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）2016年农业经营户普查微观调查数据；

（十）2013年住户收支与生活状况调查微观调查数据；

（十一）第四次全国经济普查微观数据；

（十二）2018年全国时间利用调查微观数据库。

**（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行政村普查数据和乡镇普查微观数据。

**1、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。

**2、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）2016年农业经营户普查微观调查数据**

2016年农业经营户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和农户普查微观数据。

**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。

**（十）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。

**（十一）第四次全国经济普查微观数据**

为更好使用第四次全国经济普查微观数据集，特作以下说明：

**1、关于微观数据集的对象**

第四次全国经济普查微观数据集抽取对象为普查的企业法人单位，不包括非企业单位，在企业法人单位中不包括农林牧渔业、金融业和铁路运输业单位。

**2、关于微观数据集的抽取方法**

在第四次全国经济普查数据库中，按照规模以上工业企业法人单位、规模以下工业企业法人单位、除工业外的其他企业法人单位三部分，分别采取10%的比例进行抽取。

**（1）规模以上工业企业法人单位**

在年主营业务收入达到2000万元及以上的工业企业法人单位中抽取。抽样方法是按照行业代码（大类）分层，并按主营业务收入降序排列；层内企业个数小于等于5的不参加抽样，每层前3位企业不参加抽样；按照10%的比例等距抽样后，相邻的两个样本数值型指标相加，作为一个新样本。

**（2）规模以下工业企业法人单位**

在年主营业务收入2000万元以下的工业企业法人单位中抽取。抽样方法是按照行业代码（大类）分层，并按主营业务收入降序排列；层内企业个数小于等于5的不参加抽样，每层前3位企业不参加抽样；按照10%的比例等距抽样后，相邻的两个样本数值型指标相加，作为一个新样本。

**（3）其他企业法人单位**

在除工业外的其他企业法人单位中抽取。抽样方法是按照单位所在地（分省）和行业代码（中类）分层，层内按营业收入排序后按照10%的比例等距抽样，层内企业个数小于5不参加抽样。对从业人员期末人数、营业收入、资产总计、负债合计、科技类以及不同专业类别指标设定了临界值（见表1、表2），指标数值大于临界值时，均用临界值来代替。

**3、关于微观数据集的抽取指标**

根据不同行业特点有重点地抽取指标，共抽取了83个指标，具体如下：

**（1）规模以上工业企业法人单位指标**

共39个，包括：行业代码（前两位）、资产总计、固定资产净额、固定资产原价、累计折旧、流动资产合计、应收账款、存货、产成品、负债合计、流动负债合计、应付账款、所有者权益合计、实收资本、国家资本、集体资本、法人资本、个人资本、港澳台商资本、外商资本、营业收入、营业成本、销售费用、管理费用、财务费用、利息收入、利息支出、投资收益、营业利润、利润总额、亏损企业亏损额、平均用工人数、R&D人员、R&D人员折合全时当量、R&D经费内部支出、机构数、专利申请书、发明专利申请数、新产品销售收入。

**（2）规模以下工业企业法人单位指标**

共5个，包括：行业代码（前两位）、资产总计、营业收入、负债合计、从业人员期末人数。

**（3）其他企业法人单位指标**

共39个，包括：行业代码（前三位）、单位所在地区划代码、开业（成立）年份、从业人员期末人数、登记注册类型、企业控股情况、营业收入、资产总计、负债合计、单位规模等10个通用指标。此外，有资质的建筑业企业还包括建筑业总产值、签订合同额、营业利润。房地产开发经营业企业还包括本年完成投资、商品房销售面积、商品房销售额。限额以上批发和零售业企业还包括流动资产合计、流动负债合计、营业成本、营业利润、利润总额、商品销售额。限额以上住宿和餐饮业企业还包括流动资产合计、流动负债合计、营业成本、营业利润、利润总额、营业额。限额以下批发和零售业企业还包括营业成本、营业利润。限额以下住宿和餐饮业企业还包括营业成本、营业利润。有资质的建筑业企业和规模以上服务业企业科技类指标共7个，包括R&D人员、R&D人员折合全时当量、R&D经费内部支出、机构数、专利申请书、发明专利申请数和新产品销售收入。

**4、关于微观数据集的指标单位**

资产总计、营业收入、营业成本、营业利润、利润总额、负债合计等经济类指标单位为“千元”，从业人员期末人数为“人”，R&D人员为“人”，R&D人员折合全时当量为“人年”，机构数为“个”，专利申请数和发明专利申请数为“件”。

**5、关于微观数据集的有关指标分类和分组**

所有企业行业代码均使用国民经济行业分类标准（GB/T 4754-2017）。行政区划代码前两位对应关系见表3。

开业时间分组：1949年及以前、1950-1977年、1978-1991年、1992-1996年、1997-2001年、2002-2004年、2005-2008年、2009-2013年、2014-2018年、无开业年份。

登记注册类型分组：内资：110国有，120集体，130股份合作，141国有联营，142集体联营，143国有与集体联营，149其他联营，151国有独资公司，159其他有限责任公司，160股份有限公司，171私营独资，172私营合伙，173私营有限责任公司，174私营股份有限公司，190其他。港澳台商投资：210与港澳台商合资经营，220与港澳台商合作经营，230港澳台商独资，240港澳台商投资股份有限公司，290其他港澳台投资。外商投资：310中外合资经营，320中外合作经营，330外资企业，340外商投资股份有限公司，390其他外商投资。

企业控股情况分组：1国有控股，2集体控股，3私人控股，4港澳台商控股，5外商控股，9其他。

单位规模分组：1大型，2中型，3小型，4微型。

|  |  |
| --- | --- |
| **表1** **不同行业企业临界值设定** |  |
| **类别** | **行业代码****前两位** | **从业人员期末人数（人）** | **营业收入****（亿元）** | **资产总计（亿元）** | **负债合计****（亿元）** | **建筑业总产值（亿元）** | **签订合同额（亿元）** | **营业利润（亿元）** | **流动资产合计（亿元）** | **流动负债合计（亿元）** | **营业成本（亿元）** | **利润总额（亿元）** | **商品销售额****（亿元）** | **营业额（亿元）** | **本年完成投资（亿元）** | **商品房销售面积（万平方米）** | **商品房销售额（亿元）** |
| **建筑业** | 47-50 | >1000 | >5 | >5 | >3 | >6 | >10 | >0.3 | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| **批发和零售业** | 51-52 | >200 | >1 | >1 | >1 | -- | -- | >1 | >1 | >1 | >1 | >1 | >1 | -- | -- | -- | -- |
| **住宿和餐饮业** | 61-62 | >200 | >0.5 | >0.5 | >0.5 | -- | -- | >0.5 | >0.5 | >0.5 | >0.5 | >0.5 | -- | >0.5 | -- | -- | -- |
| **房地产开发经营业** | 70 | >200 | >5 | >20 | >20 | -- | -- | -- | -- | -- | -- | -- | -- | -- | >8 | >10 | >10 |
| **服务业** | 05,54-60，63-65,70-90 | >500 | >1 | >5 | >3 | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |

注：房地产业（行业代码前两位70）中的非房地产开发经营企业按照服务业企业处理。

**表2**

**科技类指标临界值**

|  |  |  |
| --- | --- | --- |
| **指标名称** | **单位** | **临界值** |
| R&D人员 | 人 | 1000 |
| R&D人员折合全时当量 | 人年 | 500 |
| R&D经费内部支出 | 亿元 | 3 |
| 机构数 | 个 | 5 |
| 专利申请数 | 件 | 200 |
| 发明专利申请数 | 件 | 100 |
| 新产品销售收入 | 亿元 | 10 |

**表3**

**行政区划代码对应关系**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 11 | 北京 | 12 | 天津 | 13 | 河北 | 14 | 山西 | 15 | 内蒙 |
| 21 | 辽宁 | 22 | 吉林 | 23 | 黑龙江 | 31 | 上海 | 32 | 江苏 |
| 33 | 浙江 | 34 | 安徽 | 35 | 福建 | 36 | 江西 | 37 | 山东 |
| 41 | 河南 | 42 | 湖北 | 43 | 湖南 | 44 | 广东 | 45 | 广西 |
| 46 | 海南 | 50 | 重庆 | 51 | 四川 | 52 | 贵州 | 53 | 云南 |
| 54 | 西藏 | 61 | 陕西 | 62 | 甘肃 | 63 | 青海 | 64 | 宁夏 |
| 65 | 新疆 |  |  |  |  |  |  |  |  |

**（十二）2018年全国时间利用调查微观数据库**

 全国时间利用调查微观数据库是2018年全国时间利用调查个体样本数据，调查对象共48580人。微观数据已进行匿名化处理，以“家庭编码+人码”对个体进行标识，其中家庭编码为20位，已做脱敏处理以确保无法识别调查对象地理位置且每一户编码唯一。指标涵盖城乡、性别、出生年月、民族、受教育程度等基本信息和一天24小时活动时间和活动内容情况等。