查数据对贫穷进行描述。该描述通常被称为“贫穷概况”。为了确定贫穷概况，需要利用住户调查中的收入和/或家庭支出数据，来计算每个家庭的总购买力（总收入或总支出）。那些购买力低于贫穷线的家庭将被确定为贫穷家庭。

10. 上一段包含一个隐含教训和一个隐含问题。教训是贫穷分析要求住户调查数据中包括家庭总收入和/或家庭总支出方面的准确信息。没有这方面的数据，贫穷分析将十分困难，否则就必须寻找其他方法来确定家庭是否贫穷。虽然在没有这类数据的情况下也可以从一项调查中获得某些有用的信息，但如果调查收集了收入和/或支出方面的数据，那么可以从中了解到更多东西。问题是：如果一项调查包括收入和支出方面的数据，那么应该使用哪种数据呢？一般来说，使用支出数据比较合适，因为支出数据通常比收入数据更准确，并且从理论上讲，消费支出和家庭福利的关系更为密切，原因是收入有时被用以偿还债务或被存起来用于未来消费，所以不一定能够反映当前的福利状况。

11. 确定贫穷概况的首要工作就是要说明谁是贫穷者。如果没有住户调查数据，决策者和其他观测人员通常很难了解贫穷的是哪些人，贫穷人口的特征又是什么。更糟糕的是，他们所持有的某些看法可能不准确。如很多政府官员和其他观测人员的大部分时间都是在大城市度过的，所以会根据他们在这些城市地区所看到的情况来看待贫穷问题。但实际情况是，几乎所有的国家中，农

村的贫穷比率要比城市高出很多。因此，使用住户调查数据的首要工作是估计贫穷的比率；根据城市、农村以及国家的地区来描述贫穷的区域，并计算贫穷的某些基本特征。应该根据种族或宗教团体、教育水平和职业对贫穷率进行检验，这一点很重要。考察贫穷人口的住房条件以及生产性资产的所有权情况也很有用。有了这些和其他方面的信息，就可以为决策者提供有用的建议。

12. 譬如，世界银行（1999年）最近就越南贫穷情况所做的报告中就包括了贫穷的某些基本特征。根据估计，1998年越南有37%的贫穷人口。在越南，有79%的穷人从事农业方面的工作，这些人员当中，几乎所有人都是自营职业者。另一个基本事实是，少数民族的贫穷率要高出很多：尽管少数民族只占总人口的14%，但却占越南贫穷人口的29%。

13. 贫穷最重要的特征之一就是其居住地。从理想的角度来说，决策者希望了解每个城镇和农村地区的贫穷比率。遗憾的是，一般住户调查的样本规模通常为3 000-15 000个住户，这种规模太小，因而无法在这样的分类水平上对贫穷做出精确估计。然而如果有最新的普查数据，那么可以将这些数据和住户调查数据结合起来，这样就可以对更小地理区域的贫穷情况进行估计。基本思路是根据住户调查数据对各种“预测性”变量和住户收入或支出之间的关系进行估计。所使用的预测性变量应该是普查中也存在的变量。在对预测性关系进行估计后，可采用普查数据来模拟较小地理区域的支出分布情况，这样就可以对这些地区的贫穷率进行估计。采用厄瓜多尔提供的数据在这方面举出的例子见Hentschel等（1998年）。至于所采用的方法，其详细讨论见Rao（2002年）和Kalton（2002年）。

14. 确定贫穷定义和贫穷概况的最后一个要点就是，人们通常希望对同一国家中不同时间的贫穷情况进行比较，或对不同国家在同一时间的贫穷情况进行比较。在进行这种比较的时候，不同时间或不同国家用以定义支出或收入的住户调查数据应该是按同样方式收集的，这一点很重要。调查问卷设计的微小差异或数据收集方法的其他变化通常会导致估计值发生完全可以乱真的显著变化，而且这种变化通常是无法预料的。坦率地说，如果被比较的各项调查所收集的数据和/或数据分析方式不同，那么将无法进行这种比较。因此，如果为用以界定贫穷的变量收集数据，那么收集数据方式的任何变化都必须予以谨慎考虑，此举的目的是为了防止观测的变化由于统计程序的缘故而与实际变化不符。因此，通常最好不要大幅度改变收集数据的方式。

3. 利用贫穷概况进行基本政策分析

15. 为了给决策者提供建议，首先应了解贫穷人口的所在地及其某些基本特征。当然，应该在贫穷人口最为集中的地方实施为贫穷人口提供援助的特别项目，但如果能够对有关贫穷人口的简单统计进行正确分析，那么项目的实施将会更有成效。本节将对贫穷方面的四种基本信息进行介绍，这些信息可用来在各种政策对贫穷人口的影响方面汲取教训。

16. 贫穷人口如何获得收入。如上所述，政府政策对贫穷人口产生影响的一种方式就是对贫穷人口的收入施加影响。因此一个重要的问题就是贫穷人口应该靠什么谋生。也许应该询问的第一个问题就是贫穷人口是否靠自营职业而不是受雇于某个雇主来获取收入。在很多国家，绝大多数贫穷人口都是自营职业的农民、手工艺者或商人。有些政策可对工薪人员产生影响，如最低工资法律的变动或只适用于工薪人员的“社会保障”或健康保险计划的实施，根据定义，这些自营职业的贫穷人口不会直接受到这些政策的影响。

17. 由于大多数贫穷人口都属于自营职业的农民，所以一个重要的问题就是：他们生产的农作物是什么？其生产的农作物中有多少被出售？在这方面科特迪瓦提供了一个具体例子。Glewwe与de Tray（1990年）发现，象牙海岸的很多贫穷农民种植棉花，而非贫穷农民却很少种植棉花。这样，如果政府通过政策对棉花价格施加影响，将主要会对该国的贫穷人口产生影响。

18. 贫穷人口的消费结构。贫穷人口的经济福利取决于其消费的商品和服务价格。如，加纳占总人口20%的最贫穷人口中，拥有摩托车或汽车的人口不到1%（Glewwe和Twum-Baah, 1991年）。这意味着提高汽油价格对加纳贫穷人口所产生的直接影响不大，当然公交成本的上升可能会对他们产生间接的影响。

19. 从更一般的角度来看，有关食品和非食品项目消费方面的数据，以及有关供电和自来水使用情况的数据能够为决策者提供大量的信息。如果考虑对某种商品征税或实施补贴，那么应该对数据进行检查，以确定有关政策对贫穷人口的影响程度。还应该注意，汇率政策对价格也会产生影响，因此也需要了解贫穷人口消费进口产品的数量。以上有关加纳的情况就是一个恰当的例子：加纳所有的石油产品都是进口的。

20. 贫穷人口使用的服务。对健康和采取补贴政策是合理的，至少在某种程度上来说如此，理由是这种补贴为贫穷人口带来了惠益。但健康服务的类别有很多，教育的类型和水平也各异。如果通过数据显示这些服务的使用者是谁，就可以对具体政策受益者的贫穷状况进行考察。

21. 这方面的一个最新例子来自越南。Gertler与Litvack（1998年）发现，在占总人口20%的最贫穷人口中，普通人每年去政府医院门诊看病一次，去乡镇卫生所门诊看病大约两次。与此形成鲜明对比的是，在占总人口20%的最富裕人口中，普通人每年去政府医院门诊看病四五次，去乡镇卫生所门诊看病大约只有一次。这种差异的主要原因是政府医院主要在城市地区，而该国大约90%的贫穷人口生活在农村地区。这一简单数据显然说明，对乡镇卫生所实施补贴为穷人所带来的好处多于为非穷人所带来的好处，而对医院实施补贴为非穷人带来的好处多于为穷人带来的好处。

22. 方案参与。直接使用住户调查数据的最后一种情况就是考察谁参与政府的各种减贫方案。这要求住户调查列有一个或一个以上有关住户参与方案情况的具体问题，同时还应该有收入或支出方面的数据，以便将住户分为贫穷住

户或非贫穷住户。尽管过去有关这方面的数据很少，但随着调查设计者对其价值的认识，这些数据正日益变得普遍。

23. 在利用住户调查对方案目标的选择进行评估方面，牙买加提供了一个例子（Grosh, 1991年）。住户调查数据显示：更有可能使用食品券的是贫穷家庭而不是非贫穷家庭，也许这一点并不令人吃惊。但荒谬的是，从普通食品补贴中获得好处的往往是富裕家庭。这一信息于1980年代转达给了政府，1990年代初，政府中止了食品补贴，同时将食品券方案的津贴增加了一倍。

24. 基本描述性分析的最后一点就是：几乎所有住户调查都是以复杂的样本设计而非随机样本为基础。因此，将会从特别关注的分组人口群，如贫穷人口中抽选过多的样本，这说明必须使用抽样权数来获取无偏误的基本描述性统计估计值。另外，在计算标准误差时必须将样本设计考虑进去。第十六章和本书其他章节对这方面的内容进行了详细讨论，因此，读者在进行描述性分析之前，应当先查阅这些章节的讨论内容。

C. 住户调查数据的多元回归分析

25. 有关住户调查使用的上述例子以非常简单的统计为基础，能够使用简单统计软件包的任何人都可以对这种简单的统计进行计算。然而，在政策方面从中获得的教训可能过于简单，因为它们忽视了人们对这些政策做出的行为反应。譬如，如果由于某种农产品通常是由贫穷人口生产的，而决定取消该产品的税，那么当这种产品价格上升时，非贫穷家庭也有可能开始生产这种产品，这样，有关政策的一部分惠益就会转移到非贫穷家庭。³¹ 同样，如果贫穷人口对某种食品的消费量很大，那么对这种食品征税似乎会对他们产生很大的负面影响。但如果不对另外一种类似食品征税，贫穷人口可能会转而消费这种食品，这样，他们的福利仅有少许减少。另一个例子与教育有关。如果一个国家的贫穷孩子很少上高级中学，那么这种情况说明，降低这些学校的学费不能给贫穷人口带来多少好处，但这种措施却有可能大大增加贫穷孩子在这些学校上学的人数。这种可能性又说明，单纯从现有入学结构来看问题，将会低估这种政策对贫穷人口所带来的好处。

³¹ 在有税收的时候，生产者出售产品所收到的价格将低于消费者支付的价格，其差额即为税额。当税收被废取消除后，生产价格将等于消费价格。几乎在所有的情况下，这都将意味着生产者出售产品收到的价格将提高，消费者支付的价格将下降。

26. 可采用住户调查来估计政策变化所引起的住户行为变化。这项工作并不容易，因为它需要进行更复杂的分析。进行这种估计的最常用方法就是多元回归分析。最复杂的方法通常采用专门设计的住户调查所提供的数据，专门设计的调查收集进行所述分析所需的精确数据。这是有必要的，因为这些方法所需要的数据通常是普通住户调查无法提供的。本节将介绍三种常见的方法，以便利用住户调查数据，估计各种政策如何影响住户行为。详情见Deaton（1997年）。

1. 需求分析

27. 经济学家通常就价格和住户收入对商品和服务购买情况所产生的影响进行估计。这种研究被称作需求分析。这种分析的一般概念是：对于任何商品

(i) 来说, 某个住户购买该商品的数量(q_i)取决于该住户的收入(y)以及该商品的价格(p_i)和所有其他商品的价格。其公式如下:

$$q_i = f(y, p_1, p_2, \dots, p_i, \dots, p_n) + \varepsilon \approx \beta_0 + \beta_1 p_1 + \beta_2 p_2 + \dots + \beta_i p_i + \dots + \beta_n p_n + \beta_{n+1} y + \varepsilon$$

函数 $f(y, p_1, p_2, \dots, p_i, \dots, p_n)$ 非常概括地说明了收入和价格是如何影响住户需求的, 该公式中的 ε 反应了其他原因的影响作用(也可能是与其他原因无关的 q_i 中的随机变化)。一个简化的常见需求分析是假设一个线性表达式, 在此, 用“ \approx ”符号右方的数项表示, “ \approx ”说明这种简化是一种近似。如果 ε 与 y 和价格变量没有关系, 那么在这个线性关系式中可以使用简单的普通最小平方(OLS)获取收入(y)和价格(p_i)系数(β 's)的无偏估计值。在实际运用中, 这种假设可能站不住脚, 有很多其他方面的估计问题会使分析更加复杂。关于需求系统估计的其他信息, 经典的参考资料有Deaton与Muellbauer(1980年)。更新处理方法, 见Pollack和Wales(1992年)及Lewbel(1997年)。

28. 除了采用简单的描述性统计获取信息以外, 为了了解需求分析如何提供其他信息, 可以考虑税收对某种进口粮食(如小麦)的影响。(发展中经济体通常对进口物品征税, 因为这种税收较容易征收。在很多发展中国家, 由于热带气候的原因, 进口是小麦的惟一来源。)假设一公斤面粉的现有价格为10, 普通贫穷家庭每年消费60公斤的面粉。再假设进口价格是按国际价格确定的, 这样, 如果对小麦进口征收50%的税, 那么面粉的价格将会增加到15, 这就意味着普通贫穷家庭将需要额外支付300(5x60)的税金。当然, 根据简单描述性统计所进行的这种分析假设贫穷家庭将在征税后继续购买同等数量的面粉。事实上, 如果面粉价格上升, 那么各家庭将很有可能减少面粉的消费而增加其他主食的消费(如, 大米、玉米或木薯)。可以通过需求分析估计来计算这种行为反应的大小。假设前段中的等式描述的是小麦需求, 那么 q_i 代表住户每年购买面粉的公斤数, p_i 代表每公斤面粉的价格。如果 $\beta_1 = -3$, 那么面粉价格上升5将会使面粉消费量减少15, 这样一个普通贫穷家庭的年消费将为45公斤。这反过来又意味着普通贫穷家庭将需要额外支付225(5x45)的税金, 而不是300的税金。尽管这一例子非常简单, 但它说明在考察具体政策的影响时, 应考虑住户行为。

29. 至于使用需求分析来分析政府政策对贫困人口的影响, 这方面的例子见Deaton、Parikh和Subramanian(1994年)。作者根据一系列的需求等式为十多种不同食品进行了估计。他们计算了食品价格上升对国家社会福利的总体影响以及价格变化对贫困人口福利的影响程度。一个特别有趣的结果是: 大米价格的上升对贫困人口福利的负面影响要比粗粮价格上升对贫困人口福利的负面影响小, 这是因为贫困人口对粗粮的依赖程度更大。这样大米征税对贫困人口的影响不如粗粮征税对贫困人口的影响大。

2. 社会服务的使用

30. 健康和教育方案可以为贫穷家庭提供很多惠益, 但对方案的参与并不一定意味着得到了大量惠益。在这些方案中, 有些方案可能无效。就健康领域

而言，决策者可能希望了解对方案的参与是否真正改善了个人的健康状况。至于教育，决策者可能希望了解孩子们通过上学实际学到了多少东西。已经根据发展中国家的住户调查数据进行了很多研究，这些研究试图调查健康和教育方案在实现其目标时到底有多成功。

31. 说明多元回归运用的一个最新例子就是分析具体学校特征对学生学习以及未来工资的影响。Glewwe（1999年）对这一问题进行了考察，他根据加纳的住户调查数据就学校和住户特征对孩子学业成绩的影响（通过考分度量）进行了估计。所采用的等式如下：

$$T_i = \beta_0 + \beta_1 MED_i + \beta_2 FED_i + \beta_3 y_i + \beta_4 IQ_i + \beta_5 SC_{1i} + \beta_6 SC_{2i} + \dots + \varepsilon$$

式中， T_i 是孩子i的考分， MED_i 和 FED_i 分别是孩子i的父母的教育水平。 y_i 是孩子i的家庭收入，SC变量代表大量的学校和老师特征。根据这种等式进行估计非常复杂（参见Glewwe, 2002年）；但一旦估计出 β ，就可以提供信息说明不同学校和老师特征是如何影响学生的成绩。将这些影响结果与各种学校和老师特征的成本进行比较，将有助于了解什么类型的教育支出最有成效。

32. 通过对加纳数据进行分析，Glewwe发现维修教室的破漏屋顶、为没有黑板的教室提供黑板能大大提高学生的成绩和学习期限（在校年限）。针对提高学校质量进行的这种“投资”，Glewwe简单计算了其经济回报率，结果显示这些回报率很高，有时达25%或25%以上。

3. 政府具体方案所产生的影响

33. 尽管可以很容易地运用住户调查数据考察某个住户个人是否参与了某种旨在为贫穷者提供帮助的方案，但却较难确定参与有关方案实际使福利得到提高的程度。此处的问题是参与有关方案可能会带来其他影响，这些影响可能会降低福利。如，“工作换食品”的方案可以为贫穷人口提供就业，但必须根据工作成本（包括工作对其健康的影响）来度量收入增加所带来的好处。同样，如果向住户提供食品券以便增加他们对食品的消费，那么使用这些食品券购买食品不一定会增加食品的消费量，因为他们很可能会将原本用以购买该食品的一部分或全部现金挪做它用。在估计各种方案对住户行为所产生的影响时，需要认真进行深入的计量经济学分析，以便了解方案参与所带来的全部影响，从而最终了解方案参与对住户福利的总体影响。

34. 这方面的最新例子见Jacoby（2002年）的论文，文章对菲律宾学校供餐方案产生的影响进行了研究。该论文就校方为学生供应午餐是否会减少家长在家中为他们提供食物的数量进行了研究。尽管大多数经济学家预计学生在家中所吃食物会出现再分配现象，但Jacoby没有发现有任何迹象表明存在着这种膳食转移情况。相反，他发现参与学校供餐方案对孩子在家中的食品消费没有任何影响，也就是说，对于参与学校供餐方案的孩子来说，学校提供的食品使他们的食品消费总量增加了。

D. 小结

35. 住户调查可以提供大量的信息，决策者和方案设计者可以运用这些信息就各种政策和方案是否能够给贫穷家庭带来好处进行评估。为了使所提供的信息更有用，调查必须包括收入和支出数据，以便将住户划分为贫穷住户和非贫穷住户，另外调查还应该包括住户如何受某项政策或方案影响的数据。直到最近，所使用的住户调查通常都是为其他目的而设计的。但在1980年代和1990年代，出现了各种新的住户调查，其用意显然是提供此类信息。其中最突出的是世界银行有关生活水平衡量研究（LSMS）的住户调查。Grosh与Glewwe（1998年）对这些调查进行了简要介绍，而Grosh与Glewwe（2000年）则对这些调查进行了极其详细的介绍。然而，即使是用于其他目的的常规调查，也可以通过增加几个问题来使其对贫穷状况分析起到更重要的作用。譬如，在常规的住户收入和支出调查中，增加有关国家减贫方案（如农村就业方案或食品券方案）参与情况的问题将非常有用。

36. 本章向读者概要介绍如何利用住户调查设计可减少发展中国家贫穷现象的政策。由于本出版物的篇幅有限，所以只能做一些简要的讨论。读者若想了解更多详细的情况，可查阅本章所列的参考书籍和论文。

参考资料

Deaton, Angus（1997年）。《住户调查分析：发展政策的微观经济方法》。马里兰，巴尔的摩：约翰·霍普金斯大学出版社。

_____和John Muellbauer（1980年），《经济学与消费者行为》。联合国，剑桥以及纽约：剑桥大学出版社。

Deaton, Angus、Kirit Parikh和Shankar Subramanian（1994年）。“马哈拉什特拉的粮食需求模式与定价政策：用于住户一级调查数据的分析”。Sarvekshana, 第17期，第11-34页。

Gertler, Paul和Jennie Litvack（1998年）。“在转型期获得保健：私营部门在越南的作用”。《住户福利与越南的转型》，D. Dollar、P. Glewwe和J. Litvack等。哥伦比亚特区华盛顿：世界银行。

Glewwe, Paul（1999年）。《发展中国家学校质量投入经济学》。伦敦：麦克米伦出版社。

_____（2002年）。“发展中国家的学校与技能：教育政策和社会经济成果”。《经济文献杂志》，第40期，第2号，第436-482页。

_____和Dennis de Tray（1990年）。“调整时期的贫穷人口：科特迪瓦的个例研究。”《宏观经济政策改革、贫穷和营养》，P. Pinstrup-Andersen等。纽约，伊萨卡：《科内尔粮食和营养政策方案专论，第3号》。

- Glewwe, Paul和Kwaku Twum-Baah (1991年)。《1987-1988年加纳的福利分配情况》。《生活水平衡量研究工作文件, 第75号》。哥伦比亚特区华盛顿: 世界银行。
- Grosh, Margaret (1991年)。《住户调查作为政策变化的工具: 牙买加生活条件调查的教训》。《生活水平衡量研究工作文件, 第80号》。哥伦比亚特区华盛顿: 世界银行。
- _____和Paul Glewwe (1998年)。“数据监测: 世界银行的生活水平衡量研究住户调查”。《经济展望杂志》, 第12期, 第1号, 第187-196页。
- _____ (2000年)。《设计发展中国家住户调查问卷: 15年生活水平发展研究的教训》。纽约: 牛津大学出版社(为世界银行)。
- Hentschel, Jesko和其他人 (1998年)。《合并人口普查与调查数据来研究贫穷的空间维度》。《政策研究工作文件, 第1928号》。哥伦比亚特区华盛顿: 世界银行。
- Jacoby, Hanan (2002年)。“是否有住户内‘粘蝇纸效应’? 来自学校供餐方案的证据。”《经济学杂志》, 第112期, 第476号(1月), 第196-221页。
- Kalton, Graham (2002年)。“抽样调查实践中的模型(修订)”。《官方统计杂志》, 第18期, 第129-154页。
- Lewbel, Arthur (1997年)。“消费者需求系统和住户等价尺度”。《应用及经济学手册, 第二期, 微观经济学》。H. Pesaran和P. Schmidt等。牛津: 联合王国: 布莱克威尔出版公司。
- Lipton, Michael和Martin Ravallion (1995年)。“贫穷与政策”。《经济学发展手册》, 第3期。J. Behrman和T. N. Srinivasan等。北荷兰。
- Pollak, Robert和Terence Wales (1992年)。《需求系统说明与评估》。联合王国, 牛津: 牛津大学出版社。
- Rao, J. N. K. (2002年)。《小区域评估》。纽约: 威利国际出版公司。
- Ravallion, Martin (1998)。《理论与实践中的贫穷线》。《生活水平衡量研究工作文件, 第133号》。哥伦比亚特区华盛顿: 世界银行。
- 世界银行 (1999年)。《越南: 消除贫穷》。《越南政府 - 捐助方 - 非政府组织贫穷工作组的联合报告》。汉诺。
- _____ (2001年)。《2000/2001年世界发展报告: 消除贫穷》。纽约: 牛津大学出版社。

第十八章

构建指数的多变量法

Savitri Abeyasekera

雷丁大学统计结算中心

大不列颠及北爱尔兰联合王国雷丁市

摘 要

按照调查的性质划分，调查的结果是形成多变量数据结构。虽然本章承认简单做法对调查数据分析具有价值，但还是举例说明了运用多变量技术进行深入分析对选定人口亚群的益处。现有软件包使调查研究人员能够运用这些较为先进的方法。

本章对在进一步分析数据以达到具体调查目标之前，多变量法在指数构建过程中和在利用调查数据特有子集进行数据探索的初期阶段所起作用的多种情况进行了举例说明。重点主要放在同时研究若干关键变量的方法上。关于这一点，多变量法允许对数据中可能存在的类型进行比较深入的调查，使许多变量之间复杂的相互关系能够用图表表示，并且为汇总和进一步分析提供减少数据维数的方法。在有关指数构建的讨论中，使用了较为宽泛的多变量法解释，以便将回归类方法包括在内。

本章的重点自始至终是对多变量法进行概述，从而使我们能够从非常务实的角度对指数构建价值进行评价。它所针对的目标既包括参与大规模住户调查的人员，也包括参与研发项目而很少有运用本章所述分析方法经验的调查研究人员。为了说明这些方法的使用法，本章列举了一些适当的例子，并且对如何解释分析结果进行讨论。

关键词：指数构建、多变量法、主成分、类集分析。

A. 引言

1. 在对调查数据进行分析时，大多数调查分析师一般都会使用直截了当的统计方法。最常用的做法是利用单向、双向或多向图表和利用柱状图、线形图等图表来显示。Wilson和Stern（2001年）和本出版物的第十五和十六章对这些方法进行了全面阐述，并且对数据分析过程期间需要注意的方面进行了充

分讨论。但在某些情况下，超出简单汇总的分析程序才是合乎需要的程序。本章对这样一类的程序进行了讨论。

2. 多变量法涉及到若干变量的同时处理（Krzanowski和Marriott，1994年a和1994年b；Sharma，1996年）。从严格的统计意义上来说，它们关系到对一组结果变量的集体研究，因此考虑到该组内各种变量的关联结构。但许多研究人员在运用多元回归法时也使用“多变量”术语，因为它涉及到与若干解释（预测）变量和主要结果变量（例如，Ruel，1999年）。另外，合在一起探讨几种变量的好处是因为它考虑到自变量间相关问题。第十九章对回归方法进行了比较充分的讨论，这些方法主要涉及主要响应变量的建模。在这里，我们主要把重点放在讨论对几种计量变量的共同研究，以作为我们在指数构建的讨论中对多变量法做出比较宽泛解释的一个初步步骤。

3. 多变量技术常被视为一种需要具备高级统计知识的“先进”技术。尽管许多多变量程序的理论及其应用甚至让统计人员产生畏难感，但它们确实对分析发展中国家的调查数据具有有益的作用。我们首先讨论对这些方法的有效使用：(a) 作为调查数据类型的解释工具；(b) 查明人口中的自然分组，以便作进一步分析；和(c) 减少所涉变量数的维数。我们将这些使用视为从住户级变量中构建指数的初步措施，例如，生成贫穷指数[例如，见Sahn和Stifel（2000年）]。

4. B节全面介绍了作为对一组结果变量进行集体研究的多变量法。接下来的四节谈到了多变量法的应用领域，举了许多例子进行说明。最后一节介绍了有关这些方法的价值和局限性方面的一些结论。本章将尽最大可能减少技术细节，重点强调对所涉概念的理解和解释。想要深入了解这些方法的读者应参考Everitt和Dunn（2001年）以及Chatfield和Collins（1980年）。

B. 关于使用多变量法的一些限制

5. 在本章，我们的讨论重点是将多变量做法作为有价值的描述程序用于数据探索初期阶段和指数构建过程的问题。但在这些方法的运用过程中，重要的是要在一开始就要强调，用于国家住户调查的完整数据集的分析不可能得出有用的结论，因为任何国家的住户都不可避免地具有多样性。如果将城市和农村人口以及不同农业生态区的人口合并起来分析，可能会损失有价值的信息，因为这些不同社会阶层中的住户的生活差别可能会很大。因此，只有对数据结构进行仔细研究，在查明可以适用这些方法的不同部门和分层人口之后，才能使用本章所描述的方法，并且还要牢记主要调查目标。

6. 即使是在这些分层内部，或者是在需要整体样本分析的情况下，注意与被抽中单位有关的样本权数将是一件重要事情。如果被分析数据在这些方面差别相当大，那么利用不含样本权数计算功能的软件包可能会导致做出错误的结论。在这种情况下，按照样本权数来权衡样本单位，例如使用SAS（2001年）中的WEIGHT（权数）语句或STATA（2003年）中的`aweight`命令，将

会解决这种与C、D、E和F节中所述方法有关的难题。许多软件包将会考虑到与G节所述方法有关的抽样权数问题。

C. 多变量法概述

7. 在调查工作中使用多变量法所暗含的基本主题就是简化，例如，将数量巨大而且可能复杂的数据减化成具有实际意义的少数汇总量度，或者查明数据中的关键特征和一切令人感兴趣的模式。目的常是探索性质的：这种方法可能在形成研究人员所关心的假设方面有用，但在检验这些假设方面没有用。在这些方法中，有许多方法是利用无分布法，不为任何变量假定某种基础统计分布。但是，由于在所使用的数据类型（例如，间隔大小、计数、二进制）方面需要一定注意，所以我们将会在本章相关段落谈到这一问题。

8. 入手点是一个数据矩阵，行代表各种情况（样本单位），列代表各种变量。有时候，行是我们比较关心的问题，例如，如果它们代表农户，那么可能关心的是：依据数据矩阵中某些列所代表的若干社会经济标准将这些住户分成不同的财富类型组。在其他情况下，列可能是主要关心的问题，例如，如果与某一特定主题相对应的变量集需要合并成用于进一步分析的某种形式的复杂指数。

9. 在下文各节中，我们将集中论述处理发展中国家调查多变量数据的四种主要做法。前三种可以视为最终实现指数构建的探索性技术。首先，我们来了解一下有助于理解数据的图示法和汇总量度。然后，我们再看看两种流行的多变量程序，类集分析和主成分分析（PCA），因为这是两种在指数构建过程中具有有效初步作用的关键程序。后一个程序将在G节进行比较充分的讨论，该节采用的方法是许多研究人员常用的广义“多变量”解释法，一起讨论的还有其他可以用来构建指数的方法。我们始终假定，需要进行分析的调查数据的适当子集已经选定，并且假定使用多变量程序处理这些数据的目的是将一种探索性措施与一种分析结合起来，以期实现某种广义的调查目标。

10. 当然还有可能在具体情况下考虑的许多其他多变量法。表十八.1介绍了多种此类方法，并且为每种方法附了简要说明。本章只限于讨论前三种方法，因为目的是强调数据探索，它是构建指数必需的第一步。这三种方法也可能最适用于调查数据分析。加之在我们关于指数构建的讨论中比较广泛地使用“多变量”术语，使它们成为调查数据分析中宝贵的补充工具。在与调查目标有关时，表十八.1中的其余方法在具体场合可能有用。但它们并不在本章讨论的范围之内，本章打算只对一些简单的方法作一般性介绍。

表十八.1
一些多变量技术及其目的

多变量技术	技术的目的
1. 描述性多变量法	数据探索；查明数据类型和关系。
2. 主成分分析	随着多变量集中变量的线性组合，形成新的变量（主成分），减少数据维数。
3. 类集分析	查明各种情况或变量之间的自然分组。
4. 因素分析	为多变量响应集内的各种变量建立相关结构模型，方法是使这些变量与一组公因子相关。
5. 多元方差分析（MANOVA）	将单变量方差分析扩展到对几个变量同时进行研究。目的是根据实验设计结构，划分变量集中间的总平方和及叉积矩阵。
6. 判别分析	确定一种能够让两个或多个个体组分开的函数。
7. 典型相关分析	研究两个组之间的关系。它涉及到在多变量集中形成若干对线性变量组合，从而使每一对线性组合依次产生两个组的两个个体之间的最大关联。
8. 多维比例	画一个表示若干对象之间空间关系的“图”，入手点是两个对象之间的距离表。

D. 图和汇总法

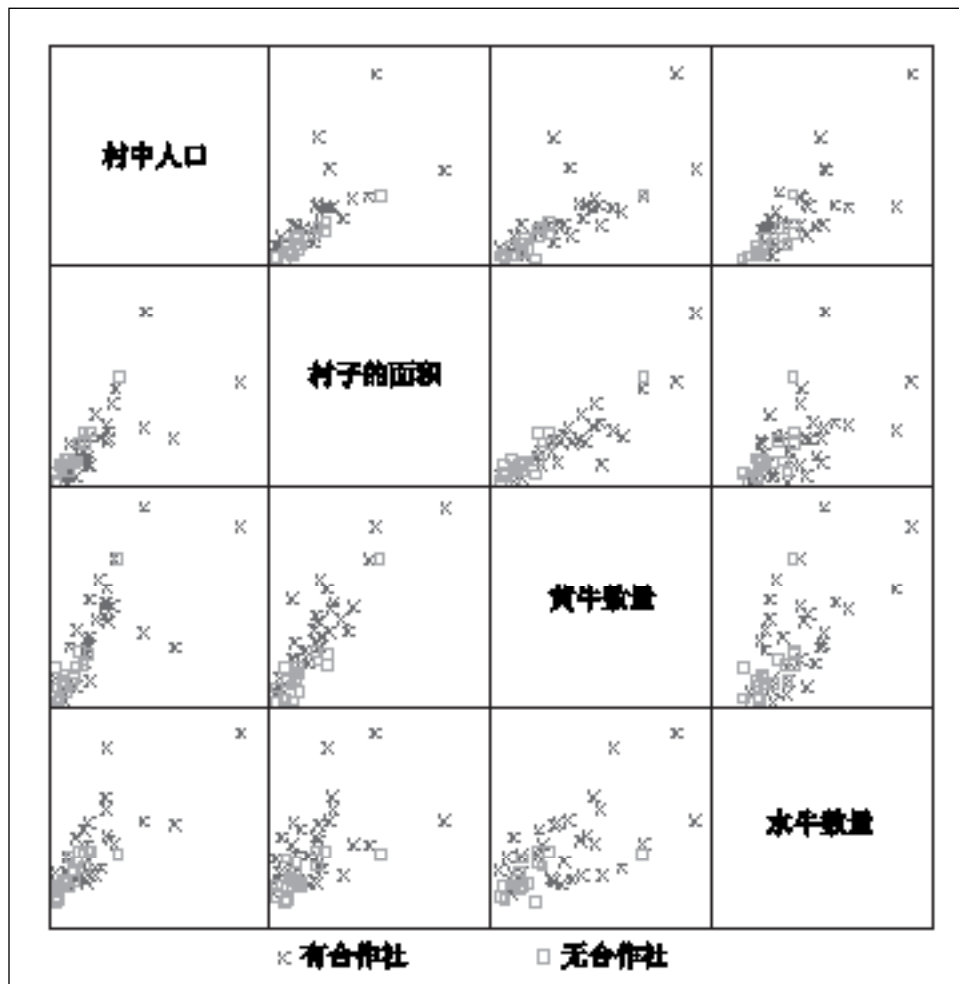
11. 如果要进行数据分析，对数据的初步了解是一个必不可少的开始阶段。仔细弄清数据将会给你一种关于数据含义和分布类型的感觉，查明可能存在的离群值（与其余数据类型不一致的观察结果），说明数据类型，并让用户了解某些变量变异是否大于其他变量[例如，见Tufte（1983年）以及Everitt和Dunn（2001年）]。

12. 就像在单变量分析集一样，计量数据的平均值和标准差以及二进制和分类数据的频率表等汇总法就是理想的分析方法。要想查明变量之间的联系，可以考虑变量对。在这一初步阶段，合理的做法是“成捆”考虑数据，可能分为两捆，一捆包括定量数据（连续或离散数据），另一捆包括质量数据（分类和二进制数据）。对于定量数据，散布图（成对）是一种有意义的做法，而对于质量数据而言，成对的双向表也是适用的，可能要与一些关系措施结合使用，并且可能用到卡方检验统计量。如有必要，也可以利用不同的符号来显示散布图，以说明由分类变量所确定的数据子集。

13. 大部分统计软件包都有矩阵图功能，例如SAS（2001年）中的PLOT程序、SPSS for Windows（SPSS，2001年）和GenStat for Windows（GenStat，2002年）中的Graph/Graphics菜单。这些都是图形显示，显示所有变量对之间能够在一起显示的散布图，因此，能够提供一种迅速判断，区分每个变量与考虑之中的多变量数据集中所有其他变量之间的关系。

14. 例如，图十八.1显示了一个用SPSS（2001年）生成的矩阵图。根据印度古吉拉特邦50个村中是否建有乳品合作社，该图说明了这些村在四种变量之间的关系。这四种变量是：村中人口数量、面积、黄牛数量和水牛数量，这四种变量只是较大一批变量中的少数几种。这些数据来自在引入一项方案之前所进行的基线调查，调查的目的是为了促进动物保健培训。每个图的横轴和纵轴都由与对角线方格平行的轴来确定。例如，在第一行中的三个图全都用村中人口作为它们的纵轴，而村的面积、黄牛数量和水牛数量则作为它们的横轴。同样的三个图也出现在第一列中，但它们的轴倒过来了，即用村中人口作为横轴，用村的面积、黄牛数量和水牛数量作为纵轴。在该数据集中可能有一个离群值，可以从第一行与人口数量很多的村子对应的方格中明显看出。从图中可以观察到所有变量对之间的某种关联。并且还可以看到，在上述所有变量中，有乳品合作社的村比没有乳品合作社的村更有可能出现较大的数值。

图十八.1
六种变量之间矩阵图示例



15. 如果矩阵图查明某些变量对显示令人感兴趣的类型或离群值，最好就像简单双向散布图那样重复这些数据类型或离群值，但要注意与每个数据点

有关的抽样权数问题。在气泡图中，每个点都由一个气泡来代表，而每个气泡则表示一个与样本权数成一定比例的面积（Korn和Graubard，1998年），这种图特别有用，能够提供比较有意义的解释。例如，抽样权数较大的离群值显然比抽样权数小的离群值有更大的影响。说明散布图中样本设计的其他方法还有很多，例如，对概率与样本权数成比例的数据进行二次抽样，然后在忽略样本权数的同时绘出图表，或利用核修匀法。读者欲知进一步详情，请参阅Korn和Graubard（1998年）。

16. 显示多变量数据的其他图示法还有很多。例如，Manly（1994年）表明，怎样才能以三种不同方式画出由若干变量描述的若干对象，以便说明各种变量值的特点。在Everitt和Dunn（2001年）中，有一章谈到许多图示问题，其中包括二变量盒状图、协同图和格状图表，而且Jongman、Ter Braak和Van-Tongeren（1995年）也介绍了双维图的使用范例。本章无法继续介绍更加详细的情况，但鼓励读者参阅上文引用的参考资料，以便作进一步了解。但重要的是要指出，在用于具体的人口分组时，这种图示法最有价值。

E. 类集分析

17. 类集分析（Everitt、Landau和Leese，2001年）是一种数据驱动法，主要目标是查明抽样单位之间的自然分组（例如，受访者、农场、住户），从而使每一个组（类集）内部的单位彼此相似，而不同组中的单位则不一样。也有变量类集相关的情形，例如，在只从每个组选择一两个变量以便在较少数变量的基础上进行进一步分析的时候。因此，它在进行数据探索和/或数据减化时是一个有用的工具。它还可能用于帮助生成假设和其他特殊情况。

例 1

18. 作为一个例子，考虑进行一项调查，目的是对某一地区资源贫乏的农户所采用的各种低成本虫害治理策略的效果进行调查。假定对那些可能参加今后农场试验的农民进行一项基准调查，其目的是：(a) 了解农业家庭的社会经济状况；(b) 了解农民当前治理虫害的做法；和(c) 了解农民关于虫害对其所种农作物影响的认知情况。在这里，我们将集中讨论这三个目标中的第一个目标，并且考虑怎样才能用类集分析来帮助确定有效选择参与农场试验主要研究的不同农民组别。

19. 在基线调查期间，对大量的社会经济变量进行了比较。目的在于根据这些变量将农户分成不同的层。例如，一种做法是选择两种主要变量，并且形成与这两种变量有关联的分类组合所定义的层。例如，如果所选择的变量是户主的性别（男/女）和住户的食品安全级别（低、中、高），则将会产生六个不同的层。

20. 这种方法的缺点是，它忽略了住户的其他社会经济特征。多变量分析法则使多种变量同时得到考虑成为可能。在所有相关的社会经济变量基础上，将类集分析用于农业住户是一种比较有效的方法，能够将住户分成若干个分

组，这样每个组就代表农业人口中与其他组不同的社会经济组。这是一种重要方法，因为与虫害治理策略有关的各种建议不一定适合所有农业住户。将农民分为不同类集的初步分类法有助于为选择不同类型的农民参加探讨各种虫害治理策略提供依据。它还有助于以不同类集的具体特征为中心，以便能够调查这些特征与所建议策略之间的相互作用。Orr和Jere（1999年）对此进行了举例说明。

21. 要想进行类集分析，必须做出两个决定。第一，必须确定类集单位之间的相似性（距离）度量法。相似性度量法就是利用源于若干变量的信息来提供一个数值，以反应每一对类集单位之间的“密切”程度。而距离度量法则相反，表示任何一对单位之间的距离相隔多远。如果所有变量都是定量变量，或者至多包括少数有规则的分类变量，除此之外，利用欧几里德³²距离矩阵可能也是比较恰当的做法。但是，调查数据经常包括二进制和无序分类变量。我们已经建议对这种数据采取各种相似性量度。例如，如果要在两个二进制变量之间产生一种相似性量度，则可以首先按照这两种变量将数据进行交叉列表，列出下文所示的 2×2 表。

³² 欧几里德距离可被看作与适用于多维空间一样只反应的“距离”的正常含义。

	0	1
0	<i>a</i>	<i>b</i>
1	<i>c</i>	<i>d</i>

22. 则一种可能的相似性量度就是被称为简单匹配系数的 $(a + d) / (a + b + c + d)$ 。另一种相似性量度是Jaccard系数 $d / (b + c + d)$ 。Krzanowski和Marriott（1994年b）还介绍了各种其他量度。如果涉及到混合数据类型，见Gower（1971年），了解适当的相似性量度。在实际情况下，如果在类集分析中使用了大量的不同类型的变量，最好进行若干个不同的类集分析，每一次考虑同一类型的变量，然后确定所出现的不同类集组是否相似。这种方法对类集成员资格进行交叉验证。

23. 一旦确定了距离量度或相似性量度，那么就必须做出关于类集分析方法的决定。统计软件同样提供了多种选择。例如，SPSS（2001年）提供了七种选择（例如，组联系、组内联系、最近邻居等方法）。在这些选择中，有些选择是凝聚式程序，其中，最初对 n 个单位进行类集分析，形成 n 个类集，每个类集一个成员，然后根据它们与其他类集中成员的相似性将它们进行连续合并。替代办法是一种分裂式过程，其中，所有 n 个单位起初都是一个单一类集，然后进行连续分类，直到找到满意的解决办法为止。另一方面，在就形成类集的方式问题做出正确决定过程中，应小心从事。Everitt、Landau和Leese（2001年）对这些问题进行了详尽的讨论。

例 2

24. 如果所有变量都是二进制变量，那么就会出现一种特殊情况。如果利用分级类集法，这一程序就可能变得比较简单。为了举例说明，我们只利用少数观测数据，这些数据来自对一个农场研究方案中涉及74位农民的一次小型调查。表十八.2中所示在农场调查期间记录的若干变量数据只是八位农民的变量数据。这些变量与是(+)和否(-)答案相对应。目的之一在于调查是否可以根据这些特征将农场分成少数类集。

表十八.2
表明有无各种农场特征的农场数据

特 征	农 场 (农 民)							
	1	2	3	4	5	6	7	8
住在高地(+)或低地(-)?	-	+	+	+	+	+	-	+
降水量是否大?	-	+	+	+	+	-	-	+
收入是否高?	-	+	+	-	-	+	-	-
家庭是否大(>10人)?	-	+	+	+	-	+	-	+
能否在2公里内弄到柴火?	+	-	-	+	+	-	+	+
10公里内有无卫生设施?	+	-	-	-	-	-	-	-
户主是否女性?	+	-	-	-	-	-	-	-
有无自来水?	-	-	-	-	-	-	+	-
农场中是否有厕所?	+	-	-	-	-	+	-	-
种不种玉米?	+	-	-	+	+	-	+	+
种不种木豆?	-	+	+	+	-	+	-	-
种不种豆类?	-	-	-	+	+	-	-	+
种不种花生?	-	-	-	-	-	-	-	+
种不种高粱?	+	-	-	-	-	-	-	-
有无牲畜?	+	+	-	-	+	-	+	-

25. 另外，为了举例说明和让构建细节简单化，考虑利用任何两种变量共同拥有的+数量，形成相似性矩阵。结果见表十八.3所示。然后，首先将八个农场视为组成部分的八个类集，再依次将最接近的类集合并，直到最后所有农场都在一个单一类集之中，这样就能形成一组类集。

26. 图十八.2以图表方式说明了上例中的相似性矩阵。这种图表叫作树枝形结构关系图。它说明怎样通过在任何一点利用水平线切割“树”，才能选择特定类集数。例如，如果将水平线放在树的顶部附近，那么就会产生三个类集，这些类集是由(1)、(7)和(2、3、4、5、6、8)三个组形成的。在大多数实际情况下，在确定分层分类形成的类集数时，需要做出主观判断。Everitt、Landau和Leese(2001年)描述了解决这一问题的正式方法。

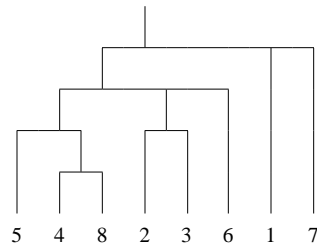
27. 利用适当软件，可以十分容易地进行类集分析，但在类集分析之前应该密切注意所使用的数据类型、相似性或距离量度，以及用来产生类集的方法。

法。如果所使用的软件只允许对一种类型的数据进行聚类，那么就需要特别小心。例如，SPSS（2001年）要求在类集中使用的所有变量都必须是连续数据、分类数据或二进制数据。如果数据属于混合类型，那么利用这种软件的比较好的做法可能是将所有变量都转化成二进制分数，并使用适合于二进制变量的相似性量度，但同时承认这种方法会导致损失一些信息。

表十八.3
八个农场之间的相似性矩阵

		农 场							
		1	2	3	4	5	6	7	8
农 场	1	-	1	0	2	3	1	3	2
	2		-	5	4	3	4	1	3
	3			-	4	2	4	0	3
	4				-	5	3	2	6
	5					-	1	3	5
	6						-	0	2
	7							-	2
	8								-

图十八.2
由农场间相似性矩阵形成的树枝形结构关系图



28. 还有另外两个问题要记住。第一个问题关系到需要明白（就作者所知），复杂样本设计对类集分析的影响尚不得而知。如果调查设计涉及到整群抽样程序，且被抽样类集之间存在相当大的差异，不注意抽样权数就将类集分析用于整个样本数据很可能会自行产生调查设计类集。因此，最好考虑对每一个调查设计类集进行类集分析，并且研究这些类集中交叉结果的一致性。另外，如果软件不能考虑权数，则应该区别调查类集内的抽样权数，而且在解释结果时应该谨慎。

29. 第二个问题涉及到由于计算内存的限制而可能导致多大的计算难度。如果利用充分调查样本进行类集分析，就会出现这种情况。如果分析与执行类集分析的目标一致，就可能仅限于较小组别被调查样本，以帮助缓解这一问题。

F. 主成分分析 (PCA)

30. 假定在一次调查中有几种变量从多个方面对一个主要问题进行衡量, 比方说12种变量。例如, 在营养调查中, 可以从几个人体测量值方面以及利用描述其家庭的社会经济特征的各种变量来衡量儿童的营养状况。这些变量可能存在相互关联, 那么就会出现的问题是, 这些变量能否以某种形式减化成数量更少的变量, 并且尽可能地反映原始数据集中变异。主成分分析的目的就是要做到这一点。这种分析技术严格适用于定量测量集或具有序数尺度的测量集。但是, 由于这种分析技术在很大程度上属于一种描述性技术, 将二进制变量和/或少量名义分类变量包括进去不可能是一种切合实际的结果。

31. 在主成分分析中, 新的变量集形成后, 是一个原始数据集的线性组合。³³ 说明最大变异量的线性组合被称为第一主成分。然后产生独立于第一主成分之外的另一主成分, 该成分尽可能地说明其余变异性。然后, 其他成分按顺序形成, 每一种新的成分都独立于前一种成分。如果前面一些成分, 比如说前3种成分, 说明12种变量的原始集合之中存在相当大数量的变异性, 比如说90%, 要分析的变量种数基本上从12种减少到3种。

32. 重要的是要指出, 当主成分分析进行非自行加权时, 如果将它适用于整个调查样本, 那么主成分估计量就可能存在严重偏差 (Skinner、Holmes和Smith, 1986年)。正如在B节所强调的那样, 一般建议只将主成分分析用于具有相同抽样权数 (至少接近) 的较小样本子集的调查数据分析。如果有关数据子集的抽样权数差异相当大, 那么在解释结果的时候应该小心从事。

³³ 如果 X_1, X_2, \dots, X_p 是 p 种变量的原始集合, 则由这些变量的线性组合形成的 Y 变量所采取的形式就是 $Y = a_1X_1 + a_2X_2 + \dots + a_pX_p$, 其中式中, a_i 的($i=1, 2, \dots, p$)是数字, 也就是说, 主成分系数。

表十八.4

主成分分析的结果

变 量	成 分		
	第一主成分	第二主成分	第三主成分
1. 家庭的整体福利	0.24	0.11	0.90
2. 渔业资源的整体福利	0.39	0.63	0.02
3. 当地收入	0.34	0.51	0.55
4. 渔业资源利用情况	-0.25	0.72	0.17
5. 资源控制	0.57	0.40	0.12
6. 参与社区事务的能力	0.77	0.13	0.29
7. 影响社区事务的能力	0.75	0.22	0.34
8. 社区冲突	0.78	0.03	0.18
9. 社区遵纪守法和资源管理	0.82	0.12	0.07
10. 传统上从水中获得资源的数量	0.38	0.66	0.12
说明的方差比例	33	19	14

例 3

33. Pomeroy等（1997年）将主成分分析法用于分析从200户调查所得的数据，在这次调查中，要求受访者对10种指数进行打分，分为1-15等，分别代表他们的等级地位，从而说明他们对由于其所在地区基于社区的沿海资源管理项目而发生的变化的认识。这些指数列于下表，而主成分分析结果则列入表十八.4。

因此，本表给出的第一主成分即为：

第一主成分 = 0.24（住户）+ 0.39（资源）…… + 0.82（遵纪守法）+ 0.38（收成）。

34. 在Pomeroy等人的著作（1997年）中，这种第一主成分被描述为关于社区成员行为的一个指数，而第二主成分被描述为与渔业资源有关的一个指数，第三主成分则被描述为家庭福利的一个指数。然后，他们将这些成分作为独立的变量用于多变量回归分析，用以调查若干说明因数在解释每一种指数的变异性方面的效果。

35. 尽管这里的变量解释是合理的，但有人可能会对为了进一步分析而采用上述计算方式利用（比如）第一主成分的价值提出质疑。只有变量5、6、7、8和9描述了社区成员的行为，并且这些变量在第一主成分中是得分较高的变量。与其在第一主成分的计算中包括全部10种变量，倒不如将原始数据集中行为变量的简单汇总重新计算一种新的变量，例如，取变量5、6、7、8和9的简单算术平均值，或这些变量的加权平均值，其中资源控制（变量5）同其他变量相比被赋予的权数稍低。同样，可以将资源变量（变量2、3、4和10）进行合并，得出一个简单汇总，而变量1仍然不变。利用这种方式，主成分分析将会查明怎样用一种简单的方式将这10种指数汇总起来，以便提供一种新的具有实际意义的量度集，以便作进一步分析，例如，Pomeroy等（1997年）通过回归分析探讨了对前三种主成分中每一种成分具有影响的因素。

例 4

36. 大不列颠及北爱尔兰联合王国政府国际开发部采用的可持续生计框架提供了另一种实例。这一框架考虑五种生计资产，即社会资本、人力资本、自然资本、有形资本和金融资本。若要进行一项研究家庭生计情况的调查，那就需要在许多辅助变量方面对所有这些资产进行调查。例如，可以从对支助网络的依赖程度、家庭汇款收入所占比例、在本组内的信誉度、参与决策的程度等方面对社会资本进行调查；可以从教育程度、健康状况等方面对人力资本进行调查；从拥有自行车或收音机、有自来水、电等方面对有形资本进行调查。

37. 在这里的目标是为五种生计资产确定单一变量，每种资产一种单一变量。对于有形资产而言，可以用直截了当的方式做到这一点，例如，通过将物品价格作为权数，得到与某个家庭是否拥有特定清单中的物品相对应的二进

制反应的简单加权平均值。另一方面，社会资本不能用这种简单的方式进行合并，因为向描述社会资产的变量分配权数是一件比较困难的事。在这里，我们可能不得不接受通过适用于社会变量集的主成分分析用数据导出的权数。结果可被用于产生一种适当的社会资本总体量度，在已知前一个或两个主成分中每个变量的相对权数之后，再求简单加权平均值。

G. 多变量指数构建法

38. 指数构建可能拥有几种不同的含义。例如，在健康调查中，儿童的营养状况一般利用从人体测量值所产生的指数进行测量，例如，年龄标准体重、年龄标准身高和身高标准体重，这些分别代表体重低于标准、发育缓慢和消瘦。

39. 在比较复杂的例子中，对母乳喂养、使用奶瓶、饮食多样化、儿童在过去七日内接受选定食物组的天数、喂养频率等项目的反应数据可以汇总起来，形成儿童喂养指数（Ruel和Menon，2002年）。这是另一类型的指数，其中，研究人员就将要分配的具体得分做出决定，保证每个变量的序数尺度始终保持高数量值，代表“好”与“不好”。如果涉及到二进制变量，例如在拥有资产的数量方面，资产的价格可被用于每一个项目赋予不同的权数，就像上文例4（F节）所说明的那样。

40. 如果调查涉及到确定态度或意见问题，比方说，对利用保健服务质量的态度或意见，则可能出现另一类型的指数。这里有几个问题可能会被问到，这就需要提供有关1-5分的记分量表的答案，1表示“很差”，5表示“很好”。另外，可以对所有相关问题的结果得分进行汇总，从而得出一个指数，以反映住户对保健服务质量的看法。

41. 在这里，我们的讨论将继续深入，包括由数据决定多变量程序所用指数的形式问题。这里仍然保留指数的共同解释，即作为一种单一价值来描述一种复合量度中若干变量所反映出来的信息，一般采用以下形式：

$$\text{指数} = a_1X_1 + a_2X_2 + a_3X_3 + \dots + a_pX_p$$

式中， a_i 表示将由数据确定的权数，而 X_i 则表示在调查中所测量的一个适当子集中的 p 种变量。我们对可以由数据来确定权数 a_i 的两种方式（见下文）进行了举例说明。哪一种方式更适合通常取决于指数构建所要达到的目标。

42. 第一种方式基于一种回归建模方法；而第二种方法则基于应用主成分分析。这两种方法在谈到用于衡量发展中国家住户财富或社会经济状况代用指标的指数问题时进行了讨论。关于这一专题的文献资料非常之多，并且在Davis（2002年）中也有全面的阐述。另见本出版物的第十七章，该章就住户调查数据用于了解贫穷状况的问题进行了有益的讨论。

1. 建立消费支出模型，构建收入代用指数

43. Henstchel等人(2000年)以及Elbers、Lanjouw和Lanjouw(2001年)介绍了一种将建立的消费支出模型作为收入代用指数的方法。它涉及到利用来源于详细住户预算调查的数据,查明代表贫穷的各种变量。这是通过把消费支出作为多线性回归模型中的因变量和一系列住户级变量(例如,住户拥有的资产、居住条件、公用设施利用情况等)作为模型中的潜在说明(预测)变量来实现的。说明反应(因)变量中最大变异的解释变量的最小子集被用于预测消费支出。如果已在普查中采集到解释变量,则可由此而得到的模型方程可用于普查数据,以预测每个普查住户的消费支出。然后,这些消费支出数据可以用于绘制全国范围内的贫穷状况分布图。如果在预计普查日期之前就进行了住户预算调查,则可以从预算调查数据中查明适当的预测变量集,并且列入普查调查问卷。为了举例说明这种方法,我们在下文直接举了一个例子。

例 5

44. 坦桑尼亚联合共和国国家统计局在2000-2001年进行了一次全国住户预算调查(HBS),调查大约涉及到22 000个住户。根据在为期28天内采集到的住户支出详细资料,将为每个住户计算每个相当于成人的人在28天内的全部消费支出。利用从住户预算调查所采集的初步数据进行的回归建模查明了一系列的潜在住户级变量(城市和农村地区单独设集),说明消费支出差异很大。这些变量被列入一个调查表,被总部设在达累斯萨拉姆的成人发病率和死亡率项目(AMMP)组用于它正在研究的三个哨点监测地的一项住户普查。目的在于利用每个成人发病率和死亡率项目哨点监测地的住户预算调查数据,确定一种反映消费支出情况的指数,并且将该指数用于成人发病率和死亡率项目在每个哨点监测地所覆盖的住户。

45. 关于建模方法的全部详情和对模型效果的评价,见Abeyasekera和Ward(2002年)。在这里,我们介绍一下某个农村地区的各种结果的汇总情况(见表十八.5),说明已被输入模型方程的变量以及用于计算消费支出指数的权数(回归系数)。

46. 从表十八.5的结果中,预测坦桑尼亚联合共和国乞力马扎罗地区住户消费支出的指数如下:

$$\begin{aligned} \text{消费支出指数} = & 9.79388 + (0.11043 * \text{灯}) + (0.19950 * \text{沙发}) + (0.12870 * \text{自行车}) + (0.11858 * \text{种子}) \\ & + (0.16254 * \text{化肥}) + (0.025824 * \text{土地面积}) + (0.088769 * \text{肉}) + (0.076132 * \text{收入4}) \\ & + (0.13451 * \text{收入3}) + (0.098303 * \text{收入2}) + (0.27985 * \text{教育4}) + (0.15878 * \text{教育3}) \\ & - (0.0091977 * \text{教育2}) - (0.0022552 * \text{年龄}) + (0.010456 * \text{住户人数2}) - (0.23902 * \text{住户人数}) \end{aligned}$$

47. 模型解释了消费支出中65%的变异率。考虑到农村住户谋生手段复杂多样，这是一个相当高的数字。在其发展阶段，判断这种指数质量的依据是：(a) 将它与消费支出真值进行比较；和(b) 考虑其查明坦桑尼亚联合共和国基本需求贫困线以下住户的真正比例的能力。方法(a)用于绘制与真值相对应的指数图表，显示出很好的对应关系。如果总体真值和总体预测值都被分到五种财富五分位数类别中，并且彼此对应列表，那么这种方法就不太好用了。只有46%的住户被分到正确的五分位数类别中。按贫困线分类比较好一些，87%的住户在贫困线以上或以下的分类正确。

48. 第十九章最后几节将介绍建模方法的其他例子。

表十八.5

坦桑尼亚联合共和国乞力马扎罗地区消费支出预测指数构建中所使用的变量及其相应的权数

预测变量	显著性概率	权数 (模型系数) (STATA估计)
住户人数	0.000	- 0.239
住户人数平方值	0.000	0.0104
户主年龄 (岁)	0.038	- 0.00226
户主学历 ^a	0.000	0, -0.00920, 0.159, 0.280
主要收入来源 ^b	0.017	0, .0983, 0.1345, 0.0761
过去一周吃肉的天数	0.000	0.0888
住户拥有的土地面积	0.000	0.0258
化肥 ^c	0.000	0.1625
种子 ^c	0.004	0.1186
拥有自行车	0.000	0.1287
拥有沙发	0.000	0.1995
拥有灯	0.001	0.1104
模型方程中的常数	0.000	9.794

样本规模 = 1 026

$R^2 = 0.651$

调整后的 $R^2 = 0.646$

^a 无；初等；中等；高等及以上。

^b 出售农作物；出售牲畜；商业/工资/薪酬；其他来源。

^c 如果在过去12个月内购买过。

2. 用于构建“财富”指数的主成分分析 (PCA)

49. 只有能够从前一次调查中获得可靠的消费支出数据 (因变量)，才能使用上文G.1节所讨论的方法。由于采集可靠的消费支出信息有一定难度，加上采集数据的成本很高，促使一些研究人员建议使用由主成分分析推导出来的、以资产为基础的贫困指数。在前一次研究结果表明资产与消费之间的关系非常密切之后，第一主成分就被用作一种社会经济地位指数 (Filmer和Pritchett, 1998年)。但是，在把资产指数视为一种贫困量度时必须小心，因为它的效果取决于对所使用资产的选择以及适用主成分分析的具体数据集。作为这种方法的一个例子，Gwatkin等 (2000年) 对主成分分析法进行了举例说明，用这一方法来确定坦桑尼亚联合共和国财富五分位数，并且利用了以下以混合资产为基础的变量以及与卫生有关的集合：

- 住户是否有电、收音机、电视机、冰箱、自行车、摩托车、汽车（每一种变量的代号编码为1=有、0=无）；
- 每个卧室住的人数（一种定量反应）；
- 住户饮用水的主要来源（七个类别）；
- 住户成员使用的盥洗室的主要类型（五种类型）；
- 住户地板材料的主要类型（六种类型）。

50. 它们所使用的数据来自通过人口与健康调查（DHS）调查问卷收集到的信息。本分析使用了适当的抽样权数。

51. 作者强调，他们的工作只是适用于全国样本的一种初步尝试，但未来通过社会经济等级调查人口差异的尝试将会产生不同的结果。他们建议，这种情况可能会由于利用一些除定义社会经济地位的资产以外的依据或由于抽样误差等而出现。一种更显而易见的理由是各地财富差异。实际上，在将他们的方用于坦桑尼亚联合共和国三个分组人口（即上文G.2节中所提到的三个地区），利用全国住户预算调查所采集到的数据（表十八.6）时，发现在财富五分位数的划分过程中存在着差异。因此，即使只是在单个国家的不同时间，或者在用于不同分层人口时，最好不要将主成分分析结果视为可移植的结果。

52. 研究人员还将主成分分析中的第一主成分作为汇总指数，以便对数据进行进一步分析。例如，为了在所进行的多变量回归分析中控制社会经济地位，确定影响儿童营养状况的因素，Ruel和Menon（2002年）从人口与健康调查数据集中构建了一个社会经济指数，以便将住户按百分位点进行分类。他们利用来自拉丁美洲五个国家的七种数据集，对城市和农村人口进行了单独分析。所使用的变量有水资源、卫生条件、住房材料（地板、墙壁、屋顶）以及拥有各种资产的情况。在进行主成分分析之前，这些变量的数值都是按照升序方式进行排列（从最差到最好）。只有主成分系数大于0.5的变量被保留到最终的指数中。这里的做法是合理的，主要目标一直是构建一种指数，以便校正在今后的分析中出现的社会经济差量。

表十八.6

将人口分成五种财富五分位数的划分点

财富五分位数	达累斯萨拉姆 ^a (住户预算调查)	乞力马扎罗 ^a (住户预算调查)	莫洛戈罗 ^a (住户预算调查)	全坦桑尼亚联 合共和国 ^a (住户预算调查)	全坦桑尼亚联 合共和国 ^b (住户预算调查)
20百分位数	-1.2993	-0.8452	-0.9190	-1.0317	-0.5854
40百分位数	-0.7709	-0.6289	-0.6180	-0.5704	-0.5043
60百分位数	-0.1054	-0.2459	-0.3645	-0.3051	-0.3329
80百分位数	1.1603	0.3239	0.4586	0.4609	0.3761

^a 2000-2001年住户预算调查。

^b 1996年人口与健康调查。

H. 结论

53. 我们在本章的目的一直是要论证利用多变量法构建指数，重点强调在此种分析过程中需要将多变量探索工作作为第一阶段。但是，这些方法的应用需要缜密的思考，适当注意它们的含义及其局限性。例如，主成分分析对减少变量的成功取决于能够仅仅通过少数成分指数就能对数据中相当大部分的变异进行汇总，并且能够为这些成分中的每一个成分提供有实际意义的解释。最好还要仔细考虑主成分分析程序的效果，在整个变量集中是否只有一小部分变异在第一主成分中得到考虑。还应该充分注意在与本分析目标有关的指数计算中所包含的各种变量的适当性。

54. 类集分析遇到的困难不仅与查明适当相似性或距离量度有关系，而且还与关于类集分析所用方法的决定有关系。在这里，各种因素都必须考虑，包括所用数据的类型、计算方法和程序对数据小变化的稳健性。

55. 还有必要再次强调的是，在具有可以划分人口的明晰结构时，本章所述的各种方法最适合用于适当的人口子集。如果需要分析的数据来自一项全国性调查，情况尤其如此。然后，必须做出有关选择所用子集的决定，但要有适当理由。所得到的一种结果就是不同的子集可能会产生不同的指数。但这本身将会是一种有用的发现，表明在所考虑的人口子集内部作进一步分析将会更有实际意义。

56. 本章对多变量技术的价值进行评估，不仅作为一种探索工具，而且（更为具体地讲）为了将它们用于指数构建。在通用统计软件[例如，SPSS（2001年）、STATA（2003年）]中现有的一些功能程序使这种分析可以以比较简单的方式进行。因此，建议研究人员考虑在调查数据分析期间利用这些软件，以便尽可能从数据中提取可能多的信息，有效地促进调查目标的实现。

鸣 谢

我要向我的同事Ian Wilson和两位匿名审阅人就本章草案初稿所提的宝贵意见表示衷心感谢。还要感谢坦桑尼亚联合共和国国家统计局允许我们将他们的数据用于本章中列举的一些例子，我还要感谢大不列颠及北爱尔兰联合王国政府国际开发部（DFID），因为它资助了许多有意思的、涉及发展中国家调查的项目，为本章的撰写提供了思路。但本章的材料仍然由作者承担一切责任，并不意味着国际开发部发表了任何意见。

参考资料

- Abeyasekera, S. 和P. Ward (2002年)。《预测哨点监测地每个成人支出的模式》。达累斯萨拉姆：“成人发病率与死亡率”，坦桑尼亚联合共和国卫生部。请登陆：www.ncl.ac.uk/ammp/tools_methods/socio.html。
- Chatfield, C. 和A. J. Collins (1980年)。《多元分析简介》。伦敦：查普曼·霍尔出版社。
- Davis, B. (2002年)。“是否有可能避免困境？关于选择贫穷研究方法的思考”。请登陆：http://www.povertymap.net/pub/Pov_mapping_methods_18-9-02.pdf。
- Elbers, C.、J. Lanjouw和P. Lanjouw (2001年)。“乡村和城镇的福利：关于贫穷与不平等的微观估计”。耶鲁大学和世界银行，Mimeo. Vrije Universiteit。
- Everitt, B. S. 和G. Dunn (2001年)。《应用多元数据分析》。伦敦：阿诺德出版社。
- Everitt, B. S.、S., Landau和M. Leese (2001年)。《聚类分析》。伦敦：阿诺德出版社。
- Filmer, D. 和L. Pritchett (1998年)。《在没有支出数据的情况下评估财富效应-或分割：应用于印度国的教育注册》。哥伦比亚特区华盛顿：《世界银行政策研究工作文件，第1994号》。
- GenStat (2002年)。《第6版GenStat软件》。联合王国，牛津：VSN国际公司。
- Gower, J. C. (1971年)。“相似处及其一些财产的折算率”。《生物测定学》，第27期，第857-872页。
- Gwatkin, D. R. 和其他人 (2000年)。《坦桑尼亚卫生、营养和人口方面的社会-经济差异》。哥伦比亚特区华盛顿：世界银行卫生、人口、营养和贫穷专题小组。请登陆：<http://www.worldbank.org/poverty/health/data/tanzania/tanzania.pdf> (2004年6月30日起可访问)。
- Hentschel, J. 和其他人 (2000年)。“合并人口普查和调查数据来探究贫穷的空间维度：厄瓜多尔的个例研究”。《世界银行经济评论》，第14期，第1号，第147-165页。
- Jongman, R. H. G.、C. J. F. Ter Braak和O. F. R. Van Tongeren (1995年)。《社区和景观生态学中的数据分析》。联合王国，剑桥：剑桥大学出版社。
- Korn, E. L. 和B. I. Graubard (1998年)。“调查数据的散点图”。《美国统计员》，第52期，第1号。
- Krzanowski, W. J. 和F. H. C. Marriott (1994年a)。《多元分析》，第一部分。《分布、排列与推论》。伦敦：阿诺德出版社。

- _____ (1994年b)。《多元分析》，第二部分。《分类、协方差结构和重复测量》。伦敦：阿诺德出版社。
- Manly, B. F. J. (1994年)。《多元统计方法：初级读本》。第二版。伦敦：查普曼·霍尔出版社。
- Orr, A. 和P. Jere (1999年)。“确定南部马拉维虫害综合防治中的小农目标群体”。《国际害虫防治杂志》，第45期，第3号，第179-187页。
- Pomeroy, R. S. 和其他人 (1997年)。“评估有助于基于社区的海岸资源管理获得成功的因素：中维萨亚地区项目-1，菲律宾”。《海洋和海岸管理》，第36期，第1-3号，第24页。
- Ruel, M. T. 和其他人 (1999年)。“良好的保健做法可以缓解贫穷以及母方接受学校教育少对儿童营养状况的不利影响：来自阿克拉的证据”。《粮食消费与营养司的讨论文件，第62号》，哥伦比亚特区华盛顿：国际粮食政策研究所。
- Ruel, M. T. 和P. Menon (2002年)。“利用人口和健康调查创建一个儿童供餐指数：拉丁美洲的事例”。《粮食消费与营养司的讨论文件，第130号》，哥伦比亚特区华盛顿：国际粮食政策研究所。
- Sahn, D. E. 和D. Stifel (2000年)。“资产作为发展中国家住户福利的衡量方法”。《第00-11号工作文件》。密苏里，圣路易斯：华盛顿大学，社会发展中心。
- SAS (2001年)。*《SAS8.2版》*。北卡罗来纳，卡里：SAS软件研究所，SAS出版社。
- Sharma, S. (1996年)。*《应用多元技术》*。纽约：约翰·威利国际出版公司。
- Skinner, C. J.、D. J. Holmes和T. M. F. Smith (1986年)。“样本设计对于主要成分分析的影响”。《美国统计学会志》，第81期，第395册，第789-798页。
- SPSS (2001年)。*《11.0版SPSS软件》*。伊利诺伊，芝加哥：LEAD技术公司。
- STATA (2003年)。*《8.0版Stata软件》*。得克萨斯，大学城：Stata公司。
- Tufte, E. R. (1983年)。*《数量信息的显示》*。康涅狄格，柴郡：制图出版社。
- Wilson, I. M. 和R. D. Stern (2001年)。*《分析调查数据的方法》*。《支持DFID自然资源项目的统计准则汇编》。联合王国，雷丁：雷丁大学，统计服务中心。请登陆：<http://www.reading.ac.uk/ssc> (2004年6月25日起可访问)。

第十九章

调查数据的统计分析

James R. Chromy

研究三角研究所

研究三角园

美利坚合众国北卡罗来纳州

Savitri Abeyasekera

雷丁大学统计服务中心

大不列颠及北爱尔兰联合王国雷丁

摘 要

在进行调查分析的时候需要考虑到这样一个事实：调查数据是利用复杂样本设计从选定的单位中获取的。这就需要使用权数来分析调查数据，并以能够体现复杂样本设计的方式来计算调查估计数方差。本章概述权数的开发及其在计算调查估计数方面的应用，并一般性地讨论调查数据的方差估计。首先讨论何谓“描述性”估计数，比如调查报告中广泛使用的“总数”、“平均数”和“比率”之类。然后讨论用于考查各种调查变量关系的调查数据“分析”方法的三种形式，即：多重线性回归模型、逻辑回归模型和多层次模型。这些模型形成一套变量工具，用以分析主要答复变量与诸多其他因子之间的关系。我们在本章用一些范例说明这些建模方法的应用，并提供解释所得结果的指南。

关键词：复杂调查设计、分析统计、回归、逻辑回归、等级式结构、多层次建模。

A. 引言

1. 住户调查利用复杂样本设计控制调查成本。通常没有列出所有个人和住户的完整抽样标架可供使用；而即便有这样的人口登记册可用，根据简单随机样本设计实施住户访问调查的成本也会高得令人望而却步。第二十三章讨论的“生活水平衡量研究”（LSMS）调查提供了一个很好的范例来说明住户调查设计的许多复杂特征。

2. 表十九.1列示住户调查设计的一种典型结构。如上文所述，住户调查采用的大多数样本设计都比较复杂，涉及到分层、多级抽样和不均等抽样率。在分析的时候需要用权数来补偿不均等抽样率，而对不答复的调整会导致更加不均等的加权。在估计调查估计数的精确度时应考虑复杂样本设计。

表十九.1

住户调查设计的典型结构

特 征	可能的定义	含 义
层次	地区	可以减少估计数的标准误差。
	社区类型（城市/农村）	控制样本分布可能导致非比例抽样。
初级抽样单位	查点区或类似地理区域	便于样本类集以控制成本。
	农村层次的村一级	便于开发仅在抽样地区内的完整居住单位住址框架。 使用与规模成比例的概率进行选择。
二级抽样单位	居住单位地址	可以包含零个、一个、或一个以上住户或无关的人。 用等概率在初级抽样单位中间选择。
第三阶段抽样单位 （如果不是所有家庭成员都自动纳入样本）	住户成员	从负责任的成年住户成员所提供的住户成员花名册中选择。可导致不均等加权以说明住户大小。
观测单位	住户 住户成员 住户成员经营的农业或工商企业 子群特别文档，比如成年劳动力 与住户成员有关的事件或一段经历 一段时期的反复度量（固定样本调查）	可能需要不止一个分析文档，以便进行专项分析。

3. 本章B节概述调查分析中所用权数的开发以及在编制广泛用于调查报告的简单“描述性”估计数（诸如总数、平均数和比率/百分比）时权数的应用。该节还概述了基于复杂样本设计的此类估计数的方差估计。

4. 其余各节重点讨论调查数据的三种“分析”应用形式，意在探讨一个关键答复变量或应变量——比如上学儿童的学习成绩、一个住户的贫困程度等——如何受到多种因素的影响；这些因素通常叫作解释性变量，或回归量变量。如果关键答复变量是一种定量计量变量，就适合采用多重线性回归模型；而如果关键答复变量是二进制变量，就是说，当答复只有两种可能的值的时候（比如是/否、有/无），则可以应用逻辑回归模型。这些回归方法可用于非成套的调查数据，或用于多级设计层次的某个单层抽样单位。或者，可能需要在分析中考虑不同等级层次发生变异的各种不同起因，然后开始进行多层次建模。这种方法顾及同一层次各抽样单位之间的相关结构，因为它们发生在不同层次的单位当中。

B. 描述统计学：权数与方差估计

5. 住户调查一般旨在产生人口总数、人口平均数、或者总数或平均数的简单比率。比如，总数可能是人口总数、劳动力中的男子人数、劳动力中的妇

女人数、或五岁及以下儿童人数。平均数的实例可能有劳动力中的个人平均收入、劳动力中妇女的平均收入、以及劳动力中男子的平均收入。为了估计总收入低于贫困线以下水平的住户的比例或主要靠女性挣钱的住户的平均家庭收入，可能需要比率估计数。

6. 住户调查可以得出全国估计数，但是也可以是为了产生地缘政治地区或其他跨部门领域的估计数。再者，可以通过重复进行住户调查获取周期估计数（比如年度或五年期估计数），可以把这种估计数看做“时间域”。只要所产生的估计数仅包含即便是针对人口域的（不论是跨部门的“域”还是时间上的“域”）“总数”、“平均数”或“比率”之类的估计数，我们都可以将产生这些估计数所需的分析定性为“描述性分析”。描述统计学不但包括估计数本身，而且也包括有关这些估计数精确度的某种度量。描述性报告可以包括各种估计数的标准误差，或者基于这些标准误差的区间估计数。在对标准误差的估计进行分析的时候，需要考虑到住户调查的样本设计问题。对于区间估计数来说，不仅需要基于设计的适当标准误差估计数，而且需要了解在计算标准误差估计数方面的自由度。在官方发表的描述住户调查成果的统计报告中，这些相当简单的描述统计类型的数据占了大多数。

7. 调查的权数³⁴和基于这些权数的统计估计提供了从住户的概率样本得出的观测值与住户人口的概括性度量或总体参数之间的联系。图十九.1说明这种联系。所有住户的人口有时叫作目标人口或称“全域”。若不应用概率抽样和加权，就没有支持样本观测值与目标总体参数之间的联系统计学理论。

图十九.1
权数的应用与统计估计数



8. 任何忽略样本设计和权数的分析都势必以各种假设条件为依据。如果样本设计的目的是产生一个等概率样本，那么用于估计平均数、比率或变量间的关系的权数即可安全无虞地予以忽略。Kish（1965年，第20-21页）把这些设计叫作“等概率抽选法”（*epsem*）设计，并且指出，即使复杂的多级样本也可以在设计的最后阶段或接近最后阶段把它设计成“等概率抽选法”的抽样单位。下文将提到，即便设计起初采用的是“等概率抽选法”，对不答复的调整可能会产生不均等的权数。如果通过权数的调整对数据加以后分层或多维校准的话，这些过程几乎总会导致不均等权数调整，从而产生不均等权数。

³⁴ 基于设计的权数一般是作为选定观测单位的选择概率倒数开发的。通常，关于住户调查分析文档的调查权数是已经对不答复做过调整的基于设计的权数。为了实现与已知的或精确得多的边际总数相吻合的后分层或校准，往往需要做更多的调整。此外，还可能对权数做某种形式的调整，以便在因为发生意外抽样事件或实地数据收集事件而导致权数大的时候限制不均等加权效应。使用“调查权数”这个术语是为了将其与严格的“基于设计的权数”加以区分。

9. 然而,有些分析者愿意假定可以不用权数或者使用均等权数来分析住户调查数据。当在数据中应用模型来研究一个应变量与诸多独立的解释变量之间的关系的时候,这些假定大多是站得住脚的。

10. 从理论上讲,在能够从所有样本成员得到完整答复的调查中,使用作为每个观测单位选择概率的倒数计算出来的基于设计的权数,可以提供无偏差的人口总数和其他线性统计估计数(Horvitz和Thompson,1952年)。可是在实际当中,住户调查总会遇到一些不答复的情况;如果不采取任何其他措施的情况下从分析得出观测值,这些不答复的情况就会导致出现估计偏差(见第八章)。现已开发出一些方法,试图减少不答复造成的偏差。最简单的方法就是把样本划分成不同的加权类别。据认为,在这些类别范围内,答复者和不答复者总体参数之间的差异就会小得多,乃至达到可以忽略的地步(Rubin,1987年)。然后在加权类别中按权数对比率加以调整。这样,每一种类别在经过调整的估计数中都有代表;其比率跟选定样本中的代表比率相同。

11. 概率抽样的过程未必能保证所选定的样本在已知特征方面的分布将与总人口的那些特征分布是一致的。在选择样本前进行分层可以确保某些特征的这种分布状况,但如果没有关于样本选择所用的标架的分类变量可用的话,其他变量也许就不可能有这种分布状况了。并不是对住户调查数据所产生的每一个估计数都进行复杂比率调整,而往往是把后分层作为一次性权数调整;之后,这一调整便自动适用于使用经调整的权数所产生的所有估计数。后分层调整的最简单方法就是使用类似于加权类别不答复调整的那种划分样本的方法。

12. 在几个方面同步控制加权样本分布的校准方法,有时候用于不答复权数调整和后分层,或两者都用这些校准方法(Deville和Särndal,1992年;Folsom和Singh,2000年)。

13. 如果权数极大,就有可能通过设计效果导致住户调查估计数方差增大(见第六章和第七章)。有时候可以任意缩减或调整这些权数,尤其是在此种大权数并非有计划的样本设计所致的情况下。

14. 根据住户调查制作的分析文档所附的最终权数可以包含如下诸因素:

- 作为总体选择概率的倒数计算的基于设计的权数;
- 不答复调整因素;
- 后分层调整因素;
- 权数调整因素。

15. 这些因素应做成文件,以便任何分析工作者都能审查它们。在基本初步设计的权数中应用的这些调整因素,在加权类别的定义、在用于后分层调整的控制总数的选择以及在用于控制设计效果的权数调整的范围方面,会涉及到一些主观的、有时是武断的判断。当调查估计数中出现意外结果或明显异常情况

况的时候，对加权过程以及总的调查设计与实施的其他方面进行彻底检查是常有的事。

16. 总的来说，由于复杂调查设计中包括使用权数和某种设计结构，所以在使用住户调查数据进行分析的时候就遇到一些特殊的挑战。本手册的好几个章节都讨论了复杂调查设计所产生的设计效果问题。特别是第二十章探讨了复杂调查设计对分析结果的影响。欲了解有关复杂设计分析的更透彻讨论和某些专题的详情，读者不妨查阅Skinner、Holt和Smith（1989年）；Korn和Graubard（1999年）；以及Chambers和Skinner（2003年）的文献。本出版物第二十章还围绕复杂调查分析提供了较为技术性的讨论；第二十一章讨论了软件问题，并提供了有关使用真实数据范例分析调查数据的一些方法实例。

17. 非线性统计。在复杂调查中，即便像平均数这样的简单统计量也会变为非线性值。为了从一次复杂调查中估计某项总体均值，有必要估计有关变量（比如家庭收入）的总体总量，并估计总体的规模（比如家庭总数）。于是便可根据这两个估计数的比率估计出平均数。平均家庭收入可以估计为：

$$\text{平均家庭收入估计数} = \frac{\text{家庭总收入的估计数}}{\text{家庭总数的估计数}}$$

这个估计的平均数证明是两个线性统计量的非线性函数（比率）。在复杂调查中，样本规模（某一特定类型观测值的数量）本身是一个随机变量。对于小样本来说，这类非线性估计值并不是没有偏差，但从微不足道的意义上来说，它们是一致的，也就是说，如果将样本规模扩展到有限总体规模的话，这种非线性估计就会恰好等于可比的有限总体值（Cochran，1977年，第21、153和190页）。如果我们能够把有限的总体看作是产生于一个假设性的无限总体的话，那么我们就可以考虑让这个样本规模无限地扩大。在这种情况下，当这个非线性估计数随着样本规模的扩大在概率方面收敛为超总体参数的时候，我们就可以断言模型具有一致性（譬如，见Skinner、Holt和Smith，1989年，第17-18页）。

18. 非线性统计的标准误差，只能使用一级泰勒级数近似值来表述。可以通过使用一级泰勒级数近似值或者采用譬如均衡重复复制法或刀切复制法之类的复制方法来获得非线性统计量的标准误差估计数。

19. 当首先在全人口层级估计所需要的应变量及自变量线性函数的时候，同样类型的论据就转而用于采用“线性”模型的分析领域。

20. 总之，使用权数可以产生无偏差的线性估计数和一致的非线性估计数。据认为，实际当中使用一致的估计数来控制估计偏差是令人满意的。其他类型的偏差和非抽样误差，比如因为不答复、采访员失误或回答有误等造成的误差，通常都具有更多实际意义，尤其是在样本规模变大的情况下。

21. 住户调查中的样本设计结构。一般来讲，人口和样本设计都会有某种结构。正如上节所讨论以及在表十九.1中所说明的那样，在住户调查样本设

计中，一般在抽样标架中采用一种嵌套结构。虽然这种结构对构建诸如总数、平均数或模型函数等一级统计估计数没有什么影响，但它确实对二级统计量（访查估计数）有影响，这样，分析人员就可以对一级统计量的误差做出估计，并可对特定假设进行具有统计显著性的检验。

22. 根据分层的多极样本充分表述的估计数方差在每个样本设计阶段都有其组成部分。譬如，如果仅在第一阶段采用分层法，那么，根据包括区段、住户和住户成员的三段设计得出的总体总量 T 的估计数就可能有如下形式的方差：

$$\text{Var}(\hat{T}) = \sum_h \left(fpc_{h1} \frac{S_{h1}^2}{n_{h1}} + fpc_{h2} \frac{S_{h2}^2}{n_{h2}} + fpc_{h3} \frac{S_{h3}^2}{n_{h3}} \right)$$

式中， h 层的各项分别界定如下： fpc_{hi} 项分别是在区段选择阶段（ $i = 1$ ）、居住单位选择阶段（ $i = 2$ ）和个人选择阶段（ $i = 3$ ）的有限总体校正系数。 S_{hi}^2 项分别是以这三个抽样阶段的加权数据为基础的方差成分。 n_{hi} 项分别是 h 层内的区段样本规模（ $i = 1$ ）、住户（ $i = 2$ ）和个人（ $i = 3$ ）。在实际当中，这些方差成分有的难以估计或不可估计并非罕见；由于规模1的子样本或其他缘故，发生这种情况是可能的。Cochran（1977年，第279页）注意到，如果第一阶段的有限总体校正系数（设定为1）可以忽略，那么方差估计数即可依据该公式的一个简单得多的模拟，这种模拟仅涉及初级抽样。关于第一阶段有限总体因子1的假设，往往被描述为近似“不带置换”的样本设计方差的“带置换”的样本设计方差估计数。

23. 如果观测值是个人，为了在前面讨论的三段设计的人口总数估计数方面应用这一假设，我们可以设定一个新变量：

$$Z_{hi} = n_{h1} \sum_j \sum_k w_{hijk} Y_{hijk}$$

式中， w_{hijk} 和 Y_{hijk} 分别是 h 层内 i 区段的住户 j 中个人 k 的权数和观测值。于是便可获得这样一个合理的方差估计数：

$$\text{var}(\hat{T}) = \sum_h \frac{\sum_i (Z_{hi} - \bar{Z}_h)^2}{n_{h1}(n_{h1} - 1)}$$

该公式之所以能够成立，是因为总体总量估计数可以列式如下：

$$\hat{T} = \sum_h \bar{Z}_h$$

³⁵ Woodruff（1971年）说明了如何开发线性化的变量以利于计算复杂泰勒级数方差近似值。

若 Z_{hi} 选择适当，则使用一级泰勒级数近似值³⁵即可估计线性及非线性统计方差。这种方法亦可扩展到回归或逻辑回归中的参数估计数。请注意，采用这种方法不需要能够估计后续阶段方差的差益。

24. 如果第一阶段有限总体校正系数明显小于1，这个公式就会过高估计方差并导致过高估计调查估计数的标准误差。一个不大的过高估计就会在一定程度上导致很大的置信区间差异，或者可能导致在进行假设检验方面减少统计

显著性的说明。在这个意义上，假设第一阶段有限总体校正系数为1在统计学上据说是偏于保守的，因为这样会有助于防止虚假的统计显著性说明。应当指出的是，通过假定第一抽样阶段的有限总体校正系数为1，可以简化基于泰勒级数的软件和基于重复的软件（见第二十一章）。

C. 分析统计

25. 在本节，我们从简单描述性估计数方面的考虑转到所谓“分析统计”，即旨在研究各种变量间关系的统计。实际上，从数据用户希望在各种域之间进行估计数比较的那一刻开始，所需要的统计在性质上就变成了“分析统计”。简单分析统计可以基于各种域之间的差异，比如通过比较两个地缘政治子区域总收入低于贫困线的家庭比率或比较过去两年的农作物产量情况所产生的差异。有时候，一个简单比较中的各项估计数是相互独立的，因而差异的标准误差可以根据单个估计数的标准误差严格确定。在这些情况下，可以从两个域平均数之间的估计差异推导出标准误差：

$$se(\bar{y}_1 - \bar{y}_2) = \sqrt{\{se(\bar{y}_1)\}^2 + \{se(\bar{y}_2)\}^2}$$

关于差异的上述标准误差公式假定两个估计数是相互独立的，因此，其估计数不是相互关联的。这种形式的差异标准误差对数据用户比较方便，因为据此可以从公布的单个估计数的标准误差推导出某项差异的标准误差。然而对复杂样本设计来说，各个域的估计数之间往往是相互关联的。这样，两个域的估计数差异的方差便包含一个协方差项：

$$se(\bar{y}_1 - \bar{y}_2) = \sqrt{\{se(\bar{y}_1)\}^2 + \{se(\bar{y}_2)\}^2 - 2\text{cov}(\bar{y}_1, \bar{y}_2)}$$

26. 这个协方差项一般是正项，因此，与上面讨论的独立估计数的情况相比，会导致差异估计数的标准误差较低。在设计住户调查的时候，可以利用差异估计数标准误差中的协方差项；通过利用一个共同而连续的个人或住户样本，纵向固定样本调查可以在年度估计数之间实现高度正协方差。因为不能从公布的个别估计数标准误差推导出差异的标准误差，所以需要预计哪些比较值最值得关注，并公布这些比较值的标准误差。

27. 在关于有限总体的严格描述统计方面，通过使用一个有限总体校正系数，可以适当减少描述性估计数的标准误差。在简单随机抽样的最简单情况下，有限总体校正系数是：

$$fpc = 1 - \frac{n}{N}$$

式中， n 是样本规模； N 是总体规模。如果分析的目的具有分析性，那么，即便在两个域平均数之间的观测差异的统计显著性这种最简单的情况下，也不宜使用有限总体校正系数（Cochran, 1977年，第34-35页）。这是因为，统计显著性检验形式需要假设这两个域总体是否产生于一个共同的无限假设总体（一个单一的超总体）。³⁶ 将有限总体校正系数用于结构复杂设计这一问题将在下文讨论。

³⁶ Cochran（1977年，第39页）指出，有限总体校正系数不适合在统计工作中用来检验域平均数的差异。当把这条指导原则运用于既包含分层又包含类集的复杂设计的时候，它的解释就变得更为含糊不清了。Chromy（1998年）讨论的问题是以较高的比率对学校进行分层和抽样时，在学校对学生进行抽样这一问题。Graubard和Korn（2002年）就此问题提供了最新评述。

D. 关于回归建模的一般性评论

28. 本章其余各节讨论的方法讲到一种建模技术，即为一个关键答复变量或应变量中的变量建模，并确定一套潜在解释性变量中的哪一个变量子集对促成变异起到最重要的作用。可以通过应用适当的变量选择程序，或者通过采用某种敏感的排序程序来选定“最佳”子集，以探讨若干种不同的模型；同时从实际出发，密切注意在分析过程的每一步录入或消除变量是否恰当。

29. 我们想要强调的是，应当把本章讨论的各种方法看作对简单分析方法的补充，而不是取代它们。使用简单描述性概要（平均数、标准偏差等）、图解法（散点图、条形图、框图等）和相关的列表对数据进行初步探索是极有价值的，应该成为数据分析的第一阶段。有时候，也许只要有这第一阶段的分析即可达到目的。然而，调查目标可能还需要对数据作进一步分析，在后面这种情况下，建模方法可能就变得重要了。

30. 这里所讨论的建模方法在需要采取整体做法的情况下具有特别重要的意义，譬如，把分析的目标定位于认识现行农业制度的合理性，或者定位于住户如何管理其有限的资源以满足生产与消费两方面的需要。本章自始至终强调适当建模方法的实际运用，同时重视发展中国家在实地调查中可能遇到的各种困难。特别强调了分析工作的各种制约因素，以确保只有在仔细考虑所采用的方法是否适合拟进行的研究之后，才将本文所讨论的方法付诸应用。

31. 回归模型的应用，目的在于更好地认识某个应变量与一套自变量或解释性变量之间的关系。可是一定要当心，往往不可能根据观察到的应变量与解释性变量之间的联系确定其因果关系，但在设计周密、有控制和随机化的实验情形下例外。³⁷ 记住这一点，就可以通过在住户调查所获观测数据的分析中运用回归方法学到不少东西。

³⁷ 调查中也可以搞随机化的实验。这些实验往往是在先期检验样本中或在某项现行调查的补充样本中进行的方法论实验。也可以利用住户调查样本补充社会实验课题，来进行一些社会实验。

32. 住户调查数据不同于利用随机化和辅助变量控制进行受控实验所产生的数据，它通常是在很少或根本不控制可能影响变量之间关系的其他因素的情况下取得的观测结果。有的时候，回归方法可以消除这些未加控制的混杂变量所产生的影响，从而获得偏差较少的真实关系估计数。

33. 回归建模往往带有探索性。可以开发多种模型来解释有关应变量的变化。模型中所使用的解释性变量限于调查数据文档中的现有变量；因此，选定用以解释某个应变量差异的变量只能是与实际产生结果的因素明显相互关联的变量。可能有多种作为成因相互关联的竞争变量，但是逻辑上似乎其中没有一个与该应变量有关联。应该用实质性理论（比如社会理论或经济理论）来指导住户调查的分析者选择解释性变量和确定关系（比如线性关系和非线性关系）的形式。

34. 如果不能根据实质性理论提出确凿的理论关系，或者，如果根据实质性理论提出了好几种相互竞争的解释性变量，那就可以应用标准（非调查）软件包中提供的变量选择方法来确定潜在的解释性变量。许多非调查软件包中都有前向及后向变量选择方法，可以帮助确定与应变量有线性关系的解释性变

量。如果非调查软件包允许的话，我们甚至极力建议使用调查权数来进行这种探索性分析。可以通过调查权数标准化来概括整个样本，以便提供较好的误差估计数和更接近正确的统计显著性检验（关于这种方法的实例，见第二十一章）。在使用非调查统计软件包或程序进行变量选择之后，一个好的做法是使用软件包对利用调查权数并认可住户调查设计的模型进行评价。

35. 模型变量可以是分类变量、计数变量或连续测量数据。当应变量是计数或连续测量数据的时候就使用线性模型；建议在计数数据方面使用对数变换。如果相关计数变量包含零值，对数变换就用不上，但是像PROC LOGLINK（SUDAAN 2001）之类的程序可以用来拟合一个计数变量的预期对数值。当应变量是在两个层次确定的分类变量的时候，使用逻辑回归；多项式回归模型也可以用于两个层次以上的分类应变量。为了讨论的目的，我们把解释性变量划分为分类变量和连续变量，因为在模拟背景下基本上用同样的方式处理计数变量和连续（测量）变量。也可以用本章没有讨论的生存模型和其他多元方法来分析调查数据。

36. 使用界定研究范畴的分类解释性变量类似于不用模型构建简单域的比较。模型的应用可以使分析者能够同步调整其他可能的解释性变量。这一般叫作协变量调整。在没有协变量调整的情况下，回归模型系数就再生成简单域比较并估计存在于总量中的域差异。如果模型中包含其他变量作为协变量，就用回归模型系数来估计假设在所有域的同层都有协变量存在的域差异。

37. 通过调整或不用调整其他协变量，也可以获得适用于连续性解释变量的回归模型系数。应根据分析的目的决定调整或不调整协变量。未经调整的协变量用于描述存在于总体中的应变量与解释性变量之间的经验关系。如果假设其他变量保持恒定的话，调整的估计数也用于描述上述关系。如果模型中包含的其他变量也是良好的应变量预测工具的话，这些变量就能提高正在研究中的主要预测工具在特定层次预测值的精确度。分析方法的选择取决于分析的目的。

38. 下面的实例中只讨论连续解释性变量的简单模型。如果解释性变量是连续的，分析者就应调查应变量与潜在解释性变量之间的关系。简单的标图表明，为了使变量恰当相关，光有线性关系是不够的。可以视观测标图的需要，补充一些项（二次项或三次项），以便更好地捕捉这些关系。这样，应变量即可与一个解释性变量、与它的二次项、三次项或更高的项有线性关系。在包含了某些潜在解释性变量之后，就可以使用剩余的标图来确定是否其他变量或纳入的变量的更高的项（二次项或三次项）可能影响模型的拟合。对于有着广泛值和对这个范围的应变量有不同影响的解释性变量来说，可以让这种关系在该范围的各子集发生变化的样条模型往往很有用。当调查样本包括青年、中年和老年人的时候，往往可以使用回归样条模型来展示年龄效应。

39. 其他诊断方法包括检查拟议模型的拟合效果好坏和检查补充变量回归参数的统计显著性。根据标准（非调查）程序编写的程序可以改写成适用于加权调查数据的程序。经过解释的变异的概念可以结合加权调查数据和线性回

归一起使用。可以用列联表方法来评价逻辑回归模型的拟合。Korn和Graubard（1999年，第3章）在将改写的诊断程序应用于一般调查数据分析方面进行了有益讨论。

40. 根据观测数据开发回归模型显然涉及到解释性数据分析的概念（Tukey, 1977年）。此类分析可以使我们对数据和观测到的变量间的关系有深刻洞察力，但是这种“无计划的”分析结论的统计显著性尚有待日后确认或由其他调查数据来验证。

E. 线性回归模型

41. 为了讨论线性回归模型（本节）和逻辑回归模型（见F节）的目的，方便的做法是假设第一阶段的抽样“带重置”。我们进一步假定观测数据的分析文档载有层次下标变量，用 h 来表示；初级抽样单位用 i 来表示。如果我们打算在样本选择的第一阶段使用上面B节所讨论的带重置的设计假设，就不需要确定更多的结构变量了。只要我们使用一个统计软件包，其中说明了包括分析权数和设计结构（层次和初级抽样单位）在内的住户调查设计，模型系数及其标准误差就包含了复杂住户样本设计的全部含义。当我们讨论多层次模型的时候，重点就转为将设计结构纳入模型，以及通过分析就可以获得与结构变量有关的效应估计数。

42. 包含一个连续解释性变量和一个分类解释性变量的线性回归模型可以表示如下：

模 型 1

$$y_{hij} = \alpha x_0 + \beta_1 x_{1hij} + \sum_{d=1}^D \gamma_d x_{2dhij} + \varepsilon_{hij}$$

43. 在模型1中，观测到的应变变量 y_{hij} 代表观测值；截距变量 x_0 始终定为1；观测的连续解释性变量用 x_{1hij} 来表示；一套指标变量 x_{2dhij} 界定了分类变量的 D 个层次。回归模型系数 α, β_1 和 γ_d （ $d=1, 2, \dots, D$ ）叫作回归系数，拟通过分析来估计。该模型的最后一项是误差项，它计量与第 h 层、第 i 个初级抽样单位、第 j 项观测值有关的模型所产生的偏差。这是一个主效应模型，因为它不包含任何交互作用效应。

44. 视软件应用的具体情况，可在模型语句中将这套指标变量规定为单一变量；可能有必要用程序语句或指令规定该变量为分类变量并确定层数。然后程序给指标变量规定一个向量。如果观测值 hij 属于 d 类，那么一个指标变量——比如 x_{2dhij} ——就定为“1”；否则就定为“0”。为了避免解释性变量之间的线性依存关系，分析程序就将分类变量的指标重新参数化。一般做法是去掉最后一类分类变量，这样，该类别就变成了参照类。³⁸表十九.2列示了一些可以为模型1估计的效应，其中，应变变量是住户的工薪收入；连续解释性变量是住户中的工薪者人数；分类变量划定了全国四个地区“域”（北部、南部、东部和西部）。

³⁸ 还有可能通过添加一个线性约束，比如要求效果值的总和等于“0”或加权效果值等于“0”，来估计分类变量的系数。

表十九.2

模型1的线性回归参数估计数的解释——设应变量为住户的工薪收入

效应 (通常为程序输出中所规定的效应)	系 数	估 计 数	解 释
截距	$x_0=1$	α	参照组或“0”层级的有薪住房收入：西区工薪者人数为0
住户中工薪者人数	x_{1hij}	β_1	每增加一位工薪者住户收入的变化（按地区调整）
地区			住户工薪收入的差别（按住户中的工薪者调整）
北部(d = 1)	$x_{21hij} - x_{24hij}$	$\beta_2 = \gamma_1 - \gamma_4$	北部对西部
南部(d = 2)	$x_{22hij} - x_{24hij}$	$\beta_3 = \gamma_2 - \gamma_4$	南部对西部
东部(d = 3)	$x_{23hij} - x_{24hij}$	$\beta_4 = \gamma_3 - \gamma_4$	东部对西部
西部(参照域d=4)	$x_{24hij} - x_{24hij} = 0$	$\gamma_4 - \gamma_4 = 0$	没有估计数

45. 估计的域变量回归系数是按照该域和参照域之间的差别确定的。关于北部域估计系数的统计显著性检验实际上是检验北部和西部是否是出自同一个超总体的随机样本。如果北部地区的系数与0差别很大（基于用0.05的显著性水平进行的假设检验），那么分析员即可断定，在调整了住户工薪者人数之后，北部和西部地区的住户工薪数据出自同一超总体样本的可能性很小（5%或更低）。调查程序软件使用户能够要么通过类别排序（以便使期望的参照类排在最后），要么通过明确指明来规定不同的参照集。这可能是获取有价值的回归参数估计数的一个有用工具。通过估计系数的函数还可以对其他比较做出估计。

46. 表十九.3根据表十九.2列示的参数估计数列出了模型1的一些可以估计的函数。表十九.3按地区和住户中的工薪人数列示了住户工薪收入估计数。可以很容易地将此表扩展到每户有三位或三位以上工薪者。

表十九.3

可估计的住户工薪收入（模型1）

地 区	每户有:	
	一位工薪者	两位工薪者
北部	$\alpha + \beta_1 + \beta_2$	$\alpha + 2\beta_1 + \beta_2$
南部	$\alpha + \beta_1 + \beta_3$	$\alpha + 2\beta_1 + \beta_3$
东部	$\alpha + \beta_1 + \beta_4$	$\alpha + 2\beta_1 + \beta_4$
西部	$\alpha + \beta_1$	$\alpha + 2\beta_1$

47. 让我们来考察分析员在使用模型1研究住户工薪收入的时候所必须做出的几种假设。或许最重要的假设就是住户的收入几乎都与每户的工薪者人数相关联。线性假设是：工薪者从0增至1人、从1人增至2人、从2人增至3人……

等等，住户工薪收入都是按照同样的增长额变化。这种假设看起来令人生疑。由于分类变量要求尽量少做有关解释性变量与应变量之间关系形式的假设，所以分析者就可能决定把工薪者人数转换成分类变量，于是就采用只有分类变量的模型。³⁹ 模型1的一个变体可以写为：

³⁹ 关于评价线性回归模型拟合优度以及适用于非线性关系的其他备选模型的方法，读者若想了解更多的讨论，可参考Korn和Graubard（1999年，第95-100页）。

模 型 2

$$y_{hij} = \alpha x_0 + \sum_{d=1}^{D_1} \gamma_{1d} x_{1dhij} + \sum_{d=1}^{D_2} \gamma_{2d} x_{2dhij} + \varepsilon_{hij}$$

48. 对于模型2来说，分析员可以根据住户中的工薪者分布情况把工薪类别划分成少到2人，或者比这大得多的人数。为了限制所要估计的参数数量，分析者可以选定如下四种工薪者住户类别：

- 1类：没有工薪者；
- 2类：一位工薪者；
- 3类：两位工薪者；
- 4类：三位以上工薪者。

49. 这个模型依然是主效应模型，但回归参数的数量已从5个增加到7个。表十九.4解释了根据模型2估计的回归系数。该模型不再要求分析者假设住户工薪收入对每户工薪者人数的线性关系。但由于模型中缺少交互作用项，所以这个模型的确假设了如下条件：

- 在所有这四种类别中，“每户工薪者人数”效应都相同；
- 在“每户工薪者人数”的所有级别“地区效应”都是一样的。

表十九.4

根据模型2得出的线性回归参数估计数的解释——设应变量为住户的工薪收入

效应（通常为程序输出中所确定的效应）	系 数	估 计 数	解 释
截距	$x_0=1$	α	在参照级别（无工薪者和西部地区）的住户工薪收入
每户工薪者人数			每增加一位工薪者住户收入的变化（按地区调整）
1人(d = 1)	$x_{11hij} - x_{14hij}$	$\beta_1 = \gamma_{11} - \gamma_{14}$	1对无
2人(d = 2)	$x_{12hij} - x_{14hij}$	$\beta_2 = \gamma_{12} - \gamma_{14}$	2对无
3人以上(d = 3)	$x_{13hij} - x_{14hij}$	$\beta_3 = \gamma_{13} - \gamma_{14}$	3对无
无（参照域d = 4）	$x_{14hij} - x_{14hij} = 0$	$\gamma_{14} - \gamma_{14} = 0$	没有估计数
地区			住户工薪收入的地区差别（按住户中的工薪者调整）
北部(d = 1)	$x_{21hij} - x_{24hij}$	$\beta_4 = \gamma_{21} - \gamma_{24}$	北部对西部
南部(d = 2)	$x_{22hij} - x_{24hij}$	$\beta_5 = \gamma_{22} - \gamma_{24}$	南部对西部
东部(d = 3)	$x_{23hij} - x_{24hij}$	$\beta_6 = \gamma_{23} - \gamma_{24}$	东部对西部
西部（参照域d = 4）	$x_{24hij} - x_{24hij} = 0$	$\gamma_{24} - \gamma_{24} = 0$	没有估计数

50. 大多数回归软件包都可以用来检验分类变量间的交互作用。在这种情况下有九个等级的交互作用自由度。虽然解释有两种分类主效应和一种交互作用的回归模型是可能的，但我们不妨推荐一种不同的方法。首先检验交互作用：在这种情况下，模型2可以扩大到包括“住户工薪者”与“地区”之间的交互作用。如果交互作用的统计检验表明存在交互作用，那就通过应用一个只有在第16层级确定的单一分类变量的简单模型将16种可估计的参数纳入整个模型。我们把这个模型叫作模型3并写为：

$$\text{模 型 3} \\ y_{hij} = \alpha x_0 + \sum_{d=1}^{16} \beta_d x_{1dij} + \varepsilon_{hij}$$

51. 这16个新分类变量层级及其估计数（括号内）分别是：

- 北部，一个工薪者 ($\hat{\alpha} + \hat{\beta}_1$)
- 北部，两个工薪者 ($\hat{\alpha} + \hat{\beta}_2$)
- 北部，三个工薪者 ($\hat{\alpha} + \hat{\beta}_3$)
- 北部，没有工薪者 ($\hat{\alpha} + \hat{\beta}_4$)
- 南部，一个工薪者 ($\hat{\alpha} + \hat{\beta}_5$)
- 南部，两个工薪者 ($\hat{\alpha} + \hat{\beta}_6$)
- 南部，三个工薪者 ($\hat{\alpha} + \hat{\beta}_7$)
- 南部，没有工薪者 ($\hat{\alpha} + \hat{\beta}_8$)
- 东部，一个工薪者 ($\hat{\alpha} + \hat{\beta}_9$)
- 东部，两个工薪者 ($\hat{\alpha} + \hat{\beta}_{10}$)
- 东部，三个工薪者 ($\hat{\alpha} + \hat{\beta}_{11}$)
- 东部，没有工薪者 ($\hat{\alpha} + \hat{\beta}_{12}$)
- 西部，一个工薪者 ($\hat{\alpha} + \hat{\beta}_{13}$)
- 西部，两个工薪者 ($\hat{\alpha} + \hat{\beta}_{14}$)
- 西部，三个工薪者 ($\hat{\alpha} + \hat{\beta}_{15}$)
- 西部，没有工薪者 ($\hat{\alpha}$)

52. 将第十六个类别确定为参照组，模型3的截距估计数 $\hat{\alpha}$ 便与该组的住户工薪收入估计数（西部，无工薪者）相对应。其余15组各组住户工薪收入估计数分别为第十六组估计数加上各该组的回归系数估计数。这16个估计数也可以通过直接估计获取。如果在适当的调查软件中应用调查权数和设计结构，分别用这两种方法（模型3或直接估计）获取的估计数及其估计的标准误差应该是相同的。使用模型3开发16个直接估计数一无所获。

53. 如果这16组当中有些组的样本规模很小，这些“小样本”组的估计数精确度就会很差。通过从边际估计数“借用”样本规模，同时围绕如何从假设

的超总体中产生有限总体再做几种假设（像前文讨论的那样），就可以使用一个主效应模型（模型1或模型2）为小规模样本组获得比较精确的估计数。

54. 分析人员一般使用模型来调整若干解释性变量。假定分析者想要调整城市或社区的特征，比如城市化程度（即城市的百分比）。分析的结果可能表明，在考虑到城市百分比并使之标准化之后，地区效应减小了。在一个主效应线性模型中，对（作为一个连续变量或分类解释性变量的）城市百分比进行调整可以提供假定每个地区内的城市百分比都一样（或者都是标准的）地区效应估计数。若不调整协变量，公式（或直接估计数）就会如实地反映地区参数；如果调整公式的协变量，公式就会反映去掉协变量效应之后的地区参数。Korn和Graubard（1999年，第126-140页）讨论了作为一种标准化方法使用预测幅度的问题。

F. 逻辑回归模型

55. 如果应变量是分类变量的话，就不适用线性回归方法。虽然多项式建模法可用，但我们将只讨论可以用逻辑回归模型分析的二元（二级）分类变量。在这个意义上，逻辑回归是多项式回归的一种特殊的简单情况。

56. 对于一个编码为0或1的两类或二元应变量来说，线性回归方法可以奏效，但是可能产生0-1范围以外的预测值。尤其是如果分析者可用的软件包将变量选择程序内置于线性回归软件而不是内置于逻辑回归软件的话，或许可以把线性回归当作一个有二元应变量的初始步骤，用以鉴别作为应变量的良好预测工具的解释性变量。

57. 数值法用于拟合逻辑回归方法的参数；因此，在这种方法集中于某种解决办法的时候可能会遇到困难。在集中使用这种方法遇到问题的时候，用户应留心软件发出的任何警告。一般来说，这些问题可以通过简化模型来解决。

58. 涉及一个连续解释性变量和一个分类解释性变量的逻辑回归模型可以表示如下：

模型 4

$$\log \left(\frac{p(x_{kij})}{1-p(x_{kij})} \right) = \alpha x_0 + \beta_1 x_{1kij} + \sum_{d=1}^D \gamma_d x_{2dkij} + \varepsilon_{kij}$$

59. 除应变量之外，模型4中各项的定义与模型1相同。为了了解逻辑变换，请考虑如下范例： $p(\mathbf{x}_{kij})$ 设为解释性变量的一个函数；为方便起见，指定它为 p 。进一步假定 p 是一个有一套既定解释性变量数值的住户收入水平低于贫困线以下的概率。那么， $p/(1-p)$ 就叫作贫困概率，而 $\log(p/(1-p))$ 就是 p 的对数概率，有时叫作 $\text{logit}(p)$ 。模型4试图使 p 的对数概率与 x 的对数概率相关。观测值是单一的住户，在这里我们不是观测贫困概率，而是观测实际状况：即是否处于贫困。另外，由于应变量是 p 的对数概率，所以参数 $[\alpha, \beta_1$ 和 γ_d

($d = 1, 2, \dots, D$)]也是关于 p 规模的对数概率；而且假定 p 的对数概率与 x 的对数概率之间的关系是线性关系（与上文的模型3相比）。

60. 分类解释性变量的再参数化和参照类别的定义也跟上面所讨论的线性回归一样。逻辑回归程序输出中的回归模型参数貌似线性回归的模型参数，但是它们的解释各不相同。表十九.5归纳了模型4的通常参数估计数的解释。请注意，有五个估计参数（一个截距 a 和四个 β ）。

表十九.5

根据模型4得出的逻辑回归参数估计数的解释——设应变量为贫困线以下的住户指标

效应 (通常为程序输出中所确定的效应)	系数	估计数	解释
截距	$x_0=1$	α	在参照组或0级别的贫困对数概率：西部地区无工薪者
每户工薪者人数	x_{1hij}	β_1	每增加一位工薪者，贫困对数概率的变化（按地区调整）
地区			贫困对数概率的地区差别（按每户工薪人数调整）
北部 ($d = 1$)	$x_{21hij} - x_{24hij}$	$\beta_2 = \gamma_1 - \gamma_4$	北部对西部
南部 ($d = 2$)	$x_{22hij} - x_{24hij}$	$\beta_3 = \gamma_2 - \gamma_4$	南部对西部
东部 ($d = 3$)	$x_{23hij} - x_{24hij}$	$\beta_4 = \gamma_3 - \gamma_4$	东部对西部
西部 (参照域, $d = 4$)	$x_{24hij} - x_{24hij} = 0$	$\gamma_4 - \gamma_4 = 0$	没有估计数

61. 还请注意，逻辑模型参数可以预测贫困的对数概率，但不直接预测贫困概率。让我们考虑表十九.5中的 β_2 。它的表示式如下，是对数概率的差别：

$$\beta_2 = \log \left(\frac{p(\text{north})}{1 - p(\text{north})} \right) - \log \left(\frac{p(\text{west})}{1 - p(\text{west})} \right)$$

根据对数的特点，它也可以表示为一个机会比率对数：

$$\beta_2 = \log \left(\frac{\left(\frac{p(\text{north})}{1 - p(\text{north})} \right)}{\left(\frac{p(\text{west})}{1 - p(\text{west})} \right)} \right)$$

通常，逻辑回归程序的标准输出也提供机会比率，因为这种比率可以很方便地计算为：

$$e^{\beta_2} = \frac{\left(\frac{p(\text{north})}{1 - p(\text{north})} \right)}{\left(\frac{p(\text{west})}{1 - p(\text{west})} \right)}$$

此外，可以根据该模型确定个别住户的贫困概率：

$$P(x_{ij}) = \frac{1}{1 + e^{-\logit[P(x_{ij})]}}$$

62. 作者们在援引逻辑模型拟合结果的时候，有时把机会比率²解释成表示某个域（如北部）发生某种情况（如贫困）的概率相当于另一个域（如西部）发生这种情况（贫困）的概率的两倍。虽然对罕见事件（ p 接近0）来说，这种语句大体符合实际，但是对一般事件来说却远远脱离实际。

G. 多层次模型的应用

63. 现在我们转而讨论多层次建模，并强调首先需要确认调查数据结构。这里有重要意义的是按多阶段设计的调查结构。譬如讲，一个国家的农业生态地区可能形成多个层次，可以从每个层次选择若干行政单位。这些行政单位将形成初级抽样单位。然后从每个初级抽样单位选择第二阶段单位；接着再从第二分阶段单位选择更次级的单位……等等。于是便形成一个等级数据结构。这可能涉及到使用一个或多个层次的分层变量。

64. 例如，一项关于某地区农业住户的调查可能需要使用该地区的行政区划作第一阶段单位，然后从每个政区选择村，再从每个村选择住户，也许在选择住户的时候需要确保包括不同的富有类别。这里必须注意在住户层级收集的数据多样性的不同来源。总体差异包括政区间的差异、村与村之间的差异、以及村内户与户之间的差异。也往往在每个层级收集数据：这里有住户级、村一级和政区级。接下来，重要的是确认并记下哪些变量是在村一级计量的（比如有无推广官员；政府给予化肥的补贴），哪些变量是在住户一级计量的（比如住户的社会和经济特征）。

65. 为了数据分析的目的，可以编制独立的“平面”电子表格文件，以分别储存村一级的信息和住户一级的信息，并且用关键标识符将这些文件联系起来。如果分析的目的需要把村一级的数据和住户一级的数据分开来分析的话，这种做法是恰当的。但如果分析需要把村一级的信息与住户一级的信息结合在一起的话，这种做法就不适合了。比这要理想得多的做法是建立一个关联数据库，就是说，可以把各个不同层级的数据储存在同一个文档里的数据库，同时设有各种链接，以便使一个层级的数据可以跟另一层级的数据相关联。分析工作必须把来自诸多层级的信息聚拢到一起，才能探索不同层级之间的相互关系，这样（譬如讲）才能做出全面的解释。

66. 就此而言，多层次建模是相关性的关键统计技术。之所以需要这种建模技术（见Goldstein, 2003年；Snijders和Bosker, 1999年；Kreft和de Leeuw, 1998年），是因为这样可以在一项多层次设计的各层级内部和层级之间探索各种关系，同时兼顾不同层级的可变性。通过应用适当的软件，比如MLwiN（Rashbash等人，2001年）和SAS（2001年），它还可以（譬如讲）对产生于复杂调查设计的估计数给出基于模型的标准误差。这种标准误差可以充

当兼顾分层和类集的更精确标准误差的合理近似值。应该指出的是，*MLwiN*还可以兼顾抽样权数。这很重要，因为在多层次抽样设计中不均等选择概率可能使关键参数的估计值产生偏差。Pfeffermann等人（1998年）以及Korn和Graubard（2003年）更透彻地讨论了这些问题。

67. 有一个问题值得简要地突出强调一下，就是在这一点上忽略等级式结构——不论数据累积到较高级别还是分解到较低级别——可能产生的结果。如果仅在一个层级需要相关的分析，不会有问题。可是要当心，任何推断都只能在该层级做出。若想根据等级式结构的某一特定层级所分析的数据在另一层级进行推导，那是不可能的。因此，忽略等级式结构的分析不允许探索跨等级的效应。如果把较高级别的单位当作一个分析因子而在较低级别分析数据，那也会遇到困难。这样做是徒劳无益的，因为不能把相关的结论推而广之，运用到总体中的所有较高级别的单位：这些结论只适用于抽样单位。

68. 下面我们以一种情况为例，说明如何运用多层次建模方法来探索关系才能有所收益。还可以从Congdon（1998年）、Langford、Bentham和McDonald（1998年）以及Goldstein等人（1993年）的著作中找到更多的范例。

范 例 1

69. 在一项关于促成太平洋岛国在社区的基础上成功地共同管理沿海资源的因素的调查研究中，从五个国家选择了31个场地，对133个小组进行重点采访，每个组有来自不同住户的受访者二至六人（世界银行，2000年）。选择了斐济、帕劳、萨摩亚、所罗门群岛和汤加五国，代表一系列沿海管理条件。据信，选择的这31个场地所涵盖的各种条件均对管理成败有影响。此项研究按照“单位努力量渔获量”（CPUE）的趋势来收集“对成功与否的评估”、生境状况、场地面临的威胁以及对遵行情况的评估。按5分制的级别计量了前三项指标（5 = 大幅度改善；1 = 大幅度退化）；按4分制的级别计量遵行情况。

70. 还从各国的渔业和环境部门收集全国范围的数据以及现场数据。此外，请每个由若干住户组成的重点受访组对多至三种资源保护（单位努力量渔获量）、三种生境、三种威胁和遵行方面的管理细则进行评估。于是，将此次调研期间收集的信息分为四个层级：国家、场地、重点受访组和特定资源、生境、威胁或规则。

71. 但值得注意的是，该项调查采用了非概率抽样；因此，在能否将任何分析结论推广到任何明确界定的目标人口的问题上可能引起争议。不过，为了本节讨论的目的，假定抽样是在概率基础上进行的，并且用多层次模型对重点受访组一级的数据进行分析——所关注的特定变量是对“单位努力量渔获量”的评估，这种评估通过对有关三种资源的评估打分的平均数获得。（在等级式结构最高级别的）国家级的效应可以作为一个因子（固定效应）纳入模型，因为它本质上是一个分层法变量。但是为了能够在跨越所有共同管理场地的范围内对分析结果加以概括，有必要将各场地作为随机变量而不是作为固定效应纳入模型。各场地范围内的重点受访组也作为一种随机效应纳入模型。多

层次模拟的本质在于包含一个由固定效应变量和随机效应变量组成的矩阵。这种模型也可以对场地一级的变量和重点受访组一级的变量之间的相互作用进行解释。

72. 为了说明如何拟定多层次模型以回答特定调查问题，我们以2000-2001年马拉维进行的《粮食生产与安全调查》为例（Levy和Barahona，2001年）。该调查旨在对一项方案进行评价，这项方案的目标是通过分发内含化肥、玉米和豆类种子的启动包增强农村小生产者的粮食安全。

范 例 2

73. 《粮食生产与安全调查》是一项以政区为层次采用两阶段分层抽样法的全国调查。马拉维全国有27个政区，每个政区选定四个村子，每个村选出大约30个住户。选择村子在住户数量方面的限定范围是：最少40户（以便每个村有足够的户数确保能够采访启动包的受领者），最多250户（以便使调查团队能够在可用资源所允许的限期内完成调查工作）。⁴⁰ 在此限度内，每个阶段进行随机抽样。整个调查期间总共采访了108个村的3 030户。

⁴⁰ 对目标人口的此项约束条件将分析推断限定在这一户数范围内的村民。

⁴¹ 资产指数是根据不同的牲畜数量和诸如收音机、自行车、牛车……等家庭财产确定的一个加权平均数。

⁴² 收入指数是根据一系列不同来源收入确定的。

74. 我们为进行多层次建模所考虑的数据来自一份在调查期间填写好的住户问卷。我们在说明中行将考虑的变量子集包括：区、村、住户的鉴定数量；户主的性别和年龄；家庭人口；该户是否领到启动包；以及反映家庭资产⁴¹和收入⁴²的两项指数。

75. 有几种模型可能适用于此类数据。编拟模型的第一步就是确定哪些变量是随机效应，那些变量是固定效应。

76. 在范例2中，政区是分层法变量，被视为固定效应。一般来说，如果抽样过程多次重复的结果都产生同样一套选择的话，任何效应均可视为固定的。另一方面，如果村子和住户是随机选择的，它们便构成模型中的固定效应。

77. 譬如讲，分析资产指数(AI)的基本模型为：

模 型 5

$$y_{ijk} = \mu + d_k + U_{jk} + \varepsilon_{ijk}$$

式中， d_k 是政区效应，($k = 1, 2, \dots, 27$)；指数 i 和 j 分别对应于第 i 户和第 j 村。有时为了方便起见，将政区参数视为体现了政区 k 的平均值AI对所有政区整体平均值AIs的偏差。可是模拟软件采用了一种不同的参数化方法，将其中一个政区效应确定为零(0)。于是，其余的效应便提供了每个区的AIs与效果值为零的那个区的AI之间的比较。

78. 在这个模型中， U_{jk} 和 ε_{ijk} 表示随机变量，分别代表政区 k 内所有村子之间的差异（假定所有政区都一样）和政区 k 、村 j 中所有住户间的差异（假定所有村和区的组合都一样）。 U_{jk} 和 ε_{ijk} 是随机变量，假定在该模型中它们分别代表

平均值为零的正常分布变量和固定变量，即 σ_u^2 和 σ_e^2 。进一步假定它们是相互独立的。因此，我们可以分别将其写成 $U_{jk} \sim N(0, \sigma_u^2)$ 和 $\varepsilon_{ijk} \sim N(0, \sigma_e^2)$ 。

79. 拟合该模型，遂得 σ_u^2 和 σ_e^2 的估计数及 d_k 的估计数，以及相关的标准误差。根据 d_k ($k = 1, 2, \dots, 27$)的参数估计数可对27个政区的AI平均数进行比较。

80. 现在，假定想要调查住户大小（一个定量变量）如何影响AI平均数的差异以及该住户是否得到启动包（一个二元变量）。按照跟标准的一般线性建模一样的做法，将上述变量纳入模型。该模型表示如下：

模 型 6

$$y_{ijk} = \mu + d_k + U_{jk} + t_{p(ijk)} + \beta x_{ijk} + \varepsilon_{ijk}$$

式中， $t_{p(ijk)}$ 代表与得到启动包相符的效应； x_{ijk} 代表住户大小；而 β 代表描述 x_{ijk} 和 y_{ijk} 关系——亦即住户大小（HHSIZE）和资产指数（AI）的关系——的斜率。

81. 这里， $t_{p(ijk)}$ 和 β 均被视为固定效应，而 U_{jk} 和 ε_{ijk} （跟以往一样）都是随机效应。形成该模型的假定条件是：在所有政区和村子HHSIZE和AI的关系都一样。

82. 将差异的两个组成成分（村和住户）纳入上面的公式就意味着，该公式考虑到等级式结构的两个层级的可变性。这就是说，与 $t_{p(ijk)}$ 和 β 相关联的标准误差计算正确，与这些参数有关联的（调查）显著性检验结果也会如此。假如一个一般线性模型把村一级当作固定效应进行拟合的话，情况就不是这样了。即便使用（包含抽样权数的）调查软件，标准回归程序也会忽略任何村中住户之间的相互关联结构。

83. 把村当作随机效应还有另一个重要好处。要是把村当作固定效应的话，那么分析结论就会仅仅适用于调查期间采访过的那一组村子。把村子当作随机效应就意味着，与住户大小和AI的关系有关的结论、领到和没有领到启动包的住户之间AI值的比较，以及各区之间的比较等都可以推而广之，适用于所有村子的更广泛人口。通过将相关的交互作用项纳入公式，还可以探索区一级的固定效应 d_k 与启动包受领者效应 $t_{p(ijk)}$ 之间的相互作用。

84. 如果将模型6中的斜率项 β 当作村与村之间各不相同的一个随机变量，还可以有另一个模型——通常叫作随机系数回归模型。于是，模型6就变成了：

模 型 7

$$y_{ijk} = \mu + d_k + U_{jk} + t_{p(ijk)} + \beta_j x_{ijk} + \varepsilon_{ijk}$$

式中，设 β_j 为 $N(\beta, \sigma_\beta^2)$ 。另外，既然 β_j 在村与村之间是随机的，也可以把它当作一个带有 U_{jk} 的协方差，比如 $\sigma_{\beta u}$ 。

85. 由此可见，在本节所讨论的分析中，通过检验 σ_{β}^2 为0的假设，我们即可有效地得知各村之间在AI对HHSIZE关系的斜率上是否存在差异。如果这一假设不可否认，则可以推断所有村子的这种关系形式都是一样的。

86. 有可能将此模型进一步扩大到包括村一级的各种变量，比如享用清洁水供应的机会或者得到农业技术推广官员提供咨询服务的机会。这里就显出了多层次建模带来的实际好处，因为接下来就可能探索村一级的变量与住户一级的变量之间的关系了。可见，通过多层次建模来研究一个等级抽样计划的不同层级各种变量之间的关系是可能的。好处在于，当把不同层级的变量合在一起进行分析的时候，能够考虑到较低层单位间相关的关系。在上述范例中，可以考虑进一步的模型（比如包含户主性别和年龄的模型）以及这些项与前面包含在模型中的那些项之间的交互作用。

87. 当然，在多层次模型的拟合方面是有限度的。跟所有其他建模程序一样，假设的多层次模型是在合理的程度上并符合样本设计的情况下被认为是“正确”的。这些假设是否符合实际，当然可以辩论。

H. 旨在支持调查过程的建模

88. 即使在住户调查被严格用于提供描述性统计数据的情况下，也需要进行旨在支持其他调查过程的建模。对不答复的调整往往直接或间接地依据统计模型：Groves等人（2002年，第197-443页）讨论了说明不答复情况的各种方法。可以用逻辑回归模型来开发预测的响应倾向（指数），以便对不答复情况做调整，或者根据类似的响应倾向确定加权等级[譬如，见Folsom（1991年）；Folsom和Witt（1994年）；或Folsom和Singh（2000年）]。也可以把预测统计模型当作估算遗漏数据的部分程序来使用[譬如，见Singh、Grau和Folsom（2002年）]。最后，可以用统计方法来评价嵌入调查中的方法论实验成果[譬如，见Hughes等人（2002年）]。

I. 结论

89. 我们在本章的目的是讨论有关调查数据分析的问题。这些问题结合采取描述性的和分析性的两种调查数据处理做法，涉及到运用调查权数和适当的方差估计方法。本章还概述了建模技术在调查数据分析中发挥作用的实际情况。建模技术是有用的工具，但是在应用这种技术的时候需要慎重思考，并且注意它们所依据的假设条件。

90. 我们讨论了在根据调查数据开发描述性和分析性统计中调查权数和样本结构认知所发挥的作用。利用调查权数并兼顾样本结构的调查数据软件，可以用来估计基于调查数据的线性和逻辑回归模型的参数。以样本为基础的估计数是通过使模型适合整个有限总体所可能获得的估计数。另外，还可以获得这些估计数的标准误差。回归模型中适用于调查数据的解释性变量几乎总能观测到，因为它们存在于总体当中，而不是根据某种实验设计随机指定的。分析者应当明白，基于调查数据的回归系数只不过体现了存在于总体中的应变量和解

释性变量之间的关系，而未必暗含因果关系。我们讨论了回归和逻辑回归模型的参数如何与简单描述统计发生关系，以及如何就某些相对简单的模型来解释这些参数。

91. 尤其是，一般认为多层次建模是一种相当“先进”的技术，最好向熟悉该技术应用和制约条件的统计学专家征求意见。目前看来，多层次模型还很少用于分析发展中国家的调查。可是它们在这些国家的应用极为可取，因为通过这些模型可以洞察不同层级各种变量之间的相互关系，并能在多层次设计中计量不同层级抽样单位的可变性。

92. 我们说明了，对于熟悉一般线性回归模型（GLMs）的人来说，多层次模型的编拟并不太困难；但是这方面也存在各种与模型有关的假设条件，需要通过残值分析来检查这些假设，就像对待一般线性回归模型那样。如果所关注的主要答复是二元性质的话，也可以采用多层次建模法，不过我们没有就此举例说明。在确定哪些是随机效应、哪些是固定效应以及如何通过模型说明来帮助回答具体调查问题的时候，也需要谨慎行事。

93. 然而，本章所讨论的建模方法跟所有统计技术一样，也需要在应用这些技术的时候承认各种制约条件。我们鼓励使用调查权数和承认样本设计结构的分析软件。必须承认，获得可供使用的顾及抽样设计的适当软件还是有困难的。第二十一章描述了几种涉及到多重回归和逻辑回归程序、把注意力放在抽样设计问题上的软件包。但遗憾的是，这些软件包没有为多层次模型的拟合提供方便。为此目的，用户需要借助比较通用的统计软件，诸如SAS（2001年）、GenStat（2002年）和SPSS（2001年），或者像*MLwiN*（Rah-bash等人，2001年）那样的专用软件包。

94. 本章提供的一些模拟方法可以充当调查数据分析的工具。我们建议调查分析人员和研究人员在适合调查目标的情况下在调查数据分析过程中认真考虑这些方法，以便从代价高昂的调查数据中获取尽可能多的信息。

鸣 谢

作者特别感谢各位审稿人和编辑提出的许多有价值的意见和建议，尤其感谢Graham Kalton博士对有关调查权数的讨论所做的改进。

同时也感谢大不列颠及北爱尔兰联合王国政府国际开发部（DFID），通过其资助发展中世界进行的令人感兴趣的诸多调查项目为本章提供观点。但作者依然对本章的材料负完全责任，而且这些材料并不意味着国际开发部发表了任何意见。

参考资料

- Chambers, R. L. 和C. J. Skinner (2003年)。《调查数据的分析》。联合王国，奇切斯特：威利国际出版公司。
- Chromy, James R. (1998年)。《有限样本对国家评估样本要求的影响》。加利福尼亚，帕罗奥多：美国研究机构，NAEP有效性研究。
- Cochran, W. G. (1977年)。《抽样调查》，第三版。纽约：约翰·威利国际出版公司。
- Congdon, P. (1998年)。“婴儿健康结果的多层次模型：产妇的风险因素与地理变化”。《统计员》，第47期，第1部分，第159-182页。
- Deville, J. C. 和C. E. Sarndal (1992年)。“抽样调查中的填补评估”。《美国统计学会志》，第87期，第87号，第376-382页。
- Folsom, Ralph E., Jr. (1991年)。“抽样和不答复误差减少的指数和对数权数调整”。《美国统计学会社会统计科的记录》。佐治亚州，亚特兰大：美国统计学会。
- Folsom, Ralph E. 和Michael B. Witt (1994年)。“测试收入和方案参与调查新损耗不答复调整方法”。《美国统计学会调查研究方法科的记录》，第428-433页。
- Folsom, R. E. 和A. C. Singh (2000年)。“极值、不答复和事后分层抽样权数校准的总体指数模型”。《美国统计学会调查研究方法科的记录》。印第安纳，印第安纳波利斯。
- GenStat (2002年)。GenSta软件，第6版，联合王国，牛津：VSN国际公司。
- Goldstein, H. (2003年)。《多层次统计模型》，第三版。伦敦：阿诺德出版社。
- _____和其他人 (1993年)。《学校考试成绩的多层次分析》。《牛津教育评论》，第19期，第425-433页。
- Graubard, B. I. 和E. L. Korn (2002年)。“超总体推论”……《统计科学》，第17期，第73-96页。
- Groves, Robert M. 和其他人 (2002年)。《调查中的不答复情况》。纽约：约翰·威利国际出版公司。
- Horvitz, D. G. 和D. J. Thompson (1952年)。“并非有限全域替换的抽样普遍化”。《美国统计学会志》，第47期，第663-685页。
- Hughes, Arthur和其他人 (2002年)。“访谈者经验对回答者药物使用报告的影响”。《重新设计现行的国家住户调查：方法问题》，J. Gfoerer、J. Eyerman和J. Chromy等。健康与人类服务部出版物，第SMA03-3768号。马里兰，罗克韦尔：应用研究办公室，药物滥用和精神健康服务局，第161-184页。
- Kish, Leslie (1965年)。《抽样调查》。纽约：约翰·威利国际出版公司。

- Korn, E. L. 和 B. I. Graubard (1999年)。《健康调查分析》。纽约：约翰·威利国际出版公司。
- _____ (2003年)。“利用调查数据估计方差分量”。《皇家统计学会志》，B辑，第66期，第175-190页。
- Kreft, I. 和 J. de Leeuw (1998年)。《介绍多层次建模》。伦敦。
- Langford, I. H.、G. Bentham和A. McDonald (1998年)。“地理总体健康数据的多层次建模：关于欧洲共同体恶性黑色素瘤死亡率和紫外线暴露的个例研究”。《医学统计》，第17期，第41-58页。
- Levy, S. 和 C. I. Barahona (2001年)。“2000-2001年目标明确的投入方案：主要报告”。尚未发表。
- Pfeffermann, D. 和其他人 (1998年)。“多层次模型中的不均等概率抽选法权重”。《皇家统计学会志》，B辑，第60期，第23-40页。
- Rasbash, J. 和其他人 (2001年)。《MLwiN 1.10.0007版》。多层次模型项目。伦敦：伦敦大学教育学院。
- 三角研究所 (2002年)。《SUDAAN使用手册，8.0版》。北卡罗来纳，研究三角园区：三角研究所。
- Rubin, Donald B. (1987年)。《调查中不答复的多重填补》。纽约：约翰·威利国际出版公司。
- SAS (2001年)。《SAS 8.2版》。北卡罗来纳，卡里：SAS软件研究所，SAS出版社。
- Singh, Avinash、Eric Grau和Ralph Folsom Jr. (2002年)。“全国家庭药物滥用调查药物使用数据的预测相邻填补”。《重新设计现行的国家住户调查：方法问题》，J. Gfroerer、J. Eyerman和J. Chromy等。健康与人类服务部出版物，第SMA03-3768号。应用研究办公室，药物滥用和精神健康服务局。
- Skinner, C. J.、D. Holt和T. M. F. Smith等。(1989年)。《复杂调查分析》。纽约：威利国际出版公司。
- Snijders, T. A. B. 和 R. J. Bosker, R. J. (1999年)。《多层次分析：基本和高级多层次建模简介》。伦敦：赛奇出版社。
- SPSS (2001年)。《11.0版SPSS软件》。伊利诺伊，芝加哥：LEAD技术公司。
- Tukey, J. W. (1977年)。《探索性数据分析》。马萨诸塞州，雷丁：阿狄森-韦斯利出版社。
- Woodruff, R. S. (1971年)。“粗略估计复杂样本方差的简单方法”。《美国统计学会志》，第66期，第411-414页。
- 世界银行 (2000年)。《乡村的呼声：太平洋岛屿海岸管理的比较研究》。《太平洋讨论文件汇编，第9号》。哥伦比亚特区华盛顿：世界银行。巴布亚新几内亚和太平洋岛屿国家管理股。

第二十章

较先进的调查数据分析方法

Gad Nathan

希伯来大学

以色列耶路撒冷

摘 要

我们在本章审议大多数样本调查在实际当中采用的复杂样本设计对调查数据分析的影响。详细说明了设计可能影响或不影响分析的各种情况，并定义了有关的基本概念。在确立了分析模型的情况下，我们考虑模型与样本设计之间可能存在的关系。如果设计可能对分析有影响而与设计有关的附加解释性变量却不能添加到分析模型的话，可以采用两个基本方法：要么采用结合设计加以修改的经典分析方法；要么专门为每项设计开发一个新的分析工具。说明了对线性回归、线性模型和分类数据分析有实际数据应用价值的各种不同的方法。

关键词：复杂样本设计、调查数据分析、线性回归、线性模型、分类数据分析、基于模型的分析。

A. 引言

1. 样本设计与数据分析

1. 不论在发达国家还是发展中国家，绝大多数样本调查的首要目的都是描述性的，亦即提供一个有限总体描述量度的点估计数和区间估计数，比如平均数、中位数、频率分布和各种定性变量的交叉表，等等。可是，正如第十五章至第十九章所说明的和第二十一章将要说明的那样，现在越来越关注有关各种调查变量之间关系的推断，而不仅仅是对现象的描述。

2. 在本章，我们将试图评估通用复杂样本设计对调查数据分析的影响。我们将努力确定设计可能在哪些情况下影响分析。通常，如果样本设计所依据的变量包含在分析模型中的话，样本设计对分析没有影响。可是，往往有些设计变量未被纳入模型，其原因要么是因为设定有误，要么是因为对作为解释性变量的设计变量缺乏兴趣。这就可能造成严重偏倚。

3. 我们将讨论两种基本方法，以便在没有将与设计有关的附加变量补充到分析之中的情况下，用以处理复杂样本所产生的数据。一种方法就是修改一种用于处理简单随机样本数据的已开发的经典分析工具。第二种方法就是为特定的复杂设计开发一种新的分析工具。

4. 下面，我们提供一些样本设计可能影响分析的实例，界定几个基本概念，并讨论设计效果在分析复杂样本数据中所起的作用。B节描述用于分析复杂样本数据的两种基本方法。在C节和D节，我们分别讨论与连续数据分析和分类数据分析有关的实例。最后一节是总结和几点结论。本章附件提供了若干正式定义和技术成果。

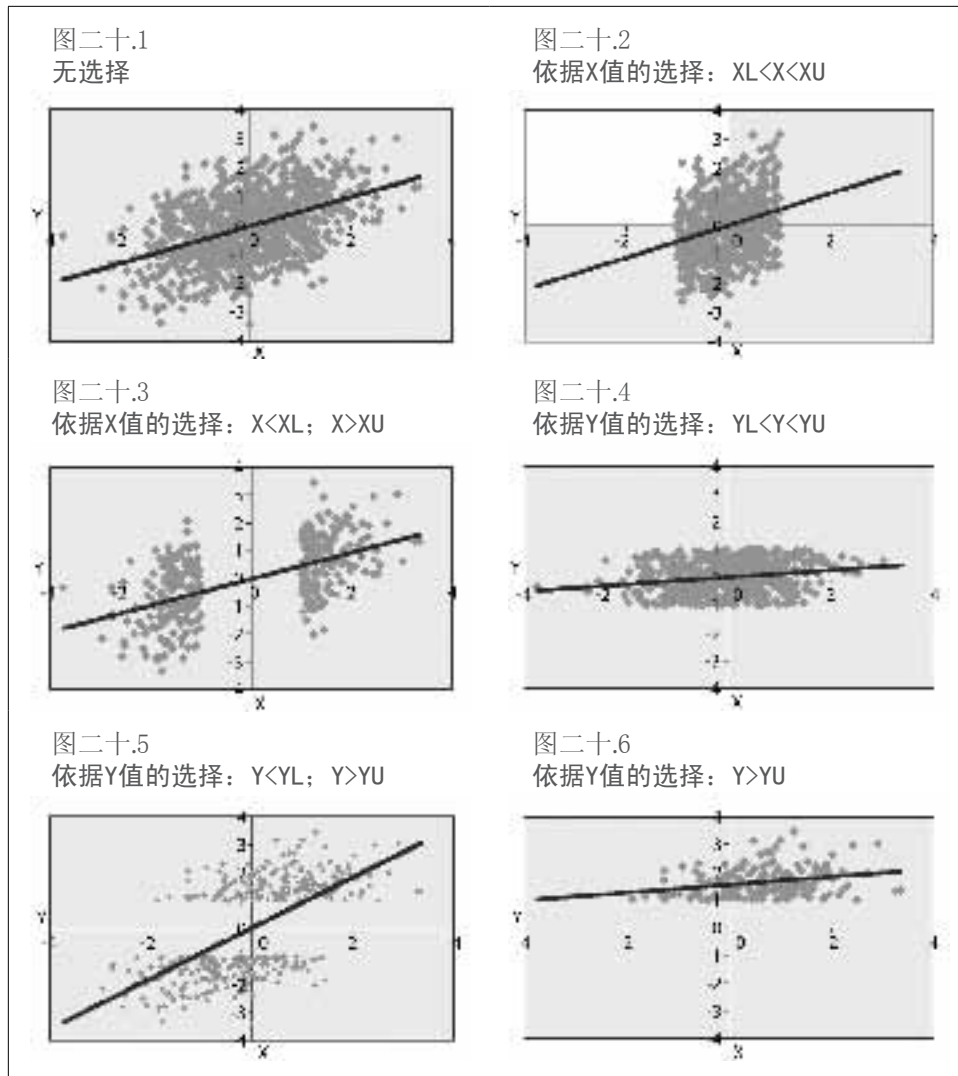
2. 样本设计对分析有影响（和无影响）的实例

5. 为了说明样本设计对分析的潜在影响，让我们考虑下面这个简单却能说明问题的实例（详见Nathan和Smith, 1989年）。设 Y 为关注变量； X 为一个辅助变量。假定线性回归模型 $Y_i = \alpha + \beta X_i + \varepsilon_i$ ，其中 $\varepsilon_i | X_i \sim N(0, \sigma_\varepsilon^2)$ ，对总体是成立的。该模型对从这个总体中选定的任何简单随机样本也是成立的。有时候，对简单随机样本来说，假定 ε_i / X_i 之间的独立性要比对从中提取样本的总体更适合。譬如讲，在一个总体当中， ε_i 值可能与同一住户的不同成员相关；而在个人简单随机样本中，如果每户选择不止一个成员的概率很小的话，这种相关就可以忽略不计。

6. 按照简单随机抽样法，回归系数的标准估计数是无偏倚的。图二十.1的显示勾绘了 Y 对应于 X 的总人口斜线；对于一个简单随机样本来说，显示应该是一样的。在图二十.2至图二十.6的显示中，分别采用了与简单随机抽样法截然不同的方法从总体中抽取样本。让我们考虑完全依据 X 值的样本选择，比如通过斜截图二十.2和图二十.3中固定极限以外（或以内）的 X 值数据点。从这些图可以清楚地看出，样本的选择对回归截距（ a ）和斜率（ β ）的参数估计数没有任何影响（尽管它可能影响到估计量的方差）。

7. 现在来考虑基于目标变量 Y 值的样本选择，比如通过斜截图二十.4至图二十.6中固定极限以外（或以内）的 Y 值数据点。

在这些情况下很明显，回归斜率的估计数出现了偏差。在最后（图二十.6）的情况下，斜截不对称，截距也有偏差。这些例证是极端的，因为在样本调查中极少发生依据这种应变量的样本选择。不过在实验或观测研究中，比如在流行病学案例控制研究或基于选择的经济学研究中，这种情况却很普通[譬如，见Scott和Wild（1986年）以及Manski和Lerman（1977年）]。但是在许多情况下，样本是根据设计变量选择的，而这些变量可能与应变量密切相关。因此，在公司企业或农场调查中广泛应用的某种通用抽样程序是按照与规模成比例的概率来选择单位。规模量度——比如说上一年的产量，很明显就会联系到作为今年生产变量的关注变量。如果忽略样本设计的话，像回归系数之类的模型参数的标准估计就可能出偏差。



8. 以上例证说明了依据误以为来自简单随机抽样而实为来自复杂样本的数据进行分析的危险性。这些例证揭示了鉴别在什么情况下设计有可能影响分析，以及当设计的确影响分析的时候将设计考虑在内的必要性。

3. 基本概念

9. 大多数样本调查都主要是为描述性（或计数的）目的而设计的。它们意在估计有限总体参数值，比如中位数住户收入或所有成年人中的艾滋病患病率。原则上，如果调查涵盖总人口，亦即如果是普查计数而不是总人口的某个随机样本计数的话，这些统计是可以精确计量的。调查抽样的标准理论确保随机样本数据可以用来提供有限总体参数及其抽样误差的无偏估计数，而不论样本设计多么复杂。它能保证样本设计是概率样本设计，就是说，人口中的每个单位都有一个已知的被抽样的正概率。这些经典的有限总体参数估计方法叫作基于设计的（或基于随机化的）方法，因为所有推断都以样本设计特征为依据，而通过样本概率分布来进行。不过应该指出的是，通常只有在拥有大量人

口信息可用的情况下才能够评价各种不同估计策略（辅以估计公式的样本设计）的效率。而实际情形通常并非这样。因此，就连经典的抽样文献（比如，Cochran, 1977年）也往往要依靠模型来论证特定抽样或估计方法的合理性。如果人口数值是基于简单回归模型推导出来的，那么在某些假定条件下，比率估计量要比简单扩展估计量效果好。由此可见，基于设计的方法往往是借助于模型的，但并非基于模型（或依靠模型）：对于借助模型的样本设计方法或描述式统计估计方法来说，几乎无偏倚的有限总体参数的估计不需要模型假设条件。

10. 基于模型（或预测理论）的样本设计和估计方法假定，有限总体数值实际上是基于一个带有超总体（模型）参数的假设模型的超总体分布实现值。关于进一步的细节和讨论，见Brewer和Mellor（1973年）；Hansen, Madow和Tepping（1983年）；Särndal, Swensson和Wretman（1992年）；以及Valliant、Dorfman和Royall（2000年）的参考资料。与借助于模型的估计方法所不同的是，基于模型的估计方法的效率提高取决于假定模型的有效性。因此，如果对模型假设的有效性有任何疑问的话，均方就会明显降低，以致可能不宜使用纯粹基于模型的分析方法。Hansen, Madow和Tepping（1983年）在这方面提供了一个很好的实例，其中假定回归模型中没有截距，即使有截距实际上也几近于0，以致基于模型的推论无效。

11. 现在，越来越多的调查成果不但用于描述目的，而且用于分析目的。往往在设计调查的时候就想着调查成果的分析用途。这是因为决策者和研究人员对原始数据所隐含的过程以及为所调查的各种变量之间关系建模感兴趣。此类分析显然需要对模型做出种种假设。分析的目的在于确认一个假设模型的有效性，并估计模型参数的有效性，而不是有限总体参数的有效性。可见，分析本来就是基于模型的。按照定义，关于模型参数的推断实际上必须以相关的模型为依据。

12. 不过应该指出，如果人口规模非常之大，而假设的模型确实成立的话，实际上在模型参数和与之对应的有限总量参数之间就没有什么差别了。譬如，如果标准线性回归模型 $Y_i = \alpha + \beta X_i + \varepsilon_i$ ，其中 $\varepsilon_i | X_i \sim N(0, \sigma^2)$ 成立，而人口规模很大，那么，根据中心极限定理，标准人口回归系数值 B （见附件）就会非常接近模型参数值 β 。因此，虽然我们应该集中估计由此产生的模型参数，但这些参数有时候会被与之对应的有限总体参数所取代。为了使陈述简明扼要，接下来的各种实例大多可以按一元分布（亦即一个单一应变量和一个单一解释性变量）列出公式。通常可以直接将其结果扩展成为多元分布。

13. 总之，假定模型是统计分析的组成部分。为了分析样本调查数据，选一个好的模型来拟合数据是分析工作的关键部分。研究者和分析者必须在应用分析方法之前透彻理解他们要研究的过程所依据的模型。我们将从下面各节得知，根据复杂样本设计所得数据的应用，不但需要了解所依据的模型，而且需要了解复杂设计如何影响分析。

4. 设计效果及其在复杂样本数据分析中的作用

14. 前面第六章和第七章围绕设计效果及其估计的话题，主要结合它们在计数调查的设计和估计中发挥的作用进行了全面讨论。我们从本章可以看出，它们在复杂样本调查所得数据的分析方面也发挥着重要作用。基本概念基于这样的事实：假定各项模型假设均成立，根据简单随机抽样，模型参数无偏估计数和这些估计数的方差估计数易于获取。这些估计数和方差估计数为检验与模型参数有关的各项假设奠定了基础。譬如，按照简单随机抽样，设简单回归模型 $Y_i = \alpha + \beta X_i + \varepsilon_i$ ，其中 $\varepsilon_i | X_i \sim N(0, \sigma_\varepsilon^2)$ ，其普通最小平方样本估计量 b 是 β 的一个无偏估计量，其方差 $v(b)$ 的一个无偏估计量是现成的（见附件）。于是，借助于中心极限定理，零假设 $\beta = 0$ 的标准检验便以检验统计量 $b / \sqrt{v(b)}$ 为基础。在复杂样本设计（比如分层类集样本）的情况下，如果回归模型成立，且样本设计不取决于 Y_i 值（譬如与图二十.4-二十.6中的情况相反），那么，估计量 b 就使模型保持无偏差。这就是说，在特定的回归模型中预期的 b 值是 β ，其中的预期是就 Y_i 值的超总量分布而言。从C.1节可知，如果模型不成立的话，情况就不是这样了。然而，即使模型设定成立， $v(b)$ 也不再是模型方差 b 的一个有效估计量，必须予以修改。一个正确模型方差的直接估计量往往是可以计算的，比如通过使用第二十一章描述的软件包，并可以用来取代 $v(b)$ 。如果没有直接的估计量可用，也时常可以获得一个好的设计效果估计量，用 $d^2(b)$ 来表示。可以用它来修改检验统计量，即用 $d^2(b) \times v(b)$ 取代 $v(b)$ 。下面，我们将进一步介绍设计效果的特定用途，用来修改针对其他应用的标准检验统计量。

B. 复杂样本数据分析的基本方法

1. 模型设定是分析的基础

15. 基础模型的正确设定是任何分析的一个基本步骤。错误设定模型——不论是排除了相关的解释性变量（或者包含了不相干的变量）还是使用了错误的函数形式（比如采用线性而不是二次形式）——所带来的后果是众所周知的，在标准教科书中已有明示。它们可能表现为模型参数的估计偏差（主要是排除了相关的变量）、效率的损失（大多与包含了错误的解释性变量有关）、以及假设检验的规模和能力的改变等方面。如果错误的模型设定直接关系到样本设计变量或关系到与设计变量相关联的变量的话，这些影响可能会更大。不过，调查设计变量也许与研究的目标无关，认识到这一点很重要。另外，就调查项目而言，也没有正当理由将它们纳入分析模型。

16. 将调查设计变量纳入模型有两种基本方法。一种是总量法，它把有关模型看作总体一级的模型，在概念上独立于用以获取数据的样本设计。按照这种方法，只将与调查项目有关的设计变量纳入模型。譬如讲，我们打算不分地域地解释关于“受教育的年限”这个解释性变量使用/不使用的二元变量。比如，样本是按地理区域分层的，与此有关的模型不尽相同。在这种情况下甚

至连简单随机抽样法都用了。结果，模型中就可能包含分层（见下面讨论的分类法）以反映模型变量间的关系在地区上的差异。对比之下，如果只是为了运作的理由（方便或成本）而进行分层和按层次分配样本的话，样本的权数就不大可能与总体模型相关。否则，若在不受分层影响的分析中包含抽样权数，就可能导致在一定程度上降低效率。可是，一个不受层次影响的模型比较容易解释；而如果某些被忽略的层效应确实存在的话，关于模型失败的原因解释起来就比较费劲。

17. 第二种方法是分类法，这种方法将分析者的模型扩展到不仅包括有关调查变量，而且包括调查设计中所使用的变量以及那些与设计中所体现的人口结构有关的变量。模型中包含与分层和类集有关的设计变量，是为了体现人口的复杂结构。譬如，在前面的例证中，每个地域层次的模型可能包含一套不同的系数（既有截距也有斜率）。按照分类法进行的推论充分考虑到样本设计，并假定所有设计变量都恰当地纳入了模型。需要用这种方法估计的大量参数可能会造成困难，并且与较为简约的总体模型参数相对而言可能会导致估计数不太准确。只有在分析者相信所设定的模型确实与其目的相关的情况下，才适合采取分类法。

18. 所要采取的方法——是总体法还是分类法——适当与否，取决于分析者的目的。总体法比较适合研究对总体有影响的因素，因此对评价国家政策行动较为有用。分类法则比较适合研究微观影响和局部影响，以及针对部门的决策。关于进一步的例证和讨论，见Skinner, Holt和Smith(1989年)以及Chambers和Skinner(2003年)。

2. 模型与样本设计之间可能存在的关系：告知性设计和非告知性设计

19. 在分析复杂调查数据的时候，区分告知性与非告知性调查设计十分重要。模型一经设定，分析者必须考虑，在确定了模型协变量的条件之后样本选择概率是否与响应变量的值有关联。在模型中的协变量值既定的情况下，如果样本观测值的联合有条件模型分布与其在总体中的有条件分布有差异，抽样过程就是告知性的。只有当这两种分布相一致的时候，样本设计才是非告知性的（或者说是可以忽略不计的）；在这种情况下可以使用标准分析方法，就好像观测值是出自简单随机样本似的。如果样本设计是告知性的，那么，适用于样本数据的模型就有别于总体模型。在这种情况下忽略抽样过程就可能产生有偏差的点估计量，并且使分析失真，就像在常规分析中将变量排除于模型之外那样。请注意，将与设计有关的变量正确地纳入模型，即可确保设计为非告知性的。

20. 在一个模型中包含所有与设计有关的变量，主要存在两个问题。首先，也许不能确切知道设计中到底使用了哪些变量，如果知道，可能也无法掌握它们的值。即便设计变量业经确定和测量，分析者也可能不知道这些设计变量与有关变量之间关系的确切形式（比如，是线性的还是指数的）。譬如，如

果是分层设计，那就需要检查某种回归关系对不同的层次有不同的斜率和截距的可能性。

21. 其次，当把设计变量正确地纳入模型的时候，由此产生的估计数也许对分析没有什么价值，因为附加的变量不具固有的项目调查重要性（请回顾前面关于总体分析和分类分析的讨论）。这就意味着，不能总是仅仅通过修改基础模型来处理复杂样本设计对分析的影响这一问题。下面，我们将考虑如何修改标准分析方法以便考虑到复杂设计，以及如何构建针对设计的专用估计和分析方法。

3. 使用适合复杂样本分析的标准分析软件包所涉及的问题

22. 几乎普遍使用的标准统计分析软件已经导致可靠的统计实践被广泛滥用。在分析复杂样本调查数据方面，这种滥用情况尤为严重。

23. 统计软件在便利分析方面所具备的优点，也令人遗憾地带来了在对有关基本原则缺乏最起码认识的情况下进行分析的可能性。这已成为定量研究工作、尤其是社会科学领域的一个严重问题。最通用的软件对待数据的方式就好像它们源自简单随机抽样似的，这一事实的存在使得问题更加复杂化了。如前所述，如果设计是告知性的，这个问题就可能推论产生严重偏差。可是只要适当地小心从事，往往可以通过调整标准软件，使其大体上捕捉到或说明复杂设计的效应。尤其是最新SAS®版本的SURVEYREG程序（版本8和9）以回归分析为特色，按照下述类似方式顾及到样本设计的因素[见An和Watts(2001年)]。

24. 例如，让我们考虑由 $Y_i = \alpha + \beta X_i + \varepsilon_i$ ，其中 $\varepsilon_i | X_i \sim N(0, \sigma_i^2)$ 界定的线性（异方差）回归模型。通常，标准计算机程序计算 b ，即普通最小二乘方（OLS）估计量 β ，或广义最小二乘方（GLS）估计量 b_G ，其中用 σ_i^2 的倒数分别对总数和积数加权，并假定已知其数值（或相对值）（见附件）。如果模型成立，二者均为参数 β 的无偏估计量；不过在异方差情况下 b_G 是一个更有效的估计量。标准程序还提供普通最小二乘方估计量 $v(b)$ 和广义最小二乘方估计量 $v(b_G)$ 的方差估计数，按照适当的模型（在 $v(b)$ 情况下为异方差模型）这两种方差估计数均不存在模型偏差。

25. 许多情况下，可能对模型的有效性存有疑问。这样的话，或许不估计 β 而估计 β 的有限总体对应值更为恰当；我们用 B 来代表这个对应值（见附件）。虽然 b （普通最小二乘方估计量）在 $\beta \approx B$ 的情况下是一个模型无偏估计量，但并非在各种情况下都是模型无偏的。在适当条件下，模型加权（Horvitz-Thompson）估计量 b_w ——它的叉积和二乘方都用列入概率的倒数加权——既是模型一致的也是模型无偏的。另外，可以通过使用 w_i 值作权数，从许多标准程序的加权回归选项中获得 b_w 值。或者，可以通过 $Y_i/\sqrt{\pi_i}$ ， $X_i/\sqrt{\pi_i}$ ，变换变量的未加权回归，其中用 $1/\sqrt{\pi_i}$ 取代截距，来获得 b_w 。可是必须强调一点：在这两种选择中不论选定哪一种，大多数标准程序所报告的方差-协方差矩阵估计数

都不正确——不论作为设计平均数二乘方误差的估计量还是作为模型方差的估计量——除非是在异常情况下。

26. 总之，除非可以确定复杂设计对估计没有严重影响，否则就应避免使用不考虑复杂调查设计的标准软件程序。一般情况下，通过恰当使用标准软件可以达到上述要求。见C.2节中的例子。推荐使用专门适用于复杂样本设计的软件包（见第二十一章）。

C. 回归分析和线性模型

1. 未纳入模型的设计变量效应和加权回归估计数

27. 在按照例程序将为简单随机样本开发的标准模型应用于复杂样本调查所得的数据时，已经普遍应用了回归分析和线性建模。如前所述，这种做法往往会导致错误的分析和结论。防止出错的一个关键性保护措施就是确定那些可能决定或影响样本设计的变量，以便将其纳入模型。但是，我们已经知道，即便确定了这些变量，从调查项目的观点来看也不宜将其列入模型。在本小节，我们将研究不把设计变量列入模型的传统估计量的影响，并调查修改这些估计量以将复杂设计考虑进去的可能性。为了便于讲解，让我们来考虑单一应变变量 Y （出于计数理由在本节用 X_1 表示）的情况，其中有关的模型有一个唯一的解释性变量(X_2)，还有一个唯一的设计变量(X_3)。因此可得相关的模型 $E(X_1)=\mu_1+\beta_{12}(X_2-\mu_2)$ ，式中， β_{12} 是相关的参数而不是整个模型，其中包括设计变量 X_3 。本小节使用的公式见附件。

28. 按照Nathan和Holt（1980年）所规定的很普通的条件， β_{12} 的标准普通最小二乘方估计量 $b_{12} = r_{12} / r_{23}$ 无论对 X_3 还是对样本 S 均可依（模型）有偏的条件而定，而且是无条件的偏差。有条件模型期望值及其无条件（联合模型和设计）期望值的表达式说明，一般来说， b_{12} 有渐近偏差，除非 X_2 与 X_3 之间的相关值 ρ_{23} 为0，或者 X_3 的简单样本方差是其真方差的一个无偏估计量。由此可见，上述第二个条件对于大量等概率样本（等概率抽选法）设计来说的确是渐近成立的，但是对非均等概率设计（比如非比例分层样本设计）却极少成立。

29. 可以根据正常情况下的最大似然估计量 $\hat{\beta}_{12}$ ，使用一个经过校正的有渐近偏差的估计量，而不用普通最小二乘方估计量。在Nathan和Holt（1980年）中给出了 b_{12} 的方差表达式和 $\hat{\beta}_{12}$ 的方差表达式。应该指出的是， b_{12} 方差的常用估计量 $v(b_{12})$ 不会接近于无偏差，即便在 b_{12} 是 $\hat{\beta}_{12}$ 的一个一致估计量的情况下也不可能。当 ε_i 值为非独立的、且在样本观测值当中恒等分布的时候，就可能发生这种情况。

30. 不论估计量 b_{12} 还是 $\hat{\beta}_{12}$ 都不取决于样本设计，不过它们的属性取决于样本设计。在没有关于整个人口的设计变量 X_3 的现成资料可用（以致不能用 S_3^2 进行估计）的情况下，或者在分析者想要确保偏离模型的稳健性时，有关样本设计的信息可能对改进这些估计量有用。可以根据对每一种未加权估计量成分的Horvitz-Thompson估计，使用样本加权估计量来改进这些估计量。我们可以

在 b_{12} 和 $\hat{\beta}_{12}$ 的表达式中用加权的样本矩取代未加权的样本矩，以获取加权估计量 b_{12}^* 和 $\hat{\beta}_{12}^*$ 。

31. 请注意，在 X_3, S_3^2 的总体方差未知的情况下，可以使用 b_{12}^* ，但在这种情况下不能使用 $\hat{\beta}_{12}^*$ 。不难看出，在相当普通的条件下，这两种估计量都是有限总体参数 B_{12} 的设计一致估计量。

32. Nathan和Holt（1980年）对 $N = 3850$ 农场的总体的这四种估计量的性能进行了经验比较，同时提供了如下相关数据：农田(X_1)；总面积（英亩）(X_2)；和上一年总值(X_3)。依据总值 X_3 对这些农场进行了分层，结果产生六个规模层次：563；584；854；998；696和155。采用了如下六种样本设计来选择规模 $n = 400$ 的样本（见表二十.1）：

表二十.1

采用各种调查设计，规模为3850个农场的总体的普通最小二乘方估计量的偏差和平均数二乘方以及无偏估计量方差

调查设计	$E(b_{12})-\beta_{12}$	$MSE(b_{12})$	$V(\hat{\beta}_{12})$	$V(b_{12}^*)$	$V(\hat{\beta}_{12}^*)$
A	0-000	0-000214	0-000197	0-000226	0-000197
B	0-000	0-000200	0-000198	0-000222	0-000196
C	0-031	0-001102	0-000160	0-000222	0-000196
D	0-027	0-000879	0-000163	0-000220	0-000195
E	0-042	0-001877	0-000152	0-000225	0-000196

资料来源：Nathan和Holt（1980年）；表1。

- (A) 简单随机抽样；
- (B) 按比例分层简单随机抽样；
- (C) 固定规模分层简单随机抽样；
- (D) 按比例层次分配高于 X_3 高值（25, 30, 60, 80, 130, 75）的分层简单随机抽样；
- (E) 按U-形层次分配（100, 80, 20, 20, 80, 100）的分层简单随机抽样。

33. 上述结果表明了非等概率抽选法设计（C, D, E）的 b_{12} 偏差；而其他估计量要么在设计上是一致的，要么在模型上是一致的（或二者兼而有之）。它们还表明了 $\hat{\beta}_{12}$ 相对于考虑中的所有设计加权估计量的优点。即使整个模型似乎对总体不成立，这一结论也成立。在 S_3^2 未知的情况下，效率较低，但仍然是是一致的， b_{12}^* 是一个合理的估计量。

34. 总之，如果数据基于不均等概率设计，就应考虑加权及未加权的最大似然估计量，而不是简单的普通最小二乘方估计量。未加权估计量似乎效率较高。可是在许多情况下，分析人员可能会得不到计算最大似然估计量所需的信息；而效率较低但是相容的样本加权估计量则是相宜的，而且的确在例行使用[见Korn和Graubard（1999年）]。

2. 检验设计对回归分析的影响

35. 许多分析人员宁愿使用能够从标准软件包获得的简单加权或未加权回归系数估计量，而不用C.1节所建议的经修改的估计量。我们发现，当设计为非告知性或设计效果可以忽略不计的时候，简单的普通最小二乘方估计量是一致的，但是如果情况不是如此，则与之对应的加权估计量是可取的。DuMouchel和Duncan(1983年)根据标准软件包建议了一种简便的检验方法，用以确定是否应当在以非类集样本数据为基础的时候使用权数。让我们来考虑一种有单一解释变量的一元情形。可以将其直接扩展到多元情形。设， $\tilde{\Delta} = \hat{b}_w - b$ 其目标是检验假设： $\Delta = E\{\tilde{\Delta}\} = 0$ 。DuMouchel和Duncan的论著说明，按照模型， $Y_i = \alpha + \beta X_i + \gamma Z_i + \varepsilon_i$ 其中 $Z_i = w_i X_i$ 而 $\varepsilon_i | X_i \sim N(0, \sigma^2)$ ， $\Delta = 0$ 的检验跟 $y = 0$ 的检验是一样的。这两位作者提供了一个数值范例来说明一种涉及到密歇根大学调研中心收入动态研究课题组提供的数据子集的多元情形。这个包括658个人的样本是按各种不同的概率选择的，由此确定的权数范围从1到83不等。用以解释教育程度的最终模型包含一个常数和17项解释性变量，诸如父母的教育程度、收入情况、年龄、种族、就业状况和相互影响等。归纳成如下方差(ANOVA)分析表：

表二十.2

ANOVA加权及未加权回归比较表

来源	df	平方和	均方	F	显著性水平
回归	17	730.6	43.0	17.35	<.0001
权数	18	43.3	2.5	.97	.494
误差	622	1542.2	2.5		
总计	657	2315.9			

36. 从总体来看，与Z_i相对应的这18个变量（17个解释性变量和一个常数乘以w_i）的F值为.97，显著性水平仅为.494。由此可见，一个未加权的回归证明是合理的，尽管会导致功效有所减弱。

37. 一般来说，分析人员对零假设 $\Delta = E\{\tilde{\Delta}\} = 0$ 的接受和拒绝给予了同样的关注：当它是虚的时候就接受它，当它实施的时候就拒绝它。因此，当显著性水平明显高于标准值.05的时候，他们可能决定进行加权分析（使用适当的软件）或者开发一个不太简约的模型。在上面的例证中，显著性水平非常接近0.5，就是说，权数可以忽略不计。可是在这个模型的较早期版本中，Z_i的显著性水平是.056，DuMouchel和Duncan在这个点上追加了一些相互作用项。最终结果就是表中所列的那些数据。

38. 上面讲的DuMouchel-Duncan检验假定ε_i值既独立又恒等分布。调查数据往往出自多级样本设计。如果同一个样本类集的ε_i观测值彼此相关，或者如果观测值具有某种未知的异方差性而不论何种设计，这个检验就不恰当。可是分析者也许觉得使用样本权数可能会给由此产生的估计量增加不必要的方差。

可以按照Fuller (1984年) 建议的方法进行沃尔德 (拟合直线法) 检验。实际当中, 这种检验需要使用像SAS/SURVEYREG这样的软件并且两次输入每个数据点: 一次是将样本权数定为1, 另一次是将样本权数定为实际权数。

39. 如果线性模型是正确的, 而且误差既独立又恒等分布, 但样本设计为告知性的话, Pfeffermann和Sverchkov (1999年) 针对这种情况推荐了一套可供选用的权数。上述检验可以用来评估其相对于样本权数的权数。关于抽样权数在建模调查数据中的作用, 详见Pfeffermann (1993年) 以及Korn和Graubard (1999年)。

3. 采用告知性样本设计的多层次模型

40. 近来, 使用多层次模型分析具有复杂等级结构的总体数据的情况增多了。比如, 在大多数住户调查中, 作为住户成员的个人成了调查单位, 并对住户个人之间的关系和住户之间的关系予以关注。对学校的学生和工商企业的雇员的调查也存在类似的等级结构。

41. 通过使用带有反映等级配置的误差结构的混合 (随机效应和固定效应) 模型, 可以很容易地将普通的单一层次线性模型加以扩展, 以便将等级结构考虑进去。例如, 通称为随机截距的模型 (在单一解释性变量情况下) 可以列式如下:

$$y_{ij} = \beta_{oi} + \beta x_{ij} + \varepsilon_{ij}; \quad \varepsilon_{ij} | x_{ij} \sim N(0, \sigma_{\varepsilon}^2); \quad (i=1, \dots, N; j=1, \dots, M_i)$$

式中, y_{ij} 是在二级单位*i* (比如住户) 中的一级单位*j* (比如个人) 的结果变量; x_{ij} 是一个已知变量; β 是一个未知参数。这里的截距 β_{oi} 是一个随机变量, 进一步建模为:

$$\beta_{oi} = \alpha + \gamma z_i + u_i; \quad u_i | z_i \sim N(0, \sigma_u^2); \quad (i=1, \dots, N)$$

式中, z_i 是一个已知的二级单位解释性变量; α 和 γ 是未知参数。

42. 按照简单随机抽样法, 这种类型的模型可以采用单层线性模型直接扩展理论来分析。遗憾的是, 没有提供模型参数估计数的封闭形式 (上面公式中的 $(\alpha, \beta, \gamma, \sigma_{\varepsilon}^2, \sigma_u^2)$)。而是采用了迭代广义最小二乘方法 (IGLS)。它所产生的估计数都集中于最大似然解决方案。可见, 在这种情况下不能使用将加权最小二乘方法调整成顾及样本设计的封闭式方法。为了获得相容的参数估计数, 开发了IGLS(PWIGLS)的样本加权版本, 它根据选择概率对一级和二级估计方程式进行加权 [见Pfeffermann等人 (1998年)]。

43. 最近, Pfeffermann, Moura和Silva (2001年) 提出一种依赖模型 (完全基于模型) 的方法, 用于说明告知性抽样的多层次分析。推荐的这种方法所依据的理念是: 推导出适合样本数据的等级模型作为总体模型和一阶样本列入概率的一个函数, 然后使用经典估计方法拟合样本模型。选择概率成为待建模并且由此强化估计量性能的附加结果变量。更详细的陈述就超出本章范围了,

但可参阅Pfeffermann, Moura和Silva(2001年)。严格效法1996年里约热内卢基本教育评价研究的一次模拟实验表明, 应用上述建议的方法获得的成果前景看好。

D. 分类数据分析

1. 修改卡方检验法以检验拟合优度和独立性

44. 旨在评估复杂样本设计对分类数据(即每个点都属于一定数量的类别或单元的数据)影响的初期尝试都涉及到对卡方检验的修改, 这些检验法通用于评估单一类变量与假设分布之间的拟合优度, 或者用以检验两类变量之间的独立性。虽然文献中推荐了若干经过修改适用于按比例分层样本随机抽样数据的卡方检验法, 但一般来说此种设计的实际效果还很小。因此, 在表二十.3[资料来源: Kish和Frankel(1974年)]中提供的一项关于出自以色列按比例分层样本的八个数据集的修正卡方统计量中, 没有一种最终迭代统计量与按照简单随机抽样(SRS)假设所得迭代统计量相差超过4%, 大多只相差不到1%。

表二十.3

三种迭代卡方检验与简单随机抽样检验的比率^a

数据集	层次数	行×列	样本规模	Nathan的三次检验					
				第一迭代			最后迭代		
				X ²	χ^2	G	X ²	χ^2	G
1	4	3×3	845	1.028	0.992	1.017	1.004	1.004	1.005
2	4	3×3	821	1.088	0.963	1.043	0.999	1.003	1.001
3	4	3×3	491	1.740	0.707	1.406	1.011	1.001	1.009
4	4	3×3	2528	1.095	0.959	1.049	1.003	1.005	1.003
5	6	2×4	500	1.079	0.967	1.040	1.004	1.003	1.003
6	3	2×2	120	1.013	0.967	1.009	1.008	0.969	1.007
7	5	2×2	269	1.076	0.989	1.043	1.011	1.015	1.011
8	2	2×4	81	1.368	0.889	1.186	1.029	1.037	1.029

资料来源: 节选自Nathan(1972年)的数据。

^a 根据以色列按比例分层样本编制的八个列联表分别是: 表1-4, 储蓄; 表5, 态度; 表6, 医院数据; 表7, 家禽药物; 表8, 观念实验。

45. 虽然一般情况下按比例分层简单随机抽样设计对分类数据分析的影响很小, 但在整群抽样中情况往往不是这样, Rao和Scott(1981年)的开创性文献就是例证。他们在检验拟合优度方面表明, 按照零假设, 通常的卡方统计量X²作为k-1独立 χ^2 加权了的随机变量呈渐近分布(亦即平方正常)。这些权重是矩阵D的特征值(见附件)。可以把这个矩阵视为对一元统计设计效果的一种自然多元扩展(见第六章和第七章)。它的特征值 λ_{0i}^2 叫作一般化的设计效果, 对于 \hat{p}_i 的设计效果 λ_i^2 (i类总体估计比率)的某些线性组合来说, 可以表现为设计效果。用标准X²统计量除以这些广义设计效果的平均估计数(用 $\bar{\lambda}^2$ 表示), 可以获得一个修正卡方统计量 X_c^2 。这种修改只需知道单元估计数的设计效果即可。虽然按照零假设 X_c^2 没有一个渐近的 χ_{k-1}^2 分布, 但它有同一个渐

近预期值 X_{k-1}^2 (亦即 $k-1$)，不过其方差比较大。实践证明，凭经验可以通过拿这个统计值与 X_{k-1}^2 的临界值作比较，用 X_c^2 来检验拟合优度。这可以从表二十.4[资料来源：Rao和Scott(1981年)]得到印证，该表列示了分别根据 X^2 和 X_c^2 对出自1971年大不列颠及北爱尔兰联合王国总住户调查的六项数据进行检验的真正规模。该调查采取了分层三段设计。

表二十.4

根据 X^2 和 X_c^2 对选自1971年大不列颠及北爱尔兰联合王国总住户调查的项目进行检验的估计渐近规模 (名义规模为.05)

变 量	k	m	$\hat{\chi}^2$	规模 (X^2)	规模 (X_c^2)
G1: 房舍建筑年龄	3	33.1	3.42	.41	.05
G2: 所有权类别	3	33.4	2.54	.37	.06
G3: 住宿类别	4	27.7	2.17	.30	.06
G4: 房间数	10	34.6	1.19	.14	.06
G5: 住户每周总收入	6	26.6	1.14	.10	.06
G6: 户主年龄	3	34.6	1.26	.10	.05

结果说明，采用标准卡方统计量 X^2 可能造成误导，而修正统计量 X_c^2 的效果却很好。

46. 在检验双向列联表中的独立性的时候，类似的结果也成立。对一个有 r 列 c 行的列联表来说，相关的零假设是 $H_0: h_{ij} = p_{ij} - p_{i+} p_{+j} = 0$ ($i = 1, \dots, r; j = 1, \dots, c$)，式中， p_{ij} 是第 (i, j) 单元中的人口比率； p_{i+} , p_{+j} 是边际总数。对于出自简单随机样本的数据来说，通常的卡方统计量 χ^2 跟零假设条件下自由度 $b = (r - 1)(c - 1)$ 的卡方统计量一样呈渐近分布。在样本设计复杂的情况下就不需要如此。实际上， χ^2 的渐近分布是 b 独立 χ^2 随机变量的一个加权和，就跟检验拟合优度时的情况一样。

47. 根据对 $\hat{h}_{ij} = \hat{p}_{ij} - \hat{p}_{i+} \hat{p}_{+j}$ [详见Rao和Scott(1981年)]估计量的完整方差-协方差矩阵的估计，可以构建一个广义的沃尔德统计量。幸运的是，一个只需要 \hat{h}_{ij} , $\hat{v}(\hat{h}_{ij})$ 的方差估计数的一阶校正似乎是适当的近似值。修正统计量被定义为 $X_{RC}^2 = X^2 / \hat{\delta}^2$ ，式中， $\hat{\delta}^2$ 是估计的设计效果 \hat{h}_{ij} 的一个加权平均数。在没有这些设计效果估计数可用的情况下 (在对发表的数据进行二次分析的时候往往有这种情况)，通过用 \hat{v} 取代 $\hat{\delta}^2$ 可以获得一个替代的修正统计量。 \hat{v} 是估计的单位比率设计效果 \hat{h}_{ij} 的一个加权平均数。这些近似值是否适当，在很大程度上取决于设计效果的相对方差。在这种相对方差很大的情况下，可使用一个二阶校正。

48. 表二十.5[资料来源：Rao和Scott(1981年)]根据大不列颠及北爱尔兰联合王国总住户调查的数据提供了15个双向列联表的经验结果。这些结果再次

说明：(a) 未校正的卡方统计量 χ^2 在许多情况下表现效果很差；(b) 根据 $\tilde{\delta}^2$ 校正的统计量 χ^2_{RC} 达到了近乎精确的名义规模；(3) 根据 $\hat{\tau}^2$ 校正的统计量过于保守。

表二十.5

根据, 和对选自1971年大不列颠及北爱尔兰联合王国总住户调查的变量交叉分组进行检验的估计渐近规模 (名义规模为.05)

交叉分组	r+c	$\tilde{\delta}^2$	$\hat{\tau}^2$	规模 (χ^2)	规模 (χ^2_{RC})	规模 ($\chi^2_{\hat{\tau}^2}$)
G1XG2	2X2	1.99	3.18	.16	.05	.01
G1XG3	2X3	1.97	2.36	.22	.05	.03
G1XG4	2X3	1.24	1.98	.09	.05	.01
G1XG5	2X6	.91	1.23	.04	.05	.02
G1XG6	2X3	.97	1.75	.05	.05	.01
G2XG3	2X3	1.94	2.49	.21	.05	.03
G2XG4	2X3	1.41	1.86	.12	.05	.02
G2XG5	2X6	1.02	1.18	.06	.05	.03
G2XG6	2X3	1.13	1.61	.08	.05	.02
G3XG4	3X3	1.26	1.72	.11	.05	.01
G3XG5	3X6	.93	1.14	.03	.05	.02
G3XG6	3X3	.96	1.51	.05	.05	.01
G4XG5	3X6	.94	1.05	.05	.05	.03
G4XG6	3X3	.93	1.21	.04	.05	.02
G5XG6	6X3	.85	.94	.03	.05	.04

2. 对数线性模型的归纳

49. Rao和Scott(1984年)已将上述双向表的结果归纳为对数线性模型, 用于分析多向表。在 $\sum_{i=1}^T r_i = 1$ (比如对 2×2 的表来说, $T = 4$) 的多向表中, 用 π 表示总体单位比率 π_i 的 T -向量。用 M_1 来表示饱和对数线性模型 (包括所有可能的相互作用)。让我们来考虑对一个缩减的嵌套子模型 M_2 是适当的这一假设进行的检验。设 $\hat{\pi}$ 为 M_1 之下 π 的虚假最大似然估计量。它被定义为 (根据人口数据已经获得的) 普查似然方程组样本估计数的解, 并且按照调查设计系以 π 的设计一致估计量为依据 (见附件)。同样, 设 $\hat{\pi}^*$ 为 M_2 之下 π 的虚假最大似然估计量。一般来说, 分别依据 $\hat{\pi}$ 和 $\hat{\pi}^*$, 用于检验 H_0 的标准个人卡方统计量在零假设情况下没有渐近卡方分布。这种情况跟双向表的情况很相似, 因为标准个人卡方统计量的渐近分布是权数为 δ_i^2 的 u 独立 χ^2 随机变量的一个加权和, 而这些权数又是广义设计效果矩阵的特征值 (详见附件)。

50. 为了将复杂设计考虑进去，拟采用修正卡方统计量 X^2_{ij} , X^2_{ij} 和 X^2_{ij} 。这里， $\bar{\delta}_i^2$ 是各特征值的平均估计数； λ^2 是设计效果 X_j 的平均估计数；而 λ_j^2 是单元设计效果的平均估计数（详见附件）。应当特别指出的是， λ^2 和 λ_j^2 不取决于零假设 H_0 ，但 $\bar{\delta}_i^2$ 取决于 H_0 。再者，当 M_1 是饱和模型的时候， λ_j^2 只需要知道单元设计效果即可； λ^2 也是如此。

51. 在承认 $\bar{\delta}_i^2$ 和 λ_j^2 显解的模型这一重要例证中，Rao和Scott(1984年)表明，在只知道单元设计效果及其边缘函数的设计效果的情况下，可以计算 $\bar{\delta}_i^2$ 。比如，对于在一个三向 $I \times J \times K$ 表—— $H_0: \pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k}$ ，其中 $\pi_{i++}, \pi_{+j+}, \pi_{++k}$ 是三向边缘函数——中具有完全独立性的假设来说， $\bar{\delta}_i^2$ 值可以作为这三向边缘函数设计效果估计数和单元设计效果估计数的一个函数明确计算出来。

52. 表二十.6[资料来源：Rao和Scott(1984年)]根据加拿大1978-1979年保健调查的一个 $2 \times 5 \times 4$ 表给出了修正的和未修正的这些统计量的相对性能。涉及的变量有性别 ($I = 2$)；吸毒状况 ($J = 5$)；和年龄组 ($K = 4$)。所检验的假设分别是：(a) 完全独立性（用 $\bar{1} \otimes \bar{2} \otimes \bar{3}$ 表示）；(b) 部分独立性（比如， $\pi_{ijk} = \pi_{i++}\pi_{+jk} \Leftrightarrow \bar{1} \otimes \bar{2}\bar{3}$ ）以及 $(\bar{2} \otimes \bar{1}\bar{3})$ 和 $(\bar{3} \otimes \bar{1}\bar{2})$ ；(c) 有条件独立性（比如， $\pi_{ijk} = \pi_{i+k}\pi_{+jk} / \pi_{++k} \Leftrightarrow \bar{1} \otimes \bar{2}\bar{3}$ ）以及 $(\bar{2} \otimes \bar{1}\bar{3})$ 和 $(\bar{3} \otimes \bar{1}\bar{2})$ 。采用了复杂设计，涉及到分层和多级抽样。另外还采用了后分层以改进估计数。

表二十.6

X^2 和修正统计量 X^2_{ij} , X^2_{ij} , X^2_{ij} ：2×5×4表的估计渐近显著性水平（SL）及名义显著性水平 $\alpha = 0.05$

	假 设						
	(a)	(b)			(c)		
	$\bar{1} \otimes \bar{2} \otimes \bar{3}$	$\bar{1} \otimes \bar{2}\bar{3}$	$(\bar{2} \otimes \bar{1}\bar{3})$	$(\bar{3} \otimes \bar{1}\bar{2})$	$\bar{1} \otimes \bar{2}\bar{3}$	$(\bar{2} \otimes \bar{1}\bar{3})$	$(\bar{3} \otimes \bar{1}\bar{2})$
SL(X^2)	0.72	0.33	0.76	0.72	0.43	0.30	0.78
SL(X^2_{ij})	0.16	0.11	0.14	0.13	0.095	0.11	0.12
SL(X^2_{ij})	0.34	0.056	0.39	0.32	0.098	0.06	0.39
SL(X^2_{ij})	0.34	0.054	0.39	0.32	0.097	0.06	0.39
$\bar{\delta}_i^2$	2.09	1.40	2.25	2.09	1.63	1.39	2.31
C.V. ($\bar{\delta}_i^2$)	1.54	1.02	1.37	1.27	0.86	1.05	1.11

53. 上述比较显示了实际显著性水平（SL）与期望的名义水平 $\alpha = 0.05$ 之间的关系。结果再次表明，未经修正的统计量的SL水平高得不可接受。经过修正的统计量 X^2_{ij} 和 X^2_{ij} 不取决于假设条件，表现类似于SL值，范围介于0.06到0.39，太高了。修改基于边际设计效果和单位设计效果 X^2_{ij} ，其性能比较稳定，SL值介于0.095到0.16，都高于名义水平，大概是因为 $\bar{\delta}_i^2$ 的变异系数（CV）很大的缘故。

54. 总之，对于分类数据分析中的标准卡方检验统计量来说，现在已有现成的校正方法。在特定的类集样本中，这些校正对确保有效的分析往往是必要的，并可以较为容易地使用估计的边际和单元设计效果加以应用。关于处理复杂样本设计对卡方检验和逻辑回归的影响所可用的软件，详见第二十一章。

E. 总结与结论

55. 我们在本章说明了通用复杂样本设计对调查数据分析的影响的评估方法。这份材料的本意主要是对有关问题作介绍性解释，而非规范性的阐述。评估和处理样本设计对分析的影响可能难以进行，并不是规定一些简便“用法”规则就能奏效的。如前所述，不同的问题可能有不同的（或好几种不同的）解决办法。这些事情在很大程度上取决于设定的模型及其基本假设的有效性，取决于样本设计的方方面面（比如不均等选择概率、类集，等等），还取决于所设想的分析类型。对有关模型与样本设计变量之间关系的了解是必要的。遗憾的是，这方面的资料并不总是随手可得。这就是说，可能需要代之以使用假定条件和近似值。

56. 任何分析的第一步也是根本的一步就是要正确地设定基础模型。这是主题调查项目分析员的责任，但模型的最终确定可以而且应该以适当的统计方法为基础。可以使用标准的图示和描述方法进行为确定适当的模型所需要的初步探索性分析，而无需将样本设计效果考虑进去。

57. 初步工作模型一经设定，就需要确定设计是否对分析有困惑性的影响。譬如，可以通过比较线性回归系数的加权估计数和未加权估计数来检验设计影响（见C.2节）。如果需要把复杂设计的因素纳入分析，那就必须选择适当的纳入方法。分类法只是对与样本设计有关的模型添加变量。

58. 然而在许多情况下，并不能通过修正模型有效而充分地反映样本设计的影响。在这种情况下若采用分类模型法，可以使用本章推荐的两种基本方法来处理潜在的样本设计影响。一种方法就是修改经典分析工具以便将设计考虑进去。这种方法最适合处理可以依据广义设计效果修改标准卡方统计量的分类数据分析。第二种方法就是开发适当定义的设计专用分析工具。这里推荐了样本加权估计数和大型样本沃尔德统计量。在使用沃尔德统计量之前需要一个可靠的协方差矩阵估计量。实际当中并不总是能够获得这种估计量。

59. 在处理复杂样本设计对分析的影响方面，就此问题开展的大量研究工作产生了各种实用方法，其中一些方法已在本章作了介绍。目前正在开展进一步的研究，现有的许多方法已经纳入了新的在用软件。遗憾的是，由于问题复杂，不大可能在将来开发出任何统一的全面方法。在应用现有的各种方法和软件时，应当慎之又慎。有关的应用不但需要掌握基本理论知识，而且需要具有实际模型构建方面的透彻认识和经验。

附 件

正式定义和技术成果

回归方法(见B.2和B.3节)

- 标准线性回归模型: $Y_i = \alpha + \beta X_i + \varepsilon_i$, 其中 $\varepsilon_i | X_i \sim N(0, \sigma^2)$

- 标准总体回归系数: $B = \frac{\sum_{i=1}^N Y_i (X_i - \bar{X})}{\sum_{i=1}^N (X_i - \bar{X})^2}$

- β 的普通最小二乘方(OLS)估计量: $b = \frac{\sum_{i=1}^n y_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$

- b 的无偏方差估计量: $v(b) = \frac{s^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$,

式中, s^2 是根据估计的回归剩余方差计算的 σ^2 的无偏估计量。

- 一般线性(异方差性的)回归模型: $Y_i = \alpha + \beta X_i + \varepsilon_i$,

式中, $\varepsilon_i | X_i \sim N(0, \sigma_i^2)$

- 加权总体回归系数: $B^* = \frac{\sum_{i=1}^N Y_i (X_i - \bar{X}_\sigma) / \sigma_i^2}{\sum_{i=1}^N (X_i - \bar{X}_\sigma)^2 / \sigma_i^2}$,

式中, $\bar{X}_\sigma = \frac{\sum_{i=1}^N X_i / \sigma_i^2}{\sum_{i=1}^N 1 / \sigma_i^2}$

- β 的广义最小二乘方(GLS)估计量: $b_G = \frac{\sum_{i=1}^n y_i (x_i - \bar{x}_\sigma) / \sigma_i^2}{\sum_{i=1}^n (x_i - \bar{x}_\sigma)^2 / \sigma_i^2}$

式中, $\bar{x}_\sigma = \frac{\sum_{i=1}^n x_i / \sigma_i^2}{\sum_{i=1}^n 1 / \sigma_i^2}$

- 广义最小二乘方估计量方差: $v(b_G) = \frac{1}{\sum_{i=1}^n (x_i - \bar{x}_\sigma)^2 / \sigma_i^2}$

- 设计加权(Horvitz-Thompson)估计量: $b^* = \frac{\sum_{i=1}^n w_i y_i (x_i - \bar{x}^*)}{\sum_{i=1}^n w_i (x_i - \bar{x}^*)^2}$,

式中, $w_i = \frac{1/\pi_i}{\sum_{i=1}^n 1/\pi_i}$, π_i 是列入概率, 而 $\bar{x}^* = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$

排除设计变量的影响（见C.1节）

- 相关的模型： $E(X_1) = \mu_1 + \beta_{12}(X_2 - \mu_2)$ ，式中， β_{12} 是相关的参数。
- 带有设计变量 X_3 的全模型： $E(X_1) = \mu_1 + \beta_{12,3}(X_2 - \mu_2) + \beta_{13,2}(X_3 - \mu_3)$
- 符号表示法：
 - 多变量分析的常用符号表示法，比如 $\beta_{12,3}$ 表示设定 X_3 ，关于 X_2 的有条件回归系数 X_1 ；

$$\bullet X_i \text{ 的第一和第二总体矩: } \bar{X}_i = \frac{1}{N} \sum_{j=1}^N X_{ij}; \quad S_i^2 = \frac{1}{N-1} \sum_{j=1}^N (X_{ij} - \bar{X}_i)^2;$$

$$\bullet S_{ik} = \frac{1}{N-1} \sum_{j=1}^N (X_{ij} - \bar{X}_i)(X_{kj} - \bar{X}_k);$$

$$\bullet \text{ 样本矩: } \bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}; \quad s_{ik} = \frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_k); \quad s_i^2 = s_{ii},$$

式中，我们假定采用取决于 X_3 的任何设计所选定的一个固定规模为 n 的样本 S 。

- β_{12} 的标准普通最小二乘方估计量： $b_{12} = \frac{s_{12}}{s_2^2}$
- b_{12} 的渐近模型有条件期望值：

$$E_M(b_{12} | X_3, \mathbf{S}) = \frac{\beta_{12} + \beta_{13}\beta_{23}(s_3^2/\sigma_3^2 - 1)}{1 + \rho_{23}^2(s_3^2/\sigma_3^2 - 1)} + O(n^{-1})$$

- 无条件（联合模型和设计）期望值：

$$E_M(b_{12}) = \beta_{12} + \frac{\sigma_1}{\sigma_2} \frac{\rho_{132}\rho_{23} \left[(1 - \rho_{12}^2)(1 - \rho_{23}^2) \right]^{\frac{1}{2}} (Q - 1)}{1 + \rho_{23}^2(Q - 1)} + O(n^{-1}), \text{ 式中,}$$

$$Q = E\left(\frac{s_3^2}{\sigma_3^2}\right) / \sigma_3^2$$

- 普通最小二乘方估计量 b_{12} 有渐近偏差，甚至是无条件的偏差，除非 $\rho_{23} = 0$ 或者

$$E(s_3^2) / \sigma_3^2, \text{ 亦即 } Q = 1$$

- 校正的渐近无偏估计量（正常情况下的最大似然估计量(MLE)）：

$$\hat{\beta}_{12} = \frac{s_{12} + (s_{13}s_{23}/s_3^2)(S_3^2/s_3^2 - 1)}{s_2^2 + (s_{23}^2/s_3^2)(S_3^2/s_3^2 - 1)}$$

- 加权估计量： $b_{12}^* = \frac{s_{12}^*}{s_2^{*2}}; \hat{\beta}_{12}^* = \frac{s_{12}^* + (s_{13}^*s_{23}^*/s_3^{*2})(S_3^{*2}/s_3^{*2} - 1)}{s_2^{*2} + (s_{23}^{*2}/s_3^{*2})(S_3^{*2}/s_3^{*2} - 1)}$,

$$\text{式中, } \bar{x}_i^* = \sum_{j=1}^n \frac{x_{ij}}{N\pi_j}; \quad s_{ik}^* = \sum_{j=1}^n \frac{x_{ij}x_{kj}}{N\pi_j} - \bar{x}_i^*\bar{x}_k^*; \quad s_i^{*2} = s_{ii}^*; \text{ 和}$$

$\pi_j = P(j \in S | X_{3j})$ 是样本列入概率。请注意，在分层简单随机抽样中 $\sum_{j=1}^n \frac{1}{N\pi_j} = 1$ ，这是此处假定的设计。对于较一般的设计，可以用 $1/w_j$ 来取代 $N\pi_j$ ，其中

$$w_j = \frac{1/\pi_j}{\sum_{k=1}^n 1/\pi_k}$$

- 结果: $E_y(k_{12}^*) = E_y(\hat{\beta}_{12}^*) = B_{12} + O(n^{-1})$, 式中, E_y 代表设计预期值 (亦即对重复样本选择的预期值)。

分类数据分析 (见D节)

• 检验拟合优度:

- 假定已知多项式分布, 概率为 $p_0 = (p_{0,1}, \dots, p_{0,k-1})$, 式中, k 是类别数, 而

$$\sum_{i=1}^k p_{0i} = 1.$$

- 按照 H_0 , 卡方统计量 $X^2 = n \sum_{i=1}^k \frac{(\hat{p}_i - p_{0i})^2}{p_{0i}}$ (式中 \hat{p}_i 是 p_{0i} 的样本估计数) 呈渐近分布:

$$X^2 = \sum_{i=1}^{k-1} \lambda_i^2 Z_i^2; Z_i \sim N(0,1), \text{ 式中, } \lambda_i^2 \text{ 是 } D = P_0^{-1} V_0 \text{ 的特征值; } P_0 \text{ 是 SRS 零假设条件下样本估计数的方差矩阵; 而 } V_0 \text{ 是其在 } H_0 \text{ 项下的真方差矩阵.}$$

- 修正卡方统计量: $X_C^2 = X^2 / \tilde{\lambda}^2$; $\tilde{\lambda}^2 = \sum_{i=1}^{k-1} (1 - \hat{p}_i) \tilde{d}_i^2 / (k-1)$, 式中, \tilde{d}_i^2 是 \hat{p}_i 的设计效果 d_i^2 的估计数。

• 双向列联表中的独立性检验:

- 相关的假设: $H_0: h_{ij} = p_{ij} - p_{i+} p_{+j} = 0$ ($i = 1, \dots, r; j = 1, \dots, c$), 式中, p_{ij} 是第 (i,j) 单元中的总体比率, 而 $p_{i+} = \sum_{j=1}^c p_{ij}$, $p_{+j} = \sum_{i=1}^r p_{ij}$ 是边际总数。

- 通常的卡方统计量: $X_I^2 = n \sum_{j=1}^c \sum_{i=1}^r \frac{(\hat{p}_{ij} - \hat{p}_{i+} \hat{p}_{+j})^2}{\hat{p}_{i+} \hat{p}_{+j}}$, 式中, \hat{p}_{ij} 代表 p_{ij} 的样本估计量。

- X_I^2 是 b 独立 χ_1^2 随机变量加权之和的渐近分布。

- 一阶校正: $X_{IC}^2 = X_I^2 / \tilde{\delta}^2$, 式中: $\tilde{\delta}^2 = \sum_{i=1}^r \sum_{j=1}^c (1 - \hat{p}_{i+})(1 - \hat{p}_{+j}) \tilde{\delta}_{ij}^2 / b$, 而

$$\tilde{\delta}_{ij}^2 = n \frac{\tilde{v}(\hat{h}_{ij})}{\hat{p}_{i+} \hat{p}_{+j} (1 - \hat{p}_{i+})(1 - \hat{p}_{+j})} \text{ 是 } \tilde{h}_{ij} \text{ 的估计设计效果.}$$

- 通过用 $X^2 = \frac{1}{rc-1} \sum_{i=1}^r \sum_{j=1}^c (n - \bar{p}_{ij}) \tilde{d}_{ij}^2$ 取代 $\tilde{\delta}^2$ 所得的备选修改。

• 对数线性模型的归纳

- 对数线性模型: $\mu = \tilde{\omega}(\theta) \mathbf{1} + X\theta$, 式中, π 是总体单元比率 π_i 的 T -向量; μ 是对数概率 $\mu_i = \ln \pi_i$ 的 T -向量; X 是一个已知的满秩 $T \times r$ 矩阵, ; $X' \mathbf{1} = 0$; θ 是参数的一个 r -向量, 而 $\tilde{\omega}(\theta) = \ln\{1/[1' \exp(X\theta)]\}$ 是一个正常化因子。

- 相关的假设: $H_0: \theta_2 = 0$, 式中, $X = (X_1, X_2)$ 而 $\theta = (\theta_1, \theta_2)$, X_1 是 $T \times s$, 而 X_2 是 $T \times u$; θ_1 是 $s \times 1$, 而 θ_2 是 $u \times 1$ 。

- 设 $\tilde{\pi}$ 是按照 M_1 设计, π 的虚假最大似然估计量, 这是虚假似然方程式 $X' \tilde{\pi} = X' \hat{p}$ 的解, 式中, \hat{p} 是按调查设计 π 的 (设计) 一致估计量。同样, 设 $\tilde{\pi}_2$ 是按照 M_2 设计, π 的虚假最大似然估计量。

- 用于检验 H_0 的标准个人卡方统计量: $X^2 = n \sum_t \frac{(\tilde{\pi}_t - \tilde{\pi}_t)^2}{\tilde{\pi}_t}$ 。

- X^2 的渐近分布: $X^2 = \sum_{j=1}^u \delta_j^2 Z_j^2$; $Z_j \underset{\text{i.i.d.}}{\sim} N(0,1)$, 式中, δ_j^2 是一个广义设计效果矩阵的特征值。

- 修正的卡方统计量： $\chi^2_{ij}{}^{(2)}$ 、 $\chi^2_{ij}{}^{(3)}$ 和 $\chi^2_{ij}{}^{(4)}$ ，式中：

$\hat{\delta}^2$ 是特征值 $\delta^2 = \frac{1}{u} \sum \delta_i^2$ 的平均估计数。

$\hat{\lambda}^2$ 是设计效果 $X\tilde{p}$ 的平均估计数。

$\hat{d}^2 = \frac{1}{T} \sum \hat{d}_i^2$ ，式中， $\hat{d}_i^2 = n \frac{\hat{v}(\hat{p}_i)}{\hat{\pi}_i(1-\hat{\pi}_i)}$ 是单位 t 的估计设计效果。

- 范例：对于一个三向 $I \times J \times K$ 表中的完全独立性假设， $H_0: \pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k}$ ，式中， $\pi_{i++}\pi_{+j+}\pi_{++k}$ 是三向边缘函数； $\hat{\delta}^2$ 值由下式给出：

$$\hat{\delta}^2 = \frac{\sum_i \sum_j \sum_k (1 - \hat{\pi}_{i++} \hat{\pi}_{+j+} \hat{\pi}_{++k}) \hat{d}_{ijk}^2 - \sum_i (1 - \hat{\pi}_{i++}) \hat{d}_i^2(r) - \sum_j (1 - \hat{\pi}_{+j+}) \hat{d}_j^2(c) - \sum_k (1 - \hat{\pi}_{++k}) \hat{d}_k^2(l)}{IJK - I - J - K + 2}$$

式中， $\hat{d}_i^2(r)$ 、 $\hat{d}_j^2(c)$ 和 $\hat{d}_k^2(l)$ 是三向边缘函数设计效果的估计数，而 \hat{d}_{ijk}^2 是单位设计效果的估计数。

参考资料

- An, A. 和 D. Watts (2001年)。《分析样本调查数据的新SAS程序》。《SAS全球用户大会文件，第23号》。北卡罗来纳，卡里，SAS软件研究所。请登陆：<http://www2.sas.com/proceedings/sugi23/stats/p247.pdf> (2004年7月2日起可访问)。
- Berthoud, R. 和 J. Gershuny等 (2000年)。《英国家庭中七年的生活：来自英国住户小组调查的社会动态证据》。联合王国，布里斯托：政策出版社。
- Brewer, K. R. W. 和 R. W. Mellor (1973年)。“样本结构对分析调查的影响”。《澳大利亚统计杂志》，第15期，第145-152页。
- Chambers, R. L. 和 C. J. Skinner等 (2003年)。《调查数据分析》。纽约：威利国际出版公司。
- Cochran, W. G. (1977年)。《抽样方法》，第三版。纽约：威利国际出版公司。
- DuMouchel, W. H. 和 G. J. Duncan (1983年)。“分层样本多级回归分析中使用调查权数”。《美国统计学会志》，第78期，第535-543页。
- Duncan, G. J. 和 G. Kalton (1987年)。“不同时期的调查设计和分析问题”。《国际统计评论》，第55期，第97-117页。
- Feder, M.、G. Nathan 和 D. Pfeffermann (2000年)。“根据不同时期的随机效果复杂调查纵向数据的时序多层次建模”。《调查方法》，第26期，53-65页。
- Fuller, W. A. (1984年)。“复杂调查设计的最小二乘方与相关分析”。《调查方法》，第10期，第97-118页。

- Hansen, M. H.、W. G. Madow和B. J. Tepping (1983年)。“对样本调查中依赖模型和概率-抽样推论的评估”。《美国统计学会志》，第78期，第776-793页。
- Kish, L. 和M. Frankel (1974年)。“根据复杂调查的推论”。《皇家统计学会志》，B辑，第36期，第1-37页。
- Korn, E. L. 和B. I. Graubard (1999年)。《健康调查分析》。纽约和联合王国奇切斯特：威利国际出版公司。
- Manski, C. F. 和S. R. Lerman (1977)。“针对基于样本选择机率的评估”。《计量经济学》，第45期，第1977-1988页。
- Nathan, G. (1972年)。“分层样本中相依表独立性测试的渐进幂级数”。《美国统计学会志》，第67期，917-920页。
- _____和D. Holt (1980年)。“调查设计对回归分析的影响”。《皇家统计学会志》，B辑，第43期，377-386页。
- _____和T. M. F. Smith (1989年)。“抽选对回归分析的影响”。《复杂调查的分析》，C. J. Skinner, D. Holt和T. M. F. Smith等。联合王国，奇切斯特：威利国际出版公司，第227-250页。
- Pfeffermann, D. (1993年)。“模拟调查数据时抽样权数的作用”。《国际统计评论》，第61期，第317-337页。
- _____和Sverchkov, M. (1999年)。“适合调查数据的参数和半参数回归模型估计”。《Sankhy, B辑》，第61期，第1部分，第166-186页。
- Pfeffermann, D.、F. Moura和N. S. Silva (2001年)。“信息概率调查下的多层次模型”。《首尔，国际统计学会第五十三届会议的特邀报告》。
- Pfeffermann, D. 和其他人 (1998年)。“多水平模型中不平等概率抽选的加权情况”。《皇家统计学会志》，B辑，第60期，第23-40页。
- Rao, J. N. K. 和A. J. Scott (1981年)。“复杂样本调查中分类数据的分析：关于双向表中适合度和独立性的卡方测验”。《美国统计学会志》，第76期，第221-230页。
- _____ (1984年)。“根据调查数据估计的细胞比例对多向相依表进行的卡方测验”。《统计学年报》，第12期，第46-60页。
- Särndal, C-E.、B. Swensson和J. Wretman (1992年)。《调查取样的模型帮助》。纽约：施普林格出版社。
- Scott, A. J. 和C. J. Wild (1986年)。“适合个例对照中逻辑模型或基于选择的抽样”。《皇家统计学会志》，B辑，第48期，第170-182页。
- Skinner, C. J.、D. Holt和T. M. F. Smith等 (1989年)。《复杂调查分析》。联合王国，奇切斯特：威利国际出版公司。
- Valliant, R.、A. H. Dorfman和R. M. Royall (2000年)。《有限人口抽样和推论：预测方法》。联合王国，奇切斯特以及纽约：威利国际出版公司。

第二十一章

调查数据的抽样误差估计*

Donna Brogan

埃默里大学

美国佐治亚州亚特兰大

* 本章包括一个附件（只有英文），里面包含利用五个统计软件包对布隆迪免疫调查数据进行的说明分析和比较分析。CD-ROM的内容包括每个软件包的程序编码和输出，该内容可以直接从联合国统计司网站(<http://unstats.un.org/unsd/hhsurveys/>)下载，也可以向联合国统计司(statistics@un.org)索取。

摘 要

复杂样本调查设计与简单随机抽样不同，包括不等概率抽样、多阶段抽样和分层等方面。对总体参数的无偏误（或近乎无偏误）估计需要进行加权分析。估计量的方差估计取决于抽样计划细节，需要近似法，一般来说，这些方法是泰勒级数线性化方法或重复方法。

一般情况下，标准统计软件包不能用于分析样本调查数据，因为这些软件包主要进行元素的简单随机抽样。这些软件包使总体参数的点估计（在无加权分析中）产生偏误和/或低估了点估计的标准误差。利用标准软件包使用抽样加权变量产生总体参数的适当点估计。但是，估计的标准误差通常仍然不正确，因为方差估计程序一般不考虑抽样计划的类集和/或分层。

本章概述能对样本调查数据进行分析的八个软件包，包括大约的费用、方差估计方法、分析选项、用户界面以及优缺点。其中四个软件包是免费的，因此，软件购置预算有限的发展中国家可能会关注这四个软件。

布隆迪的一个复杂样本调查数据集显示，用标准统计软件进行的分析是不正确的。经评价的软件包有八个，根据布隆迪调查对其中的五个（STATA、SAS、SUDAAN、WesVar和Epi-Info）进行的注解描述性分析显示了如何使用这些软件包。最后，为了实现共同的分析目标，将五个软件包的数值结果与布隆迪调查数据进行比较。不论采用的是泰勒级数线性化方法还是平衡多次重复（BRR）方法，所有五个软件包得出的方差估计结果均相同。

关键词：泰勒级数线性化、重复、最终类集、方差估计、复杂样本调查、软件包。

A. 调查样本设计

1. 正如本出版物的许多章所论述的那样，住户调查样本设计是复杂的，一般都涉及到分层多阶段抽样。使用复杂样本设计的结果是，在分析住户调查数据时，不能不加选择地应用标准统计方法和软件。尤其是，调查中答复单位被赋予权数，这些权数补偿不等抽选概率、单位不答复和不覆盖，还可以使某些变量的加权调查分布符合这些变量的已知分布。应当在调查分析中使用这些权数，而且，调查估计的抽样误差计算应考虑这样一个事实，即调查样本是采用复杂样本设计抽选的。幸运的是，现在有许多针对调查分析的专业软件包，可以正确计算出复杂样本设计中加权调查估计的抽样误差。本章对一些这样的软件包进行了说明和评价。

2. 作为对调查分析程序包进行讨论的准备阶段，下面的两节利用复杂样本设计评价加权分析问题和方差估计方法。接下来的各节对用于复杂样本调查数据导出的估计量的八个方差估计软件包进行比较，并利用布隆迪样本调查的数据举例说明如何使用其中的五个软件包。本章的结尾给出了一些结论和建议。CD-ROM包含与本出版物有关的附件，该附件为利用选出的五个软件包进行的三项分析提供了注解数据分析。

B. 复杂样本调查数据的数据分析问题

1. 加权分析

3. 在许多住户调查中，分析单位（住户或个人）都是使用不等概率选择的，在分析中，应通过加权补偿这些不等抽选概率。另外，即使使用等概率选择这些单位，往往也需要加权，用来弥补单位不答复和基准的确定，如事后分层（见第十九章）。这些加权应该用于估计总体参数的分析。未加权的估计量（不建议）对总体参数来说可能严重偏误，这当然取决于具体的调查。数据集中给定答复样本元素R的样本加权变量值用 $WTVAR$ 表示，可以解释为用R表示的总体中的元素数。数据集中所有元素R的 $WTVAR$ 值总和估计总体中的元素数量。

4. 有时候，通过用 $WTVAR$ 乘以（R总数）/（所有R的 $WTVAR$ 值的和），就可以给抽样加权变量 $WTVAR$ “加以范数”。所有R的“范数加权变量” $WTNORM$ 值的总和就是分析的样本规模（R的数量）。无论是使用样本加权变量 $WTVAR$ 还是加上范数的加权变量 $WTNORM$ 得到平均总体参数（比如均值或比例）的点估计，这一点并不重要：两种方法都会得到相同的计算值。但是，加上范数的加权变量 $WTNORM$ 不能直接用于估计总体参数总数，比如，不能用于估计总体中营养不良儿童的总人数。

2. 方差估计概述

5. 方差估计是重要的，因为它显示估计量的精确度，其结果是得到关于总体参数的置信区间和检验假设。基于复杂样本调查数据的估计量方差估计一定要确认以下因素：(a) 大多数估计量是非线性的（线性估计量的比率是常见

的)；(b) 估计量是加权的；(c) 在初级抽样之前，抽样计划将基本上已经使用了分层（而且，在接下来的抽样阶段也是如此）；(d) 由于多阶段整群抽样，样本中的元素在统计上一般都是独立的。在几乎所有的情况下，不可能获得估计方差的封闭式代数式。因此，复杂样本调查数据的方差估计的研究文献包括了若干种相似的方法，样本调查数据分析人员可以从中选择。

6. 接近估计方差最常采用的两种方法是泰勒级数线性化（TSL）（Wolter, 1985年；Shah, 1998年）和重复方法（Wolter, 1985年；Rust和Rao, 1996年）。这些方法将在c节进行更详细的讨论。用来分析样本调查数据的大多数软件包只使用其中的一种方法。对于是样本数据光滑函数的估计量（例如，总数、均值、比例、均值和比例之间的差等），以上两种方法给出可比方差估计值，但并不明显主张使用其中的一种方法。对于样本数据的非光滑函数的估计量（例如中位数），似乎主张使用特别的重复程序，即平衡的多次重复，而不主张使用泰勒级数线性化方法或另一种重复方法——刀切法（Korn和Graubard, 1999年）。关于方差技术的比较有一个重要文献，里面包括偏向两种方法中其中一种方法的特殊例子（[见Korn和Graubard（1999年）以及他们的许多参考资料，又见Kish和Frankel（1974年）]）。

3. 没有替代抽样的有限总体修正（FPC）系数

7. 为简便起见，开始就考虑根据等概率从人口规模 N 中选出样本规模 n 的人口均值的估计，然后比较两种样本设计。在一种设计中，各元素根据简单随机抽样选出，也就是说，在没有替代的情况下选出这些元素。在另一种设计中，这些元素根据无限制抽样选出，也就是说，有替代抽样（也称有替代的简单随机抽样）。这两种设计的样本均值方差的不同之处是，有限总体修正（ fpc ）项包括在简单随机样本的方差中，而不包含在无限制样本的方差中（见第六章）。 fpc 项就是 $(1 - f)$ ，其中 $f = n/N$ ，就是抽样比。 fpc 项一定能够大于1.0，反映了没有替代抽样导致的方差减少。如果抽样比 f 小， fpc 就接近1.0，对方差的影响最小，在方差估计中可以安全地忽略不计。换句话说，没有替代的样本可以和有替代的样本受到一样的对待。小的抽样比一般被看作是5%或10%。另一方面，如果 f 大，在无替代选择样本时忽略 fpc 项就会导致过高的估计方差。在具有不同分层的不同抽样比的分层随机样本设计中， fpc 项可以小到某些分层中忽略不计的程度，但其他分层中不能忽略。

8. 大多数住户调查的基础是适用于很大总体的复杂样本设计。初级抽样单位一般采用概率与规模成比例（PPS）方法抽选，没有替代抽样，这就使“抽样比”的概念更加复杂。但是，初级抽样单位的数量往往很大，每个分层中的初级抽样单位抽样比都较小，对第一阶段的所有 fpc 项来说，其值都接近1.0。这样，复杂样本调查数据分析的常见近似值就是每个分层中有替代的初级抽样单位抽样。如果该近似值是在带有大的初级抽样比的某些分层中计算出的，那么在某种程度上就会过高估计。鉴于没有近似值的方差估计的复杂性，这样的过高估计往往是可以接受的。请注意，如果任何分层的抽样第一阶段在有替代的情况下进行抽样，该分层中就没有近似值。

4. 假层和假初级抽样单位

9. 为了进行方差估计，有时候由于在抽样计划中确实使用了层和初级抽样单位，便不能确认层和初级抽样单位。为估计方差，可能会对层和初级抽样单位的定义进行修正，使得实际使用的抽样计划适合某一个软件包中某个抽样计划选择。在进行这样的修正时，为方差估计而新定义的和初级抽样单位变量有时候被称作假层和假初级抽样单位。

10. 在初级抽样之前，定义了很大数量的层，这时就出现了一个常见的例子，每个分层中只有一个抽选的（抽样的）初级抽样单位。如果每个分层只有一个初级抽样单位，那么方差估计是不可能的，因为在初级抽样单位之间，层内的变异性是不可能估计的。在这种情况下，两个层就折迭或合成一个假层，在该假层中给出两个样本初级抽样单位。折迭层是有策略地进行的，而不是任意进行的，是在了解初级抽样单位分层变量和初级抽样单位抽样方法的基础上进行的（Kish, 1965年）。

11. 隐含分层又产生了另一个例子。比如，一个国家可以根据南北分层，根据村庄定义初级抽样单位。在每个层中，初级抽样单位按照地理邻近度排列，接下来使用系统的PPES（概率与估计规模成比例）在每个层中选择许多（比如30个）初级抽样单位概率样本（Kish, 1965年）。层内总体初级抽样单位的地理顺序加上系统抽样，就导致在每一个南北层中出现隐含的村庄（初级抽样单位）地理分层。为了认识方差估计方面的隐含分层，抽样计划一般都将被描述为包括15个北方假层和15个南方假层，每个分层有两个抽中的初级抽样单位或假初级抽样单位。从北方抽样标架中抽取的前两个初级抽样单位将进入第一个假层，接下来的两个初级抽样单位进入第二个假层，等等。

12. Korn和Graubard（1999年）给出了另外几个例子，这些例子中，为方差估计之目的，形成了假层和假初级抽样单位，旨在减少重复的数量和减轻计算负担。另外，WesVar用户指南（2002年）附件D在描述基于重复法的方差估计软件的各种抽样计划方面，提供了指南和范例。

5. 描述许多复杂抽样计划的一种常见近似（WR）

13. 复杂样本调查一般使用多阶段整群抽样。另外，在初级抽样之前常常对总体初级抽样单位进行分层。而且，第二阶段和以后各阶段单位（样本初级抽样单位之内）的分层可能出现在这些阶段的抽样之前。但是，这些复杂设计常使用的方差估计近似方法不需要考虑抽样和分层的所有阶段。后来各个阶段的复杂抽样自动而适当地包含在上述讨论的抽样第一阶段“有替代”的近似之下。实际上，在没有替代近似的第一阶段这种情况下，没有几个抽样调查软件包能把方差估计中抽样的各个阶段单独包括在内。

14. 使用最终类集方差估计（UCVE）进行复杂设计是很常见的事，这种做法首先由Hansen、Hurwitz和Madow（1953年）提出，Wolter（1985年）也就此进行过讨论。最终类集方差估计可以使用泰勒级数线性化或重复法实施。最终类集方差估计方法处理初级抽样单位时，如同这些初级抽样单位在第一阶段

分层中进行过有替代品抽样，然后，每个R（数据集中样本答复元素）只需经过其选择来源的第一阶段分层和初级抽样单位（分层内）进行确认。为方差估计之目的，不需要初级抽样单位水平以下但在元素阶段之前抽样阶段的信息。这样，实际抽样计划描述就简单化了，看上去像分层的一阶段整群抽样，也就是说，像一个完全计数的最终类集的分层样本。只要有替代假设的第一阶段是合理的，那么该最终类集方法为估计方差提供了可行的近似值。该常见近似值（UCVE）在样本调查文献中有时用WR（有替代）表示，下文中的WR就是以此意义使用的。

15. 这样，当抽样计划被描述为WR，方差估计只需要三个调查设计变量：

- (a) 样本加权变量WTVAR（点估计也需要）；
- (b) 初级抽样之前使用的分层变量（或假层变量）STRATVAR；
- (c) 初级抽样单位（或假初级抽样单位）变量，用PSUVAR表示。

16. 每个样本受访者R在基本数据文档一定要有以上三个变量中任何一个变量的值。例如，一个特别的R可以代表总体中8 714个元素（WTVAR的值为8 714），而且可能是从分层或假层#6（STRATVAR值为6）和从分层6内的初级抽样单位或#3假初级抽样单位中选出的（PSUVAR的值为3，在STRATAVR = 6内）。

17. WR是大多数样本调查软件包或程序的默认值或惟一的样本计划描述。例如，WR是默认值，在SUDAAN、SAS、STATA、Epi-Info、PC-CARP和CENVAR中使用泰勒级数线性化方法。WR是默认值，在WesVar和SUDAAN中使用平衡多次重复和刀切法。要注意，元素的单一阶段抽样（如简单随机抽样或分层随机抽样）是多阶段抽样的特殊情况，在这些特殊情况中，抽样标架中的总体初级抽样单位是总体元素，每个样本的初级抽样单位只包括一个元素（换句话说，就是没有样本元素的类集）。如果使用没有元素替代抽样的单一阶段（例如SAS、STATA和WesVar），那么，只有WR抽样计划描述的软件包可以使方差估计中纳入fpc项。

18. 使用WR计算实际复杂抽样计划的近似值可能会使方差估计偏高。但是，调查数据分析人员一般都愿意接受一定程度的偏高估计，以便使WR近似值相对简单化。但是要注意，如果有几个分层中初级抽样没有替代而且有大的抽样比，那么偏高估计是显著的，在这种情况下，最好使用可以把第一阶段fpc因素纳入在内的软件。

6. 方差估计技术和调查设计变量

19. 公开发布的样本调查数据集一般采用泰勒级数线性化方法和重复方法这两种主要方法中的一种为方差估计而设定。有时候，设定分开发布的数据集以使用两种方差估计方法。变量估计的相关样本设计变量应该包括在公开发布的数据集当中，对于这些变量如何定义和如何使用有相对应的文件。

20. 如果数据集使用了泰勒级数线性化方法，在文件中查找三个调查设计变量：样本加权变量 *WTVAR*，第一阶段分层变量 *STRATVAR* 和初级抽样单位变量 *PSUVAR*。（当然，这些变量不会使用此处所用的名称。）如果数据集使用了重复方法，在文件中查找样本加权变量 *WTVAR* 和几个重复加权变量，这些变量经常被称作 REPL01—REPL52（52个重复加权变量）。如果在数据集中有重复加权变量，则没有必要知道 *STRATVAR* 或 *PSUVAR* 变量。

21. 那些实施调查并为分析准备数据集的调查者需要把相关调查设计变量包括在内，并给数据集中每一个样本受访者元素（R）分配一个值。所需要s的最小变量集是：样本加权变量 *WTVAR*、第一阶段分层（或假层）变量 *STRATVAR* 和分层内初级抽样单位（或假初级抽样单位）变量 *PSUVAR*。这三个调查设计变量近似于实际抽样计划 *WR*，允许在方差估计时直接使用泰勒级数线性化方法或允许对重复加权进行个人或软件计算。如果希望在方差估计中纳入 *fpc* 项和/或抽样或分层的其他阶段，则需要在数据集中加入其他调查设计变量以及具有这些能力的样本调查软件（如 SUDAAN）。

22. 遗憾的是，一种常见情况是不包括任何调查设计变量或重复加权变量的样本调查数据集的获得。假设使用了概率抽样，就有必要在估计中建立调查设计变量 *WTVAR*，在方差估计中建立调查设计变量 *STRATVAR* 和 *PSUVAR*。让人乐观的是，抽样计划的详细情况可以从书面文件中获得，也可以通过和抽样人员直接联系获得，这样就可以建立调查设计变量。如果信息有限，可能会得到一些粗略的近似值。例如，如果没有抽选概率可以重建，那么，假定元素的等概率抽样并且仅使用事后分层调节获得 *WTVAR* 的值也是合理的。如果初级抽样单位不能被确切地确认，如果知道某些地理标识符，也可以计算出近似的初级抽样单位。在这种情况下，如果设计变量不精确，则要清楚数据分析的局限性。

7. 复杂样本调查数据的分析

23. 除建立加权分析和正确估计估计量的方差外，在复杂样本调查数据分析中还有许多理论和实践问题。这些问题在 Korn 和 Graubard（1999年）最近的综合性著作中都得到重视和论述，包括诸如样本调查数据的拟合模型（如逻辑回归）、模型的拟合度、子总体的方差估计、综合多个调查和形成假层和假初级抽样单位等。另见本出版物本部分的其他章节。

C. 方差估计方法

1. 采用泰勒级数线性化方法估计方差

24. 假设一个带初级抽样单位分层、多阶段抽样和元素不等概率抽样的复杂抽样计划。线性估计量 $\sum w_i y_i$ 是加权总和，估计出 *y* 变量的总体总数，其中 w_i 是样本元素 *i* 的样本加权变量 *WTVAR* 的值， y_i 是样本元素 *i* 的 *y* 变量的值， Σ 是样本 $i = 1, 2, \dots, m$ 的样本元素的总和。如果 *y* 是男性糖尿病患者编码为 1 的对生变

量, 其他编码为0, 那么, 正在估计的总体总数是男性糖尿病患者的总数。估计的方差 $\Sigma w_i y_i$ 可以根据以上所讨论的WR假设直接获得。

25. 不妨设 x_i 为男性编码为1、女性编码为0的对生变量, 那么, 估计的男性糖尿病患病率用 $[\Sigma w_i y_i]/[\Sigma w_i x_i]$ 表示, 这是两个线性估计量(或两个加权)的比率。根据WR假设, 该比率估计量的估计方差不能直接获得。即使不使用复杂抽样方法, 而使用简单随机抽样, 该非线性函数即比率的方差估计也不是直接的, 而是需要某种近似方法。

26. 上述非线性估计量的代数表达式可以用一种无限的泰勒级数进行扩展, 以(估计的)预期分子值和(估计的)预期分母值为中心。非线性估计量通过只保留无限泰勒级数主要条件而用代数的方法求出近似值, 其结果是样本数据线性(或不再是线性)函数的代数表达式; 就是说, 非线性比率估计量被“线性化”。现在, 线性化函数的估计方差(包括有关协方差条件)可以根据WR假设直接获得, 正如获得 $\Sigma w_i y_i$ 估计方差一样。在此过程中, 线性化的函数方差在每个分层中单独计算(因为分层的抽样是独立的), 然后, 各层的估计方差相加得出估计量的方差。

27. 如果使用泰勒级数线性化方法, 不仅需要为每个不同的非线性估计, 而且还要为可能使用估计量的每一个可能的抽样计划(WR就是这样一个抽样计划)推导并编制出独特的近似方差估计公式。这种特征被视为方差估计泰勒级数线性化方法的一个弱点。实际上, 使用泰勒级数线性化方法分析样本调查数据的一个软件包可以不包含希望使用的具体估计量和已经使用的实际或近似抽样计划之间的综合。

28. 使用泰勒级数线性化方法的所有软件程序需要对设计变量WTVAR、STRATVAR和PSUVAR进行说明, 就像WR抽样计划近似值所需要的一样。使用泰勒级数线性化方法可以获得额外的抽样计划, 这取决于软件包; 这些抽样计划的使用可能需要额外的设计变量。

2. 方差估计的重复方法

29. 样本调查估计量方差估计的重复方法在理论上已经早已为人所知, 但随着快速计算能力的到来, 才越来越广泛地得到利用。重复方法需要计算机才能进行, 但在估计方差可以计算的不同估计量的数量方面, 这种方法要比泰勒级数线性化方法灵活。

30. 重复方法的总体设想如下: 首先, 和泰勒级数方法一样, 使用完整的或详细的样本来获得所涉总体参数的点估计; 就是说, 整个样本中都要使用总体参数估计量公式。该计算中只需要抽样加权变量WTVAR。

31. 第二, 为估计该估计量的方差, 详细样本中形成许多不同的子样本或“重复品”, 形成的方式是, 每个重复品反映出整个样本的抽样计划和加权程序及调整。每个重复品根据重复加权变量的值进行定义。例如, $REPWT_j$ 是重复品#j的重复加权变量, 其中 $j = 1, 2, 3, \dots, G$ (重复品的总数)。如果整个

样本中的某一观测值没有包含在重复品 j 中，则该观测值对 $REPWT_j$ 来说为零，如果该观测值包含在重复品 j 中，则该观测值为正值。整个样本中所有观测值的 $REPWT_j$ 值的总和是总体中元素数的一个估计值。

32. 第三，每个重复品都使用估计量公式以获得所涉及总体参数的点估计值（重复估计值），产生了同一总体参数的 G 重复估计值。

33. 第四，基于 G 重复估计值的变异性，计算出全部样本估计量的估计方差。

34. 通过不同的方式可以形成重复品，就产生了各种重复的方法。形成重复品的两种主要方法每一个都有自己的变体，这两种主要方法是平衡多次重复（BRR）和刀切法（下文将讨论这两种方法）。采用特殊重复方法为方差估计设定的公开发布的样本调查数据集通常包括有数据集的重复加权变量。在这种情况下，次要的数据分析人员必须使用方差估计软件，该软件包括在数据集内形成重复权数的特别重复方法。

35. 但是，当数据集中不存在重复加权时，可能希望在方差估计中使用重复方法。实施重复方差估计方法的某些软件包也计算重复加权，软件包形成重复加权所需要的最少调查设计变量是：样本加权变量 $WTVAR$ 、分层变量 $STRATVAR$ 和分层内的初级抽样单位变量 $PSUVAR$ 。如果整个样本由于未答复而调节和/或进行事后分层，那么软件包可以在计算重复加权（如WesVar）内将该信息接受为输入。一个人总是可以独自一人计算重复加权（无需软件包），但只建议那些了解重复方法细节的人使用这种策略。

3. 平衡多次重复（BRR）

36. 平衡多次重复（BRR）是一种特别的重复方法，可以用于一般设计，即分层多阶段抽样。但是，此种方法是每个分层确实选择有（抽选的）两个初级抽样单位的特殊情况而开发，一般使用不等概率抽样，可以有替代，也可以无替代。此种方法的使用一般带有复杂抽样计划的 WR 近似值（最终类集方差估计方法）。

37. 对于平衡多次重复，每个重复品均含有整整一半的样本初级抽样单位，每个分层有一个初级抽样单位；每个重复品往往被称作“半样本”。可能的不同重复品总数是 2^L ，其中， L 是分层的数量。但是，没有必要使用所有的 2^L 重复品，这些重复品可能需要不定的计算时间。相反，一个更小的平衡重复集会产生相同的方差估计，该方差估计可以从所有可能的重复品中获得。使用Hadamard矩阵（Wolter, 1985年）形成 G “平衡”重复品，这样，每个样本初级抽样单位都以相同数量的重复品出现，两个不同分层的每对样本初级抽样单位都以相同数量的重复品出现。所需的重复品最小数目 G 是大于或等于 L 但能被4整除的最小整数。例如，49个分层需要52个平衡多次重复品，每一分层有两个抽样的初级抽样单位。没有包含在重复品 j 中的样本初级抽样单位的观察值对重复加权变量 $REPWT_j$ 来说，其值是0，包含在重复品 j 中的样本初级抽样单

位内的观察值对重复加权变量 $REPWT_j$ 来说，其值是整个样本中抽样加权的两倍，但该值可能因为未答复和/或事后分层而调整。

38. 上述定义的平衡多次重复一个常见的变异是由Fay研究出来的（Judkins, 1990年），因为当整个样本中分母没有几种选择时，如果小范围或人口比率需要估计，标准的平衡多次重复可能就会出现问题。在Fay的方法中，没有为重复品 j 选择的样本初级抽样单位中，观察值没有被消除，就像在标准平衡多次重复中没有被消除一样。相反，它们的抽样加权因一乘法因数 $K(0 \leq K < 1)$ 而减少，而在为重复品选择的样本初级抽样单位中，观察值的抽样加权因乘法因数 $(2 - K)$ 而增加。设 $K = 0$ ，即产生标准的平衡多次重复方法。Fay方法中，一般建议值是 $K = 0.3$ 。

4. 刀切重复方法 (JK)

39. 刀切重复方法的一般设想是在某个时候删去一个样本初级抽样单位从而形成重复品，然后在必要时对每个重复品重新加权，这样对整个样本所代表的总体进行推理。在样本随机抽样或分层随机抽样中，一个样本初级抽样单位可以成为一个单一元素，或者，在近似抽样计划WR中，一个样本初级抽样单位可以包含若干元素。

40. 首先考虑这样一种情况，即在初级抽样单位抽样之前没有使用分层，而且 G 样本初级抽样单位（具有大致相同的元素数）与整个样本相似。通过在某个时间删去一个样本初级抽样单位形成总的 G 重复品数。对具有重复加权变量 $REPWT_j$ 的重复品 j ，删去的样本 $PSU\#j$ 中的观察值对 $REPWT_j$ 来说是0。剩下的（没有删去的）样本初级抽样单位对 $REPWT_j$ 来说，值等于与因数 $[G/(G - 1)]$ 相乘的观测值的抽样加权。

41. 第二个例子是每个分层中确实有两个抽选的初级抽样单位的 L 分层；即上述针对平衡多次重复讨论的设计。一次删去一个样本初级抽样单位会导致 $2L$ 重复。对每一个 $2L$ 重复来说，带删去样本初级抽样单位的分层中剩下的样本初级抽样单位将使每个观测值的抽样加权乘以2（删去的样本初级抽样单位将使每个观测值的抽样加权乘以0）。但是，此种方法通常在只有 L 重复的情况下实施，而不是在 $2L$ 重复的情况下实施，这种情况下，每一个 L 分层中，仅一个随机挑选的样本初级抽样单位被删去。对于线性估计量，仅使用 L 重复的方差估计量与使用 $2L$ 重复的方差估计量在代数上等值。

42. 最一般的抽样计划是分层多阶段抽样，共有 L 个分层（在初级抽样单位抽样之前），每个分层中有两个或两个以上抽样的初级抽样单位。每个样本初级抽样单位被删去，以形成重复；重复品数 G 等于整个样本（ n ）中抽样初级抽样单位的总数。在分层 h 中，删去的样本初级抽样单位中每个观测值的重复加权 $REPWT_j$ 的值是样本加权变量 $WTVAR$ 乘以零。删去样本初级抽样单位的分层 h 中，剩下的每个观测值变量 $REPWT_j$ 的值是样本加权变量 $WTVAR$ 乘以因数 $[n_h/(n_h - 1)]$ ，其中， n_h 是整个样本中分层 h 内样本初级抽样单位的数量。

5. 方差估计软件用户常犯的一些错误

43. 有几个软件包要求用户使用调查设计变量对输入数据集进行分类，例如，可以通过STRATVAR和STRATVAR内的PSUVAR（如第35段所述）进行。忘记分类可能会产生错误的估计方差，如果数据集没有正确分类，大多数程序将显示错误信息。

44. 由于对样本调查文件审查不充分，公开发布的数据集的用户可以说明错误的调查设计变量。错误说明的样本加权变量将导致误差估计量和错误估计的方差；也就是说，所有分析将是错误的。如果样本加权变量是正确的，但分层和/或初级抽样单位变量是不正确的，那么，点估计将是正确的，但估计的方差将是不正确的。

45. 一些公开发布的数据集有多个数据文档，这些不同的文档有不同的调查设计变量。不同的数据文档可能有不同的分析单位，这个单位可以是个人、住户或家庭，所以需要特别注意对输出的解释。一些调查变量可以仅按整个样本的概率子样本进行测量，需要不同的样本加权变量，而不需要按整个样本测量的变量。认真通读文件对所有的样本调查都很重要，不管抽样计划简单还是极为复杂。

D. 方差估计的软件包比较

46. 关于样本调查数据包的各种软件包，包括本文评价的八个软件包，其链接网址是：www.fas.harvard.edu/~stats/survey-soft/survey-soft.html，这个网站的信息量非常丰富。欲了解对复杂样本调查数据软件包的评议，另见Carlson（1998年）。注意，评价的软件包不包含SPSS。自2003年初，SPSS还不能进行复杂样本调查方差估计，但在印制本章的2003年底，它确实发布了一种附加模块。

47. 本章的后半部分对复杂样本调查数据的8个方差估计软件包进行了评价和比较：SAS、SUDAAN、STATA、Epi-Info、WesVar、PC-CARP、CENVAR及IVEware。8个软件包中的前5个通过描述分析得到了论述，使用了1989年在布隆迪开展的一项样本调查中的数据；总体比例、均值和总数得到估计，根据这些参数对值域进行了比较。布隆迪分析得出的结果在本章的表二十一.1中进行了汇总，CD-ROM上的附件给出了每个软件包的详细表格、加注范本程序和输出。附件中加注的范例可以帮助用户学习如何使用方差估计软件包的前5个软件包。

48. 使用布隆迪调查数据论述的5个软件包中，有3个（STATA、SAS和Epi-Info）包括了普通统计软件包内的样本调查程序。这三个软件包使用方差估计的泰勒级数线性化方法。论述的其余两个软件包（WesVar和SUDAAN）专为样本调查方差估计而开发，WesVar使用重复方法，而SUDAAN使用了泰勒级数线性化方法和重复方法。

表二十一.1

5个软件包中PROCS的比较：血清反应为阳性的妇女估计百分比和人数，标准估计误差，刚刚生育过的妇女，布隆迪，1988 - 1989年

软件包和PROC	%血清反应为阳性	s.e.%血清反应为阳性	95%CI %血清反应为阳性	血清反应为阳性人数	s.e.#血清反应为阳性	95%CI #血清反应为阳性
SAS 8.2 MEANS ^a 无加权	74.88% 错误	2.12% 错误	N-APP	N-APP	N-APP	N-APP
SAS 8.2 MEANS ^b 加权	67.20%	2.30% 错误	N-APP	N-APP	N-APP	N-APP
SAS 8.2 SURVEYMEANS	67.20%	3.83%	59.38%, 75.02%	142 485	8 848.10	124 415, 160 556
SUDAAN 8.0 CROSTAB和 DESCRIPT和BRR	67.20%	3.83%	N-AV	142 485	8 848.10	N-AV
STATA 7.0 Svymean	67.20%	3.83%	58.38%, 75.02%	N-AV	N-AV	N-AV
STATA 7.0 Svtotal	N-AV	N-AV	N-AV	142 485	8 848.10	124 415, 160 556
Epi-Info 6.04d CSAMPLE ^c	67.20%	3.83%	59.70%, 74.71% ^c	N-AV	N-AV	N-AV
WesVar 4.2	67.20%	3.83%	59.38%, 75.02%	142 485	8 848.10	124 415, 160 556

说明：使用的缩略词：
CI=置信区间，N-APP=不适用，N-AV=没有数据，s.e.=标准误差。

^a 错误指定的分析；忽略抽样加权、类集和分层。

^b 错误指定的分析；纳入了抽样加权，但没有类集和分层。

^c Epi-Info 6.04d给出的置信区间比其他软件包的置信区间窄。Epi-Info 6.04d使用 $z=t=1.96$ 以建立95%的置信区间，而其他软件包使用了 $t=2.042$ ，从30 ddf的学生t-分布中获得（样本调查的分母自由度，计算为初级抽样单位数减去假层的数量）。最好使用实际的调查ddf。

49. 还评价了另外3个软件包（PC-CARP、CENVAR和IVEware），但没有使用布隆迪调查数据进行论述。PC-CARP和CENVAR都使用了方差估计的泰勒级数线性化方法，而IVEware使用了泰勒级数线性化方法和重复方法。

50. 这里评价的8个软件包包括样本调查方差估计的许多（但不是所有的）可能的选择。选择3个软件包（Epi-Info、CENVAR和WesVar 2）是因为这三个软件包提供基本的描述分析，可以免费从网站上下载，这对购买软件的预算有限或没有这种预算的分析人员有吸引力。选择2个软件包（PC-CARP和WesVar 4）是因为尽管这两个软件不是免费的，但和其他软件包相比费用低廉，提供描述分析和基于设计的线性和对数回归。选择两个价格适中的软件包（SUDAAN和STATA）是因为这两个软件包除提供描述分析之外，还为基于设计的回归模式提供综合选择。尽管价格昂贵，但还是选择了SAS软件包，是因为这一软件包在数据管理和分析领域占居主导地位，而且对样本调查数据分析具有相对较新的PROCS。最后，选择最近发布的IVEware（Beta版），是因为这一软件除提供多种推算程序外，还提供综合的描述分析和基于设计的回归模式。IVEware是免费的（可以从网上下载），可随时作为SAS软件使用（所以需要SAS）。

51. 表二十一.2概述了8种软件包，介绍了这些软件包的各种特点，包括涵盖的抽样计划、方差估计方法和分析类型。

表二十一.2

具有复杂样本调查数据方差估计能力的8种软件包的属性

属 性	SAS 8.2	SUDAAN 8.0	STATA 8.0	Epi-Info 6.04d	WesVar 4.2	PC-CARP	CENVAR	IVeware
泰勒级数	有	有	有	有	无	有	有	有 Desc
重复方法BRR和JK	无	BRR JK	无	无	BRR JK	无	无	JK 模式
形成的重复权数	无	无-BRR 有-JK	无	无	有 BRR/JK	无	无	有 JK
输入数据集	SAS	SAS, SPSS, ASCII	STATA	Epi-Info	SAS, SPSS, STATA, ASCII, ODBC	ASCII	ASCII	SAS
估计总值	有	有	有	无	有	有	有	无
总值CI	有	无	有	无	有	有	有	无
总值LC	无	有	有	无	有	有	有	无
估计均值	有	有	有	有	有	有	有	有
均值CI	有	无	有	有-窄	有	有	有	有
均值LC	无	有	有	有	有	有	有	有
估计比例	有	有	有	有	有	有	有	有
比例CI	有	No	有	有-窄	有	有	有	有
比例LC	无	有	有	有-误差	有	有	有	有
估计比率	有	有	有	无	有	有	有	无
比率CI	有	无	有	无	有	有	有	无
比率LC	无	有	有	无	有	有	有	无
域分析	有	有	有	有	有	有	有	有
比较域	无-8.2 有-9.0	有	有	有	有	有	有	有
子总体分析	无	有	有	无	有	有	有	有
标准率/均值	无-8.2 有-9.0	有	有	无	有	无	无	无
卡方检验	无-8.2 有-9.0	有	有	无	有	有	无	无
对数回归	无	有	有	无	有	有	无	有
机会比率	无	有	有	有	有	有	无	有
风险比率	无	有	有	有	有	无	无	无
线性回归	有	有	有	无	有	有	无	有
其他回归模式	无	有	有	无	无	无	无	有
描述许多样本阶段	无	有	无	无	无	无	无	无
设计效果	有	有	有	有	有	有	有	有
免费试用软件	无	无	无	NA 免费	有	无	NA 免费	NA 免费
通用统计软件包	有	无	有	有	无	无	无	无

属性	SAS 8.2	SUDAAN 8.0	STATA 80	Epi-Info 6.04d	WesVar 4.2	PC-CARP	CENVAR	IVEware
管理数据能力	有	无	有	有	有	无	无	无
通过输入程序操作	有	有	有	无-6.04d 有-2002	无	无	无	有
通过简短命令操作	无	无	有	无	无	无	无	无
通过菜单选择操作	无	无	无	有	有	有	有	无
根据分层和初级抽样单位挑选数据集	无	有	无	有	无	有	有	无
开发者提供的培训	有	有	有	无	有	无	无	无
书面/在线手册	有	有	有	有	有	有	有	有
调查程序指南	无	无	无	无	有	无	有	无
费用	高	中	中	免费	低 V4 免费 V2	低	免费	免费
年度更新费	高	中	无	无	无	无	无	无
推算数据	无	无	无	无	无	有	无	有

使用的缩略词：ASCII = 美国信息交换标准代码，BRR = 平衡多次重复，CI = 置信区间，JK = 刀切，LC = 线性对比，NA = 无数据，ODBC = 开放数据库连接性，V = 版本。

E. 布隆迪样本调查数据集

52. 本章数值例子均使用了来自1989年在布隆迪开展的破伤风类毒素（TT）免疫覆盖范围样本调查的数据集。接下来是布隆迪样本调查设计的简单概述；更多细节见CD-ROM附件中第1部分。关于该调查方法及其公布结果的更多信息，见世界卫生组织（WHO）扩大免疫方案（EPI）（1996年）的报告。

1. 推断总体和总体参数

53. 该调查中推断总体是布隆迪在1988年复活节到1989年2/3月生育的妇女。涉及总体参数是对破伤风抗毒素血清反应为阳性、从而保护其新生儿免于新生儿破伤风的妇女之百分比（或比例）。

2. 抽样计划和数据收集

54. 抽样计划是根据Brogan等人（1994年）的建议而修订的为扩大免疫方案而在世界卫生组织开发的类集样本调查方法。该修订方法产生了住所或住房单位概率样本以及由此产生的妇女概率样本，而标准的世界卫生组织扩大免疫方案整群抽样方法可能不会产生这样的概率样本（同上）。

55. 布隆迪被分为两个地理区域层，即首都布琼布拉（城市分层）和该国其他地区（农村分层）。初级抽样单位（PSUs）是地理区域，农村分层内是区（*collines*），城市分层内是*quartiers*或街道（*avenues*）。每个分层的初级抽样单位抽样标架按地理距离排列，使用了系统的ppes（概率与估计规模成比例）选择每个分层的30个样本初级抽样单位。由于布隆迪农村地区居住着96%的推

断总体，而且由于每个分层分配到的样本初级抽样单位数相同，城市妇女实际上抽样过多。抽样标架的初级抽样单位特别排序以及初级抽样单位的系统 pps 抽样使得每个分层内部出现了隐含的地理分层。

56. 进行了更多阶段的样本初级抽样单位内概率抽样，以便获得有人占住的住所的样本。每个抽样的住所内所有有接受调查资格的妇女都被选作样本。根据刺破手指获得的血液样本，确定破伤风抗毒素滴定率的血清反应是否为阳性。调查答复率基本上是100%，这是罕见的高答复率。访谈了206位城市妇女和212位农村妇女。

3. 方差估计的加权程序和设置

57. 布隆迪数据集提供的样本加权变量 W 被修改为 W_2 ，这样，样本受访者 R 的 W_2 值是 R 所代表的推断总体中妇女人数的估计值。 W_2 的值是近似的，仅用于论述用各种软件包进行总体总数估计。从本章的分析中不能得出有关1989年布隆迪符合调查资格妇女总体总数的实质性结论。重要的是要注意，因为修改后的 W_2 是以前分析中使用的 W 的纯量倍数，本章所报告的估计比例和均值与该数据集以前公布的结果（扩大免疫方案，1996年）一致。农村样本妇女的 W_2 值是959.3，城市样本妇女的 W_2 值是42.0，反映了城市妇女实质上抽样过高。布隆迪抽样计划由用于方差估计目的的一般描述 WR 近似，即由第一阶段低抽样比的最终类集方差估计方法近似。

58. 由于初级抽样单位被隐含分层，每个城市和农村分层中的抽样计划都被看作是从15个假层之一中抽取的两个初级抽样单位。偏向于把抽样计划描述为30个假层，每一个假层有两个样本初级抽样单位，而不是描述为2个分层，每一个有30个样本初级抽样单位，因为前一种描述产生偏误方差估计较小，该种描述考虑到了隐含分层。假层变量 $PSTRA$ 被编码为1-30，每个假层内假初级抽样单位变量 $PPSU$ 被编码为1或2。

59. 当泰勒级数线性化方法用来进行方差估计时，只需要变量 W_2 、 $PSTRA$ 和 $PPSU$ 。但是，如果采用的是重复方法，则需要重复加权。 $WesVar$ 被用于根据变量 W_2 、 $PSTRA$ 和 $PPSU$ 计算平衡多次重复方法的重复权数。由 $WesVar$ 计算的这些重复权数同时用于 $WesVar$ 和 $SUDAAN$ 中，用来采用平衡多次重复法估计方差。

4. 调查数据分析的三个例子

60. 附件含有以下三个例子的注解数据分析，并使用了样本调查数据软件包（第2-4节）。下面的这些例子论述了就样本调查数据所做的一般描述分析和解析分析，即(a) 整个总体以及域或层的比例、总数和均值估计；及(b) 有关均值或比例的域或层比较。推断总体为1989年初布隆迪符合调查资格的妇女。

- 例1：妇女的估计人数（总人口）和血清反应为阳性的妇女百分比（总体比例/百分比）（ $IMMUNE$ 变量，1 = 血清反应为阳性，2 = 血清

反应为阴性)。变量**BLOOD** (1 = 血清反应为阳性, 0 = 血清反应为阴性) 是**IMMUNE**的重新编码。

- 例2: 在城市和农村妇女中估计例1的总体参数 (*RUR_URB*变量, 编码 1 = 农村, 2 = 城市)。确定农村/城市居民是否在统计上独立于血清反应为阳性 (*IMMUNE*)。
- 例3: 针对妇女推断总体, 按农村/城市住所, 估计每毫升 (ml) 抗毒素平均国际单位 (*IUML*)。确定农村/城市住所是否与平均国际单位有关 (*IUML*)。

注: 每毫升 (ml) 抗毒素平均国际单位 (*IUML*) 的估计可能会由于此变量的偏斜分布而产生误导结果。最好使用均值, 或在分析之前转换*IUML*, 比如, 转换为*IUML*的自然对数。本章中, 估计均值*IUML* (没有转换) 来显示5个软件包的能力, 而不是举例说明有关*IUML*的实质结果。

F. 使用非样本调查程序分析样本调查数据

61. 本节论述分析重复样本调查数据时错误地使用简单随机样本公式会产生偏误点估计值和有偏误 (一般很小) 估计标准误差。欲了解另外的论述, 见 Brogan (1998年及正在印制)。任何统计软件包都可能被用于这种论述, 可以获得与使用 (本节使用的) SAS获得的答案相似的答案。

62. 要估计的总体参数是推断总体中血清反应为阳性的妇女人口的比例。指示变量**BLOOD**得到计算, 编码为1 = 血清反应为阳性和0 = 血清反应为阴性。这样, **BLOOD**均值是血清反应为阳性妇女的比例。SAS中PROCMEANS估计出**BLOOD**均值为0.74880, 估计的标准误差是0.02124 (见表二十一.1第1行)。这两种计算都有偏误, 因为错误使用的PROC MEANS忽略了估计总体比例的抽样加权变量, 而且还忽略了计算点估计的估计标准误差的抽样加权、初级抽样单位和分层变量。

63. 然后, SAS中的PROC MEANS和抽样加权变量W2一起用于WEIGHT语句中。**BLOOD**均值估计为0.67203, 估计的标准误差, (加权) 是0.02299 (见表二十一.1第2行)。在这个分析中, PROC MEANS获得了总体比例的近似点估计值。但是, 错误使用的PROC MEANS产生了有偏误的估计标准误差, 因为它忽略了初级抽样单位变量和分层变量。

64. 最后, 使用SAS中的PROC SURVEYMEANS对复杂样本调查数据进行适当分析。(有关如何使用SURVEYMEANS的详情见下一节。) 总体比例的点估计值是0.67203, 估计的标准误差是0.03830 (见表二十一.1第3行)。PROC SURVEYMEANS考虑到了计算点估计值的抽样加权变量以及计算估计标准误差的抽样加权、初级抽样单位和分层变量。

65. 表二十一.1中的这三项分析比较表明, 未加权 (不正确) 点估计值 0.7488 (74.88%) 与加权 (正确) 估计值 0.6720 (67.20%) 之间有一点偏误。未加权的点估计值太高, 因为与农村妇女相比, 更多城市妇女的血清反应为阳性 (本章下面会有论述), 而且由于过高抽样, 样本中城市妇女的人数过多: 她

们占样本中大约一半的人数，但只占推断总体的4%。这样，在对该国进行推理的未加权分析中，给城市妇女以更多的影响和高于估计总体比例的偏误。

66. 对产生正确加权点估计值的两种分析进行比较，结果说明，即使使用了WEIGHT语句，SAS中错误使用的PROC MEANS严重低估了标准误差，错误的计算是0.02299（2.30%），而正确的PROC SURVEYMEANS计算是0.03830（3.83%）。之所以如此，主要是因为不论使用还是未使用WEIGHT语句，PROC MEANS均忽略了样本初级抽样单位中的妇女类集，而SURVEYMEANS注意到了方差估计的类集。由于组内相关系数对复杂样本调查中大多数实测变量来说是正值，把类集纳入考虑范围的正确方差估计程序通常会产生较大的估计标准误差。

67. 总而言之，如果样本调查数据没有利用合适的样本加权变量进行分析，就获得了有偏误的总体参数点估计值。另外，即使在分析中纳入了样本加权变量，产生了适宜的总体参数点估计值，当样本元素归入调查数据类集，而且该类集在方差估计中没有得到确认，则标准误差往往被低估。低估标准误差产生很窄的置信区间和很小的P-值显著性统计检验，换言之，统计显著性水平被夸大。

68. 通过忽略样本调查数据类集的低估方差量值，用公式 $[1 + p(b - 1)]$ 求出近似值。该公式中， p 是总体元素之间的组内相关系数， b 是每个样本类集（初级抽样单位）样本元素的均值（见第六章）。例如，如果公式的值为2，把类集考虑在内大约会使忽略类集获得的估计方差扩大一倍。请注意，名为PPSU的布隆迪初级抽样单位变量为软件确认（某一分层的）相同样本初级抽样单位内何种样本元素是类集在一起的。

69. 除类集对估计的方差产生影响外，答复者中间抽样加权变量值的实际变异提高了估计的方差。这样，如果分析中忽略抽样加权变量，估计的标准误差就被低估（总体参数的估计值就产生偏误）。

G. SAS 8.2中的样本调查程序

1. SURVEYMEANS和SURVEYREG概述

70. SAS中8.2版包括两种最近开发的样本调查数据分析程序（这两种程序首先出现在8.0版）：SURVEYMEANS和SURVEYREG。SAS包括普通的抽样计划描述WR，这两个程序需要三个基本的调查设计变量。单一阶段抽样设计（如分层随机抽样和简单随机抽样）可以使用有限总体修正项。泰勒级数线性化方法用于方差估计。SAS V9包括用于复杂样本调查数据的两个新PROCS，SURVEYFREQ用于分类变量的分析，SURVEYLOGISTIC用于对数回归。目前还在开发用于样本调查数据的其他SAS程序。

71. 说明WR相关调查设计变量的语法对SURVEYMEANS和SURVEYREG都是相同的。关键词STRATA用于指称分层变量，关键词CLUSTER用于指称初级抽样单位变量，关键词WEIGHT用于说明抽样加权变量（如MEANS等其他SAS程序中

的一样)。这些语句适合某一种调查,一定存在于每一个SAS样本调查程序,只要相同的样本调查数据集正处于分析之中,这些语句一般就不会发生改变。对于布隆迪数据集,下面的SAS语句描述了SAS PROC SURVEYMEANS或PROC SURVEYREG的样本调查设计:

STRATA	<i>PSTRA</i>
CLUSTER	<i>PPSU</i>
WEIGHT	<i>W2</i>

72. 如果STRATA语句缺失, SAS假定抽样计划在初级抽样之前没有初级抽样单位分层。如果CLUSTER语句缺失, SAS假定样本元素没有进行类集,就是说,每一样本类集只包括一个元素;在抽样的第一个阶段(或惟一的阶段),所有的元素都被抽样;使用了简单随机或分层随机抽样。如果WEIGHT语句缺失, SAS假定每个*R*有相同的加权变量值, SAS分配给加权变量的值是1.0。如果三个调查设计语句(STRATA、CLUSTER、WEIGHT)缺失,就相当于从无限的总体中规定简单随机抽样,这是对SAS中大多数非调查PROC做出的假设。

2. SURVEYMEANS

73. 本程序使用样本调查数据,估计连续变量和总体比例的人口均值和总数,以及分类变量的总数。对所有的点估计值都提供估计的标准误差和变化系数,为总体参数提供置信区间。可以在PROC语句中要求提供特别统计量,也可以把缺省打印输出当作统计量,或者使用PROC语句中的ALL获得所有的统计量,用SURVEYMEANS计算这些统计量。

74. VAR语句上出现需要分析的变量(连续和分类变量)。CLASS语句列出了VAR语句上的分类变量; SAS假定有关VAR语句的其他所有变量都是连续变量。

75. 带有一个或更多分类变量的DOMAIN语句用来说明VAR语句的所有变量的分析域,除域分析外, SAS自动对剩余整体(即整个整体)进行分析。不带DOMAIN语句的程序只为整个人口提供估计值。虽然SURVEYMEANS中BY语句可以用来获得域的估计值,但不建议在样本调查中使用该语句,因为如果使用BY语句,就不能使用方差估计的适当公式。使用DOMAIN语句分析域。

76. SAS V8.2没有能够分析子总体的语句,比如仅分析年长妇女的语句。但是,可以通过首先确定指示变量进行子总体分析,譬如确定*OLDERFEM*,该变量说明样本元素是否属于子总体。然后,语句*DOMAIN OLDERFEM*可以用来获得预期的分析,忽略不是年长妇女的样本元素的SAS输出。在使用PROC SURVEYMEANS之前,不要使用SAS IF语句为年长妇女的数据集建立子集,原因是,由于SURVEYMEANS可能不知道样本调查中分层和样本初级抽样单位的全部数量,从而可能无法正确地计算标准误差。

3. SURVEYREG

77. 根据基于设计的方法,本程序对样本调查数据实施线性回归(Korn和Graubard, 1999年),就是说,分析考虑到了调查设计变量。对于非调查数

据的线性回归，因变量是连续变量（或假定是连续变量），自变量可以是连续变量和分类变量的混合。MODEL语句包括因变量和所有的自变量，MODEL语句的任何分类变量必须也出现在CLASS语句内，在SAS程序中，CLASS语句一定在MODEL语句之前。对分类自变量，SURVEYREG形成虚拟指示变量（编码为1或0），定义为参照组的变量，其编码值最高。SURVEYREG中其他的选择及其输出与SAS中（非调查）线性回归相似。

78. SAS V8.2没有样本调查程序用以比较均值或比例的域，这些功能还在开发之中。在这种情况下，比如可以有这样一个问题，布隆迪推断总体中农村妇女和城市妇女在血清反应为阳性的中间IUML单位或在比例上有区别吗？V9.0中的SURVEYFREQ可以用来对两个变量居住区（农村/城市）和血清反应为阳性（是/否）两个问题进行卡方检验。在用于样本调查数据的SAS中全面开发出域比较程序之前，可以如以下所述使用SURVEYREG比较域。

79. 如果需要在均值IUML方面对推断总体中农村妇女和城市妇女进行比较，使用SURVEYREG中的MODEL语句，把连续变量IUML作为因变量，把指称农村/城市的域变量作为自变量。SURVEYREG中标准输出部分是对零假设的检验，这种零假设是，农村/城市（自由度为1）的人口回归系数等于0，这种关于回归系数的零假设相当于推断总体中农村和城市妇女具有相同均值IUML的零假设。

80. 如果需要就血清反应为阳性的比例对推断总体中的城市和农村妇女（对生变量）进行比较，则使用指示变量BLOOD（1 = 血清反应为阳性，0 = 血清反应为阴性）作为因变量。（注意BLOOD只是IMMUNE变量的再编码，其中1 = 血清反应为阳性，2 = 血清反应为阴性。）在SURVEYREG中的MODEL语句内，把BLOOD定义为因变量，把指称农村/城市的域变量定义为独立分类变量。回归系数等于0的零假设相当于推断总体中血清反应为阳性的农村和城市妇女比例相等的零假设。

4. 数值例子

81. CD-ROM的附件第2部分说明使用SURVEYMEANS和SURVEYREG处理第60段所列的三个例子。对加注解的SAS程序（用户写程序）和加注解的SAS输出进行评价，有助于读者为写出自己的SURVEYMEANS和SURVEYREG SAS程序以及解释其输出做好准备。

82. 表二十一.1第3行概述了附件第2部分的SURVEYMEANS输出，该输出用来估计布隆迪推断总体中血清反应为阳性的妇女人数和百分比，并且有估计标准误差和置信区间；这些结果中的大多数在本章F节得以讨论。表二十一.3第1行（见CD-ROM中附件的第7部分）概述了SURVEYMEANS输出，该输出用来估计农村和城市妇女这两个域中血清反应为阳性的百分比，分别为66.51%和83.50%。表二十一.4第1行（见CD-ROM中附件的第7部分）概述了SURVEYREG输出，该输出对农村和城市妇女进行了比较，得出的t值为-3.52，p值为0.0014，用来检测农村和城市妇女中血清反应为阳性者的比例没有区别这一零假设。因

此，推断人口中，农村和城市妇女中血清反应为阳性者的比例不同：城市妇女中血清反应为阳性的比例较高。

5. 优点/缺点/费用

83. 如果已经是SAS/STAT的用户，使用SAS中样本调查程序不再需要额外费用，而且这些程序使用了人们熟悉的语法。另外，还可使用SAS的全部功能进行数据管理和形成新变量。样本调查程序的技术支持和文献包含在SAS常规支持系统中。与前面所评价的其他样本调查软件包相比，SAS的费用较高。

84. SAS 8.2没有能力对范围进行比较，而SURVEYREG可以用作这种分析的临时解决办法。V9.0中的SURVEYFREQ增补版提供有关分类变量的比较。

85. SAS只使用泰勒级数线性化方法进行方差估计。对于分层多阶段整群抽样，SAS仅处理普通抽样计划描述WR。但是，它可以把 fpc 项纳入单阶段分层随机抽样或简单随机抽样。

86. SAS 8.2进行样本调查数据分析的能力是基本的、描述性的，可以满足许多用户的分析需求。V9.0中SURVEYFREQ增补版对分类变量具有描述和分析能力。还在开发之中的样本调查程序，比如对数回归，使得SAS将来可以和其他提供综合样本调查分析的软件包相媲美。

H. SUDAAN 8.0

1. SUDAAN概述

87. SUDAAN（研究三角研究所，2001年）是一种专业软件包，起初开发用来对复杂样本调查数据进行分析，但现在得以综合，利用诸如纵向数据分析和综合估计等式（GEE）等技术对相关数据进行分析。SUDAAN是SURVEY DATA ANALYSIS的缩略形式，用于描述和分析统计量的程序是DESCRIPT、CROSSTAB和RATIO。基于设计的建模程序包括线性回归、对数回归（包括多项式）、对数-线性回归和生存分析。

88. SUDAAN 8.0是用C语言编程的，其中用户提供的命令语句与SAS的命令语句相似。输入数据集可以是SAS、SPSS或ASCII文档。SUDAAN可以自行运行（独立SUDAAN），也可以与SAS一起运行（可随时调用SAS的SUDAAN）。SAS用户一般更倾向使用可随时调用SAS的SUDAAN。

89. SUDAAN是惟一包含用于方差估计的两个最常见方法的样本调查软件包：这两种方法就是泰勒级数线性化方法和重复方法。SUDAAN中后一种方法包括具有或者没有Fay调节因数的平衡多次重复（BRR）和刀切法。SUDAAN中所有重复方法假定普通抽样计划描述以前为WR。如果平衡多次重复用于方差估计，对采用平衡多次重复方法的重复加权一定要提供输入数据集；SUDAAN不产生采用平衡多次重复法的重复加权。SUDAAN将为刀切删除一（初级抽样单位）方法产生重复加权，或者接受因刀切删除一方法和该方法变化而提供了输入数据集的刀切重复加权。

90. 使用三种语句把样本调查设计描述到SUDAAN上, 采用的方式是:
(a) 为PROC语句的DESIGN关键词选择一个选项; (b) 指定NEST语句的分层和类集变量; 以及(c) 指定WEIGHT语句的样本加权变量。输入到SUDAAN的输入数据集一定通过NEST语句出现的所有变量进行分类, 这些变量一般是第一阶段分层变量以及每个分层内的初级抽样单位变量。

91. 不像拥有样本调查能力的多数其他软件包, 多阶段抽样中抽样和分层的第二和更高阶段可以描述到SUDAAN进行方差估计, 减少了一贯使用一般抽样计划描述WR的必要性。另外, SUDAAN具有广泛的能力, 可以把没有替代抽样的多阶段有限总体修正 (*fpc*) 项纳入方差估计。SUDAAN手册有印刷版本, 也有pdf文件, 该手册举了一些例子, 说明如何把抽样计划描述到SUDAAN的几个例子 (见第三章)。

92. SUDAAN的默认抽样计划是上述定义的WR, 不论是对泰勒级数线性化方法、平衡多次重复, 还是刀切法, 都是如此。在PROC语句上使用SUDAAN语法DESIGN=WR不仅要求最终类集方差估计方法和初级抽样, 该抽样有替代或没有替代, 但有初级抽样的很小小数, 还要求使用泰勒级数线性化方法。DESIGN = WR, NEST语句包括一个或多个调整变量 (通常只有一个) 以及一个初级抽样单位变量。如果PROC语句缺失选项DESIGN=, 则SUDAAN假定DESIGN = WR。

93. SUDAAN语法DESIGN=BRR需要带有用于方差估计的平衡多次重复的一般抽样计划描述WR (前面已经讨论过)。采用平衡多次重复法的重复加权变量一定要在输入数据集之内, SUDAAN程序中REPWGT语句给出了变量名称, 用于重复加权变量。

94. 在没有JACKWGTS和JACKMULT语句的情况下, SUDAAN语法DESIGN = JACKKNIFE调用使用删除一刀切法进行方差估计的一般抽样计划描述WR, 其中, SUDAAN产生刀切法重复加权。在有JACKWGTS语句的情况下, SUDAAN语法DESIGN = JACKKNIFE调用一般抽样计划描述WR, 并把刀切法权数提供给SUDAAN, 作为输入数据集中的变量。

95. 用于布隆迪调查的样本调查设计和方差估计泰勒级数线性化方法的说明按如下方式描述到SUDAAN:

```
PROC.....      DESIGN = WR.....
NEST  PSTR      PPSU
WEIGHT  W2
```

96. 用于布隆迪调查的样本调查设计和方差估计平衡多次重复方法 (平衡多次重复) 的说明按如下方式描述到SUDAAN:

```
PROC.....      DESIGN = BRR.....
WEIGHT  W2
REPWGT  REPLWT01 - REPLWT32
```


以上要注意，REPWT语句确认包括在输入数据集中的重复加权。这32个重复加权变量基于30个假层之上，每个假层有两个初级抽样单位，通过使用WesVar获得32个重复加权。还要注意，使用平衡多次重复时，NEST语句缺失，SUDAAN不需要知道分层和初级抽样单位变量，因为它仅使用用于方差估计的重复加权变量。

2. DESCRIPT

97. DESCRIPT程序估计连续变量的总体总数和均值以及分类变量的总体总数和百分比。VAR语句列出了需要分析的（相关）变量。对于一个既定的DESCRIPT程序，VAR语句上的所有变量一定是连续变量，或者所有变量一定是分类变量。如果VAR语句上有分类变量，那么，CATLEVEL语句也一定用来说明，每一分类变量中哪些等级需要估计。例如，下面两个语句估计布隆迪血清反应为阳性和血清反应为非阳性的推断总体百分比 [假设IMMUNE编码为1, 2或者(……)代表缺失]。

VAR	IMMUNE	IMMUNE
CATLEVEL	1	2

98. 使用包含一个或多个分类变量的TABLES语句，为范围提供估计值。利用CONTRAST、PAIRWISE或DIFFVAR语句，通过线性对比将对域进行相互比较。通过使用STDVAR和STDWGT语句，可以估计标准率和均值，例如按年龄调整后的疾病流行情况。关于平均值或百分比的线性和高水平（二次方程等）的趋势可以通过使用POLY（POLYNOMIAL）语句在一些分类变量的各阶层进行评估；SUDAAN使用正交多项式线性对比进行这些分析。

99. TABLES、CONTRAST、PAIRWISE、DIFFVAR、STDVAR或POLY语句上的所有变量也必须出现在SUBGROUP语句上，要求有LEVELS语句，该语句显示SUBGROUP语句上的每个分类变量分析中的最高编码值。

100. SUDAAN中SUBPOPN语句可以用在所有的PROCS中，把分析限制在子总体，比如限制在年长妇女这一群体。使用SUDAAN之前，使用把整个样本调查数据集输入到SUDAAN而不是把输入数据集变成有关子总体子集的子集SUBPOPN语句，这是因为，由于子数据集中一些样本初级抽样单位有可能缺失，SUDAAN可能会导致不正确的估计标准误差。

3. CROSSTAB

101. CROSSTAB程序仅用于分类变量。CROSSTAB中的TABLES语句表示估计人口百分比和总数的单向、双向或多向表格，用于TABLES语句的所有变量需要相应的SUBGROUP和LEVELS语句。

102. CROSSTAB中TEST语句要求卡方检验，用于检验两个分类变量在统计上是独立的这一零假设。一项卡方检验基于皮尔逊型检验（CHISQ），并采用有关估计总体总数的“观察值减去预期值”计算方法。另一卡方检验基于估计

的总体比率 (LLCHISQ)。有置信区间的机会比率和相对风险 (实际上是流行率) 是通过使用PRINT语句上的RISK = ALL为2x2表估计的。最后, Cochran-Mantel-Haenszel检验 (使用TEST语句上的CMH) 评估两个变量的统计独立性, 同时控制另一个变量的 (“分层”)。

4. 数值例子

103. CD-ROM中附件的第3部分论述了使用CROSSTAB和DESCRIPT来操作第60段所举的三个例子, 利用SAS-CALLABLE SUDAAN (SAS V8.2和SUDAAN V8.0)。方差估计使用泰勒级数线性化方法和BRR (平衡多次重复) 两种方法。对有解释的SUDAAN程序 (用户写程序) 和有解释的SUDAAN输出进行评价, 能够为读者在写自己的SUDAAN程序和解释输出时提供帮助。附件第3部分只包含在TABLES1、3、4、5和6的讨论中选出的SUDAAN分析。

104. 表二十一.1第4行概述了 (附件) 第3部分的CROSSTAB和DESCRIPT, 用于估计布隆迪推断总体中血清反应为阳性的妇女百分比和妇女人数, 并且有估计的标准误差。从SUDAAN得到的CROSSTAB和DESCRIPT对方差估计的某一给定方法来说是一致的 (和预期的一样), 泰勒级数线性化和平衡多次重复结果也是一致的 (但并非总是如此)。SUDAAN结果和SAS SURVEYMEANS的结果一致。要注意, CROSSTAB和DESCRIPT不为估计的人口百分比或总数计算置信区间。

105. (CD-ROM) 附件第7部分的表二十一.3中第2行显示了从用于估计血清反应为阳性百分比的CROSSTAB和DESCRIPT所得的一致输出 (不论使用了泰勒级数或平衡多次重复), 但用于农村和城市妇女两个域。SUDAAN CROSSTAB和DESCRIPT结果与SAS SURVEYMEANS一致。

106. (CD-ROM) 附件第7部分表二十一.4第2行概述了DESCRIPT输出 (使用泰勒级数和平衡多次重复), 该输出使用线性对比, 比较了血清反应为阳性的农村和城市妇女百分比。使用泰勒级数和平衡多次重复的估计标准误差有些不同, 但可以忽略。结论是布隆迪推断总体中城市和农村妇女在血清反应为阳性的流行率方面是不同的。要注意, 比较这两个域时, DESCRIPT线性对比结果与使用SAS SURVEYREG时是一致的。

107. (CD-ROM) 附件第7部分表二十一.5第1、2行显示CROSSTAB中两种不同卡方检验的结果, 这两种卡方检验是Pearson (CHISQ) 和log-linear (LLCHISQ)。使用泰勒级数和平衡多次重复的结果是一致的。城市妇女中估计的血清反应为阳性流行程度比农村妇女高出许多 (使用CHISQ), 城市妇女血清反应为阳性的估计率比农村妇女高出许多 (使用LLCHISQ)。

108. (CD-ROM) 附件第7部分表二十一.6第1行显示了血清反应为阳性的估计机会比率 (0.393) 和流行率 (0.797) (农村妇女与城市妇女之间的比), 各有95%的置信区间。在机会比率95%置信区间的上限中, 其差距可以忽略不计。估计的机会比率和流行率在量值上是不同的, 因为血清反应为阳性的流行率不低。

5. 优点/缺点/费用

109. SUDAAN是一种综合样本调查（和相关数据）软件包，具有描述分析和建模分析的分析能力。它具有估计和检验用户规定的总体参数对比矩阵（包括回归系数）的能力，SUDAAN可以在主机和PC环境中运行。SAS用户往往在学习SUDAAN方面具有优势，因为其语法与SAS相似。但是，SUDAAN的一些语法是深奥的，或许比其他软件包需要更多的学习时间。

110. 与本章评价的其他软件包的费用相比，SUDAAN的费用很高，特别是用作与SAS连接的SUDAAN时，因为此时也需要SAS。为经过许可的用户提供技术支持。SUDAAN V8.0用户指南基本上可以用作参考书，而不是学习SUDAAN的材料，里面有一些详细的NHANES - III（国家卫生和营养调查 - III）数据注解分析范例，而这些数据有助于学习如何使用SUDAAN。

111. 此处论述的软件包中，只有SUDAAN包括两种主要的方差估计方法，即泰勒级数线性化方法和重复方法。但是，SUDAAN对平衡多次重复（BRR）并不构成重复加权，需要用户提供这些加权。SUDAAN对刀切删除一程序构造重复权数，如果数据集中包括这些刀切重复权数，SUDAAN将接受这些权数。

112. 此处评价的软件包中，也只有SUDAAN具有为综合方差估计描述抽样阶段、分层和 fpc 项的广泛功能。另外，此软件包还为设计效果计算提供几个不同的定义，可以让人把过多抽样和/或不等加权的效应排除在设计效果之外。

113. 在SUDAAN中输入ASCII数据很麻烦，这就使人们更倾向于另外两种数据输入选择，即SAS或SPSS数据集。输入独立的SUDAAN的SAS数据集一定是SAS V6.04或SAS传送文档。可连接SAS的SUDAAN可以读出SAS能读的任何数据集。SUDAAN输出可以使用电子方式保存到SAS数据文档格式中，以备在SAS或在像EXCEL这样的电子表格软件中使用。SUDAAN对变量重新编码的功能非常有限，没有数据管理功能。这样，使用SUDAAN之前，在SAS或SPSS（取决于输入数据集的类型）进行必要的重新编码或形成新的变量时要非常谨慎。

I. STATA 7.0中抽样调查程序

1. STATA概述

114. STATA是一种普通的统计软件包，于1995年开始为样本调查数据分析增添了更广泛的功能。此处论述了STATA 7.0；V8.0是2003年发布的。只有泰勒级数线性化方法可以用于方差估计。普通的抽样计划描述WR是默认值。对于没有替代抽样计划（简单随机抽样和分层随机抽样）单阶段方差估计和没有替代整群抽样（分层或不分层）但每个分层中的类集（初级抽样单位）使用等概率抽样且每个抽样的初级抽样单位中所有元素都包含在样本中的单阶段方差估计，STATA可以把 fpc 项纳入方差估计。

115. STATA样本调查分析的范围比SUDAAN更广，具有用户规定的总体参数对比矩阵数学统计功能，上述参数包括回归系数。STATA与简短命令互动运

行，这样，此软件包就相对容易学习。但是，如有必要，用户写程序可以以批处理的形式提交。STATA对区分大小写，应用小写字母将使命令键入STATA。STATA分配内存的默认值，在内存中装载输入数据集的拷贝。如果该内存对大数据集来说显得不够，可以通过集贮存命令扩大内存。

116. STATA中样本调查命令以名称svy（代表“调查”）开始。描述性命令用来估计人口均值（svymean）、总体总数（svytotal）、总体比例（svyprop）以及双向表格中的百分比和总数（svytab）。Svytab中的总体比例置信区间使用了分对数转换，这样，估计的下限和上限被限制在(0, 1)内。Svytab内有双向表格中为样本调查数据而做的8个不同卡方检验。模拟程序包括线性回归、对数回归（包括带有一个额定或有序变量的多项式）、泊松回归和概率单位模型。

117. Svyset命令用来把抽样计划指定到STATA上。为描述一般抽样计划WR（默认值），命令svyset的三个关键词交互输入STATA，在输入分层变量名称之前，输入关键词strata，在输入初级抽样单位变量名称之前，输入关键词psu，在输入抽样加权变量名称之前输入关键词pweight。这样，布隆迪调查的抽样计划以下方式描述到STATA V7:

```
svyset    strata    pstra
svyset    psu      ppsu
svyset    pweight   w2
```

118. 如前面针对SAS中样本调查程序所示，STATA中省略strata关键词表示，在初级抽样之前没有对PSU进行分层。省略psu关键词表示包括一个阶段的元素抽样，而且没有抽样的元素类集。省略pweight关键词表示加权相等的抽样元素，加权变量的默认值为1.0。svyset语句在STATA V8中得以修改。

119. 命令svydes指导STATA输出它已经附到数据集（经svyset命令获得）中的调查设计变量，并汇总分层数、每个分层的PSU数、每个分层内部每个初级抽样单位的观测值均值。这是对样本调查设计特点的极有用的总结。

2. SVYMEAN、SVYPROP、SVYTOTAL、SVYLC

120. svymean命令估计连续变量或编码为1或0的指示变量的总体均值（就是说，估计的总体比例）。输出选择包括估计的标准误差、变异的估计系数和总体参数的设计效果和置信区间。

121. svyprop命令用于分类数据：它估计处于每级分类变量的总体比例，并有估计的标准误差。与svymean相比，svyprop的输出选择较少。

122. svytotal命令估计连续或指示(0, 1)变量的总体总数，有估计的标准误差、变异的估计系数、设计效果和置信区间。

123. 上述三种命令中每一种都可以用来估计范围的总体参数，采用的方法是使用命令行上的by选择，如使用by (stra)或使用by (urb_rur)分析布

隆迪农村和城市妇女的两个范围。STATA根据其svy命令中的by语句，使用正确的范围方差估计公式。

124. 另外，上述三个命令中每一个都可以和命令行上的subpop选项一起用于子总体的总体参数估计，比如年长妇女的总体参数估计。在分析子总体时不要使用STATA中的“if”语句，因为这样做，估计的方差可能会不正确；应该使用subpop选项。

125. svylc命令估计用户规定的范围均值线性组合、比例或总数，以及估计的标准误差、t-检验、p-值及置信区间。该命令可以用来对范围进行相互比较。在V8.0中，svylc命令由lincom替代。在V8.0中，命令svylc继续起作用，但不再记入文档。

3. SVYTAB

126. STATA中的svytab命令用于双向表格，估计带有估计标准误差的人口百分比（行、列或总数）、带有估计标准误差的表格单元格的总体总数以及置信区间。使用分对数获得总体比例的置信区间，这样，估计的下限和上限就限制在区间(0, 1)。有8个不同的卡方检验用来检验表格中两个分类变量的统计独立性的零假设。使用svytab时可使用命令subpop。

4. 数值例子

127. (CD-ROM) 附件第4部分论述使用STATA命令对第60段列出的三个例子进行操作。每个操作的例子是与STATA交互使用的日志文件。对加以注释说明的STATA日志的评价（用户命令和STATA输出）可以帮助读者更好地使用STATA中样本调查命令，更好地对输出进行解释。

128. svymean和svytotal两种命令与指示变量BLOOD（1 = 血清反应为阳性，0 = 血清反应为阴性）一起使用。表二十一.1（第5、6行）显示了血清反应为阳性的妇女估计人数和百分比，具有置信区间。STATA计算与SAS SURVEYMEANS、SUDAAN DESCRIPT及CROSSTAB一致。

129. (CD-ROM) 附件第7部分表二十一.3（第3行）显示了农村/城市地区血清反应为阳性的妇女百分比。STATA svytab点估计值和估计的标准误差与SAS SURVEYMEANS及SUDAAN DESCRIPT和CROSSTAB一致。但是，STATA svytab和SAS SURVEYMEANS之间的范围置信区间有细微的差别，因为STATA svytab使用了分对数转换以获得置信区间。

130. (CD-ROM) 附件第7部分表二十一.4（第3行）显示了血清反应为阳性的农村妇女和城市妇女在百分比方面的线性对比的STATA svylc结果，说明两个范围之间存在很大差别。STATA结果与SUDAAN DESCRIPT一致，在两个范围比较方面，与使用SAS SURVEYREG的结果一致。

131. (CD-ROM) 附件第7部分表二十一.5（第3-5行）显示了关于血清反应为阳性在统计上独立于农村/城市地区的零假说三次卡方检验的STATA

svytab结果。三次svytab卡方检验拥有相似的（而且很小的） p -值。STATA svytab（第3行）默认值卡方检验是Rao和Scott（1981年；1984年）提出的皮尔逊型卡方检验，进行了二阶校正。Svytab（第4和5行）中另外两次卡方检验与SUDAAN CROSSTAB中的卡方检验相同，STATA和SUDAAN对这两次检验的计算结果相同。

132. 因为STATA中svytab命令不产生机会比率或流行率比率，命令svylogit曾被用来估计血清反应为阳性的机会比率（城市比农村）。带有置信区间的STATA机会比率，见（CD-ROM）附件第7部分表二十一.6（第2行）所示。STATA svylogit命令计算结果与使用SUDAAN CROSSTAB进行点估计和置信区间估计的结果相同。

5. 优点/缺点/费用

133. STATA是一种综合的一般统计分析软件包，对样本调查数据具有广泛的分析能力，包括描述和基于设计的建模程序。该统计分析包作为统计软件包，已经受到好评，相对容易学习，拥有活跃的用户群。与本章评价的其他软件包相比，其费用适中。

134. STATA接受估计总体参数的用户定义的对角矩阵，包括回归系数，适合希望检验自己的特定假设或总体参数估计组合的人。一般说来，该软件使得那些拥有必要数学统计背景的人在进行统计分析时有很大的灵活性。

135. STATA仅使用泰勒级数线性化方法，局限于普通抽样计划描述WR。但是，在没有替代抽样的方差估计中，STATA包括元素单一阶段抽样和单一阶段整群抽样 fpc 项。抽取STATA分析结果（例如，未加权的样本大小、点估计值和标准误差）输入到其他数据格式上具有一定的难度，但却是可能的。

J. Epi-Info 6.04d和Epi-Info 2002样本调查程序

1. Epi-Info概述

136. Epi-Info已由疾病防治中心（CDC）和世界卫生组织（WHO）开发多年，该软件可以免费从疾病防治中心（CDC）网站<http://www.cdc.gov/epi-info/>上下载。

137. Epi-Info有两个版本：上一个DOS驱动Epi-Info V6.04d和最新Windows驱动Epi-Info 2002。

138. Epi-Info功能包括编写调查问卷或研究数据收集表格、专用数据输入、数据分析和文字处理。其分析和统计功能针对的是全世界的流行病学家。Epi-Info分析的输出（分析结果）可以发送到屏幕、打印机或电子文档。

139. Epi-Info的两个版本（DOS或Windows）具有对复杂样本调查数据进行基本描述分析的功能。只有普通抽样计划描述WR。输入数据集必须通过两到三个调查设计变量进行分类：分层变量STRATVAR和分层内部的初级抽样单位

变量*PSUVAR*。Epi-Info不会把任何*fpc*项纳入方差估计。而且，该软件包不估计总体总数，泰勒级数线性化方法用于方差估计。

140. Epi-Info具备用于复杂样本调查数据的分析功能，起初开发这种功能的目的是用于行为风险因素监视系统（BRFSS），这是由疾病防治中心（CDC）发起的美国各州年度卫生样本调查计划（Brogan, 1998年及印刷中），二是用于世界卫生组织（WHO）类集样本方法，这是扩大免疫方案（EPI）在全球使用的方法，用来估计儿童免疫覆盖率（Brogan等, 1994年）。但是，Epi-Info中的样本调查程序可以用于任何复杂样本调查，只要该调查可以用普通抽样计划描述WR进行描述。

2. Epi-Info V6.04d (DOS)、CSAMPLE模块

141. DOS Epi-Info是疾病防治中心（CDC）和世界卫生组织（WHO）共同开发的。Epi-Info 6.04d的数据输入是dBase文档或ASCII文档，Epi-Info把这种文档转换成Epi-Info数据文档*.rec。软件包就是把SAS或SPSS或其他类型的数据文档转换成Epi-Info*.rec文档，例如DBMS-COPY (<http://www.dataflux.com/conceptual/>)。Epi-Info 6.04d作为一种交互式程序运行，不能以批的模式运行。使用旧电脑、旧操作系统和/或硬盘存储空间有限的人更喜欢使用DOS版，而不是Windows版。

142. Epi-Info V6.04d中的CSAMPLE模块对复杂样本调查数据进行分析。CSAMPLE估计人口均值（针对连续变量或编码为1/0的指示变量）总体百分比（针对分类变量），以及估计的标准误差、置信区间和设计效果，还为通过各级分类变量形成的范围提供估计值。另外，CSAMPLE估计范围均值或范围百分比之间的差，具有相应的估计差的估计标准误差以及人口差的置信区间。CSAMPLE为2 x 2表估计机会比率和风险比率。注意，CSAMPLE不估计总体总数。

143. 如果Epi-Info 6.04d的CSAMPLE模块打开，在由用户指定分析所使用的变量的地方出现数据输入屏幕。用户为三个调查设计工具箱中的每一个工具箱选择一个变量：STRATA（分层变量）、初级抽样单位（初级抽样单位或类集变量）和WEIGHT（抽样加权变量）。对于布隆迪的调查，Epi-Info的规格如下：

STRATA	<i>PSTRA</i>
PSU	<i>PPSU</i>
WEIGHT	<i>W2</i>

144. 用户在叫作MAIN的工具箱中规定分析变量（或因变量），该变量可以是连续变量，如*IUML*，或分类变量，如*IMMUNE*。如果MAIN中规定的连续变量（或假定连续的变量）需要估计人口均值，那么，用户点击MEANS选项。如果MAIN中规定的分类变量需要估计人口百分比，那么，用户点击变量TABLE选项。

145. 如果各个范围需要估计均值或百分比,那么,定义这些范围的变量在叫作CROSSTAB的工具箱中进行规定,分析变量则在MAIN工具箱中规定。

146. 另外,CSAMPLE可以就分析变量的均值估计两个范围之间的差。用户可以规定定义两个可比范围的两个CROSSTAB变量级别。

3. Epi-Info 2002 (Windows)

147. Epi-Info 2002是Windows应用软件,由疾病防治中心(CDC)开发。Epi-Info 2002数据分析的数据输入通过MicroSoft Access 1997文档(*.mdb)或dBase文档进行。Epi-Info 2002还可以读出为Epi-Info的DOS版编制的*.rec文档。软件交互运行,但有以批模式运行的选项。

148. Epi-Info 2002有三个复杂样本程序,位于Advanced Statistics(高级统计学)下的Analyze Data(分析数据)部分。Complex Sample Frequencies(复杂样本频率)为分类变量估计单向百分数分布,有估计的标准误差和置信区间。Complex Sample Tables(复杂样本表)为双向分类变量表计算行和列百分比[标有陈述(行)和结果(列)],行的百分比有估计的标准误差和置信区间。如果该表是2 x 2,该程序还估计机会比率和风险比率,有置信区间。Complex Sample Means(复杂样本均值)估计连续变量的均值,有估计的标准误差和置信区间,包括分类变量形成的范围均值估计。如果域变量分两级,还可估计范围均值的差,有估计的标准误差和置信区间。

149. 在所有三个复杂样本程序中,调查设计变量在三个工具箱内确认,这三个工具箱分别标有Weight、初级抽样单位和Stratify By(样本调查分层变量)。为获得估计的标准误差和置信区间作为输出,双击OPTIONS:SET,然后选择Statistics = Advanced。

4. 数值例子

150. (CD-ROM)附件第5部分论述了使用Epi-Info 6.04d中的CSAMPLE操作第60段所举的例子。每个操作的例子包括从Epi-Info获得的输出,用说明进行解释。对加以注释的输出进行评价有利于读者解释CSAMPLE输出。

151. 表二十一.1(第7行)给出了血清反应为阳性的妇女百分比的Epi-Info 6.04d估计值。Epi-Info点估计值和估计的标准误差与SAS SURVEYMEANS、STATA svymean和SUDAAN DESCRIPT及CROSSTAB一致。血清阳性反应率的95%置信区间比SAS SURVEYMEANS和STATA svymean得出的置信区间窄。这是因为,Epi-Info在其95%置信区间计算中使用 $z = 1.96$,而不是30 df的Student-t值2.042,后者是布隆迪调查分母自由度[初级抽样单位数(60)减去假层数t(30)]。

152. (CD-ROM)附件第7部分表二十一.3(第4行)给出了农村/城市地区血清阳性反应率的Epi-Info估计值。Epi-Info点估计值和估计的标准误差与SAS SURVEYMEANS、STATA svytab和SUDAAN DESCRIPT及CROSSTAB一致。与

SAS SURVEYMEANS和STATA svytab的结果相比, Epi-Info范围置信区间要窄, 因为Epi-Info使用了 $z = 1.96$ 。

153. (CD-ROM) 附件第7部分表二十一.4 (第4行) 给出了比较农村和城市妇女血清阳性反应率的Epi-Info线性对比。估计的对比值(-16.99%)与SAS SURVEYREG、SUDAAN DESCRIPT和TATA svylc一致。Epi-Info没有给出估计差的估计标准误差, Epi-Info给出的关于对比值的95%置信区间是错误的。

154. (CD-ROM) 附件第7部分表二十一.6 (第3行) 给出了Epi-Info估计机会比率(城市比农村)以及血清阳性反应率的估计值, 置信区间为95%。Epi-Info点估计值与SUDAAN CROSSTAB和STATA svylogit完全一致, Epi-Info置信区间与SUDAAN和STATA非常接近。

5. 优点/缺点/费用

155. Epi-Info一个主要优点就是费用: 可以从疾病防治中心(CDC)的网站免费下载。另外, DOS和WINDOWS操作系统均可使用, 需要运行Epi-Info的硬件和软件有很大的灵活性。Epi-Info样本调查功能当然会吸引那些进行其他类型流行病学或统计分析的Epi-Info用户。

156. Epi-Info仅使用泰勒级数线性化方法, 仅处理普通抽样计划描述WR。DOS版本的CSAMPLE模块和Windows版本中的CSAMPLE模块(Advanced Statistics下的三个程序)足以进行复杂样本调查数据的基本描述统计。这包括整个人口和范围的人口均值或百分比的估计。Epi-Info没有估计总体总数、进行卡方检验、把 fpc (有限总体修正)项纳入方差估计之内或进行基于设计的建模分析(如对数回归或线性回归)的样本调查能力。

K. WesVar 4.2

1. WesVar概述

157. WesVar是专门用于样本调查数据分析的软件包, 方差估计使用重复方法(Rust和Rao, 1996年): 平衡多次重复(包括可选择的Fay因素)以及三种刀切变异。WesVar没有泰勒级数线性化方法的功能。使用平衡多次重复的样本调查设计有几个分层, 每个分层确实有两个样本初级抽样单位。刀切法如泰勒级数线性化方法一样, 可以用于每个分层样本初级抽样单位的任何数字($> = 2$)的设计。

158. WesVar默认抽样计划是前面提到的普通抽样计划WR。WesVar具有把 fpc 项纳入方差估计的功能, 但仅用于刀切法和元素的单一阶段抽样。

159. WesVar 4.2可以读以下类型的输入数据集: DOS的PC-SAS、SAS传送、SAS (V6-8)、SPSS、STATA、ASCII和使用ODBC的文档, 如Microsoft Excel或Access。根据假定的普通抽样计划WR, 如果需要建立重复加权, WesVar需要每个观测值的分层、初级抽样单位和加权变量。但是, 一旦文档上有重复加权, 就不需要初级抽样单位和分层标识符: 这是重复方法针对公用文

档所具有保密优点。在评价的所有软件包中，只有WesVar能为不答复、事后分层和多变量重复加权调节基本的调查权数。在完成输入数据集的准备工作之后，被保存为WesVar (*.var) 文档，用于数据分析和以后的数据管理。

160. 可以进行一系列描述性统计：估计人口均值、百分比、百分等级和总数，这些均有估计标准误差，还可以估计变化系数、置信区间和设计效果。WesVar的特别优势和总的重复方法是能够获得用户规定的总体参数功能的点估计值（有估计标准误差），比如普及比率。WesVar中有基于设计的回归分析：线性、对数和多项式对数。

161. 从WESTAT网页可以下载WesVar V4版，试用期为30天，WesVar V2版可以从网页上下载，免费无限期使用（见<http://www.westat.com/wesvar>）。与V2版相比，WesVar V4版能接受更多的输入数据集，具有更佳的文档处理和数据管理功能，能调节不答复重复权数，包括更多分析选择项。用户开始可以先使用WesVar V2版，然后在必要时升级至V4版。

2. 使用WesVar V4.2版

162. 在Windows环境中，用户通过弹出菜单使用WesVar。WesVar软件打开后，第一个菜单包括4个选择项。第一个选择项是新WesVar数据文档，(1) 读入非WesVar数据集的数据集；(2) 创建重复权数或接受输入数据集中已经存在的重复权数；(3) 对变量进行重新编码、转换、加上标识和格式；(4) 进行事后分层、多变量重复加权和答复调整；(5) 定义用于分析的子总体；及(6) 按请求修改默认值ddf，然后将数据集保存为WesVar文档。第二个选项是开放式WesVar数据文档，该选项读入WesVar数据文档，完成上述6个操作程序。

163. 第三个选项是New WesVar Notebook，接受WesVar数据文档分析请求，运行该请求，显示输出，在WesVar系统中一个“笔记本”内保存上述要求和输出，用于组织所请求的分析和所得输出。需要两类分析中的一类：表格或回归（线性、对数或多项式）。在完成表格或回归选择后，有许多选项可以用来对分析进行规定。浏览菜单屏幕进行分析以及读出输出并不是直接的，WesVar指南提供了若干有用的例子，说明如何进行菜单浏览和组织输出。

164. 如果以前的WesVar的请求和输出被保存在笔记本中，那么，就可以选择第一个菜单上的第4个选项：开放式WesVar Notebook。可以在现成的笔记本上加入新的分析要求并保存。与特别WesVar数据文档有关或与特别项目有关的所有分析都可以组织到一个或多个笔记本中。

165. WesVar中5种重复法之一一定要进行特别规定，以便建立重复加权或确认输入数据文档中已经存在的重复权数。这些重复方法有：

- (1) 平衡多次重复 (BRR) —— 每个分层确有两个样本初级抽样单位；
- (2) Fay氏与平衡多次重复扰动方法 (FAY)；
- (3) 刀切删除一，无明显分层 (JK1)；

- (4) 刀切，每个分层确有两个样本初级抽样单位（JK2）；
- (5) 刀切，每个分层有两个或多个样本初级抽样单位（JKn）。

166. WesVar用户指南中的附件A和D详细概述了这五种重复法，通过例子说明如何把不同的抽样计划转变成五种方法之一。

3. 数值例子

167. 由于布隆迪输入数据集没有包含重复加权，所以，当时有必要选择五种重复方法之一，然后要求WesVar计算出重复加权。WesVar需要的布隆迪调查设计变量为：*PSTRA*、*PPSU*和*W2*。由于布隆迪抽样计划是*WR*，共有30个假层，每个分层确有两个样本初级抽样单位，所以，平衡多次重复或JK2是最佳选择。选择平衡多次重复，没有Fay扰动因素。另外，不用进行不答复调整或事后分层或多变量重复加权，因为使用泰勒级数线性化方法时，整个数据集中没有进行这些调整。

168. (CD-ROM) 第6部分说明了使用WesVar操作第60段中列出的三个例子。每个操作的例子包括从WesVar 4.1或4.2的输出，但没有说明请求的分析所使用的输入菜单屏幕。对加注解的WESVAR输出进行评价有助于读者对WesVar输出做出解释。

169. 表二十一.1第8行显示，WesVar在估计百分比和血清反应为阳性的妇女人数方面与其他所有样本调查软件包一致（有标准误差）。WesVar置信区间与SAS和STATA一致，但与Epi-Info不一致，因为后者置信区间太窄。

170. (CD-ROM) 附件中表二十一.3第5行说明，在范围点估计值和估计标准误差方面，WesVar与其他所有软件包一致。WesVar置信区间与SAS SURVEYMEAS置信区间非常接近，但与STATA *svytab*（使用分对数转换）和Epi-Info（使用 $z = 1.96$ ，而非Student *t*-值）有细微差距。

171. (CD-ROM) 附件第7部分表二十一.4第5行显示WesVar线性对比结果，以比较农村和城市妇女血清阳性反应率。在线性对比和Student *t*-统计的估计标准误差方面，WesVar与SAS SURVEYREG、SUDAAN DESCRIPT和STATA *svylc*一致。关于线性对比的置信区间，SAS、STATA和WesVar之间存在差距，但可以忽略不计。

172. (CD-ROM) 附件第7部分表二十一.5第6和7行显示了WesVar中实施的复杂样本调查数据两个Rao/Scott卡方检验。这些计算结果与其他软件包的其他卡方检验不完全一致。

173. (CD-ROM) 附件第7部分表二十一.6第4行显示，WesVar对数回归程序产生的估计机会比率和置信区间与SUDAAN CROSSTAB和STATA *svylogit*产生的估计机会比率和置信区间一致或基本一致。表二十一.6第5行显示，WesVar估计的普及性（使用了TABLES中的单元格功能）与SUDAAN CROSSTAB和Epi-Info一致，SUDAAN和WesVar之间的置信区间差可以忽略。

4. 优点/缺点/费用

174. WesVar仅使用重复方法进行方差估计。带有重复加权的公开发布的数据集二级分析人员不必了解样本设计的细节（如调查设计变量 $STRATVAR$ 和 $PSUVAR$ ），但他们确实需要对WesVar指定用于获取重复权数（从调查文件中获得信息）的方法。如果用户需要使用WesVar建立样本调查数据集的重复加权，就需要对重复方法有一定的了解，另外，要有与普通抽样计划 WR 相关的三个调查设计变量（分层变量 $STRATVAR$ 、分层内部初级抽样单位变量 $PSUVAR$ 和样本加权变量 $WTVAR$ ）。

175. WesVar具有建立样本调查数据集重复加权的广泛功能。共有五种不同的重复方法，包括不答复调整、进行事后分层或多变量重复加重的机会。另外，WesVar可以选择把有限总体修正项用于单一阶段抽样，使用刀切法进行方差计算。

176. 对于不熟悉方差估计重复方法的用户来说，WesVar用户指南附件A对重复方法的理论和实践给出了清楚细致的概述，当然，阅读该材料需要一些数学统计背景。另外，用户指南的附件D也给出了有用的指导和若干范例，说明如何针对给定的抽样计划选择重复方法。

177. WesVar能够估计用户定义的总体参数函数，而如果使用泰勒级数线性方差估计方法，会更加困难。这样，在能够估计的总体参数方面，WesVar与本章评价的其他软件包相比，在本质上更为灵活。尽管SUDAAN具有平衡多次重复和刀切重复方法，能够用于方差估计，但SUDAAN不像WesVar那样允许用户规定需要进行估计的总体参数的函数。

178. 与其他大多数样本调查软件相比，从WesVar直接输出，操作上有些困难。对所要求表格的每一个单元，WesVar输出都会给出其中的一行（如附件第6部分所示）。但是，可以从WesVar网站上免费下载Table Viewer实用程序。该实用程序把WesVar 4输出转换成网格或列表形式，以在屏幕上显示或进行打印，还可制作这种形式的电子文档，粘贴到Microsoft Word或Excel等应用软件中。

179. 与本章评价的其他所有软件包相比，WESVAR的费用低廉。V4版可以免费下载，试用期为30天，V2版也可以免费下载，无限制使用。

L. PC-CARP

180. PC-CARP是一个独立的MS-DOS程序，由衣阿华州立大学（统计系）开发和提供。该程序处理上文讨论的普通抽样计划 WR ，对于更为简单的设计，可以将 fpc 项纳入两个抽样阶段。采用泰勒级数线性化方法估计方差。

181. 点估计值、估计标准误差和置信区间是为以下方面构造的：总体和子总体总数、均值、比例、分位数、经验分布函数、比率和比率差（以及由此导致的均值、比例和总数的差）。其中还包括基于设计的线性回归和双向列表分析，包括一项卡方检验。此外还有三个附加模块：PC-CARPL用于以基于设

计的对数回归，POSTCARP用于样本调查数据的事后分层，EV CARP用于利用解释性变量的量度误差进行回归分析。

182. 用户界面是通过键盘操纵的文本菜单屏幕；不支持鼠标。只有ASCII码文件才能被作为输入接受，输入记录的空间有限或长度固定，FORTRAN语法中有支持性的格式语句。数据集对观测数据的数量没有限制，大部分分析可以接受多达50个的变量。PC-CARP能够在安装有DOS 5.0及以上版本的老式计算机系统上运行，也可以在Windows 3.1x或Windows 95及更高版本的操作系统上运行。该程序仅占用3 Mb的硬盘空间和450 Kb的随机存取存储器空间。所有的新系统必须能够支持DOS程序以运行PC-CARP。

183. 与评价的其他软件包相比，PC-CARP的一次性购买价格很低。无需每年的升级费用。购买三个附加模块的每一个，所需的费用都很少。

184. 本章没有提供对使用PC-CARP的布隆迪调查进行的示例分析。

M. CENVAR

185. CENVAR是一个名为微型计算机综合处理系统（IMPS）的综合统计软件系统的一部分。IMPS由美国普查局设计，用于处理、管理和分析综合样本调查数据。包括CENVAR在内的IMPS可以免费从以下地址下载：<http://www.census.gov/ipc/www/imps/download.htm>。截至2003年初，部分IMPS基于Windows，部分仍然基于DOS。本章不包括对IMPS的讨论。

186. CENVAR根据PC-CARP改写而成，因此具有PC-CARP的许多特点。CENVAR支持的样本设计与PC-CARP相同，即普通抽样计划WR，以及在无替代抽样的更为简单的一阶段和两阶段设计应用中将fpc项纳入方差估计。在方差估计中使用泰勒级数线性化方法。软件为菜单驱动，没有鼠标支持。

187. 点估计值、估计标准误差、置信区间、变异系数以及设计效果是为以下方面构造的：总体和子总体总数、均值、比率和比率差（以及由此导致的均值、比例和总数的差）。不包括PC-CARP中的其他选项，即以基于设计的线性回归、双向列联表分析和分位数估计。CENVAR中没有PC-CARP的三个附加模块。

188. 《CENVAR用户指南》（1995年版）的篇幅长约100页，可以从网上下载。该指南包含的有用例子和培训练习来自普查局进行的三项样本调查。数据字典必须在运行CENVAR之前创建。因此，为了使用CENVAR，必须熟悉IMPS。CENVAR在安装有DOS 3.2或更高版本的个人电脑上运行。该程序要求10 Mb的硬盘存储空间和640 Kb可用内存。本章没有给出布隆迪调查中使用CENVAR的例证分析。

N. IVEware (Beta版)

189. IVEware（估算和方差估计软件）是密歇根大学调查方法项目最近开发的一个可调用SAS的样本调查数据应用软件。该程序处理普通抽样计划WR，并根据程序使用泰勒级数线性化方法或重复方法。

190. IMPUTE模块利用多变量序贯回归法来估计项目遗漏值,包括多重估计数据集。DESCRIBE模块估计总体和子总体的均值和比例、子群差异以及均值和比例的线性对比;使用了泰勒级数线性化方法。REGRESS模块适用于若干个基于设计的回归模型(线性、对数等);使用了刀切重复技术。SASMOD模块允许用户在使用CATMOD、GENMOD和MIXED等SAS PROCs进行数据分析时考虑复杂样本设计的特点。对三个数据分析模块(DESCRIBE、REGRESS和SASMOD)可以进行多重分配分析。

191. IVEware可以在SAS V6.12或更高版本的机器上运行,也可以在安装了微软Windows操作系统或Linux操作系统的个人电脑上运行;此外,还可以在其他平台上运行。尽管用户无需熟悉IVEware的SAS宏语言、C语言和FORTRAN语言程序块,但是他们需要具有中等程度的SAS以及SAS软件使用经验。IVEware软件和文件可以从以下地址免费下载:<http://www.isr.umich.edu/src/smp/ive/>。本章没有给出布隆迪调查中使用IVEware的例证分析。

0. 结论和建议

192. 有些数据分析人员可能会对在复杂样本调查数据方差估计中要使用专用软件感到吃惊。尽管有些数据分析人员想利用为简单随机样本开发的软件进行复杂样本调查数据的方差估计,但是我们并不建议这样做。现在有几种软件可用于方差估计,其中有些还是免费的。在这些软件中进行选择的理由主要是对软件的熟悉程度、成本、易用性以及是否仅仅对基本的描述性分析或更加综合的分析感兴趣。

193. 如果你已经使用了一个具有样本调查方差估计功能的普通统计软件包,则该软件包当然是首选,因为已经支付了购置成本而且熟悉该软件的语法。STATA用户在使用该软件包过程中已经具有了综合样本调查方差估计的能力,因此其用户无需寻找别的软件,除非要分析的数据集必须使用重复方法。与SAS 8.2版相比,使用最近发布的SAS 9.0版的用户加强了样本调查方差估计的能力,预计这方面的能力将来还能得到进一步加强。但是,如果SAS V9.0不足以满足样本调查方差估计的目的,那么同时使用免费的IVEware软件包和SAS可以满足你的需要。Epi-Info用户在使用该软件包时只具有基本的样本调查数据方差估计能力,不过,如果这可以满足你的需要,该软件就足够了。SPSS是一种广泛使用的统计分析软件包,2003年底发布了一个复杂样本调查附加模块,因此,该软件现在也是一个可行的选择。

194. 如果普通统计软件包不具备必要的样本调查方差估计能力,就需要考虑专门的样本调查软件包(例如,WesVar、SUDAAN、PC-CARP或CENVAR)或者其他普通统计软件包(例如带有/不带IVEware或SPSS或Epi-Info的STATA或SAS)。SUDAAN往往对SAS用户具有吸引力,因为其语法与SAS相似,还可以作为与SAS相联的SUDAAN运行,虽然使用独立的环境,但是它也接受SPSS输入的数据集。WesVar、PC-CARP和CENVAR都是独立的程序,具有各自独立的组织,因此熟悉其他统计软件包并不会影响对这三者的选用。PC-CARP和CENVAR可能

对那些必须或喜欢在DOS环境中操作的用户具有吸引力，对那些喜欢使用Windows环境的用户可能缺乏吸引力。

195. 如果费用是影响选择软件的主要因素，有些软件包显然具有这方面的优势。Epi-Info虽然免费，但是在样本调查方差估计的分析性项目方面受到限制，不过在基本分析方面还是很不错的。CENVAR也是免费的，而且具有比Epi-Info更丰富的分析选项，可是没有基于设计的回归程序。WesVar V2也是免费的。IVEware是免费的，可是必须与SAS一起运行。低成本的综合样本调查软件包括WesVar V4和PC-CARP。STATA和独立的SUDAAN费用低廉，而SAS则很贵。

196. 影响软件选择的另一个因素可能是使用的方差估计方法。例如，如果你正在分析一个包括平衡多次重复或刀切重复加权但没有层/初级抽样单位标识变量的公开数据集，那么仅仅使用泰勒级数线性化方法的软件包对你来说没有用处。在这里所评价的软件包中，SUDAAN和IVEware同时提供泰勒级数线性化方法和重复法，WesVar仅提供重复程序，而STATA、SAS、PC-CARP、Epi-Info和CENVAR则仅仅具有泰勒级数线性化功能。

197. 最后，软件的选择还取决于所要进行的分析。这里评价的所有8个软件包都可以进行基本分析和描述性分析。其中超出基本分析的包括STATA、SUDAAN、WesVar、PC-CARP和SAS（带有或不带IVEware）。表二十一.2对这8个软件包的许多特性进行了总结和比较。

198. 在本章中，根据经验对5个软件包进行了对比（SAS、SUDAAN、STATA、Epi-Info和WesVar），对所有的描述性和分析性例证都提供相同的点估计以及一个预期的发现。5个软件包基本上都产生相同的估计标准误差，无论使用的是平衡多次重复方法还是泰勒级数线性化方法。5个软件包对部分置信区间的计算存在细微的差异；原因在前面已经讨论过了。因此，根据本章报告的基准测试分析，对在这5个软件包中做出选择并没有令人信服的理由。

199. 专门的样本调查软件包（专门用于方差估计）市场将来可能会消失。这一趋势似乎使得标准统计软件包（例如STATA、SAS和SPSS）包含上述功能。因此，未来的数据分析人员可能更容易获得和使用利用复杂调查数据进行方差估计的适当软件。

鸣 谢

感谢以下人员：

Michael S. Deming医学博士，他提供了布隆迪数据集及其文件，多次认真阅读草稿，提出了宝贵的修改建议。

Kevin Sullivan博士，他在浏览Epi-Info方面提供了指导和宝贵的建议，多次认真阅读草稿，提出了宝贵的修改建议。

Z. T. Daniels理科硕士、工商管理硕士，他确定了WesVar输出和文本表格的格式，并把布隆迪人口数据上传到网上。

Graham Kalton博士和Ibrahim Yansaneh博士，他们提出了宝贵的结构安排和编辑建议。

James Chromy博士，他认真阅读了草稿，提出了宝贵的修改建议。

几位匿名审阅人，他们认真阅读了草稿，并提出宝贵的修改意见。

Paul Weiss理学硕士，他在浏览WesVar方面给予指导，并提出了宝贵的建议。

凡本章舛误之处，责任概由作者自负。

参考资料

Brogan, Donna (1998年, 刚印刷完毕)。“样本调查数据软件: 标准程序包的错用”。《生物统计学百科全书》的特邀章节, 主要由Peter Armitage和Theodore Colton等。纽约: 约翰·威利国际出版公司, 第5卷, 第4167-4174页。第二版刚印刷完毕的修正章节。《生物统计学百科全书》, 将于2004年出版。

Brogan, Donna和其他人(1994年)。“增加扩大免疫方案聚类调查设计的准确性”。《流行病学年报》, 第4期, 第4号, 第302-311页。

Carlson, Barbara L. (1998年)。“样本调查数据软件”。《生物统计学百科全书》, 第5卷, 主要由Peter Armitage和Theodore Colton等。纽约: 约翰·威利国际出版公司, 第4160-4167页。

Cochran, William G. 《抽样方法》, 第三版。纽约: 约翰·威利国际出版公司。

扩大免疫方案(1996年)。“估计通过血清调查保护妇女免受破伤风的情况”。《流行病学记录周刊》(世界卫生组织), 第71期, 第17-124页。

Hansen, Morris H.、William N. Hurwitz和William G. Madow (1953年)。《样本调查方法与理论》, 第一期, 《方法与应用》。纽约: 约翰·威利国际出版公司。

Judkins, D. (1990年)。“Fay方差估计法”。《官方统计杂志》, 第6期, 第223-240页

Kish, Leslie (1965年)。《抽样调查》。纽约: 约翰·威利国际出版公司。

_____和M. R. Frankel (1974年)。“根据复杂样本的推论”。《皇家统计学会志》, B辑, 第36期, 第1-37页。

Korn, Edward L. 和Barry I. Graubard (1999年)。《健康调查分析》。纽约: 约翰·威利国际出版公司。

- Krotki, Karol P. (1998年)。“发展中国家的抽样情况”。《生物统计学百科全书》，第5卷，主要由Peter Armitage和Theodore Colton等。纽约：约翰·威利国际出版公司，第3939-3944页。
- Levy, Paul S. 和Stanley Lemeshow (1999年)。《人口抽样：方法与应用》，第三版，纽约：约翰·威利国际出版公司。
- Lohr, Sharon L. (1999年)。《抽样：设计与分析》。加利福尼亚，太平洋丛林：达克斯伯利出版社，布鲁克斯/科尔出版公司。
- Rao, J. N. K. 和A. J. Scott (1981年)。“复杂样本调查中分类数据的分析：关于双向表中适合度和独立性的卡方测验”。《美国统计学会志》，第76期，第221-230页。
- _____ (1984年)。“根据调查数据估计的细胞比例对多向相依表进行的卡方测验”。《统计学年报》，第12期，第46-60页。
- Rust, K. F. 和J. N. K. Rao (1996年)。“复杂调查使用重复方法的方差估计”。《医学研究的统计方法》，第5期，第283-310页。
- Shah, Babubhai V. (1998年)。“方差估计的线性化方法”。《生物统计学百科全书》，第3卷，主要由Peter Armitage和Theodore Colton等，纽约：约翰·威利国际出版公司，第2276-2279页。
- Som, R. K. (1995年)。《实用抽样方法》，第二版。纽约、巴塞尔和香港：马塞尔·德克尔。
- Wolter, Kirk M. (1985年)。《方差估计简介》。纽约：施普林格出版公司。

软件参考资料

- CENVAR方差计算系统：《IMPS3.1版：使用者指南，1995年》。哥伦比亚特区华盛顿，美国商务部，人口普查局。请登陆：<http://www.census.gov/ipc/www/imps/download.htm>。
- 流行病学软件。关于软件和文档编制的信息，请登陆：<http://www.cdc.gov/epiinfo/>。
- IVEware。关于软件和文档编制的信息，请登陆：<http://www.isr.umich.edu/src/smp/ive/>。
- PCCARP (1986年、1989年)。《使用手册》，Wayne Fuller和其他人等。衣阿华，安姆斯，衣阿华州大学，统计实验室。请登陆：<http://cssm.iastate.edu/software>。
- 三角研究所 (2001年)。《SUDAAN使用手册，8.0版》。北卡罗来纳，研究三角园区：三角研究所。请登陆：www.rti.org/sudaan。
- SAS/STAT。关于包括样本调查数据程序在内的SAS/STAT软件的更多信息，请登陆：<http://www.sas.com/technologies/analytics/statistics/stat/index.html>。

STATA。请登陆：<http://www.stata.com>，关于STATA中svy需求的讨论情况登陆：<http://www.stata.com/help.cgi?svy>，关于参考手册登陆：<http://www.stata.com/bookstore/>。

《WesVar 4.2使用指南（2002年）》。马里兰，罗克维尔：Westat公司。另见：<http://www.westat.com/WesVar/about/>。

附 件

本章包括一个附件（只有英文），里面包含利用五个统计软件包对布隆迪免疫调查数据进行的说明分析和比较分析。附件也制作作为CD-ROM，其内容包括每个软件包的程序编码和输出，该内容可以直接从联合国统计司网站（<http://unstats.un.org/unsd/hhsurveys/>）下载，也可以向联合国统计司（statistics@un.org）索取。

第二部分

个案研究

导 言

Gad Nathan

以色列耶路撒冷
希伯来大学

1. 本出版物的第一部分试图介绍发展中国家和转型国家住户调查设计与分析中最重要层面的“前端发展水平”。重点将介绍适用于发展中国家和转型国家住户调查的有关调查设计、实施与分析的一般原则与方法，尤其是操作方面的一般原则与方法：设计效果、调查成本和非抽样误差。适用于发展中国家和转型国家住户调查的方法和技巧有许多，因此本出版物的前几章将尽可能多地对这些方法和技巧进行介绍。各种应用的许多实例都包括在各自相应的章节内，但用于发展中国家和转型国家各种调查的某些方法和技巧将单列章节进行介绍。因此，第七章将介绍7个国家中11项调查的样本设计以及设计效果数据。同样，第十一章将提供个案研究，并详细介绍巴西对非抽样误差进行报告、控制、估计和补偿的现行做法。

2. 但对于从业者来说，至关重要的是需要了解所选用的各种方法和技巧在实际应用中是如何结合使用的，并观测各种具体的实例，以了解各种方法是如何用于一个设计完善和分析周全的完整住户调查中的。各国的具体情况及其基础设施对一般原则的具体运用有着重要的影响，在一个完整的调查中，如何结合使用这些原则尤其会受到这方面的影响。一般来说，个案研究是应用科学研究的一个基本学习工具。通过详细的个案研究来探讨如何将理论上的统计概念和结果用于统计调查的设计和分析是非常有效的。正因为如此，本出版物的第二部分将专门进行个案研究。我们希望通过个案研究将第一部分所讨论的方法运用到实际生活中。这不仅是为了举例说明所研究的技术的各方面的应用情况，更重要的是为了说明如何在发展中国家和转型国家住户调查的完整设计和分析方案中结合使用这些技术。

3. 本部分的四章内容包括了世界各地的数百项住户调查，各项住户调查的主题各异，条件和设计不同，详细程度也各异。在大多数情况下，个案研究都会介绍调查的目的和范围、人口的定义和样本设计、调查工具、实地设计与实施、不答复误差与估计、分析、加权和设计效果。在某些情况下，所述调查是根据国际组织大量调查的设计参数进行标准化的。在另一些情况下，由于邻近国家（如，转型国家）具有相同的条件，所以调查设计具有相同之处。

4. 第二十二章介绍了100多项住户和个人调查中人口与健康调查（DHS）方案的一般特征和设计情况，范围涉及50多个国家。第二十三章介绍了由世界银行负责在40多个国家进行的60多项生活平衡量研究（LSMS）所具有的操作特征。第二十四章对住户预算调查（HBS）所特有的、与样本设计和衡量有关的若干问题进行了讨论，讨论是以若干个发展中国家和转型国家中的此类调查经验为基础的。在对老挝1997-1998年期间支出与消费调查进行个案研究时，还对调查工作的一般条件、调查工具、衡量方法、样本设计与实地工作进行了详细介绍。在对这些方面的经验进行评估后得出了各种有趣的结论。最后第二十五章对东欧14个转型国家住户调查的主要设计和实施层面进行了回顾，并通过详细的案例分析对其中6个国家的住户调查进行了介绍。

5. 所介绍的有些特征存在着许多共同之处。如，所有调查都是住户调查，或本身具有住户的元素。但在许多情况下，分析单位主要为个人——每个住户中的个体（如，人口与健康调查中的妇女）或住户中的所有个人（如，劳动力调查），通常通过一个代表来做出回答。在所介绍的调查中，几乎所有调查的基本样本设计都非常相似——多阶段整群抽样通常将大地理单位作为初级抽样单位（PSU）。通常会对初级抽样单位进行某种分层。在大多数情况下，住户一级的设计都是进行自行加权的。但在每户抽取一个人的情况下，个人样本不再是自行加权。实际上，所有设计都是完全的概率设计，但捷克斯洛伐克共和国和斯洛伐克的住户预算调查仍然在使用定额抽样。

6. 各项调查的目的和宗旨差别很大。如，人口与健康调查旨在“为各国提供用以监督和评估人口、健康和营养方案的相关数据。”生活平衡量研究方案的重点是对生活条件进行了解、衡量和监督。住户预算调查方案旨在衡量日常住户预算的重要方面——收入和支出。转型国家的各种住户调查则是为了对生活条件、消费价格指数的构成以及劳动力统计数据进行分析，当国有经济向市场经济过渡时，需要进行这种分析。

7. 一般来说，这些调查所使用的调查工具仍然是以实地访谈调查为基础，通过笔和纸问卷进行的。据报道，爱沙尼亚劳动力调查（第二十五章）首次尝试使用了计算机辅助的电话访谈（CATI）。所报告的许多调查都对访调员的培训和管理给予高度重视，并采取各种措施来减少不答复和答复误差。人口与健康调查所报告的答复率很高：住户为88-99%，妇女为87-99%。生活平衡量研究调查所报告的总体答复率也很高（74-99.7%）。根据报告，收入数据缺失的比率很高，失业人员的收入数据尤其如此。老挝住户预算调查的不答复率只有3.1%。另一方面，转型国家住户预算调查所报告的不答复率在8-49%之间。这些国家劳动力调查的答复率稍微高一些，不答复率在4-29%之间，有些国家的不答复率一直不到10%。

8. 在许多个案研究中，都对数据整理、编辑和插补的工作进行强调。大多数处理与分析都是通过标准的软件包进行的，通常没有加权。转型国家确实广泛使用了加权和校准方法。有许多研究则努力通过标准方法对设计效果进行估计。这些估计可用于分析以及未来设计的改进。对生活平衡量研究的设计

效果进行的审查表明，有必要将这些估计用于分析中，但由于调查的主题各不相同，不同重要变量在设计效果方面也存在着巨大的差异，所以无法就样本设计得出有用的结论。

9. 本部分提供了从不同国家的各种调查中获取的大量经验，各份报告除提供学习这些经验的机会外，还对所涉各类调查得出了各自的重要结论。其中包括：应当不断更新样本框架；不断重视实地培训和访调员的管理；高质量数据准备的重要性；数据要求和分析的拟订和更新；设计效果的使用及其他。这些个案研究连同本出版物第一部分所介绍的各种方法，是可以从本出版物学到的一部分重要知识。

第二十二章

人口与健康调查

Martin Vaessen

ORC Macro

美国马里兰州卡尔弗顿

Mamadou Thiam*

联合国教科文组织

加拿大蒙特利尔

Thanh Lê*

WESTAT

美国马里兰州罗克维尔

* Mamadou Thiam和
Thanh Lê以前在ORC Macro都
参与过人口与健康调查方案。

摘 要

本章概要介绍国际人口与健康调查（DHS）方案在进行大规模住户和个人调查时所遵循的主要程序，以及调查、抽样程序、答复率和设计效果的一般内容，并对所有重要调查部分——从培训到数据处理以及报告的编写——遵循的程序和方法进行了描述。本章还介绍了迄今为止从实施本调查方案中吸取的主要教训。

关键词：住户调查、答复率、调查抽样、抽样误差、设计效果、调查实地工作。

A. 引言

1. 人口与健康调查（DHS）方案自1984年以来一直在世界各地的发展中国家进行住户调查。人口与健康调查的主要目的是为各国定期提供用以监督和评估人口、健康和营养方案所需要的相关数据。捐赠者和各国越来越注重采用客观指标来对这方面的进展情况进行衡量，这就增加了对定期住户调查数据的需要，因为行政统计机构以及收集数据的其他系统不能提供合适的信息。在人口与健康调查中，住户样本是在整个国家抽取的，之后通过住户调查问卷对住户样本进行访谈以收集住户特征，并确定所有住户成员及其基本特征。还通过妇女调查问卷对年龄在15-49岁之间的妇女进行访谈，主要是为了就背景特征、生育行为、避孕知识及其工具的使用、孩子和妇女的健康以及其他方面的问题收集相关信息。每次访谈的时间平均为35-40分钟左右，各访谈时间一般在10-90分钟之间，当然，有些访谈时间会更长一些。样本规模的差异也很大，妇女人数一般在5 000-30 000名之间。在有些国家，还对年龄在15-59岁之间的男性样本进行了访谈，通常情况下，这是用来抽选妇女样本的子样本。对男性的访谈平均需要25分钟左右。以下各节将对人口与健康调查方案的历史及其调查的一般内容进行介绍，并对单位不答复进行分析。另外还对样本设计效果、调查的不同阶段以及发展中国家人口与健康调查的有关教训进行了介绍。

B. 历史

2. 人口与健康调查是在两个早期的住户调查方案之后进行的：世界生育率调查（WFS）和避孕普及调查（CPS）。世界生育率调查是在1973-1984年期间进行的，避孕普及调查是在1977-1985年期间进行的。世界生育率调查方案在41个发展中国家进行了调查，并与20个发达国家合作进行了调查。世界生育率调查主要是为了获得生育、计划生育以及儿童健康方面的信息。该方案由美国国际开发署（USAID）和联合国人口基金（UNFPA）共同提供经费。此外，大不列颠及北爱尔兰联合王国、荷兰和日本政府也提供了援助。

3. 避孕普及调查方案在33个国家进行了43项调查，调查范围更窄，主要是计划生育方面的调查。该方案由美国国际开发署提供资金，各项调查限于已经从美国国际开发署获得发展援助的国家。

4. 人口与健康调查始于1984年。截止2003年，已经在大约70个国家进行了约150项妇女调查、75项男性调查以及10项健康设施调查。调查通常每5年进行一次，当然有几个国家调查的时间间隔更短一些。调查大多在那些从美国国际开发署获得援助的国家中进行，当然参与的某些国家由世界银行或联合国人口基金提供资金。在许多国家，调查还由美国国际开发署之外的其他捐赠机构提供支持，如英国国际开发部（DFID）、联合国儿童基金会（UNICEF）、日本和瑞典政府以及其他机构等。人口与健康调查全面概述参与国的人口与母婴健康问题，各机构随时可以获取相关数据，用于监测和评估。为适应不断变化的环境和优先事项，调查内容随着时间的流逝发生了变化。

C. 内容

5. 为最大限度地使有关信息具有兼容性，各国每轮人口与健康调查的核心内容都具有标准性。除了这些核心内容外，各国还可以做出决定，增加调查问卷单元，来处理本国特别关心的问题。在撒哈拉以南非洲地区的国家，调查问卷的核心内容与其他国家有所不同，主要是调查问卷的复杂程度不同。

6. 1997-2002年期间的核心调查问卷如下：

住户调查问卷。该调查问卷旨在获取年龄、性别、父母生存状况以及住户成员上学方面的基本数据。另外，调查问卷还包括水供应和住户娱乐设施方面的信息。住户调查问卷还针对年龄在15-49岁之间的妇女以及年龄在5岁以下的儿童收集了身高、体重及其血色素（用以衡量是否贫血）方面的信息。

妇女调查问卷。本调查问卷适用于育龄妇女，包括以下部分：

- 受访者的背景特征
- 生育史
- 避孕
- 怀孕、产后护理以及母乳喂养
- 免疫接种、卫生与营养

- 婚姻与性生活
- 生育偏好
- 丈夫的背景以及妇女的工作
- 艾滋病毒/艾滋病及其他性传播感染

有些调查包括艾滋病毒/艾滋病、梅毒或其他生物标记的检测。

7. 还有男性调查问卷。这种调查问卷中的某些题项和妇女调查问卷相同，它不适用于所有国家。此外，还有针对计划生育和保健服务提供者的调查问卷，但这类调查问卷是和住户调查分开的，适用对象是服务提供者。这种调查问卷称作服务提供评估（SPA）调查问卷，调查问卷通过向服务提供者和顾客提问以及对服务的提供进行观测，涉及提供服务的所有方面。

8. 人口与健康调查方案开发了许多单元，各国可以将这些单元加在其调查问卷中。可以提供以下单元：

- 切割女性生殖器官
- 孕产妇死亡率
- 口服避孕药的情况
- 绝育史
- 血亲（有血缘关系的亲属结婚）
- 口头分析（死因方面的详细问题）
- 艾滋病毒/艾滋病
- 子女教育
- 妇女状况
- 家庭暴力
- 疟疾
- 住户健康支出

9. 由于核心文书的篇幅问题，一个国家一般无法增加两个或三个以上的单元，但这要视所选单元的长度而定（关于调查问卷及其他材料，请访问www.measuredhs.com）。

D. 抽样标架

10. 显然，在对人口与健康调查进行规划的初期阶段，应该解决是否有适宜抽样标架这一问题。人口与健康调查需要收集个体住户中个人的数据，但一般却没有有关该个人或住户方面的最新清单。根据定义，大多数人口与健康调查所使用的抽样标架是一张覆盖整个国家领土的不重叠地区单位的清单。用于

抽样标架之目的的这些单位，其基本特征是边界的界限明确，地图描绘清晰。每个地区单位还有一个独特的识别码，还必须有一个衡量（人口和/或住户数量）规模的现行量度或估计量度。其他特征如，每个区域单位通常会有城市/农村分类，这些分类可用来进行分层。

11. 在大多数国家，预期的地区单位与普查查点区（EA）一致，这样就可以为第一抽样阶段提供方便的框架。在有些国家，这些查点区的人口规模可能很大，而在其他国家则可能很小。不管规模如何，查点区通常是初级抽样单位（PSU）。在有些国家，如果查点区足够小的话，那么也会将其作为最终的地区单位。如果将查点区用作初级抽样单位，而结果发现（住户或人口）规模太大的话，那么在样本设计中会进行分割，作为抽选样本的中间阶段。

12. 如上所述，不管是包括查点区（EA）的抽样标架还是其他单位的抽样标架，都有可能不是最新的。通常需要采取措施(a)更新整个框架；或(b)在抽样的倒数第二个阶段通过编制一个最新的住户清单来更新部分框架。

13. 在有些调查中，如果认为已有的标准样本在设计方面能够实现人口与健康调查的测量目标，可以将其用作抽样标架。

E. 抽样阶段

14. 对于任何样本设计来说，抽样阶段的数量取决于抽样标架的特征以及调查的目标。尽管各国的样本设计并没有标准化，但每项人口与健康调查的样本设计所遵守的一般原则都是一样的：简单、概率抽样（已知的非零抽样概率）、类集和分层。在人口与健康调查中，通常需要两个或两个以上的抽样阶段，这主要取决于抽样标架中的地区单位规模。

15. 基本的设计包括在第一阶段通过概率与规模成比例（PPS）的抽样方法抽取地区单位，规模是指每个地区单位的人口数或住户数。这个第一抽样阶段确定了一个界点，越过这个界点之后，抽样工作就应走出办事处，到实地绘图，或在必要的情况下在选定的区域单位编制住户清单。绘图包括绘制草图，以显示每个被抽中的初级抽样单位的边界以及每个初级抽样单位内住所的位置。如果国家有详细准确的初级抽样单位图，绘图工作只涉及住所位置的更新。如认为框架不是全新的，那么可以针对所有有人占住的住所列出所有住户清单（包括编制清单的小组在进行访谈时无人在家的住户），并据此来编制每个被抽中的初级抽样单位的最新住户清单。所编制的清单可以作为抽样标架，用于在第二阶段进行系统的住户抽样。

16. 任何住户调查的类集规模（每个初级抽样单位或类集将要抽取的住户/妇女数量）取决于被调查的变量。对于类集程度很高、通常需要进行地区间比较（避孕普及情况及其决定因素）的变量，理想的类集规模为每个类集15-20名妇女。其他生育变量的类集程度则要低一些，如果需要进行非地区方面的比较（如，年龄组之间或教育水平之间的比较），那么理想的类集规模可以更大一些。人口与健康调查对于农村地区采用大约30-40名妇女的类集规

模。在城市地区，大规模类集的成本优势一般弱一些，人口与健康调查采用了20-25名妇女的类集规模。如果有最新的住户清单，这些数字减少了，原因是在清单编制方面不存在需要大类集规模的因素（ORC Macro, 1996年）。人口与健康调查还收集儿童健康方面的数据，这些儿童是被抽中妇女的孩子，类集规模应该足够大，以便有足够数量的孩子可用于分析。

17. 被抽中住户中所有有资格的个人都被纳入最终样本。尽管在大多数的人口与健康调查样本中，各个初级抽样单位中被抽中的住户数量各不相同，但在有些调查中则抽取了固定数量的样本。

18. 通常情况下，被抽中初级抽样单位的规模都太大，因而无法直接编制清单。设计时应该考虑采用分割的方法减少清单的量，并在各初级抽样单位之间保持均衡的工作量。每个大的初级抽样单位都被分割为若干个分段，其中有一个分段保留在概率与规模成比例（PPS）样本中。

19. 人口与健康调查的绝大多数样本设计都是类集和分层样本设计。显性分层通常以地理标准为基础，如城市/农村细分，只有在第一抽样阶段使用。初级抽样单位是在每层中独立选取的。隐性分层则采用系统抽样的方法。通常情况下，初级抽样单位的数量较大，对于10 000个住户的样本，初级抽样单位大约在300-550之间。

20. 为方便设计的准确实施，人口与健康调查努力让其样本设计尽可能地简单。但为了满足国家具体情况的需要，对基本设计进行了修改。这些修改包括采用有或没有紧密类集的标准分段设计；紧密类集是指类集中每个被抽样住户在地理上相互邻近，而非紧密类集是指类集中被抽样住户在地理上是分散的。这是样本设计的一种变化形式，在这种样本设计中，一个事先确定好的标准分段规模，也就是说，所列明的最终地区单位应尽可能地小。国家的每个初级抽样单位或查点区 i 都被分配了若干个分段 s_i ——用普查人口除以标准分段规模。然后通过概率与规模成比例（PPS）的抽样方法抽取初级抽样单位，在这种情况下，规模的大小等于分段 s_i 的数量。再在被抽中的每个初级抽样单位内，随即抽取一个分段。至于紧密类集的标准分段，其情况是分段由平均规模 T 组成，而 T 是预期的类集规模。这样，采用“全取”方法，可以避免编制清单的工作（ORC Macro, 1996年）。

21. 在人口与健康调查中，既有整个国家的估计也有具体地理域（如城市、农村和地区）的估计。由于人口规模中的域通常是不定的，所以为了提供分析所需要的足够样本规模，通常在设计样本时，对小域进行过度抽样。当然，这样可能会使国家估计值出现偏差，可以通过对样本数据进行适当的加权来纠正这种偏差。样本权数的主要部分为设计权数，设计权数以抽选概率为基础。在进行加权时还要考虑到住户和个人一级的不答复情况。可采用最后阶段的加权，在这种情况下，只要用于抽样的区域抽样框架过时，就应该进行事后分层调整，并且使用来源可靠的人口预测数据。

F. 不答复情况的报告

22. 在人口与健康调查中，不允许置换不答复单位（住户或个人），这和许多其他调查不同。为了获得目标样本单位数量，需要在进行样本设计时根据过去或类似的调查，对样本单位的不答复率进行估计，然后将不答复率用来确定所需要抽取的单位数量。此外，为确保高水平的答复率，在实地工作中采取了各种措施。之后将对人口与健康调查的答复率进行介绍，包括对这些答复率在不同时间和不同地区的情况进行比较。

23. 如前所述，人口与健康调查的数据是在两级进行收集的：住户与个人。合格的个人大多数为育龄妇女，但某些国家还对年龄在15-59岁之间的男性进行了访谈。在人口与健康调查中，不答复是指：没有对被抽作样本的住户或个人进行访谈。住户或个人的答复率是通过准确记录所有住户和合格个人的方式来测量的。答复率通过输入调查问卷中的答复代码计算。住户调查问卷用以确定每个住户内的所有合格个人，只有合格的个人才会领到一张个人问卷。

24. 住户一级的答复代码为：

- 1H 完全访谈
- 2H 住户成员不在家，或没有合格的受访者在
- 3H 在延展的期限内整个住户成员都不在家
- 4H 推迟
- 5H 拒绝
- 6H 住所无人居住或地址所在地不是住所
- 7H 住所被毁
- 8H 住所没有找到
- 9H 其他

那么住户答复率为：

$$R_H = \frac{1H}{1H + 2H + 4H + 5H + 8H}$$

25. 在人口与健康调查中，其代码为3H、6H、7H和9H的住户为不合格的住户，因此不包括在分母中。⁴³代码9H通常由监督员重新编码，将其列为某个明确的编码，但这一编码几乎都是不存在的。代码为9H的住户中，有少数的情况可以归入不合格之列。应该指出的是，由于许多国家都缺乏完善的地址体系，所以人口与健康调查中编制清单的工作首先是要根据有人占住住户的名称来确定住所，然后用名称来代替地址。如果在编制清单和访谈期间，有一家新的住户搬入某个住所，这并不意味着发生了抽样单位置换这一情况，因为住所是抽样的真实基础。另外，如果在编制清单后，某个住户搬出去了，而没有另一个住户搬进来，这并不构成不答复情况。

26. 个人一级的答复代码为：

⁴³ 由于在人口与健康调查中，代码为3H的住户（在延展期间整个住户成员都不在家）为不合格的住户，所以这种计算住户答复率的情况和美国民意研究协会2000年标准所确定的RR5方法类似，这一方法夸大了真实的答复率，因为少数代码为3H的住户是合格的但却没有纳入计算中。

- 1I 完全访谈
- 2I 不在家
- 3I 推迟
- 4I 拒绝
- 5I 部分访谈
- 6I 无能力
- 7I 其他

个人答复率为：

$$R_f = \frac{1I}{1I+2I+3I+4I+5I+6I+7I}$$

27. 为每层或每个报告区域分别计算了未经加权的住户和个人答复率，这些答复率和总体答复率一起被编入人口与健康调查的国家报告中。总体答复率是根据住户和个人一级的答复率确定的。在人口与健康调查中，各区域的答复率是类似的。由于样本通常是在每个区域内大致经过自行加权的，所以从整体上讲，一个国家的加权答复率和未加权答复率是非常接近的。应该指出的是，大多数人口与健康调查中都使用了上述答复代码，但考虑到特定国家的具体情况，有些调查对这些代码进行了修改。

G. 不答复率比较

28. 通过上述公式为1990-2000年期间44个国家进行的66项调查计算了住户答复率和妇女答复率。其结果见以下地区的附件：亚洲、欧亚大陆、拉丁美洲、近东和撒哈拉以南非洲地区。

29. 数据显示，这些调查的住户答复率在87.9%和99.5%之间，平均答复率为97.5%。这说明对人口与健康调查样本中所确定的绝大多数住户进行了成功的访谈。同样的调查中，妇女答复率在86.5%-99.3%之间，平均答复率为95%。这说明对大多数合格妇女进行了完全访谈。

30. 拉美的总体住户答复率为95%，除该地区以外，其他所有地区的平均住户答复率为98%左右。至于住户，拉美的平均妇女答复率低于人口与健康调查方案所覆盖的其他地区：分别为92%和97%。在每个地区内，各国的住户和妇女答复率差异不大，变异系数在0.4%-3.7%之间。

31. 在人口与健康调查方案的最后三个阶段（DH II、DHS III和MEASURE-DHS+），⁴⁴ 平均住户答复率一直高达97%，而同一期间，平均妇女答复率则略有上升，从94%上升到96%。

32. 在人口与健康调查中，住户和个人一级高水平的答复率得益于对实地人员的严格培训以及实地工作的严密监督。此外，在每项调查中，都认真确保了编制清单工作的时间和访谈的时间不会相隔太长。另外，与发达国家调查不

⁴⁴ MEASURE是美国国际开发署的重要项目，MEASURE DHS+是其中的一部分。“MEASURE”的全称是“对获取和使用结果进行监督和评估”。

同的是，发展中国家的住户调查通常能够得到潜在受访者的通力合作。随着时间的推移，各地区的平均住户答复率和个人答复率已非常接近。

H. 人口与健康调查的样本设计效果

33. 本节将简单介绍人口与健康调查中的某些设计效果以及组内相关系数值 ($\bar{\rho}$) [详情见Lê和Verma (1997年)：对世界生育率调查抽样误差进行的类似分析见Kish、Groves和Krotki (1976年) 以及Verma、Scott和O'Muircheartaigh (1980年)]。

34. 设计效果等于通过复杂样本设计获得的任何估计的抽样方差与通过简单随机样本或同等样本规模的非限制性样本获得的同一估计的抽样方差之比 (Kish, 1965年)。即：

$$D^2(y) = \frac{Var_{complex}(y)}{Var_{unrestricted}(y)}$$

35. 设计效果是分层、不等抽样概率、样本加权调整（用于不答复）、人口加权调整（用于未覆盖情况和提高精确度）以及对复杂样本设计所有元素进行类集的结果。

36. 由于加权产生的估计设计效果可以从样本中计算，公式为：

$$d^2(\hat{y}) = 1 + cv^2(w_j)$$

式中， cv^2 为抽样权数 w_j 的变异系数平方。

37. 因类集效果产生的设计效果可按以下公式计算：

$$D^2(\hat{y}) = 1 + (b-1)\rho$$

式中， b 为平均类集规模， ρ 为组内相关性。

38. 本出版物第六章对设计效果、组内相关系数——定义、设计效果的组成部分、抽样调查设计中设计效果 and 组内相关系数的使用等问题进行了全面的讨论。为了了解复杂样本设计对标准误差的影响，通常使用设计效果的平方根 $d(\hat{y})$ 。

39. 如前所述，人口与健康调查以全国范围内的代表性住户样本为基础，采用了标准的多阶段分层概率样本设计，设计包括相当数量的初级抽样单位。对于许多国家的城市和农村地区，以及通常与行政区相当的较小地区，其估计值一般在国家一级获得。

40. Lê与Verma (1997年) 对1985-1993年期间48项人口与健康调查的抽样误差进行了研究。就全国的总估计而言，平均根设计效果 $d(\bar{y})$ （其中 \bar{y} 通常是按37个变量和48项调查进行平均的一个比率）大约为1.50，平均值在1.13（特立尼达岛和多巴哥）和2.07（尼日利亚）之间。这是说，相对于同样规模

的非限制性样本系数而言，类集、加权和设计的其他方面使估计的标准误差平均增加了1.5个系数（或估计方差的2.25个系数）。

41. 大多数国家的城市和农村地区中都使用了类似的类集规模（城市地区的平均类集规模为24，农村地区为30）。因此，城市和农村地区的 $d(\bar{y})$ 平均值差异很小，城市地区为1.4，而农村地区则为1.5。地理区域的 $d(\bar{y})$ 值也是如此，在每个国家内，不同地区的 $d(\bar{y})$ 值非常接近，只是比整个国家相应的 $d(\bar{y})$ 值略小一点，这一点再次说明国家所有地区使用了相同的设计。与之相反的是，根据个别受访者的人口和社会经济特征所界定的子群，其 $d(\bar{y})$ 值则要比整个国家相应的值明显小得多。因为这些子群涉及整个初级抽样单位，有关的类集规模（ b_d ）小于整个样本（ b ）的类集规模，因此子群设计效果通常要小一些。如，在突尼斯的人口与健康调查中，对于工作女性和非工作女性这两个子群而言，变量“理想家庭规模”的 $d(\bar{y})$ 值分别为1.56和1.70，而总样本的 $d(\bar{y})$ 值则为1.79。

42. 由于在人口与健康调查中，城市和农村地区或地理区域的抽样率不同，所以需要对本样本数据进行加权。为补偿不答复方面的差异以及样本实施过程中的其他缺陷，也有必要进行加权。这种加权往往会加大抽样误差。针对人口与健康调查中基于总样本的估计，对其归因于变量权数的设计效果进行了计算。在1985-1990年进行的早期调查中，归因于加权的平均 $d(\bar{y})$ 值为1.08（相当于方差增加了17%）；在1990-1993年进行的近期人口与健康调查中，这一数值增加到1.15%（相当于方差增加了32%），这与为考虑区域估计值在城市和农村地区采用epsem（等概率）样本的惯例相差许多。

43. 正如表二十二.1所示，各个国家总样本的平均 $d(\bar{y})$ 值在各个变量之间的差异很大， $d(\bar{y})$ 值的范围从婴儿死亡率这一变量约为1.1或1.2的低值到儿童是否在医院出生这一估计量的2.5这一高值不等。这说明现有医疗的地区类集内具有较高的相关性。在对这些 $d(\bar{y})$ 值的差异进行考察时，应该注意表中不同部分样本基础的差异。如，最上一组估计值是以年龄在15-49岁之间的所有妇女为基础的，第二组只以这一年龄组中目前已婚的妇女为基础，下一组则以在过去五年中的所有分娩为基础。样本基础的变化导致了类集设计效果的 b 值不同，而这一因素又导致表二十二.1中 $d(\bar{y})$ 值发生变异。

44. 计算同质性 ρ 要比计算归因于类集的设计效果对未来调查计划更有用，因为设计效果取决于 ρ 和类集规模 b 。只有当这些参数相同时，过去调查的设计效果才能适用于新的调查。但应该考虑到改变 b 的可能性，因为类集规模可以由抽样人员进行控制，但组内相关性却不能。如果已经有 ρ 的估计值，那么可以通过计算不同 b 值的类集设计效果，来考察改变 b 的效果。这样， ρ 就成为我们关注的关键因素。根据人口与健康调查计算了平均 $\bar{\rho}$ 的估计值，其结果也显示在表二十二.1中。如表中所示， $\bar{\rho}$ 的值差异很大，从0.01的低值到0.22的高值不等。正如所预期的那样，各种取决于地方卫生设施提供情况的估计，其 $\bar{\rho}$ 值往往较大。

表二十二.1.

48项人口与健康调查的平均 $d(\bar{y})$ 值和 $\hat{\rho}$ 值, 1984-1993年

比例/均值	$d(\bar{y})$	$\hat{\rho}$
15-49^a 岁的所有妇女		
目前已婚	1.43	0.03
已生育子女数量	1.35	0.02
最近5年中的出生数量	1.44	0.03
5岁以下存活的孩子数量	1.41	0.02
40-49岁妇女生育子女的数量	1.26	0.02
15-49岁目前已婚的妇女		
不想再要孩子	1.32	0.02
希望将下次生育时间推迟两年或两年以上	1.24	0.01
了解一种避孕方法	2.01	0.14
了解一种现代避孕方法	2.08	0.15
了解一种避孕药供应来源	1.94	0.12
目前使用任何一种避孕方法	1.50	0.05
目前使用一种现代避孕方法	1.43	0.04
目前使用宫内避孕器 (IUD)	1.42	0.04
目前服用避孕药	1.41	0.04
目前使用避孕套	1.38	0.03
目前使用公共来源的避孕药	1.36	0.03
绝育	1.36	0.03
过去五年中的所有生育		
母亲生产时是否接受医疗保健	2.54	0.22
母亲是否注射破伤风类毒素	2.02	0.12
5岁以下儿童		
过去两周是否有过腹泻	1.34	0.03
上述儿童是否接受过ORS ^b 治疗	1.25	0.12
6-35个月的儿童		
年龄标准身高比规定标准低两个标准差	1.33	0.05
年龄标准体重比规定标准低两个标准差	1.29	0.04
身高标准体重比规定标准低两个标准差	1.19	0.02
12-23个月的儿童		
是否有健康卡	1.33	0.15
上述儿童是否接受了全部的免疫接种	1.31	0.21
1-4年或5-9年前出生的儿童		
进行调查前的1-4年内婴儿死亡率	1.23	0.02
进行调查前5-9年内婴儿死亡率	1.14	0.01

^a 在大约四分之一的调查中, 样本都限于曾经结婚的妇女, 因此该组的所有变量也都限于曾经结婚的妇女。

^b 口服体液补充盐。

45. 从人口与健康调查方案的抽样误差分析中获得的一个重要发现是，如果样本设计具有可比性，那么某一特定估计对象的 \bar{p} 估计值在各国之间具有很强的通用性。这样，一个国家在设计一项新的调查时，如果有必要，并且能够认真检查可比性的话，就可以采用邻国的类似调查在抽样误差方面的经验数据。

I. 调查实施⁴⁵

46. 在着力注重科学抽样和抽样误差计算的同时，不应忘记调查中误差的来源有许多。与抽样变异性有关的误差通常可以量化，而其他误差一般不能轻易地量化。不过，通常情况下，非抽样误差很可能要比抽样误差大。如果在培训和征聘实地工作人员和数据处理人员时不给予足够的重视，尤其会出现这种情况。这样，每一项人口与健康调查中，非抽样误差的控制都是一项主要的目标。

47. 至于实施问题，许多人口与健康调查都是在难以征聘到高素质实地工作人员的国家中进行。在这些国家中，实地工作面临着交通、住宿、卫生和食品供应等方面的众多问题。实地工作人员需要前往全国各地，而这又会带来安全和监督方面的问题。这些问题以及其他方面的问题正是人口与健康调查方案高度重视实地人员培训、实地和办事处监督的重要原因。然而，尽管对监督给予了高度重视，但还是会出现一些诸如系统得不到正确实施、数据质量出现问题之类的情况。以下将对人口与健康调查的主要实施步骤进行介绍，并强调了细致准备、全面培训和监督的必要性。

48. 调查的另一个重要方面是提供调查数据的及时程度，以及调查数据可供决策者、方案管理人员以及分析人员使用的程度。有许多调查从来就没有被深入分析过或适当发布过，发展中国家尤其如此。人口与健康调查方案旨在确保所有调查都能及时得到分析，调查结果能够被妥当发布，数据可供人们进行进一步的研究。为实现这一目的所需要采取的办法描述如下。

J. 编制与翻译调查文件

49. 每个参与国的调查文件通常都包括一份住户调查问卷、若干份用于女性和/或男性的个人调查问卷以及相应的手册。调查问卷包括：人口与健康调查的核心问题、针对具体国家所进行的修改内容以及可选择的单元。人口与健康调查人员应和地方人员一道进行调查问卷的修改，一定要记住国家的具体需要。人口与健康调查的示范调查问卷很长，所以要根据调查问卷的总体长度对增加的内容进行认真考虑。如果调查问卷变得难以使用或需要很长的时间才能填好，数据质量很可能会受到影响。每个国家都应对核心的访调员和监督员手册进行改编，以体现调查问卷的内容是针对具体国家编制的。

50. 人口与健康调查的政策是将调查问卷翻译并印刷成所有主要的地方语言，以便使用受访者的语言进行访谈。任何一个语言群体，只要其在样本中占了10%或10%以上的比率，就应该有其自己语言的调查问卷。通常情况下，由访调员或其他人进行现场翻译的情况是无法完全避免的，因为对于样本内的某些

⁴⁵ 在有关调查组织、人口与健康调查的特征的章节中，许多材料摘自人口与健康调查组织手册草案，该手册由本章的一位作者起草。

受访者来说，可能无法提供足够语言版本的调查问卷供他们使用。但还是应该尽量减少现场翻译的需求量。

51. 翻译不是一件容易的工作，不仅需要很强的语言能力，而且需要了解人口与健康调查中常见的术语和表达方法。很少有一个人能掌握所有这些技能，当一个国家使用多种语言时尤其如此。

52. 在翻译方面，人口与健康调查需要有一人将人口与健康调查问卷翻译成所需要的当地语言，可采用英语、法语或西班牙语版本的核心调查问卷。如果以前有翻译过的人口与健康调查或类似的调查，那么肯定应该考虑到这种翻译过的调查。通常情况下，如果所问的问题相同，需要翻译的内容也会相同，如果先前的翻译有缺陷，则属于例外情形。

53. 然后由一名独立的翻译人员将已经翻译过的调查问卷回译原来的语言。为保证两个版本具有完全的独立性，在回译时不要参考原有调查问卷，这一点很重要。下一步是让两位翻译人员以及高级调查人员一起研究原始版本的调查问卷和回译后的调查问卷，以便解决这两个版本所存在的不符问题。这是一个重要的步骤，当一种语言不是通常书写的语言时，这一步尤其重要，因为这种情况下的翻译不是一件简单的工作。

54. 这一过程所产生的调查问卷应该很容易被那些将要用自己的语言接受访谈的受访者理解。但确定调查问卷的最后译文之前，没有必要进行大量的实地访谈，但应当使用每种语言至少进行三到五次访谈。很重要的一点是要记住，翻译的目的是为了确保向所有受访者提出的问题都相同。但这并不是意味着翻译应该是逐字逐句的翻译。一个好的翻译尽管不是逐字逐句进行翻译，但所传达的意思却是相同的。人口与健康调查通常在国家中重复进行，当然各轮调查问卷的内容可能有些不同。因此，可以采用大多数问题以前的译文，并借鉴从先前的事先检验和实地工作中获得的经验。

55. 如果不能使用英语、法语或西班牙语版本的文件，访调员和监督员手册等调查文件应该翻译为可以被所有实地工作人员理解的语言。

K. 事先检验

56. 事先检验是检验译文、调查问卷跳过模式、访调员手册、监督员手册以及其他调查程序的重要手段。通过这种方式还可以让高级调查人员在进行主要培训课程前获取培训实地工作人员的经验。人口与健康调查的国家管理人员通常参与事先检验中的访谈。

57. 为了进行事先检验，需要对少量的实地工作人员进行培训，培训通常为两周。培训由地方人员进行，并由人口与健康调查的国家管理人员提供协助。人口与健康调查的惯常做法是培训未来的监督员作为事先检验的访调员，这些人将作为监督员参与以后的访调员培训。这样就可以确保他们能够得到全面的培训，他们的角色在访调员培训期间就已经得到确定，并有足够的人员对访调员培训期间的实践活动和测试提供指导和纠正意见。

58. 事先检验通常包括100-200个住户，访谈大约需要一周的时间。事先检验的访谈是在那些没有被主要调查抽取的城市和农村地区进行的，其目的是为了防止调查结果受到影响。目前人口与健康调查方案在这类调查方面积累了非常全面的经验，因此事先检验规模可以小一些，没有必要覆盖国家的许多不同地区。

59. 事先检验的实地工作所采取的步骤与主要实地工作中所采取的步骤一样。因此，需要列出住户清单，以便小组人员熟悉跟踪程序和使用控制表格。高级调查人员应积极监督事先检验的所有阶段，以便了解所遇到的各种问题，并提出各种解决方案。

60. 事先检验所获得的经验是修订调查问卷和相关手册的基础。需要根据事先检验期间所观测的工作情况对错误进行纠正和改进。这一活动的关键是记录培训期间发现的所有问题、各种实践和实际访谈。在随后期间发现的问题应通过对事先检验访谈进行观测的调查人员所做的报告来进行记录，或通过事先检验访调员的日常工作报告来进行记录。参与事先检验的所有人员都应该对其所做的观测进行记录，这一点很重要。

61. 还应该注意确保事先检验后所做的任何修订不应引发新的错误。事实上，如果需要对调查问卷进行全面修订，需要采用新的工具进行一些实地访谈，以确保所进行的修订是正确的，不会引发新的问题。

L. 实地工作人员的征聘

62. 住户调查的质量在很大程度上取决于实地工作人员的素质。因此，应该尽可能征聘最出色的人员来从事这一工作。在发展中国家，很少有机构拥有常设的访调员和监督员实地工作队，即使有，访调员的绝大多数也通常都是男性。除非属于男性方面的调查，否则人口与健康调查将需要女性的访调员。因此，人口与健康调查的实地人员一般是该项工作特别征聘的人员。由于收集数据或实地工作阶段通常需要三到六个月，因此所征聘的人员通常是那些目前没有工作、愿意而且能够离家几个月的人员。在有些国家，由于调查涉及到更广泛的健康内容，因此借调了为卫生部工作的医务人员，让他们充任访调员和监督员。

63. 征聘工作应考虑需要征聘多少能够讲调查所用每种语言的工作人员。由于可能会出现自然减员和不合格候选人员被解聘的情况，因此，征聘的见习员人数应该比实地工作所需要的人数至少高出10-15%。需要根据候选人能力的客观测试而非其他特点来进行征聘。候选人应该体面、能够长途跋涉并且能够和访谈对象建立友善关系。另外一项必备条件是具有良好的团队精神。绝对不能按照候选人和调查人员的关系、个人偏好或其他不能接受的征聘做法来进行征聘。

64. 监督员和实地编辑人员的职位需要那些可以成为小组领导的人员。他们应该自信、有较强的上进心和良好的团队精神。所有这些特征也都是访调员候选人所需要的特征。但一个好的访调员所具有的主要特征是能够自然而然地流利提问，让受访者放松并能正确记录受访者做出的回答。

M. 访调员培训

65. 访调员培训和事先检验培训很类似，只是它一般需要三到四周的时间，这主要是因为学员人数更多的缘故。在培训期间，访调员候选人应在实地进行至少5-10次的实践访谈。培训由地方人员进行，并由为事先检验培训的工作人员以及人口与健康调查国家管理人员提供协助。

66. 访调员的最终甄选取决于他们在一系列书面测试中的成绩、在办事处进行实际访谈期间的表现以及进行事先检验访谈的质量。甄选标准应该客观，这一点尤其重要。在许多地方，调查人员面临着来自其他人的巨大压力，他们需要根据这些人的特定选择来填补现有的工作职位。但甄选人员的惟一方法是对人员胜任相关工作的资格进行审查，并对这些人员在培训期间的表现做出客观的评价。事实上，在培训期间进行客观的书面测试有助于调查人员记录某些候选人不能入选的原因。

N. 实地工作

67. 人口与健康调查的政策要求实地工作采取分组办法。分组工作的原因有许多，但其中一个主要原因是能够对工作进行更加严密的监督。另一个理由是因为需要为大多数访调员提供专用运输工具。在许多国家还需要维护实地工作人员的安康，这是另外一个重要的理由。

68. 小组一般包括一名监督员（组长）、一名女性实地编辑人员、一名卫生技术员以及三到四名女性访调员。如果还包括对男性的调查，那么小组还应包括一名男性访调员。许多国家都为每个小组派一辆专车，并配备一名司机。小组人员的多少有时受所用车辆的乘载量限制。

69. 监督员对小组、小组工作的日常组织和监管进行全面负责。实地编辑人员主要负责检查访谈的质量。在实际工作中，为了建立和维护一个好的访谈小组，需要由监督员和实地编辑人员分担责任。

70. 确定小组人数需要考虑的主要因素为：初级抽样单位的数量、类集的规模以及预期的实地工作时间。需要考虑的其他重要因素包括：现有车辆的数量、可以征聘到的合格访调员和监督员的人数以及该国所用语言的数量。实地工作应持续三到六个月。有时，时间还可以更短一些。但为了取得高质量的数据，访调员的数量保持在较低的水平，因为人员数量受培训、合格候选人数等方面的限制。而这反过来又限制了可用来决定实地工作期限的小组数量。

71. 如果可能，所有小组都在常规的另一地理区域（如同一省份）开始实地工作，以便使高级调查人员能够在最需要监督的期间对所有小组进行监督。如果从一开始小组就分散在全国各地，将很难立即走访所有小组。

72. 在对调查小组分配抽样地区时，应该考虑到所使用的语言和其他要求，并应尽量减少每个小组的旅行次数。一般来说，小组每周工作六天，每次离家工作几个星期或几个月。

73. 如果第一次访问时未完成访谈, 在将此种情况列为不答复情况之前, 应做进一步努力, 对被抽中的住户或受访者进行访谈, 最多三次, 并分三天进行。之后的联络应该安排在受访者最有可能在家的时间。如果大多数小组成员已经完成工作, 但还需要一天进行一两次回访, 那么小组通常会转向一个新的类集, 而留一位访调员做最后的“清理”工作。如果新的类集距离不远, 而且小组车辆可以接送负责“清理”工作的访调员, 那么这是有可能的。在其他情况下, 整个小组则需要一直留在实地, 直到有关类集的所有工作都完成为止。正如前面所述, 不允许置换那些拒绝访谈或被列为不答复的住户或个人。

74. 应该为小组提供足够的调查问卷和材料以保证工作能够在所有时间全速进行。应将已完成的调查问卷打包, 保证它们的安全并防止它们受到天气的影响, 直至将它们送到总部办事处为止, 这通常由定期访问每个小组的实地巡视监督员负责。

75. 高度重视监督是人口与健康调查的特点。经验显示, 如果没有不断监督, 数据质量将会受到很大的影响。因此, 调查采取了数级的监督。小组监督员和实地编辑人员需要不时地对访调员进行观测, 并对每份调查问卷的完整性和准确性进行彻底检查。如发现重大问题, 访调员应回访受访者以获得正确的信息。另外, 监督员通常还需要负责对子样本(大约为被抽中住户的10%)进行再次访谈, 以确保首次访谈已经进行过, 所有合格妇女的确定都是正确的。

76. 调查主管以及人口与健康调查工作人员将在实地工作中提供进一步的监督。应定期到实地小组进行访问, 以检查访调员、编辑和监督员的工作。在这种检查期间, 应该在实地编辑对调查问卷进行检查之后, 至少对每位访调员的一两份调查问卷进行认真检查。这样就可以同时发现访调员和编辑的错误。进行监督性的实地访谈尤其重要, 有些监督员和编辑所做的工作是不理想的, 这种情况并非少见。这将会影响访调员的工作质量, 因此应尽快予以纠正。实地访问是进行这种纠正的主要手段。在这些实地访问期间, 一个有用的工具是“数据质量表”。实地工作期间, 定期采用这种“数据质量表”来查明具体问题以及与具体调查小组和访调员有关的问题。数据质量表含有受访者年龄以及小孩年龄方面的信息, 这些信息可用来检查访调员所选择的受访者是否正确。另外, 数据质量表还包括婴儿和儿童死亡方面的情况, 因而可以检查死亡儿童被遗漏的情况。此外, 为了衡量每个小组和访调员的工作效率, 以及检查是否存在有意忽略住户和/或受访者的情况, 数据质量表还包括了住户和个人答复率方面的信息。在这些数据质量表的检查期间所发现的问题将传达给实地, 以便未来的工作可以避免这些问题(另见下文0节)。

77. 编制住户清单是住户抽样阶段的一部分, 有关实地工作的本节没有对此进行介绍。如E节所介绍的那样, 这是一项单独的工作, 是在实地工作之前由编制住户清单的专业人员进行的, 一般需要两到三个月的时间。将编制清单的工作和主要的实地工作分开是为了确保对编制清单的工作进行严格监督, 并在主要实地工作开展之前由办事处的合格人员对住户进行抽样。将抽样作为办

事处的一项工作有助于避免由实地人员抽取住户时可能产生的偏差，当“编制清单的人员”和“抽样人员”属于同一人时，尤其会产生这种偏差。

0. 数据处理

78. 在人口与健康调查中，数据处理一般在实地工作开始后的一两周开始，通常在实地工作完成之后的一个月内完成。需要对数据录入人员进行调查问卷方面的培训，让他们参与部分访调员培训课程，或接受两三天的专门培训。数据处理协调人员通常参加访调员的全部培训课程。

79. 数据录入是在一个单独的房间内进行的，这样可以保证调查问卷的安全以及数据录入人员不被打扰。该房间应该靠近已完成的调查问卷的储存地。在数据录入和编辑期间，所有调查问卷都要处理好几次，因此如果储存地靠近数据录入设施，将能大大减少工作量和压力。数据录入人员每天的工作不超过6小时，因为数据录入是一种机械性的紧张工作。根据用于数据录入工作的计算机的数量，数据录入人员的值班数可能在一个以上，以便在实地工作结束后尽快完成数据录入和编辑工作。如果可能，应避免采取两班制，因为多个监督员和办事处编辑人员可能会导致不一致问题的出现。

80. 人口与健康调查政策是：对所有调查问卷的数据进行两次录入（“双录入”），对录入结果进行比较，解决任何不一致的问题。这种100%的检查大大减少了为解决不一致问题而进行二次编辑的工作量，能够获得更纯净、更准确的数据集。双数据录入由两位不同的数据录入人员进行，目的是为了获得最佳的录入效果。在数据录入期间，需要对每份调查问卷进行范围、跳过及一致性检查。

81. 数据录入和编辑的一个方面与数据质量控制直接有关。人口与健康调查的程序是要在数据录入和编辑期间定期获得一组选定的表格，以检查手册编辑和个别调查问卷数据录入期间不容易发现的问题。如，这些“实地检查表”是为了发现访调员是否为了减少工作量而在受访者或其孩子的年龄方面弄虚作假，是否少报了婴儿和儿童的死亡率，是否准确记录了死亡的年龄等。一旦录入了足够数量（如300份）的调查问卷，那么就可以启用这些实地检查表，之后则两周启用一次，其目的是为了访调员或访谈小组能够发现不正常回答或受访者的不正常特征。由来自实施机构以及人口与健康调查方案的人员对这些表进行检查。问题应该传达给有关的小组，以便能够采取纠正措施。

82. 为每个国家所编制的基本表格是根据核心调查问卷中所收集的数据设计的。如果数据来自核心调查问卷中添加的问题，那么有关这种数据的表格应该和那些需要这些特别表格的人员/机构一起进行设计。这种工作应该在早些时候进行，以确保制表过程得以顺利进行。应该由人口与健康调查人员和国家相关人员对所有表格进行全面检查。

83. 由于数据录入、编辑、插补和制表程序很复杂，所以这些程序由人口与健康调查的数据处理人员开发，并由这些人员前往有关的国家安装有关程序。通常情况下，数据处理专家将于数据处理结束后返回，以协助检查最后的

数据集，对某些变量重新编码，插补缺失的日期，为样本配上加权系数，启用以前为初步报告和最后报告制定的表格。尽管所做的计算总是采用最后的样本权数，但在为数据编制表格时，各种情况的加权和未经加权数据都应该显示在报告中。

P. 分析及报告的编写

84. 分析以人口与健康调查的一组示范表格为基础，示范表格由人口与健康调查国家管理人员以及东道国人员进行修改以适合所采用的调查问卷。除这些表格外，还可以增补适合国家具体需要的表格，以显示每个国家所收集的额外数据。分析后可编制有关调查数据的综合报告。

85. 另外还可就重要发现编制一份小型的报告，以尽可能广泛散发数据。有关重要发现的报告是在主要调查报告之后或者与该报告同时编制的，国家研讨会上可以获得这种报告（有关散发的内容，见下文Q节）。

86. 除了编制这些调查报告外，人口与健康调查还协助各国对调查数据进行更深入的分析。这些分析通常会得出30-60页的研究论文，主要是解决国家或供资机构特别关注的问题。但这些论文也可以包括特别的表格和简短的分析性说明，以便使国家可以对政策性问题和/或其他问题做出反应。

Q. 传播

87. 向所有相关的受众传播调查结果是调查方案的主要目标。调查报告在地方进行了广泛的分发，各自国家的合作机构和其他机构可以获取调查报告，人口与健康调查网站上也可以查阅和下载调查报告。为了更广泛地传播调查结果，除国家研讨会外，还以挂图、图集、日历、海报和其他材料的形式来进行传播。

88. 另外，还举行了国家研讨会向决策人员、方案管理人员、研究人员和捐赠组织代表介绍调查的主要发现。研讨会一般由媒体进行报道，这样就有助于将这些数据用于政策和方案之目的。有些国家还组织地区研讨会，以确保调查结果除用于国家政策和方案外，还能被人们了解并用于其他方面。

89. 人口与健康调查的所有数据都被输入人口与健康调查的数据档案。几乎所有参与该方案的国家都已经授权世界各地的责任研究员使用他们的数据。由人口与健康调查的数据档案小组对数据要求进行跟踪，并为那些有权使用这些数据的人们提供数据和文件。如今经过适当的电子注册和对每个用户授权后，可以通过互联网免费获得相关的数据。截止2002年底，ORC Macro为人口与健康调查数据文件和子文件提供了80 000多次的访问量。网址为：www.measuredhs.com。有关人口与健康调查方案的其他信息也可从该网站获得。

R. 人口与健康调查数据的使用

90. 人口与健康调查数据通常用于监督和评估参与国母婴健康和人口方案的进展。重复调查信息的存在为国家提供了用以衡量进展所需要的趋势数据。

有时数据被用于近期的行动计划，如在贫血严重的地区提供补铁物质。更多的时候，这些数据被用于制定政策、修改干预方案的目标和制定长期的健康和人口计划。人口与健康调查数据有助于促进人们为撒哈拉以南非洲以及其他地方的计划生育方案提供支持，因为调查数据显示甚至连某些最贫穷的国家也可以发生变化，而且正在发生着变化。

S. 能力建设

91. 人口与健康调查方案的目标之一是通过大规模的全国住户调查来提高参与国收集和分析数据的能力。实现这一目标的主要手段是编制先进的基本文件，如调查问卷和手册；开发能够促进发展中国家调查处理的软件程序；以及在国家调查的所有阶段对有关的地方人员进行在职培训。

92. 有助于能力建设的一个主要方面是新软件的开发。最初，人口与健康调查为调查处理开发了综合调查分析系统（ISSA）程序。这一软件的出现有助于早日获得“纯净”数据文件和报告。为了适应新型基本硬件和软件的发展需求，人口与健康调查和美国普查局以及一家软件开发公司一道推出了新的调查数据处理软件“普查与调查处理”（CSPPro）。该软件有望被广泛使用，并将代替不同机构处理大规模调查时所使用的各种程序。美国普查局已经为该软件使用方面的广泛培训方案提供了支持，预计该软件将成为大多数发展中国家的标准软件，这将会大大促进能力建设的举措。

93. 人口与健康调查方案一直通过有关调查实施的详细基本文件、定期技术支持访问（每个国家10-14次）以及合作编写调查报告等方式，为地方相关人员提供长期培训和信息反馈。基本文件包括有关所有重要调查实施阶段的手册。上述三种机制仍然是参与国能力建设的主要手段。

T. 吸取的教训

94. 在实施人口与健康调查及其以前方案的过程中，吸取了发展中国家住户调查的许多宝贵教训，如：

- 许多国家的抽样标架需要进行成本昂贵的实地更新才能用于收集高质量数据的调查。住户清单通常是过时或不存在的。为了进行质量控制，并确保所有住户都按已知的概率抽取，有必要在办事处抽取住户，不能将这种工作交给实地人员处理。在办事处抽取住户可以避免访调员只访问那些更方便的住户而不访问那些距离较远的住户。在办事处从住户清单中抽取住户可以确保对列入清单的住户进行无偏抽样，还可以方便实地抽样的监督。
- 在倒数第二个抽样阶段进行样本更新时，需要进行密切的监督，以确保全面地将所有住户列入清单。根据对若干项调查的观测发现，编制住户清单的人员可能会漏列边远、艰苦或危险地区的住所。如果没有行之有效的监督，由编制住户清单的人员所编制的清单可能会出现偏误。

- 住户和受访者个人一级的答复率普遍不错（见有关答复率的F节）。
- 必须针对每项调查中一组具有代表性的调查项目，进行抽样误差和设计效果方面的计算，以便对样本设计的有效性以及调查估计的精确性进行评估。
- 人口与健康调查中，理想的类集规模为15-20人。在这种调查中，需要在人口与健康的各种项目（有些项目的类集程度比其他项目高，有些涉及被抽样妇女的孩子）和数据收集成本之间进行平衡。
- 归因于类集的设计效果是类集规模 b 和组内相关系数 p 的递增函数。由于在样本设计具有可比性的情况下， p 在各国之间具有很强的通用性，因此正如第六章所介绍的那样，一项调查中的 p 、 b 和设计效果可用于另一个国家中一项具有可比性的新调查。
- 复杂调查中访调员和监督员的培训需要三到四周的时间。人口与健康调查培训通常需要三周的时间。但为了让实地工作人员做好充分准备，将培训时间延长一周或一周以上的情况有许多。调查的大多数问题源于实地工作人员，而非受访者。进行适当的培训和监督是避免这些问题的主要手段。
- 访调员和监督员可能会给调查带来严重的问题。因此有必要进行不断的严格控制，以避免某些访调员和监督员为了减轻工作负担而马虎了事和/或有意在样本或调查方面弄虚作假。人口与健康调查已经有充分的证据表明，访调员可能会将妇女和/或儿童列入不合格年龄的范畴，这样就可以不必对这些人进行访谈。尽管这种问题一般不会发生在所有实地工作人员身上，但却确实存在，而且通常情况下只限于访谈小组的少数人。因此，在实地工作的整个过程中都应不断保持警觉。
- 应努力让访谈时间平均不超过一小时。这一结论不是根据不同调查期限的实际实地试验得出，而是根据实地人员的反馈得出的。各人口与健康调查的期限差异很大，期限的长短取决于受访者的特征及其回忆有关日期和时间的容易程度。访谈时间少则10分钟，如对没有孩子和性活动的单身女性所进行的访谈；多则一个半小时以上，如对孩子很多且不易想起调查内容的女性进行的访谈。
- 实地后勤方面的一个重要问题是是否有合适的车辆接送调查小组人员。取得和使用实地工作车辆的费用很高，因为车辆应该是能够走山路且能容纳全部调查小组的大型车辆。如果没有合适的车辆，将会耗费更多的时间，并对小组士气产生不良的影响。即使有了合适的车辆，访调员和监督员也需要走很长的路程才能到达某些住所。因此，应在尽量减少行路困难的情况下将他们送到调查区。
- 实地后勤最困难的一个方面是：在访谈使用多种语言的国家，如何为受访者配备合适的访调员和调查问卷。为保证访调员的访谈大多数都按照受访者自己的语言以及该语言的调查问卷进行，有必要根据语言

能力来确定小组的成员，同时还应制定详细的部署计划以考虑各小组的语言要求。

- 数据录入人员需要参与访调员的培训课程，以便有能力进行数据录入和编辑工作。人口与健康调查问卷很复杂，参与访调员的培训能够让数据录入人员深入了解调查问卷的流程以及调查问卷各部分之间的关系。为了在交互式数据录入和编辑过程中采取相应的纠正措施，数据录入人员需要这方面的知识。
- 双数据录入可以节省编辑的时间，当然采取这种方法的费用看起来可能会很高。在早期的人口与健康调查中，数据只录入一次。稍后的调查则采用双录入的方法来发现那些无法通过范围和一致性检查程序发现的错误，并将编辑期间需要进行纠正的调查问卷数量降到最低程度。人口与健康调查负责数据处理的人员认为双数据录入对数据编辑所带来的好处远远超出了其所要花费的成本。
- 为获得高质量的数据，有必要在数据录入期间将已完成调查问卷中所遇到的问题不断反馈给实地。尤其在调查的早期阶段，应该立即告诉实地人员他们所犯的错误是什么，这样才能避免未来发生有关的错误。交互式数据录入是早期发现实地问题的很好办法。
- 有必要通过某些表格来显示编辑个人调查问卷时不易发现的答案模式。如，访调员是否为了避免访谈而有意将潜在的受访者列为更年长或更年轻的范畴？只有通过研究几百个访谈中受访者的年龄模式才能清楚发现这一性质的问题。
- 在许多国家，编制调查报告是最具有挑战性的工作之一。调查研究中的能力建设是人口与健康调查方案的目的之一。为加强能力建设，编写报告是需要做出重大努力的领域之一，可以通过和地方编写人员一道共同进行。最近，报告编写讲习班已经被认为是传授能力的一种更有效的方法，在讲习班期间，所有编写人员以及人口与健康调查人员一道对报告的各个章节进行研究。然而，报告的编写工作属于一种技术性的工作，并不是所有人（不管其在人口或健康方面的学识有多高）都能做好这一工作。
- 在抽样、数据处理和编写报告过程中迫切需要技术援助。在其他领域，这种协助通常是为了确保不同调查步骤得到及时实施。在许多（如果不是绝大多数的话）人口与健康调查中，给地方人员带来最大困难的是上述领域。相比较而言，许多地方机构的培训和实地工作都开展得很好。因此，有必要提供所需要的技术援助，以便克服一个或更多困难领域中所存在的薄弱环节。
- 各国愿意和责任研究员共享其调查数据。在调查实施前应该就这方面的计划达成协议。人口与健康调查方案已经非常成功地获得了参与国的同意，允许责任研究人员在未来的研究项目中共享他们的数据。一

个独一无二的多国数据库由此得以产生，这对于各国和捐赠机构来说极其宝贵。为实现这一目的，需要在达成调查协议之际和参与国的有关当局达成协议。如果不在这个时候达成这种协议，通常无法在事后和他们进行磋商，因为政府可能已经发生更迭，开始对调查进行规划时，现行政府部门的负责人也许已经换人。

附 件

1990-2000年期间44个国家66项调查的住户答复率和妇女答复率（部分地区）

地 区	国 家	调查年份	阶 段	住户数量	住户答复率 (%)	妇女数量	妇女答复率 (%)
亚洲	孟加拉国	1994	DHS III	9 255	99.1	9 900	97.4
	孟加拉国	1997	DHS III	8 762	99.1	9 335	97.8
	印度尼西亚	1991	DHS II	27 106	99.1	23 470	97.6
	印度尼西亚	1997	DHS III	34 656	98.8	29 317	98.3
	巴基斯坦	1991	DHS II	7 404	97.2	6 910	95.7
	菲律宾	1993	DHS III	13 065	99.5	15 332	98.0
	菲律宾	1998	DHS III	12 567	98.7	14 390	97.2
欧亚大陆	哈萨克斯坦	1995	DHS III	4 232	98.7	3 899	96.7
	哈萨克斯坦	1999	MEASURE	5 960	98.1	4 906	97.8
	吉尔吉斯斯坦	1997	DHS III	3 695	99.4	3 954	97.3
	土耳其	1993	DHS III	8 900	96.8	6 862	95.0
	土耳其	1998	MEASURE	8 596	93.8	9 468	90.6
	乌兹别克斯坦	1996	DHS III	3 763	98.4	4 544	97.2
拉美	玻利维亚	1994	DHS III	9 335	97.6	9 316	92.3
	玻利维亚	1997	DHS III	12 281	98.6	1 831	94.6
	巴西	1991	DHS II	6 416	94.5	6 864	90.7
	巴西	1996	DHS III	14 252	93.2	4 579	86.5
	哥伦比亚	1990	DHS II	8 106	91.4	9 715	89.0
	哥伦比亚	1995	DHS III	11 297	89.5	2 086	92.2
	哥伦比亚	2000	MEASURE	11 747	92.8	2 531	92.5
	多米尼加共和国	1991	DHS II	8 131	87.9	8 200	89.3
	多米尼加共和国	1996	DHS III	9 026	97.8	9 034	93.2
	危地马拉	1995	DHS III	11 754	96.1	3 388	92.6
	海地	1994	DHS III	4 944	97.5	5 709	93.8
	尼加拉瓜	1997	DHS III	11 726	98.3	4 807	92.1
	巴拉圭	1990	DHS II	5 888	96.5	6 262	93.1
	秘鲁	1992	DHS II	13 711	98.3	17 149	92.6
近东	埃及	1992	DHS II	10 950	98.3	9 978	98.9
	埃及	1995	DHS III	15 689	99.2	14 879	99.3
	摩洛哥	1992	DHS II	6 635	99.1	9 587	96.5
	也门	1991	DHS II	12 934	99.2	6 515	92.2
撒哈拉以南非洲地区	贝宁	1996	DHS III	4 562	98.6	5 719	96.0

地 区	国 家	调查 年份	阶 段	住 户 数量	住 户 答复率 (%)	妇 女 数量	妇 女 答复率 (%)
	布基纳法索	1992	DHS II	5 283	97.3	6 848	92.8
	布基纳法索	1999	DHS III	4 871	98.8	6 740	95.6
	喀麦隆	1991	DHS II	3 647	97.0	4 147	93.3
	喀麦隆	1998	DHS III	4 791	98.0	5 760	95.5
	中非共和国	1994	DHS III	5 583	99.4	6 005	98.0
	乍得	1997	DHS III	6 930	98.7	7 705	96.7
	科摩罗	1996	DHS III	2 277	98.9	3 160	96.5
	科特迪瓦	1994	DHS III	5 977	99.3	8 271	97.9
	加纳	1993	DHS III	5 919	98.4	4 700	97.1
	加纳	1999	MEASURE	6 055	99.1	4 970	97.4
	几内亚	1999	MEASURE	5 216	97.6	7 117	94.9
	肯尼亚	1993	DHS III	8 185	97.1	7 952	94.8
	肯尼亚	1998	DHS III	8 661	96.8	8 233	95.7
	马达加斯加	1992	DHS II	6 027	98.6	6 520	96.0
	马达加斯加	1997	DHS III	7 349	97.6	7 424	95.1
	马拉维	1992	DHS II	5 409	98.4	5 020	96.6
	马里	1996	DHS III	8 833	98.7	10 096	96.1
	莫桑比克	1997	DHS III	9 681	95.9	9 590	91.5
	纳米比亚	1992	DHS II	4 427	92.6	5 847	92.7
	尼日尔	1992	DHS II	5 310	98.7	6 750	96.3
	尼日尔	1997	DHS III	6 007	98.7	7 863	96.4
	尼日利亚	1990	DHS II	9 173	98.1	9 200	95.4
	尼日利亚	1999	MEASURE	7 736	98.8	10 529	93.2
	卢旺达	1992	DHS II	6 292	99.4	6 947	94.3
	塞内加尔	1993	DHS II	3 563	99.0	6 639	95.0
	塞内加尔	1997	DHS III	4 855	98.3	9 186	93.5
	多哥	1998	DHS III	7 620	98.6	8 964	95.6
	乌干达	1995	DHS III	7 671	98.4	7 377	95.8
	坦桑尼亚联合共和国	1992	DHS II	8 560	97.3	9 647	95.8
	坦桑尼亚联合共和国	1996	DHS III	8 141	97.9	8 501	95.5
	赞比亚	1992	DHS II	6 245	99.4	7 247	97.4
	赞比亚	1996	DHS III	7 365	98.9	8 298	96.7
	津巴布韦	1994	DHS III	6 075	98.5	6 408	95.6
	津巴布韦	1999	MEASURE	6 512	97.8	6 208	95.2

鸣 谢

作者对联合国为讨论本出版物草案而召集的专家组成员、外聘评审员以及ORC Macro的Alfredo Aliaga博士表示感谢，感谢他们为本出版物提供了宝贵的意见。

参考资料

- Cleland, J. 和C. Scott等（1987年）。《世界生育率调查：评估》。纽约：牛津大学出版社。
- 资源开发研究所/Macro系统国际公司。（1990年）。《人口与健康调查-I中数据质量的评估》。《方法报告，第1号》。马里兰，哥伦比亚。
- _____（1994年）。《人口与健康调查-I调查中健康数据质量的评估》。《方法报告，第2号》。马里兰，哥伦比亚。
- Kish, L.（1965年）。《抽样调查》。纽约：威利国际出版公司。
- _____ R. Groves和K. Krotki（1976年）。《生育率调查中的抽样误差》。《世界生育率调查专题文件，第17号》。海牙：国际统计学会。
- Lê, T. 和V. Verma（1997年）。《人口与健康调查中的样本设计和抽样误差分析》。《人口与健康分析报告，第3号》。马里兰，格鲁曼卡沃顿：Macro国际公司。
- ORC Macro咨询公司（1996年）。《抽样调查》。《人口调查-III的基本文件，第6号》。马里兰，格鲁曼卡沃顿。
- _____（2001年）。《调查安排手册》。《草案》。马里兰，格鲁曼卡沃顿。
- Scott, Christopher和其他人（1988年）。“逐字问卷与实地翻译或表格：实验性研究”。《国际统计评论》，第56期，第3号，第259-278页。
- Vaessen, Martin和其他人（1987年）。“将问卷翻译为本地语言”。《世界生育率调查：评估》。John Cleland和Chris Scott等。纽约：牛津大学出版社。
- Verma, V. 和T. Lê（1996年）。“人口与健康调查中的抽样误差分析”。《国际统计评论》，第64期，第265-294页。
- Verma, V.、C. Scott和C. O'Muircheartaigh（1980年）。“世界生育率调查中的样本设计和抽样误差”。《皇家统计学会志》，A辑，第143期，第431-473页。

第二十三章

生活水平衡量研究调查

Kinnon Scott

Diane Steele

Tilahun Temesgen

世界银行

世界银行

世界银行

美国哥伦比亚特区华盛顿 美国哥伦比亚特区华盛顿 美国哥伦比亚特区华盛顿

摘 要

为改善发展中国家制定、实施和评估社会经济政策所需要的住户统计数据，生活水平衡量研究（LSMS）方案应运而生。生活水平衡量研究方案关注的重点一直是：生活水平的了解、衡量和监测；政府支出和方案与住户行为之间的相互作用；政策的事先和事后评估；社会部门观测结果的起因。为此而进行的生活水平衡量研究调查采用了多种调查手段来获取实现这些目的所需要的数据，并采取大量质量控制措施以确保获取高质量的相关数据。尤其在最近几年中，生活水平衡量研究方案一直强调由数据用户参与调查的设计，并一直致力于可持续性问题的研究。本章将对生活水平衡量研究调查的内容、调查中的关键设计与实施方法以及提高分析能力的举措进行全面介绍。另外，本章还对调查成本以及所获取的数据质量进行了评估，这是对生活水平衡量研究调查数据进行政策使用的例子。本章还根据若干项生活水平衡量研究调查，讨论了某些住户和个人变量的平均样本设计效果和组内相关系数的计算。

关键词：贫穷衡量、生活水平、调查方法、设计效果、组内相关、质量控制。

A. 引言

1. 无论是绝对数还是相对数，社会服务和基础设施方面的公共部门支出都占用了相当数量的资源份额。卫生和教育支出各占国内生产总值（GDP）的3-4%，这并不足为奇。依国家而定，这种支出从数百万美元到数亿美元不等。有关税收和价格方面的重大经济政策变化将会对福利的相对水平和绝对水平产生重大影响。然而由于通常缺乏相关的数据，在制定、实施和修改各种政策时，很少了解这些政策对改善国家人口生活水平的总体效果如何。由于缺乏合适的住户数据，决策者不得不依靠行政数据。尽管这些行政数据可用于某些目

的，但如果要了解住户行为、政府政策对住户和个人的影响情况以及社会部门观测结果的决定性因素，了解这些情况的能力就会受到严重的限制。住户调查的任务是要弥补这方面的差距。

2. 政府可以采用而且也确实采用了生活水平衡量研究（LSMS）调查，作为更好地了解观测结果起因以及政府政策效果的一个手段。生活水平衡量研究调查不仅仅是对结果进行衡量，而且还对影响或导致这些结果的各种因素之间所存在的关系进行了确定。单一主题的住户调查能够提供相关主题方面具有一定深度的重要信息，但却不足以说明为什么会存在某些结果以及影响这些结果的因素有哪些。生活水平衡量研究调查一方面是为了探讨各种资源与住户特征之间的关系，另一方面是为了对政府的行为进行探讨，并以此来了解对每个部门产生影响的势力因素、各种行为或结果。让政府深入了解影响生活条件的各种因素有助于各项政策和方案的改善，而这反过来又有助于更有效地利用稀有的政府和私人资源，改善生活水平。

3. 生活水平衡量研究调查需要负责调查管理的国家政府、国家的主要数据用户、世界银行以及其他双边和多边捐赠组织的通力合作。⁴⁶ 尽管每项生活水平衡量研究调查都是以一组核心的概念为基础，但它们实质上都是定制的调查，以便满足各国政府在某个时期的特殊需要。主要的实施机构通常为国家统计局（NSO），它利用生活水平衡量研究发现的最有效技术，牵头开展调查问卷设计、样本设计和实地工作方法方面的工作。

4. 本章将对生活水平衡量研究进行全面的介绍。首先简单介绍方案的历史，其次介绍生活水平衡量研究调查的关键特征，接下来的章节则解释生活水平衡量研究的设计特征如何对所收集的数据质量产生影响。最后一节举例说明了生活水平衡量研究调查数据的使用方法。

B. 为什么进行生活水平衡量研究调查？

5. 生活水平衡量研究是为满足决策者对高质量数据的需要所做出的努力，于1980年开始进行。在经过五年的工作后，于1985年以科特迪瓦和秘鲁作为试点首次进行了生活水平衡量研究调查，这五年期间的工作包括：对现有住户调查进行审查；向研究人员和决策者进行广泛的咨询，确定所需数据的类型；向调查方法学者进行广泛咨询，了解如何进行实际实地工作程序方面的最佳设计等。首次进行的这两次调查属于特别进行的研究项目，目的是为了对所有方法进行检验，以确定所获数据的质量和有用性。⁴⁷ 由于这两项调查获得了成功，自1985年以来已经在40多个国家进行了60多项生活水平衡量研究调查（全部清单见附件一）。

C. 生活水平衡量研究调查的主要特征

6. 以下对生活水平衡量研究的主要特征进行了总结。欲详细了解调查及实施调查的方法，读者可参考1996年生活水平衡量研究手册。⁴⁸

⁴⁶ 此外，合作进行生活水平衡量研究调查的其他机构还包括：泛美开发银行、联合国的各种组织（如，联合国开发计划署、联合国儿童基金会、联合国人口基金）以及来自加拿大、丹麦、大不列颠及北爱尔兰联合王国、日本、挪威、瑞典和美国的双边捐赠机构。

⁴⁷ 生活水平衡量研究的详细介绍，见Grosh和Glewwe（1995年）。

⁴⁸ 见Grosh和Muñoz（1996年）。

1. 内容及所采用的工具

7. 生活水平衡量研究调查的独立调查工具有四个之多。这些工具是：(a) 收集住户、个人以及住户经济活动（农业和家庭企业）信息的住户调查问卷；(b) 收集有关住户所处环境的数据的社区⁴⁹ 调查问卷，重点收集现有服务、经济活动以及市场准入方面的信息，最近还包括社会资本方面的信息；(c) 在住户所在地区发放的价格调查问卷，以便对生活费进行调整；⁵⁰(d) 发放给地方服务提供者的设施调查问卷，以获得关于提供给住户的服务类型和服务质量的信息。图二十三.1显示的是所用工具与生活水平衡量研究调查的政策目的以及所需变量之间的关系。

图二十三.1
生活水平衡量研究目的与调查工具之间的关系

目的	指标	工具
个人与住户福利的衡量		
水平、分布及相关性	消费	
	收入	住户调查问卷
	财富、储蓄	价格调查问卷
	人力资本	
	人体测量	
分析政策		
谁从方案/公共支出中受益	服务的使用	
	谁接受服务	
	转移	
公共支出/方案的影响	服务成本	住户调查问卷
	政策的影响	社区调查问卷
服务的提供	离最近服务的距离	设施调查问卷
服务质量	所提供服务的类型	价格调查问卷
服务价格	人员、预算以及其他投入	
经济政策的效果	部门之间的转移净额	
确定决定性因素		
所观测的结果为什么会发生	住户构成	住户调查问卷
影响住户行为的是什么	人力资本、福利、现有的服务等。	社区调查问卷
		设施调查问卷
		价格调查问卷

⁴⁹ 注意这并不是社会学意义上的“社区”，而是收集在调查期间被选中的住户所在地区相关信息的一种手段。

⁵⁰ 国家消费价格指数通常不足以达到这一目的，因为这些指数通常是城市的指数，即使将农村地区包括进去，价格并不是在适当分类的级别取得的。

8. 调查工具的内容反映了在特定时期实施调查的国家所需要的首要数据。由于关心的重点是为了衡量生活水平的方方面面，所以住户调查工具将重

点收集对这些方面产生影响的各种主题信息。表二十三.1显示的是典型生活水平衡量研究调查的内容，摘自1997-1998年的越南调查。

表二十三.1

越南住户调查问卷的内容，1997-1998年

首次访问	第二次访问
住户花名册	生育
教育	农业、林业、渔业
健康	非农业自营职业
劳动力	食品费用与生产
移民	非食品与耐用品
住房及公用事业	汇款收入
	借款、贷款与储蓄
	人体测量

9. 为适应每个国家的具体需求，调查问卷进行了很大程度的改变，因此在过去几年中调查工具的总体内容发生了变化，而且还纳入了新的单元和主题。譬如，波斯尼亚和黑塞哥维那2001年的健康单元增加了有关抑郁的问题，目的是为了衡量这种精神疾病的发病率，并确定该疾病和其他福利以及劳动力市场参与之间的关系。危地马拉2000年增加了有关社会资本的单元，以便就贫穷的社会因素收集相关信息，如对社区/政府方案和集体活动的参与、被排除在社会之外的原因、对福利的认识、对正义的理解以及伸张正义的情况。阿尔巴尼亚、巴西、尼泊尔、牙买加、南非和塔吉克斯坦则针对贫穷方面的主观措施增加了有关问题，目的是为了考察这些措施和其他措施之间的关系。⁵¹ 表二十三.2举例说明了最近几年增加的单元样本。总之，尽管存在着一整套的标准单元，但每个国家的生活水平衡量研究调查都反应了调查时该国的重点、数据需要或所关注的问题。世界银行最近开展了一个研究项目，内容是“改善生活水平衡量研究调查的政策相关性”，该项目编写的一本新书根据主题列出了可以通过生活水平衡量研究数据解决的政策问题大纲，并为调查问卷的设计提供了指导。⁵²

⁵¹ 要详细了解危地马拉的社会资本工作，见世界银行(2002年b)。关于贫穷方面的主观措施，详情见Pradhan和Ravallion(2000年)、Ravallion和Lokshin(2001年)、Ravallion和Lokshin(2002年)。波斯尼亚和黑塞哥维那数据的分析正在进行中。

⁵² 见Grosh和Glewwe(2000年)。

10. 调查问卷设计阶段是一个旨在发现和纳入相关政策问题的过程。大多数国家都成立了数据用户小组或指导委员会，成员来自不同部委、捐赠机构、学术机构以及国家统计局(NSO)。该小组负责确定具体政策的数据需求，以确保收集适宜的数据。调查问卷设计阶段平均需要8个月左右，有尽可能多的人员参与。这一漫长过程带来的其他益处是产生了对相关数据的需求，并确定了对相关数据的所有权。其结果是提高了数据在政策领域的运用，而不是以其他方式获得数据。

2. 样本问题

11. 通常情况下，生活水平衡量研究调查属于国家调查，调查采用的是多阶段住户概率样本。⁵³ 总体样本规模较小（相对于其他几项调查而言），通

⁵³ 实际上，和大多数住户调查一样，所抽取的是住所，之后对选定住所内的所有住户进行访谈。请注意，在使用固定样本设计时，住所或住户是否被追踪将取决于固定样本的目的和后勤问题。

表二十三.2
新增单元示例

主 题	国家及年度
日常生活的活动	科索沃（2000年）、吉尔吉斯斯坦（1993年、1996年、1997年、1998年）、牙买加（1995年）、尼加拉瓜（1993年）
残疾	尼加拉瓜（1993年）
与艾滋病有关的死亡率所造成的影响	坦桑尼亚联合共和国-卡盖拉（1991-1994年）
识字和/或识数测试	越南（1997-1998年）、牙买加（1990年）、摩洛哥（1990-1991年）
精神卫生	波斯尼亚和黑塞哥维那（2001年）
私有化	波斯尼亚和黑塞哥维那（2001年）、吉尔吉斯斯坦（1996年、1997年）
冲击/脆弱性	玻利维亚（1999年、2000年）、危地马拉（2000年）、巴拉圭（2000-2001年）、秘鲁（1999年）
社会资本	危地马拉（2000年）、科索沃（2000年）、巴拿马（1997年）、巴拉圭（2000-2001年）
贫穷方面的主观措施	阿尔巴尼亚（2002年）、巴西（1996年）、牙买加（1997年）、尼泊尔（1996年）、南非（1993年）、塔吉克斯坦（1999年）
时间利用	危地马拉（2000年）、尼加拉瓜（1998年）、牙买加（1993年）、巴基斯坦（1991年）、摩洛哥（1990-1991年）、坦桑尼亚联合共和国-卡盖拉（1991-1994年）

通常为2 000-5 000个住户。限制样本规模的理由主要有两个。首先，为了关注质量以及权衡抽样误差与非抽样误差之间的关系（有关这一点的详细讨论，见下文第C.4节）。其次，生活水平衡量研究调查是为了重点分析各种决定性因素或住户特征之间的关系，而非准确估计具体的比率或均值。鉴于这些原因，生活水平衡量研究的样本规模相对较小，通常情况下，样本规模还不足以将调查结果分为若干个小地理区域，如国家、自治市或部门。

12. 尽管实际设计因国家和情况的不同而各异，但所有生活水平衡量研究调查都采用了概率抽样。⁵⁴ 调查对研究域进行了确定（城市/农村），而在每个域内采用了分层的两阶段类集设计。⁵⁵ 正如大多数住户调查一样，生活水平衡量研究调查采用了类集设计而非简单的随机样本（SRS）。这是出于成本方面的考虑，当然类集设计降低了估计的精确度（要详细了解采用多阶段抽样的样本设计效果，见下文第E.4节和附件三）。初级抽样单位（PSU）属于地理上定义的区域单位，是根据概率与规模成比例的方法抽取的。样本框架通常采用国家最新的普查资料，但在没有现成普查资料或普查资料不相关的情况下，则使用了其他资料。（见尼加拉瓜1993年生活水平衡量研究的基本信息文件，在该研究中，用投票登记册对过时的普查信息进行了补充；以及波斯尼亚和黑塞哥维那2001年生活水平衡量研究，在该研究中，由于内战的缘故，需要进行大量的清单编制工作。）

⁵⁴ 每项调查的基本信息文件能够提供单项调查样本设计的详细情况。有关内容可以访问生活水平衡量研究网站：<http://www.worldbank.org/lsms/>。

⁵⁵ 但在有些国家必须进行三阶段设计。

13. 一旦选定初级抽样单位，就应对这些初级抽样单位进行普查统计，以准确获得有关所有住所和住户的最新清单。编制清单这项工作在时间上应尽可能与实际调查的实地工作接近。为避免可能出现的任何偏差，这种工作不是由访调员进行的，而是由国家统计局的绘图部门进行。在获得有关初级抽样单位内所有住户的最新清单后，则系统抽取二级抽样单位（住户），通常情况下，每个初级抽样单位内的住户数量相同，通常在12到18户之间，然后向住户所有成员收集数据。尽管生活水平衡量研究调查的样本设计旨在覆盖整个国家，但在某些情况下，由于内部冲突或自然灾害的缘故，可能会将特定的地区排除在外。

14. 生活水平衡量研究调查估计一般要求使用样本权数。即使原有样本设计要求进行自行加权，如加纳、尼加拉瓜（1993年）和塔吉克斯坦，但由于不答复率存在着差异，所以也有必要在对数据进行分析时，采用不同的权数。事实上，大多数样本设计都是自行加权的。通常情况下，特定国家的样本设计要受该国分析因素的影响。如，规模小但属于国家关注的人口子群（少数民族、边远地区、从事特定经济活动的人群或在某个重要政府方案领域的人群）可能需要进行过采样，以确保有足够的样本来对他们进行单独分析。此外，这种样本设计需要在数据分析中使用样本权数。鉴于在生活水平衡量研究调查中采用的样本设计，应该记住的最后一点是，对数据的显著性进行统计检验时，也应该考虑设计的多阶段性质（有关这一问题的详细情况，见本出版物有关样本设计效果的各章节）。

3. 实地工作组织

15. 如上所述，生活水平衡量研究调查的目的决定调查的结构与内容：它们还体现在所使用的实地工作方法中。生活水平衡量研究调查的实地工作设计应该确保机动访谈小组收集数据时兼顾数据录入活动和严密的监督工作。⁵⁶应该至少对每个住户访谈两次，访谈间隔为两周。图二十三.2显示了实地工作的方式。实地工作的设计应该确保每个访谈小组每月完成两个选定社区（初级抽样单位）的访谈。小组于每月的第一周和第三周在第一个社区工作，而第二周和第四周在第二个社区工作。调查问卷的前半部分在第一次访问期间完成，第一次访问在第一周或第二周进行，但要视社区而定。在两次访问期间，录入第一次访问所获得的数据，并对误差进行检查。第二次访问是为了纠正第一次访问的误差，进行后半部分的调查，并留出一段固定的时间收集食品支出方面的信息。⁵⁷收集数据通常在12个月的期间内进行，以便在必要的时候进行季节性调整，但许多国家选择了更短的期限。

⁵⁶ 要了解访谈小组的详情，见附件二。

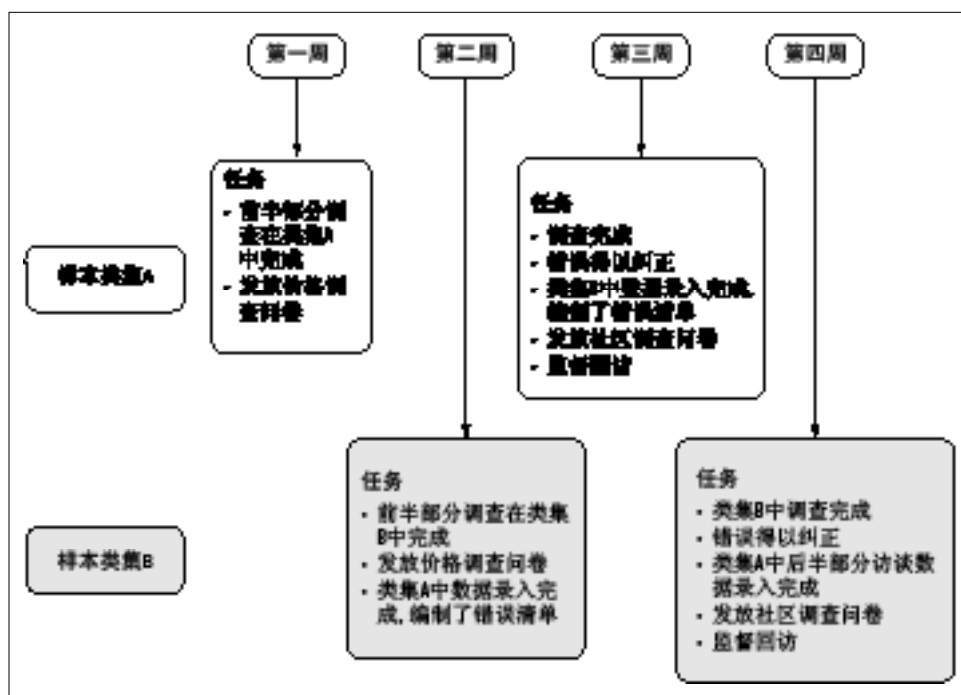
⁵⁷ 尽管正式计划的访问为两次，但调查问卷的所有部分均由答卷人直接回答，这说明为了对住户所有人进行访谈，访调员对每个住户的访谈次数实际上是根据需要确定的。

16. 监督员负责与收集初级抽样单位内住户数据的访谈小组一道对社区调查问卷和价格调查问卷进行管理。设施调查可能需要额外的人员进行管理。

4. 质量

17. 生活水平衡量研究调查目前最关注的一个问题是确保获得高质量的数据。调查的复杂性使得质量控制机制尤其重要。如表二十三.3所示，质量控

图二十三.2
各小组的一个月活动计划



制的形式有许多, 从最简单的形式(利用逐字逐句的问题、显性的跳过模式、被翻译为某个国家语言的调查问卷、尽量减少访调员误差的封闭式问题)到更加复杂的形式(包括同时进行数据录入、对住户立即进行重访以纠正不一致性错误或发现缺失数据)。很显然, 并不是所有这些质量控制措施都是生活水平衡量研究调查所特有的。但由于生活水平衡量研究调查所具有的复杂性, 所以一直强调在调查实施中采取一整套的质量控制措施。除了上述控制措施以外, 也许更具有争议的是, 生活水平衡量研究方案为了尽量减少非抽样误差而选用了小规模的样本。这么做的理由是, 尽管使用小规规模样本时的抽样误差可能会大一些, 但至少这种误差是可以量化的。相反, 非抽样误差的来源有许多, 也无法对其量值进行测量。但众所周知, 非抽样误差的总量往往随着样本规模的增大而增加。这样, 就决定限制这些非抽样误差, 即便在可能利用调查数据进行地理分类时, 这种做法也会限制地理分类的水平。生活水平衡量研究调查着重探讨生活水平各方面的关系, 而不是对具体指标或比例进行非常精确的衡量, 这说明上述决定对本调查的妨碍不如其他调查那么大。⁵⁸ 最后, 最新的方法使生活水平衡量研究调查数据(以及其他数据)和普查数据挂钩, 这就能够在普查数据内估算贫穷值, 这些方法有助于至少在贫穷和不等量度方面, 在某种程度上减少小样本规模问题。⁵⁹

18. 生活水平衡量研究调查所采用的另一个质量控制机制是使用直接答卷人, 又称受访者本人。这种机制主要有两个优点: 能减少受访人员的负担,

⁵⁸ 譬如, 劳动力调查为了说明在一段时间内失业率的变化很小, 需要的样本规模与分析失业决定性因素所需要的样本规模相比要大得多, 这是生活水平衡量研究调查分析的重点。

⁵⁹ 要了解这一方法的详情, 见下文第F节。

表二十三.3
生活水平衡量研究调查中的质量控制

质量控制领域	控制措施
调查问卷	<ul style="list-style-type: none"> • 逐字逐句的问题 • 显性跳过模式 • 尽量少用开放式问题 • 书面翻译为相关的语言^a • 敏感性问题的放在末尾 • 一整套：所有住户和个人数据使用一个格式
试点阶段	• 对调查问卷和实地工作进行正式试点测试
直接答卷人	• 个人以及最知情的人
同时录入数据	<ul style="list-style-type: none"> • 进行范围检查和一致性误差检查 • 回访住户以进行纠正
两轮模式	<ul style="list-style-type: none"> • 减少厌倦感 • 设定有限制的回访期间 • 可以对数据录入进行检查并对住户数据采取修正措施
培训	• 对访调员（一个月）、监督员和数据录入人员进行强化培训
分散式实地工作	• 机动小组，成员有监督员、两到三位访调员、携带计算机和打印机的数据录入人员以及司机和车
监督	• 每两到三个访调员配一名监督员
小样本规模	• 限制非抽样误差
数据存取政策	• 数据对所有研究人员和机构开放

^a 在一个国家中，如果某些语言没有书面形式（如，巴拿马的土著语言），那么将使用双语访调员。但这不是完善的解决方案，除非绝对必要，否则应避免使用。

⁶⁰ 即使对于“最知情的”受访者而言，实际访谈时间也被控制在一个小时以内，一般认为访谈有关人员的时间最多不能超过这个时限。但对于某些具体的住户来说，时间可能会超过这个限度，在这种情况下，需要谨慎行事，避免让答卷人感到厌倦，以及由此引起数据质量的下降。

⁶¹ 如果儿童年龄在10岁或12岁以下，或者如果家庭成员不能进行交流，那么可以使用代理受访人。如果使用代理受访人，应当注明实际受访者的身份编码。

并进而减轻受访者的厌倦感。住户调查问卷实际上是一系列短时间（10–15分钟）的个人访谈。只有最了解消费、农业和住户的受访者需要接受较长时间的访谈。⁶⁰ 使用直接答卷人还可以提高数据的质量，因为这样可以保证问题是由最了解情况的人员回答。⁶¹ 如果指望住户的任何一个人都能在健康、教育、劳动力、移民、信贷状况、生育状况方面或有关活动方面替其他所有家庭成员做出完整准确的回答，那么是不现实的。因为信息量太大。另外，在一个家庭中某些信息可能不愿意让家庭的其他成员知道（信贷、储蓄、收入、避孕工具的使用等活动都有可能属于不愿意让其他成员知道的信息）。在这种情况下，使用直接答卷人是准确获得有关每个家庭成员信息的惟一方式。应尽可能对访调员进行培训以便他们能够私下进行个别访谈。

19. 对每项生活水平衡量研究调查涉及的所有人员进行培训是进行质量控制的另一种机制。对于国家统计局的人员应该采取“在职”培训的形式，另外还需要有更正式的课程。调查方案投入了大量的资源对实地人员、访调员、监督员和数据录入人员进行正式培训。通常情况下，实地人员的培训时间为四周，既有理论，又有实践。完成培训后，根据培训课程的考试情况挑选实地工

作人员。最理想的方法是根据学员在课程期间参与活动的成功情况以及培训结束时正式考试的情况来挑选人员。

20. 提高数据质量的最后一个方法往往被忽视，这个方法是促进调查微观数据的公开使用。确保数据集能够被一系列研究人员和决策者广泛使用将有助于对现有数据进行认真检查。通过创建一个反馈环，以便向数据生产者进行反馈将有助于提高未来调查的质量。已经就大多数生活水平衡量研究调查数据集达成了公开使用数据的协议，并且采取了各种措施来帮助政府发布这种数据。尽管世界银行并不拥有生活水平衡量研究调查的数据集，但已经允许世界银行直接发布半数以上的调查数据集（事实上，有30%的数据集可以直接从生活水平衡量研究网站上下载）。⁶² 至于其余的数据集，一旦政府批准了个人请求，其中绝大多数都是可以分发使用的。根据申请这类许可使用权的人提供的反馈，大约有90%的申请获得了批准。

⁶² <http://www.worldbank.org/lsms/>。

5. 数据录入

21. 同时数据录入需要使用复杂的数据录入软件检查范围误差、记录内以及记录之间数据不一致的情况，如果可能甚至可以根据外部参照表（如，提供人体测量表、农作物产量数据和价格表）来对数据进行检查。数据是在收集数据阶段实地录入笔记本电脑的，数据录入人员是机动调查小组不可分割的一部分。每次访谈后立即对数据进行录入，数据录入过程中将产生一张有关误差、不一致情况和缺失信息的清单。之后由访调员返回住户向住户成员澄清所有问题，并填补任何缺失的信息。这种方法有助于防止在调查结束后对一大批数据进行整理。最好避免进行这种整理：尽管这种方式通常可以获得内部一致的数据集，但这些数据集却不能最好地反应每个人的情况。此外，这种方法还需要花费大量的时间，因此会耽误数据的使用，最糟的是，会使某些数据过时。随着既廉价、功能又强大的计算机的问世，以及新软件的开发，有些生活水平衡量研究调查将很可能采用计算机辅助个人面谈（CAPI）方式完全以电子的方式进行。由于这种方式可以缩短实地工作和数据公布之间的时间，并可以获得较高质量的数据，所以目前正在探讨这一方法。⁶³

⁶³ 采用计算机辅助个人面谈系统是美国普查局能够在实地工作的10天内公布其每月劳动力调查（当前人口调查）结果的因素之一。计划在2003年针对阿尔巴尼亚做一项实验，实验的内容是对计算机辅助个人面谈的成本和利得与生活水平衡量研究调查的同时数据录入进行比较。

6. 可持续性

22. 从最简单的角度来讲，可持续性、生活水平衡量研究调查的长期实施以及决策中对调查数据的使用所面临的三个最大障碍是预算限制、人员轮换以及分析能力的缺乏。尽管没有确保可持续性的计划，但生活水平衡量研究的经验已经为如何提高可持续性的可能性提供了若干项指导。首先，有必要让决策者和数据用户参与设计和分析阶段。这方面的基本入手点是创造有关生活水平衡量研究结果的需求以及在决策中使用有关数据。因为从有关数据中受益的正是这些终端用户（而不是国家统计局本身），而最愿意在政府预算分配过程中确保未来调查预算需求得以满足的也正是这一群体。通常情况下，在统计系统外寻找或发现一个或更多调查和数据“支持者”是保持可持续性的关键。⁶⁴

⁶⁴ 在这方面的一个例子是牙买加。需求最初来自总理办公室，计划部一直和实施调查的统计局参与调查设计和使用的每个阶段。在牙买加，自1980年代末以来，每年都进行生活水平衡量研究，有关该例子的详细内容见Grosh（1991年）。

23. 应该吸取的第二个重要教训是，实现可持续性是一个长期的过程：一次性调查的投资几乎没有什么长期影响。还需要经过若干年的努力，培训一批骨干人员，验证各种工具的有效性和使用情况，在生产者和用户之间建立联系，对有关方法进行修改以适应具体国家的需要和技能。此外，在做好调查工作的文件编制以及数据和传播活动的存档工作方面进行投资，有助于确保机构资料不因个别职员的离开而丢失。在已经进行过生活水平衡量研究调查的国家中，有近40%的国家已经进行过多项调查。

⁶⁵ 关于生活水平衡量研究调查在建设分析能力方面所吸取的各种教训，见Blank和Grosh（1999年）。

⁶⁶ 将数据收集人员和数据分析人员分开一直是个令人关心的问题。在消除障碍后必须牢记可信度的问题。

⁶⁷ 在这方面，为改进数据而提出的“21世纪统计促进发展伙伴关系（PARIS21）”倡议强调了可持续数据收集、分析和使用的重要性。

⁶⁸ 泛美开发银行-世界银行-拉加经委会项目题为“改进生活条件调查”，但更为人们所熟知的是西班牙文缩略语：MECOVI（拉加改进方案）。

⁶⁹ 见Ryten（2000年）。

⁷⁰ 譬如，在秘鲁，调查问卷为私营企业或研究人员留有一定的空格，这些企业和人员付费用后可以在调查问卷的任何指定部分增列特定问题。

⁷¹ 论述成本问题的本节是以Grosh和Muñoz（1996年）以及Juan Muñoz 在世界银行2002年2月26-28日举办的贫穷和不平等课程上的发言为基础。

24. 最后，建设分析能力应当是一个明确的目标。⁶⁵ 这将会扩大数据的用途，从而有助于创建人们对未来数据集的需求。另外，提高国家统计局人员的技能，以及国家统计局在政府中的形象也有助于留住有关人员。⁶⁶ 最后，外部力量也有助于提高数据的需求。从世界银行、国际货币基金组织（货币基金组织）以及千年发展目标获得优惠贷款的国家制定了各种消除贫穷战略，所有这些战略都需要用以衡量和监测贫穷和关键社会指标的数据。在这方面制定长期的目标有助于建立一个监测和评估系统，而监测和评估系统则在很大程度上依赖于住户调查，如生活水平衡量研究调查、行政数据和项目数据。⁶⁷ 最近对泛美开发银行-世界银行-拉丁美洲和加勒比经济委员会（拉加经委会）改善住户调查的项目进行了一次评估，⁶⁸ 评估强调可持续性具有长期性，并提出必须解决从捐赠筹资向地方筹资过渡这一问题。⁶⁹

D. 进行生活水平衡量研究调查的成本

25. 关注质量对实地调查的时间成本和资源成本都有重大影响。生活水平衡量研究调查成本在40万美元到150万美元之间，具体成本因国家和调查年度而定。按照每个住户计算，这与其他复杂调查（如，收入与支出调查、人口与健康调查）相当。当然，成本的变化也取决于国家统计局的能力、现有统计基础设施的状况、调查目标以及有关国家的交通状况。如果实施机构已经有了很好的基础设施和经验丰富的工作人员，那么将会大大降低调查的成本。每项调查的资金通常来自各个方面：政府预算（国家统计局的预算，以及来自其他机构的预算）、双边捐赠、多边捐赠以及信贷等。在有些情况下，私人部门也为调查提供部分费用。⁷⁰

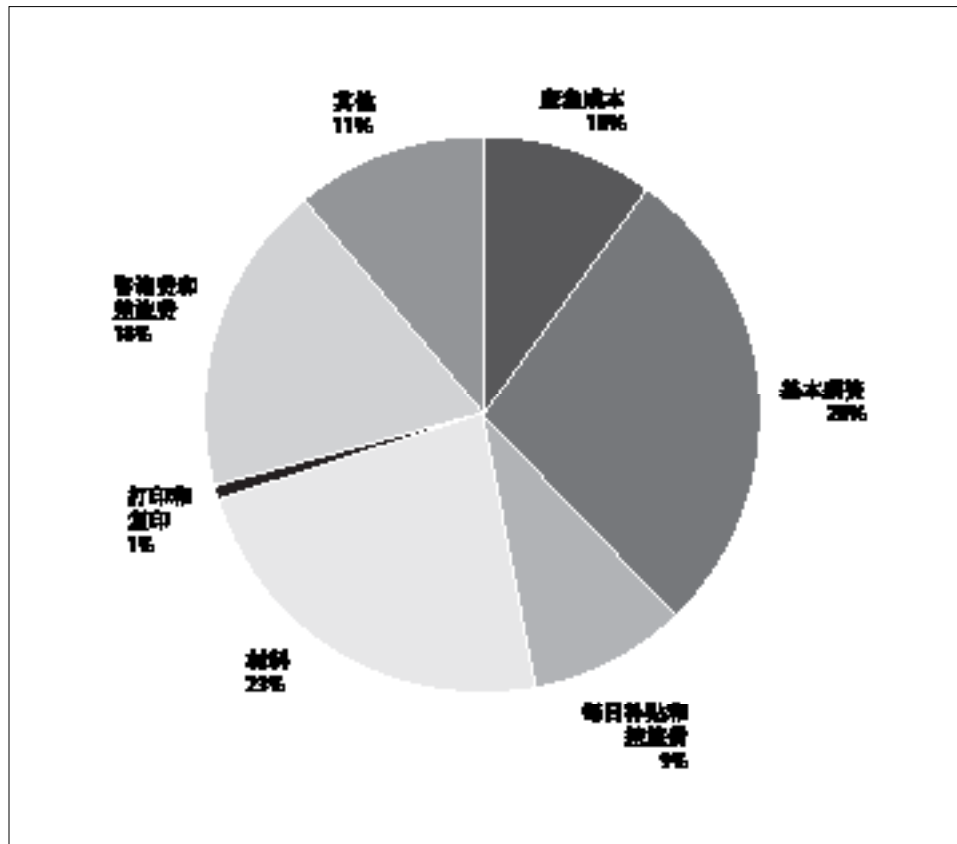
26. 一般来说，生活水平衡量研究调查的成本取决于所采取的方法、样本规模以及实地工作的复杂性。图二十三.3显示了生活水平衡量研究调查的成本构成以及各自所占的相对权数。⁷¹（帮助读者为生活水平衡量研究调查编制预算的简单入门练习见附件二。）

27. 在成本中占最大份额的是薪资。而在这一成本当中实地工作人员的薪资几乎占四分之三，实地工作人员包括：访调员、监督员、数据录入人员、人体测量人员和司机。生活水平衡量研究调查的实地工作人员许多（相对于样本规模而言），其原因包括：监督员和访调员之间的比例高（通常为1:3）；调查问卷内容的多寡以及直接答卷人的使用（这将会限制每天访问的住户数量）；将数据录入纳入实地工作小组；向每个小组成员提供交通以确保小组的流动性

和完整性。其他为办事处人员的薪资：通常为国家统计局的工作人员，当然，如有必要，可以聘请项目协调员。

图二十三.3.

生活水平衡量研究调查的成本构成（各部分在总成本中所占的份额）



资料来源：依据Grosh和Muñoz(1996年)，表8.2。

28. 在成本构成中所占的比例位居第二的是材料和设备费用。其中包括（购买或租赁的）计算机和车辆及其维护以及其他办公设备。这一部分的费用差异很大，主要取决于国家统计局或实施机构的现有基础设施。另外，如果禁止购置车辆，资金来源就会增加成本，因为租赁所需车辆的成本有时非常高昂。

29. 成本比例位居第三的是技术援助。技术援助成本的差异也很大，主要取决于实施机构的现有技能和经验。进行第二次或第三次生活水平衡量研究调查的国家所需要的技术援助和设备显然要少得多。通常情况下，技术援助方面最需要的技能是：抽样、调查问卷设计、数据录入定制、实地组织和分析技术。

30. 如果调查结果的数据质量确实很高，可以用来改进政策，那么生活水平衡量研究调查方面的支出是值得的。尽管从绝对数来说，这方面的费用很高，但相对于与社会政策支出的数额来说，生活水平衡量研究调查的费用并不高。下一节将使用最近的生活水平衡量研究调查的数据来说明生活水平衡量研

究调查数据的质量问题。主要从缺失数据、数据对于生活水平衡量研究的效用、内部一致性和设计效果等方面，举例说明质量。

E. 生活水平衡量研究设计对质量的效用如何？

1. 答复率

31. 衡量质量的首要量度是调查的总体答复率：样本中的住户是否对调查做出反应？或者，是否有大量的住户没有纳入调查，从而使最终结果可能出现偏差？⁷² 考察答复率很有用，因为这是衡量培训、调查问卷设计、访调员以及抽样程序（普查查点、地图的更新以及其他）质量的一个指标。许多国家的生活水平衡量研究调查都获得了很高的答复率。表二十三.4显示的是最近完成的生活水平衡量研究调查答复率情况。但生活水平衡量研究并非不受国家具体情况的影响。在冲突后国家，如果预期的信任度较低，答复率也就相应较低，波斯尼亚和黑塞哥维那以及科索沃的生活水平衡量研究调查就属于这种情况。但牙买加较低的答复率也许能更好地说明质量控制机制所起的作用。牙买加较少采用标准的生活水平衡量研究实地技术和质量控制措施：看来，这种做法确实使答复率较低。⁷³ 危地马拉较低的答复率之所以低，可能是由于编制住户清单和实地调查之间的间隔时间过长。在后来进行的访谈中，这一间隔为9-10个月。

⁷² 参与率系指整个家庭，而非该家庭的个别成员。

⁷³ 牙买加生活条件调查实地工作方面的详细情况，见世界银行（2001年）。

表二十三.4.

最近生活水平衡量研究调查的答复率

国 家	年 份	被抽中住所的数量	实际样本规模	答复率 ^a (%)
波斯尼亚和黑塞哥维那	2001	5 400	5 402	82.6
加纳 ^b	1998-1999	6 000	5 998	97.4
危地马拉	2000	8 940	7 468	83.5
牙买加	1999	2 540	1 879	74.0
科索沃	2000	2 880	2 880	82.0
吉尔吉斯斯坦	1998	2 987	2 979	99.7
尼加拉瓜	1998-1999	4 370	4 209	96.3
塔吉克斯坦	1999	2 000	2 000	..
越南	1997/1998	5 994	5 999	93.9

^a 波斯尼亚和黑塞哥维那、加纳、科索沃、塔吉克斯坦和越南采用了置换住户。答复率等于已完成访谈减去置换住户除以规划的样本规模。波斯尼亚和黑塞哥维那的置换住户为938；加纳为155；科索沃为519；越南为372。作者无法确定塔吉克斯坦的置换住户数量。

^b 在加纳调查中，对每个住户访谈七次。最终样本规模为参与所有七次访问的住户数量。

说明：两个点（..）表示没有提供资料。

2. 项目不答复

32. 计算项目不答复的百分比是质量的另一个指标。根据最早的生活水平衡量研究调查，对这一问题进行的审查显示项目不答复无关紧要（10个主要变量不答复率不到1%）。⁷⁴ 将生活水平衡量研究调查中的项目不答复率和没有采用相同质量控制措施的其他调查进行比较也令人关注。通常情况下无法进行这种比较，但本文将在这方面稍做比较。1998年对拉美各国的劳动力调查进

⁷⁴ 见Grosch和Glewwe（1995年）。

行了审查，这次审查就有薪工人、自营职业人员和雇主劳动收入缺失值的频数编制了资料。⁷⁵ 在被援引的国家中，有三个国家还在劳动力调查的当年进行了生活水平衡量研究调查。从表二十三.5中可以看出，在这些国家中，生活水平衡量研究调查的大多数比较指标确实比劳动力调查好（或至少一样好）。尽管例子有限，但这也似乎说明生活水平衡量研究在质量控制方面的投资起到了积极作用。

表二十三.5.

生活水平衡量研究和劳动力调查中收入数据缺失的频率

国 家	调 查	收入数据缺失百分比			直接答复者的百分比
		有薪工人	自营职业者	雇 主	
厄瓜多尔	劳动力调查, 1997年	6.3	6.7	13.2	..
	生活水平衡量研究, 1998年	3.6	8.5	6.5	96.5
尼加拉瓜	城市劳动力调查, 1997年	1.0	1.4	5.7	..
	生活水平衡量研究, 1998年	1.1	1.0	4.7	84.6
巴拿马	劳动力调查, 1997年	2.9	36.2	26.0	..
	生活水平衡量研究, 1996年	1.0	3.5	8.4	98.7

33. 也许检验数据质量的一个更好方法是数据被使用的程度而不只是对缺失响应的数目进行加总。对于生活水平衡量研究调查而言，其主要目标是对福利进行衡量，所以最重要的是要确定所收集数据适用于这种目的的程度。以货币形式衡量福利的最常用方式是家庭消费总量，无论在理论上还是实际当中，这种方式都有其优势。这是一种复杂的衡量方法，需要从调查问卷的一系列单元中获取数据：包括个人和住户的数据。通常情况下，消费数据来自住宅单位（采用住房、设施和其他住房支出的值）、耐用品单元（计算服务流的值）、教育单元（私人、现金支出）、食品消费单元（购买的、家庭生产的以及赠送的食品）、农业单元（没有纳入食品消费单元中的、家庭所消费的自产食品）以及非食品支出单元（从肥皂到家庭室内陈设等项目）。

34. 表二十三.6显示的是可以汇总消费总量指标的住户百分比。对于大多数调查来说，由于缺乏数据，所以没有包括在分析中的住户很少。其中的一个例外是加纳，1998年，加纳的主要问题究竟是什么尚不明确：样本规模比其他国家大一些，但不如危地马拉，危地马拉的规模要比其他国家大许多。有些食品消费数据可能是通过日记收集的（而不是采用标准的生活水平衡量研究方法收集），这可能是其中的一个因素：遗憾的是，调查记录并没有涉及这一问题。⁷⁶

3. 内部一致性检查

35. 确保数据内部的一致性也是至关重要的。由于调查工具很复杂，所以访调员难以在访谈过程中对内部的一致性进行监督，这也说明为什么需要采取

⁷⁵ 见Feres（1998年）。尽管收入不是劳动力调查或生活水平衡量研究调查的中心，但这两类调查采用了类似方式收集收入信息。

资料来源：劳动力调查的资料来自Feres（1998年）；生活水平衡量研究调查资料由作者计算得出。

说明：在尼加拉瓜1998年生活水平衡量研究调查中，缺失数据的百分比不包括零，因为根据访调员说明，如果有关人员得到的不是现金而是实物，那么访调员应该将其编码为零反应。本类别中的农民是自给农民，他们的收入在农业生产单元中的其他地方进行了计算。

两个点（..）表示没有提供数据。

⁷⁶ 见加纳统计署（2000年）。

如此多的质量控制措施来解决一致性问题。表二十三.7列示了内部一致性检查的三个例子。第一项检查用以确定社区调查问卷和住户数据的关联程度；第二项检查确定花名册中具有完整的上学/学前信息的学前儿童或学龄儿童的百分比；第三项检查确定在劳动力单元中被确定为自营职业的人员是否在非农业家庭企业单元中报告了他们的活动详情。

表二十三.6.

有完整消费总量指标的住户：最新的生活水平衡量研究调查提供的例子

国 家	年 份	最终样本规模	有完整消费总量指标的住户 (%)
波斯尼亚和黑塞哥维那	2001	5 402	99.9
加纳	1998-1999	5 998	87.7
危地马拉	2000	7 468	97.4
牙买加	1999	1 879	99.8
科索沃	2000	2 880	100.0
吉尔吉斯斯坦	1998	2 979	99.4
尼加拉瓜	1998-1999	4 209	96.0
塔吉克斯坦	1999	2 000	100.0
越南	1997-1998	5 999	100.0

表二十三.7

数据的内部一致性：单元之间的成功联系（百分比）^a

国 家	住户调查与社区调查 ^a	单元之间的正确联系：		
		花名册与教育单元 ^b		就业单元与非农业家庭企业单元 ^c
		学 前	小 学	
波斯尼亚和黑塞哥维那	...	99.5	99.8	90.4
加纳	99.9	..	96.5	70.2
危地马拉	100	100	100	93.0
牙买加	96.4	..
科索沃	100	..	100	58.6
吉尔吉斯斯坦	100	86.5	98.4	93.1
尼加拉瓜	..	97.9	97.5	62.0
塔吉克斯坦	100	..	99.9	..
越南	100	..	99.6	98.1

说明：表提及正确联系的百分比。波斯尼亚和黑塞哥维那、牙买加和塔吉克斯坦未包括社区调查问卷。牙买加、科索沃、塔吉克斯坦和越南未包括有关学前的特别单元。牙买加和塔吉克斯坦没有收集有关非农业家庭企业的信息。

两个点(..)表示没有提供资料。

^a 住户和住户所在社区相比较。

^b 花名册中的年龄变量与教育单元中的个人情况相比较。

^c 在就业单元中表示自己为自营职业者的人员与非农业家庭企业单元中的信息相比较。

36. 正如表中所显示的那样，前两项检查显示数据质量很高。但第三项检查却发现了问题。这说明在对住户进行的两次访问期间缺乏实地控制。只有

越南对问题进行了明确提问，让访调员在第二次访问中确保完成这一单元。显然，所有调查都需要类似的检查。

4. 样本设计效果

37. 判断生活水平衡量研究调查的最后一个标准是样本规模和设计。如果所使用的住户调查数据属于多阶段、分层和类集的复杂设计，在计算估计值的真实方差时，采用的方法是考虑样本设计的这些特征以及加权。设计效果等于估计值的真实方差（考虑到多阶段样本）与采用同等规模的简单随机样本所获得的估计值方差之比。⁷⁷ 因此，设计效果为“1”说明并没有因为使用多阶段设计而影响样本估计值的精确度；设计效果大于“1”说明采用多阶段设计降低了样本的有效性和估计值的精确度。

⁷⁷ 欲详细了解样本设计问题，见本出版物本章附件三或其他章节。

38. 作为生活水平衡量研究活动的一部分，还对某些早期生活水平衡量研究调查中关键变量和指标的设计效果进行了审查。由Temesgen和Morganstein（2000年）进行的该项审查强调，在采用生活水平衡量研究调查数据（当然还有采用多阶段样本设计的其他住户调查数据）和设计有关样本时，必须考虑几个要点。⁷⁸ 其中主要的一点是，由于生活水平衡量研究调查涉及多个主题，所以设计一个有效样本的过程会更加复杂。个人变量和住户变量的设计效果差异很大，这一点可以从表二十三.8中看出，该表摘自Temesgen和Morganstein（2000年）。总之，尽量降低一个变量的设计效果将很可能会提高其他变量的设计效果。其次，非抽样误差与抽样误差之间的折衷很明显。生活水平衡量研究调查中的设计效果值可能很高。下表显示，如果将生活水平衡量研究调查用以计算均值、比率和点估计，至关重要的是要对样本设计进行考虑，并应慎审注意数据的正确使用。

⁷⁸ 摘自Temesgen和Morganstein报告的若干个表载于本章附件三。

表二十三.8

生活水平衡量研究调查设计效果示例

国 家	人均消费			保健机会			失 业 率		
	全部	农村	城市	全部	农村	城市	全部	农村	城市
科特迪瓦, 1988年	6.7	3.6	5.5	6.3	5.7	2.2	7.0	4.4	5.7
加纳, 1987年	1.9	3.1	1.8	2.9	3.0	5.0	1.7	1.5	2.0
加纳, 1988年	3.2	2.9	2.9	2.2	2.5	3.6	1.3	1.1	1.4
巴基斯坦, 1991年	1.6	1.1	2.6	5.0	4.0	5.2	4.6	4.7	2.5

资料来源: Temesgen和Morganstein (2000年)。

39. 值得一提的是，正如其他调查一样，不仅各个变量的设计效果不同，而且一个国家内，不同地区的同一变量和不同时间的特定变量，其设计效果也不同。最后，不同国家之间的设计效果可能存在着很大的差异。如果以前曾经有过相关调查，对这些调查的组内相关性和设计效果进行认真审查将有助于改

进未来生活水平衡量研究调查的设计。在编制和解释生活水平衡量研究调查以及采用多阶段样本的其他调查所得结果时应该谨慎，因为所采用的样本设计可能很复杂。

F. 生活水平衡量研究调查数据的使用

40. 在过去几年中，生活水平衡量研究调查数据已经被广泛用于各种政策和研究之目的。其他一些出版物已经按照年代顺序对其中一些调查进行了记录，⁷⁹ 有关读者可以在生活水平衡量研究网站上找到各种基于生活水平衡量研究调查数据的论文和报告的书目，但只有一部分书目。该书目显示了采用生活水平衡量研究数据进行分析的范围，但数据的使用肯定不只限于书目中所列的范围。研究的不断进行以及调查问卷的修订说明数据使用的范围正在不断发生变化。为了说明使用生活水平衡量研究数据的各种使用方法以及将该数据和其他数据结合使用的方法，也许更有必要集中对其中一种用途（如，政府方案的目标针对的是贫穷人口）进行说明，而不是试图对这些数据的用途进行全面考察。

41. 首先，来自牙买加的一个早期例子显示了一项简单的分析如何能够为政府提供明确信息，说明采用备用方案将目标针对贫穷人口所产生的影响。正如Grosh（1991年）所概括的那样，在牙买加，有关营养的方案主要有三个：普遍食品补贴、食品券和学校供餐方案。根据牙买加进行的生活水平衡量研究调查，可以对贫穷家庭从三个方案中所获得的福利价值进行量化。分析结果显示：和其他两个方案不同的是，食品补贴为贫穷家庭所带来的福利价值明显下降。这种分析结果也是政府决定取消补贴而为另外两个方案增加资源的原因之一。

42. 可以利用生活水平衡量研究调查数据产生的第二个工具用以确定贫穷地区的地理目标。通过结合使用普查数据，生活水平衡量研究调查数据可用于绘制贫穷图，从而有助于向贫穷地区分配资源和出台相关的方案。⁸⁰ 采用这种方法，需要有生活水平衡量研究调查和相隔只有几年的普查数据。⁸¹ 生活水平衡量研究调查能够提供有效的福利指标（住户消费总量），但由于样本规模很小，所以有关贫穷数据的分类只能限于城市和农村地区，以及有关国家中少数的几个大地理区域。很显然，这不能满足政府向贫穷地区集中分配资源的所有需要，而在分散管理的体制下，这也无助于向地方政府分配资源。另外，在大区域内，人口贫穷水平的异质性程度很高，小规模样本的住户调查通常不会发现这种异质性。

43. 为了能够在较小的归并程度上提供贫穷信息，需要一个数据集，其样本规模比生活水平衡量研究的样本规模大若干个数量级。对于任何国家而言，最大的数据集当然是普查。但因为普查涉及整个人口，所以从每个住户收集的信息很有限，而且普查也通常是每10年才进行一次。因此无法根据普查数据构建有效的贫穷量度。一项创新性的工作能够使调查数据和普查数据相联系，目前正在对这项工作进行检查。这种方法利用了生活水平衡量研究所提供的福利

⁷⁹ 譬如，见Grosh（1997年）。

⁸⁰ 至于如何利用调查中的福利指标来绘制贫穷图以及如何结合使用普查数据，详细情况见：Hentschel等（2000年）；Elbers、Lanjouw和Lanjouw（2002年；2003年）；Elbers等（2001年）；Demombynes等（2001年）。正在做出进一步努力，以便采用这种方法将两种调查结合起来，但无法根据这种结合来准确估计标准误差。

⁸¹ 可以采用其他住户调查，但条件是这种调查能够提供货币形式的可靠福利指标，如消费总量或总收入。

指标和普查所提供的覆盖面。该方法利用普查和调查中同时出现的一组变量来对生活水平衡量研究调查数据中的贫穷进行估计。然后将这种方式所估计的参数和普查数据结合在一起，来对每个住户贫穷的概率进行预测，并利用普查数据来计算小地区的贫穷人口比。得出的贫穷图可以作为政府进行资源分配的一种工具。如，厄瓜多尔、危地马拉、马达加斯加、尼加拉瓜、巴拿马和南非就有这种贫穷图。

44. 将生活水平衡量研究调查数据用以改进社会方案目标制订的第三个例子产生于尼加拉瓜对应急社会投资基金（或西班牙首字母缩写词：FISE）进行的评估。评估涉及目标制订问题以及应急社会投资基金对社区用水、公共厕所、教育、保健和排水设施的投资所产生的影响。⁸² 在本例中，尼加拉瓜所计划的是全国性的生活水平衡量研究调查，调查对住户进行过采样，既包括应急社会投资基金项目区的住户，也包括没有应急社会投资基金方案的类似地区的住户。其他数据来源是用以评估项目行政支出的项目和行政记录。

⁸² 评估目标、所采用方法及其结果的详细情况，见世界银行（2000年）。

45. 在应急社会投资基金社区和类似的非应急社会投资基金社区进行的采样有助于创建控制组和处理组来对应急社会投资基金投资的影响以及相关目标的有效性进行衡量。另外，还根据生活水平衡量研究调查中的国家样本，采用倾向对照法创建了另外一个控制组，这加强了评估范围的深度。在社区一级（应急社会投资基金投资在实施方案的社区是逐步增加的吗？）和个人一级（在实施应急社会投资基金方案的社区内，较贫穷的人口是否会或多或少地从应急社会投资基金投资中受益呢？）都对确定目标的有效性进行了评估。

46. 评估利用在统计上显著的结果，显示了确定目标的总体有效性，并进而有助于根据确定目标的标准对主要项目类型进行评估。研究显示，排水项目的有效性明显下降，而公共厕所和基本教育项目的有效性则在有系统地提高，最贫穷人口中有17%的人从中受益。评估的直接结果是暂停了排水项目，决定集中增加最贫穷社区的投资，扩大最贫穷社区的受益范围。对应急社会投资基金方案进行的这种复杂评估，其花费的成本占评估之日前项目投资的1%。

G. 结论

47. 生活水平衡量研究调查的结果已经显示了该方法所具有的价值。政府已经利用有关数据来了解现行政策的效果；重新制定各种政策；改进向团体和地区分配资源的目标。注重质量已有成效，错误的情况减少了，数据的用途也更广了。但在使用这种方法时，需要进行适当的权衡。其费用相对较高，较小的样本规模限制了分类的水平，事先需要花许多时间进行计划和设计。但一旦工作开始，就可以迅速产生有关数据，与决策者的联系增加了数据的用途。

48. 显然，将生活水平衡量研究调查纳入国家住户调查系统有各种优点。至于多长时间进行一次这样的调查将取决于若干个因素。首先，国家在分析方面的需求应该足以让政府决定在一定时间内进行一次或多次调查。尽管可以根据截面数据（目标的确定、发生率、采用倾向对照分数方法的影响）对许多政

府方案进行评估，但在对不同时间的变化以及政策和事件的影响进行其他方面的分析时，需要有重复的截面数据和固定样本数据集。

49. 有关生活水平衡量研究调查频率的第二个考虑因素与国家的分析能力有关。为了向决策者提供相关信息，以便未来进行的每一轮调查都可以根据以前调查的结果进行改进，需要对数据进行分析。如果不能迅速对数据进行分析，那么对多轮调查进行的投资可能就没有效果。在这种情况下，就有必要在各轮调查之间留出时间间隔（如，三年）。

50. 最后，在决定具体调查的频率和时间时，预算和后勤问题也和各种实质性问题一样重要。因此，在决定一项调查的频率时，需要权衡该调查结果和其他调查结果的重要性。另外，应该记住任何一种来源的数据都无法满足所有的需要。政府在制定宏观经济政策和微观经济政策时，需要行政记录、项目管理信息系统（MIS）数据以及住户调查系统。将生活水平衡量研究调查和国家整个调查系统结合起来将大大促进人们了解政府政策和支出如何影响人们的生活。

附件一

生活水平衡量研究调查清单

国 家	年 份	住 户 数
阿尔巴尼亚	1996	1 500
阿尔巴尼亚	2002	3 600
亚美尼亚	1996	4 920
阿塞拜疆	1995	2 016
玻利维亚	1999	..
玻利维亚	2000	5 032
玻利维亚	2001	..
波斯尼亚和黑塞哥维那	2001	5 402
巴西	1996-1997	4 940
保加利亚	1995	2 500
保加利亚	1997	2 317
保加利亚	2001	2 633
柬埔寨	1997	6 010
中国：河北和辽宁	1995和1997	780
科特迪瓦	1985	1 588
科特迪瓦	1986	1 600
科特迪瓦	1987	1 600
科特迪瓦	1988	1 600
厄瓜多尔	1994	4 500
厄瓜多尔	1995	5 500
厄瓜多尔	1998	5 801
厄瓜多尔	1998-1999	5 824
冈比亚	1992	1 400
加纳	1987-1988	3 200
加纳	1988-1989	3 200
加纳	1991-1992	4 565
加纳	1998-1999	5 998
危地马拉	2000	7 276
几内亚	1994	4 705
圭亚那	1992-1993	5 340
印度：北方邦和比哈尔邦	1997-1998	2 250
牙买加	1988-2000（年度）	2 000-7 300
哈萨克斯坦	1996	1 996
科索沃	2000	2 880

国 家	年 份	住 户 数
吉尔吉斯斯坦	1993	2 000
吉尔吉斯斯坦	1996 (春)	..
吉尔吉斯斯坦	1996 (秋)	1 951
吉尔吉斯斯坦	1997	2 962
吉尔吉斯斯坦	1998	2 979
马达加斯加	1993	4 504
马拉维	1990	6 000
毛里塔尼亚	1987	1 600
毛里塔尼亚	1989	1 600
毛里塔尼亚	1995	3 540
摩洛哥	1991	3 323
摩洛哥	1998	..
尼泊尔	1996	3 373
尼加拉瓜	1993	4 200
尼加拉瓜	1998-1999	4 209
尼加拉瓜	2001	4 290
尼日尔	1989	1 872
尼日尔	1992	2 070
尼日尔	1995	4 383
巴基斯坦	1991	4 800
巴拿马	1997	4 945
巴布亚新几内亚	1996	1 396
巴拉圭	1997-1998	4 353
巴拉圭	1999	5 101
巴拉圭	2000-2001	8 131
秘鲁	1985	5 120
秘鲁 (仅利马)	1990	1 500
秘鲁	1991	2 200
秘鲁	1994	3 500
俄罗斯联邦 ^a	1992	6 500
南非	1993	9 000
塔吉克斯坦	1999	2 000
坦桑尼亚联合共和国:卡盖拉	1991-1994	840
坦桑尼亚联合共和国:全国	1993	5 200
突尼斯	1995-1996	3 800
乌干达	1992	9 929
越南	1992-1993	4 800
越南	1997-1998	5 999

注意:两个点(..)表示无数据。

^a 1992年俄罗斯纵向监测调查由世界银行提供资金。之后的调查没有世界银行的参与。更详细的信息,请访问卡罗莱纳人口中心网站:http://www.cpc.unc.edu/projects/rlms/rlms_home.html。

附件二

生活水平衡量研究调查的预算编制

正如第二十三章所述，任何两项生活水平衡量研究调查都不会完全一样，任何两个国家统计局不会完全一样，不同国家的薪资、交通和设备等费用也不一样。因此，无法提供相关的信息来说明特定时间特定地方的生活水平衡量研究调查将需要多少费用。本章将举例说明不同成本在调查总成本中所占的份额。为帮助读者对预算编制有个初步的了解，下面将进行一次小练习，该练习将为实地工作中最基本的薪资成本估算提供快速指导。根据该项指导以及有关国家的实际成本，就可以对生活水平衡量研究调查的大致成本进行估算。

鉴于调查工具和使用直接答复者的复杂性，平均而言，每个访调员每天可以完成两个“半场访谈”（如何实施调查的内容见图二十三.2）。换句话说，访调员每天可以完成两个住户的一轮访谈。如果假设工作日为六天（不管是每周“休假”还是其他方式的每月休假），每个访调员每月可以完成对24个住户的访谈。

假设需要4 000个住户的样本，这样，如果每个访调员每月可以访谈24个住户，总共需要167个访调员对4 000个住户进行访谈。如果实地工作需要12个月的期限，那么将需要14个访调员。每对访调员需要有一位监督员、一位数据录入人员、一位司机和一辆车。这样，实地工作人员的总人数（不包括国家统计局派出的区域监督员）包括：

14名访调员
7名监督员
7名数据录入人员
7名司机

如果规划人员采用以下参数，调查中实地工作的薪资成本将为：

项 目	每人每月成本	月 数	总 成 本
14名访调员	500	13	91 000
7名监督员	575	13.5	54 338
7名数据录入人员	525	14	51 450
7名司机	300	13	27 300
实地薪资成本的粗略估计数			224 088

说明：尽管实地工作时间仅为12个月，但还应加上一个月的培训费用（培训期间通常要给予实地工作人员某种报酬）和/或可能发生的调查工作延期费用。如有必要，还应为数据录入人员留出一个月，以便对数据集进行最后的整理。

根据图二十三.3的数据，实地工作人员的费用占调查总薪资成本的四分之三，占总调查成本的28%。在这种情况下，通过简单的计算，调查成本的大致估计数为1 067 086。

显然，这只是一个大致估计的数字。另外，还需要有关其他成本（如技术援助成本等）的详细数字。但在着手对实际调查编制预算时，这种简单的入门练习很有用。至于如何为生活水平衡量研究调查设计切实可行的预算，其详细情况见Grosh和Muñoz（1996年）第8章。

附件 三

⁸³ 本附件在很大程度上利用了Temesgen和Morganstein (2000年) 以前的工作。

生活水平衡量研究调查中样本设计对精确度和有效性的影响⁸³

A. 引言

本出版物的其他章节对抽样问题，尤其是复杂样本设计或多阶段抽样设计对估计值方差的影响进行了详细介绍。这种所谓的设计效果适用于不是采用简单随机样本的所有调查，如生活水平衡量研究调查。设计效果是总抽样误差的一部分：通过多阶段类集设计获得的估计值与采用简单随机样本设计获得的估价值之差。在本附件中，我们将对主要问题进行总结，并说明样本设计对若干项生活水平衡量研究调查的实际影响。

B. 抽样误差、设计效果和相关部分的计算

在简单的随机样本中，所有被抽中单位都有相同的独立抽选概率。但出于后勤和成本方面的考虑，住户调查几乎从不采用简单随机抽样的方法，而是像生活水平衡量研究调查那样，采用包括分层和类集在内的较为复杂的多阶段抽样设计。这将会影响估计值方差的计算以及样本本身的有效性。为了计算一个阶段以上的样本设计的抽样误差，有必要了解设计中用以确定分层、初级抽样单位 (PSU) 和加权步骤 (如果有的话) 的变量。一旦确定这些变量，就可以采用各种统计软件包来计算所需要的量度。⁸⁴

⁸⁴ 此处的计算采用了统计软件WESVAR。可用以估算复杂样本设计的抽样方差和各种相关统计量的其他程序软件有：CENVAR、CLUSTERS、Epi-Info、PC CARP、SUDAAN、VPLX和STATA。其中一些软件包可以免费从万维网上下载。

在本报告中，生活水平衡量研究调查中被抽中住户和个人变量方面的抽样误差量度包括：标准误 (*SE*)，在计算标准误时需要考虑样本设计的复杂性；变异系数 (*CV*(%))；样本规模 (*n*)；设计效果；组内相关系数 (ρ)；置信区间 (*CI*) 的上下限；以及有效的样本规模 (*EFFn*)。第二章、第六章、第七章以及其他各章对这些术语进行了界定。

C. 根据生活水平衡量研究调查计算的标准误、设计效果和组内相关

计算调查变量抽样误差的一个重要方面是要对样本设计的有效性 (精确度) 进行相互比较；并将该样本设计的精确度和假设的同等规模简单随机样本设计可能获得的精确度进行比较。除了显示现有调查数据的可靠性外，这种比较还可以帮助分析人员评估特定设计的实施情况的好坏，并可以为未来调查的设计提供相关的信息。以下三个表格对若干变量的设计效果和相关量度进行了比较，以便显示 (a) 同一国家不同变量之间的差异；(b) 同一国家不同时间的差异；以及 (c) 不同国家之间的差异。⁸⁵

⁸⁵ 完整的报告见Temesgen和Morganstein (2000年)。

正如表A三.1所示，在一个国家内，同一调查的不同变量在设计效果方面存在着很大的差异。该表的数据以1987年加纳生活水平衡量研究的数据为基础，采用的是住户和个人一级的变量。如表所示，某些变量 (如人均总支出) 的组内相关性较低，其设计效果也不好 (1.9)；但另一些变量 (如卫生设施和水的使用情况) 的组内相关性很高 (基础设施通常集中在具体的空间内)，其设计效果也较好 (分别为7.8和8.0)，城市或农村子总体群的设计效果则更好。

表A三.2也以加纳的数据为基础，该表显示设计效果可以随时间和变量的不同而不同。在本例中，两项调查的时间只相差一年，基本的样本设计没有变化，但设计效果却发生了变化：在卫生服务使用方面，估计值的精确度大大提高了 (设计效果从5.01下降到3.64)，失业的设计效果也下降了，但下降的幅度不大。在调查的第二年，表中的另一项变量——成人识字率的精确度降低了。

最后，正如所料，不同国家的设计效果差异很大。表A三.3显示科特迪瓦的调查与巴基斯坦的调查怎样使其相同的变量在设计效果方面产生了巨大差异。这种结果是由于两个国家采用的样本设计不同及这些国家的特征不同造成的。

表A三.1

按变量划分的设计效果的差异, 加纳, 1987年

变 量	估 计 值	SE	CV(%)	置信区间		n	设计效果	EFFn	ρ	
				下 限	上 限					
用电情况	总计	0.267	0.019	7.265	0.229	0.305	3 138	6.034	520	0.300
	农村	0.078	0.022	28.744	0.034	0.121	2 023	14.063	144	0.787
	城市	0.611	0.041	6.714	0.530	0.691	1 115	7.888	141	0.403
住户规模	总计	4.940	0.083	1.682	4.777	5.103	3 136	2.089	1 501	0.065
	农村	5.147	0.097	1.877	4.958	5.336	2 022	1.735	1 165	0.044
	城市	4.565	0.165	3.615	4.241	4.888	1 114	3.291	339	0.134
土地所有权	总计	0.591	0.024	4.018	0.544	0.637	3 138	7.315	429	0.376
	农村	0.747	0.033	4.393	0.683	0.811	2 023	11.520	176	0.634
	城市	0.308	0.035	11.413	0.239	0.376	1 115	6.453	173	0.319
人均总支出	总计	82 745.2	1 902.2	2.3	79 017.1	86 473.4	3 104	1.883	1 648	0.053
	农村	70 908.1	2 526.4	3.6	65 956.3	75 859.8	2 001	3.100	646	0.127
	城市	104 219.5	3 702.1	3.6	96 963.6	111 475.4	1 103	1.759	627	0.044
人均食品支出	总计	56 779.3	1 309.2	2.3	54 213.2	59 345.3	3 104	1.927	1 611	0.055
	农村	52 382.3	1 777.9	3.4	48 897.6	55 867.0	2 001	2.577	776	0.095
	城市	64 756.0	2 147.9	3.3	60 546.2	68 965.8	1 103	1.580	698	0.034
安全的垃圾处理	总计	0.019	0.003	16.647	0.013	0.026	3 135	1.724	1 818	0.043
	农村	0.010	0.003	29.044	0.004	0.016	2 020	1.704	1 185	0.042
	城市	0.037	0.009	23.481	0.020	0.054	1 115	2.347	475	0.079
安全厕所的使用情况	总计	0.590	0.025	4.159	0.542	0.638	3 135	7.808	401	0.405
	农村	0.659	0.034	5.091	0.593	0.725	2 020	10.114	200	0.549
	城市	0.465	0.038	8.092	0.392	0.539	1 115	6.357	175	0.313
安全用水情况	总计	0.395	0.025	6.251	0.347	0.443	3 135	7.994	392	0.416
	农村	0.224	0.031	13.818	0.164	0.285	2 020	11.150	181	0.611
	城市	0.704	0.046	6.482	0.615	0.793	1 115	11.144	100	0.593

资料来源: Temesgen和Morganstein (2000年)。

说明: 有关变量的说明, 见下文的表A三.4和A三.5。

总之, 由于生活水平衡量研究调查所采用的是小规模样本和多阶段抽样, 因此需要根据样本估计值的精确度进行权衡。譬如, 在1988年科特迪瓦数据中, 针对所有个人的“成人识字率”, 其设计效果高达6.7, 这一设计效果说明样本规模(n)为1 660时所获得的估计值精确度和只采用249这个简单随机抽样样本所获得的估计值精确度相等。但如果我们只考虑“城市”人口, 那么设计效果则要低一些(2.7), 但仍然高于1, 这说明样本规模为915时所获得的估计值精确度和采用344简单随机样本所获得的估计值精确度相等。由于设计效果可能会很大, 而且这种设计效果会随着变量、时间和国家的不同而发生变化, 所以在使

用数据时，尤其是对显著性进行统计检验时，分析人员亟需认识并考虑到样本设计问题。这也突出强调了在为多主题住户调查设计有效样本时有难度。努力降低一个变量的设计效果很可能会使其他变量的设计效果增加。在此，一个可以借鉴的经验是尽可能首先考虑对调查具有重要性的变量。

表A三.2

不同时间设计效果的差异, 加纳, 1987年和1988年

		加纳, 1987年								
变 量		估 计 值	SE	CV(%)	置信区间		n	设计效果	EFFn	ρ
					下 限	上 限				
成人文化水平	女性	0.402	0.021	5.103	0.362	0.442	1 339	2.342	572	0.080
	男性	0.613	0.018	2.953	0.578	0.649	1 381	1.910	723	0.054
	总计	0.509	0.016	3.192	0.477	0.541	2 720	2.875	946	0.112
保健服务机会	女性	0.443	0.016	3.625	0.411	0.474	2 756	2.876	958	0.112
	男性	0.423	0.017	4.017	0.390	0.457	2 542	3.011	844	0.120
	总计	0.433	0.015	3.517	0.403	0.463	5 298	5.013	1 057	0.239
失业	女性	0.039	0.004	10.063	0.031	0.047	4 011	1.655	2 424	0.039
	男性	0.047	0.004	9.136	0.038	0.055	3 543	1.454	2 437	0.027
	总计	0.043	0.003	7.666	0.036	0.049	7 554	1.983	3 810	0.059

		加纳, 1988年								
变 量		估 计 值	SE	CV(%)	置信区间		n	设计效果	EFFn	ρ
					下 限	上 限				
成人文化水平	女性	0.390	0.022	5.526	0.348	0.432	1 289	2.519	512	0.090
	男性	0.587	0.020	3.397	0.548	0.626	1 226	2.013	609	0.060
	总计	0.486	0.018	3.654	0.451	0.521	2 515	3.179	791	0.130
保健服务机会	女性	0.375	0.013	3.558	0.348	0.401	2 921	2.215	1 319	0.072
	男性	0.365	0.015	4.118	0.335	0.394	2 606	2.539	1 026	0.092
	总计	0.370	0.012	3.346	0.346	0.394	5 527	3.635	1 521	0.157
失业	女性	0.036	0.003	9.593	0.029	0.042	3 852	1.307	2 946	0.018
	男性	0.034	0.003	9.885	0.027	0.041	3 260	1.123	2 904	0.007
	总计	0.035	0.003	7.306	0.030	0.040	7 112	1.372	5 185	0.022

资料来源: Temesgen和Morganstein (2000年)。

说明: 有关变量的说明见下文表A三.4和A三.5。

表A三.3

不同国家设计效果的差异

科特迪瓦, 1988年										
变 量		估 计 值	SE	CV(%)	置信区间		n	设计效果	EFFn	ρ
					下 限	上 限				
成人识字率	总计	0.567	0.031	5.538	0.506	0.629	1 660	6.676	249	0.378
	农村	0.411	0.042	10.212	0.329	0.493	745	5.415	138	0.294
	城市	0.738	0.024	3.217	0.691	0.784	915	2.661	344	0.111
保健服务机会	总计	0.417	0.029	6.883	0.361	0.473	1 849	6.260	295	0.351
	农村	0.303	0.034	11.174	0.236	0.369	1 051	5.693	185	0.313
	城市	0.622	0.025	4.078	0.572	0.671	798	2.181	366	0.079
失业	总计	0.038	0.007	18.837	0.024	0.052	4 979	6.991	712	0.399
	农村	0.007	0.003	50.457	0.000	0.013	2 529	4.357	580	0.224
	城市	0.081	0.013	16.218	0.055	0.107	2 450	5.679	431	0.312

巴基斯坦, 1991年										
变 量		估计值	SE	CV(%)	置信区间		n	设计效果	EFFn	ρ
					下 限	上 限				
成人识字率	总计	0.5	0.013	2.5	0.48	0.53	6 834	4.335	1 577	0.222
	农村	0.42	0.017	3.95	0.39	0.45	3 249	3.669	885	0.178
	城市	0.68	0.018	2.616	0.64	0.71	3 585	5.156	695	0.277
保健服务机会	总计	0.5	0.012	2.329	0.48	0.52	9 238	5.02	1 840	0.268
	农村	0.46	0.015	3.177	0.43	0.49	4 752	4.048	1 174	0.203
	城市	0.61	0.017	2.74	0.57	0.64	4 486	5.185	865	0.279
失业	总计	0.03	0.003	9.735	0.02	0.03	18 232	4.633	3 935	0.242
	农村	0.02	0.003	14.955	0.02	0.03	8 934	4.706	1 898	0.247
	城市	0.03	0.003	8.956	0.03	0.04	9 298	2.539	3 662	0.103

资料来源: Temesgen和Morganstein (2000年)。

说明: 有关变量的说明, 见下文表A三.4和A三.5。

表A三.4

分析变量的说明：个人一级

变 量	说 明	人口基数
失业	目前失业但有能力工作而且正在寻找工作的成人	15-64岁之间的人
保健服务机会	在访谈前的当月生病并且前往现代保健设施，如医院、诊所和卫生所（但不是助产士、信仰疗法术士或其他传统执业医师）看病的人的比率	前一个月生病的人数
成人识字率	能够识字的成人（定义为可以读报的人）比率	15-24岁之间的人

表A三.5

分析变量的说明：住户一级

变 量	说 明
安全水的使用情况	使用安全饮用水的住户比率。在住户一级，如果住户的饮用水来源为水龙头、管道或装有水泵的水井等，那么该变量的值为“1”，如果住户的饮用水来源（如，河流、运河、敞开的水井、湖或沼泽地等）可能会对健康带来危害，那么该变量的值为“0”。
土地所有权	拥有土地的住户比率。对于一个住户来说，如果住户拥有土地，这个变量的值为“1”，否则为“0”。
用电情况	能够用上电的住户比率。对于一个住户来说，如果住户以电作为照明和/或能源，那么这一变量的值为“1”，否则为“0”。

参考资料

Blank, Lorraine和Margaret Grosh (1999年)。“运用住户调查来建设分析能力”。《世界银行研究观察家》，第14期，第2号（8月），第209-227页。

Demombynes, Gabriel和其他人（2001年）。《编制改善的贫穷地理概况：三个发展中国家的方法与证据》。《发展经济学所讨论文件，第2002/39号》。赫尔辛基：世界发展经济学研究所/联合国大学。

Elbers, C.、J. Lanjouw和P. Lanjouw（2002年）。《微观一级的福利分析》。《政策研究工作文件，第2911号》。哥伦比亚特区华盛顿：世界银行。

_____（2003年）。“关于贫穷与不平等的微观估计”。《计量经济学》，第71期，第1号，第355-364页。

Elbers, C. 和其他人（2001年）。《巴西的贫穷与不平等情况：通过PPV-PNAD数据得出的新估计值》。哥伦比亚特区华盛顿：世界银行，发展经济学研究小组。

Feres, Juan Carlos (1998年)。Falta de respuesta a las preguntas sobre el ingreso: su magnitud y efectos en las Encuestas de Hogares en América Latina. In Conference Proceedings from the 2° Taller Regional del Medición del Ingreso en las Encuestas de Hogares, Buenos Aires. November. Santiago de Chile:

Comisión Económica para América Latina y el Caribe (CEPAL). Document LC/R.1886。

加纳统计局（2000年）。《1998-1999年加纳生活水平调查，第四轮：数据使用手册》。阿克拉。

Grosh, Margaret（1991年）。《住户调查作为一项政策变更工具：牙买加生活条件调查的教训》。《生活水平衡量工作文件，第80号》。哥伦比亚特区华盛顿：世界银行。

_____（1997年）。“决策中多题目住户调查数据的使用：初级读本”。《世界银行研究观察员》，第12期，第2号，第137-160页。

_____和Paul Glewwe（1995年）。《生活水平衡量研究调查指南及其数据集》。《生活水平衡量研究工作文件，第120号》。哥伦比亚特区华盛顿：世界银行。

_____等人（2000年）。《设计发展中国家住户调查问卷：15年生活水平衡量研究调查的教训》。哥伦比亚特区华盛顿：世界银行。

Grosh, Margaret和Juan Muñoz（1996年）。《规划和实施生活水平衡量研究调查手册》。《生活水平衡量研究工作文件，第126号》。哥伦比亚特区华盛顿：世界银行。

J. Hentschel和其他人（2000年）。“合并住户数据与人口普查数据来编制分解贫穷地图：厄瓜多尔的个例研究”。《世界银行经济学评论》，第14期，第1号（1月）。

Kish, Leslie（1965年）。《抽样调查》。纽约：约翰·威利国际出版公司。

Pradhan, Menno和Martin Ravallion（2000年）。“运用消费充足的质量感来衡量贫穷程度”。《经济学与统计学评论》，第82期，第462-471页。

Ravallion, Martin和Michael Lokshin（2001年）。“运用主观问题来确定福利效应”，《经济》，第68期，第335-357页。

_____（2002年）。“俄罗斯的自测经济福利”。《欧洲经济评论》，刚印刷完毕。

Ryten, Jacob（2000年）。“MECOVI方案：未来的想法：中期评估”。为美洲开发银行编制的文件，尚未公布。12月。

Skinner, C. J.、D. Holt和T. M. F. Smith（1989年）。《复杂调查分析》。联合王国，奇切斯特：约翰·威利国际出版公司。

Temesgen, Tilahun和David Morganstein（2000年）。《抽样误差的测量：应用于生活水平衡量研究调查中的选定变量》。哥伦比亚特区华盛顿：世界银行，发展经济学研究小组。

世界银行（2000年）。《尼加拉瓜：紧急情况社会投资基金的事后影响力评估》。20400-NI号报告。哥伦比亚特区华盛顿。

_____（2001年）。《1988-1998年牙买加生活条件调查：基本信息》。哥伦比亚特区华盛顿：发展经济学研究小组。

_____ (2002年a)。《基本信息文件：波斯尼亚和黑塞哥维那生活水平衡量研究调查》。哥伦比亚特区华盛顿：发展经济学研究小组。

_____ (2002年b)。《危地马拉贫穷评估》。24221-GU号报告。哥伦比亚特区华盛顿。

_____ (2002年c)。《1993年尼加拉瓜生活水平衡量文件》。哥伦比亚特区华盛顿：发展经济学研究小组。

第二十四章

住户预算调查的调查设计和样本设计

Hans Pettersson

瑞典统计局

瑞典斯德哥尔摩

摘 要

本章论述住户预算调查中调查设计与样本设计方面的某些问题。重点为发展中国家的调查。B节就消费和收入计量方面的问题进行了一些详细讨论，C节将对某些关键的样本设计问题，如分层、空间（地理）和时间（整个季节）方面的样本分配等问题进行讨论。D节将对老挝支出与消费调查1997/1998（LECS-2）进行介绍。E节将对来自LECS-2的某些经验进行讨论。

关键词：住户预算调查、支出与消费调查、支出计量、日记法。

A. 引言

1. “住户预算调查”是许多调查大类的通称。这些调查可以称之为“家庭支出调查”、“支出与消费调查”或“收入与支出调查”。但这些调查的一个共同点是要努力掌握住户每日“预算”中的重要部分。原先打算作为住户预算调查的有些调查承担起了多目的调查的角色。有关住户消费、支出和收入的核心问题也增加了其他方面的单元，如，健康、营养和教育。将若干个主题纳入一个多目的调查中的这种方法正变得越来越常见。本章将重点讨论将住户预算计量作为重要组成部分的调查，这种调查可以是多目的调查，也可以是更加专业的预算调查。

2. 有关住户消费、支出和收入的数据有许多用途。可根据调查数据对人口的社会经济特征及其分布（如贫穷状况）进行各种研究。如果调查是定期进行的，可以用来对各人口群的福利进行监测。世界银行的生活水平衡量研究（LSMS）调查是为衡量人口贫穷状况和生活水平差异而特别设计的。在最近几年中，人们对如何利用调查来对政府干预的结果，尤其是减贫项目的效果进行评估表现出了极大的兴趣。这些数据还可用来在福利和财政领域做出决策。

3. 住户预算调查数据能够为国民经济统计系统，尤其是国民账户提供非常重要的信息。这些调查对住户部门的消费进行了衡量，还可以获得家庭企业和农业活动方面（在贫穷国家，农业生产在国民生产中占有很大一部分）的生产信息。经济统计系统所关注的重点是国民总量。如果一项调查主要为了满足经济统计系统的需求，则应该提供国家一级的各种总量的估计值。在有些情况下，如果将这种调查数据用于政策性分析和干预措施的评估，这种调查设计所起的作用不大，因为在这种情况下，所关注的是各人口群或各地区之间的差异。

4. 本章将重点讨论发展中国家中为经济统计系统提供数据的预算调查。本章共分四节。B节将讨论与调查设计有关的某些重要问题，尤其是难度较大的计量问题（特别是住户消费的计量问题）。C节将讨论住户预算调查的样本设计问题。D节将以个案研究的形式讨论老挝支出与消费调查1997/1998（LECS-2）。E节将讨论从老挝支出与消费调查中所吸取的经验与教训。

B. 调查设计

1. 住户预算调查中收集数据的方法

5. 住户预算调查的主要目标是为了对住户总消费及其构成进行计量。解决计量问题的传统方法，也是目前许多调查中依然在使用的一个方法是收集详细的信息。要求住户对大量项目的购买数量和货币金额分别进行报告。另一种方法是将所收集的消费数据限制在一个不那么详细的项目清单内。世界银行的生活水平调查通常采取这种方法（Deaton, 1997年）。

6. 消费数据基本上可以通过两种方式进行收集：

- 通过住户访谈，包括消费方面的回顾性问题。
- 采用住户日记，住户日记每日记录消费和支出。

7. 日记法通常需要对住户进行至少两次的访谈，一次在日记期开始时，一次在日记期结束后。通常情况下，还安排一次中期访谈，以确保日记报告的顺利进行。回顾性访谈可以在一次住户访问中完成，但通常进行两次访问。

2. 计量问题

8. 在使用回顾性问题进行访谈的情况下，如何对住户消费进行计量呢？是对大量项目进行详细的衡量呢，还是进行不太详细的衡量呢？第一种方法所获得的信息要比第二种方法更详细，但成本要高许多。如果我们不需要详细的情况，而主要估计消费总量，那么通过只有少量问题的第二种方法，能否获得和详细调查问卷一样准确的估计呢？对于准确性，并没有一个定论。Deaton援引了最近几年的研究，其中包括一项对印尼8 000名住户进行的试验调查（Deaton, 1997年），这次调查对两种调查问卷进行了测试。详细调查问卷包括218个食品项目和102个非食品项目，简略调查问卷中包括15个食品项目和8个非食品项目。两种调查问卷的总食品支出估计值差异不大。详细调查问卷

的非食品支出估计值大约高出15%（世界银行, 1992年, 附录4.2）。但在其他试验中没有复制这些结果。萨尔瓦多（Joliffe和Scott, 1995年）和牙买加（牙买加统计研究所和规划研究所, 1996年, 附录三）进行的类似试验显示这两种调查问卷之间的估计结果差异更大。在萨尔瓦多进行的试验中, 详细调查问卷的总支出高出40%, 食品支出高出27%。在牙买加的试验中, 详细调查问卷的总支出高出26%。Deaton得出结论说: “尽管简略调查问卷有时会大幅度降低调查成本并缩短调查时间（在印尼, 从80分钟缩短至10分钟）, 但这种节省是以精确度为代价的”（Deaton, 1997年）。

9. 日记法最大限度地减少了对受访者记忆的依赖。但如果一大部分人口为文盲, 就难以采用这种方法。即使人口的识字率很高, 日记法也存在着一些问题。譬如, 较贫穷的住户能够记日记的可能性不大, 而能够使用日记的许多住户事实上又不记日记（Deaton和Grosh, 2000年）。越南统计局发现, 在城市地区, 许多住户不愿意为1995年越南多目的住户调查记日记（Glewwe和Yansaneh, 2001年）。记日记的期限也是应该考虑的一个问题, 许多调查为期两周, 但有些调查则为期一个月。研究显示, 在为期两周的日记中, 第一周和第二周之间所报告的支出较低, 这很可能是由于疲劳所致。

10. 许多住户预算调查还收集住户收入方面的数据。住户收入的计量要比支出的计量困难更大。对许多受访者来说, 收入是个敏感的话题, 富裕家庭尤其如此。有时, 受访者中会有一种疑虑, 担心有关收入的信息将被用于征税之目的, 在一个住户拥有家庭企业的情况下, 尤其如此。

11. 所有家庭成员以及所有类型的收入都应该记录（来自家庭企业或农业的收入, 来自兼职活动、资产回报等方面的非正式收入）。赠金和实物礼品、汇款和贷款则使收入的计算更加复杂。小农住户的农业收入则带来了特殊的问题, 因为这些住户的部分食品来自生活物质的生产。另外, 有些现金收入可能来自断断续续的农产品销售, 而在访谈期间很难掌握这种收入的情况。

12. 无论是从概念上还是实际的角度上来讲, 计量住户收入所面临的这些困难可能会低估住户收入, 而且在有些情况下, 这一点已经得到证明。收入与支出调查的经验也证明了这种说法。人们经常会发现, 调查收入的估计值往往大大低于消费估计值, 其水平如此之低, 以致单凭“住户将储蓄用于消费”这一理由都难以对整个差异进行解释。另一种理由是高估了消费, 但这种解释的可能性不大。研究显示, 消费更有可能被低估而不是高估。因此, 有理由相信许多调查中的收入估计都偏低。

3. 标准期限

13. 与有关计量工具（“详细”调查问卷或“简略”调查问卷、食品消费日记法或回顾问题等）的决定密切相关的是有关标准期限的决定。要求受访者回忆的标准期限不能太长, 因为这样会使回忆的错误增加。加纳生活水平调查所进行的一项试验对延长标准期限所带来的影响进行了研究。研究显示: 回忆期每延长一天, 有关13种常买物品所报告的支出平均下降2.9%

(Scott和Amenuevegbe, 1990年)。研究人员对不同回忆期所带来的影响有不同的看法。有关印度全国抽样调查的一项早期研究似乎显示, 对于某些食品项目来说, 一个月的标准期限所产生的偏差低于一周的标准期限(Mahalanobis和Sen, 1954年)。最近几年有关生活水平调查的研究证实了Scott的结果, 但之所以出现这样一个结果, 是因为在长期数据方面的回忆结果随着时间的推移而有所误, 还是由于在短期数据方面的边缘效应(套叠效应)(Deaton, 1997年), 目前尚不明确。

14. 常买物品(如食品)的标准期限通常较短, 至多为一个月的回忆期。不常买的物品则不然, 如家用耐用品之类的物品, 有关支出的回忆期应该相对长一些, 因为期限太短会使总量的估计值产生较大的差异。因此, 不同物品组的标准期限是不同的。

4. 访谈次数

15. 大多数收入和支出调查都通过对样本住户的反复访问来收集数据。对每个住户的访问次数取决于计量的方法。回顾性方法的标准程序是进行两次访问, 大约相隔两周。在采用日记法的调查中, 建议对住户的追踪访问时间间隔为一周或两周。

16. 对同一住户进行反复访问可能会使受访者厌倦, 从而影响报告的质量。至于对住户进行长期跟踪以及为控制数据质量而对同一住户进行频繁访问所带来的好处, 应该根据这种做法可能产生的厌倦问题进行权衡。

17. 另一种反复访问的调查是: 一年中在两个或两个以上的标准期限内对住户进行访谈。这方面的一个例子是1995/1996年埃塞俄比亚住户收入、消费和支出调查, 在该调查中, 分两个不同的季节对住户进行了两次访问, 并向住户询问上一个月的情况。以下有关抽样的部分将对这种情况做进一步的讨论。

5. 不答复

18. 住户预算调查的一个显著特征是样本住户的回答负担很重。因此预算调查中拒绝回答的比率一般高于其他调查, 在某些人口群中, 这种比率可能会很高。除了从一开始就拒绝回答的情况以外, 调查中还会出现中途退出的情况。在访调员反复进行访问并对收入和支出进行详细追问的情况下, 住户会感到厌倦, 这样, 在预算调查期间中途退出的比率可能要比其他调查高。

19. 至于发展中国家预算调查的不答复情况, 目前还没有较好的类似研究。生活平衡量研究调查的不答复率在20%以下(Deaton和Grosh, 2000年), 这比西欧住户预算调查的不答复率水平低许多, 西欧住户预算调查的不答复率可高达40-50%。发展中国家的不答复率很可能存在着很大的差异。在对地方社区实施严格行政管理的国家中, 不答复率水平很可能会低一些。

C. 样本设计

20. 预算调查对样本设计的要求和其他住户调查的要求没有多大的不同。通常情况下，采用多阶段抽样，初级抽样单位（PSU）为查点区（EA）或行政单位，如公社、村庄或分区。本节将对预算调查样本设计方面的问题进行讨论。

1. 分层、层的样本分配

21. 通常根据行政地区（省等）对初级抽样单位进行分层，而在区域内则根据城市/农村分层。对于住户预算调查来说，如果根据收入水平做进一步的分层，那么将会提高效率。在城市和较大的城镇，通常可以确定两到三个收入层，并可以将初级抽样单位大致分为这些层（如高收入、中等收入和低收入地区）。

22. 住户预算调查的用户有许多，各用户对调查结果的要求不同。如果预算调查是多目的住户调查的一部分，情况更是如此。通常情况下，调查规划人员必须处理重要用户在需求方面的矛盾。住户预算数据的一个重要用途是用于国民账户（NA）。国民账户首先需要的是为国民账户提供可靠的国家总量估计。在这种情况下，就需要通过样本设计使样本在人口中平均分布（自行加权样本），或对经济活动较频繁的中、高收入住户进行某种程度的过采样。

23. 其他重要的用户是政府规划人员和政策分析人员，这些人员利用有关数据进行计划、福利监督和政策分析。对于这些用途来说，需要提供国家不同地区或不同人口群方面的可靠估计值，而不是可靠的国民估计值。调查中，所有地区和重要人口群中都应有足够数量的住户（如边远地区或贫穷村庄的住户）。在这种情况下，就需要通过样本设计使区域的样本分布大致相同，如果可能，应该在重要的人口群获得足够的样本。

24. 必须采取某种折衷措施来解决上述需求矛盾。有时，在这种情况下采取的折衷办法是平方根分配，根据这种方法，各层（区域）的样本是根据各层规模（人口或住户数量）的平方根按比例进行分配的。越南住户生活水平调查以及南非住户调查的样本使用了平方根分配法。

2. 样本规模

25. 预算调查的总样本规模在各国之间有所不同。许多调查的样本规模为3 000-10 000个住户。但在较大的国家，样本规模可能要比这大许多。地方主管部门可能强烈需要详细地理区域的调查结果，在有些情况下，这种要求甚至会影响调查数据的质量。如果样本规模很大，那么就有可能从将非抽样误差保持在容许水平的同等重要工作中“窃取”资源。具有挑战性的问题是需要平衡次国家行政机构的需求以及将样本规模和非抽样误差保持在可控水平上的预算要求。通常情况下，调查设计者必须面临的一个艰巨任务是向用户解释在抽样和非抽样误差之间保持平衡的必要性。

3. 在不同时间进行抽样

26. 较大人口群的支出和收入模式在不同季节可能存在着很大差异。调查最好包括各个季节，并获取足够的样本。应对重大假期予以特别注意，因为这期间的消费模式通常和其他期间的消费模式差异很大。

27. 处理季节性问题的一个方法是采用一年的标准期限。正如我们所看到的那样，对于大多数物品来说，这并不是一个可行的解决方案，对于食品项目来说，尤其如此。更好的办法有：

- 在一年中的不同期间（包括所有季节）对同一住户进行反复访问（采用重复标准期限）。
- 对住户进行一个期间的调查，如一个月（在该期限内可能进行若干次访问）。根据抽样计划将各住户分布在一年中的不同期间，以便充分获取所有季节的样本。这种设计假设：将每月的横截面数据加总（乘以12），就可以重新获得该年的统计数据。

28. 第二种方法很可能是解决季节性问题的最常见方法。这种方法用于支出调查，如老挝人民民主共和国、纳米比亚和莱索托。

29. 譬如，1995/1996年埃塞俄比亚住户收入、消费和支出调查就采用了第一种方法，这次调查分两个季节对住户访问两次，并向受访者询问上一个月的情况。

30. 对于第二种方法，我们需要注意季节性差异，但只注意季节性的总量差异。各种总量，如均值、住户年度总收入或支出将能得到正确的估计。一般的离度量度将会有所偏差。按个别住户的每月总量乘以12得出一年的总量中会存在季节性差异（因为只对一个月进行调查）和随机的非季节性差异（因为住户在各月的收入和支出不同，而这种不同与季节效应无关）。根据每月总量所获得的年度总量中所存在的季节性和非季节性差异会使差异增大，相对于按年度观测的总量来说，这种差异更大。如果将每月总量的离度量度作为估计值，那么年度总量的离差估计值将会有偏差。可以根据有关数据来估算季节性差异，并采用所估计的季节性差异来减少偏差。但如果偏差是由于住户在各月之间的差异造成，那么就无法减少偏差，因为有关每个住户的数据只是一个月的数据。

31. 对于希望了解不同住户年度支出分布情况的分析人员来说，为期一个月的调查设计就会带来问题，因为一般性离度量度存在着偏差（如标准差）。譬如，这些问题会影响贫穷分析，在进行贫穷分析时，将住户分为贫穷线以下的住户或贫穷线以上的住户，然后对这些群体的特征进行分析。如果不进行修正的话，那么在贫穷人口不到总人口一半的情况下，就有可能高估贫穷的程度，而在贫穷人口占总人口一半以上的情况下，则有可能低估贫穷的程度。Scott通过模型计算发现，在一项只对住户收集一个月数据的调查中，其年度支出的标准差高估了36%（Scott, 1992年）。

D. 个案研究：1997/1998年老挝支出与消费调查

32. 老挝人民民主共和国在过去十年中进行了两次支出与消费调查。第一次老挝支出与消费调查（LECS-1）于1992/1993年进行。第二次调查（LECS-2）于1997/1998年进行（国家计划委员会，老挝人民民主共和国国家统计局中心，1999年）。第三次调查（LECS-3）正在进行中。

1. 调查工作的一般环境

33. 在最近一次普查中（1995年），老挝人民民主共和国的人口为450万。从面积上来说，老挝比英国大。北部和东部为山区。国家许多地方的交通都不方便：根据1995年的普查，有57%的农村家庭所在的村庄没有公路。老挝人民民主共和国仍然以农业为主。人口的绝大多数为农村自营职业者。成人识字率大约为60%。尽管老挝人民民主共和国有许多方言，但人口的大多数都懂官方语言——老挝语。村庄为界限分明的行政单位，村庄甚至被正式细分为“住户小组”，每10-15个住户为一组。对实地工作环境进行（主观性的）大致评估就会发现，和一般发展中国家相比，在老挝人民民主共和国，更难进入农村地区的住户，但一旦进入，住户进行配合的可能性更大。

2. 调查以及调查问卷的主题

34. “增值”和“生产中的劳动投入”这两个宏观经济指标的大部分都与住户的农业生产或非正式家庭活动有关。为了获得住户生产数据，在第二次老挝支出与消费调查中引进了三个新的单元：(a) “轻松”的时间日记，该日记用以记录一个家庭成员的时间利用情况，以便对老挝经济中的劳动时间投入进行计量；(b) 有关农业和住户商业活动的两个单元。这样就可以计算农业和非正式商业活动中家庭生产的增值情况。

35. 有关住户构成、教育、就业、生育和孩子营养方面的一般单元在第一次访谈中进行。日记单元用以记录一个月内家庭的所有交易。住房、耐用品的使用、土地和牛方面的单元在第二次访谈中进行。有关住房方面的问题被用作插补租金值的基础。月末向住户询问其在前12月期间购买耐用品的情况。还向村长发放了村庄调查问卷。该调查问卷包括道路、交通、水、电、卫生设施、地方市场和学校等内容。

3. 计量方法

36. 在第一次老挝支出与消费调查中采用了日记法来计量家庭交易情况，如果这种方法有效，那么就应该在新一轮的老挝支出与消费调查中采用这种方法。改变计量方法将会使各调查之间的可比性受到影响。第一次老挝支出与消费调查中，日记法很有效，但这是因为访调员为住户提供了大量支持。如果不是访调员频繁地提供密切的支持，那么许多住户不能（或者不愿意）正确地记日记。在这些情况下，日记法似乎就不太理想。但我们必须考虑到，老挝人民民主共和国许多村庄的交通都很困难，所以一旦访调员到达村庄，就应该让访

调研员就地完成其需要为每个住户进行的三次访谈，而不是让调研员在村庄和总部之间往返数次。另外，在村庄逗留期间，调研员应该随时和住户保持经常联系。国家统计局最后决定在第二次老挝支出与消费调查中采用“调研员支持的日记法”。由调研员在村庄停留一整月，向住户提供日记记录方面的所有协助。

37. 对大米日常消费的计量采取了特别的办法。对家庭每个成员一天的大米消费进行了计量，以获得每个人每顿进餐的准确数量。向受访者出示一张印有图片的纸，纸上印有六个盘子，每个盘子大米的数量不同（一个“球”，两个“球”等），并要求受访者指出哪张图片上的数量是准确的。

38. 在这个月期间，选择24小时作为记录家庭时间利用情况的期限。第二次老挝支出与消费调查中所使用的时间利用日记由瑞典统计局和艾塞克斯大学经济与社会研究理事会（ESRC）宏观社会变化研究中心联合开发，其主要目标是为了开发一种“轻松”的日记——这种日记格式可以和其他调查工具一道使用，并且不会给受访者带来过多的负担。只要求一个年龄在10岁或10岁以上的住户成员（随机抽取）填写某一指定日期的时间利用日记。调研员随机选择受访者，一周中每天选择的受访者数量不变。

39. 时间利用日记包括22项事先确定的活动，重点为经济活动。调研员在收集日记时还对其中的某些活动收集了补充信息。如果受访者回答其工作为“雇员”，那么就会要求这些人员回答他们是否为农场工人，是在政府部门、私人部门还是在其他部门工作。如果受访者的回答是“在自营企业工作”，则要求这些人员回答其在自营企业中担任了什么角色。然后根据一张大约有50个类别的清单对有关答案进行分类，该清单上的类别依据的是所有经济活动的国际标准行业分类（经济活动行业分类）和1993年国民账户体系[欧洲共同体委员会、国际货币基金组织、经济合作与发展组织、联合国和世界银行（1993年）]。

4. 样本设计、实地工作

40. 查点区作为初级抽样单位（PSU）。初级抽样单位根据18个省进行分层，各省内又按照城市/农村分层。农村查点区进一步被分为“有公路”的查点区和“没有公路”的查点区。25个初级抽样单位样本被分配给每个省，然后又根据城市/农村进行进一步分配，城市地区被分配的样本要比农村地区多50%。每个省则根据系统概率与规模成比例（PPS）调查法抽出初级抽样单位，并抽出450个初级抽样单位样本。

41. 调查前对被抽中初级抽样单位内的住户编制了清单，并通过系统的抽样在每个初级抽样单位内抽取了20个住户，这样共有9 000个住户样本。为了在不同时间进行抽样，将省级样本按12个月的期限随机进行分配，每个月为两个村庄（有一种情况下，为三个村庄）。

42. 村庄的工作需要包括两位访调员的小组。访调员从省统计局的长期雇员中抽调。许多人员都曾参加过第一次老挝支出与消费调查。培训期限为两周。

E. 吸取的经验教训

1. 计量方法、不答复

43. 访调员花费了大量时间帮助受访者记录所有与住户有关的交易、家庭企业和农业活动。有理由相信这种单调乏味而又耗时的工作提高了受访者回答的质量。事实证明，在许多情况下，由于访调员对住户进行了频繁的访问，所以各方建立了一种轻松和信任的关系。而且这也使访调员有足够的时间理清农业或家庭企业中住户消费和生产之间的复杂关系。

44. 对质量进行了几次检查。将调查中获得的大米消费估计值与外部农业生产数据进行了核对，并发现核对结果比较一致。另外还对第一个和第二个“两周”日记期间的消费水平进行了核查。没有迹象表明第二个期间低报了数据，两个期间日记记录的数量有细微差异。另外，两个期间的总消费估计值也类似。

45. 第一个“两周”日记期间和第二个“两周日记”期间的消费总量差异很小，这就提出了一个问题：更短的日记期间是否足以获得消费方面的数据。

46. 报告的不答复率很低，只有3.1%。城市地区的不答复率非常低，只有0.6%；农村地区的不答复率要高一些，为3.9%，但这一比率仍然较低。不答复率水平低估了一些。采用了不答复替代，但由于对访谈结果进行报告的程序不理想，因此难以对正确的不答复水平进行评估，也难以对没有联系的情况和拒绝回答的情况进行区分。拒绝回答的数量很少。老挝住户调查的所有经验显示，住户认为有责任参与政府调查。另外，由村书记告诉他们参与调查。

2. 样本设计、抽样误差

47. 对方差和成本结构所进行的分析显示初级抽样单位内的最佳样本规模（查点区）为8-12个住户。这样，调查中所使用的20个住户这一样本规模大于最佳水平（Pettersson, 2001年）。

48. 计算还显示样本在各省的平均分配使国家估计值产生的抽样误差比按比率分配所获得的估计值大约高20%。国家估计值的变异系数(CV)一般低于5%。城市样本规模小于农村样本规模（分别为2 008个住户和6 874个住户）。但城市估计值的变异系数和农村估计值的变异系数类似，其部分原因是城市地区的设计效果较低。

49. 农村地区的设计效果相对较高，要比城市地区高许多（见表二十四.1）。这说明农村的村庄在社会经济方面具有同质性。由于大多数农村初级抽样单位包括一个村庄，所以初级抽样单位也具有同质性。在城市和乡镇当中，较少按收入将人口分为富人区和穷人区：富裕家庭和贫穷家庭比邻而居，

分散在城市的各个地区。因此，许多城市初级抽样单位包括非常富裕的住户和非常贫穷的住户，这样城市初级抽样单位的异质性程度就相对较高。

表二十四.1

住户消费和耐用品拥有情况的设计效果

	全 国	城 市	农 村
以老挝基普作为货币单位的每个住户的月消费总量	5.4	3.8	7.7
以老挝基普作为货币单位的每个住户的月食品消费量	5.8	4.4	6.8
住户拥有机动车的比率	2.1	1.3	3.3
住户拥有电视机的比率	5.4	3.1	6.8
住户拥有收音机的比率	4.5	2.7	4.8
住户拥有录像机的比率	5.5	3.9	6.1

50. 对每个样本住户进行了一个月的调查，样本在12个月的期间平均分布。在根据调查对贫穷率进行估计时，这就带来了问题（见第C.3节）。根据数据对季节性差异进行了估算，并利用季节性差异来消除估计值中的季节性差异。但无法估计住户在各月之间的随机非季节性差异。结果是在某种程序上过高估计了住户消费量的离差，并过高估计了贫穷率。

3. 使用时间利用日记的经验

51. 由本人完成以及访调员完成的日记数量不清楚。但有迹象显示，一般情况下，访调员会为大多数受访者提供大量的帮助，当然各地区可能存在着差异。

52. 对一个住户成员进行随机抽样的效果不佳。对填写了时间利用表的人员所计算的年龄/性别分布情况显示，在随机抽样原则的执行方面，访调员和监督员不是很成功。在许多情况下，可能出于实际的考虑，访调员似乎没有坚持采用随机抽取的人员，而是允许采用替代人员。计算结果显示，抽样中，处于活跃年龄（15-64岁）的男性比率偏高，男女少年（10-14岁）、老年（65岁或65岁以上），尤其是老年妇女的比率偏低（Johansson, 2000年）（见表二十四.2）。需要对程序进行修改，以确保时间利用数据更具备代表性。如果时间利用的调查单元主要是为了获得经济活动方面的数据，就可以不包括最年少和最年老的人员。但如果社会方案重点是为了了解童工和老年人状况，就有必要将这一类人包括进去。

4. 采用第二次老挝支出与消费调查估计国内生产总值

53. 将计量住户生产值、投入成本和时间利用的单元纳入调查获得的各种经验令人振奋。这种做法大大加强了对国内生产总值(GDP)进行估计的统计基础。如今调查为国民账户提供了以下方面的重要数据：(a) 住户生产的增值；(b) 整个经济中的劳动力投入；以及(c) 私人消费的水平 and 结构。

表二十四.2

时间利用日记的样本中实际人数和预计人数之间的比率

年 龄	实际/预计比率		
	男 性	女 性	全 部
10-14	0.41	0.49	0.45
15-64	1.33	1.04	1.18
65+	0.59	0.29	0.43
所有年龄组	1.11	0.90	1.00

54. 在对1997年的国内生产总值进行新的基本估计方面，农业和非正式经济活动中的住户生产占国内生产总值的64%，从消费角度来看，其在国内生产总值中所占的比率更大。整个经济中有大约80%的劳动力投入都来自农业和非正式经济活动中的住户生产（Johansson, 2000年）。

F. 结论

55. 本章解决与旨在计量“住户预算”的调查设计有关的问题。重点讨论了那些对住户消费总量和生产进行估计、并且其估计可以用于国民账户和一般国家经济统计的调查。有关设计问题的更全面讨论，感兴趣的读者可以参考其他出版物[如，见Deaton和Grosh（2000年）以及联合国（1989年）]。

56. 在本章的个案研究中，访调员花在每个住户上的时间多少有些不同寻常。老挝这种耗费资源的调查设计是出于统计准确性和实地工作环境方面的考虑。在识字率较低的人口中采用日记法意味着需要或多或少地每天为许多住户提供帮助。一般认为，在老挝，为准确采集住户消费方面的数据，有必要采用由访调员提供帮助的日记法。但在其他国家，采用一些费用较低的其他方法也可能得出令人满意的质量估计数据。

参考资料

- 经济合作与发展组织、国际货币基金组织、欧洲共同体委员会。联合国和世界银行（1993年）。《1993年国民账户体系》。出售品编号：E.94.XVII.4。
- Deaton, A.（1997年）。《住户调查分析》。《发展政策的微观经济方法》。马里兰，巴尔的摩以及伦敦：约翰·霍普金斯大学出版社。
- _____和M. Grosh（2000年）。“消费”。《设计发展中国家住户调查问卷：15年生活水平衡量研究的教训》，M. Grosh和P. Glewwe等。哥伦比亚特区华盛顿：世界银行。
- Glewwe, P. 和Yansaneh, I.（2001年）。《2002年至2010年多种目的住户调查的建议》。特派团向一般统计科提交的报告，越南。
- Johansson, S.（2000年）。《老挝人民民主共和国住户调查方案》。《2000年8月7日至21日万象短期特派团的报告》。斯德哥尔摩：瑞典统计局，国际咨询办公室。
- Joliffe, D. 和K. Scott（1995年）。“住户消费措施对调查设计的敏感度：萨尔瓦多经验的成效”。哥伦比亚特区华盛顿：世界银行，政策研究部。
- Mahalanobis, P. C. 和S. B. Sen（1954年）。“印度国家样本调查的一些情况”。《国际统计学会通报》，第34期。
- Pettersson, H.（2001年）。《住户调查的样本设计：特派团向国家统计局提交的报告，老挝人民民主共和国，2001年2月19日至3月2日》。斯德哥尔摩：瑞典统计局，国际咨询办公室。
- Rydenstam, K.（2000年）。“‘轻松’时间日记式记录方法：关于老挝人民民主共和国和瑞典一些行动和经验的报告”。为2001年10月23日至27日举行的联合国开展时间利用调查方法专家组会议编制的文件。
- Scott, C.（1992年）。“根据住户调查中一个月跨部门的数据进行的年度支出估计”。《Stat内部通报》，第8期，第57-65页。
- _____和B. Amenuvegbe（1990年）。《回忆时间对报告住户支出的影响：加纳的实验性研究》。《撒哈拉以南非洲调整的社会维度工作文件，第6号》。哥伦比亚特区华盛顿：世界银行。
- 老挝人民民主共和国国家统计局，国家规划委员会（1999年）。《老挝人民民主共和国住户：社会与经济指标：1997/1998年老挝支出和消费调查》。万象。
- 牙买加统计和规划研究所（1996年）。《1994年牙买加生活条件调查》。金斯敦。
- 联合国（1989年）。《国家住户调查能力方案：住户收入和支出调查：技术研究》。DP/UN/INT-88-X01/6E。技术合作发展部和统计办公室。
- 世界银行（1992年）。“印度尼西亚：公共支出、价格与贫穷人口”。印度尼西亚常驻特派团11293-IND，雅加达。Deaton中援引（1997年）。

第二十五章

转型国家住户调查

Jan Kordos

华沙经济学院
中央统计局
波兰华沙

摘 要

本章对过去十年中，即1991-2000年期间转型国家住户抽样调查（住户抽样调查）设计和实施的主要方面进行了回顾。此外，本章还将提供14个转型国家在这些调查操作方面的信息。这些国家的统计局于2001年通过填写特别问卷提供了这些信息，在某些情况下，他们随后还对提供的信息进行了更新。

本章包括两节：A节对转型国家的住户调查进行了总体评估。B节包括部分转型国家住户抽样调查的个案研究。

A节综合介绍了转型国家住户调查的主要特点。文章特别对两个主要类型的住户调查进行了介绍：住户预算调查（HBS）和劳动力调查（LFS）。并对下列调查特点进行了论述：抽样标架、样本设计、样本规模、估计方法、抽样误差估计、不答复率、调查成本和设计效果。虽然每个转型国家都需要对其住户预算调查进行重新设计，但它们已经具有住户预算调查的某些经验。而劳动力调查则是一种全新的调查形式，只是在过去10年中才引进到各转型国家的。在某些国家，是利用国外技术援助引进的。A节在结论部分还对改进转型国家住户抽样调查提出建议，并考虑到2000年人口与住房普查。

B节提供了下列国家的个案研究：爱沙尼亚、匈牙利、拉脱维亚、立陶宛、波兰和斯洛文尼亚。文章概括了每个国家住户预算调查、劳动力调查及其他住户调查的主要特点。

关键词：住户预算调查、劳动力调查、调查成本、设计效果、抽样误差、不答复率。

A. 对转型国家住户调查的综合评估

1. 引言

1. 本节的目的在于介绍过去10年中某些转型国家，尤其是中东欧国家和俄罗斯联邦住户调查设计和实施方面的某些情况。由于不同住户抽样调查（住户抽样调查）在调查项目、响应单位、周期、样本设计和收集方法方面存在很大差异，从而导致调查成本和不答复率也有很大不同。本章重点放在如下两种住户调查的设计和实施的实施上：住户预算调查（HBS）和劳动力调查（LFS）。但对转型国家过去10年进行的其他住户调查也有所介绍。

2. 在对转型国家过去10年进行的住户抽样调查进行考虑之前，本章将对这些国家在转型前的住户调查作一个概述，作为理解这些国家住户调查进一步发展情况的基础。

3. 在编写本章时，作者制作了一份特殊问卷，分发到下列14个转型国家：

白俄罗斯、保加利亚、克罗地亚、捷克共和国、爱沙尼亚、匈牙利、拉脱维亚、立陶宛、波兰、罗马尼亚、俄罗斯联邦、斯洛伐克、斯洛文尼亚和乌克兰。

4. 8个国家编写了综合文件，并在《转型期统计》[第5卷第4号（2002年6月）]公布了这些文件。

5. 本章特别介绍了这些国家住户抽样调查的设计和实施的实施，重点放在抽样标架、样本设计、样本规模、参数和抽样误差的估计方法、不答复率、调查成本和成本构成、设计效果及它们在统计分析中的应用等问题上。除此之外，本章还对2000年人口与住房普查之后如何改进调查的未来计划进行了阐述。

2. 转型前（1991-2000年）中东欧国家和苏联住户抽样调查

6. 要对转型前的中东欧国家及苏联住户抽样调查进行客观评估并非易事。众所周知，这些国家实施的是集中统计制度，完全报告或普查是数据收集的主要形式。但是，有些出版物对这些国家在此阶段的住户调查进行了报道，而且据了解还有专门的会议、研讨会和工作组会议对这些国家的调查方法进行了讨论。

7. 前共产主义国家（即中东欧国家和苏联）的住户抽样调查体系中最重要调查为住户预算调查（HBS）。除此之外，它们还定期开展过大规模的生活条件调查，以及收入调查、微型普查、卫生调查、时间利用调查和不同种类的社会与人口抽样调查。

8. 从1950年代开始，按照苏联法确定了住户预算调查，苏联法是以所谓的部门法（Postnikov, 1953年）为基础的。根据这种方法，从每个部门中抽取若干企业，再从企业的职工中抽取住户。由参加调查的被抽中住户连续几年记收支日记。样本不进行轮换，且只有那些在社会主义企业工作的职工住户才有

资格入选为样本，但离被选企业太远的住户排除在外。在每个部门中，住户按照二阶段设计进行抽选。在第一阶段的抽选中，按照概率与企业职工总人数成比例的方法从整个国家中选取一定数量的企业（或工作场所的其他单位）。第二阶段，在每个被抽选的企业中，从职工清单中系统抽取同等数量的住户，职工清单根据经济组别进行分层。每个经济组首先根据薪资或薪金的多少进行排序。在每个抽选阶段，从“抽样间隔”中间开始系统抽取抽样单位。这种抽样方法被认为能够对每个部门进行自行加权。在样本抽取之后，利用平均薪资和薪金数据，通过特殊程序对样本的代表性进行检查。在开始阶段，不答复率很低，但各国之间差别很大。

9. 在1959 - 1962年期间，各国对住户预算调查（HBS）的改进与统一给予了特别关注。为此，经济互助委员会（CMEA）统计常设委员会建立了一个特别工作组，成员来自中东欧国家和苏联。工作组在诸如概念定义、分类和问卷设计等方法领域取得了一些进展。有些国家对部门法提出了质疑，指出了同一住户连续几年接受调查的弊端。在某些国家，由于不答复率不断攀升，因此建议采用轮换抽样方法，缩短同一住户参与调查的时间。1960年代，有些国家开始尝试一种“地域方法”，这实际上是一种地区概率设计。根据这种方法，住户从查点区和按区域分层的居住地中抽取。样本住户的轮换缩短了住户参与调查的期限[Główny Urząd Statystyczny (GUS), 1971年a; Kordos 1985年, 1996年]。⁸⁶ 在某些中东欧国家（保加利亚、捷克斯洛伐克、匈牙利、波兰、罗马尼亚），住户预算调查方法开始发生变化。

⁸⁶ GUS = 波兰中央统计局。

10. 经过某些试验之后，波兰于1971年在其住户预算调查中采用了地域法，并于1982年开始采用轮换法[Główny Urząd Statystyczny (GUS), 1971年a; Kordos, 1982年, 1985年; Lednicki, 1982年]。

11. 在匈牙利，继1970年代中期制订统一住户调查制度（USHS）后，住户预算调查在1976 - 1982年期间成为一种连续调查，随后在1983 - 1991年期间两年开展一次，但从1993年起，又成为连续调查。1963年开始的收入调查每10年进行两次。在统一住户调查制度下还进行了其他许多住户调查，尤其是1980年代，如进行了时间利用调查、声望调查（各种职业的声望）、生活条件和社会分层调查等（Mihalyffy, 1994年; Éltető和Mihalyffy, 2002年）。

12. 上述国家在转型前还进行了其他一些住户调查。经济互助委员会统计常务委员会在其1968-1970年工作计划中纳入了“在经济互助委员会成员国统计调查中扩大使用更多抽样方法的可能性”这一议题。1970年4月，波兰负责组织了这方面的研讨会，并编制了主要文件（Kordos, 1970年）。九个国家（保加利亚、捷克斯洛伐克、德意志民主共和国、匈牙利、蒙古、波兰、罗马尼亚、苏联和南斯拉夫）参与了这次研讨会。每个国家提交了一份俄文文件，事后以特刊的形式用波兰文出版了这些文件[Główny Urząd Statystyczny (GUS), 1971年a]。另外还提交了方法论方面的文件，并在第二期上予以公布[Główny Urząd Statystyczny (GUS), 1971年b]。根据这些文件，可以对1970年以前这些国家所进行的住户调查类型进行一般性的评估。

13. 还召开了几次有关住户调查尤其是住户预算调查的国际会议。参与这种国际会议的波兰统计人员编制了综合报告，该报告已通过波兰统计杂志公布。作者曾参加过1961年在维也纳召开的专门讨论住户调查问题的欧洲统计研讨会（Kordos, 1963年），另外，还曾参加过1981年在日内瓦召开的专门讨论住户调查方法的第二次国际会议（Kordos, 1981年）。

14. 转型经济体经济统计国际会议：1990年代的东欧，于1991年2月14-16日在华盛顿召开，这次会议也讨论了上述国家的住户调查（Garner等, 1993年）。

15. 根据上述出版物，可以确定各种抽样方法还用于：加快人口与住房普查的数据处理（保加利亚、捷克斯洛伐克、德意志民主共和国、波兰、南斯拉夫）；微型普查（捷克斯洛伐克、匈牙利、波兰、苏联、南斯拉夫）；生活条件（保加利亚、匈牙利、波兰、罗马尼亚、苏联）；人口与住房普查之后的普查后调查（保加利亚、捷克斯洛伐克、匈牙利、蒙古、波兰、罗马尼亚、苏联、南斯拉夫）；以及时间利用调查（保加利亚、匈牙利、波兰、罗马尼亚）。这些国家的统计发展水平存在着很大的差异，而这对转型期住户调查的进展会有一些影响。

3. 转型期间住户调查

16. 本节包括了转型期间，即1991 - 2000年期间，住户抽样调查所采取的方法及实施情况。与1990年以前进行的调查相比，这一期间的调查在内容上进行了很大的扩充和修改。住户预算调查已经并将进一步得到改进，而且首次在每一个国家已经推行或即将推行一种新的调查，即劳动力调查。另外，其他一些新的样本住户调查——人口福利与健康调查、人口生活条件调查及其他人口与社会调查——也正在推出。

17. 首先，我们将讨论住户预算调查和劳动力调查。并概括介绍一下其他定期或一次性住户调查。然后，我们再特别关注一下所有住户调查都涉及的某些方法问题，如抽样标架构建、样本设计、评估方法、抽样误差、设计效果、调查成本、不答复以及住户调查的未来改进计划等。

18. 过去10年在几乎所有转型国家中，都对住户预算调查进行了重新设计，并引进了新的调查。由于在转型之前没有劳动力调查，所以这些转型国家设计并实施了新的调查。表二十五.1列出了新的住户预算调查和新劳动力调查的开始年份、周期以及重新设计的最后年份。

19. 如表二十五.1所示，按照欧洲共同体统计局（欧统局）要求（Eurostat, 1997年）重新设计和修改后的新型住户预算调查通常为连续调查。在1992-1999年期间，转型国家引进了劳动力调查。

4. 住户预算调查

20. 转型国家进行的住户预算调查是一项传统调查，有着悠久的历史。由于它们在人口生活条件分析及消费物价指数计算中所发挥的特殊作用，住户

预算调查一直倍受重视。为改善调查方法和组织形式，转型期的各个国家进行了多种调查方法试验，做出了各种尝试。在某些国家，如保加利亚、匈牙利、波兰和罗马尼亚，调查方法的改进从1970-1980年代就开始了。1990年代初，其他国家也开始改变其住户预算调查方法，按照欧统局要求（Eurostat，1997年）对其调查进行重新设计和修改。欧统局负责向成员国和其他有关国家提供帮助，通过提供准则和直接技术支持，帮助这些国家改进其调查方法和调查程序（Eurostat，1995年、1996年、1998年a、1998年b）。为此，新的概念、定义和分类被采纳，新的日记簿和调查问卷建立了起来。调查也首次为国家账户的建立提供了协助，目的是计量住户最终消费的总体水平。

21. 所有住户预算调查仅限于个体住户中的人口调查。集体或机构住户（医院、老年公寓、寄宿公寓、监狱、兵营等）不包括在内。所有14个转型国家，除捷克共和国和斯洛伐克外，都对住户预算调查进行了重新设计。

表二十五.1

某些转型国家新的住户预算调查和劳动力调查，1992 - 2000年：
开始年份、周期、重新设计的最后年份

国 家	开始年份		周 期		重新设计的最后年份	
	住户预算调查	劳动力调查	住户预算调查	劳动力调查	住户预算调查	劳动力调查
白俄罗斯	1995	-	季度	-	1995	-
保加利亚	1992	1993	连续	季度	2000	2001
克罗地亚	1998	1996	连续	两次	2000	2000
捷克共和国	1991	1993	连续	连续 ^a	1999	2000
爱沙尼亚	1995	1995	连续	连续	1999	2000
匈牙利	1976	1996	连续	连续	1997	1997
拉脱维亚	1995	1995	连续	两次	1998	1999
立陶宛	1992	1994	连续	两次	1996	1997
波兰	1982	1992	连续	连续 ^b	2000 ^c	1999
罗马尼亚	2001	1994	连续	连续 ^d	2000	2001
俄罗斯联邦	1997	1992	连续 ^e	季度	1996	1998
斯洛伐克	2003	1993	连续 ^f	连续	2002	1999
斯洛文尼亚	1999	1993	连续	连续	1997	每年 ^g
乌克兰	1999	1999	连续	季度	2000	1999

资料来源：数据来自部分国家提交的调查问卷。

说明：“季度”表示每季度进行一次，“连续”表示连续进行，“两次”表示每年进行两次，连字符“-”表示数据不适用。

^a 自2000年开始。

^b 自1999年第四季度开始。

^c 自1982年开始，重新设计过三次。

^d 自1996年开始。

^e 自1952年开始连续调查，但1996年进行了重新设计。新的调查从1997年开始实施。

^f 自1957年开始连续调查，但2002年进行了重新设计。新的调查从2003年开始实施。

^g 每年添加临时问题。

5. 劳动力调查

22. 对于转型国家来说，劳动力调查是一个新的概念，是在1992年之后才发展起来的。欧统局及各国国家统计局和劳工部代表每年定期参加几次在卢森堡举行的就业统计工作组会议（Eurostat，1998年a，1998年b），以便对这些调查的技术问题展开讨论。这样，劳动力调查得以按照国际劳工组织（ILO）的建议和欧统局（Eurostat，1998年a）规定的方法和定义实施。

23. 自1989年起，国际劳工组织统计局便开始积极帮助中东欧国家和前苏联国家从根本上修改和调整其劳动力调查体制，以满足这些国家向市场经济转型过程中出现的新要求。技术援助的形式包括一系列培训、研讨会、会议以及专家视察。

24. 关于劳动力调查，国际劳工组织专家曾前往俄罗斯联邦（1992年为准备劳动力调查试点前往俄罗斯二次，1993年5月前往俄罗斯进行全面劳动力调查）、乌克兰（1991年11月和1992年11月为准备试点劳动力调查分别前往乌克兰，1993年11月进行了一次试验调查）、保加利亚（分别为1991年12月、1992年7月和10月、1993年4月和1994年2月）、斯洛文尼亚（1993年10月）、白俄罗斯（1993年11月和1994年9月前往白俄罗斯准备试点调查和跟踪调查）、哈萨克斯坦（1993年3月和6月前往哈萨克斯坦检查开展试点劳动力调查的可行性）执行任务。此外，他们还于1991年在挪威、1991年和1992年在德国为俄罗斯和乌克兰专家举办了三次有关劳动力调查准备与实施的在职培训课程。

25. 1994年（8月31日至9月2日），国际劳工组织在明斯克召开了经济转型国家劳动力调查改革国际会议。此次会议的具体目标是要对已取得的成就和有待努力的方面进行评估，以便为经济转型国家的决策和所需的信息提供可靠和持续的劳动力市场统计资料。为本次会议所准备的所有文件全部发表在《转型期统计》杂志特刊（1995年3月第2期第1号）上。

26. 14个转型国家中劳动力调查设计与实施的某些方面值得关注。从表二十五.1中可以看出，13个转型国家已开始进行劳动力调查，白俄罗斯计划不久也开始实施该项调查。14个国家中有7个（捷克共和国、爱沙尼亚、匈牙利、波兰、罗马尼亚、斯洛伐克和斯洛文尼亚）进行了连续调查，也就是说，基准周在全年中平均分布。另有3个国家（保加利亚、俄罗斯联邦和乌克兰）每季度开展一次调查，其他国家（克罗地亚、立陶宛和拉脱维亚）每年开展两次（每半年进行一次）。爱沙尼亚在1999年之前每年开展一次调查（春季），但从2000年开始，持续每季度进行一次。所有国家都计划在不久的将来对其劳动力调查进行重新设计，以人口与住房普查结果为基础，对其抽样标架、样本设计以及评估方法进行改进。

6. 住户预算调查及劳动力调查中抽样设计及实施的共同特征

27. 住户预算调查和劳动力调查是两种截然不同的住户调查。但由于某些方法和实施特征（如抽样标架、样本设计、评估方法、抽样误差估计、设计效果、成本、不答复率以及改进调查的未来计划）为两种调查所共有，因此，将它们放在一起考虑是有助益的。

28. 在访调员的征聘与培训方面，不同国家所遵循的程序非常相似。通常情况下，这些访调员不是专为住户预算调查或劳动力调查而征聘和培训的，他们还要进行该国的其他住户调查。在所有住户预算调查中，数据采集包括两部分：(a) 住户或个人所记的日记（通常按日进行）和(b) 一次或多次访谈。

29. 对劳动力调查来说，面对面访谈是主要的数据采集方式。由“参照人员”提供住户信息，每人填写一份个人调查问卷。由代理人进行的访谈非常少见，但大多数国家认为这是一种有效的数据来源。在无法与受访人员个人取得联系的情况下，大多数国家允许“自填问卷”，即调查人员将调查问卷留给受访者自己填写。自填问卷模式优于代理访谈。由于调查问卷内容方面的原因，电话访谈尚未得到广泛推广，但已有国家开始尝试采用计算机辅助电话访谈（CATI）（爱沙尼亚）。目前绝大多数国家仍采用传统的“纸笔”方式进行调查访谈。

住户预算调查和劳动力调查的抽样标架

30. 人口与住房普查在某些国家（保加利亚、匈牙利、波兰和罗马尼亚）是构建住户调查抽样标架的基础。以查点区（CEAs）为基础，采用普查数据来创建初级抽样单位（PSU），通常情况下需要根据调查的具体要求进行调整。在大多数情况下，都将住所作为二级抽样单位（SSU）。通常情况下，被选为初级抽样单位中的住所每年更新一次。更新工作包括：由于新建筑物的落成而导致住所单位数量增加的情况，由于房屋拆除以及国家行政区域变更后地区边界发生变更而引起住所单位数量减少的情况[见Glowny Urząd Statystyczny（GUS），1999年；Kordos, 1982年, 1996年；Lednicki, 1982年；Martini, Ivanova和Novosyolva, 1996年；Mihalyffy, 1994年]。

31. 某些前苏联国家（如白俄罗斯、爱沙尼亚、拉脱维亚、立陶宛）则利用人口登记（PR）和从人口登记中得到的地址以及其他能够得到的行政文件作为抽样标架（见Lapins和Vaskis, 1996年；Martini、Ivanova和Novosyolova, 1996年；Šniukstiene、Vanagaite和Binkauskienė, 1996年；Traat、Kukk和Sostra, 2000年）。

32. 在俄罗斯联邦，1994年的微型调查被有效地用作住户预算调查和劳动力调查的抽样标架（见Goskomstat, 2000年）。

33. 一般来说，覆盖的目标人口包括一个国家领土内的所有个体住户，仅有很少例外。在某些情况下，有些小规模人口群体便被排除在外，这主要是现有抽样标架覆盖面有限造成的结果。

34. 有些国家计划将2000年度的普查结果作为住户预算调查和劳动力调查以及未来其他住户调查的抽样标架（见Éltető和Mihalyffy, 2002年；Kordos、Lednicki和Zyra, 2002年；Lapins等, 2002年；Kurvits、Söstra和Traat, 2002年）。

样本规模和分配

35. 2000年度的住户预算调查样本规模因国家而异，斯洛文尼亚为1 028户，斯洛伐克为1 300户，波兰为36 163户，俄罗斯联邦为48 675户。表二十五.2列出了2000年经济转型国家住户预算调查和劳动力调查的样本规模。

36. 一般来说，由于较大的国家对分类结果有更大需求，并且具有较强的能力，所以它们通常要求较大的样本规模。当然，这并非与国家的大小成比例。在某些国家，样本在各个地理区域按比例分配，以便在国家一级尽量获得最准确的估计值。但也有三个国家，它们是匈牙利、波兰和俄罗斯联邦，则选择不按比例进行分配，它们以更高的比率来对更小的区域进行抽样，从而来保证国家每个区域都有一个最低限度的样本规模。

37. 2000年：

(a) 住户预算调查：俄罗斯联邦的样本规模最大（48 675户），其后依次为波兰（36 163户）、乌克兰（12 534户）和匈牙利（11 862户）。样本规模最小的国家为斯洛伐克（1 300户）和斯洛文尼亚（1 028户）。

(b) 劳动力调查：俄罗斯联邦的样本规模最大（123 041），其后依次为乌克兰（38 695）、匈牙利（36 500，季度调查）、波兰（24 400，季度调查）、捷克共和国（31 800）、保加利亚（24 000）、罗马尼亚（17 600）和斯洛伐克（10 250）。其他所有国家的样本规模都在10 000以下。

样本设计和抽样

38. 在过去10年中，转型国家对住户预算调查和劳动力调查进行了不同的样本设计。抽样前，初级抽样单位分层也采用了不同的标准。最为常用的分层标准为地理区域和/或城/乡环境。也有许多国家（如匈牙利、波兰、俄罗斯联邦和乌克兰）则按地区人口规模进行分层。

39. 大部分调查都是分两个阶段进行的：第一阶段选择初级抽样单位，然后第二阶段在每个选取的初级抽样单位内选择少量的住处或住户。通常情况下，两个阶段的选择概率要进行平衡，以在区域内获得经过“自行加权”的住户样本，即初级抽样单位应按照概率与规模成比例（PPS）的方法抽取，通常与住户的数量成比例；然后在选取的初级抽样单位中选择同等数量的二级抽样单位（SSU）。在拉脱维亚和立陶宛的大城市采用了对住处、住户或人员的直接（单阶段）抽样。与此相反的是，在匈牙利对小的地方采取了三阶段抽样方法：大的地区放在第一阶段、较小聚集区归第二阶段，住址或住户放在最后阶段。

样本轮换

40. 住户的回答负担可通过定期样本轮换得以缓解。但是由于额外的样本维护、对调查人员可能进行额外的培训、首次收集基线信息的额外成本、以及对提供数据的新单位进行培养存在着困难等因素，单位轮换会使调查成本增加。按照一定的比率对抽样单位进行部分轮换则不失为一种好的折衷方法，也就是说，100%地完全置换样本单位，其成本太高且对变化的估计不准，而根本不轮换（即固定样本调查）则会导致无法接受的回答负担。根据轮换计划，在

某一特定期限内将某一单位保留在样本中，在此期限过后的最短期限内，该单位没有资格被同一调查再次选为抽样单位。

41. 在大多数转型国家的住户预算调查和劳动力调查中，还采取了其他一些样本轮换方式。譬如，爱沙尼亚、波兰和罗马尼亚采用了2-(2)-2模式，也就是说，两个季度轮换为样本，两个季度不作为样本，接着又在两个季度轮换为样本，随后退出调查。

结果的加权

42. 住户预算调查中的不答复率通常很高，这在很大程度上改变了样本住户的社会经济结构。为尽量减少这种影响，样本结果应当加权。如果有强有力的辅助信息，并通过校准方法将这些信息用于再加权之中，抽样误差和不答复误差都将大大降低。匈牙利是在两种调查中使用校准方法的惟一国家（见Éltető和Mihalyffy, 2002年；另见Deville和Särndal, 1992年）。

43. 框架中有关单位基本特点的信息有助于样本设计和抽样。更重要的是，这些信息可用于计算用以降低不答复影响的权数。为此，必须提供调查中有关答复和不答复单位特征的相关信息。

44. 首先，样本中的每个住户根据其被抽中概率的倒数进行加权。对不答复的加权包括：将样本划分到合适的加权类别中，在每个加权类别中，对受访者进行加权以便对该类别中的不答复情况进行调整。在某些情况下，可能会采用来源于外部的有关权数。此外，对于住户预算调查来说，可能也会采用来自劳动力调查中有关（住户规模及城乡关系）的权数（波兰）。

45. 波罗的海国家还采用了特殊的程序从人口登记中获取自行加权的住户样本（见Lapins和Vaskis, 1996年；Šniukstiene、Vanagaike和Binkauskienė, 1996年；Traat、Kukk和Sõstra, 2000年）。

46. 劳动力调查数据还被同时用作住户级和个人级的分析。因此，在涉及两类单位的分析中，如有必要的话，可以使用能够确保分析绝对连贯性的加权程序。在住户一级对原始样本进行全面加权，也就是说，这种程序可以确保住户中的每个人得到同样的权数。

47. 权数应按顺序得出。第一步之后，任何一步的权数都应按照已加权的样本值（根据以前所有步骤的结果进行加权）进行计算。一个单位的最终权数将根据每一步确定的加权因数来确定。每一步计算得出的权数都被标准化，也就是说，它们的比例应该能够使每个样本单位的平均值都等于1.0，权数总和等于原始样本规模。

48. 虽然所有调查中都使用一套共同的程序，但每个步骤所包含的具体变量和使用数据的来源又因调查不同而有所不同。不过，某些变量实际上在所有情况下往往都很重要，如住户地理位置、住户规模和构成、按年龄、性别及其他基本特征划分的人口分布等（Verma, 1995年）。

表二十五.2

2000年若干个转型国家中住户预算调查和劳动力调查的样本规模、样本设计和估计方法

国 别	样本规模		样本设计		估计方法	
	住户预算调查	劳动力调查	住户预算调查	劳动力调查	住户预算调查	劳动力调查
白俄罗斯	6 000	-	2阶段	-	加权	-
保加利亚	6 000	24 000	2阶段 PPS	2阶段 PPS	直接	加权
克罗地亚	2 865	12 843	2阶段 PPS	2阶段 PPS	加权	直接
捷克共和国	3 250	31 800	配额	2阶段 PPS	上次 microc.	加权
爱沙尼亚	9 840	9 127	PR PPS	PR Eq.Pr.	加权	加权
匈牙利	11 862 ^a	36 500 ^b	3阶段 PPS	3阶段 ^c PPS	校准	校准
拉脱维亚	3 847	7 940	2阶段 PPS	2阶段 PPS	加权	加权
立陶宛	10 680	6 000	PPS 人	PPS 人	加权	加权
波兰	36 163 ^d	24 400 ^e	2阶段 PPS	2阶段 PPS	加权LFS	加权人口统计
罗马尼亚	17 827	17 600	2阶段 PPS	2阶段 PPS	加权	加权
俄罗斯联邦	48 675	123 041	2阶段 PPS	2阶段 PPS	加权 microc.	加权 microc.
斯洛伐克	1 300	10 250	配额	2阶段 PPS	-	加权
斯洛文尼亚	1 028	7 000 ^f	2阶段 PPS	1阶段 人	加权	加权
乌克兰	12 534	38 695	2阶段 PPS	2阶段 PPS	加权	加权

资料来源：数据来源于部分国家提交的调查问卷。

说明：“上次microc.”系指1995年微型普查；“加权microc.”=加权微型普查；“PR”=人口登记；“加权LFS”=用于劳动力调查的权数；“加权人口统计”=人口预测-事后-分层控制数据所采用的权数；“校准”=校准方法；“Eq Pr.”=等概率。连字符(-)=数据不适用。

^a 配合调查的住户数量为10 191。为取得此结果，访调员须拜访多达17 243个地址。

^b 按季度抽取。

^c 采取二阶段抽样的自我代表城市除外。

^d 该样本规模只在2000年取得过。在之前的一年中，样本规模约为32 000个住户。

^e 按季度抽取的住户数量。每个季度抽取同等数量的住户。

^f 季度数字。

标准误差估计

49. 大多数国家的住户预算调查和劳动力调查都采用复杂的样本设计，这就需要将复杂的特征纳入抽样方差的计算中(Wolter, 1985年)。没有哪种分析方差表达式可用来估计复杂估计值的抽样误差，为此需要采用估计法。有些国家采用了随机分组法(如波兰，2000年前的住户预算调查和1999年前的劳动力调查均采用了这种方法)，有的采用刀切法(匈牙利)，有的采用泰勒级数法(波兰1999年第四季度后的劳动力调查)，还有的则采用了对称半样本法(波兰2001年后的住户预算调查)和定制分析法(俄罗斯联邦)。另有一些

国家（爱沙尼亚、拉脱维亚和斯洛文尼亚）则依赖于“相关数据统计分析软件”（SUDAAN），这是一种用于计算复杂设计标准误差的著名软件包。

住户预算调查和劳动力调查的不答复率

50. 如果看一下某些转型国家在过去4年中住户预算调查的平均不答复率情况，就能从表二十五.3所给出的数据确定以下三组：

- (a) 高不答复率组（超过40%）：爱沙尼亚（43.6%）、波兰（43.4%）、保加利亚（41.7%）和匈牙利（40.0%）；
- (b) 中度不答复率组（20%到30%之间）：俄罗斯联邦（25.6%）、乌克兰（25.0%）、拉脱维亚（24.5%）和立陶宛（22.2%）；
- (c) 低不答复率组（20%以下）：克罗地亚（19.0%）、斯洛文尼亚（18.5%）和罗马尼亚（11.0%）。

51. 从表二十五.3和二十五.4中可以看出，各国住户预算调查的不答复率远远高于劳动力调查的不答复率。此外，在某些国家中，随着时间的推移，两种调查的不答复率明显上升。就劳动力调查而言，不答复率上升的国家有：

- (a) 波兰（从1992年的4.5%上升到2000年的22.1%）；
- (b) 白俄罗斯（从1993年的10.1%上升到2000年的17.2%）；
- (c) 捷克共和国（从1993年的16%上升到2000年的24%）；
- (d) 克罗地亚（从1996年的6.3%上升到2000年的15.7%）；
- (e) 罗马尼亚（从1994年的2.6%上升到2000年的8.9%）；
- (f) 斯洛文尼亚（从1992年的9.0%上升到2000年的12.0%）。

52. 表二十五.4中的数据表明各国之间不答复率相差很大，根据不答复的程度可分为三组：

- (a) 高不答复率组（15%以上）：乌克兰（28.8%）、捷克共和国（21.5%）、保加利亚（16.1%）、克罗地亚（15.7%）和波兰（15.4%）；
- (b) 中不答复率组（10%到15%之间）：爱沙尼亚（12.5%）、斯洛文尼亚（12.2%）、匈牙利（11.2%）和拉脱维亚（10.4%）；
- (c) 低不答复率组（10%以下）：立陶宛（9.1%）、罗马尼亚（7.7%）、斯洛伐克（5.6%）和俄罗斯联邦（5.4%）。

住户调查的成本

53. 任何抽样调查都应该回答两个重要问题：

- (a) 调查总成本是多少？
- (b) 主要估计值的准确程度如何？

表二十五.3

1992 - 2000年间某些转型国家住户预算调查的不答复率

国 家	年度不答复率								
	1992	1993	1994	1995	1996	1997	1998	1999	2000
	百 分 比								
白俄罗斯	..	33.0	34.2	35.6	37.9	49.0	41.1	39.7	37.0
克罗地亚	19.0	21.0	17.0
捷克共和国	无报告								
爱沙尼亚	44.4	50.2	44.9	46.6	47.5	35.2
匈牙利	..	36.7	40.4	32.6	43.3	40.6	40.9	39.6	39.0
拉脱维亚	26.1	24.1	21.9	23.1	28.7
立陶宛	24.0	20.3	22.7	22.8	22.8
波兰	23.2	27.6	25.3	25.1	31.4	34.3	40.7	49.4	49.2
罗马尼亚	8.0	10.2	9.6	10.4	11.6	13.4
俄罗斯联邦	10.4	10.5	5.9	11.5	31.4	47.5	25.0	13.9	16.0
斯洛伐克	无报告								
斯洛文尼亚	..	24.6	22.1	28.0	34.6	19.5	18.4	17.6	18.6
乌克兰	24.2	25.7

资料来源：国家特别调查问卷。

说明：两个点 (..) 表示无数据。

表二十五.4

1992 - 2000年某些转型国家中劳动力调查的不答复率

国 家	年度不答复率								
	1992	1993	1994	1995	1996	1997	1998	1999	2000
	百 分 比								
白俄罗斯	..	10.1	8.8	8.5	11.0	14.4	16.7	16.2	17.2
克罗地亚	6.3	14.0	18.1	15.0	15.7
捷克共和国	..	16	16	18	20	19	21	22	24
爱沙尼亚	7.4	..	13.5	13.4	13.2	9.9
匈牙利	..	10.3	8.1	11.4	13.6	14.3	12.2	8.9	9.2
拉脱维亚	13.7	13.3	12.4	9.8	9.4	10.1
立陶宛	9.6	9.0	8.7	8.9
波兰	4.5	5.3	8.9	9.9	10.0	9.6	11.6	18.2	22.1
罗马尼亚	2.6	2.3	6.4	6.7	7.4	7.9	8.9
俄罗斯联邦	4.6	6.8	5.9	4.5	5.5	5.8	5.8	5.3	4.5
斯洛伐克	6.2	5.9	5.1	5.0	5.6	5.9	5.7
斯洛文尼亚	9.0	7.9	9.8	9.7	10.0	12.5	12.4	11.7	12.0
乌克兰	29.2	28.3

资料来源：国家特别调查问卷。

说明：两个点 (..) 表示无数据。

54. 要对转型国家住户调查的成本进行评估并非易事。有些国家只提供收集数据的总直接成本，包括访谈成本、差旅费、材料成本以及与数据收集有关的服务费用，但未包括其他成本部分，如调查准备、方法性插补手段、数据处理、报告编写和公布等。

55. 尽管预算的编制非常重要，但成本估计却是调查计划中最薄弱的环节。成本估计存在的问题之一是保持详细的成本记录往往是一件繁重的事情。此外，要将联合行动中的成本，尤其是将管理费和其他间接费用分开也很困难。但无论如何，制定并维持一套完整的成本报告制度将对未来的规划有益，并可以为数据方案争取必要的支持（联合国，1984年）。

56. 在波兰（Kordos、Lednicki和Zyra，2002年），2000年住户预算调查的直接成本为4 567 000欧元，其中访谈成本为3 571 394欧元（78.2%），差旅费为146 144欧元（3.2%），奖励成本为429 298欧元（9.4%）。就2000年调查而言，被调查住户的样本规模为36 163户，也就是说，每户的平均调查成本为126.3欧元。

57. 2000年的劳动力调查也进行了类似的计算。劳动力调查的总直接成本为1 094 200欧元：878 642.6欧元（80.3%）为访谈费，45 956.4欧元（4.2%）为差旅费。在劳动力调查中没有奖励成本。就2000年劳动力调查而言，有近80 000个住户接受了访谈，一次访谈费用平均为13.7欧元。我们发现住户预算调查成本约为劳动力调查成本的10倍，这主要是由于住户预算调查十分耗时，包括对同一受访者进行若干次访谈以及需要使用日记和证明资料。而劳动力调查只包括一次访谈。

58. 匈牙利提供了2000年住户预算调查和劳动力调查成本方面有用的数据（Éltető和Mihalyffy，2002年）。表二十五.5和表二十五.6列出了对住户预算调查和劳动力调查成本结构的详细分析。劳动力调查支出（432 000欧元）超过了住户预算调查支出（326 000欧元）。但考虑到在劳动力调查中，每个住户每年要接受4次访谈，而且对配合的住户不提供任何奖励措施，所以对每个住户的支出费用远低于住户预算调查的相关支出（住户预算调查中每个住户的调查费用为27.5欧元，而劳动力调查中每个住户的调查费用仅为8.4欧元）。表二十五.5和表二十五.6显示住户预算调查和劳动力调查的成本结构，包括绝对数字（欧元）和所占百分比。

设计效果

59. 从某些转型国家的住户调查描述中可以看出，几乎所有的住户样本调查都是多阶段调查。这就意味着对这些调查数据进行统计分析时需要进行设计效果计算（Kish和Frankel，1974年）。

60. 以俄罗斯联邦为例（Goskomstat，2000年，第219-220页），在俄罗斯联邦，每个地区的季度劳动力调查样本规模都是单独确定的。样本规模是根据不同水平的实际失业率确定的。对俄罗斯联邦、大区域和中等区域以及小区域而言，预期的估计准确率分别定为1.5%、5%和8%。设计效果按照第六章方

程式(7)，根据1998年就业和失业样本调查数据进行计算。计算出的设计效果在1.52到2.14之间。为住户预算调查和劳动力调查的若干个特征计算了设计效果。某些设计效果列于本章的附件。

表二十五.5

匈牙利2000年住户预算调查的成本结构

成本构成	成 本	
	欧 元	百 分 比
月度日记	148 650	45.6
年底问卷	35 865	11.0
造访不答复住户	4 345	1.3
对配合的住户的鼓励	75 855	23.3
访调员奖金	18 585	5.7
材料费	42 700	13.1
合 计	326 000	100.0

资料来源:Éltető和
Mihalyffy (2002年)。

表二十五.6

匈牙利2000年劳动力调查的成本结构

成本构成	成 本	
	欧 元	百 分 比
造访住户	22 032	5.1
住户调查问卷	65 232	15.1
活动调查问卷	212 110	49.1
补充调查问卷	42 336	9.8
调查员奖金	33 695	7.8
材料费	56 595	13.1
合 计	432 000	100.0

资料来源:Éltető和
Mihalyffy (2002年)。

7. 结论

61. 在本章中，我们介绍了样本设计和住户抽样调查实施的不同方面，并重点介绍了最重要的调查：住户预算调查和劳动力调查。从住户调查的一般回顾中，我们可以得出某些结论。转型国家的住户抽样调查根据市场经济新的要求和欧统局建议（1995年；1996年；1997年；1998年a）得到了重新设计和协调，并根据各国以前的经验和当前情况，保持了一定的差异。虽然住户调查发展取得了明显的进步，但仍有许多问题值得统计部门给予进一步关注，如标准误差的计算和提出、成本构成的估算、设计效果的计算和公布以及它们在统计分析当中的应用。此外，还有一些影响特定国家的具体问题，如低答复率和区域的样本规模不充分。对于国家间结果的比较来说，这些都是非常重要和严重

的问题。由于这些问题影响到各国住户抽样调查的整体性和一致性，解决这些问题便成为欧统局的任务。转型国家都有自己进一步发展其住户调查的计划。其中的一项计划是利用2000年人口与住房普查结果。普查数据为改进抽样标架、样本设计和估计方法，主要是小区域的改进提供了机会。

62. 以下对部分转型国家若个案例进行的研究详细说明了这些国家的住户调查在抽样设计、实施和分析方面存在的问题。个案研究附有大量参考书目，这些参考资料可用来对转型国家住户调查的不同方面进行研究。

B. 转型国家的住户抽样调查：个案研究

63. 本节提供的个案研究由下列转型国家的作者编写：爱沙尼亚、拉脱维亚、立陶宛、波兰和斯洛伐克。8个转型国家的作者撰写的更多的综合文章发表在《转型期统计》杂志第5期第4号（2002年6月）上。下文提供了有关每个国家住户预算调查、劳动力调查以及其他住户调查主要特点的信息，作为A节有关内容的补充。

1. 爱沙尼亚住户抽样调查⁸⁷

导 言

64. 爱沙尼亚统计局于1995年进行了两次重大住户调查：爱沙尼亚劳动力调查（LFS）和爱沙尼亚住户预算调查（HBS）。

65. 住户预算调查是一项连续调查；其结果按季度和年度公布。1999年，根据世界银行的方案对调查进行了重新设计，调查的日记得修改，抽样和加权程序更接近最新的可用数据。Traat、Kukk和Sõstra（2000年）对调查进行了介绍，Traat（1999年）则做了更详细的描述。

66. 劳动力调查是于1995年进行的一次性调查。1997年进行的下一次调查改变了方法。随后，劳动力调查改为每季度进行一次，直到2000年，劳动力调查又变成一种连续调查。Kurvits、Sõstra和Traat（2002年）及爱沙尼亚统计局（1999年）对该调查进行了介绍。

67. 此外，爱沙尼亚统计局还进行了许多其他住户或人口调查。这些调查与其他欧洲国家进行的类似调查一致，调查取得的信息被用于进行国家和国际比较。这些调查包括：1997年成人教育调查、1999-2000年时间利用调查、1994年和1999年生活条件调查以及2000年爱沙尼亚成人卫生行为调查（Kurvits、Sõstra和Traat，2002年）。

68. 现将爱沙尼亚住户预算和劳动力调查的情况简单介绍如下。

数据内容

69. 住户预算调查是一种日记式调查。每个被抽样的住户提供其一周的食品消费和支出信息以及一个月所有其他支出和收入方面的信息。此外还有一次

⁸⁷ 由Tartu大学数学统计学院的Imbi Traat编写（电子邮箱：imbi@ut.ee）。

有关住户成员构成和其他背景信息的事先访谈，和一次有关住户成员构成变化的简短事后访谈。

70. 劳动力调查的数据收集方案要比一般劳动力调查广泛得多，1995-1999年收集回顾性信息时尤其如此。自上次调查之后，对受访者在一定期间内的劳动力状况（就业、失业或待业情况）进行了记录。对有关每种状况的起始日期和其他相关数据也进行了记录。标准单元的劳动力调查将重点放在基准周上，向就业人员询问有关其职业、通常和实际工作时间以及企业/组织经济活动等方面的问题。向失业人员询问的问题包括求职措施、求职连续性以及想找的工作的特点等。

数据收集

71. 爱沙尼亚统计局访调员司（1994年建立）负责为各种调查采集数据。由15个县协调员负责组织全国各地130名访调员的工作。在农村地区，每个访调员在一个月内进行调查时平均每次调查进行10-15次访谈，而在城市地区则为15-20次；但实际上，他们的工作量根据区域样本规模不同也有所不同。对于大约一半的访调员来说，调查工作是他们的第二职业。他们按完成的访谈量以及同不答复人员进行联系的次数取得报酬。

72. 数据录入和编码由统计局利用名为“Blaise”的调查处理系统进行。数据录入方案包括第一次逻辑检查。在劳动力调查中，利用统计分析系统（SAS）软件进行数据处理和更复杂的检查，而在住户预算调查中，则利用FoxPro进行处理。

73. 2000年住户预算调查和劳动力调查总成本分别为153 000欧元和128 000欧元。访调员的薪资、交通和通讯费用大约占总成本的70%，而数据录入约占总成本的15%。

样本设计

74. 住户预算调查的目标人口包括所有爱沙尼亚住户，组织机构除外。劳动力调查目标人口包括15 - 74岁的爱沙尼亚居民。

75. 两种调查的抽样标架为人口登记册。抽样单位是人，根据人口登记中的记录清单对他们进行系统抽样。采用了抽样率不同的分层（住户预算调查3层，劳动力调查4层），以便覆盖更多区域。

76. 两种调查都采用了框架中的辅助信息。地址在框架中出现的频率决定了该地址的入选概率。抽样分为两部分，分别按不同的规则进行：地址样本（具有详细地址记录）和人员样本（无地址记录或记录地址不全）。无地址或地址不全的情况发生在农村地区，在那里地址除了村庄的名字外，没有任何其他信息。

77. 在地址样本中, 居住在被抽样地址的所有住户都将被纳入样本之中。在人员样本中, 只有那些有成员被抽中的住户才被纳入样本中。在一个县内对有关住户进行跟踪。通过被抽中人员, 约有15%的住户被抽样。

78. 样本设计采用概率与规模成比例(PPS)的抽样方法, 其中, 规模要么是地址在框架中出现的频率, 要么是(从住户中得到的)住户规模。对住户预算调查来说, 这是最终样本。其概率与规模成比例入选概率用以推算估计值。对劳动力调查来说, 这是第一阶段住户/地址样本, 如果样本中工龄人员的数量不能从人口登记中获得, 可以在当地政府的帮助下进行确定。第二阶段抽样的目的是要获取住户(及其成员)的等概率样本。拥有一个工作年龄成员的所有住户, 以及根据系统抽样计算, 拥有两个工作年龄成员的半数住户、拥有三个工作年龄成员的三分之一的住户等被纳入最终样本。

79. 目前的住户预算调查每月抽取820个住户样本。

80. 自2000年开始, 劳动力调查中的住户按照2-2-2轮换计划进行轮换。住户接受4次访谈: 连续访谈两个季度, 隔两个季度后, 然后在下一年度对应的两个季度中再进行访谈。根据这种轮换计划, 在任何一个季度, 有25%的住户第一次参与访谈, 50%的住户为上季度接受过访谈的住户。这样, 相邻季度中就出现50%的重叠, 而且相邻年度中的相同季度也有50%的重叠。

不 答 复

81. 在日记式调查中, 由于回答的负担加大, 不答复率往往会更高(见本章A节)。

82. 一般来说, “拒绝访谈”的情况约占不答复总数的50%, “不在家”的情况占25%。

83. 劳动力调查中的不答复率总是比住户预算调查中的不答复率低得多。而且, 由于向连续调查方式的转变使得实地调查的时间受到限制, 因而拒绝率不断升高。此外, 劳动力调查采纳对住户进行四次访谈这一做法也导致不答复率提高。

加 权

84. 住户预算调查和劳动力调查都使用了加权。在住户预算调查中, 为住户计算了权数, 而在劳动力调查中, 为人员计算了权数。

85. 在住户预算调查中, 答复率和收入/支出水平确定了6个加权组。初始抽样权数乘以组答复率的倒数。然后以已知人口统计为基础, 根据性别/年龄分布(5类)对答复权数进行校准。

86. 在劳动力调查中, 权数按步骤的顺序确定(Verma, 1995年)。一个受访者的初始权数为目标人口(15-74岁人口)规模除以对四个分层中每层进行计算得出的受访者人数。然后组成规模基本相同但答复率 R_j 不同的6个区域加

权组。在每组中，某一个人受访者权数的校正因子为 $w_j^{(0)} = \bar{R} / R_j$ ，其中 \bar{R} 为总体（平均）答复率。然后采用5次迭代加权比例法，按性别、年龄（5岁年龄组）和居住地点（15个县和首都）等人口基准对样本分布进行校准。

参数与估计量

87. 在住户预算调查和劳动力调查中，估计出的大多数参数是总数和比例。其中采用了加权的霍维茨-汤普森估计量或其比率。

88. 利用SUDAAN软件对方差进行估算。由于该软件不能准确处理劳动力调查和住户预算调查设计，因此，在SUDAAN软件中，使用了最接近的设计，即将住户作为类集，通过允许置换的不等概率整群抽样法进行分层。由于假定抽样允许置换，方差估计值要比实际方差略高。

未来发展

89. 2000年普查提供了大量有关爱沙尼亚住户和个人的信息。住户预算调查和劳动力调查加权法将按照这些可利用的普查数据加以评价，因为这些数据要比以前使用的数据更准确地反映出爱沙尼亚的人口状况。

90. 此外，爱沙尼亚还不断努力，对其他调查阶段进行改进。譬如在2002年，一种新的数据采集方法——使用10台膝上型电脑的计算机辅助电话访谈（CATI）——在劳动力调查中得以试用。在第一次访谈期间，将试用面对面访谈法，在随后的三次访谈中则采用电话访谈法。

2. 匈牙利住户预算调查和劳动力调查的设计与实施⁸⁸

住户预算调查

91. 住户预算调查在匈牙利具有很长的传统。它始于1950年代，最初以定额样本为基础。到1970年代中期改用概率设计，以此作为住户调查统一系统（USHS）的一部分。抽样标架一直包括查点区（ED），每十年普查更新一次，最近一次更新在2002年。在1976年至1982年期间，住户预算调查为连续调查；1983年至1991年期间，两年开展一次；从1993年之后，再次改为连续调查。

92. 住户预算调查样本分三个阶段抽取，但自我代表的城市例外，也就是说住所在7 000和7 000以上的城市，其抽样程序只包括两个阶段。在非自我代表地区，初级抽样单位（PSU）为整个地区；第二阶段抽样单位（SSU）为查点区；最终抽样单位为住所。在自我代表城市中，查点区即为初级抽样单位。

93. 地区按规模共分为8层，此外还可按县分层。样本一般不按比例分配到各层。小地区的抽样率要比大城市，尤其是布达佩斯的抽样率低。年度样本规模平均分配到各个月份中。

⁸⁸ 由匈牙利布达佩斯H-1525第51号信箱中央统计局的Ödön Éltető和László Mihalyffy编写。

94. 要求同意接受调查的住户提供一个月的收入和支出日记。在此期间，访调员采集该住户的其他数据，如住户家人的年龄和职业结构、住所类型、大小和设备以及耐用消费品数量等。此外，在下一年度开始时，访调员再次拜访该住户，询问住户成员有关全年中购买次数较少但价值高的支出以及某些类型的年收入情况。

95. 访调员每两年对其查点区的每个住户进行拜访，收集人口和经济数据，如住户规模、年龄、教育程度、户主经济活动等，这一事实是住户预算调查的重要方面。这些数据主要用于替代之目的：由于不答复率太高，使用替代住户（较大城市中两个，其他地区一个）是允许的。从同一分层中选取替代住户，作为初始选取的住户，还可从原始访调员的查点区中选取替代住户。

96. 每年有三分之一的住户进行轮换，轮换的结果应使每个查点区的样本规模保持不变（6个住所）。由于可以替换不答复住户，调查中给予配合的实际住户数量可能比最初的6个多或少。因此，某一特定查点区的轮换率可能高于或低于三分之一。连续三年参与调查的住户将被永远轮换出去。

97. 2000年住户预算调查样本涵盖了262个地方的1 980个查点区，最初选取的住户数量为11 862个。

98. 由于访调员在替代地点常常也会遭到拒绝或其他形式的不答复情况，因此最终访谈的数量要比计划的样本规模要小。例如，在2000年的调查中，只有10 191个住户完成了调查，而不是计划的11 862（ $1\ 977 \times 6$ ）个住户，而要取得该结果，访调员必须造访多达17 243个住户地址。1993年以后，不答复率开始上升，到1996年达到43.3%，随后开始略有下降。2000年，总不答复率为39%，其中拒绝调查的住户占近27%，其余的为空户、不在家、地址错误以及其他因素。鉴于不答复问题，要在首都和某些大城市取得所计划的样本规模尤其困难。尽管到2002年年底之前，配合调查的住户均因提供数据而得到了现金奖励，但该奖励额还不足以调动众多住户来配合调查，因此该奖励计划被取消。然而，访调员薪酬制度改革为访调员带来了实惠，这就鼓励访调员尽力说服住户配合调查。从总体上讲，拒绝率从1996年的34.4%下降到2000年26.9%。

99. 住户预算调查的样本设计保证了使用霍维茨-汤普森估计值所需要的条件。总值为加权观测总数，而设计权数为被抽选概率的倒数。在住户预算调查样本98个设计层中的每一层，设计权数都是独特的，可将其定义为人口分层中非空置住所数量与完成访谈数之比。由于单位不答复和可能存在覆盖不足，设计权数不适合计算住户预算调查数据，因此，应采用校准权数。在校准过程中，设计权数通过下列辅助变量进行调整：

- 年龄-性别组（ 2×4 类）
- 经济活动（9类）
- 教育程度（3类）
- 住户类型（3类）

100. 对于季度数据，需要对三个主要方面进行校准：首都城市、县级城市和国内其他地区。对于年度数据，校准地区细分更加详细。全国七个区域——按欧统局标准划分的二级查点区（NUTS）——也得到考虑。

101. 住户预算调查校准权数利用广义加权比例加权程序计算。

102. 至于从住户预算调查数据中得出的详细收入和支出项目，应定期计算和公布其抽样误差估计。采用的计算方法是利用由R.E.费开发的VPLX软件的分层刀切选项。在未来应用中，尤其对于估计分位数，可采用自助法。

103. 住户预算调查是中央统计局（CSO）成本最高的统计之一。2000年，调查的直接支出（不包括中央统计局中央和县办事处人员薪资）包括访调员报酬、对合作住户的奖励和材料成本共计84 769 000匈牙利福林，约合326 000欧元。

104. 2000年，净收入的设计效果大约为2，食品支出为2.5，个人总支出为2。

105. 调查结果连同对数据的简单分析每年用两种语言（匈牙利语和英语）公布。包含2001年住户预算调查数据的最新出版物于2002年出版，题为《住户预算调查：2001年匈牙利年度报告》（中央统计局，布达佩斯，2002年）。该出版物还配有光盘。

劳动力调查

106. 劳动力调查是中央统计局1992年引进的新住户调查种类。其样本是以1990年普查作为框架于1991年抽取的。自我代表城市被定义为居民在15 000人或以上的城市。一个季度的初始样本规模包括670个地方的9 960个查点区，每个查点区3个地址，这样一个季度的样本地址为 $9\ 960 \times 3 = 29\ 200$ 个。

107. 在1990年代后半期，需要更加详细可靠的劳动力调查数据，因此样本规模增加了40%。样本地区，尤其是查点区的数量也有所增加。2000年，样本包括754个地方的12 829个查点区，每季度有近36 500个住户接受访问。有关样本扩大的更详细资料见Éltető（2000年）。

108. 目前，数据采集每月进行，将包含12日的那一周作为标准期限，随后一周作为数据采集周。虽然有计划增加电话访谈数量，尤其是通过电话进行重复访谈，但劳动力调查数据采集主要还是利用纸问卷通过面对面的访谈形式进行。在样本地址，所有年龄在15-74岁的个人都有资格接受劳动力调查，都应接受访谈。

109. 根据劳动力调查所采用的轮换制度，选取的住户连续6个季度内保持在样本内，然后退出。也就是说每季度有六分之一的样本轮换退出。

110. 劳动力调查样本的设计权数按照与住户预算调查相同的方法计算。劳动力调查样本的最终权数也是利用加权比例法确定。在对劳动力调查样本权数进行校准中，19个县和首都采用了下列辅助变量：

- 年龄-性别 (2 × 10类)
- 县级市或其他地区的居所 (2类)

111. 利用VPLX软件,按照分层刀切模型计算“主表”中劳动力调查季度数据抽样误差。月度数据抽样误差也得到计算,但不公布。对于抽样误差,劳动力调查符合1998年3月9日第577/98号欧委会法规中欧统计局对准确度的要求。

112. 劳动力调查中的不答复率,尤其是拒绝率,比住户预算调查中的不答复率低得多。从开始进行劳动力调查到1997年,不答复率略有攀升,最高达14.3%。1997年之后,总的答复率开始下降,到2000年回落到9.2%。开始的时候拒绝率也有所上升,1996年和1997年曾一度达到7%,之后开始下降,到2000年下降到3.2%。

113. 劳动力调查是一项费用高昂的活动。2000年劳动力调查的直接支出为109 802 000福林,约合422 000欧元。但是,由于一个住户在一年内接受4次造访,而且对合作的住户无任何奖励,因此每户调查的费用要比住户预算调查费用低得多。

114. 主要由于劳动力调查样本比住户预算调查包含更多的初级抽样单位,所以其设计效果要低得多。2001年,总失业率设计效果为1.4,其中女性参与率的设计效果为0.8。

115. 劳动力调查通过专题单元进行补充,如工作妇女状况、与体育事假母亲相关的问题等。平均有四分之三的季度中包括这些单元。三个单元中,通常为全年第二季度单元,其中有一个单元包括欧统计局对该年度推荐的主题。无论基本的劳动力调查问卷还是欧统计局单元问卷都包括了欧统计局要求的所有信息。

116. 季度和年度劳动力调查数据将用双语公报公布。

117. 可以这么说,无论住户预算调查还是劳动力调查都是匈牙利中央统计局(CSO)非常重要的住户调查。住户预算调查数据不仅被用来计算消费物价指数权数,而且在对季度和年度国内生产总值(GDP)进行的国民账户计算中还被用来估计住户消费水平。此外,其数据对各研究领域也具有举足轻重的意义,如:不同社会阶层的生活条件、不同类型住户的支出结构及其变化、消费者对不同商品的需求等。另外必须指出的是,为了加强对欧统计局要求的遵行,2001年的支出是根据“按目的划分的个人消费分类(COICOP)体系”进行分类的(联合国,2000年,第三部分)。

118. 虽然注册失业人数的信息可从其他来源得到,但无论在概念还是在详细情况方面,劳动力调查数据都有别于其他来源的信息。劳动力调查所提供的劳动力市场的实际状况和变化信息对中央和地方政府以及研究人员都是不可或缺的。根据劳动力调查数据得出的官方失业率是最重要的经济指标之一。

⁸⁹ 由拉脱维亚银行统计处的Janis Lapins; 拉脱维亚中央统计局的Edmunds Vaskis和Zaiga Priede以及里加拉脱维亚大学的Signe Balina编写。

3. 拉脱维亚住户调查的设计与实施⁸⁹

拉脱维亚住户预算调查

119. 住户预算调查(HBS)自1995年后成为连续调查。该调查于2001年5月进行了重新设计。

120. 1995年9月,在世界银行的技术援助下,拉脱维亚引进了住户预算调查。当制定的一项要求规定结果应符合欧统局的要求时,住户预算调查已经处于筹备阶段。

调查范围

121. 住户预算调查的目标人口为拉脱维亚所有个体住户。生活在机构住户(敬老院、残疾儿童收养所、学生公寓、饭店、兵营、医院、疗养院以及刑事机构等)中的人以及无家可归人员不在当前调查之列。

抽 样

122. 样本代表了整个国家的人口和最典型的人群。每个月有342个住户接受调查。被纳入样本的每个住户只接受一次调查。

123. 调查采用了分层两阶段概率抽样法。住户按照城市化程度及地理位置进行分层。各层之间的样本分配根据各层的人口规模按比例进行。在城市地区,调查将人口登记作为抽样标架,而在农村地区,则使用住户清单。

124. 首都里加的6个行政区连同其他6个主要城市构成12个自我代表层。其他所有城镇则作为其余城市地区的初级抽样单位,分布在通过综合5个区域和2个规模组而确定的10个层中。在初级抽样中,每个分层中根据概率与居民总数成比例的方法抽取初级抽样单位。在第二阶段则按照随机抽样法,在15及15岁以上的居民中抽取。

125. 在农村地区,住户样本在5个分层或地理区域内分配。原则上讲,社区(民政区;最小的农村行政区)被用作初级抽样单位;某些小社区被加到邻近的地区。在每一层中,初级抽样单位按照概率与住户数量成比例的方法抽取、在第二阶段则采用随机抽样法抽取住户。

调查成本

126. 住户预算调查是费用最高的调查之一。在2001年的住户预算调查中,每个住户的调查费用为24拉脱维亚拉特(LVL)(约合40美元)。主要费用花在实地工作上。对访调员的补偿占总调查成本的44%,其后依次为对受访者的奖励(16%)、监督员的薪资(14%)和交通费(8%)。

抽样误差

127. 在住户预算调查中，主要查点区（首都和其他6个主要城市、其他城镇及农村地区）的某些估计值方差利用SUDAAN软件进行估算。根据这些估计值，就可以估算出农村一级的方差和设计效果。

不答复

128. 2000年，总不答复率为26.1%。不答复的主要原因是拒绝回答，包括那些在调查期间停止参与调查的住户的情况（占总不答复的46.0%），其次是“不在家（31.8%）”和“因病或过于年迈无法参与（11.6%）”。城市地区的不答复率（31.9%）远远高于农村地区（12.2%）。

129. 拒绝参与调查的住户，或对调查问题不答复的住户以及在给定地址找不到人的住户可能会对不应忽视的调查结果准确性产生影响。为了在被选定的一级保持有效的样本规模，调查采用了序贯抽样法。拒绝或不答复住户由备用清单的另一住户代替，并接受调查。

2001-2002年住户预算调查的重新设计

130. 最近对住户预算进行的重新设计是2000年春季在普查的基础上完成的。调查工具有很大变化，对耐用品、偶然采购与付款、从有偿工作取得的季节性收入、住户在农业生产中的现金收入和支出等方面统一采用了12个月的回顾性标准期限。上次住户预算调查于2000年年底结束。

131. 从2002年1月开始，对住户预算调查和劳动力调查的样本进行了协调。年度住户样本按时间（在全年52个星期内，每个星期分配同等数量参与调查的住户）平均进行分配。初级抽样单位也按季度在各地区内平均分配。

132. 新的住户预算调查和连续劳动力调查使用了相同的访调员网络。而过去的住户预算调查和劳动力调查却使用不同的访调员网络。而且，农村地区的访调员从当地人口中征聘。在新的设计中，访调员是机动的，他们可以在不同的行政区域内进行工作。这就使得样本可以更广泛地分布于农村地区（在新的住户预算调查中，年度样本分布在208个不同的农村初级抽样单位中）。同时，访调员工作量的分配也更加均衡了，交通的安排更加经济。中央统计局（CSB）重新组织的访调员结构于2002年1月启动。

拉脱维亚劳动力调查

133. 1995-2001年期间，拉脱维亚劳动力调查每半年进行一次，分别在5月和11月进行。重新设计的连续性劳动力调查从2002年1月开始实施。

134. 拉脱维亚劳动力调查是根据国际公认的国际劳工组织（ILO）劳动力调查方法制定的，这就确保了其劳动力调查与其他国家调查信息的可比性（Eurostat, 1998年a; 1998年b）。

调查范围

135. 劳动力调查人口包括15岁和15岁以上住在个体住户中的所有拉脱维亚居民。居住在机构住户，如敬老院、残疾儿童收养所、饭店、兵营、医院、疗养院以及刑事机构等的人员，以及无家可归人员不在调查之列。

136. 为遵守欧统局建议、降低调查成本，与被抽中人员住在同一家庭、符合上述年龄段的所有个人都接受调查。每次调查的全国样本规模为7 940个住户。

137. 调查中所有问题是有关访谈日之前的日历周（星期一至星期天）所发生的活动。通常情况下，通过面对面采用纸笔的方法收集数据。如果受访者不愿开门，就要求他或她接受访调员的电话访谈。

抽 样

138. 城市地区的样本从人口登记中抽取。农村地区的样本则根据完整的住户清单进行抽取。自1998年起，农村地区的样本便根据在拉脱维亚中央统计局进行的住户登记进行抽取。

139. 劳动力调查覆盖7个城市、32个城镇及所有社区。在每次调查中，约有16 000人接受调查。关于样本的建立，一般根据国家行政区域划分进行分层，通过初级抽样（城市和农村地区）或第二阶段抽样（城镇）的程序进行。在城市地区，在每个选定的初级抽样单位内对年龄在15岁及15岁以上的人群采取简单随机抽样。在农村地区，在每个社区内抽选住户的简单随机样本。

140. 根据劳动力调查的样本轮换计划，每个住户中的人员要接受三次调查。在每次调查中，样本替换率为每个城市、城镇或社区住户的三分之一。

不 答 复

141. 2000年总不答复率达到10.1%。农村地区的不答复率（只有8.5%）比城市地区（11.4%）低。农村地区的拒绝率尤其低，只约占0.5%。作为提高答复率的一种方法，调查采用了代理访谈。约有三分之一的访谈是通过代理受访者完成的。

框架不完善

142. 并非所有被抽样人员都居住在其登记的住处地址。由于对这些人员进行固定样本调查费力又耗时，而且有时实际上是不可能的事，因此，访调员不得不对实际居住在抽样地址的住户进行调查。对不参与情况进行的分析显示，在所有不参与情况中农村占2.3%，仅有2.0%的情况与框架的不完善有关（空住所、房屋拆迁、地址不存在等）。

2001-2002年劳动力调查的重新设计

143. 劳动力调查于2001年完全按照欧盟（EU）要求进行了重新设计。劳动力调查目前作为一种连续调查进行。

144. 从2002年1月开始, 劳动力调查还对样本设计进行了重大改革。劳动力调查和住户预算调查采用了相同的访调员网络。结果从2002年1月开始, 住户预算调查和劳动力调查两种调查样本得到协调。预计住户预算调查和劳动力调查这两种主要住户调查的样本的协调将促进更有效地利用调查资源。

145. 2001年12月, 对访调员进行了培训。连续劳动力调查从2002年1月份开始实施。

其他住户调查

146. 1994年和1999年, 拉脱维亚与Faf0研究所(奥斯陆应用社会科学研究所)密切合作, 在挪威政府资助的NORBALT方案框架基础上进行了生活水平调查(LCS)。

147. 拉脱维亚从1990年代后半期还相继进行了其他一些住户调查, 其中包括家庭和生育调查(1995年)、时间利用调查(1996年)、消费者信心调查(1993-1999年)、住户能源消耗调查(1996年)、国内旅游调查(1998年)、住户个人计算机使用调查(1998年)、特殊贫穷单元调查(1998年)以及对待自杀问题态度调查(1999年)等。

148. 从1996年开始, 旅行者边界调查每年进行3次或4次。对从国外回来的拉脱维亚居民和离开拉脱维亚的外国旅行者这两种旅行者的流动状况进行了调查。

149. 从原则上讲, 调查结果用拉脱维亚文和英文公布, 并提供印刷和电子文本两种形式。为了研究的目的, 中央统计局保证拉脱维亚的国内外数据用户能够访问匿名微数据文件。

一些结论意见

150. 预计专业性强、机动性高的新型访调员服务的发展将会使新的抽样调查和各种特别调查的规划和实施更加灵活。

151. 中央统计局正计划引进现代数据采集方法。在首次采取的各种措施之中, 其中一项措施是在未来几年内采用计算机辅助个人访谈(CAPI)技术。

4. 立陶宛住户抽样调查⁹⁰

导 言

152. 住户预算调查(HBS)是立陶宛进行的第一种抽样调查。1936-1937年进行了第一次调查, 为期12个月。住户预算调查是惟一被用来为立陶宛计划经济获取统计数据的定期抽样调查。在立陶宛1990年获得独立之后, 国民经济转向市场经济。为获取更多的数据, 立陶宛采用了新的调查问卷, 并有必要设计新的样本覆盖私人部门, 公布的结果也必须进行重新设计以便向用户提供能与其他国家调查结果相比较的数据。住户预算调查的重大重新设计在

⁹⁰ 由维尔纽斯LT 2600, Akademijos大街4号数学与信息学院以及维尔纽斯2746, Gedimino大街29号立陶宛统计局的Danute Krapavickaite编写。

世界银行专家的帮助下于1996年完成，具体情况见Šniukštiene, Vanagaite和Binkauskiene（1996年）。样本设计和估计方法保持不变。

153. 另一项定期住户调查为1994年开始的劳动力调查（LFS）。人口登记于1996年实现了现代化。自此开始，人口登记便被用以抽取样本，作为大多数住户调查，包括劳动力调查的抽样标架。

154. 其他住户调查多为一次性调查，调查主题包括生活条件（1997年）、时间利用（1998年）、老年人状况（1999年）、住户能源消耗（1997年）、享受医疗服务（1998年）以及向住户提供计算机（2000年）。

劳动力调查的估计值和误差

样本设计

155. 劳动力调查人口包括年龄在15岁和15岁以上的立陶宛居民。样本构造程序如下：从人口登记册上抽选约3 000人的简单随机样本，然后将其家庭成员添加到样本当中，即使这些成员未登记在人口登记册上也是如此。受访者当中妇女占52.5%。

样本轮换

156. 为避免两次调查之间的调查结果出现重大变化，每次调查中只有三分之一的样本得到轮换。每个被选取的住户参与两次调查，然后轮空一次，随后再接受一次调查，最后轮换退出调查。

估计值及其准确性

157. 根据城市/农村地区、年龄及性别进行的受访者分配与根据普查数据进行的相应分配略有差别。按12个年龄组、2个性别组和10个县（共240个加权组）对样本进行了事后分层。

158. 为了对就业和失业人员进行估计，采用了不同的加权法。为提高失业人员估计数的准确性，事后分层中使用了职业介绍所的指数。Krapavickaitė、Klimavicius和Plikusas（1997年）对固定规模样本设计的方差估计进行了介绍。

调查成本

159. 一次调查的成本约为70 000立陶宛立特。⁹¹ 问卷印刷费及将问卷分发到受访者手中的费用占总成本的14%，其余86%为访调员报酬、访调员交通费以及将填好的问卷邮寄到维尔纽斯所需的费用。与样本设计和问卷编制、样本选择、数据录入、编辑和处理等方面工作有关的费用不包括在内。

⁹¹ 汇率（2000年美元）：
1美元 = 4立特。

住户预算调查

样本设计

160. 住户预算调查采用连续调查形式。样本每年抽取一次，分为12份，平均分摊到每个月中。每个住户参与一个月的调查。立陶宛个体住户人口根据居住类型分为三层。调查从最大的城市：维尔纽斯、考纳斯、克莱佩达、希奥利艾和帕聂韦日斯的人口登记中抽出4 476名年龄在16岁及16岁以上的简单随机样本。在小城镇中，概率与规模成比例的20个类集的随机样本是从140个概率与规模成比例类集中抽取的，在第一阶段从463个类集的总体中抽取了33个概率与规模成比例类集的随机样本。从选取的每个类集中抽取一个人员简单随机样本。所有居住在被选住户中的人都要接受调查。如果同一入选地址有若干个住户，生日离抽样日期最近人员的住户将被列入样本。

估计值及其准确性

161. 住户预算调查估计中使用设计权数。估计值的设计效果大于1。这就意味着为取得更准确的估计值，将来应使用辅助信息。

调查成本

162. 住户预算调查年度总成本约为900 000立特，具体组成如下：访调员报酬占61%、税18%、支付住户14%、交通费5%及其他费用2%。

结果的发布

163. 调查结果由立陶宛统计局公布。主要结果公布在《立陶宛经济与社会发展》月刊上。所有结果均以特刊形式公布，这些特刊涉及的专题有劳动力、就业与失业（调查数据）、住户收入与支出等。

结 论

164. 根据2001年进行的人口与住房普查的初步结果，估计立陶宛的常住总人口为3 491 000。该数字比2001年1月1日公布的人口数据显示的数字少202 000。等上述普查最终结果出来后，立陶宛统计局将拥有更可靠的人口数据，可以作为改进未来住户调查的基础。到时，这些调查中的系统错误有望减少。

5. 波兰在转型期间的住户调查⁹²

导 言

165. 波兰住户调查的历史相对较长[Główny Urząd Statystyczny（波兰中央统计局，（GUS），1987年，1998年a，1999年；Kordos，1985年，1996年；Lednicki，1982年）]。1980年代，所谓的综合住户调查系统

⁹² 由华沙经济学院的Jan Kordos以及中央统计局（Al. Niepodleglosci 208号，华沙，00-925）的Bronislaw Lednicki和Malgorzata Zyra编写。

(ISHS) 逐步得到使用。该系统于1982年启用, 1992年停用 (GUS, 1987年; Kordos, 1985年)。

166. 综合住户调查系统最重要的部分为住户预算调查 (HBS), 该调查采用二阶段抽样, 在一年中对住户进行季度轮换, 并在随后的三年中对住户实行三分之一轮换。即三分之二的住户连续四年作为固定样本。子样本调查的周期也是四年。这种调查方案于1992年停止实施。同时, 在1983-1992年期间, 为住户预算调查选取的子样本被用于30多个专题社会调查中。

167. 1980年代进行的综合住户调查尝试极大地促进了住户调查朝着欧洲标准进行调整 (GUS, 1997年)。住户调查方法仍然需要进一步的一体化和改进 (Kordos, 1998年)。

转型期间的住户调查

168. 1990年之后, 对住户调查进行了较大的改进和扩展, 而且目前仍在不断进行改进。1992年, 新的劳动力调查首次被引进。此外, 其他新的住户调查也开始了, 包括生活条件调查、住户健康状况调查、时间利用调查、人口微型普查和各种普查后调查。

住户预算调查

169. 住户预算调查始于约45年前 (GUS, 1999年; Kordos, 1996年; Lednicki, 1982年)。其间尝试过各种调查方法, 并试图对其实施进行改进。1990年代初, 调查方法发生改变。在1992年引进的住户调查新方法中, 收入和支出的分类方法以及调查中社会经济的分类方法都发生了变化。首次将涵盖32 000个住户的波兰所有类型的个体住户纳入调查。1997年, 波兰政府为促进住户调查的一体化做出了积极的努力。⁹³ 2000年, 对住户预算调查进行了重新设计, 并对某些方法的内容进行了修改 (Kordos、Lednicki和Zyra, 2002年)。波兰正计划进一步改进住户预算调查, 加强该调查与其他住户调查的一体化。其中大部分工作是在欧统局建议下进行的 (Eurostat, 1997年)。

⁹³ 见1997年10月30日中央统计局局长关于设立工作组改进住户调查方法、促进住户调查一体化的第20号内部章程。

劳动力调查

170. 1992年5月波兰首次开展了人口经济活动调查, 随后每季度重复一次, 直到1999年第三季度 (Szarkowski和Witkowski, 1994年)。该调查是按照国际劳工组织的建议设计的。在每个季度中, 约有24 000个住户及这些住户中年龄在15岁及15岁以上的家庭成员接受了调查。有时, 还将选取的社会主题单元包括在该调查中, 从而大大提高了社会及经济分析的机会, 扩大了已公布结果的范围。

171. 调查结果每季度公布一次。1999年, 波兰对劳动力调查进行了重新设计, 以便使调查适应新的国家行政区域划分, 并根据欧统局要求提高其效率 (Eurostat, 1998年b; Verma, 1995年)。

1995年人口和住房微型普查

172. 波兰在上世纪最后10年中开展了若干项特别住户调查，其中最大的一项调查是1995年进行的微型普查。1995年5月，大规模的人口和住房抽样调查（微型普查）展开（Bracha, 1996年；GUS, 1998年a）。此次为第三次微型普查，前两次分别在1974年和1984年进行。需要指出的是，该普查为获取残疾人、移民以及其他社会科学方面的数据提供了机会。

173. 1995年的微型普查涵盖了人口的5%，也就是说将近600 000个住户。人口与住房普查于2002年5月和6月进行；上一次普查的时间是1998年。

生活条件调查

174. 除住户预算调查外，从1997年开始，波兰决定对人口生活条件进行一次多方面调查（Kordos、Lednicki和Zyra, 2002年）。在法国国家统计和经济研究所（统计经研所）的合作下，经过精心准备之后，调查于1997年中期首次进行了大规模抽样。随后利用固定样本子样本每年进行一次更小规模的调查，隔几年进行一次较大规模的调查。

175. 总共有12 524个住户参与了调查，其中住户答复率为87%，成人答复率为86%。在1998年中期又进行了一次更小规模的调查。

176. 1999年的样本包括两种子样本：1998年选取的子样本（固定样本）和新的子样本，新的子样本规模与1998年的固定样本子样本相同。这样，每年都有一个固定样本子样本和一个从更新抽样标架中抽取的新子样本。

177. 2001年，波兰进行了一次新的大规模生活条件调查，其样本规模约为24 000个住户，共有18 052个受访者，不答复率为25%。调查将继续进行下去，直至2005年实行新的收入和生活条件调查（EU-SILC），这项新的调查按照欧统局方案（Eurostat, 2001年）筹备。

人口健康状况调查

178. 调查于1996年4月进行，共覆盖了192 000个住户。答复率为88.6%。这是波兰首次进行规模如此之大的健康状况调查。

179. 人口健康调查是在世界卫生组织（WHO）建议的基础上进行的，因此，调查结果能够与其他欧洲国家的结果，尤其是欧盟成员国和欧洲经济委员会（欧洲经委会）区域国家的结果进行比较。

时间利用调查

180. 波兰中央统计局于1969年、1976年和1984年进行了三次时间利用调查（Kordos, 1988年b）。1996年，波兰中央统计局开展了一次小规模的时间利用调查，样本规模为1 000个住户，包括年龄在10岁或10岁以上的人员。调查目的之一除其他外是检验欧统局建议的方法是否适用（GUS, 1998年b）。一次大规模的时间利用调查将于2004年进行。

住户调查在方法上的共同之处

抽样标架

181. 普查是波兰住户调查中抽样标架的基础。初级抽样单位（PSU）利用调查查点区（ESD）或查点区（CEA）进行确定，而调查查点区或查点区则通常根据调查的具体要求进行调整。住所通常作为二级抽样单位（SSU）。调查查点区和查点区内的住所每年更新一次，更新涉及：由于新建筑物的落成而导致住所存量增加，因房屋拆除以及国家行政区域变更后地区边界变更而引起住所存量减少。在每个地区，抽样标架包括地址信息和对人口成员数量及住所数量的估计（GUS, 1998年a）。

182. 在住户预算调查和劳动力调查中，为了抽取样本，有必要将邻近的调查查点区（ESD）或查点区（CEA）进行合并，以满足每个初级抽样单位对规模的最低要求。譬如，从33 023个调查查点区中为住户预算调查构建了29 172个初级抽样单位（城市地区，初级抽样单位至少有250个住所，农村地区有150个住所）。

住户调查样本设计

183. 通常情况下，每个住户调查采用第二阶段样本抽取法，初级抽样单位（PSU）按照概率与规模成比例（PPS）的方法抽取。分层则按照地区（省）、城市/农村地区及某些情况下的地方规模进行。在连续调查中，如住户预算调查和劳动力调查中，采用了不同的轮换方式，对最终结果进行了加权以最大限度地减小不答复造成的影响。

住户预算调查的样本设计

184. 在过去的45年中，住户预算调查采用了不同的样本设计（GUS, 1999年；Kordos, 1996年；Lednicki, 1982年）。再次，我们将讨论2000年开始采用的住户预算调查的最新样本设计。至于1992-2000年期间的样本设计，其详细说明见Kordos、Lednicki和Zyra（2002年）。

185. 从2001年开始，从29 172个初级抽样单位中抽取了两个子样本，每个子样有675个初级抽样单位。初级抽样单位按16个省进行分层，每个省则按照地方规模类别进行分层。大城镇为单独的一层。每个省的层数从3个到12个不等。合在一起共有96层。各层的样本分配根据每层住所的总人口按比例进行。根据哈特利·罗的计划，按照概率与住所数量成比例的方法选取初级抽样单位。在初级抽样单位中，选取24个住所作为样本接受2年调查（每月2个住所，两年中对同样住所进行调查）。此外，在每个初级抽样单位中，单独选取150个住所作为备用子样本，以便在不答复的情况下使用。每年将选取一个包括675个初级抽样单位的新子样本，并对其进行2年的调查。

住户预算调查的加权

186. 住户预算调查中的不答复率通常很高，它们对样本中住户的社会经济结构带来很大的影响。为最大限度地减小这种影响，应对样本结果应进行加权。

187. 首先，样本中的每个住户按照它被抽样时所采用的概率的反比进行加权。其中还要采用来源于外部的权数。在住户预算调查中，另外采用了从劳动力调查中获得的相关权数（住户规模及城乡人口比例）。

标准误差的估计方法

188. 2000年前，标准误差的估计一直采用随机分组法。2001年后，一直采用对称半样本法。

劳动力调查的样本设计及其1999年的重新设计

189. 劳动力调查样本分两个阶段分层选取。初级抽样单位为城镇中的查点区（CEA）；在农村地区，初级抽样单位为调查查点区（ESD）（在某些情况下，抽样单位通过合并两个或更多的相邻查点区或调查地区取得，以便达到预先规定的最低规模要求）。住所作为二级抽样单位（Szarkowski和Witkowski, 1994年）。

1999年对调查的重新设计

190. 从1999年第四季度开始，劳动力调查便作为一项连续性调查进行。初级抽样单位和二级抽样单位的抽取方法与以前调查的抽取方法相同，但样本按16个省进行分配的做法发生了改变。为使各省的估计值更加准确，每个省的样本规模根据该省住所数的平方根按比例进行分配。各省中的分层规模则根据地方规模按比例进行。

191. 各层中的初级抽样单位按照概率与初级抽样单位中住所数量成比例的方法抽取。然后，从每个初级抽样单位中选取特定数量的住所（4到9个）。一个季度中每13个星期，⁹⁴ 访调员对随机抽取的一定数量的住所（1 880-1 900个）进行访问，收集访问前一周与经济活动有关的数据。调查覆盖被抽中住所中所有15岁及15岁以上的人员。访问的住所样本每周进行更换。季度样本按照随机抽样分为13个周样本。季度样本规模从24 440个到24 700个住所不等（GUS, 2000年）。

⁹⁴ 根据欧统区的规定，劳动力调查有使用的“季度”一词与日历季度略有不同：劳动力调查中的每个季度有13个周，起始时间始终是星期一。因此，2000年第一季度的期限为1月3日至4月3日。

192. 采用的住户轮换方式如下：二个季度接受调查，轮空二个季度，然后再接受二个季度调查，最后退出调查[2 - (2) - 2轮换模式]。

劳动力调查结果的加权

193. 加权分三个阶段进行（详细介绍请见Kordos、Lednicki和Zyra, 2002年）。

标准误差的估计

194. 1999年之前，估计值的标准误差一直按照随机分组法进行计算。从1999年第四季度对劳动力调查进行重新设计之后，一直采用泰勒线性化方法。

住户调查成本

195. 波兰中央统计局拥有一套住户调查成本估计系统。在每次抽样调查中，根据之前在该领域的经验和某些行政建议对调查的直接成本进行估计。该调查成本估计包括实地访谈成本、差旅费、材料费、与调查有关的服务费、为增加调查参与率而设置的奖励、税金等（GUS，2001年）。但不包括编码与编辑、计算机运行以及与方法有关的间接成本和同时负责若干个项目的人员的成本。

196. 作为例子，本章A节对2000年波兰住户预算调查和劳动力调查的成本要素进行了介绍。

设计效果

197. 对波兰住户调查，也就是住户预算调查和劳动力调查某些特征的设计效果进行了计算（Kordos、Lednicki和Zyra，2002年）。作为一项练习，同时也为了与其他国家进行比较，对2000年和2001年的若干参数的设计效果进行了计算。

198. 住户预算调查某些特点的设计效果与相关标准误差（括号中为百分比）如下：总收入：4.24（1.1）；总支出：4.16（1.0）；食品支出：3.53（0.4）；服装和鞋：2.72（1.5）；住房维护：4.04（1.3）；人员医疗：3.28（1.7）；交通和通讯：2.16（4.5）；教育：2.50（3.9）。

199. 劳动力调查则根据城乡地区、地方规模（城镇类别）和教育水平，对2000年和2001年交叉类别的不同人群中失业总人数的设计效果进行了计算。城镇类别的离散度最高，设计效果在1.7到3.55之间。

200. 从上述估计值可以看出，住户预算调查和劳动力调查的设计效果通常都大于1，某些特征的设计效果甚至大于4。因此，根据简单随机样本假设得出的标准误差往往低估了应用复杂样本设计推导出的标准误差。

住户调查中的不答复

201. 如A节所述，在过去十年间，无论是住户预算调查还是劳动力调查，不答复率都有所上升。不答复率上升的主要原因为拒绝和“不在家”。在住户预算调查中，拒绝率从1992年的10.2%上升到2000年的25.0%，不在家率从1992年的4.5%上升到2000年的14.5%。

202. 在1992 - 2000年期间的大部分时间内，劳动力调查的不答复率逐年稳步上升，从1992年的4.5%上升到2000年的22.1%，其中拒绝率从1992年的2.0%上升到2000年的10.9%。1992 - 2000年，不答复率大幅度上升，主要原因是拒绝回答和不在家的比率上升。

203. 劳动力调查中的不答复率因地方规模的不同而不同，最高的地方为华沙，最低的地方为农村地区。2000年，根据地方规模加权的年度不答复率如下：华沙：54.5%；城市（居民在50万到100万之间）：32.6%；城市（居民在10万到50万之间）：33.3%；城镇（居民在20 000到10万之间）：23.1%；城镇（居民在20 000以下）：19.0%；农村地区：11.1%。

结 论

204. 本节对波兰中央统计局（GUS）在转型期间进行的住户调查进行了介绍，并重点介绍了两种连续调查，即住户预算调查和劳动力调查。波兰中央统计局具有很长的住户调查传统，并在这一领域具有丰富的经验。这有助于在转型初期对调查进行重新设计和设计新的调查。

205. 吸收和利用2002年人口与住房普查的结果是未来几年中住户调查的最重要任务。普查结果将不仅为住户调查提供更新的抽样标架，而且为提高估计精确度的工作和目前正在研究的小地区估计方法提供了辅助信息。

206. 我们已经开始筹备新的住户调查——EU-SILC，将于2005年推行（Eurostat, 2001年）——并且将对当前的调查进行改进，以使它们符合欧盟标准。

6. 斯洛文尼亚的劳动力调查和住户预算调查⁹⁵

导 言

207. 斯洛文尼亚共和国于1990年代初独立。在此之前，斯洛文尼亚属于前南斯拉夫，统计活动由联邦统计局集中进行。但是，住户调查在国家统计方案中并不占重要位置。独立之后，斯洛文尼亚统计局迅速从一个地方统计局变为一个国家统计机构。过渡过程相对比较顺利，部分原因是因为高层管理人员保持未变，在整个过渡期间一直在位。

劳动力调查（LFS）

背 景

208. 首次劳动力调查于1989年由卢布尔雅纳大学社会科学系进行（Vehovar, 1997年）。1995年斯洛文尼亚共和国统计局接管了劳动力调查的所有职责。

209. 1989-1995年期间的劳动力调查样本以一种相当特别的方式进行设计和实施，这主要是因为年度预算不确定。从1992年开始，采用了三阶段类集样本设计，每年抽取3 000个新住户作为样本。被选单位连续三年作为样本，总样本规模约为8 500个单位。

⁹⁵ 由卢布尔雅纳大学社会科学系的Vasja Vehovar以及斯洛文尼亚共和国统计局的Metka Zaletel、Tatjana Novak、Marta Arnež和Katja Rutar编写。

重新设计

210. 1997年，由于对（区域）结果的要求更频繁和更详细，进行了一项重大的重新设计。欧统局准则对促进该项重新设计也很重要（Eurostat, 1998年a）。

211. 劳动力调查经修改后成为一种连续性固定样本调查，每季度进行样本抽选并公布结果。每季度样本分为6个两周区间。访谈标准期限为访谈前的一周（星期一至星期天）。调查采用3-1-2轮换方式，即被选住户连续接受三个季度的调查，然后轮空一个季度，接着再接受两个季度的调查。这种模式可以保证两个连续季度有60%的样本重复，连续两年有40%的样本重复。

212. 劳动力调查抽样标架为中央人口登记和分层信息。层定义的基础是6个类型的定居点（根据定居点大小和农民人口的比例）和12个地理区域。层折迭后，共有47个层。

213. 在每层中，开始按照随机的方法，然后利用系统抽样抽取样本。根据定居点、街道和建筑物牌号对数据进行分类，以此进行隐性分层。对每层中的抽样率进行调整，以便将预计的不答复率情况考虑进去。抽样中没有采用实地替换，因为证据显示实地替换弊大于利（Vehovar, 1999年）。

214. 每个季度选取2 000个新单位。此外，以前四个季度中的约5 000个（有反应的）住户被纳入样本。这样，每个季度约有7 000个住户样本（2 000个为新抽取样本，5 000个为连续样本）。预计其中约有6 000个住户为有反应的住户。完成个人访谈的总人数约为20 000人。

实 施

215. 所有新抽取的季度样本住户都会接受（面对面的）计算机辅助个人访谈（CAPI）。全国约有30名经验丰富的劳动力调查访调员，他们全部配备了便携式计算机。重复访谈则利用CATI在统计局的电话中心进行，无电话住户和无法参与电话访谈的住户除外。全国电话普及率约为95%。在访谈之前，每个住户会收到一份带有调查说明的预先通知和一份包括以前劳动力调查结果的小册子。电话访谈不提供奖励。

216. 对新选样本进行面对面访谈，其不答复率在17-18%之间，拒绝率在12-13%之间。在对已经在固定样本中的住户反复进行的电话访谈中，不答复率略低（10-11%），拒绝率也较低（6-7%）。从1991年开始，劳动力调查的不答复率大幅度上升，但在最后四年中稳定下来。

抽样误差及公布

217. 在不等抽样概率和单位不答复的情况下，对有关数据进行了加权。根据已知人口的年龄（8组）、性别和地区（12个地区）分布情况进行事后分层。由于事后分层是在个人层面上进行的，因此同一住户的成员可能会得到不同的权数。

218. 按照常规做法，只对重要变量的抽样误差和设计效果进行估计：失业率和就业/人口比率。设计效果相对较低，譬如，失业率的设计效果为1.3。

219. 根据常规做法，对估计值的变异系数（CV）进行了计算。变异系数的估计值小于10%，其公布不受任何限制；变异系数在10-20%之间的估计值，其公布要加单括号；变异系数在20-30%之间的，其公布要加双括号。变异系数在30%以上的，其结果以点（.）代替，表示“非零，但不可靠”。

220. 调查结果每季度在《统计快报》、《统计年鉴》和其他几家斯洛文尼亚出版物上公布。《调查结果》特别从刊则提供详细的调查结果和调查方法。调查数据还会在其他组织的出版物上公布，如世界银行、联合国儿童基金会（儿童基金会）和欧统局出版物。统计局以外的研究人员也可以对微数据进行分析。

住户预算调查（HBS）

背 景

221. 第一次住户消费调查是在1960年代进行的。在1997年之前，调查一直按照南斯拉夫联邦统计局设计的较先进和具有创新性的方法进行。样本设计包含了二阶段类集样本，在第一阶段对初级抽样单位进行分层。初级抽样单位为查点区（EA），按概率与规模成比例（PPS）的方式进行抽样。在第二阶段，从中央人口登记中选取人员；这些人员还决定被抽中的住户。在每个初级抽样单位中，对5个住户进行访谈。1993年之前，调查采取替代程序来获得5个应答单位；但从1994年开始，每个类集中的“抽取率”从每个初级抽样单位的6人增加到8人，这种设计要求通过权数进行额外修正。定期开展了两种不同的住户预算调查：一种为季度调查，另一种为5年一次的年度调查。最后一次年度住户预算调查覆盖了3 270个住户，最后一次季度调查覆盖了1 000个住户。在年度调查中，访谈是在全年的年末进行的。而在季度调查中，每年对被抽中的住户进行四次访谈。

住户预算调查样本的重新设计

222. 失去样本重新设计的主要因素是欧统局的新准则（Eurostat，1997年）。

223. 人口登记是抽取个人受访者的基础。这些人还决定着被抽中的住户。通过权数对人员和住户的不等抽选概率进行调整。机构住户不包括在内。年度样本规模包括1 200个受访住户。由于这一样本规模太小，所以无法采用“北欧”模式，为此，将连续三年的样本数据合并起来，重新计算后列入中间一个年度。这样，可以确保3 600个住户样本规模。

224. 按比例将样本分配到47个分层。由于样本规模较小，层数较多，因此只采取隐性分层。在小定居点（居民不足1 000人），查点区作为初级抽样单位，并按概率与规模成比例（PPS）方式选取。在每个初级抽样单位中选取

4个受访住户。在较大的城镇和城市，则采用简单随机抽样（SRS）法。其结果是设计效果相对较低，重要变量的设计效果约为1.2。每个季度的样本单位单独选取，并分配到相应季度中的12个星期中。第13周用来处理不答复者方面的工作。

实 施

225. 在对住户进行访谈之前，提前一周向住户寄送预先通知书和一份奖品：一个袖珍计算器。由于这是一种连续调查，调查可由较少的访调员负责实施（如，20名）。

226. 访调员在一张特殊表格上对一个住户中所有联系人/试图联系的人进行登记。这样每个单位的住所、住户和参照人员等情况以及尝试联系的次数、已填写日记的数量和不答复的潜在原因等都一目了然。

227. 数据通过访调员填写调查问卷和住户成员记日记方式进行收集。几乎所有访谈都通过计算机辅助个人访谈（CAPI）系统进行。

228. 住户记14天的日记。在此期间，他们定期填写各自的每天支出情况。如果住户至少能填写基本的调查问卷，就将其视为做出了回答，因为三分之二的数据来自调查问卷。通过调查问卷能够取得较高和较稳定的答复率（约81%）。但是，完整答复率，包括日记的完整率则较低，约为70%。

抽样误差及公布

229. 如果某一特定住户单元的所有日记都丢失，可根据类似住户日记利用热卡插补法对数据进行插补。缺失方案的不答复也通过热卡插补程序进行插补。每个缺失值由同一插补类别中以前受访者的相应数据替代，插补类别根据住户规模和社会人口特征确定。尤其是，缺失的个人收入数据由那些在就业状况和教育程度方面与其接近的个人收入替代。

230. 设计权数和事后分层权数的计算方法与劳动力调查的方法相似。此外，还确定了具体的膨胀系数，以便对不同标准期限进行补偿。重新计算的系数基本上为调查标准期限（一年）与个体变量标准期限之比。在结合使用连续三年的数据时，也需要采用特殊的权数。因此，某一具体日期的计算用到三年的数据时，其中一半为该日期之前的数据，一半为该日期之后的数据。

231. 住户预算调查方法和结果在上文为劳动力调查援引的出版物中介绍。

结 论

232. 1991年独立之前，住户抽样调查不是斯洛文尼亚收集数据的一种通用手段。但与其他转型国家不同的是，斯洛文尼亚自1960年代开始便定期进行住户预算调查，从1980年代末开展了一系列年度劳动力调查。

233. 独立之后，斯洛文尼亚统计局顺利有效地进行了过渡。目前，国家统计局定期进行标准系列的住户调查。基本的社会经济调查（劳动力调查和住户预算调查）十分接近欧统局的要求（斯洛文尼亚共和国统计局，2001年）。除此之外，斯洛文尼亚还进行了其他许多住户调查：1997年的住户能源消耗调查（HECS）、2000/2001年的时间利用调查（TUS）、月度消费者态度调查（CAS）、季度国内人口旅行调查（QSTDP）和年度犯罪与受害情况调查（2002年、2001年）。

234. 住户抽样调查系统仍有待于进一步改进。斯洛文尼亚具有一套建立在丰富和准确登记基础上的统计系统（税收数据、雇员数据库、保险数据库等），这套系统可快速有效地与地理系统和普查数据联系起来，因而能为最佳样本的设计和数据的估计提供更多优势。

鸣 谢

以下国家的中央统计局局长同意提交有关其国家住户抽样调查的资料：白俄罗斯、保加利亚、克罗地亚、捷克共和国、爱沙尼亚、匈牙利、拉脱维亚、立陶宛、波兰、罗马尼亚、俄罗斯联邦、斯洛伐克、斯洛文尼亚和乌克兰。在此，作者对这些国家的许多同行提供的数据和有益意见表示感谢，并对那些为B节提供资料的人员表示感谢，这些材料为A节的内容提供了补充。

作者还对一些匿名审阅人以及2002年10月8-11日在纽约举行的“发展中国家和转型国家调查实施特点分析”专家组会议的成员表示感谢，感谢他们提供了有益的意见和建议。

参考资料

- Bracha, Cz. (1996年)。Schemat losowania próby do Mikrospisu 1995 (“1995年小型人口普查的样本设计”)。Wiadomości Statystyczne, 第3号, 第12-18页。
- Deville, J.-C. 和C. E. Särndal (1992年)。“抽样调查中的校准估值”。《美国统计学会志》，第87期, 第376-382页。
- Éltető, Ö. (2000年)。“扩大匈牙利劳动力调查以获得劳动力市场指标的可靠小区域估计值”。《转型期统计学》，第4期, 第4号, 第549-563页。
- _____和L. Mihalyffy (2002年)。“匈牙利的住户调查”。《转型期统计学》，第5期, 第4号, 第521-540页。
- 欧洲联盟 (1998年)。“1998年3月9日理事会关于在共同体内组织一次劳动力样本调查的第577/98号条例”。《欧洲共同体公报》，14/3/98, I. 77/3-I. 77/7页。
- Eurostat (1995年)。《欧洲社会统计学的未来：指导方针和战略》。卢森堡。
- _____ (1996年)。《欧洲社会统计学的未来：行政登记册与传播战略的使用》。卢森堡。
- _____ (1997年)。《欧盟的家庭预算调查：协调的方法与建议》。“人口与社会条件3, 方法E”。卢森堡。
- _____ (1998年a)。《劳动力调查：方法与定义》。卢森堡。
- _____ (1998年)。“中欧和东欧国家的劳动力调查、方法和定义(暂定)”。卢森堡。
- _____ (2001年)。“收入和生活条件统计工作队的会议, 2001年12月10日和11日”。卢森堡, 工作文件。
- Garner, T. I. 和其他人 (1993年)。“东欧经济状况住户调查：评估”。《转型期经济体的经济统计：1990年代的东欧》。哥伦比亚特区华盛顿：美国劳动力统计局, 第309-353页。
- Główny Urząd Statystyczny (GUS) (波兰中央统计办公室) (1971年a)。Badania statystyczne metodą reprezentacyjną w krajach socjalistycznych (“社会主义国家的抽样调查”)。Biblioteka Wiadomości Statystycznych, tom 14 (华沙), 第220页。
- _____ (1971年b)。Wybrane problemy metodologiczne badań reprezentacyjnych (“抽样调查的一些方法问题”)。Biblioteka Wiadomości Statystycznych, tom 15 (华沙), 第151页。
- _____ (1987年)。Problemy integracji statystycznych badań gospodarstw domowych (“统计住户调查一体化问题”)。Biblioteka Wiadomości Statystycznych, tom 34 (华沙)。

- _____ (1997年)。Stan dostosowania polskiej statystyki publicznej do standardów Unii Europejskiej. Harmonogram prac dostosowawczych (“波兰统计学与欧洲标准的协调情况”)。华沙。油印。
- _____ (1998年a)。Metodologia i organizacja mikrospisów (“小型人口普查的方法与安排”)。Statystyka w Praktyce。华沙。
- _____ (1998年b)。Budżet czasu ludności 1996 (“1996年时间利用调查”)。Studia i Analizy Statystyczne。华沙。
- _____ (1999年)。Metodyka badania budżetów gospodarstw domowych (“住户预算调查的方法”)。Zeszyty Metodyczne i Klasyfikacje。华沙。
- _____ (2000年a)。Aktywność ekonomiczna ludności Polski: I kwartał 2000 (波兰劳动力调查: 2000年第一季度)。Informacje i Opracowania Statystyczne。华沙。
- _____ (2001年)。Zasady wyceny kosztów prac statystycznych realizowanych przez służby statystyki publicznej w roku 2003 (“2003年官方统计中统计工作成本评估的原则”)。华沙。油印。
- Goskomstat (2000年)。《统计学的方法原则》第三版, 莫斯科, 第294页。
- Groves, R. M. (1989年)。《调查误差与调查成本》。纽约: 约翰·威利国际出版公司。
- _____和M. P. Couper (1998年)。《住户访谈调查中的不答复情况》。纽约: 约翰·威利国际出版公司。
- Kish, L. 和M. R. Frankel (1974年)。“根据复杂调查的推论(正在讨论)” 。《皇家统计学会志》, B辑, 第36期, 1-37页。
- Kordos, J. (1963年)。Seminarium statystyczne w Wiedniu (维也纳统计讨论会)。Przegląd Statystyczny, 第2号, 307-310页。
- _____ (1970年)。Mozliwości szerszego stosowania metody reprezentacyjnej w badaniach statystycznych krajów-członków RWPG (“经济互助委员会成员国统计调查中更广泛应用抽样方法的可能性”)。Przegląd Statystyczny, 第1号, 33-50页。
- _____ (1981年)。Problemy badań gospodarstw domowych: drugie spotkanie statystyków europejskich w Genewie (“住户调查的问题: 欧洲统计员第二次会议”)。Wiadomości Statystyczne, 第11号, 36-38页。
- _____ (1982年)。Metoda rotacyjna w badaniach budżetów rodzinnych w Polsce (“波兰住户预测调查中的轮调方法”)。Wiadomości Statystyczne, 第9号。
- _____ (1985年)。“实现波兰住户调查综合系统”。《国际统计学院通报》(阿姆斯特丹), 第51期, 第2卷, 13-18页。

- _____ (1988年a)。Jakość danych statystycznych (“统计数据的质量”)。Państwowe Wydawnictwo Ekonomiczne (华沙), 204页。
- _____ (1988年b)。“波兰的时间利用调查”。《联合国欧洲经济委员会统计期刊》, 第5期, 第159-168页。
- _____ (1996年)。“波兰四十年住户预算调查”。《转型期统计学》, 第2期, 第7号, 1119-1138页。
- _____ (1998年)。“波兰社会统计学及其与欧洲联盟标准的协调情况”, 《转型期统计学》, 第3期, 第4号, 第617-639页。
- _____ (2001年)。“波兰统计出版物中的一些数据质量问题”。《转型期统计学》, 第5期, 第3号, 第475-489页。
- _____ B. Lednicki和M. Zyra (2002年)。“波兰的住户样本调查”。《转型期统计学》, 第5期, 第4号, 第555-589页。
- Krapavickaitė, D. (2002年)。“立陶宛的住户样本调查”。《转型期统计学》, 第5期, 第4号, 第591-603页。
- _____, G. Klimavicius和A. Plikusas (1997年)。“聚类抽样中的一些估值”。《立陶宛数学学会第三十八届会议的记录》, 第298-303页。
- Kurvits, M. A.、K. Sõstra和I. Traat (2002年)。“爱沙尼亚住户调查: 着力于劳动力调查”。《转型期统计学》, 第5期, 第4号。
- Lapins, J. 和E. Vaskis (1996年)。“拉脱维亚的新住户预算调查”。《转型期统计学》, 第2期, 第7号, 第1085-1102页。
- Lapins, J和其他人 (2002年)。“拉脱维亚的住户样本调查”。《转型期统计学》, 第5期, 第4号, 第617-641页。
- Lednicki, B. (1982年)。Schemat losowania i metoda estymacji w rotacyjnym badaniu budżetów gospodarstw domowych (“住户预算轮调调查中的抽样设计与估计方法”)。Wiadomości Statystyczne, 第9号。
- Martini, A.、A. Ivanova和S. Novosyolova (1996年)。“白俄罗斯收入与支出调查: 设计和实施”。《转型期统计学》, 第2期, 第7号, 第1063-1084页。
- Mihalyffy, L. (1994年)。“1992年至2001年十年住户调查统一系统”。《转型期统计学》, 第1期, 第4号, 第443-462页。
- Postnikov, S. (1953年)。《选择其预算调查工作者、非工作者职工以及合作职工家庭的方法》。Vestnik Statistiki, 第3号, 第14-25页。
- Särndal, C-E.、B. I. Swensson和J. Wretman (1992年)。《调查取样的模型帮助》。纽约: 施普林格出版社。
- Šniukstiene, Z.、G. Vanagaite和G. Binkauskienė (1996年)。“立陶宛住户预算调查”。《转型期统计学》, 第2期, 第7号, 第1103-1117页。

- 爱沙尼亚统计办公室（1999年）。《1998年爱沙尼亚劳动力调查：方法报告》，塔林。
- 斯洛文尼亚统计办公室（2001年）。《斯洛文尼亚统计系统：全球评估》，2001年。法尔项目。卢布尔雅那。
- Szablowski, J.、J. Wesolowski和R. Wieczorkowski（1996）。*Indeks zgodności jako miara jakości danych: na podstawie wyników spisu kontrolnego do Mikrospisu 1995*（“相称指数作为一项数据质量措施：以小型人口普查事后调查为基础，1995年”）。*Wiadomości Statystyczne*，第4号，第43-49页。
- Szarkowski, A.和J. Witkowski（1994年）。“波兰劳动力调查”。《转型期统计学》，第1期，第4号，第467-483页。
- Traat, I.（1999年）。《重新设计住户预算调查：抽样小组的最后报告》。塔林：爱沙尼亚统计办公室。
- _____ A. Kukk和K. Sõstra（2000年）。“爱沙尼亚住户预算调查中的抽样和估计方法”。《转型期统计学》，第4期，第6号，第1029-1046页。
- 联合国（1964年）。《编制样本调查报告的建议（暂定问题）》。《统计丛书》，C辑，第1号，修订2。ST/STAT/SER.C/1/Rev.2。纽约。
- _____（1984年）。《住户调查手册（修订版本）》，方法研究，第31号。纽约。出售品编号：E. 83. XVII. 13。
- _____（2000年）。《根据目的进行的支出分类：政府的职能分类》。《根据目的进行的个人消费分类》。《服务住户的非营利机构目的分类》。《根据目的进行的生产者支出分类》。《统计丛书》，第84号。出售品编号：E. 00. XVII. 6。
- Vehovar, V.（1997年）。“斯洛文尼亚的劳动力调查”。《转型期统计学》，第3期，第1号（6月），第191-199页。
- _____（1999年）。“实地替换和单位不答复情况”。《官方统计杂志》，第15期，第2号，第335-350页。
- _____和M. Zaletel（1995年），“斯洛文尼亚的不答复趋势”。《转型期统计学》，第2期，第5号，第775-788页。
- Verma, V.（1995年）。“土耳其劳动力调查技术报告”。TUR/86/081号项目。日内瓦：国际劳工组织。
- Wolter, K. M.（1985年）。《方差估计简介》。纽约：施普林格出版社。

كيفية الحصول على منشورات الأمم المتحدة

يمكن الحصول على منشورات الأمم المتحدة من المكتبات والمؤسسات التعليمية والبحثية في جميع أنحاء العالم. يمكن أيضًا الحصول على منشورات الأمم المتحدة من المكتبات والمؤسسات التعليمية والبحثية في جميع أنحاء العالم.

如何索取联合国出版物

联合国出版物可在世界各地区的图书馆、研究机构、书店、大学、政府机构和民间组织等处索取。

HOW TO OBTAIN UNITED NATIONS PUBLICATIONS

United Nations publications may be obtained from libraries, institutions and distributors throughout the world. Consult our Issuance of Sales to the 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 spécialisées du monde entier. Pour en savoir plus, consultez notre adresse postale : Nations Unies, Section des ventes, New York ou Genève.

КАК ПОЛУЧИТЬ ПУБЛИКАЦИИ ОРГАНИЗАЦИИ ОБЪЕДИНЕННЫХ НАЦИОН

Издание Организации Объединенных Наций можно приобрести в книжных магазинах и в библиотеках во всем мире. Для получения информации о порядке приобретения обращайтесь по адресу: Организация Объединенных Наций, Секция продаж, Нью-Йорк или Женева.

CÓMO CONSEGUIR PUBLICACIONES DE LAS NACIONES UNIDAS

Las publicaciones de las Naciones Unidas están en venta en librerías y casas de venta de libros en todas las partes del mundo. Consulte el folleto sobre "Naciones Unidas, Sección de Ventas, Nueva York o Ginebra".