中国教育追踪调查(CEPS)

基线数据使用手册

王卫东 郝令昕 张宪

中国人民大学中国调查与数据中心

中国教育追踪调查(CEPS)项目组

2015 年 6 月

目录

一、研究设计

中国教育追踪调查(China Education Panel Survey, 简写为 CEPS)是由中国人民大学中国调查与数据中心设计与实施的、具有全国代表性的大型追踪调查项目，旨在揭示家庭、学校、社区以及宏观社会结构对于个人教育产出的影响，并进一步探究教育产出在个人生命历程中发生作用的过程。

中国教育追踪调查(CEPS)以 2013-2014 学年为基线，以初中一年级（7 年级）和初中三年级（9 年级）两个同期群为调查起点，以人口平均受教育水平和流动人口比例为分层变量从全国随机抽取了 28 个县级单位（县、区、市）作为调查点。调查的执行以学校为基础，在入选的县级单位随机抽取了 112 所学校、438 个班级进行调查，被抽中班级的学生全体入样，基线调查共调查了约 2 万名学生。

中国教育追踪调查(CEPS)以问卷调查为主要手段，对全体被调查学生及其家长或监护人、班主任老师、主课任课老师以及学校负责人进行问卷调查。问卷的内容包括：学生的基本信息、户籍与流动、成长经历、身心健康、亲子互动、在校学习 、课外活动、与老师/同学的关系、社会行为发展、教育期望、以及家庭成员基本信息，家长的基本信息、生活习惯、亲子互动、家庭教育环境、家庭教育投入、社区环境、对学校教育的看法、与老师的互动、对孩子的教育期望，班主任老师对学生行为的评价、与学生家长的互动、对本地与外地户籍学生的比较，包括班主任在内的任课老师的个人基本信息、教育理念、日常教学工作、工作压力与满意度，学校负责人基本信息、教育理念，学校的基本信息、教学设施、招生入学、在校师生情况以及日常教学管理。

中国教育追踪调查(CEPS)同时还对学生进行综合认知能力测试、基本人格测试，并收集学生的重要考试（期中考试、中考、高考等）成绩，并计划组织对学生的健康与体格检查，采集生物医学指标，综合利用各种手段和技术全面采集高质量的数据。

中国教育追踪调查(CEPS)在初中阶段逐年进行追踪调查，计划在学生初中毕业后的第 1 年、第 3 年、第 4 年、第 7 年、第 8 年、第 17 年、第 27 年进行追踪调查，整个调查周期长达 30 年，并计划第 10 年新起一个从 7 年级开始的同期群。

中国教育追踪调查(CEPS)所产出的内容广泛的多层次追踪数据将向学术界全面开放，为教育学、经济学、管理学、社会学等相关学科的研究者提供基础数据资源，同时也服务于教育政策的制定者和学校的管理者，为其决策提供可靠的实证数据依据。

二、抽样设计

（一）多阶段抽样设计

中国教育追踪调查（CEPS）采用多阶段的概率与规模成比例（PPS）的抽样方法，抽样过程分为四个阶段。

表 1　CEPS 抽样设计的四个阶段

抽样阶段	抽样单元
第一阶段(PSU)	在全国县（区）级行政单位中抽取 28 个县（区）
第二阶段(SSU)	在每个入样县（区）所辖地理范围内分别抽取 4 所开设了七年级和/或九年级的学校
第三阶段(TSU)	在每所入样学校中分别抽取 4 个班级，包括 2 个七年级班和 2 个九年级班
第四阶段	入样班级的所有学生、家长、班主任、主科目（语数英）任课教师以及学校领导[1]构成最终调查样本

（二）抽样框设计

1.第一阶段

抽样第一阶段，以县（区）级行政单位作为 PSU。根据 2010 年全国第六次人口普查数据，将全国（31 个省、自治区、直辖市，不含港澳台）共 2870 个有常住人口的县（区）级行政单位形成 3 个抽样框，共抽取 28 个县（区），其中核心样本 15 个县（区），补充样本 13 个县（区）。

抽样框 1：全国所有 2870 个县（区）级行政单位。从中抽取 15 个区（县）作为核心样本。

抽样框 2：上海市所辖 18 个县（区）。为充分反映特大城市上海市的特殊情况，本次调查将上海市所辖 18 个县（区）单独作为一个抽样框，从中抽取 3 个县（区）作为补充样本。

具体而言，在抽样框 2 中，由于上海市多个县（区）拥有大量外来流动人口，为了更好地反映这种特殊性，在抽样时首先从上海市全部 18 个县（区）中抽取

[1] 在计算权数时只考虑学生，而不考虑家长、班主任、主科目（语数英）任课教师以及学校领导。

2 个县（区）作为抽样框 2 的核心样本，再将 18 个县（区）中拥有大量流动人口[2]的 13 个县（区）作为一个子抽样框，从中抽取 1 个县（区）作为抽样框 2 的补充样本。

抽样框 3：全国拥有大量流动人口的 120 个县（区）。为使更多流动儿童、随迁子女进入样本，从而充分反映流动人口的特殊属性对教育过程和教育不平等的影响，本次调查从全国拥有大量流动人口的 120 个县（区）中抽取 10 个县（区）作为补充样本。

表 2 CEPS 的 PSU 抽样框设计

样本类型	PSU 抽样框	抽样框说明	PSU 总体	PSU 样本
核心样本	抽样框 1		2870	15
补充样本	抽样框 2	2a 上海核心样本	18	2
		2b 上海补充样本	13	1
	抽样框 3		120	10

2.第二阶段

抽样第二阶段，以学校作为 SSU。由地方合作单位通过入样县（区）教育部门收集当年（2013-2014 学年）最新统计的学校名单[3]、学校类型[4]和学校规模[5]等基础资料。项目组基于以上资料，构建第二阶段抽样框，采用概率与规模成比例（PPS）的抽样方法，从每个入样县（区）抽取 4 所学校。

特别说明一点，在这一阶段的抽样以及后面的学生个体权数计算过程中，所使用的都是入样县（区）教育部门提供的 2013/2014 年统计数据，而未使用全国初中学校总名单及其统计的各校学生数量，原因在于各地新建学校或撤销现有学校的情况比较普遍，学生人数逐年变化，而能够获得的全国学校最新名单为 2009 年，信息滞后较为严重，无法作为抽样框使用。以抽样框 2 的 4 个 PSU 样本为

[2]拥有大量流动人口的县（区）定义为跨省流入人口超过 19.6 万人或者省内流入人口超过 22.2 万人的县（区）。

[3]本阶段抽样框的学校指的是入样县（区）所辖地理范围内的所有开设七年级/九年级阶段教育的学校，包括三年制或四年制初级中学、开设了三年制或四年制初中部的完全中学以及开设了三年制或四年制初中部的九年一贯制学校、十二年一贯制学校，但不包括职业初中。

[4] 学校类型包括公立学校、私立学校或打工子弟学校，指的是该校初中部的情况，不包括与其同属一校的高中部或小学部。

[5] 学校规模指该校初中部的在校学生数量，不包括与其同属一校的高中部或小学部。

例，将 2009 年全国初中学校总名单所列出的学校与 2013 年入样县（区）教育部门提供的学校名单所列出的学校进行对比，对比结果如表 3 所示。

表 3 以抽样框 2 为例的 2009 年和 2013 年学校名单对比

PSU	2009 年学校总数	2013 年学校总数	2009 和 2013 年都存在的学校数量	2009-2013 年间撤销学校数量	2009-2013 年间新建学校数量
1	22	26	15	7	11
2	47	52	42	5	10
3	30	30	29	1	1
4	31	35	30	1	5

3.第三阶段

抽样第三阶段，以班级作为 TSU，此阶段的抽样工作由地方合作单位在实地调查开始前完成。如果入样学校的被调查年级只有 1 个班或 2 个班，则全部入样；如果该年级有 3 个及以上班级，则使用班级抽样页中的随机数表抽取 2 个班级。

4.第四阶段

完成第三阶段的班级抽样后，被调查班级的所有学生全部入样。

（三）相关变量说明

（1）抽样框

变量名 frame，表示 PSU 样本在哪一个抽样框被抽中，取值为 1 表示来自抽样框 1（全国核心样本），取值为 2 表示来自抽样框 2（2a 上海核心、2b 上海补充样本），取值为 3 表示来自抽样框 3（全国补充样本）。

（2）年级

变量名 grade9，表示 2013-2014 学年学生所在年级，取值为 0 表示七年级，取值为 9 表示九年级。

（3）学期

变量名 fall。基线调查分秋季学期和春季学期两批开展，17 个 PSU 于 2013 年秋季学期完成调查，11 个 PSU 于 2014 年春季完成调查。变量 fall 表示调查进行于哪个学期，取值为 1 表示 2013 年秋季学期，取值为 0 表示 2014 年春季学期。

三、测量工具设计

（一）学生问卷

- 个人基本信息
- 户籍与流动
- 成长经历
- 身心健康
- 亲子互动
- 在校学习
- 课外活动
- 与老师/同学的关系
- 社会行为发展
- 教育期望
- 家庭成员基本信息
-

（二）家长问卷

- 家长的基本信息
- 生活习惯
- 亲子互动
- 家庭教育环境
- 家庭教育投入
- 社区环境
- 对学校教育的看法
- 与老师的互动
- 对孩子的教育期望

-

（三）班主任老师问卷

- 个人基本信息
- 教育理念

- 日常教学工作
- 工作压力与满意度
- 对学生行为的评价
- 与学生家长的互动
- 以及对本地与外地户籍学生的比较

（四）主课（语数外）老师问卷

- 个人基本信息
- 教育理念
- 日常教学工作
- 工作压力与满意度

（五）学校负责人问卷

- 个人基本信息
- 教育理念
- 学校的基本信息
- 教学设施
- 招生入学
- 在校师生情况
- 日常教学管理

（六）认知能力变量

除了调查问卷以外，CEPS 还为七年级和九年级学生分别设计了一套认知能力测试题，该测试题的内容不涉及学校课程所教授的具体识记性知识，而是测量学生的逻辑思维与问题解决能力，并且具有国际可比性、全国标准化的特点。在基线调查的认知能力测试中，七年级试卷共 20 个问题，九年级试卷共 22 个问题，答题限时均为 15 分钟。

测试题共分 3 个维度、11 个构念，包括：

（1）语言：词组类比、语言文字推理

（2）图形：图形规律分析、折纸类题目、几何图形应用

（3）计算与逻辑：数学应用、自定义运算规则、数列应用、抽象规律分析、概率、数值大小逆向思维。

变量 stcog 表示学生在认知能力测试中答对的题目数量，七年级取值为[0,20]，九年级取值为[0,22]。

变量 cog3pl 表示用三参数的 IRT 模型估计出的学生认知能力测试标准化总分。

（七）主课（语数外）期中考试成绩

学生在 2013 年秋季学期的期中考试成绩由被调查学校直接提供，非学生自填。变量 tr_chn/tr_mat/tr_eng 分别表示语文、数学、英语的原始得分。变量 stdchn/stdmat/stdeng 分别表示语文、数学和英语的标准化得分。标准化得分按学校、年级分别计算，并调整为均值=70，标准差=10 的得分。

四、数据结构

CEPS 数据为多层次面板数据。数据文件包括学校数据、班级数据、学生数据和家长数据 4 部分。

其中，学校数据包括学校负责人问卷变量（变量名前缀为 pl）、学校层面派生变量（变量名前缀为 sch）以及县（区）层面派生变量（变量名前缀为 cty）。班级数据包括班主任问卷变量（变量名前缀为 hr）、语文老师/数学老师/英语老师问卷变量（变量名前缀分别为 chn/mat/eng）以及班级层面派生变量（变量名前缀为 cls）。学生数据包括学生问卷变量、认知能力变量、主课（语数外）期中考试成绩和个人/家庭层面派生变量（变量名前缀为 st）。家长数据包括家长问卷变量（变量名前缀为 b）。

不同层次的数据通过 ids 系列变量进行合并匹配。

1.县（区）编号：变量名 ctyids，取值范围 1-28，表示 28 个入样县（区）的唯一 ID。

2.学校编号：变量名 schids，取值范围 1-112，表示 112 所入样学校的唯一 ID。

3.班级编号：变量名 clsids，取值范围 1-438，表示 438 个入样班级的唯一 ID。

4.个人编号：变量名 ids，取值范围 1-19487，表示 19487 名学生及其对应的家长的唯一 ID。

五、学生个体权数计算

（一）抽样框重组子样本及其对应总体

根据第一阶段的抽样框设计（如表 2 所示），3 个抽样框之间部分重合，并非相互独立，因此在计算学生个体权数时，采用如下方式将不同抽样框的 PSU 样本重组为 3 个子样本，重组后的子样本相互独立，并由此计算最终的权数。

子样本 1：对应的总体为全国除上海 18 个县（区）和拥有大量流动人口的 120 个县（区）以外的所有县（区），共计 2745 个。属于此范围的 PSU 样本共 14 个。

子样本 2：对应的总体为上海市所辖 18 个县（区）。属于此范围的 PSU 样本

共 4 个。

子样本 3：对应的总体为全国拥有大量流动人口的非上海市的县（区），共计 107 个。属于此范围的 PSU 样本共 10 个。

表 4　CEPS: JH 抽样框与子样本的 PSU 分布及对应关系

	子样本 1（s1）	子样本 2（s2）	子样本 3（s3）	合计样本	合计对应总体
抽样框 1（f1）	14	0	1	15	2870
抽样框 2a（f2a）	0	2	0	2	18
抽样框 2b（f2b）	0	1	0	1	13
抽样框 3（f3）	0	1	9	10	120
合计样本	14	4	10	28	--
合计对应总体	2745	18	107	--	--

（二）计算权数采用的总体数据

计算各抽样阶段七年级学生和九年级学生入样概率和权数时所使用的总体数据如表 5 所示：

表 5　CEPS: JH 计算入样概率和权数的总体数据来源

总体数据	数据来源
全国 5-19 岁青少年人口总数	2010 年全国第六次人口普查数据
全国初中七年级、九年级在校学生总数	教育部 2012 年教育统计数据[6]
全国初中七年级、九年级在校女学生总数	教育部 2012 年教育统计数据

[6]教育部 2012 年教育统计数据：http://www.moe.edu.cn/publicfiles/business/htmlfiles/moe/s7567/index.html

入样县（区）初中在校学生总数	入样县（区）教育部门 2013-2014 学年统计数据
入样学校七年级、九年级学生总数	首先使用校领导问卷中的自填分年级学生人数，如该信息缺失，则使用班级抽样页[7]中的抽中班级人数均值乘以该年级班级数量作为替代
入样班级学生总数	首先使用班级抽样页中入样班级的实际人数，如该信息缺失，则使用现场记录表[8]中的当日实到学生数作为替代

（三）入样概率及权数的计算

A.按照抽样框设计分别计算入样概率

按照抽样框 1（全国核心样本）、抽样框 2a（上海核心样本）、抽样框 2b（上海补充样本）、抽样框 3（全国补充样本），采取如下步骤，分别计算七年级和九年级的入样概率。

（1）设计抽样概率

使用四阶段的 PPS 抽样方法，在可提前获知各阶段的总体规模的理想状况下，调查样本的入样概率 p 全部相等，可以表示为：

$$p=\frac{n}{N}=\left(\frac{N_j}{N/J}\right)\times\left(\frac{N_{j,k}}{N_j/K}\right)\times\left(\frac{N_{j,k,l}}{N_{j,k}/L}\right)\times\left(\frac{n/(J\cdot K\cdot L)}{N_{j,k,l}}\right) \qquad ①$$

其中，n 和 N 分别为样本规模和总体规模；

J 为抽样第一阶段抽取的 PSU 数量，为常数 28；

N_j 为第 j 个 PSU 的总体规模（学生数量），$j=1,...,J$；

K 为每个入样 PSU 中抽取的 SSU 数量，为常数 4；

[7] 班级抽样页：在第三阶段抽取班级时，由实地调查团队使用带随机数表的班级抽样页现场抽样，其中需要填写被调查年级的班级数量和抽中班级的学生数量。

[8] 现场记录表：实地调查进行过程中由访问员填写，记录被调查班级在调查当日的实到学生数、调查开始和结束时间等信息。

$N_{j,k}$ 为第 j 个入样 PSU 中的第 k 个入样 SSU 的总体规模（学生数量），$k=1,...,K$；

L 为每个入样 SSU 中抽取的 TSU 数量，为常数 2；

$N_{j,k,l}$ 为第 j 个入样 PSU 中的第 k 个入样 SSU 中的第 l 个入样 TSU 的总体规模（学生数量），$l=1,...,L$。

（2）实际抽样概率

在实际抽样过程中，无法提前获知各阶段的总体规模，但完成每阶段的抽样后，可以得到相应阶段入样单元的准确的总体规模。因此，根据实际抽样过程得到的学生个体入样概率 p_l 可以表示为：

$$p_l=\left(\frac{N_j}{N/J}\right)\times\left(\frac{N_{j,k}}{N_j^{'}/K}\right)\times\left(\frac{N_{j,k,l}}{N_{j,k}^{'}/L}\right)\times\left(\frac{n/(J\cdot K\cdot L)}{N_{j,k,l}^{'}}\right) \quad ②$$

其中，$N_j^{'}$、$N_{j,k}^{'}$ 和 $N_{j,k,l}^{'}$ 分别表示第二、三、四阶段抽样完成后得到的相应阶段准确的总体规模，分别用以代替①式中的 N_j、$N_{j,k}$ 和 $N_{j,k,l}$，其他参数含义与①式相同。

（3）根据应答率进行事后调整

由于存在学生当天未到校、无能力填答或拒答问卷等情况，在计算学生个体权数时，根据每个班级的学生问卷应答率对入样概率进行事后调整，调整后的入样概率 rp_l 可表示为：

$$rp_l=\frac{\tilde{N}_{j,k,l}}{N_{j,k,l}^{'}}\cdot p_l$$

③

其中，$\tilde{N}_{j,k,l}$ 为调查班级实际填答问卷学生数。

由于每个调查班级中的所有学生在此处的个体权数相同，因此通过

$rw_l = \dfrac{1}{rp_l}$，将概率取倒数，得到调查班级的第 i 个学生的权数 $rw_i = rw_l$。

（4）根据人口构成比例进行事后调整（事后分层）

根据教育部 2012 年教育统计数据分年级的女学生总数计算出相应的比例，并进行事后调整。以女学生为例，调整后的学生个体权数 crw_i 可表示为：

$$crw_i = \left(\frac{N_{girl}}{\sum_i mrw_i} \right) \cdot mrw_i$$

⑤

B.按照重组后的子样本计算最终权数

按照抽样阶段和抽样框设计分别计算七年级和九年级学生入样概率和权数并进行事后调整后，再按照重组后的三个子样本分别计算并调整七年级和九年级学生最终的权数。

（1）子样本 1（s1）最终权数

子样本 1 的所有 PSU 样本均来自抽样框 1，因此对于子样本 1，最终的学生个体权数即等于根据抽样框设计进行计算和事后调整后的权数。最终的学生个体权数 wgt_i 可表示为：

$$wgt_i = crw_i(s1f1)$$

（2）子样本 2（s2）最终权数

子样本 2 的 4 个 PSU 样本中，有 2 个属于抽样框 2a（上海核心样本，f2a），1 个属于抽样框 2b（上海补充样本，f2b），1 个属于抽样框 3（全国补充样本，f3）。因此，对于子样本 2，需要按照各自所占比例分别计算上述三部分样本最终的学生个体权数。

对于属于抽样框 2a（上海核心样本，f2a）的 2 个 PSU 样本，最终的学生个体权数 wgt_i 可表示为：

$$wgt_i = crw_i(s2f2a)\cdot\left(\frac{\sum_{i\in s2} crw_i(s2f2a)}{\sum_{i\in s2} crw_i(s2f2a)+\sum_{i\in s2} crw_i(s2f2b)+\sum_{i\in s2} crw_i(s2f3)}\right)$$

对于属于抽样框 2b（上海补充样本，f2b）的 1 个 PSU 样本，最终的学生个体权数 wgt_i 可表示为：

$$wgt_i = crw_i(s2f2b)\cdot\left(\frac{\sum_{i\in s2} crw_i(s2f2b)}{\sum_{i\in s2} crw_i(s2f2a)+\sum_{i\in s2} crw_i(s2f2b)+\sum_{i\in s2} crw_i(s2f3)}\right)$$

对于属于抽样框 3（全国补充样本，f3）的 1 个 PSU 样本，最终的学生个体权数 wgt_i 可表示为：

$$wgt_i = crw_i(s2f3)\cdot\left(\frac{\sum_{i\in s2} crw_i(s2f3)}{\sum_{i\in s2} crw_i(s2f2a)+\sum_{i\in s2} crw_i(s2f2b)+\sum_{i\in s2} crw_i(s2f3)}\right)$$

（3）子样本 3（s3）最终权数

子样本 3 的 10 个 PSU 样本中，有 1 个属于抽样框 1（全国核心样本，f1），9 个属于抽样框 3（全国补充样本，f3）。因此，对于子样本 3，需要按照各自所占比例分别计算上述两部分样本最终的学生个体权数。

对于属于抽样框 1（全国核心样本，f1）的 1 个 PSU 样本，最终的学生个体权数 wgt_i 可表示为：

$$wgt_i = crw_i(s3f1)\cdot\left(\frac{\sum_{i\in s3} crw_i(s3f1)}{\sum_{i\in s3} crw_i(s3f1)+\sum_{i\in s3} crw_i(s3f3)}\right)$$

对于属于抽样框 3（全国补充样本，f3）的 9 个 PSU 样本，最终的学生个体权数 wgt_i 可表示为：

$$wgt_i = crw_i(s3f3) \cdot \left(\frac{\sum_{i \in s3} crw_i(s3f3)}{\sum_{i \in s3} crw_i(s3f1) + \sum_{i \in s3} crw_i(s3f3)} \right)$$

（四）相关变量说明

1.子样本

变量名 subsample，表示 PSU 样本所属的子样本，取值为 1 表示属于子样本 1，取值为 2 表示属于子样本 2，取值为 3 表示属于子样本 3。

2.权数

变量名 sweight，表示学生个体样本对应的学生个体权数。

六、派生变量说明

派生变量分为四个层面，分别是：县区（PSU）层面、学校层面、班级层面和个人/家庭层面。其中县区（PSU）层面的派生变量共 5 个，变量名为 cty*，位于学校数据库中；学校层面共 13 个，变量名为 sch*，位于学校数据库中；班级层面共 16 个，变量名为 cls*，位于班级数据库中；个人/家庭层面共 19 个，变量名为 st*，位于学生个人数据库中。所有层面的派生变量均排列在相应数据库的最后面。以下将对四个层面的派生变量分别作简要说明并列出频数分布表。

（一）县区（PSU）层面派生变量

县区（PSU）层面的派生变量反映入样 PSU 在人均受教育水平、行政区划、地理位置和样本量等方面的性质，共有 5 个，其变量名和变量标签分别为：

变量名	变量标签
ctyedu	PSU 平均受教育年限
ctyeduz	PSU 平均受教育年限的 z 值
ctyeduc	PSU 平均受教育年限分类
ctytype	PSU 所属行政级别
ctyplace	PSU 所属地区

1.ctyedu

数据来自第 6 次全国人口普查，入样的 28 个 PSU 的平均受教育年限（单位：年），均值为 9.54 年，最小值为 6.8 年，最大值为 12.19 年，PSU 层面的频数分布如下表：

ctyedu　PSU 平均受教育年限

	频数	百分比
6.800	1	3.570
7.560	1	3.570
7.810	1	3.570
8.030	1	3.570
8.200	1	3.570
8.330	1	3.570
8.460	1	3.570
8.580	1	3.570
8.620	1	3.570
8.690	1	3.570
8.860	1	3.570
9.070	1	3.570
9.190	1	3.570
9.390	1	3.570
9.690	1	3.570
9.840	1	3.570
9.920	1	3.570
10.13	1	3.570
10.26	1	3.570
10.47	1	3.570
10.67	1	3.570
10.86	1	3.570
11.01	1	3.570
11.16	1	3.570
11.51	1	3.570
11.66	1	3.570
12.13	1	3.570
12.19	1	3.570
Total	28	100

对应的学生个体层面频数分布如下表：

ctyedu　PSU 平均受教育年限

	频数	百分比
6.800	991	5.090
7.560	612	3.140
7.810	904	4.640
8.030	845	4.340
8.200	708	3.630
8.330	816	4.190
8.460	828	4.250
8.580	570	2.930
8.620	755	3.870
8.690	670	3.440
8.860	714	3.660
9.070	861	4.420
9.190	508	2.610
9.390	778	3.990
9.690	601	3.080
9.840	606	3.110
9.920	680	3.490
10.13	523	2.680
10.26	612	3.140
10.47	724	3.720
10.67	792	4.060
10.86	657	3.370
11.01	467	2.400
11.16	418	2.150
11.51	977	5.010
11.66	752	3.860
12.13	665	3.410
12.19	453	2.320
Total	19,487	100

2.ctyeduz

数据来自第 6 次人口普查，通过全国所有县级行政单位的平均受教育年限均值和标准差计算 28 个入样 PSU 平均受教育年限的 z 值。PSU 层面的频数分布如下表：

ctyeduz PSU 平均受教育年限的 z 值

	频数	百分比

-1.305	1	3.570
-0.787	1	3.570
-0.616	1	3.570
-0.466	1	3.570
-0.351	1	3.570
-0.262	1	3.570
-0.173	1	3.570
-0.0916	1	3.570
-0.0643	1	3.570
-0.0166	1	3.570
0.0993	1	3.570
0.242	1	3.570
0.324	1	3.570
0.461	1	3.570
0.665	1	3.570
0.767	1	3.570
0.822	1	3.570
0.965	1	3.570
1.054	1	3.570
1.197	1	3.570
1.333	1	3.570
1.463	1	3.570
1.565	1	3.570
1.667	1	3.570
1.906	1	3.570
2.008	1	3.570
2.328	1	3.570
2.369	1	3.570
Total	28	100

对应的学生个体层面的频数分布如下表：

ctyeduz　PSU 平均受教育年限的 z 值

	频数	百分比
-1.305	991	5.090
-0.787	612	3.140
-0.616	904	4.640
-0.466	845	4.340
-0.351	708	3.630

-0.262	816	4.190
-0.173	828	4.250
-0.0916	570	2.930
-0.0643	755	3.870
-0.0166	670	3.440
0.0993	714	3.660
0.242	861	4.420
0.324	508	2.610
0.461	778	3.990
0.665	601	3.080
0.767	606	3.110
0.822	680	3.490
0.965	523	2.680
1.054	612	3.140
1.197	724	3.720
1.333	792	4.060
1.463	657	3.370
1.565	467	2.400
1.667	418	2.150
1.906	977	5.010
2.008	752	3.860
2.328	665	3.410
2.369	453	2.320
Total	19,487	100

3.ctyeduc

根据变量 ctyeduz 将入样的 28 个 PSU 按照平均受教育年限划分为 3 类：低于平均水平、高于平均水平 1 个标准差内、高于平均水平 1 个标准差以上，PSU 层面的频数分布如下表：

ctyeduc　PSU 平均受教育年限分类

	频数	百分比
1 低于平均水平	10	35.71
2 高于平均水平 1 个标准差内	8	28.57
3 高于平均水平 1 个标准差以上	10	35.71
Total	28	100

对应的学生个体频数分布如下表：

ctyeduc PSU 平均受教育年限分类

	频数	百分比
1 低于平均水平	7,699	39.51
2 高于平均水平 1 个标准差内	5,271	27.05
3 高于平均水平 1 个标准差以上	6,517	33.44
Total	19,487	100

4.ctytype

根据入样 PSU 在 2010 年所属行政区划的级别分为 4 类，分别是：直辖市（包括北京、天津、上海）、省会城市市区、地级市市区和县/县级市，PSU 层面的频数分布如下表：

ctytype PSU 所属行政级别

	频数	百分比
1 直辖市	6	21.43
2 省会城市市区	5	17.86
3 地级市市区	5	17.86
4 县、县级市	12	42.86
Total	28	100

对应的学生个体频数分布如下表：

ctytype PSU 所属行政级别

	频数	百分比
1 直辖市	3,684	18.90
2 省会城市市区	3,815	19.58
3 地级市市区	2,866	14.71
4 县、县级市	9,122	46.81
Total	19,487	100

5.ctyplace

根据入样 PSU 所属地区的地理位置分为东、中、西部三类，PSU 层面的频数分布如下表：

ctyplace PSU 所属地区

	频数	百分比
1 东部地区	17	60.71
2 中部地区	5	17.86
3 西部地区	6	21.43

Total	28	100

其对应的学生个体频数分布如下表：

ctyplace　PSU 所属地区

	频数	百分比
1 东部地区	9,997	51.30
2 中部地区	4,073	20.90
3 西部地区	5,417	27.80
Total	19,487	100

（二）学校层面派生变量

学校层面的派生变量反映入样学校在学生户口类型、学校性质、县区内排名、学校所在地等方面的性质，共有 13 个，其变量名和变量标签分别为：

变量名	变量标签
schhkplace	学校本县（区）户口学生所占比例
schhkplace_4c	学校本县（区）户口学生比例分类
schhktype	学校农业户口学生所占比例
schhktype_4c	学校农业户口学生比例分类
schstlb	学校父母不全在家的学生比例
schstlb_4c	学校父母不全在家的学生比例分类
schtype_4c	学校性质（四分类）
schtype_2c	学校性质（二分类公立/民办）
schrank_5c	学校在本县（区）排名（五分类）
schrank_3c	学校在本县（区）排名（三分类）
schbrd	学校是否提供学生宿舍
schloc_5c	学校所在的地区类型（五分类）
schloc_3c	学校所在的地区类型（三分类）

1.schhkplace

根据学生和家长在问卷中填写的学生户口登记地，计算出每个入样学校所有被调查学生中，目前户口登记地在本县（区）的学生所占比例，学校层面的均值为 0.79，标准差为 0.21，中位数为 0.87，最小值为 0.05，最大值为 1，其分组后的频数分布表见变量 schhkplace_4c 的频数分布表。

2.schhkplace_4c

将变量 schhkplace 的取值由低到高分为四组，表示入样学校所有被调查学生

中户口登记地在本县（区）的比例范围，学校层面的频数分布如下表：

schhkplace_4c 学校本县（区）户口学生比例分类

	频数	百分比
1 低于 50%	10	8.930
2 50%到 70%	23	20.54
3 70%到 90%	31	27.68
4 高于 90%	48	42.86
Total	112	100

对应的学生个体频数分布如下表：

schhkplace_4c 学校本县（区）户口学生比例分类

	频数	百分比
1 低于 50%	1,376	7.060
2 50%到 70%	3,251	16.68
3 70%到 90%	5,474	28.09
4 高于 90%	9,386	48.17
Total	19,487	100

3.schhktype

根据学生和家长在问卷中填写的学生户口类型，将“非农户口”和“居民户口”合并为“非农户口”，计算出每个入样学校所有被调查学生中，目前户口类型为农业户口的学生所占比例，学校层面的均值为 0.54，标准差为 0.28，中位数为 0.61，最小值为 0.01，最大值为 0.92，其分组后的频数分布表见变量 schhktype_4c 的频数分布表。

4.schhktype_4c

将变量 schhktype 的取值由低到高分为四组，表示入样学校所有被调查学生中户口类型为农业户口的比例范围，学校层面的频数分布如下表：

schhktype_4c 学校农业户口学生比例分类

	频数	百分比
1 低于 25%	25	22.32
2 25%到 60%	30	26.79
3 60%到 80%	25	22.32
4 高于 80%	32	28.57
Total	112	100

对应的学生个体层面的频数分布如下表：

schhktype_4c　学校农业户口学生比例分类

	频数	百分比
1 低于 25%	4,103	21.06
2 25%到 60%	5,330	27.35
3 60%到 80%	4,147	21.28
4 高于 80%	5,907	30.31
Total	19,487	100

5.schstlb

根据学生和家长在问卷中填写的学生父亲和母亲分别是否在家与学生同住，计算出每个入样学校所有被调查学生中，父母一方或双方长期不在家与子女共同生活的比例，学校层面的均值为 0.22，标准差为 0.16，中位数为 0.18，最小值为 0，最大值为 0.75，其分组后的频数分布表见变量 schstlb_4c 的频数分布表。

6.schstlb_4c

将变量 schstlb 的取值由低到高分为四组，表示入样学校所有被调查学生中父母不全在家的学生的比例范围，学校层面的频数分布如下表：

schstlb_4c　学校父母不全在家的学生比例分类

	频数	百分比
1 低于 10%	20	17.86
2 10%到 30%	67	59.82
3 30%到 50%	17	15.18
4 高于 50%	8	7.140
Total	112	100

对应的学生个体层面的频数分布如下表：

schstlb_4c　学校父母不全在家的学生比例分类

	频数	百分比
1 低于 10%	3,288	16.87
2 10%到 30%	11,231	57.63
3 30%到 50%	3,471	17.81
4 高于 50%	1,497	7.680
Total	19,487	100

7.schtype_4c

根据校领导问卷 A1 题，将学校性质分为四类：公立学校、民办公助、普通民办学校和民办打工子弟学校，学校层面的频数分布如下表：

schtype_4c 学校性质（四分类）

	频数	百分比
1 公立学校	104	92.86
2 民办公助	1	0.890
3 普通民办学校	5	4.460
4 民办打工子弟学校	2	1.790
Total	112	100

对应的学生个体的频数分布如下表：

schtype_4c 学校性质（四分类）

	频数	百分比
1 公立学校	18,043	92.59
2 民办公助	110	0.560
3 普通民办学校	1,049	5.380
4 民办打工子弟学校	285	1.460
Total	19,487	100

8.schtype_2c

将变量 schtype_4c 的四分类学校性质合并为两大类：公立学校和民办学校，其中民办学校包括变量 schtype_2c 的民办公助、普通民办学校和民办打工子弟学校，学校层面的频数分布如下表：

schtype_2c 学校性质（二分类公立/民办）

	频数	百分比
1 公立学校	104	92.86
2 民办学校	8	7.140
Total	112	100

对应的学生个体的频数分布如下表：

schtype_2c 学校性质（二分类公立/民办）

	频数	百分比
1 公立学校	18,043	92.59
2 民办学校	1,444	7.410
Total	19,487	100

9.schrank_5c

根据校领导问卷 A4 题，将学校初中部在本县（区）的排名分为五类：最差、中下、中间、中上和最好，学校层面的频数分布如下表：

schrank_5c　学校在本县（区）排名（五分类）

	频数	百分比
1 最差	1	0.890
2 中下	8	7.140
3 中间	15	13.39
4 中上	67	59.82
5 最好	21	18.75
Total	112	100

其对应的学生个体的频数分布如下表：

schrank_5c　学校在本县（区）排名（五分类）

	频数	百分比
1 最差	171	0.880
2 中下	1,341	6.880
3 中间	2,225	11.42
4 中上	11,366	58.33
5 最好	4,384	22.50
Total	19,487	100

10.schrank_3c

将变量 schrank_5c 的五分类合并为三分类：中等及以下、中上、最好，其中，“中等及以下”包括变量 schrank_5c 的最差、中下、中间三个取值，学校层面的频数分布如下表：

schrank_3c　学校在本县（区）排名（三分类）

	频数	百分比
1 中等及以下	24	21.43
2 中上	67	59.82
3 最好	21	18.75
Total	112	100

对应的学生个体的频数分布如下表：

schrank_3c　学校在本县（区）排名（三分类）

	频数	百分比
1 中等及以下	3,737	19.18
2 中上	11,366	58.33
3 最好	4,384	22.50
Total	19,487	100

11.schbrd

根据校领导问卷 A10 题，该变量指示学校是否提供学生宿舍以及学生住宿比例，共有 3 个取值，学校层面的频数分布如下表：

schbrd 学校是否提供学生宿舍

	频数	百分比
1 是,全部学生寄宿	19	16.96
2 是,部分学生寄宿	42	37.50
3 否	51	45.54
Total	112	100

对应的学生个体的频数分布如下表：

schbrd 学校是否提供学生宿舍

	频数	百分比
1 是,全部学生寄宿	3,824	19.62
2 是,部分学生寄宿	7,621	39.11
3 否	8,042	41.27
Total	19,487	100

12.schloc_5c

根据校领导问卷 A23 题，该变量指示学校所在的地区类型，共有 5 个取值，学校层面的频数分布如下表：

schloc_5c 学校所在的地区类型（五分类）

	频数	百分比
1 市/县城的中心城区	42	37.50
2 市/县城的边缘城区	12	10.71
3 市/县城的城乡结合部	17	15.18
4 市/县城区外的镇	20	17.86
5 农村	21	18.75
Total	112	100

对应的学生个体的频数分布如下表：

schloc_5c　学校所在的地区类型（五分类）

	频数	百分比
1 市/县城的中心城区	7,481	38.39
2 市/县城的边缘城区	2,472	12.69
3 市/县城的城乡结合部	2,439	12.52
4 市/县城区外的镇	3,576	18.35
5 农村	3,519	18.06
Total	19,487	100

13.schloc_3c

将变量 schloc_5c 的五分类合并为三分类：市/县城的中心城区、边缘城区及城乡结合部、乡镇农村，其中，“边缘城区及城乡结合部”包括变量 schloc_5c 的“市/县城的边缘城区”和“市/县城的城乡结合部”两个取值，“乡镇农村”包括变量 schloc_5c 的“市/县城区外的镇”和“农村”两个取值，学校层面的频数分布如下表：

schloc_3c　学校所在的地区类型（三分类）

	频数	百分比
1 市/县城的中心城区	42	37.50
2 边缘城区及城乡结合部	29	25.89
3 乡镇农村	41	36.61
Total	112	100

对应的学生个体的频数分布如下表：

schloc_3c　学校所在的地区类型（三分类）

	频数	百分比
1 市/县城的中心城区	7,481	38.39
2 边缘城区及城乡结合部	4,911	25.20
3 乡镇农村	7,095	36.41
Total	19,487	100

（三）班级层面派生变量

班级层面的派生变量反映入样班级在户口所在地、学生户口类型、父母不在家比例、年级排名、班主任个人情况、班级规模等方面的性质，共有 16 个，其变量名和变量标签分别为：

变量名	变量标签

clshkplace	班级本县（区）户口学生比例
clshkplace_4c	班级本县（区）户口学生比例分类
clshktype	班级农业户口学生比例
clshktype_4c	班级农业户口学生比例分类
clsstlb	班级父母不全在家的学生比例
clsstlb_4c	班级父母不全在家的学生比例分类
clsrank_5c	班级目前在全年级的排名（五分类）
clsrank_3c	班级目前在全年级的排名（三分类）
clshrsex	班主任性别
clshrbyr	班主任出生年份
clshrage	班主任年龄
clshrage_4c	班主任年龄分组
clshrexp	班主任教龄
clshrexp_3c	班主任教龄分组
clsn	班级学生人数
clsn_5c	班级学生人数分类

1.clshkplace

根据学生和家长在问卷中填写的学生户口登记地，计算出每个入样班级所有被调查学生中，目前户口登记地在本县（区）的学生所占比例，均值为 0.80，标准差为 0.22，中位数为 0.88，最小值为 0，最大值为 1，其分组后的频数分布表见变量 clshkplace_4c 的频数分布表。

2.clshkplace_4c

将变量 clshkplace 的取值由低到高分为四组，表示入样班级所有被调查学生中户口登记地在本县（区）的比例范围，班级层面的频数分布如下表：

clshkplace_4c 班级本县（区）户口学生比例分类

	频数	百分比
1 低于 50%	38	8.680
2 50%到 70%	81	18.49
3 70%到 90%	122	27.85
4 高于 90%	197	44.98
Total	438	100

对应的学生个体频数分布如下表：

clshkplace_4c 班级本县（区）户口学生比例分类

	频数	百分比
1 低于 50%	1,335	6.850

2 50%到 70%	3,120	16.01
3 70%到 90%	5,293	27.16
4 高于 90%	9,739	49.98
Total	19,487	100

3.clshktype

根据学生和家长在问卷中填写的学生户口类型，计算出每个入样班级所有被调查学生中，目前户口类型为农业户口的学生所占比例，均值为 0.54，标准差为 0.29，中位数为 0.60，最小值为 0，最大值为 1，其分组后的频数分布表见变量 clshktype_4c 的频数分布表。

4.clshktype_4c

将变量 clshktype 的取值由低到高分为四组，表示入样本班级所有被调查学生中户口类型为农业户口的比例范围，班级层面的频数分布如下表：

clshktype_4c　班级农业户口学生比例分类

	频数	百分比
1 低于 25%	106	24.20
2 25%到 60%	111	25.34
3 60%到 80%	110	25.11
4 高于 80%	111	25.34
Total	438	100

对应的频数分布如下表：

clshktype_4c　班级农业户口学生比例分类

	频数	百分比
1 低于 25%	4,462	22.90
2 25%到 60%	5,182	26.59
3 60%到 80%	4,833	24.80
4 高于 80%	5,010	25.71
Total	19,487	100

5.clsstlb

根据调查数据，计算出每个入样班级所有被调查学生中，父母不全在家的学生所占比例，班级层面的均值为 0.23，标准差为 0.17，中位数为 0.18，最小值为 0，最大值为 0.9，其分组后的频数分布表见变量 clsstlb_4c 的频数分布表。

6.clsstlb_4c

将变量 clsstlb 的取值由低到高分为四组，表示入样班级所有被调查学生中父母不全在家的学生的比例范围，,班级层面的频数分布如下表：

clsstlb_4c 班级父母不全在家的学生比例分类

	频数	百分比
1 低于 10%	91	20.78
2 10%到 30%	243	55.48
3 30%到 50%	66	15.07
4 高于 50%	38	8.680
Total	438	100

对应的学生个体频数分布如下表：

clsstlb_4c 班级父母不全在家的学生比例分类

	频数	百分比
1 低于 10%	3,889	19.96
2 10%到 30%	10,605	54.42
3 30%到 50%	3,237	16.61
4 高于 50%	1,756	9.010
Total	19,487	100

7.clsrank_5c

根据班主任问卷 A4 题，将班级在本年级中的排名分为五类：最差、中下、中间、中上和最好，若班主任问卷本题未回答，则使用任课教师本题答案的均值（四舍五入为整数）。班级层面的频数分布如下表：

clsrank_5c 班级目前在全年级的排名（五分类）

	频数	百分比
1 最差的	16	3.650
2 中下等	60	13.70
3 中等	147	33.56
4 中上等	166	37.90
5 最好的	49	11.19
Total	438	100

对应的学生个体的频数分布如下表：

clsrank_5c 班级目前在全年级的排名（五分类）

	频数	百分比
1 最差的	638	3.270
2 中下等	2,630	13.50
3 中等	6,793	34.86
4 中上等	7,283	37.37
5 最好的	2,143	11
Total	19,487	100

8.clsrank_3c

将变量 clsrank_5c 的五分类合并为三分类：中下及最差、中等及中上、最好，其中，“中下及最差”包括变量 clsrank_5c 的最差和中下两个取值，“中等及中上”包括变量 clsrank_5c 的中等和中上两个取值。班级层面的频数分布如下表：

clsrank_3c　班级目前在全年级的排名（三分类）

	频数	百分比
1 中下及最差	76	17.35
2 中等及中上	313	71.46
3 最好	49	11.19
Total	438	100

对应的学生个体的频数分布如下表：

clsrank_3c　班级目前在全年级的排名（三分类）

	频数	百分比
1 中下及最差	3,268	16.77
2 中等及中上	14,076	72.23
3 最好	2,143	11
Total	19,487	100

9.clshrsex

根据班主任问卷 C1 题，班主任的性别频数分布如下表：

clshrsex　班主任性别

	频数	百分比
1 男	149	34.02
2 女	289	65.98
Total	438	100

对应的学生个体的频数分布如下表：

clshrsex 班主任性别

	频数	百分比
1 男	7,154	36.71
2 女	12,333	63.29
Total	19,487	100

10.clshrbyr

根据班主任问卷 C2 题，班主任的出生年份均值为 1976，标准差为 7，中位数为 1977，最小值为 1953，最大值为 1995，将此变量根据调查年份转化为年龄并分组后对应的频数分布见变量 clshrage_4c 的频数分布表。

11.clshrage

根据调查年份，将班主任出生年份变量 clshrbyr 转化为班主任年龄，班主任的年龄均值为 37.16，标准差为 7.01，中位数为 37，最小值为 19，最大值为 60，其分组后的取值区间对应的频数分布见变量 clshrage_4c 的频数分布表。

12.clshrage_4c

将班主任年龄变量 clshrage 分组，分组后共有 4 个取值，频数分布如下表：

clshrage_4c 班主任年龄分组

	频数	百分比
1 30 岁以下	61	13.93
2 30-39 岁	221	50.46
3 40-49 岁	134	30.59
4 50 岁及以上	22	5.020
Total	438	100

对应的学生个体的频数分布如下表：

clshrage_4c 班主任年龄分组

	频数	百分比
1 30 岁以下	2,536	13.01
2 30-39 岁	10,077	51.71
3 40-49 岁	5,919	30.37
4 50 岁及以上	955	4.900
Total	19,487	100

13.clshrexp

根据班主任问卷 C7 题，班主任的教龄均值为 15.47，标准差为 7.7，中位数为 15，最小值为 0，最大值为 38，分组后所对应的频数分布见变量 clshrexp_3c 的频数分布表。

14.clshrexp_3c

将班主任教龄变量 clshrexp 分组，分组后共有 3 个取值，频数分布如下表：

clshrexp_3c　班主任教龄分组

	频数	百分比
1 10 年以下	109	24.89
2 10-19 年	202	46.12
3 20 年及以上	127	29
Total	438	100

对应的学生个体的频数分布如下表：

clshrexp_3c　班主任教龄分组

	频数	百分比
1 10 年以下	4,590	23.55
2 10-19 年	9,267	47.55
3 20 年及以上	5,630	28.89
Total	19,487	100

15.clsn

根据实际调查人数计算的每个入样班级参加调查的学生数量。根据抽样设计，入样班级当日到校的所有学生均参加调查，因此该变量基本上可看做是班级的实际人数，班级规模均值为 44.49，标准差为 13.11，中位数为 45，最小值为 9，最大值为 88，分组汇总后所对应的频数分布见变量 clsn_5c 的频数分布表。

16.clsn_5c

将变量 clsn 分组，共有 5 个取值，频数分布如下表：

clsn_5c　班级学生人数分类

	频数	百分比
1　30 人以下	60	13.70
2　30-40 人	90	20.55

3	40-50 人	132	30.14
4	50-60 人	102	23.29
5	60 人以上	54	12.33
Total		438	100

其对应的学生个体频数分布如下表：

clsn_5c　班级学生人数分类

		频数	百分比
1	30 人以下	1,412	7.250
2	30-40 人	3,168	16.26
3	40-50 人	5,882	30.18
4	50-60 人	5,390	27.66
5	60 人以上	3,635	18.65
Total		19,487	100

（四）个人/家庭层面派生变量

个人/家庭层面的派生变量反映学生性别、户口、迁移流动、父母是否长期在家与子女同住、是否住校、家庭经济条件、家庭成员、家庭环境等方面的性质，共有 19 个，其变量名和变量标签分别为：

变量名	变量标签
stsex	学生性别
sthkplace	学生户口登记地
sthktype	学生目前的户口类型
stmigrant	学生迁移流动状态
stlb_4c	父母是否在家同住(四分类)
stlb_2c	父母是否在家同住(二分类)
st0	学生类型:户口性质/户口登记地/父母是否同住
stbrd	学生平时是否住校
steco_5c	学生家庭经济状况(五分类)
steco_3c	学生家庭经济状况(三分类)
stonly	是否独生子女
stsib	兄弟姐妹数量(不含自己)
stsibrank	非独生子女在兄弟姐妹中的排行
stmedu	学生母亲的教育程度
stfedu	学生父亲的教育程度
stprhedu	学生父亲/母亲的最高教育程度
stfdrunk	学生父亲是否经常喝醉酒
stprfight	学生父母是否经常吵架

stprrel	学生父母是否关系很好

1.stsex

根据学生问卷 A1 题，频数分布如下表：

stsex 学生性别

	频数	百分比
0 女	9,445	48.47
1 男	10,042	51.53
Total	19,487	100

2.sthkplace

根据学生问卷 A5 题，频数分布如下表：

sthkplace 学生户口登记地

	频数	百分比
0 外县(区)	3,500	17.96
1 本县(区)	15,987	82.04
Total	19,487	100

3.sthktype

根据学生问卷 A6 题，原变量的“非农户口”和“居民户口”合并为“非农业户口”，频数分布如下表：

sthktype 学生目前的户口类型

	频数	百分比
0 非农业户口	8,800	45.16
1 农业户口	10,687	54.84
Total	19,487	100

4.stmigrant

根据学生问卷 A5 题的选项及填写的具体户口所在地，归纳出学生是否流动及流动范围，频数分布如下表：

stmigrant 学生迁移流动状态

	频数	百分比
1 本地非流动	15,987	82.55
2 省内流动	1,600	8.26
3 跨省流动	1,779	9.19
Total	19,366	100

5.stlb_4c

根据学生问卷 B4 和 B5 题，归纳出学生的父亲和母亲是否在家与孩子同住，共有 4 个取值，频数分布如下表：

stlb_4c　父母是否在家同住(四分类)

	频数	百分比
1 父母都在家	14,953	76.73
2 只有母亲在家	1,809	9.280
3 只有父亲在家	674	3.460
4 父母都不在家	2,051	10.52
Total	19,487	100

6.stlb_2c

将变量 stlb_4c 合并为 2 个取值，其中“只有一方在家或都不在家”包括变量 stlb_4c 中的 2、3、4 三个取值，频数分布如下表：

stlb_2c　父母是否在家同住(二分类)

	频数	百分比
0 父母都在家	14,953	76.73
1 只有一方在家或都不在家	4,534	23.27
Total	19,487	100

7.st0

根据学生的“户口性质”（农业/非农）sthktype、“户口登记地”（本县区/外县区）sthkplace 和“父母是否在家与孩子同住”stlb_2c 这三个变量，将学生划分为 8 个类型，频数分布如下表：

st0　学生类型:户口性质/户口登记地/父母是否同住

	频数	百分比
1 非农流动父母双全	1,042	5.350
2 非农流动缺父母	231	1.190
3 非农本地父母双全	6,057	31.08
4 非农本地缺父母	1,470	7.540
5 农业流动父母双全	1,893	9.710
6 农业流动缺父母	334	1.710
7 农业本地父母双全	5,961	30.59
8 农业本地缺父母	2,499	12.82
Total	19,487	100

8.stbrd

根据学生问卷 C18 题，频数分布如下表：

stbrd　学生平时是否住校

	频数	百分比
0 不住校	13,194	67.71
1 住校	6,293	32.29
Total	19,487	100

9.steco_5c

根据家长问卷 E19 题和学生问卷 B9 题，首先使用家长问卷数据，当家长问卷此题未回答时，则采用学生问卷数据，共有 5 个取值，频数分布如下表：

steco_5c　学生家庭经济状况(五分类)

	频数	百分比
1 非常困难	735	3.780
2 比较困难	3,341	17.20
3 中等	14,188	73.03
4 比较富裕	1,101	5.670
5 很富裕	63	0.320
Total	19,428	100

10.steco_3c

将变量 steco_5c 合并为 3 个取值，其中“困难”包括变量 steco_5c 中的“非常困难”和“比较困难”两个取值，“富裕”包括变量 steco_3c 中的“比较富裕”和“很富裕”两个取值。频数分布如下表：

steco_3c　学生家庭经济状况(三分类)

	频数	百分比
1 困难	4,076	20.98
2 中等	14,188	73.03
3 富裕	1,164	5.990
Total	19,428	100

11.stonly

根据学生问卷 B1 和 B2 题，频数分布如下表：

stonly　是否独生子女

	频数	百分比

1 是	8,460	43.41
2 不是	11,027	56.59
Total	19,487	100

12.stsib

根据学生问卷 B1 和 B2 题，所有独生子女此变量均为缺失值，频数分布如下表：

stsib　兄弟姐妹数量(不含自己)

	频数	百分比
1	8,043	74.94
2	2,068	19.27
3	436	4.06
4	113	1.05
>=5	72	0.67
Total	10,732	100

13.stsibrank

根据学生问卷 B1 和 B2 题，所有独生子女此变量均为缺失值，频数分布如下表：

stsibrank　非独生子女在兄弟姐妹中的排行

	频数	百分比
1 排行最大	5,158	48.06
2 排在中间	1,034	9.63
3 排行最小	4,540	42.30
Total	10,732	100

14.stmedu

根据学生问卷 B8 题和家长问卷 E7 题，频数分布如下表：

stmedu　学生母亲的教育程度

	频数	百分比
1 没受过任何教育	726	3.730
2 小学	4,113	21.16
3 初中	7,967	40.98
4 中专/技校	1,156	5.950
5 职业高中	361	1.860
6 高中	2,608	13.41

7 大学专科	1,070	5.500
8 大学本科	1,259	6.480
9 研究生及以上	181	0.930
Total	19,441	100

15.stfedu

根据学生问卷 B8 题和家长问卷 E7 题，频数分布如下表：

stfedu 学生父亲的教育程度

	频数	百分比
1 没受过任何教育	125	0.640
2 小学	2,874	14.78
3 初中	8,578	44.12
4 中专/技校	1,123	5.780
5 职业高中	422	2.170
6 高中	3,328	17.12
7 大学专科	1,152	5.930
8 大学本科	1,557	8.010
9 研究生及以上	282	1.450
Total	19,441	100

16.stprhedu

将学生父亲和母亲的教育水平比较并取其中相对较高的值，频数分布如下表：

stprhedu 学生父亲/母亲的最高教育程度

	频数	百分比
1 没受过任何教育	64	0.330
2 小学	1,779	9.150
3 初中	8,559	44.03
4 中专/技校	1,032	5.310
5 职业高中	434	2.230
6 高中	3,965	20.40
7 大学专科	1,311	6.740
8 大学本科	1,939	9.970
9 研究生及以上	358	1.840
Total	19,441	100

17.stfdrunk

根据学生问卷 B10 题，频数分布如下表：

stfdrunk　学生父亲是否经常喝醉酒

	频数	百分比
1 否	17,245	91.48
2 是	1,606	8.520
Total	18,851	100

18.stprfight

根据学生问卷 B10 题，频数分布如下表：

stprfight　学生父母是否经常吵架

	频数	百分比
1 否	16,943	89.95
2 是	1,892	10.05
Total	18,835	100

19.stprrel

根据学生问卷 B10 题，频数分布如下表：

stprrel　学生父母是否关系很好

	频数	百分比
1 否	3,137	16.52
2 是	15,856	83.48
Total	18,993	100