中国人口学自主知识体系构建——完善社会科学量化研究方法

样本结构与方法论陷阱：基于中国大型社会调查数据的比较分析*

刘文博 周 皓

【内容摘要】认知世界需要基于无偏且有效的经验知识。利用不同调查数据开展同一主题的研究可能得到不同结果，进而影响对社会现实的反映以及理论检验。本文以 2015 年全国 1%人口抽样调查为参照，比较中国 6 个大型社会调查的抽样设计方案、样本结构及相同模型设定下的统计分析结果。研究发现，各调查的抽样方法相似，但抽样过程存在差异；各调查数据在部分基础特征变量的结构上存在差异，且均与 2015 年小普查存在偏离；样本结构偏离导致各调查数据在相同模型设定下的统计分析结果存在差异；目标总体限定、权重设置、变量设置及操作化方式的调整均会通过改变变量联合分布而影响统计分析结果。理论分析表明，需要充分认识样本结构在统计推断中的基础性作用。为避免陷入样本结构偏离带来的方法论陷阱，建议研究者仔细阅读调查技术文件、充分考虑抽样设计对分析结果的可能影响。

【关键词】大型社会调查；样本结构；抽样设计；方法论陷阱

【作者简介】刘文博，哈尔滨工程大学人文社会科学学院准聘副教授；周皓（通讯作者），北京大学中国社会与发展研究中心、社会学系教授。电子邮箱：zhouh@pku.edu.cn

1 引言

定量研究的目标是检验理论。当前许多针对同一主题的研究，利用不同调查数据得到不同的分析结果，进而得出不同甚至完全相悖的研究结论，无法真正检验理论的可靠性与适用性。周皓（2023）分析指出：“由此得到的经验主义的因果关系无法真正地对应现实世界中的统计性因果，反过来还会导致以经验为基础的认知过程出现问题……由于经验性知识的偏差，同样也会无法真正地认识世界，无法得到理念世界中的纯粹知识。”

自我国社会学学科恢复以来，社会科学问卷式抽样调查已经过 40 多年的发展，积累了大量研究素材，不仅为了解变迁中的中国社会提供了丰富的定量资料，还为国家政策的制定提供了有力依据。理论上讲，针对相同总体的大型抽样调查在样本结构上不应存在显著差异，但事实上，由于调查目标、问卷设计、抽样框及抽样方式存在差异，不同大型抽样调查所得到的样本结构可能完全不同。然而，当前的定量研究或多或少地忽视了样本代表性等基础性问题，而更侧重所谓“高级”的统计方法。这种倾向可能会导致研究陷入方法论的陷阱，既不利于准确理解社会现实，也无法为理论检验和政策制定提供有力依据。

样本的结构性偏离可能源于抽样调查过程的各个步骤，既可能来自抽样框设定，也可能

* 本研究得到教育部人文社会科学重点研究基地重大项目“中国人口长期均衡发展关键问题研究”（22JJD840001）和黑龙江省哲学社会科学研究规划青年项目“黑龙江省人口流动变化趋势及其对人口高质量发展的影响研究”（24SHC004）的支持。

来自具体的抽样方法设计与调查实施过程。已有相关研究综述了近 40 年来中国社会学抽样调查的发展历程(风笑天,1989;李炜,2016);或从不同角度反思了社会科学抽样调查中的各种关键议题,如调查实施中的问题(严洁等,2012;孙妍等,2025)、针对某些特定人群的抽样设计(马侠,1999;庄亚儿、李伯华,2014)等;还有研究通过借鉴和反思国外调查中的经验和教训,为我国的抽样调查实践提供参考信息(史毅、刘鸿雁,2020)。这些研究对科学规范地进行抽样调查具有重要的指导意义。

虽然已有少许研究讨论了社会科学定量研究中的数据误用问题(吴肃然、张春泥,2025),反思了数据质量和前提假设对因果推断的影响(许加明、陈友华,2020),但鲜有研究系统性检视我国社会科学定量研究所运用的大型社会调查数据的样本结构,并深入到方法论层面讨论样本结构偏离的危害。为此,本文将基于国内 6 个重要的大型社会调查展开分析:首先,根据 6 个调查各自的技术说明文件,比较其抽样设计的异同;其次,以 2015 年全国 1% 人口抽样调查(后文简称"2015 年小普查")数据为参照,比较这 6 个调查的基础变量结构与 2015 年小普查数据的偏离状况;再次,在相同的模型设定下,检验基于这 6 个调查的样本数据得到的统计分析结果之异同,以呈现样本结构偏离可能带来的危害;最后,讨论因样本结构偏离而引发的方法论陷阱,并提出促进调查数据规范使用的相关建议。

需要特别指出的是,本文提及的"样本结构偏离"是指:由抽样设计等原因导致的样本分布对总体分布的系统性的、超越了随机误差允许范围的偏离。这种偏离表现在多个变量间的联合分布上,而不仅仅表现在单个变量上。

本文的研究目的并非评判不同调查数据的优劣,而是揭示样本结构偏离所隐含的方法论风险,以提醒研究人员谨慎使用数据,避免过度追求统计方法的复杂性,却忽视更为根本的样本代表性问题。

2 研究素材

本文的研究素材是当下国内最具代表性、使用最广泛的 6 个全国性大型社会调查:中国家庭追踪调查(China Family Panel Studies,CFPS)、中国综合社会调查(Chinese General Social Survey,CGSS)、中国家庭金融调查(China Household Finance Survey,CHFS)、中国健康与营养调查(China Health and Nutrition Survey,CHNS)、中国劳动力动态调查(China Labor-force Dynamics Survey,CLDS)、中国社会状况综合调查(Chinese Social Survey,CSS)。

本文以 2015 年小普查为参照展开分析,选择的大型社会调查开展的年份应与 2015 年小普查开展的年份相同或相近,故本文最终选择 CFPS2016、CGSS2015、CHFS2015、CHNS2015、CLDS2014、CSS2015 作为分析样本。

需要说明的是,虽然 2015 年小普查数据也可能存在问题,但这已经是可获得的最具代表性的数据,并且本文仅将其作为参照标准,而非绝对标准。即便某大型社会调查与 2015 年小普查在样本结构上存在差异,也不能说明该调查数据质量之好坏。本文无意评价不同调查数据的优劣。

3 抽样设计方案对比分析

了解调查数据的抽样设计是考察样本结构的基础。本文将从数据收集模式、目标总体、抽样方法、覆盖范围、抽样前分层方式、各级抽样单元的设定与抽取规模等方面比较各大型社会调查的抽样设计方案(见表 1)。

从数据收集模式来看,CGSS 和 CSS 为多期横截面调查,而 CFPS、CHFS、CHNS、CLDS 为追踪调查。截面调查的样本结构通常在每轮调查中有其独立代表性,但追踪调查的样本结构则可能因样本流失与新增样本异质性等问题,在多轮调查后逐步偏离总体,这在对比分析时尤需注意。

从目标总体来看,CGSS 和 CSS 主要考察全国的一般社会状况,其目标总体为全国成年人口,其中,CGSS 调查 18 岁及以上人口,CSS 调查 18~69 岁人口;CFPS 和 CHFS 作为家庭调查,其目标总体是家庭内全年龄段人口;CHNS 聚焦营养健康状况,其目标总体是全年龄段人口;CLDS 聚焦劳动力的现状与变迁,其目标总体是全国劳动力人口(15~64 岁)。目标总体的界定将直接影响调查设计,并进一步塑造样本的年龄结构及其所关涉的其他社会人口特征结构。

从抽样方法来看,虽然各调查的表述并不完全一致,但思路基本相同,大多采用多阶段、分层 PPS 抽样。各阶段的抽样单元分别为区县(直辖市直接为街道)、村/居委会和家庭户 3 级,略过了省级和地区级行政区划单位,提高了样本的分散性和覆盖性。

从覆盖范围来看,6 个调查中,只有 CSS 覆盖了中国大陆 31 个省级行政区划单位;CFPS、CGSS、CHFS、CLDS 都调查了 20 多个省级行政区划单位,CHNS 涉及 15 个省级行政区划单位。不同的覆盖范围背后隐含着样本所对应总体的差异。

从抽样前分层方式看,在抽取初级抽样单位之前,各调查都根据调查目标先进行分层以提高抽样精度。分层方式可分为 3 类。第一类,根据特定标准将所有省级行政区划单位分成两层,分别建立抽样框,如 CFPS 与 CGSS。CFPS 将上海等 5 个省级行政区划单位设定为大省层,使其具有省份代表性,其余为小省层;CGSS 对“那些发展处于国内领先水平的大城市特殊对待,作为必选层”,其余为抽选层。第二类,按照地理区域分层,如 CLDS 与 CSS。CLDS 将全国划分为东、中、西三大区域后,还在区域内根据人口规模进一步划分“人口大省”和“人口小省”,同时,为广东省设置了两层(非珠三角层和珠三角层),作为补充样本;CSS 按地理-行政区域划分出东北、华北、华东、中南、西北、西南 6 层。第三类,按照人均 GDP 或收入水平对区县进行分层,如 CHFS 与 CHNS。其中,CHFS 对城镇地区和富裕家庭进行了过度抽样。不同的抽样前分层方式既塑造了样本的区域结构,也关涉到各层样本对应的子总体及后续分析时的权重计算。

从初级抽样单元(Primary Sampling Unit,PSU)及其抽取规模来看,多数调查以区县一级为抽样框。CGSS 因划分了城市层(必选层)和其他省层(抽选层),故 PSU 也分为两种:街道和区县。CFPS 的设计中,仅上海市的 PSU 为街道/乡镇,其余省级行政区划单位的 PSU 均为区县。在抽取 PSU 时,大多数调查以人口规模、GDP、非农人口比例为依据,采用内隐分层的 PPS 抽样法,以提高抽样精度;而 CHNS 则采用加权抽样,CLDS 采用等距抽样。最终,CHFS 抽取的 PSU 数量为 351 个,CFPS 为 176 个,CGSS、CSS、CLDS 均在 150 个左右,CHNS 仅为 60 个。PSU 的数量与分布将直接影响样本的分散性及覆盖性,进而影响样本结构。

从次级抽样单元(Secondary Sampling Unit,SSU)及其抽取规模来看,6 个调查的 SSU 均为村/居委会。CFPS 和 CGSS 采用系统 PPS 抽样法,在城市层(必选层)和其他省层(抽选层)中分配 SSU 抽取数量;CHNS 和 CLDS 按城市(市辖区)和县(农村)分开抽取,其中,CHNS 在城市内抽取市区和郊区,在县内抽取镇和村,CLDS 先对农村和市辖区及其内部的村/居委会按不同指标排序,再按劳动力规模进行 PPS 抽样;CHFS 对村/居委会按非农人口比例排序后再进行抽取;CSS 则直接采用 PPS 抽样。最终,CHFS 抽取的 SSU 数量为 1396 个,CFPS 为 640 个,CSS 为 604 个,CGSS 为 480 个,CLDS 为 404 个,CHNS 为 330 个。

表1 各调查抽样设计方案比较

Table 1 Comparison of Survey Sampling Designs

调查项目	抽样方法	覆盖范围	抽样前分层方式	PSU	PSU层内排序指标	PSU层内抽样规则	PSU规模
CFPS	内隐分层的、多阶段、多层次PPS抽样	25个省份(不含港澳台地区、新疆、西藏、青海、内蒙古、宁夏、海南)	分大省层和小省层:大省层包括上海、辽宁、河南、甘肃、广东;小省层为其余20个省份	区县(仅上海市为街道/乡镇)	行政区划、社会经济地位(同级行政层内按人均GDP排序,无GDP指标则按非农人口比例或人口密度排序)	含内隐分层的系统PPS抽样	144个区县;32个街道/乡镇
CGSS	多阶段分层PPS随机抽样	第二期(2010~2019年)调查的抽样方案计划覆盖31个省份(不含港澳台地区),但2015年调查实际覆盖28个省份(不含港澳台地区、西藏、海南、新疆)	分必选层和抽选层:必选层包括上海、北京、深圳、广州、天津;抽选层为必选层以外的地区,根据因子分析划分出19个区层和31个县层	必选层:街道 抽选层:区县	必选层:按城市人口规模排序 抽选层:按因子得分(人口密度、非农人口比例、人均GDP)排序	必选层:与人口规模成比例的PPS抽样 抽选层:与各PSU人口规模成比例的系统PPS抽样	100个区县;40个街道
CHFS	分层三阶段PPS抽样	从2013年起,调查覆盖29个省份(不含港澳台地区、新疆、西藏)	2011年调查按照人均GDP分成10层	市县	2013年调查按人均GDP排序	PPS抽样	2011年调查为80个市县,2015年调查扩展至351个市县
CHNS	多阶段随机整群抽样	15个省份:北京、上海、重庆、广西、贵州、黑龙江、河南、湖北、湖南、辽宁、陕西、山东、云南、浙江、江苏	按照收入水平分层(具体层数不详)	市县	—	加权抽样	基线(1989年)调查时为36个市县;根据基线调查方案(每个省份抽取4个市县),后期调查应扩展至60个市县
CLDS	多阶段、多层次与劳动力规模成比例的概率抽样	29个省份(不含港澳台地区、西藏、海南)	采用综合考虑地理区域(东部、中部、西部、广东)与人口规模的分层抽样设计,共分为8层:东部人口大省、东部人口小省、中部人口大省、中部人口小省、西部人口大省、西部人口小省、广东非珠三角层、广东珠三角层	区县	在每个层内,按照省份、城市/县级市/县的顺序,对全部区县按GDP排序	随机起点,依据劳动力规模进行等距抽样	160个区县
CSS	多阶段分层混合抽样	31个省份(不含港澳台地区)	按照地理-行政区域划分为6层:东北、华北、华东、中南、西北、西南	区县	按经济发展、人口结构、教育水平3类共10个指标综合排序	分层比例抽样+PPS抽样	151个区县

表1 各调查抽样设计方案比较(续1)

Table 1 Comparison of Survey Sampling Designs (Continued 1)

调查项目	SSU	SSU抽选方法	SSU规模	末端抽样单元	末端抽样框	末端抽样方法	设计抽样家户数	实际访问家户数(或按应答率重新调整设计后抽取的家户数)
CFPS	村/居委会	采用系统PPS抽样(上海:每个PSU中随机抽取2个村/居委会;其他省层:每个PSU中随机抽取4个村/居委会)	640个	家庭户	地图地址抽样框	随机起点的循环等距抽样	25户	按应答率扩大样本规模,在不同地区实际调查28~42户
CGSS	村/居委会	必选层:采用与各居委会人口规模成比例的系统PPS抽样抽取2个居委会 抽选层:先按城市化水平分配村/居委会比例,再采用与人口规模成比例的PPS抽样在每个PSU中抽取4个村/居委会	480个	家庭户	地图地址抽样框	按门牌号排序的系统抽样	25户	必选层:实际调查50户 抽选层:居委会实际调查38户,村委会实际调查30户
CHFS	村/居委会	2011年调查:在每个PSU内,按非农人口比例分配村/居委会样本数,保证抽取的村/居委会之和为4个 2013年调查:在每个PSU内,先按非农人口比例对各乡镇/街道、村/居委会排序,然后使用以人口为权重的PPS等距抽样抽取4个村/居委会	2011年调查为320个,2015年调查扩展至1396个	家庭户	地图地址抽样框	等距抽样	2011年调查为20~50户	—
CHNS	村/居委会	城市:随机抽取60个市区和60个郊区 县:随机抽取30个镇和180个村	330个	家庭户	—	—	—	—
CLDS	村/居委会	市辖区:先按人均GDP或非农人口比例对市辖区排序,再按外来人口比例对居委会排序,最终按劳动力规模进行PPS抽样 农村:先按行政等级对县内的乡/镇/街道排序,再按外来人口比例对村/居委会排序,最终按劳动力规模进行PPS抽样	404个	家庭户	地图地址抽样框(排除空址和商用地址)	随机起点的循环等距抽样	19户	考虑追踪损耗(10%)以及多变量多层模型分析的需要,每个社区实际调查35户
CSS	村/居委会	PPS抽样	604个	家庭户	地图地址抽样框	简单随机抽样或等距抽样	—	对接触样本量进行了扩大(具体不详)

表 1　各调查抽样设计方案比较(续 2)

Table 1　Comparison of Survey Sampling Designs (Continued 2)

调查项目	户内抽样方法	基线家户数	基线个体数	基线应答率	2015 年(前后)家户数	2015 年(前后)个体数
CFPS	按照“同灶吃饭”原则识别家庭,并用 T 表调查所有符合条件的家庭成员	14960 户	57155 人	总体:81.25% 居委会:69.35% 村委会:89.16%	14763 户 (2016 年)	45319 人 (2016 年)
CGSS	使用 Kish 表在家庭内所有 18 岁及以上人口中随机抽取 1 人	11783 户	11783 人	74.32% (未分城乡)	10968 户	10968 人
CHFS	识别家庭的原则为:家庭中至少有 1 人为中国国籍,在本地居住满 6 个月以上,并且满足“共享收入”或“共担支出”中任一条件。对于多人家庭,可直接访问所有家庭成员。对于单人家庭,若无其他家人,可直接访问本人;若其他家人在外地且经济独立,则不视为本地家庭成员。	8438 户	29324 人	总体:88.4% 城市:83.5% 农村:96.8%	37289 户	133183 人
CHNS	调查所有家庭成员	3795 户	15907 人	—	7319 户	23937 人
CLDS	访问家庭中所有 15~64 岁的劳动力个体和 65 岁及以上目前有工作的同住成员。此外,农村地区的非同住家庭成员中符合一定筛选条件者,需由家人代答外出成员问卷。	10612 户	16253 人	总体:74.02% 市区:67.96% 县级市:78.14% 县:77.17%	14214 户 (2014 年)	23594 人 (2014 年)
CSS	采用简单随机抽样法在 18~69 岁的家庭成员中抽取 1 位	10206 户	10206 人	—	10243 户	10243 人

资料来源:CFPS 信息来源于《中国家庭动态跟踪调查抽样设计》[①]和《中国家庭追踪调查用户手册(第三版)》[②];CGSS 信息来源于项目官网[③]和《中国综合社会调查(CGSS)第二期(2010~2019)抽样方案》[④];CHFS 信息来源于《中国家庭金融研究(2016)》(甘犁等,2019),该文档显示,CHFS 的调查范围从基线(2011 年)调查到第二轮(2013 年)和第三轮(2015 年)调查逐步扩大,但该文档中仅对 2011 年调查的抽样设计进行了详细说明,对 2013 年调查的介绍不完整,有关 2015 年调查的抽样设计信息则基本缺失,故表 1 中呈现的是 2011 年与 2013 年调查的抽样设计,以及 2011 年和 2015 年调查的各级抽样单元规模;CHNS 信息来源于项目官网[⑤];CLDS 信息来源于项目官网[⑥];CSS 信息来源于项目官网[⑦]以及《中国社会状况综合调查:数据介绍及应用说明》(杨标致、李炜,2024)和《全国居民纵贯调查抽样方案设计研究》(李炜、张丽萍,2014)。

① 获取网址为 https://www.isss.pku.edu.cn/cfps/docs/20200520161539050175.pdf。
② 获取网址为 https://www.isss.pku.edu.cn/cfps/wdzx/yhsc/。
③ 获取网址为 http://cgss.ruc.edu.cn/index.htm。
④ 获取网址为 http://cgss.ruc.edu.cn/xmwd/cysj.htm。
⑤ 获取网址为 https://chns.cpc.unc.edu/。
⑥ 获取网址为 http://css.sysu.edu.cn/,目前暂未开放。
⑦ 获取网址为 http://sociology.cssn.cn/css_sy/。

从末端抽样方法来看,如果说前两个阶段的抽样主要影响调查的覆盖范围,那么末端抽样则会从根本上影响样本的随机性和代表性。一旦末端抽样的随机性被违反,即便前两个阶段严格遵循随机原则,整个调查的代表性也无法保证。几乎所有调查都选择了地图地址抽样法,且多数调查(CFPS、CGSS、CLDS、CSS)都根据应答率对实际访问的家户数进行了扩大,以保证获得足够的样本量。

从户内抽样方法来看,CFPS 和 CHFS 都为家庭调查,因此被抽中家庭户内的所有成员均为调查对象;CHNS 虽未说明户内抽样方法,但从调查结果来看,被访对象是所有家庭成员;CLDS 的调查对象是家庭中 15~64 岁的劳动力个体和 65 岁及以上目前有工作的同住成员,以及农村地区的非同住家庭成员中符合一定筛选条件者。以上调查都不需要户内再抽样,而 CGSS 和 CSS 则分别利用 Kish 表和简单随机抽样法抽取 1 名家庭成员进行访问。家庭成员的界定及调查对象的选取方式会进一步影响样本结构。

上述描述分析简单勾勒了各调查的特点,能够看出各调查的抽样设计有着明显差异,而这些差异成为样本结构偏离的潜在根源。

4　样本结构对比分析

本文将从区域、城乡、性别、年龄、教育、婚姻、民族等基础变量维度描述各调查的样本结构,并与 2015 年小普查进行比较。比较过程中,本文对部分数据进行加权,以尽可能保证其代表性:CGSS 和 CSS 按“weight”变量进行加权;CLDS 按调查调整权重“wpr”进行加权①。此外,CFPS 将选择家庭关系库中的全国再抽样样本进行分析,不作加权处理②;CHFS 在 2015 年数据中仅有家庭权重而无个体权重,CHNS 数据未提供权重③,故均未加权。

4.1　区域结构与城乡结构

表 2 展示了各调查与 2015 年小普查的样本在各基础变量上的分布情况。本文首先考察与抽样设计关联最为直接的区域结构。区域结构根据“省份”变量划分得到④。结果显示,CSS 的区域结构与 2015 年小普查较为接近,其次是 CLDS。相比之下,其他调查的区域结构与 2015 年小普查存在较明显差异。其中,CFPS 和 CHNS 的东部地区比例低于 2015 年小普查,而中、西部地区比例高于 2015 年小普查;CGSS 和 CHFS 则是东部地区比例高于 2015 年小普查,而西部地区比例低于 2015 年小普查。

结合抽样设计来看,各调查的省份覆盖范围及抽样前分层方式均不一致,这可能会影响样本的区域结构:CSS 和 CLDS 都是按地理行政区划进行分层,因而其样本结构更接近 2015 年

---

① CLDS 的调查调整权重根据受访者的拒答情况计算得出,可以在一定程度上重新还原数据的代表性。

② 根据《中国家庭追踪调查用户手册(第三版)》中“8. CFPS2010 年基线调查数据初步统计分析和评估”部分,在将 2010 年 CFPS 调查与 2010 年全国人口普查数据进行对比时,采用的是基于全国再抽样样本的非加权结果。

③ 详见 CHNS 项目方发布的技术文档“Weights for the China Health and Nutrition Study”,获取网址为 https://dataverse.unc.edu/file.xhtml?fileId=7505036&datasetVersionId=30770。

④ 东、中、西部地区划分参照国家统计局划分标准,详见 https://www.stats.gov.cn/sj/zxfb/202302/t20230203_1900530.html。

小普查;CFPS 和 CHNS 的东部地区比例偏低,而中、西部地区比例偏高,可能是因为分别受到大省/小省分层和覆盖省份较少的影响;CGSS 和 CHFS 的东部地区比例偏高,可能是因为 CGSS 在分层时将 5 个东部发达大城市分为一层,提高了东部地区样本的入样概率,而 CHFS 则是对富裕家庭进行了过度抽样。

从城乡结构来看,CSS 的城乡结构与 2015 年小普查较为接近。相比之下,其他调查的城乡结构与 2015 年小普查存在较明显差异。其中,CGSS 和 CHFS 的城镇居民比例相对较高,而 CFPS、CHNS、CLDS 的农村居民比例相对较高。这些差异同样可以部分追溯至抽样设计:CGSS 将 5 个东部发达大城市分为一层,而 CHFS 对城镇地区和富裕家庭进行了过度抽样,这些都可能导致样本中城镇人口比例偏高;CHNS 则可能是由于抽取的乡村数在 SSU 中的占比较高(330 个 SSU 中有 180 个村),导致其样本中农村人口比例偏高。

表 2 各调查与 2015 年小普查样本结构对比

Table 2 Comparison of Sample Structures between Different Surveys and the 2015 Mini-census

单位:%

指标	2015 年小普查	CFPS2016	CGSS2015	CHFS2015	CHNS2015	CLDS2014	CSS2015
区域结构							
西部地区	27.12	29.70	24.35	25.28	28.61	27.08	27.37
中部地区	31.43	35.29	33.92	27.10	36.32	32.58	30.81
东部地区	41.45	35.01	41.73	47.62	35.07	40.33	41.82
城乡结构							
城镇	55.88	48.84	60.33	63.99	32.91	41.33	56.97
农村	44.12	51.16	39.67	36.01	67.09	58.67	43.03
教育结构							
未上过学	6.24	25.70	15.04	10.19	10.04	11.34	8.58
小学	21.76	20.65	24.33	21.22	20.82	22.77	20.58
初中	40.08	27.66	26.65	31.08	32.18	32.94	32.84
普通高中	12.24	15.01	11.73	14.04	14.68	12.15	13.04
中职	4.39	—	5.99	5.68	7.94	5.66	5.69
大学专科	7.79	6.39	7.16	7.64	—	7.09	9.68
大学本科	6.83	4.18	8.01	9.19	13.79	7.50	8.68
研究生	0.68	0.41	1.08	0.96	0.55	0.54	0.91
婚姻结构							
未婚	16.52	14.06	12.99	14.88	10.88	12.58	16.57
有配偶	75.96	78.38	69.76	78.67	81.66	84.21	79.07
离婚	1.79	2.16	3.14	1.31	1.36	1.34	1.78
丧偶	5.73	5.40	14.12	5.15	6.10	1.87	2.58
民族结构							
汉族	91.46	87.58	92.95	—	88.02	89.55	91.61
少数民族	8.54	12.42	7.05	—	11.98	10.45	8.39

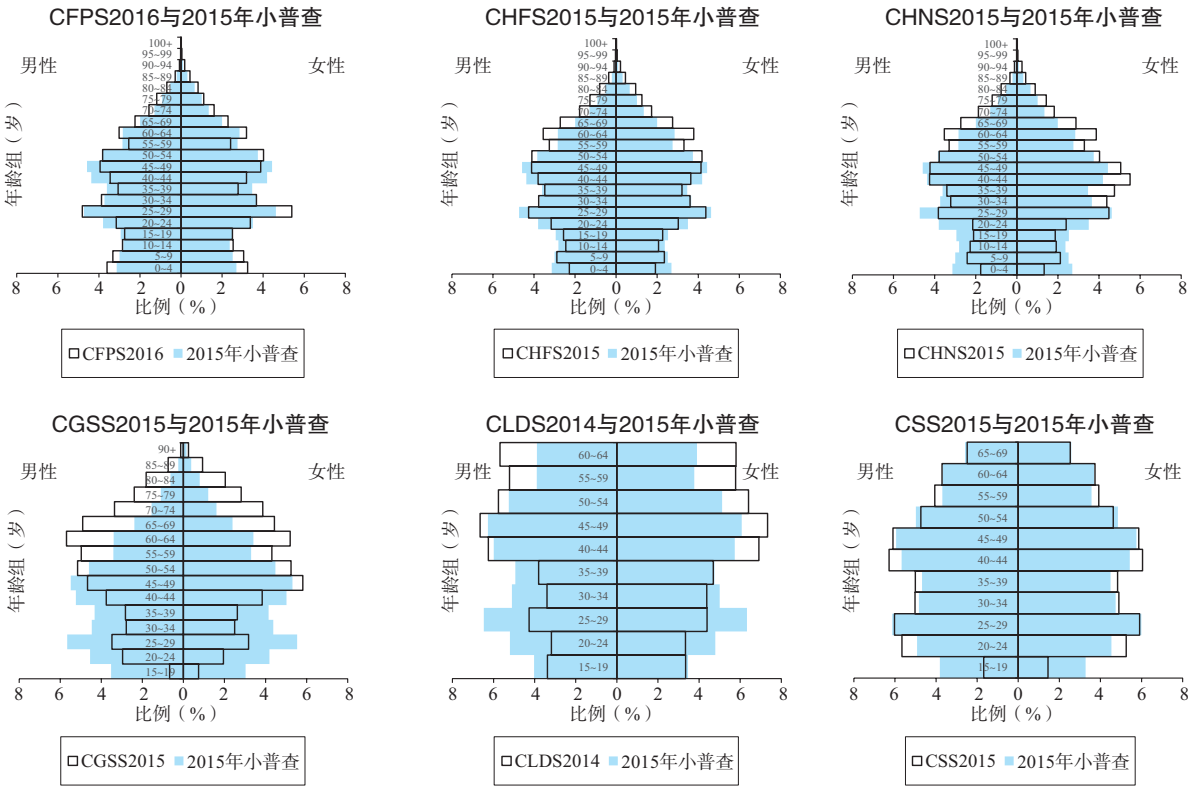
资料来源:根据《2015 年全国 1%人口抽样调查资料》以及 CFPS2016、CGSS2015、CHFS2015、CHNS2015、CLDS2014、CSS2015 中相关数据计算整理得到。

4.2　社会人口特征结构

图 1 展现了各调查样本的性别年龄金字塔与 2015 年小普查样本性别年龄金字塔的对比情况。由于各调查样本的年龄范围差异较大,此处仅对比 2015 年小普查样本与各调查样本对应的年龄段。

图 1　各调查与 2015 年小普查样本的性别年龄结构对比

Figure 1　Comparison of Age and Gender Structures between Different Surveys and the 2015 Mini-census



资料来源:同表 2。

从 3 个收集了全年龄段样本的调查与 2015 年小普查的对比情况来看,CHFS 中 49 岁及以下样本的比例略低于 2015 年小普查,而 50 岁及以上样本的比例则略高于 2015 年小普查,但其性别年龄金字塔的总体轮廓与 2015 年小普查相似;CFPS 中 10 岁及以下样本的比例略高于 2015 年小普查,25~29 岁女性的比例明显高于 2015 年小普查,35~49 岁样本的比例低于 2015 年小普查,60 岁及以上样本的比例略高于 2015 年小普查,但其性别年龄金字塔总体轮廓亦与 2015 年小普查相似;相比之下,CHNS 的性别年龄结构与 2015 年小普查差异较为明显,其 29 岁及以下样本的比例明显低于 2015 年小普查,30~54 岁女性的比例明显高于 2015 年小普查,而该年龄段男性的比例则略低于 2015 年小普查,55 岁及以上样本的比例明显高于 2015 年小普查。从其余 3 个调查与 2015 年小普查的对比情况来看,CSS 的性别年龄结构与 2015 年小普查较为接近,仅 15~19 岁组因样本抽取原因(CSS 只调查了 18 岁及以上者)差异较大;相比之下,CGSS 和 CLDS 的性别年龄结构均与 2015 年小普查存在较明显差异,具体表现为低龄组样本的比例偏低,而高龄组样本的比例偏高。

作为最具代表性的基础变量,年龄结构将直接影响其他变量的联合分布。各调查的年龄结

构都与 2015 年小普查存在一定偏离,这可能源于 5 个方面的原因。一是目标总体不同。CLDS 主要针对劳动力人口展开调查,其余调查虽都面向全国人口总体,但 CGSS 主要调查 18 岁及以上人口,CSS 则将调查对象年龄限定在 18~69 岁,其年龄结构自然不同。二是抽样框不同。各调查的抽样框所覆盖的省份各不相同,各省人口发展进程差异及由此导致的人口结构差异,会使各调查的样本结构与全国总体结构产生偏离。三是分层设计不同。6 个调查采用了 3 种分层模式,且每种模式内部的分层标准也不同,导致各调查在区县一级的抽样误差各不相同,进而会影响样本的年龄结构。四是层内抽样规则不同。多数调查都使用 PPS 抽样,并抽取 140~160 个区县,但 CHNS 使用加权抽样仅抽取了 60 个市县,这会导致其区县一级的分散性不足,进而影响样本结构。五是末端抽样不同。各调查的户内抽样方法存在差异,有的调查采用 Kish 表法,而该方法可能会导致年龄结构偏离(张丽萍,2009)。可见,在使用数据时,必须考虑调查的覆盖范围及抽样设计不同所带来的年龄结构偏离问题。

在教育结构方面,由于各调查样本的年龄范围不一致,本文统一考察各调查中 18 岁及以上人群的受教育水平。结果显示,各调查样本教育结构的共性在于:未上过学的人群比例均高于 2015 年小普查,初中学历的人群比例均低于 2015 年小普查,大学本科学历的人群比例大多高于 2015 年小普查(CHNS 未区分大学本科与专科,导致表 2 中呈现的大学本科比例偏高),其余教育阶段的人群比例均与 2015 年小普查相差不大。相比之下,在教育结构上与 2015 年小普查差异较为明显的是 CFPS,其未上过学的人群比例相对较高,而大学专科及以上学历的人群比例则相对较低。从抽样设计角度来看,由覆盖范围、分层原则等方面的原因所导致的各调查样本的年龄结构偏离,会进一步影响到教育结构。CFPS 由于覆盖了全年龄段人群,包含更多没上过学的老年人,故其样本中未上过学的人群比例较高①;而同为家庭调查的 CHFS 虽也覆盖了大量老年人口,但其为了收集高收入样本,对富裕家庭进行了过度抽样,故可能会拉高高学历人群的比例。

在婚姻结构方面,本文的对比分析同样针对各调查中 18 岁及以上人群。结果显示,CFPS 和 CHFS 的婚姻结构与 2015 年小普查较为相似;CHNS 中的未婚者比例较低、有配偶者比例较高,其他情况则与 2015 年小普查相似;CLDS 和 CSS 均表现为有配偶者比例较高、丧偶者比例较低,而 CGSS 中的丧偶者比例则过高。婚姻结构的偏离可能与末端人群选择有关。CFPS 和 CHFS 由于是家庭调查,访问户内所有人②,所以其婚姻结构与 2015 年小普查基本一致。CLDS 和 CSS 的丧偶比例偏低,可能是因为其主要针对劳动力人口或 70 岁以下人群展开调查,未收集到足够的老年丧偶人群。CGSS 虽覆盖了 18 岁及以上人口,但入户时只调查 1 人,可能使其与全国总体的婚姻结构有些偏离。因此,在研究婚姻家庭问题时应慎重选择分析数据。

从民族结构来看,在包含民族选项的 5 个调查中③,CSS 的民族结构与 2015 年小普查基

① 这也可能与变量测量方式有关,CFPS 的“个人最高学历”变量中对应“未上过学”的类别是“文盲/半文盲”,即该调查将“未读完小学”的人也划入此类,从而可能会提高此类别的占比。

② 当然,对家庭户成员的界定也可能会影响婚姻结构,此处暂不讨论。

③ CHFS2015 并未调查受访者的民族身份,亦无法从先前轮次的调查中完全匹配相关信息,故本文未呈现其民族结构数据。

本一致;CGSS 的少数民族比例略低于 2015 年小普查,其余 3 个调查(CFPS、CHNS、CLDS)的少数民族比例则高于 2015 年小普查,均超过 10%。这些差异可能与各调查覆盖的省份有关。

综上所述,各调查的样本在区域、城乡、性别、年龄、教育、婚姻、民族等结构上均与 2015 年小普查存在一定的差异,并且这些差异均可不同程度地追溯至各调查的调查目标和抽样设计等方面。当然,除了这些因素之外,抽样实施过程、无应答情况、加权调整等都可能影响样本结构。此外,对于追踪调查而言,由于样本追踪相对困难,经过多轮调查后,其样本结构与调查时点的总人口结构产生一定偏离亦属正常现象。归结来看,抽样设计(特别是 PSU 和 SSU 抽中数量及在地域空间的分散情况)会直接影响样本对总体的代表性,而调查过程中其他可能引发误差的环节(如户内访问对象的选择、被访者的应答情况等)会进一步导致样本结构的偏离。

上述比较揭示了各调查可能存在的样本结构偏离情况。当然,作为参照标准的 2015 年小普查数据同样会由于抽样框设定、概念定义(测量)和漏报误报等原因而存在结构偏离。因此,并不能据此判断各调查数据质量的好坏。

5　样本结构偏离对统计分析结果的影响

为考察样本结构偏离的后果,本文将选取各调查数据中共有的变量,构建相同的模型,比较统计分析结果。此处构建模型的目的不在于讨论因果关系,而仅在于揭示样本结构偏离可能给统计分析结果带来的影响。

5.1　样本界定、变量选择及模型设定

由于各调查的目标总体存在差异,在进行对比分析时,应考虑研究总体不同对分析结果的影响。为此,本文在选择样本时采取了两种策略。一是直接使用全样本。尽管目标总体存在差异,但在既往研究中,各调查数据均被广泛视为能够代表全国总体情况。本文旨在揭示这种做法的潜在问题。二是限定统一年龄范围。将各调查纳入分析的样本限定在 18~64 岁①,以构建一个更具可比性的研究总体,从而在一定程度上控制因总体差异而导致的样本结构影响。

本文选择受教育年限和自评健康作为因变量。二者均为基础变量,且分别为连续变量和分类变量,可以呈现不同模型设定下样本结构偏离的影响。受教育年限由最高学历转化得到。自评健康的测量题干在不同调查中略有差异:CFPS、CGSS 和 CLDS 中对应的问题为“您认为自身的健康状况如何?”;CHNS 为“与同龄人相比,您的健康状况如何?”;CHFS 则是让被访者评价家庭成员“与同龄人相比的健康状况”②。本文统一将各调查中的自评健康变量转化为“健康与否”的二分类变量。CSS 因在 2015 年未测量自评健康,无法参与这一比较。

受限于“各数据共有”的要求,自变量仅选择性别、年龄、民族、户口、婚姻、所在区域这些基础变量(自评健康模型中的自变量还包括受教育年限)。需要说明的是,由于本文构建模型仅作比较之用,而非真正检验受教育年限和自评健康的影响因素,所以尽管模型设定可能存在缺陷,但对模型结果的比较仍是有意义的。

在模型构建过程中,本文综合考虑了各调查数据的权重设置、变量设置、变量处理方式以

① “18~64 岁”这一年龄范围是综合了所有调查中的最高年龄下限(18 岁)和最低年龄上限(64 岁)的结果。

② 测量方式(包括问题的题干表述和选项设置(如对称或单侧)等)也会影响统计分析结果,此处暂不予讨论。

及目标总体,建立了几组不同模型。首先,由于加权与否可能会影响分析结果,本文先建立了一组不加权模型,比较各调查数据分析结果之异同,再建立一组加权模型,考察加权后各调查数据分析结果之异同。加权时,本文控制了各调查的 PSU。其次,考虑到 CHFS 中没有民族变量,本文还设置了包含民族变量的一组模型和不含民族变量的一组模型,以考察变量设置的影响。再次,年龄变量(在自评健康模型中则是年龄变量和受教育年限变量)可以有连续型和类别型两种处理方式,其意义并不相同。为此,本文分别采用连续变量和分类变量两种操作化方式各构建一组模型,并比较分析结果的异同。最后,考虑到目标总体差异的影响,本文还比较了各调查数据在全样本与统一限定年龄范围的样本之间的分析结果之异同①。

在相同的模型设置下,本文考察的是:同一自变量回归系数的显著性和方向在不同调查样本中的异同。显著性反映的是自变量对因变量的作用在总体中是否存在,事关因果关系能否成立,是研究关注的核心问题;方向的异同则关涉研究结论及理论检验。本文不讨论回归系数具体数值的大小。

5.2 受教育年限的影响因素模型

表 3 展现了不同调查样本中受教育年限影响因素模型的回归结果。在相同的模型设置下,民族和婚姻两个变量回归系数的显著性和方向在不同调查样本间存在差异。民族变量在 CHNS 样本中呈现不显著的正向影响,在其他调查样本中则均呈现显著的正向影响。婚姻变量中,有配偶变量的系数在 CFPS、CGSS、CHFS、CSS 的样本中显著为负,在 CHNS 的样本中显著为正,而在 CLDS 样本中为负但不显著;离婚或丧偶变量的系数在 CHNS 样本中为负但不显著,在其他调查样本中则显著为负。正如前文所述,各调查样本的民族结构及婚姻结构存在明显差异;而回归结果表明,在相同的模型设置下,民族和婚姻两个变量在不同调查数据中呈现出不同的显著性和作用方向,这体现了样本结构对统计分析结果的影响。

表 3 受教育年限的影响因素模型

Table 3 Models of Determinants of Years of Schooling

变量	CFPS2016	CGSS2015	CHFS2015	CHNS2015	CLDS2014	CSS2015
性别(参照组=女性)	1.277***	1.395***	1.100***	1.347***	1.379***	1.474***
	(0.055)	(0.067)	(0.021)	(0.062)	(0.045)	(0.068)
年龄	−0.139***	−0.131***	−0.120***	−0.128***	−0.129***	−0.134***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.003)
民族(参照组=少数民族)	1.301***	0.281*		0.098	1.028***	0.994***
	(0.095)	(0.127)		(0.105)	(0.077)	(0.133)
户口(参照组=农业)	3.818***	3.821***	3.844***	3.455***	4.231***	3.866***
	(0.063)	(0.071)	(0.021)	(0.064)	(0.049)	(0.072)
婚姻(参照组=未婚)						
有配偶	−0.214*	−0.413***	−0.109**	0.797***	−0.134	−0.623***
	(0.090)	(0.120)	(0.035)	(0.147)	(0.078)	(0.131)

① 为节省篇幅,本文仅呈现了每个因变量的一组模型结果,其他模型结果可联系作者获取。

续表3

变量	CFPS2016	CGSS2015	CHFS2015	CHNS2015	CLDS2014	CSS2015
离婚或丧偶	−0.721***	−0.952***	−0.975***	−0.005	−0.724***	−0.995***
	(0.146)	(0.168)	(0.059)	(0.193)	(0.148)	(0.197)
所在区域(参照组=西部地区)						
中部地区	0.590***	0.521***	0.448***	0.545***	0.282***	0.446***
	(0.074)	(0.088)	(0.029)	(0.082)	(0.064)	(0.093)
东部地区	1.012***	1.327***	0.889***	1.343***	0.472***	0.931***
	(0.073)	(0.090)	(0.026)	(0.079)	(0.057)	(0.087)
截距项	10.389***	12.410***	12.314***	12.049***	11.271***	12.085***
	(0.114)	(0.171)	(0.040)	(0.175)	(0.098)	(0.177)
观测值	20762	10867	103960	12739	22162	10178
R^2	0.376	0.466	0.425	0.399	0.407	0.411

资料来源:根据 CFPS2016、CGSS2015、CHFS2015、CHNS2015、CLDS2014、CSS2015 中相关数据计算整理得到。

注:①*** p<0.001,** p<0.01,* p<0.05,后表同。②括号内数据为标准误,后表同。

表 3 中的模型均未经加权处理,为此,本文还建立了一组加权模型进行对比。首先,在加权后,同一变量在不同调查样本间的回归结果差异发生变化,各调查样本不仅在民族和婚姻两个变量的回归结果上呈现出差异,还在所在区域变量的回归结果上出现差异。其次,即便是同一调查样本中的相同变量,其作用也随加权调整而发生了显著变化。例如,在 CGSS 样本中,民族变量在加权前呈现显著的正向作用,加权后其作用则变得不再显著;在 CFPS 样本中,有配偶变量在加权前呈现显著的负向作用,加权后则变为不显著的正向作用。加权调整仅改变了案例权重,也即样本结构,因而由此导致的结果差异能更直观地反映样本结构对统计分析结果的影响。

在上述比较中,由于 CHFS 样本中缺少民族变量,导致基于该样本建立的模型与基于其他调查样本所建立的模型仍不完全一致,这也可能影响比较结果。为消除这种影响,本文在加权模型的基础上继续建立了剔除民族变量的受教育年限影响因素模型。结果显示,相较于完整模型,在剔除民族变量的模型中,婚姻变量的回归结果在不同调查样本间依然存在差异,但所在区域变量的回归结果差异却消失了。可见,是否加入民族变量会影响整个样本的联合分布,进而影响统计分析结果。这揭示出样本结构(特别是变量的联合分布)对统计分析结果的影响。

上述模型将年龄视为连续变量,本文在加权模型的基础上进一步将年龄由连续变量转换为分类变量,考察改变变量处理方式后的统计分析结果差异。结果显示,将年龄变为分类变量后,各调查样本在婚姻变量和所在区域变量上的回归结果差异也较先前有所变化。这表明,对变量的不同处理方式也会影响统计分析结果,而且不仅会影响被调整的变量自身的回归结果,还会影响到其他相关变量的回归结果。

本文还对比分析了统一限定各调查样本年龄范围(18~64 岁)后的模型结果,以考察在目标总体保持一致的情况下样本结构差异对统计分析结果的影响。结果显示,统一限定年龄范围

后,各调查样本在婚姻变量和所在区域变量的回归结果上仍存在显著差异,但这种差异又不同于基于各调查全样本的回归模型之间的差异。这也体现了样本结构(多变量联合分布)的影响,即限定年龄范围后,各调查样本内部的年龄结构,以及与此相关的婚姻、所在区域、受教育水平等结构,乃至这些变量之间的联合分布都发生了变化,进而改变了统计分析结果。

基于上述分析可见,如果某项研究的核心自变量是民族,并想要讨论各民族间教育获得的差异问题,那么由于各调查样本得出的统计分析结果不一致,该研究便无法得出确切的最终结论。如果在模型中继续增加变量,样本结构随着社会分组的改变将发生更多变化,基于不同调查样本的统计分析结果也将显现出更加复杂的相互矛盾现象。这或许就是基于不同调查数据的研究会得到不同甚至相悖结果的重要原因。

值得注意的是,在这些模型中,3 个基础特征变量——性别、年龄、户口在不同调查样本中均呈现出完全相同的显著性和作用方向,即男性、年轻、具有非农户口的群体普遍有更高的受教育水平。这种一致性反映了社会现象及其背后结构的稳定性,这本身就是一种社会规律的体现。从这一视角来看,各调查数据间统计分析结果的比较可以被视为一种稳健性检验,若在样本结构存在差异的情况下,基于各调查数据的统计分析结果仍然有共性,则不仅说明结果稳健,更表明其真实地反映了潜在的社会现实。这有助于加深对社会结构的理解。

5.3 自评健康的影响因素模型

本文将再以分类的自评健康为因变量构建一组二元 Logistic 回归模型,以考察不同模型设定下样本结构对统计分析结果的影响。由于 CSS2015 未包含自评健康变量,以下分析仅比较基于其他 5 个调查数据的回归结果。

如表 4 所示,各调查样本同样在多个变量的显著性和作用方向上呈现出差异。其中,性别变量在 CHNS 样本中呈现不显著的正向影响,而在其他调查样本中则均呈现显著的正向影响;民族变量在 CHNS 样本中呈现显著的负向影响,但在其他调查样本中,其影响均不显著;户口变量的系数在 CFPS 样本中不显著,但在其他调查样本中均显著为正;婚姻状况对自评健康的影响仅在 CHNS 样本中显著,在其他调查样本中则均不显著;所在区域变量中,中部地区变量的系数在 CFPS 和 CHFS 样本中不显著,但在其他调查样本中均显著为正。这些结果同样表明,样本结构可能是导致统计分析结果出现差异的重要原因,而且这种有差异的统计分析结果无法用于检验相应的理论。比如,在讨论离婚或丧偶对自评健康的影响时,基于 CHNS 数据的分析结果表明前者对后者存在显著的正向作用,而基于其他调查数据的分析结果则表明二者之间不存在显著的相关关系。此时,离婚与丧偶对自评健康的影响到底如何,就形成了“公说公有理,婆说婆有理”的局面,自然无益于检验相关理论。

与受教育年限模型相似,本文针对自评健康模型也建立了加权模型、剔除民族变量的模型、将年龄与受教育年限转换为分类变量的模型,以及统一限定各调查样本年龄范围的模型。在这些模型中,变量处理方式和各变量联合分布的变化同样导致部分变量的显著性和作用方向较先前有所改变。然而,年龄变量和受教育年限变量的回归结果在基于各种调查数据的各类模型中均表现出一致性,即年龄越小、受教育年限越长的群体的自评健康状况越好。这种一致性也可以被视为社会因素影响的稳定性。

表 4　自评健康的影响因素模型

Table 4　Models of Determinants of Self-Rated Health

变量	CFPS2016	CGSS2015	CHFS2015	CHNS2015	CLDS2014
性别(参照组=女性)	0.384***	0.222***	0.140***	0.104	0.218***
	(0.043)	(0.057)	(0.019)	(0.072)	(0.046)
年龄	−0.041***	−0.040***	−0.037***	−0.035***	−0.056***
	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)
受教育年限	0.062***	0.066***	0.085***	0.055***	0.077***
	(0.005)	(0.007)	(0.003)	(0.009)	(0.006)
民族(参照组=少数民族)	0.124	0.078		−0.334**	0.032
	(0.070)	(0.096)		(0.120)	(0.070)
户口(参照组=农业)	0.103	0.211**	0.312***	0.175*	0.393***
	(0.053)	(0.065)	(0.022)	(0.079)	(0.060)
婚姻(参照组=未婚)					
有配偶	−0.129	0.027	−0.075	0.610**	0.197
	(0.098)	(0.134)	(0.043)	(0.204)	(0.104)
离婚或丧偶	−0.015	0.102	0.007	0.501*	−0.176
	(0.123)	(0.157)	(0.056)	(0.236)	(0.141)
所在区域(参照组=西部地区)					
中部地区	−0.030	0.175**	0.003	0.309***	0.166**
	(0.055)	(0.066)	(0.024)	(0.084)	(0.057)
东部地区	0.142*	0.510***	0.555***	0.658***	0.594***
	(0.056)	(0.073)	(0.023)	(0.088)	(0.055)
截距项	3.204***	2.685***	2.582***	3.266***	3.330***
	(0.121)	(0.188)	(0.054)	(0.258)	(0.140)
观测值	20758	10862	103589	12628	22161
Pseudo R^2	0.112	0.117	0.127	0.070	0.120
LL	−7780	−4511	−38079	−3218	−7156

资料来源:根据 CFPS2016、CGSS2015、CHFS2015、CHNS2015、CLDS2014 中相关数据计算整理得到。

6　结论、讨论与建议

6.1　结论和讨论

本文从抽样设计方案、样本结构、统计分析结果 3 个方面对比分析了我国 6 个最具代表性的大型社会调查,得到 4 个主要结论。第一,各调查采用的抽样方法基本相似,大多为多阶段、分层 PPS 随机抽样法,但各调查在具体的抽样过程(包括抽样框覆盖范围、分层原则、各级抽样单元的抽取方式与数量、户内抽样原则等)上存在较大差异。第二,各调查数据在一些

基础特征变量的结构上存在明显差异，并且各调查的样本结构均与 2015 年小普查的样本结构存在一定偏离。第三，样本结构偏离会导致统计分析结果的差异，在相同的模型设定下，基于各调查数据的统计分析结果既存在反映共性社会现实的一致之处，也在某些变量的显著性和作用方向上呈现出差异性。第四，目标总体限定、权重设置、变量设置及操作化方式的调整均会改变样本内变量的联合分布，并显著改变统计分析结果。若样本结构本身存在差异，这些调整可能会进一步改变基于不同调查数据的统计分析结果之间的差异。

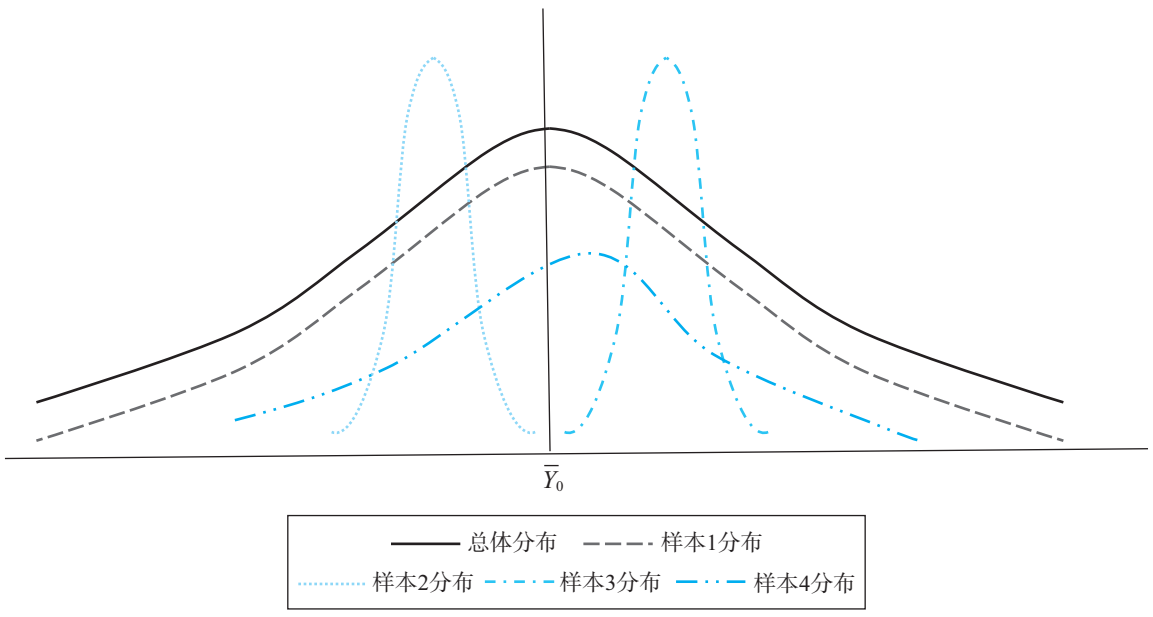
上述分析表明：国内各大型社会调查的样本结构均与全国人口总体结构存在一定偏离，这使得基于相同模型但不同调查数据的统计分析结果呈现出显著差异。这在一定程度上回答了这一问题：在利用均具有全国代表性的不同调查数据分析同一研究问题时，为何会得出不同的结论。

从方法论角度来看，基于抽样调查数据的社会科学定量研究主要遵循“假设—演绎”的实证范式，其核心目标在于理论检验，即利用抽样调查数据检验由理论推导得到的假设，并回应理论的适用性。然而，假设检验的基本前提是样本结构与总体结构一致；若二者之间存在较大偏离，则无论采用何种高级统计方法，得到的结论都无法真正反映社会现实。

以图 2 为例：假设 $\bar{Y}_0$ 是虚无假设对应的假设值，黑色实线为研究总体的分布，样本 1 代表与总体结构一致的某个样本，样本 2、样本 3、样本 4 则分别表示 3 种与总体结构存在偏离的样本。如图 2 所示，在虚无假设为真的条件下：样本 2 中假设显著且为负向；样本 3 中假设显著且为正向；样本 4 中假设为正向但不显著；基于样本 1 所得出的结论则与总体无显著差异。这 4 种差别较大的结论事实上与假设的正确性及其背后的理论无关，仅与样本和总体之间的偏离程度有关。可见，样本结构在假设检验过程中具有基础性和重要性。

图 2　不同样本分布下的假设检验结果

Figure 2　Results of Hypothesis Tests under Different Sample Distributions



资料来源：作者自行绘制。

再进一步,样本结构偏离对定量研究的方法论影响主要可以归纳为以下 3 个方面。

其一,导致统计推断失效。假设检验中存在两类错误:一类是弃真错误;另一类是纳伪错误。这两类错误都与抽样误差相关。通常情况下,抽样总误差共有 3 种来源:抽样变异、抽样偏倚、非抽样偏倚(加里·T·享利,2008)。其中,后两种来源共同构成偏差(Bias),其通常是非随机的,并决定着总误差的大小。利用结构存在偏离的样本进行统计推断时,未被识别的偏差会被全部纳入残差中,导致用于推断估计的抽样方差不再是纯粹的随机抽样变异,而是包含了抽样变异和偏差,并且由偏差决定大小(必然会大于抽样变异)。由此造成的后果是:假设检验中的接受区间由于抽样方差受偏差影响而扩大,统计推断中的显著性水平 α 不再真实反映仅由随机抽样变异引起的弃真错误概率,进而得到有偏的统计结果与研究结论,最终导致统计推断失效。因此,避免这两类错误的前提是样本对总体具有代表性。基于有代表性的样本计算得到的统计量(包括抽样方差)才能被用于推断总体。

其二,掩盖或扭曲真实的变量关系。基于结构有偏的样本得出的结论只能反映总体中与该样本对应的部分人群(子总体)的共性,而无法反映真正意义上的社会共性,即其所揭示的并非社会的普遍规律。若结构有偏的样本遗漏了研究总体中的某些群体(尤其是少数群体),那么研究结论将无法捕捉到作用于该群体的社会机制,从而掩盖其独特性。若样本结构的偏离程度较高,那么基于偏离样本所得的结论甚至可能严重扭曲真实的社会规律。总之,样本结构偏离不仅会导致统计推断失效,还会使研究者得出有悖于社会现实的结论,阻碍人们正确认识社会规律。

其三,导致因果推论失灵。探索社会现象中的因果规律是社会科学的重要目标。定量研究不仅发展出被称为因果推论“黄金准则”的实验方法,还针对难以实施实验的大样本情境,发展出工具变量、倾向值匹配、断点回归、双重差分等准实验方法,这些方法共同构成社会科学因果推论的重要路径。然而,样本结构偏离对因果推论构成了巨大挑战。本研究团队曾利用仿真实验数据考察样本结构对因果推论的影响,发现样本结构偏离会使因果推论叠加抽样过程、调查过程和数据处理过程的选择性以及进入实验的样本选择性,从而使分析样本中的实验组和控制组均无法代表其各自对应的总体,进而导致因果推论中的实验效应估计量产生偏差,并且这一问题无法通过统计手段得到根本解决(周皓,2023)。因此,在样本结构有偏的情况下,无论采用何种巧妙的研究设计都无法避免因果推论失灵。

总之,样本结构偏离会引发多维度的方法论陷阱。基于有偏样本得出的结论,既无益于理论检验与实践应用,也会损害定量研究的学术公信力。这也提醒数据使用者:样本结构对统计分析的影响是基础且关键的,若忽略样本结构而一味追求高级的统计方法,很可能陷入方法论陷阱。换言之,定量研究不能简化为秀方法、雕模型、玩设计;相反,方法、模型和设计发挥作用的重要前提之一是确保样本本身的重要属性——代表性与随机性,尤其是代表性。

6.2　建议

基于上述分析,本文从研究人员(数据使用者)和调查机构(数据提供者)两个方面提出建议,以期规范学界对调查数据的使用,避免定量研究陷入样本结构偏离带来的方法论陷阱。

对于研究人员,本文具体提出5条建议。第一,应仔细研读各类调查提供的技术文件,从目标总体、覆盖范围、抽样方法等方面系统理解调查的抽样设计方案对样本结构的潜在影响,避免陷入样本结构偏离带来的方法论陷阱。第二,应基于研究目标选择合适的调查数据,应特别注意到目前已有的各种大型社会调查数据可能只适用于总体性的因果关系讨论,而并不适用于对特定子群体的推断。比如,在讨论作为子群体的流动人口的问题时,利用各调查数据估计的流动人口总体规模与现实规模存在一定偏离。第三,应重视数据的缺失处理和加权调整。数据的缺失处理和加权调整均会改变样本结构,处理不当可能会造成样本结构的"二次偏离"。因此,应科学评估数据缺失的影响,选择适当的缺失处理方式。第四,应重视对分析结果的稳健性检验。考虑到当前各种抽样调查普遍存在的局限,研究者可通过"样本替换"等方式对比基于不同调查数据的统计分析结果,以检验研究结论的稳健性。第五,需对所用调查数据的样本代表性进行充分评估或说明。样本在多大程度上能够代表研究总体,直接决定了从统计分析结果中得出的结论可以推广到怎样的范围。

对于调查机构,本文建议其提供有关调查各阶段抽样单位及相应权重的更详细的信息,使研究人员在分析过程中能够尽可能地还原抽样过程。同时,调查机构可提供更翔实的技术文档,包括调查设计、抽样设计、数据质量评估、缺失值填补以及数据使用注意事项等内容,使研究人员能更恰当地使用数据。

全国性大型社会调查的实施总是十分困难的。已有的这些大型社会调查不仅为记录我国社会转型时期的社会变迁提供了充分的信息,也为繁荣我国社会科学的定量研究作出了重要贡献。本文无意否定各项调查的贡献与作用,只希望提醒数据使用者深入理解、正确使用调查数据,正确研判数据分析结果,以准确认识世界。

参考文献/References:

1 风笑天.我国社会学恢复以来的社会调查分析.社会学研究,1989;4:12-18
Feng Xiaotian. 1989. Analysis of Social Surveys since the Restoration of Sociology in China. Sociological Studies 4:12-18.

2 甘犁,吴雨,何青,何欣,弋代春.中国家庭金融研究(2016).成都:西南财经大学出版社,2019:1-13
Gan Li, Wu Yu, He Qing, He Xin, and Yi Daichun. 2019. Research of China Houshold Finance (2016). Chengdu:Southwestern University of Finance and Economics Press:1-13.

3 [美]加里·T·亨利著.沈崇麟译.实用抽样方法.重庆:重庆大学出版社,2008:40
Henry G. T. 2008. Practical Sampling. Translated by Shen Chonglin. Chongqing:Chongqing University Press:40.

4 李炜.与时俱进:社会学恢复重建以来调查研究的发展.社会学研究,2016;6:73-94+243
Li Wei. 2016. Advancing with Time:A Review of the Development of Sociological Surveys in the Past Three Decades in China. Sociological Studies 6:73-94+243.

5 李炜,张丽萍.全国居民纵贯调查抽样方案设计研究.科研信息化技术与应用,2014;6:17-26
Li Wei and Zhang Liping. 2014. A Study on Sampling Design for Longitudinal National Residential Survey. Frontiers of Data & Computing 6:17-26.

6　马侠.74 城镇人口迁移调查回顾——兼论代表性选点与概率抽样相结合的调查方法.人口研究,1999;1:36-43
Ma Xia. 1999. A Review of the 74-City Urban Population Migration Survey:With a Discussion on the Survey Method Combining Representative Site Selection and Probability Sampling. Population Research 1:36-43.

7　史毅,刘鸿雁.家庭追踪调查质量控制的策略与方法.统计与决策,2020;18:5-10
Shi Yi and Liu Hongyan. 2020. Strategies and Methods for Quality Control of Family Follow-up Survey. Statistics & Decision 18:5-10.

8　孙妍,王赫,潘修明,张春泥.大型追踪调查样本流失规律及模式——以中国家庭追踪调查为例.调研世界,2025;3:63-76
Sun Yan, Wang He, Pan Xiuming, and Zhang Chunni. 2025. The Patterns and Trends of Sample Attrition in Large-scale Panel Surveys:A Case Study of China Family Panel Studies (CFPS). The World of Survey and Research 3:63-76.

9　吴肃然,张春泥.让数据贴近事实:解析大型社会调查数据的应用问题.浙江社会科学,2025;8:102-112+159-160
Wu Suran and Zhang Chunni. 2025. Making Data Reflect Reality:Analyzing the Application Issues of Large-Scale Social Survey Data. Zhejiang Social Sciences 8:102-112+159-160.

10　许加明,陈友华.数据质量、前提假设与因果模型——社会科学定量研究之反思.社会科学研究,2020;2:130-139
Xu Jiaming and Chen Youhua. 2020. Data Quality, Premises, and Causal Models:Reflections on Quantitative Research in Social Science. Social Science Research 2:130-139.

11　严洁,邱泽奇,任莉颖,丁华,孙妍.社会调查质量研究:访员臆答与干预效果.社会学研究,2012;2:168-181+245
Yan Jie, Qiu Zeqi, Ren Liying, Ding Hua, and Sun Yan. 2012. A Study on the Quality Control of Social Survey:Intervention in "Curb-Stolen" Behaviors and Its Effects. Sociological Studies 2:168-181+245.

12　杨标致,李炜.中国社会状况综合调查:数据介绍及应用说明. 社会研究方法评论,2024;1:116-140
Yang Biaozhi and Li Wei. 2024. Chinese Social Survey:Data Introduction and Application Guide. Social Research Methods Review 1:116-140.

13　张丽萍.应用 Kish 表入户抽样被访者年龄结构扭曲问题研究.社会学研究,2009;4:177-195+245
Zhang Liping. 2009. Age Structure Distorted Problem of Applying Kish Table for the Household Interview. Sociological Studies 4:177-195+245.

14　周皓.样本结构性偏差与因果推论——基于实验数据的分析. 社会研究方法评论,2023;2:126-188
Zhou Hao. 2023. Sample Structural Bias and Causal Inference:An Analysis Based on Experimental Data. Social Research Methods Review 2:126-188.

15　庄亚儿,李伯华.流动人口调查抽样的实践与思考.人口研究,2014;1:30-36
Zhuang Yaer and Li Bohua. 2014. Reflections on Sampling of Migrants Surveys. Population Research 1:30-36.

Sample Structure and Methodological Pitfalls: A Comparative Analysis Based on Large-Scale Social Survey Data in China

Liu Wenbo Zhou Hao

Abstract: Understanding the world requires unbiased and valid empirical knowledge. Numerous studies on the same topic, employing different survey data, often produce divergent analytical results and even contradictory conclusions, which undermines the effective testing of theoretical reliability and applicability. However, existing studies predominantly focus on refining statistical methods while overlooking foundational issues such as sample representativeness. There is also a scarcity of systematic examinations into the sample structures of widely used large-scale social surveys and their impact on statistical findings.

To address this gap, this study draws on six most extensively used national large-scale social surveys among Chinese scholars. It compares their sampling designs and empirically investigates the similarities and differences in their sample structures. Using a consistent model specification, this study investigates the impact of deviations in sample structure on statistical analysis results, and reveals the underlying logic by which sample structure influences statistical inference.

The main findings are as follows. First, although almost all surveys employ a multi-stage, stratified Probability Proportional to Size (PPS) random sampling method, they exhibit significant differences in sampling frame coverage, stratification principles, the sampling methods and quantities of sampling units at each stage, and within-household sampling procedures. Second, notable disparities exist in the distributions of key demographic variables across the surveys. Moreover, each survey's sample structure deviates to some extent from that of the 2015 National 1% Population Sample Survey. Third, differences in sample structure lead to variations in statistical results. Under identical models, analyses based on different survey data yield both a consensus component reflecting shared social realities and significant discrepancies in the significance and direction of effects for certain variables. Fourth, adjustments in population definitions, weighting schemes, variable selection, and operationalization alter the joint distribution of variables within a sample, thereby significantly affecting statistical outcomes. When sample structures differ initially, such adjustments may further amplify discrepancies in results across different survey datasets. Fifth, the foundational role of sample structure in the methodology of statistical inference must be fully acknowledged.

Based on these findings, the study recommends that researchers should meticulously review survey technical documentation, prudently select appropriate survey data based on research objectives, appropriately address data missingness and weighting, prioritize robustness checks of analytical results, and thoroughly evaluate or explain the sample representativeness of the survey data used. Survey institutions, on the other hand, should provide more detailed weighting information and comprehensive technical documentation to enable researchers to use the data more appropriately.

The primary contributions of this study are as follows. (1) It employs empirical methods to systematically examine the sample structures of six large-scale social surveys and the impact of sample structure deviations on statistical results, revealing methodological pitfalls that offer a new perspective for understanding the contradictory conclusions drawn from different datasets in existing literature. (2) Theoretically, it extends methodological reflection in quantitative research from model specification back to the data-collection stage, broadening scholarly discourse. (3) Practically, it provides empirical guidance for standardizing data usage in quantitative research, thereby enhancing the comparability and robustness of research conclusions.

Keywords: Large-Scale Social Survey, Sample Structure, Sampling Design, Methodological Pitfalls

Authors: Liu Wenbo is Tenure-Track Associate Professor, School of Humanities and Social Sciences, Harbin Engineering University; Zhou Hao (Corresponding Author) is Professor, Center for Sociological Research and Development of China, Department of Sociology, Peking University. Email: zhouh@pku.edu.cn

(责任编辑:陈佳鞠)