

SAS

数据支持与管理的 SAS编程技术及实现案例

——基于计算机辅助调查模式

孙玉环 著



科学出版社

数据支持与管理的 SAS 编程技术及实现案例 ——基于计算机辅助调查模式

孙玉环 著

本书由东北财经大学统计学院、中国博士后科学基金
(项目编号: 20110490216) 共同资助出版

科学出版社
北 京

内 容 简 介

本书基于笔者多年从事 SAS 软件教学及参与综合性社会调查项目数据处理实践的经验,首先阐述计算机辅助调查模式下综合性调查项目对数据管理和数据支持的需求,然后梳理常用的实现数据管理与数据支持任务的 SAS 过程、语句及其功能,重点从实际应用出发,基于中国家庭动态跟踪调查项目真实和复杂的实践情境,结合 12 个典型 SAS 编程案例,分别从数据需求、难点分析、解决思路、具体 SAS 实现程序以及程序编写技巧等角度进行详细解析,探讨实现综合性调查项目数据管理与数据支持实时化、规范化和高效化的 SAS 编程技术。

本书适合统计学、社会学、市场调查、社会调查等专业的本科生、研究生和教师,以及综合性调查项目执行机构、市场调查公司中承担数据管理与数据支持服务的技术人员等阅读和参考。

图书在版编目(CIP)数据

数据支持与管理的数据支持技术:基于计算机辅助调查模式 / 孙玉环著. —北京:科学出版社,2014

ISBN 7-03-040819-8

I. ①数… II. ①孙… III. ①统计分析—应用软件—研究 IV. ①C819

中国版本图书馆 CIP 数据核字(2014)第 115304 号

责任编辑:马 跃 / 责任校对:黄江霞
责任印制:阎 磊 / 封面设计:无极书装

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

骏杰印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2014 年 8 月第 一 版 开本:720×1000 B5

2014 年 8 月第一次印刷 印张:11 3/4

字数:234 000

定价:56.00 元

(如有印装质量问题,我社负责调换)

前言

为了把握社会变迁的长期趋势,使社会科学研究建立在全面系统的数据基础之上,自20世纪50年代以来,基于长期规划和周密设计而定期、系统地搜集社会各方面的数据,逐渐成为国际社会科学研究的趋势。国际上,综合性的社会调查数据在社会科学研究领域中的使用率仅次于人口普查数据。近年来中国也开始有研究机构致力于综合性社会调查数据的搜集与开发工作,并且已经取得了一定的社会影响。随着社会各界对民生问题的普遍关注,社会调查数据在社会科学研究领域和政府决策中发挥的作用越来越大,各方对调查数据的应用和质量也越来越重视。但由于综合性社会调查项目覆盖范围广,样本量大,参与人员众多,调查内容往往涉及政治、经济、文化、社会生活等各个方面,从项目设计到调查实施的各个环节,不仅难度大、内容繁杂,而且还几乎不可避免地存在着影响数据质量的各种不确定因素。因此,科学、高效、规范地实施综合性调查项目管理,尤其是加强调查执行过程中的数据质量核查过程,及时发现可能影响调查数据质量的各种因素,积极采取相应改进措施纠正系统偏差,成为保证综合性社会项目的调查效率和调查数据质量不可或缺的一环。

早期的调查执行模式主要采用纸笔面访、邮寄问卷等方式,随着电话、电脑、视讯和网络技术的普及,调查执行模式日趋多元化,开始出现计算机辅助调查(computer assisted interviewing, CAI)技术。计算机辅助调查技术的应用,一方面大大提高了数据采集的时效性,丰富了数据采集的内容和规模,另一方面也对调查项目各个环节的数据支持与管理提出了更为繁杂的任务需求和更高的技术要求。因此,对于采用计算机辅助调查模式的综合性社会调查项目来说,选择一种高效的分析工具软件参与数据管理与数据支持工作就显得尤为重要。可以实现数据管理和分析功能的常用软件有 SAS(statistical analysis system)、SPSS、STATA、R 等,其中, SAS 系统因其具有完备的数据存取、数据管理、数据分析和展现功能,高效、灵活地处理海量数据的能力,以及严谨的程序语言和输出结果的可重复验证等功能,而具有非常大的优势。

不过,目前涉及社会调查研究领域的文献大多是基础理论与方法的介绍,而针对性地研究调查执行各个阶段数据管理和数据支持实现方法与技术方面的文献并不多见。另外,可供社会科学研究者参考的 SAS 软件应用参考资料,多是泛泛地介绍 SAS 系统的基本语句和简单实现程序的例题,尚缺乏具有深刻实践背

景的综合性社会调查应用案例的支持。而且,由于国内综合性社会调查项目尚处于起步阶段,调查项目执行机构中承担数据支持与管理任务的技术人员,多为计算机、统计学或社会调查等单一学科或专业背景,大都不熟悉 SAS 软件的编程技术,而且往往缺乏处理实际大规模复杂数据的经验,而基于计算机辅助调查系统的各个执行环节,对数据支持与管理的任务需求却多种多样,如问卷电子化设计、样本调配、历史调查数据加载、数据质量核查、数据库清理与整理、衍生变量生成等,尤其是对于一些全国性的、连续或追踪调查的综合性调查项目来说,执行的各个环节对数据管理与数据支持任务的需求会十分复杂。正是由于存在以上种种困难,目前国内综合性社会调查项目各执行机构中,承担数据管理与数据支持任务的技术人员,往往只是利用自己相对熟悉的统计分析软件,如 SPSS、R、STATA 等,采用菜单操作的方式,或者编写或繁或简的一些小程序,实现一些基本的数据管理与数据支持任务。其实,这种做法很难保证分析过程的效率、准确性以及可重复验证性,甚至一些相对复杂的数据整理任务需求可能根本无法实现,同时也势必增加不同部门之间、不同项目之间的交流难度。因此,缺乏系统介绍计算机辅助调查模式下,综合性调查项目数据支持与数据管理的任务需求及其实现技术的相关文献,尤其是关于调查执行过程中数据质量核查内容和核查任务实现技术的参考资料,一直是困扰很多综合性社会调查项目执行机构的技术难题。

2010 年 4 月~2012 年 3 月,笔者有幸在北京大学中国社会科学调查中心(Institute of Social Science Survey, ISSS)从事博士后研究工作,期间全面参与了由该中心负责实施的中国家庭动态跟踪调查(China Family Panel Studies, CFPS)项目各个执行环节的工作,并基于 SAS 系统的编程功能,承担了其中数据库构建、数据质量核查、数据清理等数据管理与数据支持任务,积累了大量的具有实际应用价值的 SAS 程序代码。CFPS 项目是中国首个针对家庭、个人和社区层面的动态跟踪调查,也是中国首个使用计算机辅助面访调查(computer assisted personal interviewing, CAPI)系统的全国最大规模的抽样调查,项目采用分层、多阶段、概率与规模成比例的抽样方法,样本覆盖全国 25 个省区、160 个区县、640 个村居,2010 年基线调查的规模为 14 960 户、近 5 万人,目前已成功完成 2008 年和 2009 年两轮的试调查以及 2010 年基线调查和 2012 年追访调查。由于 CFPS 项目覆盖范围广、样本量大、调查内容烦琐、数据库结构复杂,而且依托荷兰统计局所开发的 Blaise 调查支持系统实现数据采集和基础数据库的建立过程,对数据库格式和加载信息提出了更为严格和繁杂的要求,因此加大了开展数据管理与数据支持的复杂程度和编写实现程序代码的难度,给 SAS 程序的实现设置了很多不可预见的障碍。CFPS 项目对承担数据管理与数据支持任务的技术人员提出了非常高的要求,不仅要十分熟悉整个调查实施的流程和问卷内

容,还要能够熟练掌握 SAS 编程技巧和具备一定的统计分析能力与逻辑能力。另外,由于项目管理过程几乎涉及参与调查项目的全部管理者、实地执行人员、数据清理人员以及报告撰写人员,管理数据和支持数据需要能够实现在不同部门之间的共享和交流,因此在编写完成数据管理与数据支持任务的 SAS 程序时,除了需要保证能够把数据整理和数据核查的结果等“算”出来,更要能够做到规范、友好地“呈现”出来,同时在保证计算结果准确性的基础上,还需要考虑程序运行的速度,并要兼顾程序的可读性、可共享性和可重复性。

本书基于笔者多年从事 SAS 软件教学及参与综合性社会调查项目数据处理实践的丰富经验,首先介绍计算机辅助调查技术的产生、发展以及主要模式的优点及适用范围,其次系统介绍国内外知名综合性社会调查项目的内容及应用计算机辅助调查技术情况,阐述计算机辅助调查模式下综合性调查项目执行过程中对管理数据的需求与数据支持的需求,其中重点介绍计算机辅助调查模式下质量控制的数据核查技术及核查方案,最后在梳理综合性社会调查项目中实现数据管理与数据支持任务常用的 SAS 过程、语句及其功能的基础上,以 CFPS2010 项目实验数据库和项目实施过程中的实际数据管理与数据支持需求为例,详细介绍 12 个实际综合应用案例的数据需求、难点分析、解决思路、具体 SAS 实现程序以及程序编写技巧等,为国内各高校开设 SAS 软件应用课程、社会调查和市场调查等课程的案例教学活动,为综合性社会调查机构规范数据管理和数据支持工作的内容并提高工作效率,提供可以直接借鉴的 SAS 编程实际应用案例。

全书的具体章节安排如下。

第 1 章为计算机辅助调查概述。对计算机辅助调查技术的产生及其发展、特点、计算机辅助调查系统的主要实施模式以及各模式的优点与适用范围进行系统介绍。

第 2 章为综合性社会调查项目简述。首先阐述综合性社会调查项目在社会科学领域中的作用,其次介绍国内外知名综合性社会调查项目的内容以及应用计算机辅助调查技术的情况。

第 3 章为计算机辅助调查模式下综合性调查项目数据管理与数据支持需求。首先介绍计算机辅助调查模式下综合性调查项目工作流程,其次从调查实施前的数据准备、调查实施阶段的数据需求以及数据整理阶段的实现目标三个方面,具体阐述计算机辅助调查模式下实现综合性调查项目数据管理与数据支持的技术需求内容。

第 4 章为计算机辅助调查模式下综合性调查项目数据质量核查技术及核查方案。具体介绍计算机辅助调查模式下数据质量控制的工作流程、数据核查方法与技术以及核查策略与方案设计等。

第 5 章为计算机辅助调查模式下 SAS 系统的优势及常用过程与语句。为了

帮助读者理解和掌握本书第 6 章中所介绍综合应用案例中的 SAS 程序编写技巧, 该章在简单介绍 SAS 系统的应用领域、基本模块、数据处理流程以及计算机辅助调查模式下应用 SAS 系统参与数据管理与服务的优势的基础上, 从数据获取与数据集操作、SAS 变量操作、SAS 观测值操作、SAS 数据集管理等几个方面, 对综合性社会调查项目中常用的实现数据管理与数据支持任务的 SAS 过程、SAS 语句及其具体实现功能进行梳理和介绍。

第 6 章为计算机辅助调查模式下 SAS 编程综合应用案例及技巧分析。该章将基于 CFPS2010 项目测试数据库, 针对项目执行过程中具体的数据支持与数据管理任务需求, 选取无回答率的计算、态度量表信度的计算、离群值的确定、家庭关系数据库的分解与重构、代际码等衍生变量的计算、追访调查预加载数据的提取、编码手册的制作、受访者个人信息的提取、外出人口分布状况及外出原因列联表分析、不同来源背景变量信息不一致情况的核查等共 12 个典型 SAS 编程案例, 分别从数据需求、难点分析、解决思路、具体 SAS 实现程序以及程序编写技巧等角度进行详细解析。

科学出版社对本书的出版给予了大力支持, 东北财经大学统计学院为本书的出版提供了资助, 特此致谢。

其实, 本书所详细阐述的 12 个 SAS 编程实际应用案例中, 有多个案例的内容笔者曾经在南京财经大学承办的中国高等教育学会高等财经教育分会统计学专业协作组 2013 年协作会、东北财经大学承办的 2013 年全国应用统计方法与案例系列培训班、江西财经大学主办的 2013 年全国经济统计高端论坛暨青年教师师资培训班以及湖北经济学院承办的中国商业统计学会市场调查与教学研究分会第 24 届年会等公开场合的特邀大会报告中, 做过或深入浅出或提纲挈领的讲解和介绍, 很多专家和同行对此表现出了极大的兴趣, 提出希望能够看到这些实践应用案例结集出版的建议, 这也是最终推动笔者整理并出版本书的外在动力。不过, 由于时间仓促和笔者的知识水平所限, 本书不足之处仍在所难免, 恳请各位同行专家批评指正, 以期在后续的研究中能够日臻完善。

孙玉环

2014 年 3 月

于辽宁省教育厅人文社科重点研究基地

东北财经大学国民核算研究中心

目 录

第 1 章	计算机辅助调查概述	1
1.1	计算机辅助调查技术的产生及其发展	1
1.2	计算机辅助调查系统的特点	1
1.3	计算机辅助调查系统的主要实施模式	2
1.4	计算机辅助调查系统各模式的优点与适用范围	4
第 2 章	综合性社会调查项目简述	6
2.1	综合性社会调查项目在社会科学研究领域中的作用	6
2.2	国外综合性社会调查项目简述	6
2.3	国内综合性社会调查项目简述	8
第 3 章	计算机辅助调查模式下综合性调查项目数据管理与数据支持需求	12
3.1	计算机辅助调查模式下综合性调查项目工作流程	12
3.2	计算机辅助调查模式下综合性调查项目数据管理与数据支持的技术实现需求	14
第 4 章	计算机辅助调查模式下综合性调查项目数据质量核查技术及核查方案	18
4.1	计算机辅助调查模式下综合性调查项目数据质量核查的重要性	18
4.2	质控数据准备及工作流程	19
4.3	数据质量核查方法与实现技术	20
4.4	数据质量核查策略与方案设计	25
4.5	数据核查分析报告的一般内容	26
第 5 章	计算机辅助调查模式下 SAS 系统的优势及常用过程与语句	28
5.1	SAS 系统介绍	28
5.2	计算机辅助调查模式下应用 SAS 系统参与数据管理与服务的优势	30
5.3	计算机辅助调查模式下常用 SAS 过程及语句介绍	32
第 6 章	计算机辅助调查模式下 SAS 编程综合应用案例及技巧分析	60
6.1	计算机辅助调查模式下 SAS 编程的技术需求	60
6.2	实验数据库介绍	61
6.3	应用案例	62
参考文献		173
后 记		177

第1章

计算机辅助调查概述

1.1 计算机辅助调查技术的产生及其发展

随着计算机技术的出现和逐步成熟,计算机辅助调查的理念和技术出现了。计算机辅助调查是指在问卷调查过程中,将传统的调查方式与计算机相结合,完成数据的搜集、录入与整理等工作,工作形式通常为通过电脑屏幕展现问题,由访问者或被访者依据交互式的电脑程序所呈现出的特定规则的问题,通过计算机直接录入答案。计算机辅助调查技术发源于欧美。电话调查在美国发展较早,20世纪70年代美国市场研究领域就出现了计算机辅助电话调查(computer assisted telephone interviewing, CATI)系统,而CAPI技术则是20世纪80年代率先在欧洲出现。

1.2 计算机辅助调查系统的特点

与传统的数据搜集方法相比较,计算机辅助调查系统的特点是以应用计算机为主,把传统的统计工作步骤,包括从数据搜集、录入到数据编辑,都在调查员的个人电脑上来完成,在回答者、调查者与数据用户之间形成一个有机结合的相互作用的循环。

相比于传统的纸笔访问,基于计算机辅助调查系统的采访模式具有以下明显优势。

(1)可以有效避免因问题跳转而选错路径所导致的错误。在计算机辅助调查模式下,借助于计算机程序设定的逻辑判断语句,对于有跳转的选项问题,通过恰当的逻辑关系设定,由计算机程序引导访员和被访者,依据前面回答的结果,由计算机程序来判断下面应出现什么问题,同时不需要被特定被访者回答的问题也可以自动跳过,从而避免因路径选择错误而造成漏题或错答等情况。因此,自动路径控制可以大大减少无效数据的数量。

(2)可以实现对所获取调查信息的同步有效性检验。在传统面访中,有效性检验通常只能在数据搜集后的数据审核阶段进行,这时即使发现错误,也只能进行有限的事后弥补,因为重新找到当时的被访者再进行验证和调查几乎不具有可操作性。在计算机辅助调查模式下,设计调查问卷时,可以事先对每一个问题设定允许的值域范围,甚至还可以进行更为复杂的多个答案相关联的有效性判断。数据采集过程中,系统会自动校验采集到的信息是否在允许的范围或者是否符合逻辑关系检验。当检查结果与设定逻辑不一致时,系统会出现校验提示。根据校验提示强度的不同,可分为强制性校验(hard check)和提示性校验(soft check)两种。其中,提示性校验更多地被运用于校验单一问题答案的异常情况。当提示性校验出现时,访员可以根据实际情况选择“强制通过”、“关闭修改当前答案”、“修改相关题目答案”三种操作模式。

(3)可以提供更加生动的表现方式。在计算机辅助调查模式下,图片、音频、视频、Flash、3D 等多媒体手段和形式的应用,可以大大丰富和加强数据采集的手段。另外,通过控制问题、问题组和选项的随机显示,还可以调整每个被访者所接受的提问顺序,从而有效避免访问顺序给访问结果带来的影响。计算机辅助实现的丰富多彩的人机交互功能,是在没有计算机辅助的情况下很难实现的。

(4)省却数据的录入过程。在计算机辅助调查模式下,搜集数据后,可以将其直接上传到数据中心,不仅缩短了数据录入时间,还可以有效避免数据录入环节的登记性误差。

(5)可以基于调查系统产生的并行数据开展访问行为管理等多方面的研究。并行数据是信息、通信技术与调查技术结合的产物。例如,CATI 中,访问日程安排系统记录下来的联系过程信息;CAPI 中,访员记录下来的观察信息;计算机辅助网络调查(computer assisted web interviewing, CAWI)中,后台服务器自动记录下来的被访对象鼠标移动的轨迹;计算机辅助可视调查(computer assisted visual interviewing, CAVI)中,系统自动记录的每个被访对象面对同一个语音问题的反应时间等。这些数据为从事综合性调查研究人员开展访问行为、访问质量等影响因素的研究提供了难得的基础数据。

1.3 计算机辅助调查系统的主要实施模式

计算机辅助调查技术的应用主要有五种方式,即 CATI、CAPI、计算机辅助自填式调查(computer assisted self interviewing, CASI)、CAWI 和 CAVI。

1. CATI 模式

CATI 即计算机辅助电话调查,是将近年高速发展的通信技术及计算机信息处理技术应用于传统电话访问的产物,问世以来已经得到越来越广泛的应用。

CATI 使用一份按计算机设计方法设计的问卷,利用电话对被调查者进行访问。访员坐在计算机前,面对屏幕上的问卷,向通话另一端的被访者读出问题,并将被访者回答的结果通过鼠标或键盘记录到计算机中去;督导在另一台计算机前借助局域网和电话交换机的辅助对整个访问工作进行现场监控。

2. CAPI 模式

CAPI 即计算机辅助面访调查,与 CATI 的基础原理基本一致,但其调查问卷的表现比 CATI 模式丰富很多,如支持照片、图片、概念卡显示,支持声音文件、影音文件播放,支持 Java、3D、Flash 等,支持虚拟购物、图片选择、刻度题等。

CAPI 系统通常的工作形式是,问卷由计算机管理与呈现,访员可依照计算机屏幕上的问题进行访问工作,并且将受访者回答的答案直接输入计算机中;如果受访者不希望通过访员填写问卷,亦可由受访者直接将答案输入计算机中以保护受访者的隐私。当访问结束或告一段落时,访员通过互联网直接将问卷调查结果传回到调查项目实施单位,实施单位在接收到调查数据后即可进行分析。

3. CASI 模式

CASI 即计算机辅助自填式调查,是指只有被调查者自己能够听到或阅读到计算机上的题目并且自行通过语音或键盘作答的一种调查方式。该模式没有访员辅助,被访者完全由计算机程序引导完成访问过程。在美国等国家的一些公益性调查中,CASI 这种辅助访问方式得到了广泛接受。在访员将问卷程序转交给被访者的同时,被访者依旧可以获得相关解释和帮助,因此,CASI 在某种程度上也可被视作 CAPI 的一部分。

4. CAWI 模式

CAWI 即计算机辅助网络调查。它的特点与 CASI 相似,是被访者通过网络回答在线问卷中的问题,受访者只要能够接入互联网或局域网,就可以回答问卷,是一种比较经济的调查方式。

利用互联网技术进行调查,收集数据的方法主要有网站(页)问卷调查(web-survey)、电子邮件调查(e-survey)、弹出式调查(pop up)、网上固定样本调查(int-survey)等。

5. CAVI 模式

CAVI 即计算机辅助可视调查,它是通过一种完全可视的方式收集数据的访谈数据采集系统,并可实现无纸化操作。提出问题后,答案会以按钮的形式出现,并且可能会是图片或文字。被访者答题时会感觉是在玩游戏,而不是填表格,特别是在需要随机选择或排除答案的时候。每当选定一个选项,被访者会听到轻微的太空时代的背景声音,还有可能听到画外音以及其他刺激听觉的声音。采用 CAVI 能够打破以往文本的限制,提供一个几乎全部是图像或可视的环境。

对于被访者来说,访问过程完全不同于大多数充满文字的网络或 CAPI 调查,整个受访过程独具格调、丰富多彩、引人入胜。

1.4 计算机辅助调查系统各模式的优点与适用范围

1. CATI 模式的优点与适用范围

CATI 的优点在于速度快、样本代表性强、访问容易控制。CATI 系统与传统面访方式相比,具有以下几个特点:①速度快,效率高。CATI 省去了面访所必需的上街或入户、问卷复核、问卷编码、数据录入等费时又费力的过程,研究人员在调查结束后几分钟或几十分钟内即可拿到调查数据。这在某些时效性要求较高的调查中,更突显其优越性。②质量高。由于事先可对计算机进行设置,可避免一些因跳问路线或选择答项错误而导致的数据差错或丢失。此外,由于所有访问记录(包括录音记录)全程都可以保存在系统中,而且访问过程始终处于督导人员的全程监控之中,因而一般不会出现访员作弊现象。③成本低。采用 CATI 系统,可省去交通费、礼品费和问卷印刷费等。据测算,完成同样的调查项目,CATI 所花费的费用要比面访调查低 30% 左右。

CATI 模式一般适用于电话普及率较高的中心城市,访问时间以不超过 20 分钟为宜;访问内容应该为被访者所熟悉或理解,不宜过多和过于复杂,也不宜过于敏感或涉及被访者隐私;不应有调查员向被访者出示照片、卡片、实物等背景资料的访问要求。

2. CAPI 模式的优点与适用范围

CAPI 的优点主要表现在以下几个方面:①可以利用计算机控制跳转逻辑,输入校验等,既能提高访问效率,又可以保证数据的合法性和完整性。②数据可以实时输入电脑,然后以电子方式传给服务器,无须再进行事后的数据录入,降低了数据审核和整理强度,提高了整体的数据采集效率。③无须印刷、邮寄纸面问卷及准备卡片、冲印照片,节省了成本,还可以支持逻辑关系复杂的问卷。④支持更多题型,尤其是对图片、音频、视频、3D 动画等多媒体的支持。⑤可以在移动环境下进行访问,如在机场对国际游客进行访问等。

CAPI 模式一般适用于包含很多图片的传统项目[如 U&A(消费者使用习惯和态度研究)、BEM(品牌价值研究)等],具有复杂逻辑控制或者条件的传统项目[如多概念的 CT(产品概念测试),CAPI 可以很方便地实现多个概念随机测试部分概念;多品牌的 BEM,CAPI 可以同时测试 20 个以上的品牌,使之能够覆盖整个行业],需要播放视频、声音等的测试项目(如广告测试、网页测试、外观测试等),需要实现独特功能的项目(如利用联合分析进行品牌、价格、包装研究;通过映像地图进行货架研究、平面广告测试等;应用拖曳题完成模拟消费者

的购买决策过程；通过记录被访者反应时间进行信息传达、消费心理挖掘等)。

3. CASI 模式的优点与适用范围

CASI 模式的优点是，可以在最大限度上保护被调查者的隐私，使得被调查者愿意在匿名的方式下提供真实的答案，以提高数据的质量和有效性。而且 CASI 模式采用统一的形式向被调查者提供调查问题，使得被调查者无论是以听到还是看到的方式接收问题都完全一样，消除了外界对受访者的影响。这种调查方式对被调查者文化水平的要求也大大降低，扩大了可接受调查的人群总体，所以目前 CASI 在国际领域得到广泛应用。

CASI 主要适用于对敏感问题以及要求能够记录被调查者回答问题时间的调查。

4. CAWI 模式的优点与适用范围

与传统的面访调查以及比较先进的 CATI 和 CAPI 相比，利用互联网技术进行的 CAWI 主要具有方便、快速的优点，而且能够节约费用，能有效提高问卷编写、调查发布的效率和采集数据的质量，是传统问卷调查方式最佳的替代品。

由于网络调查本身的特点以及现阶段我国网络调查客观上存在的不足，目前我国网络调查的适用范围还受到一定限制，但是已经在网络基础数据调查、电子商务个人消费市场研究调查及与电子商务用户密切相关的调查、对企业客户的网上调查、企业的广告或品牌追踪调查及客户满意度监控、政府网络调查等领域得到具体运用。

5. CAVI 模式的优点与适用范围

CAVI 具有工作效率高、客观性强、访问效果好和费用适度的优势，而且可以通过拦截定点访问、预约式访问等多种方式运作。这样，任何需要出示广告片、包装、概念的项目都可通过 CAVI 来完成，同时访问的质量也能得到有效控制。

CAVI 广泛应用于广告测试、概念测试、包装测试等需要互动的调查项目中，因为采用 CAVI 能够打破以往文本的限制，提供一个几乎全部是图像或可视的环境，在逻辑关系复杂或需要出示图片及视频资料的项目中，常常会被用到。

第2章

综合性社会调查项目简述

2.1 综合性社会调查项目在社会科学研究领域中的作用

为把握社会变迁的长期趋势，自20世纪50年代以来，基于长期规划和周密设计而定期、系统地搜集社会各方面数据，并应用纵贯数据(longitudinal data)开展社会科学研究，正逐渐成为国际上社会学、经济学、人口学等领域的主流。美国、英国、日本、德国、法国等发达国家以及一些发展中国家都先后实施了连续性的大型社会调查，并以此为基础建立了规模庞大的、开放性的社会调查数据库。而这些在中国社会科学研究领域一直是一个有待填补的空白。20世纪80年代以来，中国的体制改革和现代化建设使中国社会发生了一系列重大变革。在这样一个大变革的时代，无论是国家的政策制定，还是社会科学和社会政策研究的发展，都迫切需要通过定期、持续进行的全国性综合社会调查，以最小的成本、最快的速度、最优的质量，全面、系统地了解中国社会变迁的现状和趋势，揭示社会成员、社会群体的相对地位、角色和观念的变化，描述和分析社会分层和分群的实际状况以及存在的问题，并借助历时性和标准化的数据，推动历时性比较和横向跨国比较分析，为社会科学和社会政策的创新发展、为国家的社会政策制定和修改提供坚实的基础。

2.2 国外综合性社会调查项目简述

美国在综合性社会调查领域一直处于国际领先地位。美国密歇根大学调查研究中心(Survey Research Center, SRC)和芝加哥大学美国全国民意研究中心(National Opinion Research Center, NORC)是美国著名的两家学术调查与研究机构，代表着国际社会调查的最高水平。

收入动态跟踪调查(Panel Study of Incoming Dynamics, PSID)^①是美国密歇根大学调查研究中心于1968年开始开展的世界上最早的综合性社会调查项目,最初的设计样本为8000人,内容涉及家庭和家庭成员的几乎所有方面。项目的目的是通过跟踪搜集家庭的各项社会、经济和人口调查数据,反映美国家庭就业、退休、收入、支出、财富、养老金、医疗、保险、婚姻和生育、心理健康和慈善事业等方面的特征及变迁趋势。PSID项目现已完成连续46年的追踪调查,共涉及由6.5万多名成员组成的前后跨越四代的8000个美国家庭。

综合社会调查(General Social Survey, GSS)^②是由芝加哥大学美国全国民意研究中心负责实施的定期的大型社会调查项目,也是芝加哥大学美国全国民意研究中心最重要的品牌标志性调查项目,从1972年首创至今已有四十多年的历史。GSS调查作为一种社会调查类型,是一类具有全面性、系统性、标准化、定期性和连续性的调查项目。在美国,GSS数据是社会科学领域使用率除人口普查数据之外位居第二位的数据库。目前世界各主要国家都有自己的GSS调查。1984年,美国GSS调查联合其他国家的社会调查项目发起成立了国际社会调查协作项目(International Social Survey Programme, ISSP),至2012年已有49个国家(包括中国)和地区加入了ISSP。ISSP的成员国家及地区约占全球人口的60%、占全球国内生产总值(GDP)的80%。ISSP的各成员逐年提出和议定进行调查的各个标准化模块,然后根据各自的具体情况在本国或本地区确定调查内容和实施年度调查,并将调查结果和数据全面开放,进行切实有效的国际协作和比较研究。

青年人历时调查(National Longitudinal Surveys of Youth, NLSY)^③是反映美国的年轻人群体进入劳动力市场状况的一个权威性跟踪调查项目,由俄亥俄州立大学设计、芝加哥大学美国全国民意研究中心具体实施。该项目分别在1979年和1997年开始了对两批年轻人的长期跟踪调查,1979年的调查样本由12686名14~22岁的年轻人组成,1997年的调查样本约由9000名12~16岁的年轻人组成。NLSY的研究主题集中于年轻人的人力资本和劳动力市场活动,调查内容包括学校教育、雇佣状况、职业培训、工作时间、收入与资产、态度与行为、健康、政治参与等多个方面。

美国其他知名的综合性社会调查项目还有威斯康星追踪调查(Wisconsin Longitudinal Survey, WLS, 1957年至今)、健康与养老调查(Health and Retirement Study, HRS, 1992年至今)、青少年健康调查(Add Health, 1994年

① 关于PSID的介绍,参见<http://psidonline.isr.umich.edu/default.aspx>。

② 关于GSS的介绍,参见<http://www3.norc.org/gss%20website/>。

③ 关于NLSY的介绍,参见<http://www.bls.gov/nls/nlsy79.htm>, <http://www.bls.gov/nls/nlsy97.htm>。

至今)、全国健康访问调查(the National Health Interview Survey, NHIS, 1957 年至今)、美国家庭成长全国调查(National Survey of Family Growth, NSFG, 1973 年至今)等。

英国的理解社会调查项目(Understanding Society)^①是到目前为止样本量最大的跟踪调查项目,样本家庭户达到 4 万户,覆盖英格兰、苏格兰、威尔士以及北爱尔兰地区,受访对象为受访家户中年满 10 周岁的全部人口。项目从 2009 年 1 月开始实施,每年定期搜集关于社会环境、经济环境以及受访者个人观念等的信息,并从受访者中随机抽取 2 万人,进一步调查了解他们个人的健康情况等信息。

其他连续性的综合性社会调查还有,德国 1984 年开始的社会经济跟踪调查(German Socioeconomic Panel Study)、法国 1985 年开始的家庭户跟踪调查(French Household Panel Study)、英国 1991 年开始的家庭动态跟踪调查(British Household Panel Study)、日本 1994 年开始的消费者跟踪调查(Japan Panel Study on Consumers)。

目前,欧美国家已经使用计算机辅助调查技术的知名调查项目很多,如美国的 PSID、HRS、青少年健康调查,英国的家庭动态跟踪调查、理解社会调查,德国的社会经济跟踪调查,法国的家庭户跟踪调查,日本的消费者跟踪调查等。

2.3 国内综合性社会调查项目简述

近年来也开始有学者和研究机构致力于中国社会调查数据的搜集与开发工作,虽然起步较晚,但也获得了一定的社会影响力。比较有代表性的有:由中国疾病预防控制中心和美国北卡罗来纳大学联合主持开展的中国营养和健康调查(China Health and Nutrition Survey, CHNS),由中国人民大学和香港科技大学合作实施的中国综合社会调查(Chinese General Social Survey, CGSS),由中国社会科学院组织实施的中国社会状况综合调查(Chinese Social Survey, CSS),由北京大学组织实施的 CFPS 和中国健康与养老追踪调查(China Health and Retirement Longitudinal Study, CHARLS),以及由北京师范大学中国收入分配研究院负责组织实施的中国家庭收入项目(China Household Income Projects, CHIP)等。

1. CHIP 项目简介^②

在 20 世纪 80 年代末,Keith Griffin 和赵人伟联合带领由中国以及国外研究

① 关于理解社会调查的介绍,参见 <https://www.understandingsociety.ac.uk/about>。

② 关于 CHIP 项目的介绍,参见 <http://www.ciidbnu.org/chip/index.asp>。

者组成的团队,首次组织了一系列全国范围内的家庭住户调查,该调查即后来被人熟知的CHIP调查。该项目旨在收集家庭调查数据,用于支持关于中国改革开放以后的收入分配问题、贫困问题以及劳动力市场问题等的经验分析。在20世纪90年代中期,赵人伟和Carl Riskin牵头组织了第二轮的CHIP调查,而在21世纪初期,由Björn Gustafsson、李实和Terry Sicular组织了第三轮的CHIP调查。在2000年的中期,李实教授和澳大利亚国立大学孟昕教授一起组织了第四轮的CHIP调查。在1988年、1995年、2002年、2007年住户调查的基础上,形成的被国际学术界称为“CHIPS”的数据库,被称为迄今中国收入分配与劳动力市场研究领域中最具权威性的基础性数据资料。

2. CGSS 项目简介^①

由中国人民大学中国调查与数据中心和香港科技大学组织实施的CGSS,是中国最早开始执行的全国性、综合性、连续性的大型社会调查项目。从2003年开始每年一次,对全国125个县(区)、500个街道(乡、镇)、1000个居(村)委会、10000户家庭中的个人进行调查。通过定期、系统地收集中国人与中国社会各个方面的数据,总结社会变迁的长期趋势,探讨具有重大理论和现实意义的社会议题,推动国内社会科学研究开放性、共享性,为国际比较研究提供数据资料。至2008年止,CGSS一共进行了5次年度调查,完成了项目的第一期。自2010年起,CGSS开始项目的第二期,计划从2010年开始到2019年为止,每两年进行一次调查,共进行5次调查。CGSS于2006年被国际社会调查合作组织(International Social Survey Programme, ISSP)接纳为代表中国的会员单位,与全球近50个主要国家一起,对某个重要的社会议题进行每年一次的共同调查。2006年,CGSS还与日本的JGSS(Japan General Social Survey)、韩国的KGSS(Korea General Social Survey)、中国台湾的TGSS(Taiwan General Social Survey)一起,共同发起了东亚社会调查(East Asian Social Survey, EASS),该调查计划每两年进行一次,就一个共同模块进行调查。

3. CSS 项目简介^②

CSS是中国社会科学院社会学研究所于2005年发起的一项全国范围内的大型连续性抽样调查项目,目的是通过对全国公众的劳动就业、家庭及社会生活、社会态度等方面的长期纵贯调查,获取转型时期中国社会变迁的数据资料,从而为社会科学研究和政府决策提供翔实而科学的基础信息。调查采用概率抽样的入户访问方式,全国样本量约为7000~10000个家庭。自2011年第三期起,该项

^① 关于CGSS项目的介绍,参见<http://www.nsrcruc.org/index/article?id=527>。

^② 关于CSS项目的介绍,参见http://www.cssn.cn/sjxz/dept/zxxm/201312/t20131210_899020.shtml。