



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

# “中国家庭追踪调查” 数据更新介绍

吴琼  
2014年7月

# 主要内容

**CFPS背景**

**2012年数据介绍**

**常见问题**

# CFPS

- 访问对象为被抽取家庭的所有家庭成员
- 跨学科

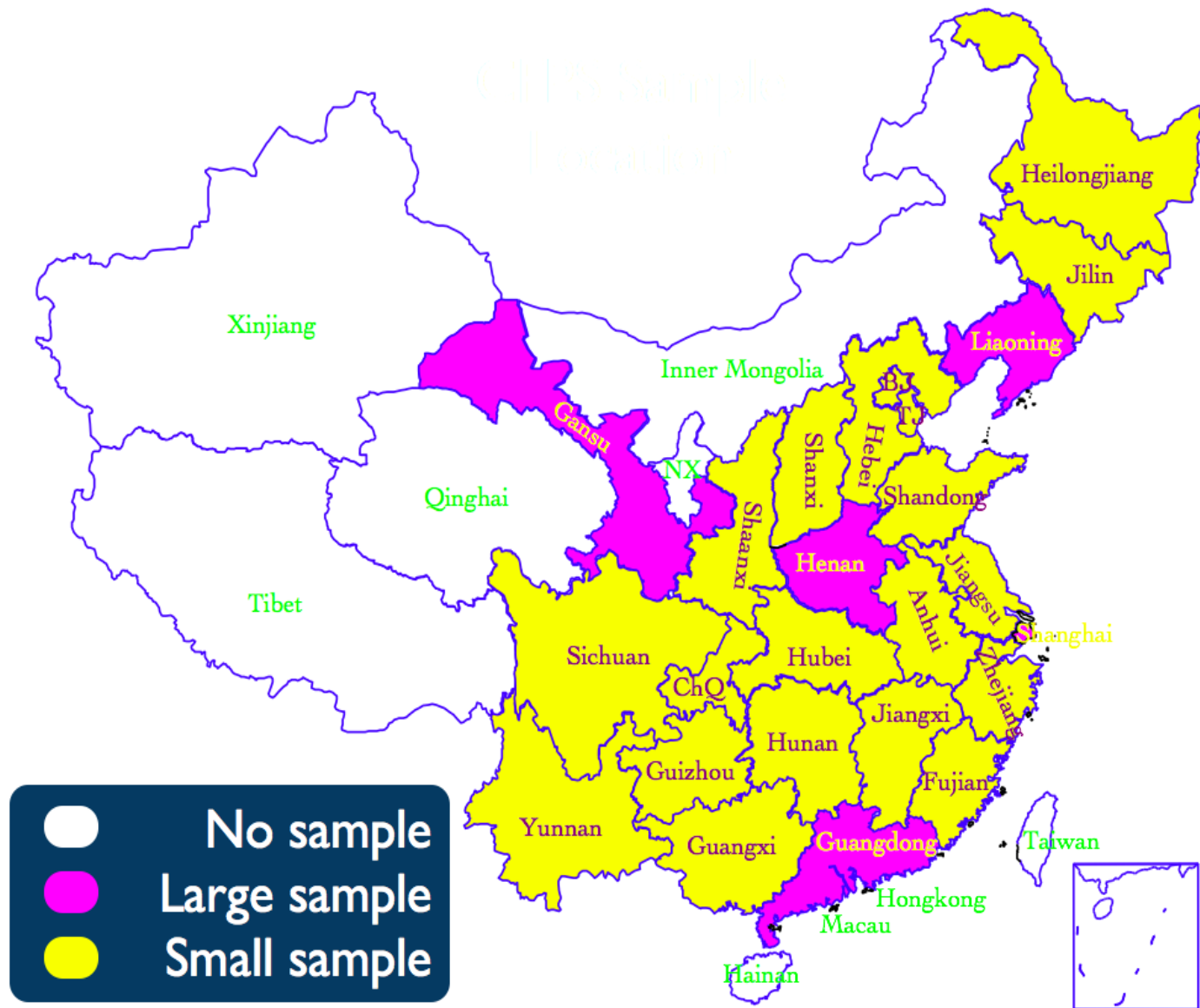
# 问卷内容涵盖的领域



# CFPS

- 访问对象为被抽取家庭的所有家庭成员
- 多领域、跨学科
- 全国代表性样本和五个“大省”省级代表性样本

# 样本来源区



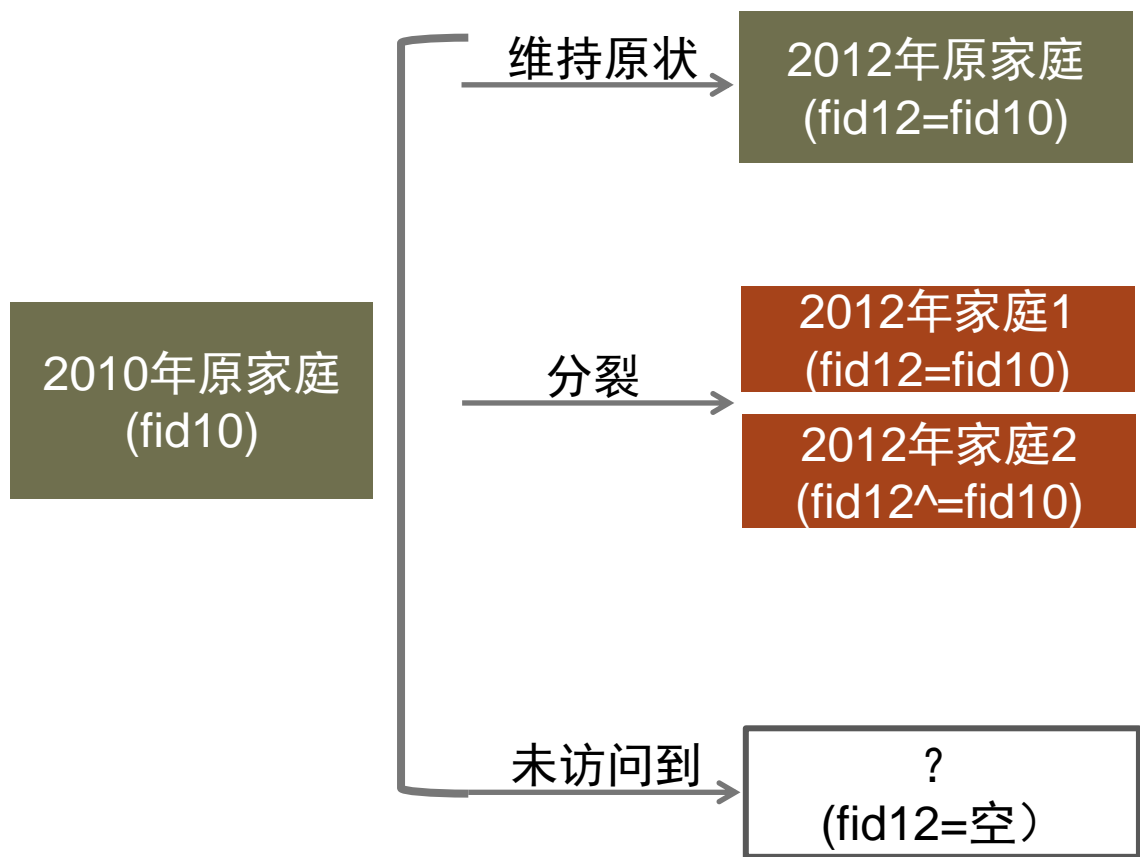
# CFPS

- 访问对象为被抽取家庭的所有家庭成员
- 多领域、跨学科
- 全国代表性样本和五个“大省”省级代表性样本
- 长期动态追踪
- 数据开放与共享

# 数据结构

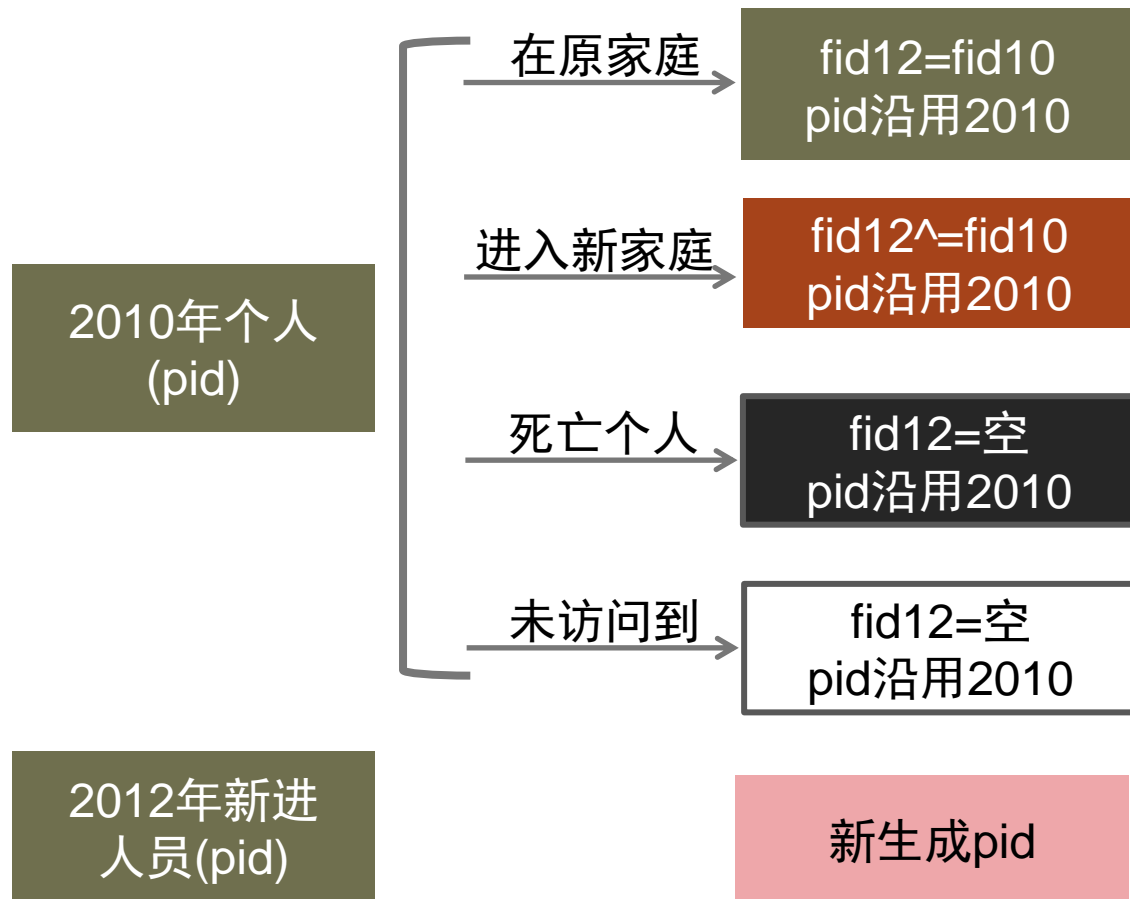


# 跨年数据结构：家庭层面



家  
户  
号  
跟  
家  
庭  
走

# 跨年数据结构：个人层面



个人号恒久唯一

# 几种类型举例

pid	fid12	fid10
100021101	100021	100021

基线成员  
2012年在原家庭

pid	fid12	fid10
110001104	223200	110001

基线成员  
2012年进入另组家庭

pid	fid12	fid10
100001401	100001	-8

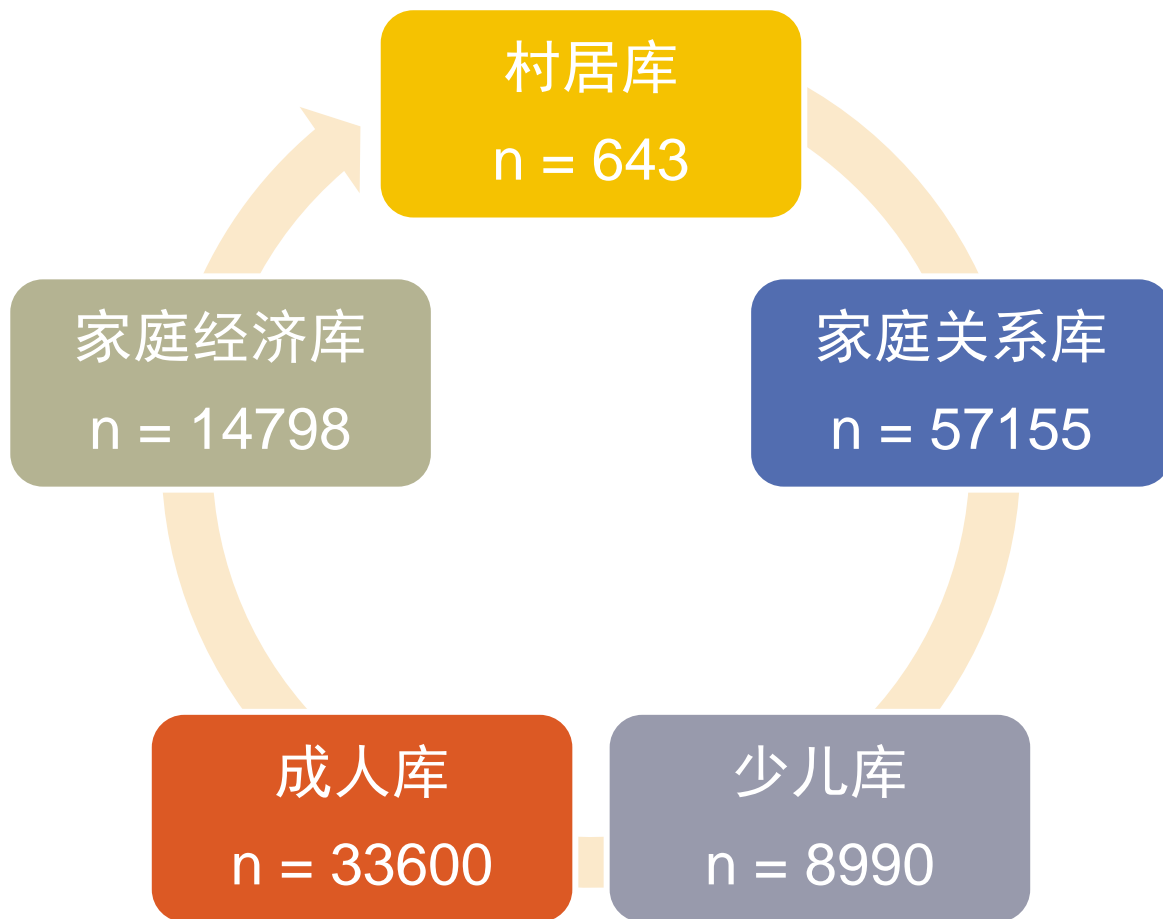
新进个人

pid	fid12	fid10
120011106	-8	120011

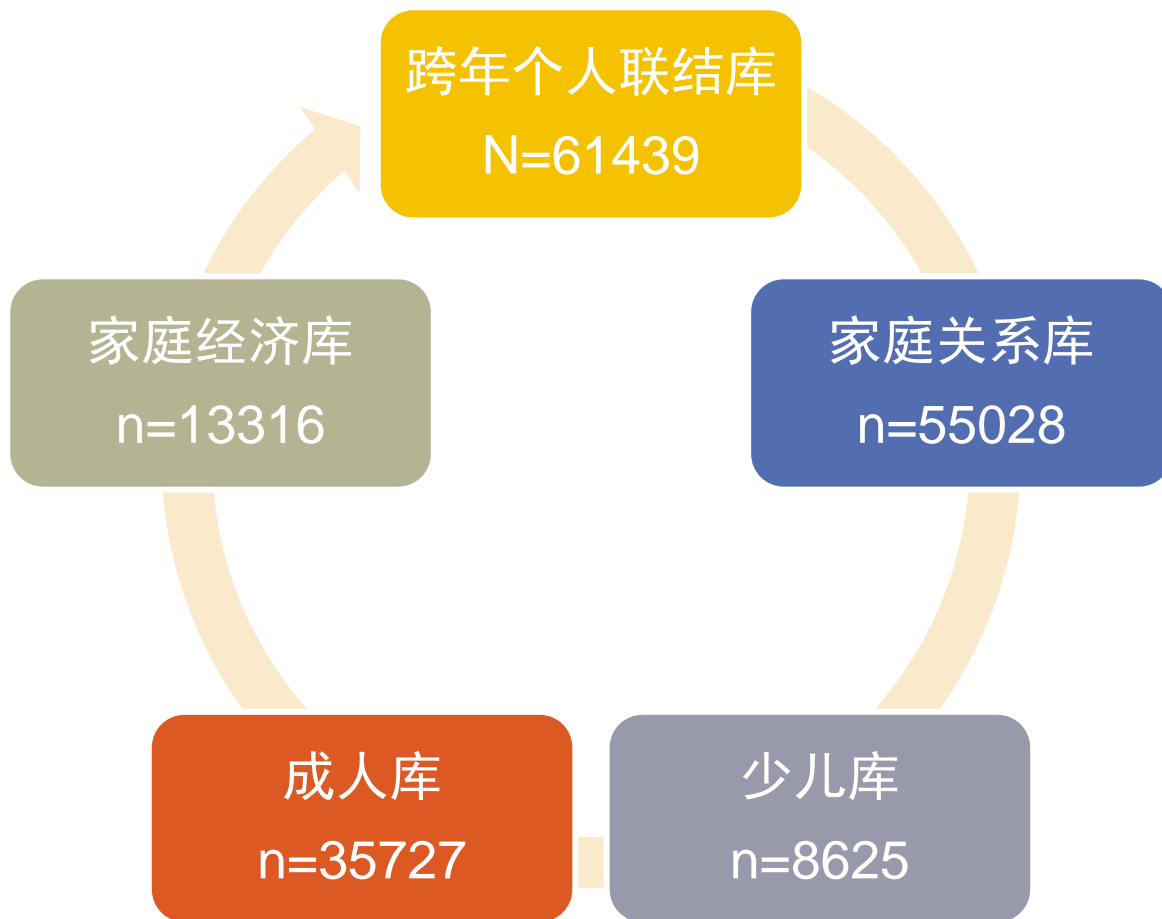
未接触到或死亡



# 数据库：2010年



# 数据库：2012年



# 最佳变量 ( **\_BEST** )

## □ 适用类型

- 存在不同数据来源
- 数据逻辑检验发现需要修正

## □ 举例

- 年龄

cfps2012\_birthy\_best: 受访者出生年-修正

- 性别

cfps2012\_gender\_best : 受访者性别-修正

# 综合变量

## □ 适用类型

- 分项加总（收入）
- 信息综合（就业状态）
- 复合测量（认知水平）

## □ 举例

- fincome1: 2011-2012全部家庭纯收入
- expense: 家庭总支出
- employ: 就业状态
- IWR: 短期记忆得分
- NS\_W: 数列题W分数

# 特殊变量

## 访问形式 (lwmode)

- 入户访问
- 电话访问

数值	标签值	频率	百分比
1	面访	34731	97.21
2	电访	998	2.79

## 问卷形式 (longform, shortform)

- 长问卷 (自答)
- 短问卷 (代答)

数值	标签值	频率	百分比
0	否	2568	7.19
1	是	33161	92.81

数值	标签值	频率	百分比
0	否	31388	87.85
1	是	4341	12.15



# 常见问题

# 常见问题

Q1

“做纵贯性分析(longitudinal analysis),如何匹配个人?”

关键变量: pid

个人pid在不同年份保持不变

# 常见问题

## Q2

“为什么将2012和2010年少儿库连接后有很多样本匹配不上？”

1. 2012年新进成员在2010年没有观测
2. 2010年13周岁以上成员在2012年成人库中
3. 样本流失

# 常见问题

## Q3

“如何连接子女与父母的信息”

1. 家庭成员关系库中直接给出直系亲属的基本信息（年龄、婚姻、学历、户口、在家状态等）
2. 对于存在个人问卷的子女与父母，通过家庭成员关系库中给出的相应pid在个人问卷库中链接其它信息

# 常见问题

Q4 “有关具体金额的题目，为何很多变量缺失？问卷中所提及的逼近法（unfolding）是如何操作的？”

1. 敏感性问题（如收入多少）预期缺失率较其它问题高
2. 逼近法针对具体数额缺失人群，试图找到其区间。
3. CFPS设计：从特定的中间点出发，向右问是否**高于**，向左问是否**低于**
4. 数据库中提供根据区间生成的估计值

fid12	FL3	FL3_max	FL3_min	FL3est
100001	5000	缺失	缺失	缺失
100002	不知道	5000	2500	3750

# 常见问题

Q5 “为什么CFPS不提供具体的区县码？”

1. 为了更好的保护受访者信息
2. CFPS区县数据库提供各样本区县的宏观变量（GDP、人均GDP、人口数、就业率、平均受教育年限、劳动年龄人口比例、老年人口比例、10到19岁人口性别比例、非农业户口人口比率），各变量数值均经过模糊化处理，可供统计分析

网站

<http://www.iss.edu.cn/cfps/>

电子邮件

[iss.cfps@pku.edu.cn](mailto:iss.cfps@pku.edu.cn)