



北京大学 中国社会科学调查中心
Institute of Social Science Survey, Peking University



数据更新简介与用户常见问题

戴利红

中国家庭追踪调查
项目办公室

2016-07-17

1 CFPS2014数据

- ◆ 数据结构简介
- ◆ 问卷设计更新
- ◆ 数据要点提示

2 往期数据（CFPS2010）更新

- ◆ 更新原有变量
- ◆ 添加新变量

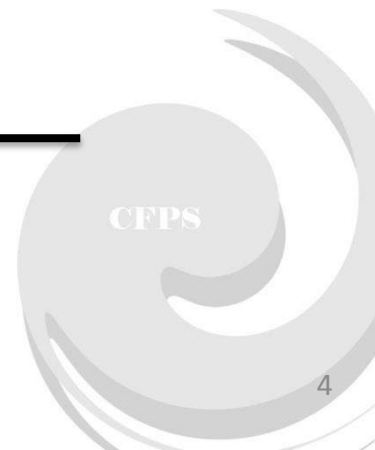
3 用户常见问题



第一部分：CFPS2014数据

1 数据结构简介

数据库	样本量	变量数
村居库	622	226
家庭关系库	57739	278
家庭经济库	13946	398
成人库	37114	2250
少儿库	8594	990



1 概念

➤ 家庭层面

原家庭：当期调查前至少完访一次家庭成员问卷的家庭

另组家庭：当期调查至上次调查之间由原家庭分裂出来的新家庭

➤ 个人层面

基因成员：2010年所有家庭T1表成员，及其在最近两次调查之间新生儿或是领养的不超过10岁子女。（有个人问卷）

核心成员：关系人（父母、配偶、子女）中至少有一个是基因成员。（有个人问卷）

非核心成员：关系人都不是基因成员的家庭成员。（无个人问卷）

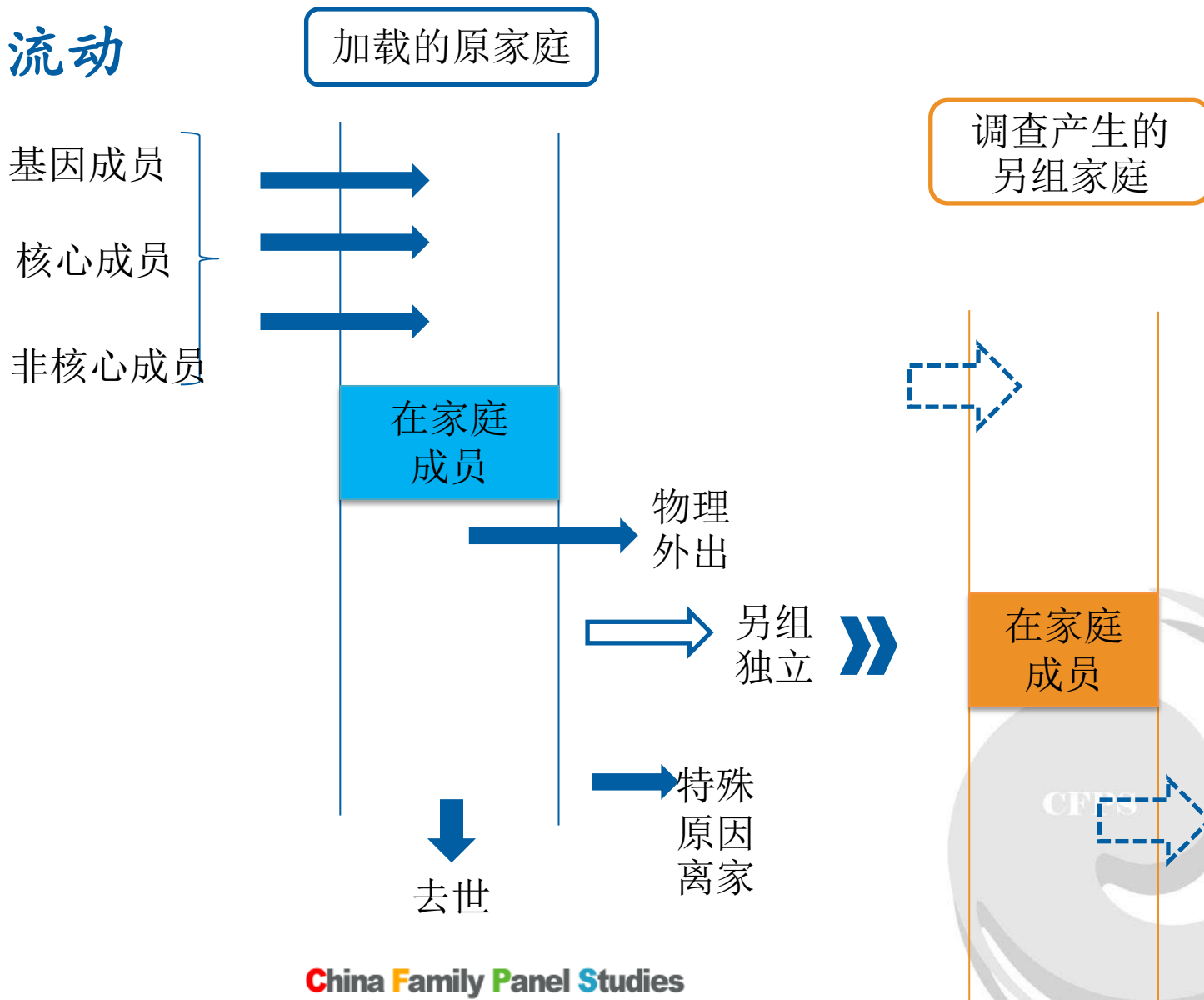
家庭关系库

家庭编码	个人编码及基本信息 (vars_p)	父亲编码及基本信息 (vars_f)	母亲编码及基本信息 (vars_m)	配偶编码及基本信息 (vars_s)	孩子们编码及基本信息 (vars_ci)	综合变量: 成员类型、家庭规模、地址信息
家庭A	A1	A3	A2
	A2	A3	A1
家庭B	A2 B1

家庭成员
及离家人员信息

家庭关系
网及关系
人信息

2 人员流动



家庭成员列表：同灶吃放的所有成员
Pid, fid14 -- co_a14_p(=1)

住在家庭中的成员列表：
Pid, fid14 -- tb6_a14_p(=1)

个人的历年家庭归属：
Pid -- Fid14, fid12, fid10

a1	A	A	A
a2	A	-8	-8
a3	A	B	B



2 问卷设计更新

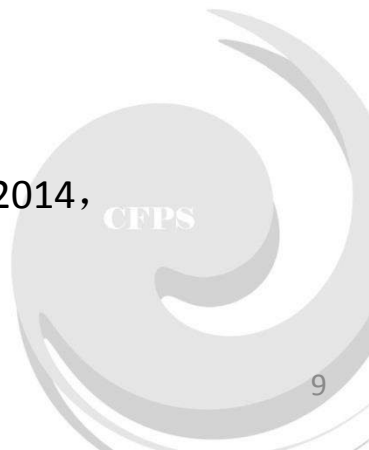
1 离家人员类型的界定

离家人员类型	CFPS2010	CFPS2012	CFPS2014 
物理外出	原家庭的判断 (经济有联系)	原家庭的判断 (经济有联系)	原家庭和离家 人员双重判断 (经济有联系)
另组 独立	null	原家庭的判断 (经济独立)	原家庭和离家 人员双重判断 (经济独立)

CFPS2014

调查时，根据原家庭的判断结果会实时生成家庭成员列表FML2014，供其他问卷加载使用

清理时，根据离家人的判断结果，会调整家庭成员构成。

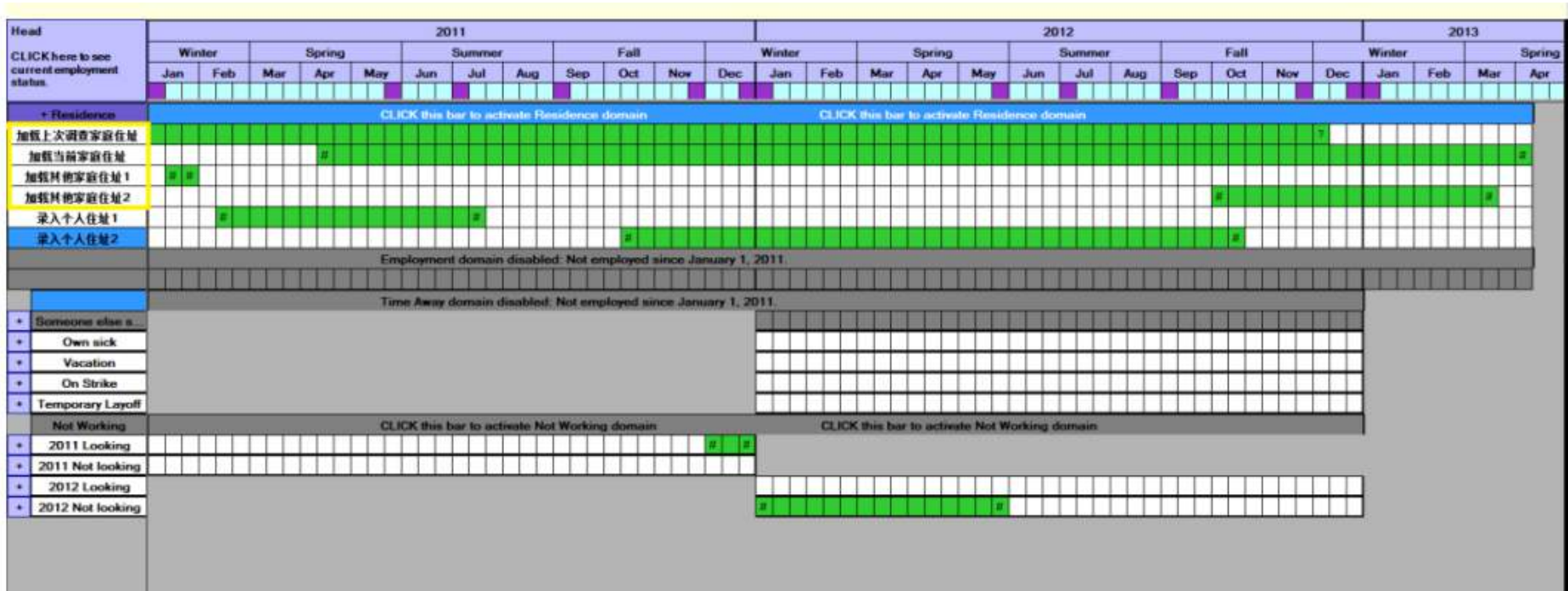


清理原则:

- 如果成功追访离家人员，则以离家人员自己界定为准
- 如果未追访到离家人员，则以原家庭界定为准

		离家人员的判定	
原家庭的判定		另组独立（是成员）	物理外出（不是成员）
另组独立 （不是成员）	是 另组家庭 ，调查版本与清理版本成员构成一致	原家庭的物理外出成员，调查版本的家庭成员少于清理版本	
物理外出 （是成员）	是 另组家庭 ，调查版本的家庭成员多于清理版本	原家庭的物理外出成员，调查版本与清理版本一致	

② 事件日历调查(Event History Calendar, EHC)方法的引入



EHC是指通过勾勒出时间轴线的模式，帮助受访者更加准确地回忆在一定时间段内某些事件的发生结点。

3 数据要点提示

1 离家人员类型的界定对经济数据的影响

CFPS2014家庭经济问卷中的描述：

根据之前的回答，您家目前是“【CAPI】加载FML2014家庭成员人数”口人，本问卷的问题包括以下所有家庭成员：

【CAPI】加载家庭成员名单FML2014。

【CAPI】加载家庭成员人数FML2014NUM。

调查时实时加载的数据，会和清理版本有差别

Q2 过去12个月，您家哪些人帮其他农户做农活或外出打工挣钱？【可多选】

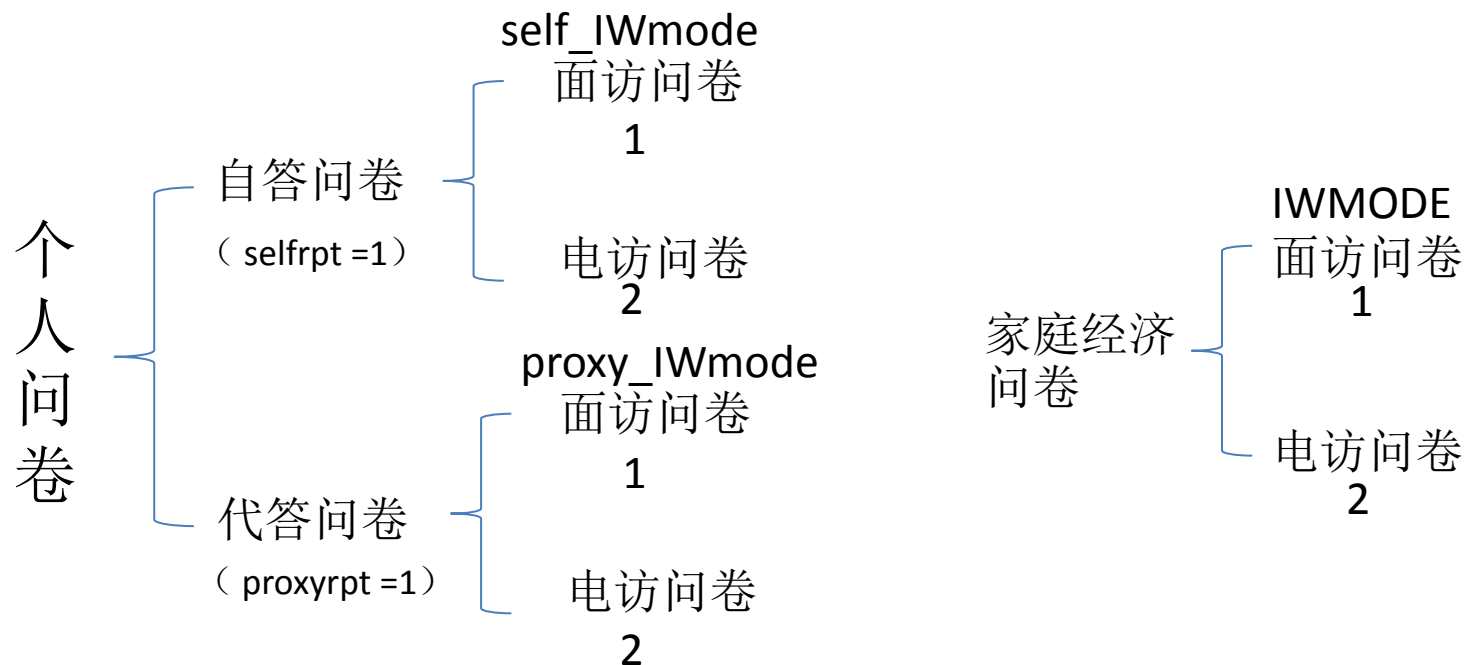
【CAPI】

#1 加载家庭成员名单FML2014 + “76.所有家庭成员”；

辅助调整的变量

变量名	变量标签	值标签
overlapfid1	经济问卷关联家户1	
overlapfid2	经济问卷关联家户2	
overlapfid3	经济问卷关联家户3	
overlapfid4	经济问卷关联家户4	
overlapfid1type	与关联家户1相关类型	1=fid14与关联家户完全相等 2=fid14与关联家户交叉；即fid14与关联家户有部分共同成员，但是不等 3=fid14完全包含于关联家户 4=fid14完全包含关联家户
overlapfid2type	与关联家户2相关类型	
overlapfid3type	与关联家户3相关类型	
overlapfid4type	与关联家户4相关类型	

2 问卷类型和访问模式在数据库中的体现



面访、自答的形式为主，电访、代答的形式为辅



③ 数据库的合并

➤ 面访与电访数据（成人库和家庭经济库）

1. CFPS2014中家庭经济、成人、少儿问卷中电访所占比例分别为3.91%、7.82%、8.55%。
2. 变量的处理规则：
统一相同问题不同变量名
区别不同问题相同变量名，电访独有的变量在面访变量后面。

➤ 成人自答与代答数据

1. 成人库中只有自答或代答问卷的比例分别是85.69%、9.88%，剩下4.42%的人既有自答问卷，又有代答问卷。
2. 变量的处理规则：
统一相同问题不同变量名
区别不同问题相同变量名，代答独有的变量在电访变量后面。



4 综合变量

➤ 家庭经济库

家庭收入部分：

全部家庭纯收入、人均家庭纯收入、人均家庭纯收入分位数、及其各分项收入

14年信息

——

与2010年可比

家庭支出部分：

家庭总支出、居民消费性支出-加总、及其各项分项支出等。

➤ 成人与少儿库

个人收入部分

个人收入、工资性收入、个人累加工资性收入

认知水平部分

词组测试题得分：14年问卷算法

词组测试题得分：与10年可比算法

数学测试题得分：14年问卷算法

数学测试题得分：与10年可比算法

第二部分： 往期数据CFPS2010更新

更新原有变量

1. 个人教育综合变量

成人库、少儿库、家庭关系库：

完成的最高学历，完成的受教育年限，离校/上学阶段

2. 婚姻相关变量

成人库： 婚姻状态_修订、初婚的结婚年份_修订、初婚配偶出生年份_修订、前任去世年份_修订、与前任的结婚年份_修订、前任配偶出生年份_修订

3. 父母背景变量

成人库、少儿库、家庭关系库：

父母分别的最高学历、出生年-综合变量、主要职业编码（国标）、主要职业ISCO88、政治面貌

4. 父母与子女关系顺序码转换为户内三位码变量

成人库： 您为哪些子女提供经济帮助：选择1... ..哪些子女为您提供经济帮助：选择1

添加新变量

1. 第一阶段抽样单位信息PSU，涉及各库
2. 部分编码信息

成人库：父母去世原因变量;行政管理职务编码

少儿库：少儿职业期望相关变量

第三部分：用户常见问题

一、收入相关的综合变量

什么是未调整收入？什么是调整后收入？什么是可比性收入？
什么是不可比收入？

在2014成人问卷中，个人收入变量中“0”的比例占到47.63%。为什么较多的人没有收入呢？

个人收入是工资、奖金、各种福利、退休金、勤工助学金、奖学金的总和。

对于农民、做生意的人，没有问他们的个人收入（这部分收入在家庭问卷中已经问过了），但他们占得比例较大，所以样本中个人收入的缺失较大。

二、城乡相关变量

社区问卷

请访员记录现在所在的是居委会还是村委会？

1. 居委会 5. 村委会

经济问卷

A0: 访员注意：无需提问，直接记录目前受访家户所在社区是居委会还是村委会？

1. 居委会 5. 村委会

国家统计局分类

Urban：根据国家统计局网站城乡分类进行整理的。

CFPS2012：特殊说明

- 城乡变量（urbancomm）

A0 + A0缺失的填补

- 居住社区类型（typecomm）

主要根据访员观察采集结果，辅助A0和A0缺失的填补。

Z7：被访村/居的类型

1. 城市（跳至Z701） 2. 城镇（跳至Z702） 3. 农村（跳至Z703） 4. 郊区（结束）

Z701、Z702：请选择一项最能描绘被访村/居具体类型的选项

三、关于不适用（-8）和缺失（.）

在2014家庭经济数据库中，住房面积这一变量中“不适用”的比例达到81.97%。此处的“不适用”含义是什么呢？

请问一下cfps2012成人库中的qa501 qa601等反应居住地迁移的变量中为什么有很多不适用？怎样的情况会被标注为不适用呢？

Q801 您家现住房的建筑面积是多少平方米？ ____ 1.0 .. 10,000.0平方米。

【CAPI】 若CFPS2014_interview=0（初访家户），跳至Q2；否则继续提问Q1。

Q1 FQ1”是否同一住房”您家现在住的房子与“【CAPI】加载上一次调查年月”调查时住的是同一个房子吗？

1. 是 5. 否

∴ ∴

【CAPI】 若Q1=1且CFPS2012_Q701=1，提问Q8，否则直接提问Q801。

Q8 FQ8”住房面积是否变化”从“【CAPI】加载上一次调查年月”至今，您家现住房的住房面积有没有发生改变？

1. 没变化（跳至R部分） 3. 变大 5. 变小

【CAPI】 Q8=“不知道”或拒绝回答跳至R部分。

2014年成人教育史问题，按照跳转判断的回答人却有缺失的情况，为什么？

注意成人电访面访、自答代答问卷数据合并造成的影响，独有的问题是缺失（.）的状态。

四、顺序码的使用

U1 外出工作模块中成员1-21和其他成员1-3分别是谁，如何对应到其各自的个人信息？

Q3 您家现在居住的这所房子的房产证（或土地证）上有哪些家庭成员的名字，请您按照房产证（土地证）上的顺序依次提供他们的姓名？【可多选】

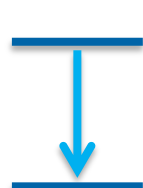
访员注意：需要按照房产证上名字的顺序进行输入。

【CAPI】加载FML2014成员名单。+ 79. 没有房产证

例子

问卷问题： U1外出工作的是哪些人？【可多选】
【CAPI】#01列出T1表格所有成员供选择。

加载T1列表 Name1 Name2 Name3

 存储值 1 2 3

转换值 pid1 pid2 pid3

五、跨库使用

成人库的pid与家庭关系库进行连接，为什么会出现重复观测？

另组家庭成员的影响，用是否同灶吃饭co_a14_p来区分。

个人库与家庭关系库有相同问题，但是信息不一致，应该使用哪个？

个人问卷信息优先，其次是家庭关系库。

end

THANKS

CFPS 项目办公室