微观数据开发应用数据说明

开发应用的微观数据范围为：

（一）规模以上工业企业财务状况年度调查微观数据；

（二）住户收支调查微观数据；

（三）2010年第六次全国人口普查微观数据；

（四）2015年1%人口抽样调查微观数据；

（五）第三次全国经济普查微观数据；

（六）第三次全国农业普查微观数据；

（七）2014-2016年企业跟踪调查微观调查数据；

（八）2000年第五次全国人口普查微观调查数据；

（九）2013年住户收支与生活状况调查微观调查数据。

**（一）规模以上工业企业财务状况年度调查微观数据**

规模以上工业企业财务状况调查数据包括2012-2016年北京、上海、浙江三个省区的年度数据。在三个省区的规模以上工业企业中，按10%的比例随机抽取样本企业，并对企业数据进行了脱敏处理。数据包括资产负债类、损益分配类和其他三大类，主要有流动资产合计、固定资产原价、资产总计、流动负债合计、负债合计、实收资本、主营业务收入、主营业务成本、营业利润、利润总额和平均用工人数等32个指标。

**（二）住户收支调查微观数据**

住户收支调查微观数据包括2005年、2008年和2010年城乡住户收支调查微观数据各1万户，共6万户。微观数据均已进行匿名化处理，仅保留省份信息。指标涵盖家庭基本情况、户人均收入情况、户人均消费情况、家庭基本设施等方面，城镇为40个指标，农村为32个指标。其中，农村住户调查为自加权，无权数信息。

**（三）2010年第六次全国人口普查微观数据**

2010年第六次人口普查微观数据库为中国2010年人口普查个体样本数据。2010年人口普查采用了长、短两种普查表，短表包括反映人口基本状况的项目，长表包括所有短表项目和人口的经济活动、婚姻家庭、生育和住房等情况的项目。长表抽取了10%的户填报，短表由其余的户填报。本数据库是对普查表长表数据进行系统抽样获得，抽样比为0.995%，占总人口（不包含现役军人和难以确定常住地的人口）的比重为0.95‰。数据库包括户记录405660条，人记录1267381条。

数据库使用常住人口口径,数据涵盖性别、年龄、民族、户籍情况、受教育程度、行业、职业、迁移流动、社会保障、婚姻、生育、死亡、住房情况等59个相关指标。

为防止个人特征的泄漏，数据库已进行匿名化处理。

**（四）2015年1%人口抽样调查微观数据**

2015年1%人口抽样调查微观数据库为中国2015年1%人口抽样调查个体样本数据。2015年全国1%人口抽样调查以全国为总体，以各地级市（地区、盟、州）为子总体，采取分层、二阶段、概率比例、整群抽样方法，在全国31个省、自治区、直辖市中抽取了2977个县（市、区），33671个乡（镇、街道），85365个村（居）委会的89147个调查小区。共调查登记常住人口2131万人，占全国总人口的1.55%。

本数据库是对2015年1%人口抽样调查数据进行系统抽样获得，占总人口（不包含现役军人和难以确定常住地的人口）的比重为1‰。数据库包括户记录432447条，人记录1371252条。

为便于用户正确使用数据，数据库提供了权数变量（户权数和人权数）。经过加权处理后，各地区的数据已经按全国统一的抽样比换算，可以直接对比。

数据库使用常住人口口径，数据指标涵盖姓名、性别、年龄、民族、户籍情况、受教育程度、行业、职业、迁移流动、社会保障、婚姻、生育、死亡、住房情况等60个相关指标。

为防止个人特征的泄漏，数据库已进行匿名化处理。

**（五） 第三次全国经济普查微观数据**

第三次全国经济普查微观数据是对第三次全国经济普查中的全部企业法人单位抽样取得，不包含金融和铁路系统单位，按照10%的比例进行抽样。为防止个人特征的泄漏，数据库已进行匿名化处理。数据分非工业企业、规模以下工业企业、规模以上工业企业三部分。

**1、非工业企业**

非工业企业的指标包括：行业大类代码、单位所在地行政区划代码、开业时间、从业人员期末人数、登记注册类型、企业控股情况、营业状态、营业收入、主营业务收入、资产总计、R&D人员合计、R&D人员折合全时当量、R&D经费内部支出、R&D项目数。经济类指标单位为千元。其中R&D人员合计、R&D人员折合全时当量、R&D经费内部支出、R&D项目数这四个指标只有纳入一套表调查单位的重点服务业企业才包含。

**2、规模以下工业企业**

规模以下工业企业指年主营业务收入在2000万元以下的工业法人企业。指标包括：行业大类代码、资产总计、实收资本、主营业务收入、从业人员期末人数等。经济类指标单位为元。

**3、规模以上工业企业**

规模以上工业企业指年主营业务收入达到2000万元及以上的工业法人企业。

 **（六）第三次全国农业普查微观数据**

第三次全国农业普查微观调查数据包括农业经营单位普查微观数据、农业经营户普查微观数据、农户普查微观数据、行政村普查数据和乡镇普查微观数据。

**1、农业经营单位普查微观数据**

第三次全国农业普查农业经营单位普查表微观数据公开样本是在第三次全国农业普查登记的全国200多万农业经营单位普查表数据记录中随机抽取的，包含约2万个农业经营单位的4206万笔普查数据。

样本包含了农业经营单位基本情况、聘用从事农林牧渔业和农林牧渔服务业人员情况、拥有或经营的耕地及耕地流转情况、农作物种植情况、确权或经营的林地及林地流转情况、牲畜或家禽饲养情况、确权或经营的牧草地（草场）情况、水产品养殖或捕捞情况、拥有的农用机械情况、农林牧渔生产经营特征情况等方面的指标。

**2、农业经营户普查微观数据**

第三次全国农业普查规模农业经营户普查表微观数据公开样本是在第三次全国农业普查登记的全国近400万规模农业经营户普查表数据记录中随机抽取的，包含约4万户规模农业经营户的3.2亿笔普查数据。

样本包含了规模农业经营户家庭基本情况、住房及生活情况、雇请从事农林牧渔业和农林牧渔服务业人员情况、确权或经营的耕地及耕地流转情况、农作物种植情况、确权或经营的林地及林地流转情况、牲畜或家禽饲养情况、确权或经营的牧草地（草场）情况、水产品养殖或捕捞情况、拥有的农用机械情况、农林牧渔生产经营特征情况等方面的指标。

**3、农户普查微观数据**

第三次全国农业普查中农户普查表微观数据公开样本是在第三次全国农业普查登记的全国约2.3亿农户普查表数据记录中随机抽取的，包含约23万户的9.6亿笔普查数据。

样本包含了农户家庭基本情况、住房及生活情况、确权或经营的耕地情况、农作物种植情况、确权或经营的林地情况、牲畜或家禽饲养情况、确权或经营的牧草地（草场）情况、水产品养殖或捕捞情况、拥有的农用机械情况、生产经营特征情况等方面的指标。

**4、行政村普查微观数据**

行政村普查数据样本是在全国约60万个行政村普查表数据记录中抽取的部分样本，包含约6万条村级单位数据。

本数据样本包含了行政村类型、地形地貌、全国特色景观旅游名村、基础设施配置等基本属性和基础设施的指标。数据涵盖了行政村基本情况、年末人口、社会保障、基本社会服务、土地经营及流转、农田水利、特色种养业、畜禽集中养殖小区、村集体经济组织财务、村干部情况等方面的指标。

**5、乡镇普查微观数据**

乡镇普查微观数据样本是在全国约4万个乡级单位普查表数据记录中抽取的部分样本,包含约4千条乡级单位数据。

本数据样本指标包含乡级类型、乡级属性、地形地貌等基本属性指标，还涵盖了乡级单位基本情况、交通设施、人口、经济、贸易市场、教育文化卫生、生活保障、公共事业等方面的指标。

上述所有微观数据均已进行脱敏处理。

**（七）2014-2016年企业跟踪调查微观调查数据**

1.数据来源：自2014年10月起，国家统计局会同原国家工商总局从工商登记制度改革实施后在工商行政管理部门新登记注册的小微企业和个体经营户中，抽取部分样本单位，按季度开展跟踪调查。本微观数据库数据来源于2014年3季度至2016年4季度的调查数据。

2.企业范围：在全国跟踪调查的单位中，挑选部分连续经营且正常填报报表的单位，并对单位代码进行了脱敏处理。

3.行业分类标准：《国民经济行业分类》（GB/T4754-2011）。

4.指标：包括单位基本情况、经济指标和问卷指标三大类，主要有行业代码、企业个体标识、资产总计、营业收入、从业人员期末人数、本季度的经营情况、本季度用工情况、本季度享受优惠政策情况、最关注的政策措施共9个指标。

**（八）2000年第五次全国人口普查微观调查数据**

2000年人口普查微观数据为中国2000年人口普查个体样本数据。2000年人口普查采用了长、短两种普查表，短表包括反映人口基本状况的项目，长表包括所有短表内容和有关迁移、受教育、经济活动、婚姻家庭、生育和住房等情况的项目。长表抽取了10%的户填报，短表由其余的户填报。本数据库是对普查表长表数据进行等距抽户的方法获得，抽样比为0.95%，（样本人口1180111人，实际登记人口1242612226人）。数据的主要结构指标对总体具有较好的代表性，能够满足大多数研究的需要。

 数据库使用常住人口口径。数据指标涵盖性别、年龄、民族、受教育程度、行业、职业、迁移流动、婚姻、生育、住房情况等。

为防止个人特征的泄漏，数据库已进行匿名化处理。

**（九）2013年住户收支与生活状况调查微观调查数据**

2013年住户收支调查微观数据库为全国住户收支与生活状况调查户级样本数据。全国住户收支与生活状况调查以省为总体，综合采用分层、多阶段、与人口规模大小成比例（PPS）和随机等距抽样相结合的方法抽选村级单位和调查户。全国共抽选约16万个调查户，覆盖约1800个县（区）。

本数据库是对2013年全国住户收支与生活状况调查数据样本户进行系统抽样获得的一个子样本，共包括2万户。每户数据指标共37个，涵盖城乡标识、户权数、户人均可支配收入、户人均消费支出、户耐用品拥有量等。

为防止个人特征的泄漏，数据库已进行匿名化处理，删除了能够辨识到住户的地理信息，仅保留省份信息。