REPORT ON CHINA LABOR-FORCE DYNAMIC SURVEY (2013)

中国劳动力动态调查：2013年报告

中山大学社会科学调查中心

社会科学文献出版社
SOCIAL SCIENCES ACADEMIC PRESS (CHINA)

中国劳动力动态调查：

2013年报告

REPORT ON CHINA LABOR-FORCE DYNAMIC SURVEY (2013)

梁　宏

社会科学文献出版社
SOCIAL SCIENCES ACADEMIC PRESS (CHINA)

图书在版编目(CIP)数据

中国劳动力动态调查. 2013 年报告/中山大学社会科学调查中心编. —北京：社会科学文献出版社，2013. 12
ISBN 978 - 7 - 5097 - 5275 - 3

Ⅰ. ①中… Ⅱ. ①中… Ⅲ. ①劳动力 - 调查报告 - 中国 - 2013 Ⅳ. ①F249. 21

中国版本图书馆 CIP 数据核字（2013）第 265134 号

中国劳动力动态调查：2013 年报告

编　　者 / 中山大学社会科学调查中心
出 版 人 / 谢寿光
出 版 者 / 社会科学文献出版社
地　　址 / 北京市西城区北三环中路甲 29 号院 3 号楼华龙大厦
邮政编码 / 100029

责任部门 / 社会政法分社（010）59367156　　责任编辑 / 赵慧英　关晶焱
电子信箱 / shekebu@ ssap. cn　　责任校对 / 张俊杰
项目统筹 / 王　绯　　责任印制 / 岳　阳
经　　销 / 社会科学文献出版社市场营销中心（010）59367081　59367089
读者服务 / 读者服务中心（010）59367028

印　　装 / 三河市尚艺印装有限公司
开　　本 / 787mm × 1092mm　1/16　　印　　张 / 15. 25
版　　次 / 2013 年 12 月第 1 版　　字　　数 / 247 千字
印　　次 / 2013 年 12 月第 1 次印刷
书　　号 / ISBN 978 - 7 - 5097 - 5275 - 3
定　　价 / 58. 00 元

“中国劳动力动态调查”项目执行团队（按姓氏笔划排名）：

才国伟、万向东、王进、叶林、连玉君、吴少龙、周欣悦、赵新元、梁玉成、梁宏、蒋廉雄、韩露、蔡禾、潘俊豪

“中国劳动力动态调查”海（境）外学术顾问委员会：

主任委员：郝令昕（美国霍布金斯大学社会学系教授）

委员（按姓氏笔划排名）：刘远立（美国哈佛大学公共卫生学院教授）、边燕杰（美国明尼苏达大学社会学系教授）、吴晓刚（香港科技大学社会科学部教授）、周雪光（美国斯坦福大学社会学系教授）、周敏（美国加州大学社会学和亚裔研究学系教授）、唐文方（美国爱荷华大学政治学系教授）、谢宇（美国密西根大学社会学系教授）、谭康荣（香港中文大学社会学系教授）

“中国劳动力动态调查”校内学术顾问（按姓氏笔划排名）：

马俊（行政管理）、王宁（社会学）、王军（经济学）、丘海雄（社会学）、李伟民（社会心理学）、李若建（人口学）、李新春（工商管理）、何高潮（政治学）、凌莉（公共卫生）、黄葳（教育学）

中国劳动力动态调查”调查合作机构（按名称笔划排名）：

山西医科大学人文社会科学学院、广西师范大学社会工作系、天津理工大学文法学院社会工作系、云南民族大学社会学系、中山大学社会学与人类学学院、中国青年政治学院社会工作学院、内蒙古大学民族学与社会学学院社会学系、长春工业大学人文学院、西北大学社会工作系、西北师范大学社会学系、成都理工大学文法学院、华中科技大学社会学系、华东师范大学社会发展学院、华北电力大学人文社会科学院法政系（保定）、江西财经大学人文学院社会学系、安徽农业大学人文学院社会学系、沈阳工程学院政法系、青海师范大学社会工作系、郑州轻工业学院社会工作系、南京理工大学社会学系、贵州民族学院社会工作系、重庆工商大学社会与公共管理学院、济南大学法学院社会工作系、浙江工商大学社会工作系、黑龙江工程学院社科部、集美大学政法学院、湖南农业大学人文学院、新疆师范大学社会工作系

前　言*

《中国劳动力动态调查：2013 年报告》（以下简称《报告》）是基于中山大学社会科学调查中心完成的 2012 年“中山大学社会科学特色数据库——中国劳动力动态调查”（China Labor-force Dynamic Survey，以下简称 CLDS）全国数据的第一份报告。《报告》的目标在于对 CLDS 所收集的劳动力数据进行描述性分析，为政府、企业界、社会及学界提供针对中国劳动力现状的可靠信息。为了让读者对《报告》和“中国劳动力动态调查”有更为清晰和准确的了解，下文将对 CLDS 的目的、设计、执行、数据等方面进行简要的介绍。

一　调查目的

当代社会科学发展的一个重要特征是数学与计算技术的运用日益普遍，社会科学研究通过数量化、模型化、计算机模拟化，在研究范式上发生了根本变化，社会科学研究成果的“科学性”得到提高，社会科学服务社会的应用价值和研究成果的可操作性越来越明显。而社会科学“科学化”的基础就是将科学知识编码化，此特点在社会科学研究中的体现就是通过收集和运用系统的调查数据阐释具有理论或实践意义的议题。因此，开展大规模、可持续的社会科学调查，建立开放的、可以共享的社会科学数据库，不仅是当今世界社会科学研究普遍采用的组织形式，具有极为重要的学术价值，也是发达国家的政党和政府及时了解民情国情、客观评价社会政治经济、科学分析发展趋势和问题、正确制定国家发展政策的重要手段。

“中国劳动力动态调查”是中山大学三期“985 社会科学特色数据库建设”的专项内容之一，目的是通过对中国城乡以社区为追踪范围的家庭、家

---

* “前言”执笔：梁玉成、倪希。

庭劳动力开展每两年一次的动态追踪调查，系统地监测村/居社区的社会结构和家庭及其劳动力的变化与相互影响，建立劳动力、家庭和社区三个层次上的追踪资料数据库，从而为进行实证导向的高质量的理论研究和政策研究提供基础数据。

选择以劳动力调查为数据库建设目标，是因为：一定数量和质量的劳动力是一个国家经济乃至综合国力得以发展的前提，是引导教育发展速度和教育布局的重要问题；劳动力的空间分布和空间流动是一个国家城市化发展和城市管理最重要的问题；劳动力的收入和财富分配是形塑一个国家社会分化、阶级阶层关系和社会矛盾的重要因素；劳动力的就业水平、居住状况、健康卫生、社会保障、家庭救助等则是一个国家制定社会政策最主要的关注点。总之，在任何一个国家里，劳动力问题都是核心的经济、社会和政治问题。在目前国内大规模综合数据库建设已经开展并基本成熟的情况下，围绕着城市和农村劳动力的家庭、教育、培训、就业、收入、健康、保障、管理、劳资纠纷等问题开展有关劳动力的专门性数据库建设，不仅能凸显数据特色，而且也可聚焦经济学、政治学、社会学、心理学、教育学、人文地理和区域经济学、公共卫生、管理学等学科在同一领域的关注，为形成具有中国特色的社会科学做出贡献。

二　调查抽样

（一）抽样设计

随着统计学多层次模型的建立，强调收集个体数据的同时收集宏观背景数据逐渐成为一个重要的学术要求。在多层次观察的数据能够用来检验宏观社会条件下的微观社会行为的同时，跟踪数据能够提供给我们宏观和微观层面的社会历史信息，因此，多层次观察下的跟踪数据为检验涉及微观到宏观以及宏观到微观转变的理论模型提供了机会。

基于此，“中国劳动力动态调查”的数据收集在以下 3 个层次上开展：第一个层次是 15 ~64 岁劳动力的个体状况（以中国教育制度计算，初中毕业一般为 15 岁，故以 15 岁为劳动年龄起点。由于中国劳动力紧张状况会持续，所以劳动

年龄延长是发展趋势，为了具有前瞻性，劳动年龄上限设定在64岁）；第二层次是包含15~64岁劳动力家庭的基本情况；第三层次是劳动力所在社区（城市社区以居委会辖区为空间，农村社区以村委会辖区为空间）的情况。

一般而言，如果随着时间的发展，调查对象的变化较为缓慢，那么应该选择追踪调查；如果随着时间的变化，调查对象变化较快，甚至发生显著的变化，那么调查总体也会随之发生显著变化，这时就应该使用横截面调查或者轮换样本调查，不断更新调查样本，以便及时反映不断变化的总体。而且这样的变化越快，更新样本的速度也应该越快。

“中国劳动力动态调查”是一项连续性调查，计划每两年开展一次。其设计是以社区为追踪范围，每个社区以及社区中的样本家庭和劳动力连续调查四轮（6年），然后该社区以及社区中的样本家庭和劳动力退出调查，同时一个新的轮换社区样本以及社区中的样本家庭和劳动力将产生并替代退出的轮换样本。在连续调查的四轮期间，如果样本家庭整体迁移出样本社区，将不再跟踪并从样本框中产生新的家庭样本。其操作是，将社区样本总体随机分成4份，按照表1的顺序进行轮换。

表1 “中国劳动力动态调查”社区样本轮换

	1				2			
2012	1	2	3					
2014	1	2	3	4				
2016		2	3	4	1′			
2018			3	4	1′	2′		
2020				4	1′	2′	3′	
2022					1′	2′	3′	4′

选择以社区为追踪范围的轮换样本调查，是基于以下考虑：第一，保证社区层次的观测不会因为家庭和劳动力的空间流动而消失，同时也防止因社区样本的衰老而难以反映处在快速变迁中的中国城乡社区；第二，兼顾横截面调查和追踪调查的特点；第三，调查经费可控。

为了保证样本的全国代表性，劳动力动态调查样本覆盖了除港、澳、台地区和西藏、海南以外的大陆地区。在抽样方法上，采用多阶段、多层次并

与劳动力规模成比例的概率抽样方法（multistage cluster，stratified，PPS sampling）。

分层抽样是把中国劳动力按分层因素进行分层。在中国，劳动力及其家庭户的社会经济地位差异首先来自地区和城乡。因此，在分层操作中，用常住人口规模作为 PPS 的依据，在 PSU（初级抽样单位）的抽取中，同级行政层内以人均 GDP 作为社会经济地位排序的指标（无法获得人均 GDP 指标时用非农人口比例或人口密度作为社会经济地位排序的替代指标）；另外，为了突出反映农村劳动力的城乡流动状况，在社区层次抽样时以非户籍（外来）人口比例作为排序指标。

而在中国大陆范围内，东、中、西部地区差异很大，广东是改革开放的前沿省；同时人口数量与劳动力数量在不同省之间存在较大差异。因此，我们采用东、中、西、广东分区与考虑省人口规模的分层抽样设计。共计有 6 个抽样框：（1）东部人口大省层，（2）东部人口小省层，（3）中部人口大省层，（4）中部人口小省层，（5）西部人口大省层，（6）西部人口小省层。这 6 个层共同构成全国代表性样本。同时，为了使样本对广东的代表性精度上升，我们有广东补充样本：（7）广东非珠三角层，（8）广东珠三角层。这 8 个层最后按照相应入样概率加权得到全国的总样本。为了使样本集中，以及对城乡均具有代表性，我们对每个层内，按照省份/省内分城乡对县区排序后进行抽样。

1. 第一阶段分层抽样：市辖区、县级市、县的抽取

（1）东中西人口大省和人口小省内抽取 PSU

在每个层中，按照省份、城市、县级市、县的顺序，对全部县区按照 GDP 排序，随机起点，依据劳动力规模进行等间距抽取 PSU。

（2）广东省补充样本抽样方案

将广东省首先区分为珠三角和非珠三角，对其中的县区、市辖区[①]和县

① 市辖区是城市的组成部分，为直辖市和地级市划定的行政分区，通常指不包括远郊区、县（县级市）的城市中心区及近郊区。从行政级别上来说，本方案所使用的直辖市和地级市的市辖区都属于县级行政区；从统计口径来说，市辖区的统计结果为定义范围内的多个区域的合计值。

（包括县级市）[①] 抽样框内部分别按照各自的人均 GDP[②]（或非农人口比例）降序排列。按经济水平 GDP（或非农人口比例）分层是为了提高对各种经济水平的市辖区和县（县级市）的代表性。最后以劳动力规模（或常住人口规模[③]）为依据进行 PPS 系统抽样。

2. 第二阶段抽样：村（居）的抽取

第二阶段抽样单位（SSU）为村（居）的行政单位，第一阶段获得的每个市辖区和县（包括县级市）样本的所有村（居）委会组成第二层的抽样框。为了提高经济水平和流动人口的样本代表性，对第二阶段抽样框按照如下方法进行排序。

对于被抽中的城市 PSU 即市辖区，首先，按照抽中市辖区的人均 GDP（或非农人口比例）进行降序排列；[④] 其次，在每个市辖区内将所有居委会按照非户籍（外来）人口比例进行降序排列，由此获得每个市辖区的居委会（SSU）抽样框列表；最后，依据劳动力规模进行 PPS 系统抽样。

对于被抽中的农村 PSU 即县（包括县级市），由于在现有的行政体制框架中，县（包括县级市）与村（居）之间还有一个行政层级，即乡、镇、街道。

---

① 从城乡分布来说，由于市辖区内部的社区全部为居委会，所以，本方案将市辖区视为城市部分；由于县和县级市内部的社区以村委会为主，且为劳动力的主要流出地，所以，本方案将县和县级市视为农村部分。换言之，本方案按照市辖区和县（县级市）的分类将初级抽样单位（PSU）分为城、乡两部分。

② 是否采用人均 GDP 作为排序依据要看资料的可获得性，如果无法得到调查前一年的人均 GDP 指标，本方案将采用非农人口比例作为反映初级抽样单位（PSU）社会经济发展水平的指标。需要说明的是，为了保证标准一致，本方案设定同级抽样框的排序指标相同。

③ 本调查针对的是劳动力，理应以劳动力规模作为 PPS 的依据。因此，本方案将各市辖区、县及社区的劳动力规模作为 PPS 的首选依据。但是，除人口普查资料外，其他官方统计资料不会公布全国各个市辖区、县（县级市）的 15 岁以上人口数量（即劳动力规模）。这种情况下，本方案设计以前一年各市辖区、县（县级市）的常住人口规模为 PPS 的依据。

④ 采用该排序标准的目的在于增加样本在各市辖区的散布度，而不会随机地集中于某些城市市辖区的居委会。在操作过程中，也可省略该排序标准，直接对抽中市辖区的所有居委会按照非户籍（外来）人口比例进行降序排列。究竟采用哪种方法，要根据非户籍（外来）人口比例指标的准确性来确定。具体来说，如果能够获得抽中市辖区所有居委会非户籍（外来）人口比例的准确指标，则可省略市辖区的排序；如果不能获得抽中市辖区所有居委会非户籍（外来）人口比例的准确指标（如概数或者大致排序），则需要按照人均 GDP（或非农人口比例）对抽中的市辖区进行降序排列，然后在各个市辖区内部对所有居委会再按照非户籍（外来）人口比例进行降序排列。

因此，首先以行政级别街道、镇、乡为序排列（默认社会经济地位由高到低）；[①] 其次，在每个街道、镇、乡内部，再依据各自村（居）的非户籍（外来）人口比例分别对其进行降序排列，由此获得每个县（包括县级市）的村（居）（SSU）抽样框列表；最后，以各村（居）的劳动力规模为依据进行 PPS 系统抽样。

特殊情况的处理如下。

（1）将虚拟村（居）委会（指有行政登记但几乎没人居住的工矿、经济开发区或科研机构等）和准村（居）委会（指未经授权且很少有人居住）从抽样框中删除。

（2）常住人口规模小于 120 人的村（居）委会按照左手原则，与邻近常住人口规模小于 120 人的村（居）委会合并，使新的 SSU 的常住人口规模超过 120 人。

（3）常住人口规模超过 10000 人的村（居）委会按照地理分片进行拆分，使每片常住人口规模不少于 4000 人；然后随机从中抽取一片，作为该村（居）委会的代表。

3. 第三阶段抽样：家庭户的抽取

第三阶段（末端）样本（TSU）为家庭户。对所有村（居）样本，采用地图地址法建立末端抽样框。所获得的抽样框为村（居）行政区划排除了空址、商用地址后，有人居住的居住地址列表，并且，按照随机起点的循环等距抽样方式，抽取一个固定大小的样本家庭户地址。另外，在末端抽样框中，对于多址一户或一址多户的特殊情形，一是在制作末端抽样框时尽量有效筛选，二是利用计算机辅助调查系统专门设计的住宅过滤模块和住户过滤模块两种方式处理。需要做的工作包括：准确界定每一个 SSU 抽样单元的行政边界，获取基图或绘制参考底图，绘制调查地图；二是制作住户清单列表，此时需要做的工作包括给每一个村（居）委会的所有住宅建筑物编号，列出所有的住宅和住户信息。需要注意的是：两项工作是同时进行的，绘图员在绘制调查地图的时候，列表员也同时了解每一栋住宅建筑物的具体信息。最后，家庭户中的劳动力（15 岁以上）全部进入个人样本。

① 也可省略此排序标准，直接将所有的村（居）委会按照非户籍（外来）人口比例排序，然后在各县及县级市中依据常住人口规模（或劳动力规模）进行 PPS 系统抽样。

4. 追踪过程中社区的更新

在追踪调查的过程中，必然有些社区会更新（如拆分、重组、取消）。我们的原则是跟踪稳定。如果一个样本社区拆分为二，两个社区都跟。如果一个样本社区与一个非样本社区重组为一个，跟踪这个重组的社区并补抽新部分的家庭户。取消的社区则不在跟踪样本里。

（二）样本规模与分配

1. 样本规模

此调查跟踪社区及社区内家庭所有居家和外出流动的劳动力。根据多阶段随机抽样（multistage cluster sampling）的理论和经验，首先考虑尽量扩大社区数以扩大社区代表性及统计效力（statistical power），我们在经费允许的情况下预测社区数量在400个左右。每个社区内最佳家庭数则由社区层面的开支与家庭层面的开支之比（C/C_1）和解释变项系数的SSU方差（σ^2）决定。$C=3000$，$C_1=700$，$\sigma^2=0.05$，根据Cochran（1977）和Raudenbush（1997），横截面调查的最佳家庭数为：

$$n_{opt}=2\sqrt{\frac{C}{C_1\sigma^2}}=2\sqrt{\frac{3000}{700\times0.05}}\cong19$$

考虑到追踪的损耗（10%）、多变量多层模型分析（为单变量单层模型的1.7倍）的需要，每个社区第一轮抽取的有效家庭样本量为 $n=19\times1.7/0.9=35$ 户。总家庭样本数不超过 $N=400\times35=14000$ 户。

根据随机抽样理论，对上述经费约束下的样本规模进行检验。以描述一个比例数据为例，这个比例的总体概率用0.5估计；假定总体概率的相对误差控制在10%，则 $\delta^2=0.5\times10\%=0.05$；取95%可信度，则 $z=1.96$；多阶段抽样设计效应一般为2~2.5，假定设计效应为2.0。根据简单随机抽样样本量计算公式 $N=deff\frac{z^2p(1-p)}{\delta^2}$，每层需要的样本数为 $N_j=2.0\frac{1.96^2\times0.5(1-0.5)}{0.05^2}=768$。考虑到追踪的损耗（10%）、多变量多层模型分析（为单变量单层模型的1.7倍）的需要，每层需要的样本量扩大至 $N_j=768\times1.7/0.9=1451$。考虑到6层，这样需要的第一轮家庭户样本为 $N=1451\times6=8706$；考虑到追踪损耗，则需

要：8706/0.9 =9673 户。因此，平均每个 SSU 抽 35 个家庭户，一共抽取 14000 个家庭户可以满足统计推断需求。

2. 样本分配

（1）东、中、西部的样本分配：为了使样本对东（不包括广东省）、中、西部地区具有独立的代表性，调查设计将三个地区按照 112∶148∶84 的比例分配 SSUs。①

（2）广东省分配到 50 个 SSUs。

3. 样本在各抽样框中的分布②

此次调查设定平均每个市辖区抽取 4 个居委会，平均每个县（包括县级市）抽取两个村（居）委会，每个村（居）委会抽取 35 个家庭，因此，可以得到表 2——样本在各抽样框中的分布。

表 2　样本在各抽样框中的分布

单位：个

抽样框	城乡分类	初级抽样单位(PSU)	次级抽样单位 村(居)委会(SSU)
广东	城市:市辖区	7	7×4=28
	乡村:县(包括县级市)	16	16×2=32
东部	城市:市辖区	13	13×4=52
	乡村:县(包括县级市)	30	30×2=60
中部	城市:市辖区	13	13×4=52
	乡村:县(包括县级市)	48	48×2=96
西部	城市:市辖区	9	9×4=36
	乡村:县(包括县级市)	24	24×2=48
市辖区总数		42	168
县(包括县级市)总数		118	236
总数		160	404

注：从统计口径来说，市辖区比县高一级。

① 在三个地区合并为全国总样本时，我们会根据 2010 年第六次人口普查的东、中、西部劳动力规模比例给定各自的权数，以保证全国总样本的代表性。

② 由于抽样时无法得到 2010 年全国第六次人口普查资料，因此，样本的城乡分配暂且按照 2000 年第五次人口普查资料设定。在“六普”资料或最近年份统计年鉴资料的基础上，我们会分别计算东、中、西和广东省的市辖区和县（包括县级市）人口规模，并以此作为分配各自城、乡样本的依据。

（三）抽样原则

根据抽样设计，我们在编制抽样程序时，制定了以下抽样原则。

1. 将全国的省份分为东、中、西三个层（东部为北京市、上海市、江苏省、天津市、辽宁省、浙江省、福建省、山东省、广东省；中部为黑龙江省、吉林省、河北省、河南省、山西省、安徽省、江西省、湖北省、湖南省、广西壮族自治区、重庆市、四川省；西部为内蒙古自治区、贵州省、云南省、陕西省、甘肃省、青海省、宁夏回族自治区、新疆维吾尔自治区）。

2. 根据各个省份的人口规模，在每一层内将其分为大省子层和小省子层。大省子层的标准，东部为人口在6000万以上，中部为人口在5000万以上，西部为人口在3000万以上（表3）。

表3 大省层与小省层的分布

	大省层	小省层
东部	江苏省、山东省、广东省 劳动力人口1.86亿人，262个PSU	北京市、上海市、天津市、辽宁省、浙江省、福建省 劳动力人口1.23亿人，210个PSU
中部	黑龙江省、河南省、河北省、四川省、湖南省 劳动力人口2.79亿人，610个PSU	吉林省、山西省、安徽省、江西省、湖北省、广西壮族自治区、重庆市 劳动力人口2.39亿人，508个PSU
西部	内蒙古自治区、甘肃省、青海省、宁夏回族自治区、新疆维吾尔自治区 劳动力人口0.912亿人，297个PSU	贵州省、云南省、陕西省 劳动力人口0.627亿人，309个PSU

3. 由于近年来中国省会城市的“县改区”现象严重，因此，将4个直辖市和各个省会城市的区，以1994年的资料为基础，该年之前的城区设为老城区，之后成为城区的设为新城区。

4. 城市作为一个单独的PSU，与县、县级市均作为PSU。城市抽取4个SSU，县和县级市抽取2个SSU。

5. 修订抽样框（添加新的3个缺失数）。[①]

① 抽样框数据纠错。

6. 4 个 rotation group，每个均需要代表全体，因此，将 2 个县或者 2 个县级市合并成一个虚拟县组，这样构成全部的包含抽取 4 个 SSU 的 PSU。

7. 把四川省放入中部。东中西部劳动力比例为 2.5 亿∶5.2 亿∶1.5 亿，因此东部抽取 32 个 PSU，中部抽取 36 个 PSU，西部抽取 20 个 PSU。

8. 在每个层中，按照省份、城市、县级市、县的顺序，对全部县区按照 GDP 排序，随机起点，按照等间距抽取 PSU。这样获得一个全国性的 PSU 代表样本。

9. 将每一个被抽到的 PSU 随机分为 4 份，构成 4 个 rotation group，每一份中，按照 PPS 的方法，随机抽取一个 SSU。

10. 两年以后新进入的 rotation group，只在 SSU 层次抽取，即 PSU 一旦抽取，则在一定时间内不变。初步计划是每次国家普查之后更新抽样框，并重新抽取 PSU。在不更新 PSU 的时候，新 rotation group 在上次抽取获得的 SSU 所在的街道（城市 PSU）或者乡镇（农村 PSU）中按照 PPS 方法抽取。即重新收集这些上次抽取获得的 SSU 所在的街道（城市 PSU）或者乡镇（农村 PSU）的村（居）委会资料，按照 PPS 方法重新抽取 SSU。

（四）补充说明

1. 广东补充样本：为了保证广东样本具有独立的代表性，在广东抽取 8 个 PSU，其中分布在珠三角的 6，珠三角以外的 2 个。

2. 大城市补充样本：该样本对大城市的反映有所不足。因此决定，如果人数在 0.5 个抽样间距以上的大城市被抽中，则加抽 4 个 SSU。

3. 替代备用方案：如果很难进入县级单位（行政上的难度，自然因素导致的执行过难），则允许采用临近的原则，按照 GDP 排序，抽取临近县；按照原来抽取的 SSU 数量 PPS 抽取。如果很难进入村（居）委会，则从其所属的 rotation group 中 PPS 抽取一个村（居）委会。

（五）抽样结果及样本在各抽样框中的分布[①]

此次调查设定平均每个市辖区抽取 4 个居委会，平均每个县（包括县级

① 在“六普”资料或最近年份统计年鉴资料的基础上，我们会分别计算东、中、西和广东省的市辖区和县（包括县级市）人口规模，并以此作为分配各自城、乡样本的依据。

市）抽取2个村（居）委会，每个村（居）抽取35个家庭，因此，可以得到样本在各抽样框中的分布（表4）。

表4 样本在各抽样框的分布情况

单位：个

	总县区数量	抽取县区数量	村(居)委会数量			
			市辖区	县级市	县	合计
东部小省层	210	26	40	16	16	72
东部大省层	173	17	12	16	12	40
中部小省层	508	30	32	4	40	76
中部大省层	610	31	20	12	40	72
西部大省层	221	15	20	0	20	40
东部小省层	385	18	16	0	28	44
广东省	89	26	28	16	16	60
合 计			168	64	172	404

2012调查年度，执行全部样本的3/4，也就是一共抽取并调查了404×3/4=303个村（居）委会。

三 问卷设计及CAPI

（一）问卷结构

根据调查对象的不同，考虑多个问卷模块：社区发展历史模块、社区现状模块、家庭历史模块、农村家庭现状模块、城市家庭模块、农村成人模块、农村青少年模块、农村外出务工人员模块（由家人作答）、城市流动人口模块、城市劳动力模块。在调查时，根据情况，由合格的被调查者回答相应的问卷。

1. 个体问卷结构

劳动力个体问卷主要是了解每个家庭中每个劳动力个体的情况。符合劳动力个体的条件为：年龄是15~64岁的人，65岁以上并且仍然在工作的人。回

答到第三部分后结束访问。以下是劳动力个体问卷的基本内容。

（1）基本情况

该部分主要是为了了解被访者（劳动力）的基本情况，包括性别、出生地、父母亲受教育程度、政治面貌、户口迁移情况、社会保险等。

（2）教育经历

这部分主要是为了详细了解被访者的教育经历，包括全日制的教育经历和培训经历。全日制教育经历包括每一阶段教育的开始和结束时间、毕业证、学历证、专业、学习等级、所在地区等情况。培训经历主要了解培训的时间、内容、目的和费用情况。该部分还进一步了解了职业资格证书情况和全日制教育结束至工作前的流动情况。

（3）工作状况

在该部分中，工作情况包括有工作和无工作两种情况：有工作主要了解被访者的基本工作情况，包括工作时间、工作收入、具体职业信息等内容。雇员部分主要了解工资形式、合同、福利、管理权、工会、权益受侵犯情况、工作体力要求、工作交往情况。雇主部分主要了解经营的行业、所有权情况、投入资金及渠道、雇员工作时间、工资、加班情况、生意成本等内容。自雇非体力劳动者部分主要了解其工作时间、顾客与服务对象、技能要求等情况。自雇体力劳动者主要了解被访者的工作时间、工作设备和与顾客及政府机构打交道等情况。有工作但不在岗这一部分主要了解被访者不在岗的状态，包括不在岗原因、不在岗是否有工资、回去的原因等情况。无工作且未找工作者主要了解没有找工作的原因、期间生活费来源以及是否打算找工作等情况。求职情况主要了解被访者做了哪些与求职相关的事情、找工作的时间长度、最后一份工作怎么结束的等情况。

（4a）非农工作史

这一部分主要了解被访者的工作单位变换史和换工作史，包括单位的行业、单位类型、单位归属、单位规模等情况。工作史包括各个工作的职业、就业方式、职务、级别等情况。

（4b）农村地区农民工作史

这一部分主要关注农民的职业流动情况，包括外出务工的情况，具体有在

每一个省份的换工作数、职业、每份工作的时长、管理职位、工作单位的性质等情况。

（5）求职与创业过程

这部分主要了解职员的求职过程和雇主（自雇劳动者）的创业过程，包括求职的渠道、收集就业信息的渠道、应聘的具体笔试面试情况、找工作过程中关系的运用等情况。创业过程主要包括生意的来源、生意中运用到的关系等情况。

（6）社会参与与支持

这一部分主要了解被访者的政治参与和获得社会支持的情况，包括投票情况，获得个人社会关系支持、社会或政府组织支持情况，社区信任度情况，也包括流动人口居住地的本地人状况、方言水平和返回家乡的意愿等情况。

（7）劳动者状态

该部分主要了解被访者的民间信仰行为与观念、宗教信仰行为与观念、工作情况的满意度（包括工作收入、工作环境、晋升机会等）、工作价值观、生活满意度、幸福感、信任度、责任感、自评社会地位、公平感、未来工作预期等情况。

（8）生育史

这一部分主要包括孩子的出生时间、性别、母乳喂养等情况。

（9）健康状况

健康状况主要包括被访者的身高、体重、自评身体状况、过去两周患病情况、健康状况对工作与生活的影响、吸烟与饮酒情况、14 岁以前身体状况、职业病与工伤情况等。

2. 家庭问卷结构

家庭问卷的设计主要是为了了解劳动力家庭的基本情况，也即劳动力再生产的初级单位的基本环境。回答家庭问卷的被访者主要是家庭情况的知情者，如果同时有几名知情者，则选择最年轻的家庭情况知情者做家庭问卷。家庭问卷的主要内容如下。

（1）基本情况

主要是了解被选定的受访者在现居住地有血缘关系的家庭人员和符合条件