



中国人口普查
CHINA POPULATION CENSUS

2010 年人口普查 方法研究

国务院人口普查办公室 编
国家统计局人口和就业统计司



中国统计出版社
China Statistics Press



2010 年人口普查 方法研究

国务院人口普查办公室 编
国家统计局人口和就业统计司

 中国统计出版社
China Statistics Press

图书在版编目(CIP)数据

2010 年人口普查方法研究 / 国务院人口普查办公室,
国家统计局人口和就业统计司编. — 北京: 中国统计出版
社, 2013. 5

ISBN 978—7—5037—6813—2

I. ①2… II. ①国… ②国… III. ①人口普查—调查
方法—研究—中国—2010 IV. ①C924. 25

中国版本图书馆 CIP 数据核字(2013)第 107681 号

2010 年人口普查方法研究

作 者/国务院人口普查办公室 编
国家统计局人口和就业统计司

责任编辑/徐 涛 刘勇利

封面设计/李雪燕

出版发行/中国统计出版社

通信地址/北京市丰台区西三环南路甲 6 号 邮政编码/100073

电 话/邮购(010)63376909 书店(010)68783171

网 址/http://csp. stats. gov. cn

印 刷/河北天普润印刷厂

经 销/新华书店

开 本/890×1240mm 1/16

字 数/778 千字

印 张/27

版 别/2013 年 6 月第 1 版

版 次/2013 年 6 月第 1 次印刷

定 价/128.00 元

版权所有。未经许可,本书的任何部分不得以任何方式在
世界任何地区以任何文字翻印、拷贝、仿制或转载。
如有印装差错,由本社发行部调换。

《2010 年人口普查方法研究》

编委会和编辑部工作人员名单

编 委 会

主 任 张为民

副 主 任 冯乃林

编 委 冯乃林 严建辉

编 辑 部

总 编 辑 冯乃林

副总编辑 武 超 崔红艳 叶礼奇

编 辑 刘勇利 徐 岚 徐 涛 李 睿 陈潇潇

前 言

2010年,在国务院和各级政府的领导下,在有关部门的通力协作及广大人民群众的积极配合下,经过全国近千万普查工作人员的艰苦努力,成功地进行了第六次全国人口普查。第六次全国人口普查与前五次人口普查相比,无论是调查环境、调查项目还是技术手段都发生了很大的变化。这次普查在借助以往成功经验的基础上,大胆创新,首次将境外人口纳入普查对象范围,继续将人口普查与住房调查相结合,并采用了长短表的方法和光电录入、汉字识别等先进的数据处理手段,使中国的人口普查与国际接轨迈进了一大步,不仅获得了极其丰富的人口和住房数据,同时也积累了宝贵的人口普查经验,为今后搞好人口普查和人口抽样调查奠定了坚实的基础。

为了使中国人口普查体系更加科学化、系统化和规范化,在做好人口普查资料开发和利用的同时,各地对人口普查方法进行了总结和研究。截止到2012年3月底,共收到全国各地报来的技术论文160余篇,这些论文涉及到人口普查工作的各个环节,从不同的角度和侧面对人口普查的方法进行了探讨,既有理论的概括又有实际操作的总结。在此基础上,我们于2012年10月在江西南昌召开了第六次全国人口普查技术业务研讨会。各省、自治区、直辖市人口普查办公室的主要领导和业务骨干及相关部门和大专院校的领导和学者应邀出席了会议,并在会上宣读论文,发表自己的意见和建议,受到各界的一致好评。这次研讨会全面系统地总结了我国在社会主义市场经济条件下进行人口普查的经验和教训,初步建立了具有中国特色的人口普查科学方法体系。我们从研讨会的论文中精选出105篇,编辑成册,以飨关注我国人口统计和普查事业的读者。这本书的出版标志着中国人口普查制度进一步完善和规范,这些成果必将在未来的人口普查中发挥重要作用。

借《2010年人口普查方法研究》问世之机,我们向为本书出版付出辛勤劳动的所有同志表示衷心感谢。并希望更多的人口统计工作者、人口学界的专家学者参与对人口普查方法的研究,共同努力,促进中国人口普查方法的不断发展,为不断完善我国人口普查的科学体系作出新的贡献。

编者

2012年11月

目 录

专题一 普查组织、方案设计

人口普查长表抽样方法探讨	(3)
强化部门配合 彰显政府行为	(10)
天津市第六次人口普查技术业务总结	(13)
天津市北辰区第六次人口普查技术总结	(15)
第六次人口普查中存在的问题及建议	(17)
搞好城市(镇)人口普查的思考	(19)
内蒙古自治区第六次人口普查技术业务总结	(23)
沈阳市第六次人口普查技术业务总结	(27)
查准总人口是人口普查的第一要务	(30)
对就业失业指标设置及数据使用的思考	(40)
市场经济条件下开展人口普查的思考	(44)
整合资源 依托中介 多种手段采集数据	(48)
关于普查行政后勤保障工作管理规范化的思考	(52)
关于乡镇一级人口普查组织模式的探讨	(55)
从人口普查看人口抽样调查制度方法改革	(59)
论第六次人口普查方案和方法的得与失	(62)
有关人口普查中若干问题的建议	(67)
广东省第六次全国人口普查业务技术总结	(70)
广州市第六次全国人口普查业务技术总结	(75)
对第六次全国人口普查一些思考与建议	(83)
海南省第六次人口普查技术业务总结	(86)
“六普”工作后主要层面的反思	(94)
认真做好省会城市人口普查工作	(98)
对人口普查工作中存在问题的思考	(106)
甘肃省 2010 年第六次全国人口普查技术业务总结	(108)
新疆第六次全国人口普查业务技术总结	(112)
第六次全国人口普查境外人员普查工作回顾与思考	(117)

专题二 区域划分、摸底

第六次全国人口普查区域划分和地图绘制技术总结	(127)
重新认识摸底的作用及《户主姓名底册》的价值	(131)
从人口地理信息系统建设看区域划分工作	(135)
区域划分、地址编码与制图工作的实践与思考	(139)
晋中市人口普查清查摸底之浅析	(142)
鞍山市人口普查技术业务总结	(146)
夯实基础,步步为营,为全面开展第六次全国人口普查工作做好准备	(154)
人普区域划分和地图绘制工作的组织与实施	(160)
科学划分普查区域	(165)

摸底方法研究与总结	(168)
浅谈人普摸底表的设计和居民身份证号的运用	(175)
浙江省第六次全国人口普查工作中的测绘保障服务	(178)
浙江六人普电子地图绘制工作的回顾和思考	(182)
对人口普查清查摸底质量控制工作几点建议	(185)
对提高清查摸底工作质量的建议和思考	(189)
如何做好人口普查区域划分和地图绘制工作	(196)
浅谈如何做好人口普查摸底工作	(200)
人口普查区域划分与绘图工作的几点体会	(204)
人口普查多阶段摸底的有益尝试	(207)
人口普查城乡属性划分的实践反思	(210)
普查区域划分和普查摸底存在的问题及对策探讨	(213)
对区域划分与绘图的思考	(217)
区域划分与地图绘制之我见	(219)
人口普查摸底方法研究	(222)

专题三 流动人口、入户登记

人口普查的登记原则	(227)
充分利用部门资料 搞准人户分离情况	(230)
突破思维定势 另辟蹊径	(236)
浅谈人口普查中流动人口的登记	(240)
确保外来人口如实申报是提高数据质量的一个重要方面	(243)
第六次人口普查双向登记制度的实践与思考	(246)
人口基础电子台帐在第六次人口普查中的应用	(250)
重构普查流程 查准查实流动人口	(252)
破解人口普查中人户分离人口登记工作难点的实践与思考	(256)
破解外出人口登记难的实践与探索	(259)
对第六次人口普查过程中存在问题的思考	(262)
对流动人口登记的再思考	(270)
对有效解决普查员“入户难、进门难、调查登记难”问题的思考	(273)
提高人口普查入户调查登记质量的方法探讨	(276)
新形势下有效破解普查员“进门难”问题的思考	(279)
关于县(市、区)内跨乡、镇、街道“流动”(人户分离)人口普查登记问题的思考	(282)
“剔重”、“补漏” 着力提高人口普查数据质量	(285)
居民身份证号码列入人口普查项目的思考	(288)
掌握人口流动特点 查准流动人口数量 保证普查登记质量	(291)
掌握流动人口特点 保证普查登记质量	(299)
采取有效措施 查准查实人户分离人口	(303)
如何有效解决普查员“入户难、进门难”的问题	(306)
把握流动人口特点 保证普查登记质量	(308)

专题四 宣传培训、质量控制、数据处理

人口普查 1.2% 的总人口漏报率是如何得到的?	(313)
第六次全国人口普查各阶段质量控制技术总结	(319)
做好人口普查宣传工作的几点体会	(324)
六普热校方法的总结与探讨	(330)

破解第六次全国人口普查难点问题的途径 (336)

第六次人口普查中存在的问题及建议 (340)

对人口普查工作中一些重点难点的把握和建议 (342)

创新思维 借势借力 努力营造人口普查良好社会氛围 (345)

人口普查中的难点及解决办法探讨 (348)

市场经济条件下有效解决普查重点、难点问题的大胆探索与实践 (350)

市场经济条件下特大城市人口普查数据处理工作做法探讨 (354)

人口普查“两员”选聘难原因分析及对策探讨 (357)

破解新时期基层“六人普”宣传动员工作难题之对策思考 (359)

强化政府行为 解决“两员”补贴 (363)

关于普查业务培训重点及难点的思考 (366)

全面数据审核 系统数据评估 (368)

浙江省“六人普”宣传工作创新实践与思考 (370)

在人口普查社会系统工程中进行全员、全程、全方位质量控制工作的探索与实践 (373)

创新模式 强化责任 高质量完成第六次全国人口普查数据处理工作 (376)

齐抓共管 上下一心 着力打造意志坚定、业务精通的普查队伍 (380)

浅谈如何提高人口普查业务培训效果 (384)

采取有效措施防止人口登记不重不漏 (386)

对重庆市总和生育率的评估 (390)

剖析本次人口普查业务的重点和难点问题 (393)

以数据质量求生存 高标准完成人口普查任务 (396)

精心组织 统一部署 认真做好人口普查工作 (402)

浅谈西藏自治区第六次全国人口普查宣传工作 (407)

高质量的人口普查数据源于有效的质量控制 (409)

提高基层人口普查数据质量的思考 (414)

我看普查登记差错的关键环节 (416)

利用部门资料 核实死亡人口 确保普查数据质量 (418)

专题一

普查组织、方案设计

人口普查长表抽样方法探讨

2010年人口普查与2000年相同采用了长短表技术。人口普查表长短表技术是指将人口普查表分为“普查表短表”和“普查表长表”两种。“普查表短表”，其特点是表中项目少，“普查表长表”，是表中项目多。“普查表长表”是抽取一定比例的户填报，“普查表短表”是由其余的户填报。在普查总体中抽取填报长表户，称为普查表长表抽样。人口普查表长表抽样设计与其它全国范围内的抽样设计不同，它不是一项独立的抽样设计，而是人口普查工作中的一部分，具有普查和抽样调查相结合的特点。

新中国成立后，中国相继于1953、1964、1982、1990、2000、2010年进行了六次全国人口普查，前四次对全国所有属于普查对象的人口登记了一种普查表，填报项目相同，项目相对较少。1953年人口普查项目有6个，人口为5.9亿人。1964年项目有9个，人口为6.9亿人。1982年普查项目有19个，人口为10.1亿人，1990年普查项目有21个，人口为11.3亿人。到2000年第五次人口普查时，中国人口已经超过12亿人，普查项目增加，内容更加丰富，普查项目涉及到人口结构、人口经济活动、人口住房等基本情况。为了从普查中尽可能多地获得信息，同时为了节省人力物力，避免普查工作量增加过多，提高普查数据的质量。2000年第五次全国人口普查首次采用了普查表长短表技术。90%户填报普查表短表，短表共有19个项目，其中户记录10项，人记录9项。10%的户填报普查表长表，长表共有49个项目，其中户记录23项，人记录26项。2000年普查结果显示，长表填报数据质量好，代表性强，长短表技术在人口普查中应用取得了较好的效果，这是中国人口普查一项开拓性工作。

根据第六次全国人口普查方案，第六次全国人口普查仍采用长短表技术。90%的户填报普查表短表，第六次全国人口普查短表项目18个，其中户记录6项，人记录12项。10%的户填报普查表长表，长表项目45个，其中户记录17项，人记录28项。

一、2000年普查长表抽样方法简介

2000年普查表长表抽样工作在国务院人口普查办公室的统一部署下，省、自治区、直辖市人口普

查办公室负责培训指导，各乡、镇、街道负责组织实施，广大普查指导员和普查员从画普查区图到编制《户主姓名底册》，采用了科学的抽样方法，抽取了10%的长表调查户，圆满地完成了长表抽样和现场登记工作。

长表代表性确定。毋庸置疑，2000年10%的长表户分布在全国31个省、自治区、直辖市的2800多个县级单位，400多万个普查小区，约有3500多万户，1.2亿人。其样本量大、分布广是全国任何抽样调查都不可比拟的。长表抽样设计对全国和省、地级市都有很好的代表性，同样也考虑了对县级单位（县、区、县级市）的代表性。中国县级单位具有城乡多种经济的综合性和地域宏观性，同时也具有稳定性和规范性。县级单位平均人口规模为40万人，按10%抽样比，县级单位抽取长表调查户的平均样本量在4万人左右。长表主要指标不仅对各省、地市级而且对县级也有较好的代表性。

长表抽样框编制。长表抽样框是由普查小区中的住户（家庭户和集体户）名单组成的《户主姓名底册》。《户主姓名底册》中的户名单，应能涵盖所有符合普查对象的住户，保证住户的不重不漏。同时住户名单应该是最新的和相对稳定的，具备普查员入户登记的可操作性。普查前的大量工作如“画地图、数房子，点人头”，编制《户主姓名底册》，为进行长表抽样做准备。在2000年11月1日普查登记前，普查员将整个普查区域划分成若干个普查区和普查小区，每个普查员负责一个普查小区，在摸清普查小区边界和基本情况的基础上，绘制普查小区图，在地图上标明每栋可能住人的建筑物，给出建筑物内所有住户相应的房屋编号，并将符合普查对象住户所对应的房屋编号转变为“户编号”，“户编号”按001开始连续升序排列。不符合普查对象的户给予900以上的特殊“户编号”。

长表抽样方法。长表调查户抽样采用随机等距的抽样方法。在每个普查小区的《户主姓名底册》“户编号”栏中，随机等距抽取长表调查户。具体抽样步骤是，首先在“户编号”栏的1—10户中，根据《普查表长表调查户随机起点表》查随机起点，与随机起点号对应的“户编号”为第一个长表调查户，然后往下数每隔10户距离抽一户，直到下一户的组距

小于 10 为止。

长表调查组织实施。长表调查户的抽样工作在普查时点 2000 年 11 月 1 日的前两天进行。乡级人口普查办公室负责组织各普查区指导员抽取长表户。为了保证长表抽样框《户主姓名底册》中的户编号不重不漏,普查区指导员在抽取长表调查户前,组织普查员对《户主姓名底册》进行核查,发现问题,及时纠正。长表调查户抽取后,由乡级人口普查办公室审核确认,一经确认后任何人不得随意改动,都必须严格按照抽取的长表户进行长表登记,以防止随意更换样本。

长表抽样误差分析。抽样误差是抽样调查总误差的基本组成部分,是调查总误差的下限,也是对抽样总体进行统计推算的基本依据。对于一项复杂的抽样设计,通过对其调查指标抽样误差的计算分析,可积累大量的信息与经验,为今后人口调查和其它

相关调查的抽样方案设计提供宝贵的参考资料。2000 年第五次全国人口普查计算了长表主要指标的抽样误差,使长表数据更全面地反映实际情况,以供数据使用者参考。

普查表长表指标抽样误差评估范围在省、地级和县级单位。由于中国的县级单位人口规模差异大,用相同的抽样比,样本量差异大,指标的精确度会有所不同。因此,县级单位长表指标抽样误差的估计尤其重要。第五次全国人口普查长表指标计算的抽样误差内容包括了四个方面:一是人口基本情况指标;二是人口经济活动特征指标;三是迁移流动指标;四是住房项目。

长表指标抽样误差估计采用了比率估计方法,计算了比率、标准误、变异系数、设计因子等四个指标。抽样误差的汇总表式及计算结果举例见表 1,分地区按职业大类分的人口比重抽样误差。

表 1 分地区按职业大类分的人口比重抽样误差

地 区	在业 人口	一、国家机关、党群组织、企业、事业单位负责人					二、专业技术人员				
		人数	比重	标准误	变异 系数	设计 因子	人数	比重	标准误	变异 系数	设计 因子
廊坊市	208277	4864	2.335	0.032	1.371	1.020	11454	5.499	0.050	0.921	1.069
安次区	18648	332	1.780	0.097	5.418	1.050	910	4.879	0.155	3.195	1.042
广阳区	18838	869	4.613	0.146	3.171	1.009	2522	13.387	0.246	1.844	1.049
固安县	23980	238	0.993	0.062	6.255	1.022	877	3.657	0.124	3.401	1.082
永清县	22182	259	1.168	0.073	6.281	1.072	808	3.642	0.132	3.649	1.114
香河县	17537	319	1.819	0.101	5.541	1.053	871	4.966	0.163	3.286	1.049
大城县	23674	309	1.305	0.069	5.250	0.979	844	3.565	0.127	3.572	1.114
文安县	23853	759	3.182	0.110	3.452	1.019	920	3.857	0.123	3.201	1.044
大厂回	6465	153	2.367	0.184	7.787	1.028	403	6.233	0.301	4.839	1.057
霸州市	29046	1101	3.791	0.108	2.845	1.014	1632	5.618	0.140	2.498	1.095
三河市	24054	525	2.183	0.092	4.235	1.034	1667	6.930	0.170	2.454	1.095

地 区	六、生产、运输设备操作人员及有关人员					七、不便分类的其他劳动者				
	人数	比重	标准误	变异 系数	设计 因子	人数	比重	标准误	变异 系数	设计 因子
廊坊市	33232	15.955	0.088	0.554	1.161	369	0.177	0.011	6.225	1.261
安次区	2033	10.902	0.263	2.415	1.216	60	0.321	0.061	19.050	1.557
广阳区	3642	19.333	0.309	1.598	1.131	4	0.021	0.010	50.729	1.069
固安县	943	3.932	0.142	3.627	1.198	2	0.008	0.005	62.321	0.929
永清县	1705	7.686	0.212	2.767	1.253	5	0.022	0.013	59.129	1.393
香河县	3337	19.028	0.327	1.722	1.165	6	0.034	0.016	47.665	1.230
大城县	1296	5.474	0.148	2.712	1.058	4	0.016	0.010	61.644	1.299
文安县	4085	17.125	0.271	1.583	1.171	5	0.021	0.008	40.724	0.960
大厂回	1829	28.290	0.576	2.037	1.084	—				
霸州市	7315	25.184	0.281	1.119	1.166	283	0.974	0.065	6.717	1.197
三河市	7047	29.296	0.329	1.123	1.182	—				

注：“—”表示无有效数。

其中:比率估计值(比重): $r = \frac{y}{x}$

比率估计值 r 表示相对率指标。如失业率、文盲率、各职业、行业人口比重。 x 表示相对率的分母,如经济活动人口、15 岁以上人口、职业行业人数。 y 表示相对率分子,如失业人数、未识字人口、各职行业分类人数。

抽样方差:

$$\text{var}(r) = \frac{1-f}{x^2} [\text{var}(y) + r^2 \text{var}(x) - 2r \text{cov}(y, x)]$$

(详细公式略)

标准误: $\text{Se}(r) = \sqrt{\text{Var}(r)}$

变异系数: $cv(r) = \frac{\sqrt{\text{Var}(r)}}{r} \times 100\%$

相对误差: $t \times cv(r) = t \frac{\sqrt{\text{Var}(r)}}{r} \times 100\%$

设计因子: $deff = \sqrt{\text{设计效益 } deff \text{ 值}}$

其中,设计效应:

$$deff = \frac{\text{复杂抽样方差 } \text{var}(r)}{\text{简单随机抽样方差 } \text{var}_{\text{SRS}}(p)}$$

简单随机抽样估计值: $p = \frac{y}{x}$

抽样方差: $\text{var}_{\text{SRS}}(p) = (1-f) \frac{p(1-p)}{x-1}$

简单随机抽样标准误: $\text{Se}(p) = \sqrt{\text{var}_{\text{SRS}}(p)}$

2000 年人口普查长表指标抽样误差以及相对误差,就全国和省级范围来讲很小,推算总体的精确度高,能够满足推算总体的要求,绝大多数指标以及细分组其相对误差能控制在 5% 以内,分析数据时,可以忽略抽样误差的影响。对地级范围总体来讲,所列的长表指标的相对误差,能控制在 10% 以内,县级单位长表指标相对误差绝大多数指标能控制在 15%—20% 以内,长表数据抽样误差达到了设计要求。

二、2010 年长表抽样技术改进

(一)重新定义了长表调查户的抽样总体

《第六次全国人口普查方案》要求,普查表长表抽取 10% 的户填报,普查表短表由其余的户填报。尽管长表的抽样比同 2000 年相同,但 2010 年人口普查采用的登记原则与 2000 年前几次普查采用常住地登记的原则不同。2010 年人口普查采用按现住地登记原则,以“现有人口”加“户籍外出人口”为登记对象,即每个人必须在现住地进行登记。普查

对象不在户口登记地居住的,在户口登记地也要填报普查表。2000 年人口普查是采用常住地登记原则,普查人口为常住人口。常住人口是指户籍在本地,并居住在本地的人口和外来人口在本地已居住了半年以上或离开户籍登记地半年以上的人口。从理论上讲,常住人口的定义是科学、严谨的,要求每个人都在常住地登记。但实际操作难度大,普查员要对每个人判断是否是本地的常住人口,而被调查的外来人口处于各种目的,不如实申报,再加上流动人口多,人户分离情况严重,使普查员现场判断是否常住人口难度增大,容易造成部分人口特别是流动人口的漏登。按现住地登记原则,减少由于普查员判断错误和被填报人不如实申报等原因造成的漏报,有利于查清总人口和流动人口。但这一登记原则,增加了普查员现场登记的工作量,普查员不仅要登记常住本地的人口,还要登记人户分离,外出半年以上的人口。这意味着普查员要将所有类型的户,全部列入 2010 年《户主姓名底册》,即包含全户暂住户,也包含外出半年以上、人户分离的户,寄挂户口的户,而根据 2000 年常住地登记的原则,这些户不属于当地登记的对象。

《户主姓名底册》是普查登记工作中的主要依据和工具,也是抽取长表调查户的抽样框。如果将《户主姓名底册》包含的所有类型的户都列入长表抽样框,将会带来三个问题:一是长表的调查项目填写可操作性差。2010 年长表的调查项目与 2000 年长表项目类同,有 45 项。抽中的外出半年以上的户、寄挂户口户,人户分离的户,由于见面询问困难,影响长表填写的质量。二是减少了样本的有效性。长表中抽取外出半年以上的户、寄挂户口户比例大,会影响推算总体常住人口的特征和结构,产生较大的偏差。三是作为抽样框总体,“现有人口加户籍外出人口”普查口径在总体中有重复,没有保持抽样框一一对应的关系。

基于上述考虑,2010 年人口普查对长表抽样总体重新进行了定义。长表调查户包括居住在本地家庭户和集体户,也包括离开本地不满半年的家庭户。全户外出半年以上的户、因房屋拆迁已离开本地的户、无人居住且没有户口寄挂的空房户,全户死亡户不列入长表抽样框总体,并给出“户编号”600 以上。短表总体是包括了上述所有的户。2000 年长表抽样框总体和短表登记人口总体定义是一致的,而 2010 年这两者的定义是不同的,这种定义是基于实际情况考虑,是在总结 2000 年人口普查难点后,为查准总人口所采用的重要手段。

(二)长表抽样由 2000 年的抽户改成了抽住户组

2000 年长表抽样技术实施所积累的宝贵经验,为 2010 年长表抽样方法的改进提供了依据。同 2000 年相比,2010 年长表抽样设计在考虑长表样本代表性的同时,针对减少长表填报误差,减轻普查员的工作负担,对长表抽样方法做了进一步改进。

2000 年长表抽样在每个普查小区直接抽取 10% 的户,利用了普查摸底阶段所编制的《户主姓名底册》,每隔 10 户抽一户,样本户分散,在样本的代表性方面无可非议。但从现场调查组织情况看,增大了普查员调查工作的难度,长表调查户每隔 10 户有一户,长表户之间距离跨度大,入户登记难,找人难,常需要多次入户,普查员往返于户之间的时间过多,影响了普查员的精力和入户登记的质量。再是不利于现场长表登记质量的监控,也难以杜绝随意换户现象。

不同于 2000 年每隔 10 户抽一户,2010 年长表抽样方法设计采取抽取长表住户组的方法。即在每个普查小区的《户主姓名底册》“户编号”上按 10% 的比例,随机等距抽取长表调查住户组,每个住户组由相邻的 4 户组成。这种抽取住户组方法的优点:首先,能保证样本分布在每个普查小区,又能使被抽中的长表调查户相对集中,节省普查员入户往返的时间,便于普查员再次入户或数据核查。其次,有利于分别选调长短表普查员(边远地区和交通不便的地区除外)。如 1600 户的普查区,抽取 10% 的长表调查户,如果每隔 10 户抽一户,普查员要路途在 160 个工作点停顿,加上多次入户,普查员完成入户登记的难度会大大增加。按照抽住户组的设计,住户分布在普查区内的 40 个相对集中的居住点,普查员在登记期间,路途集中停顿在 40 个工作点,也有易于长短表普查员分别抽调。再是加大调查中随意换户的难度,便于质量监控。

2010 年长表抽样采用抽取住户组尽管在现场操作中有明显的优点,但也存在调查小区抽样比波动较大的问题。抽取住户组同每隔 10 户抽一户不同,如果组距不能整除,舍去或进入仅涉及到 1 户,舍去或进入一个住户组会涉及到 4 户。为防止抽取住户组产生的系统偏差,影响抽样比。2010 年长表抽样对各普查小区抽取住户组的取舍做了统一规定,制定了《第六次全国人口普查长表抽样细则》,《第六次全国人口普查长表调查户抽取方法》,编制了《长表住户抽取参考表 1》(普查小区最大户编号为 20 户及 20 户以上)《长表住户抽取参考表 2》(普

查小区最大户编号为 20 户以下),供基层普查员参考,以保证普查区能按 10% 的抽样比抽取住户。

三、第六次人口普查长表基本数据分析

(一)抽样比

长表抽样细则中,要求按 10% 的抽样比例,在普查小区家庭户和集体户中抽取长表户。表 2 数据显示,在实际操作中,户的抽样比没有达到 10%,为 9.8%,其中家庭户为 9.8%,集体户为 9.1%。分城乡看,城镇抽样比为 9.6%,其中家庭户为 9.7%,集体户为 9.2%。乡村抽样比为 9.9%,其中家庭户为 9.9%,集体户为 8.5%。

数据表明,乡村户的抽样比为 9.9%,最接近设计要求。城镇抽样比为 9.6%,低于设计抽样比 0.4 个百分点。其原因初步分析为,长表调查户包括居住在本地家庭户和集体户,也包括离开本地不满半年的家庭户。当长表抽样抽取了离开本地不满半年的家庭户,如调查员无法填写被调查户长表中的调查项目,就有可能放弃填写长表,或用短表代替,造成长表抽样比低,这在城镇表现的比较突出。相比之下,乡村的调查员对离开本地不满半年的家庭户情况了解比较多,能够完成被调查户长表项目的填写。

同 2000 年相比,2010 年长表抽样比没有明显的变化,2000 年和 2010 年总抽样比均为 9.8%。2000 年家庭户、集体户抽样比分别为 9.9% 和 7.9%,2010 年家庭户、集体户分别为 9.8% 和 9.1%,集体户抽样比明显高于 2000 年 1.2 个百分点,这表明按住户组抽取集体户,能减少集体户的漏登,保证集体户在总体分布更均匀,代表性更强。

(二)家庭户规模

表 2 数据显示,2010 年长表平均每个家庭户人口为 3.08,普查总体为 3.09,二者非常接近。其中城镇长表每个家庭户人口为 2.86,城镇总体为 2.85,乡村长表平均每个家庭户人口为 3.32,乡村总体为 3.34。无论是全国城镇乡村,长表家庭户人口与总体的差异仅在 0.01—0.02 的范围内。而 2000 年,长表家庭户的平均户规模与总体的差异全国为 0.08,其中乡村相差 0.11。实践证明,2010 年长表按住户组抽取,能有效地解决现场调查随意换户,抽选人口少的家庭户填写长表的问题。

表 2 2010 年人口普查全国城镇乡村长表抽样比和平均户规模

单位:户,人,人/户,%

类 型		合 计	短 表	长 表	抽 样 比 例
全 国	户数合计	417722698	376955054	40767644	9.8
	家庭户数	401934196	362606917	39327279	9.8
	集体户数	15788502	14348137	1440365	9.1
	人口合计	1332810869	1205471284	127339585	9.6
	家庭户人口数	1239981250	1118721047	121260203	9.8
	集体户人口数	92829619	86750237	6079382	6.5
	家庭户规模	3.09	3.09	3.08	
	集体户规模	5.88	6.05	4.22	
城 镇	户数合计	221004022	199731596	21272426	9.6
	家庭户数	207189173	187188689	20000484	9.7
	集体户数	13814849	12542907	1271942	9.2
	人口合计	670005546	607472348	62533198	9.3
	家庭户人口数	590123513	532938446	57185067	9.7
	集体户人口数	79882033	74533902	5348131	6.7
	家庭户规模	2.85	2.85	2.86	
	集体户规模	5.78	5.94	4.2	
乡 村	户数合计	196718676	177223458	19495218	9.9
	家庭户数	194745023	175418228	19326795	9.9
	集体户数	1973653	1805230	168423	8.5
	人口合计	662805323	597998936	64806387	9.8
	家庭户人口数	649857737	585782601	64075136	9.9
	集体户人口数	12947586	12216335	731251	5.6
	家庭户规模	3.34	3.34	3.32	
	集体户规模	6.56	6.77	4.34	

(三)年龄构成

表 3 数据显示,长表 0—14 岁,15—64 岁和 65 岁以上的年龄段构成同总体存在差异。分长短表看,0—14 岁比例长表比短表要高 0.6 个百分点,15—64 岁的比例长表比短表要低 1 个百分点,65 岁的比例长表比短表要高 0.4 个百分点。分城镇和乡村看,长表和短表的对比都存在相同特点,长表 0—14

岁低年龄段和 65 岁以上高年龄段比例高,15—64 岁年龄段比例低。

分析其原因,一是长表抽样调查范围未包含一些依法服刑的特殊人口,也没有包含普查时点时,新进入调查小区的人口。二是由于长表调查项目多,更容易受到漏报偏差的影响,特别是 20—24 岁、25—29 岁、30—34 岁的青壮人口很容易漏报,其漏报的比例也高于其他年龄段(见图 1)。

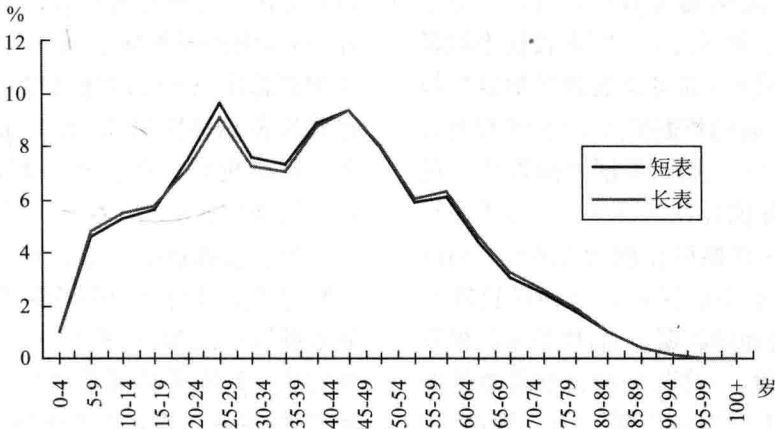


图 1 2010 年长表和短表年龄构成的比较

表 3 2010 年人口普查长表短表年龄结构比较

单位：%

类 别		长表	短表	合计
全国	0—14 岁	17.12	16.55	16.61
	15—64 岁	73.60	74.56	74.47
	65 岁以上	9.29	8.88	8.92
城镇	0—14 岁	14.62	14.02	14.08
	15—64 岁	77.31	78.21	78.12
	65 岁以上	8.07	7.77	7.80
乡村	0—14 岁	19.53	19.12	19.16
	15—64 岁	70.01	70.86	70.78
	65 岁以上	10.46	10.02	10.06

(四)性别构成

表 4 数据显示：2010 年长表的性别比低于短表。长表的性别比为 103.47，短表为 105.05。这说明长表的男性人口明显漏报，尤其在 20—24 岁，25

—29 岁，30—34 岁，城镇男性人口更容易漏报，性别比均低于 100，比短表低了近 4 个百分点。长表数据性别比偏低同上述长表年龄结构漏报的问题是相互交织的。

表 4 2010 年人口普查长表短表性别比

单位：%

类 别	合计	10—14 岁	15—19 岁	20—24 岁	25—29 岁	30—34 岁	35—39 岁	40—44 岁
全国合计	104.90	116.24	108.17	100.95	101.32	104.00	104.78	104.03
城镇	104.92	116.66	106.22	102.93	100.93	103.09	105.28	106.16
乡村	104.87	115.92	110.47	98.52	101.83	105.25	104.17	101.85
长表合计	103.47	116.37	108.50	98.95	98.84	101.44	102.31	101.84
城镇	102.69	116.92	106.08	99.94	97.39	99.73	102.31	103.43
乡村	104.22	115.96	111.15	97.77	100.74	103.76	102.32	100.27
短表合计	105.05	116.23	108.14	101.16	101.57	104.27	105.04	104.26
城镇	105.15	116.63	106.24	103.22	101.29	103.44	105.59	106.44
乡村	104.95	115.91	110.40	98.60	101.94	105.41	104.37	102.02

四、推算总体

用长表数据推算总体，通常有四种方法。方法一用抽样比推算总体。理论上讲，普查表长表抽样采用自行加权的样本设计，即将长表数据乘以抽样比的倒数，也就是将长表的数据扩大 10 倍推算总体中具有某种特征的人口。2010 年设计抽取住户的抽样比为 10%，但实际抽样比为 9.8%。由于住户规模的差异，按人计算抽样比例为 9.6%。同样在总体的各个子总体中不能保证人口抽样比都达到 10%。2000 年长表抽样数据表明，用抽样比推算总体和实际总体相比较，一些长表项目推算总体的误差在 2%—8%，最大误差会达到 19%。因此，用长表指标推算总体及总体的分类指标时，不能简单

地使用长表抽样比推算总体。方法二是比例推算，即用长表各项指标在长表调查总体中结构及比例来推算在整个普查总体中的比例。方法三是区间估计方法。计算长表指标比例，并计算该指标的抽样误差，用长表指标加减抽样误差，用区间估计方法推算总体。方法四是直接对长表数据进行加权，得到长表总体推算值，使读者使用长表资料更便利。目前我们出版普查资料，长表普查资料仍是 10%的抽样数据。

用长表数据推算总体时，采用那种方法，取决于数据使用者研究问题的需要和数据推算的范围。如果推算全国和省、地级区域，可以采用上述第二种方法。如推算县级单位(县、区、县级市)以下总体的某一指标，同时对数据精度要求高，应同时采用几种方法，对其结果进行比较，已取得可靠的推

算结果。

五、长表抽样方法的进一步探讨

第六次人口普查与前几次人口普查相比遇到了更大的挑战和困难,但这并没有影响普查表长表抽样技术的改进。第六次人口普查长表抽样技术采用抽住户组的方法在全国范围内顺利实施,其现场操作优点也得到了实践证明。

分析表明,2010年长表抽样方法明显优于2000年,主要体现在:一是提高了长表平均家庭户规模的精度,全国和分省、区、市的平均家庭户规模同总体的误差明显小于2000年。二是抽取集体户的比例明显高于2000年。

数据显示,2000年和2010年的长表抽样,全国户的抽样比都未达到设计要求的10%。分省看,两次普查全国都只有1/3的省长表户抽样比达到了设计要求10%,其他2/3的省小于10%。其次,虽然2010年集体户抽取比例超过9%,比2000年有了明

显的提高,增加了集体户在总体的分散度,但长表中抽取集体户人口在总体中的比例同2000年相比,误差增加了0.5个百分点。分析其原因,主要是同长表集体户抽样方法规定有关。长表集体户抽样方法规定“对于普查小区内抽中长表调查户为大集体户,即居住在同一个房间的人数超过了10人,由普查员在该户中任选10人登记长表,其余的人登记短表”。再是,长表0—14岁低年龄段和65岁以上高年龄段比例高,15—64岁年龄段比例低,长表数据青年人口比短表更易漏报的问题在2010年的长表抽样中仍未得到很好的解决。另外,2010年普查长表数据汇总还未完成长表主要指标抽样误差的估计。2010年长表指标抽样误差内容同2000年相同,包括了四个方面:一是人口基本情况指标;二是人口经济活动特征指标;三是迁移流动指标;四是住房项目。由于县级单位长表抽样误差的估计,对小区域总体数据推算非常重要,因此,待2010年长表指标抽样误差估计工作完成后,需进一步评估2010年长表数据质量和抽样方法。

参考文献:

- [1] 国务院人口普查办公室. 第五次全国人口普查工作细则. 2000年.
- [2] 国务院人口普查办公室. 第五次全国人口普查普查员手册. 北京: 中国统计出版社, 2000年.
- [3] 国务院人口普查办公室. 2000年第五次全国人口普查主要数据. 北京: 中国统计出版社, 2001年.
- [4] 张为民, 崔红艳. 对中国2000年人口普查准确性的估计 // 国务院人口普查办公室. 第五次全国人口普查科学讨论会论文集. 2004年.
- [5] 胡英. 2000年人口普查长表抽样和误差分析 // 国务院人口普查办公室. 第五次全国人口普查科学讨论会论文集. 2004.
- [6] 徐岚. 人口普查对象的界定. 中国统计, 2008, 9.
- [7] 国务院人口普查办公室. 第六次全国人口普查普查员手册. 北京: 中国统计出版社, 2010年.
- [8] Leslie Kish, Survey Sampling. John Wiley & Sons, Inc.
- [9] Philip M. Gbur, Steven P. Hefter, and Lisa D. Fairchild, Long form design for the U. S census 2000 dress rehearsal and plans for census 2000. U. S. Bureau of the Census.

国务院人口普查办公室 胡英