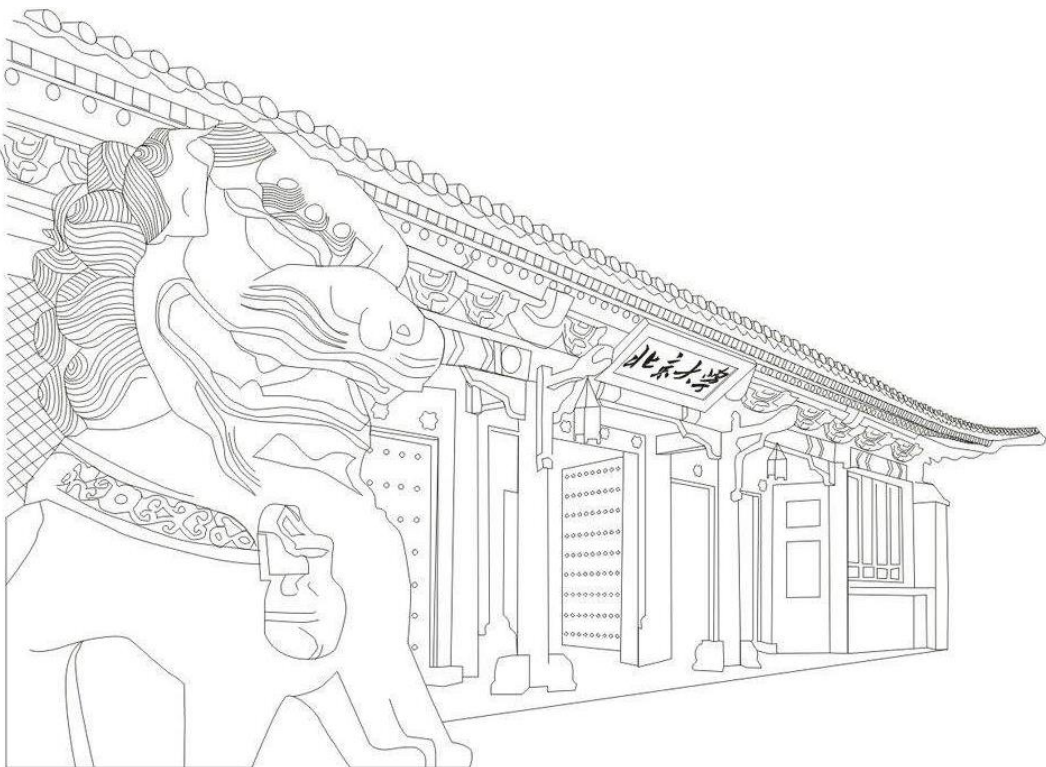




家庭与婚姻



CFPS项目办公室

2017.09.22 · 北京大学



主要内容

信息来源

数据特点

信息整合



信息来源一：家庭关系库

- 两期调查之间家庭成员的流动过程

新进成员、离家人员

- 当期的家庭成员构成

经济上一家人（co_a14_p=1），物理上同住（tb6_a14_p=1）

- 成员的基本个人信息



- 家庭关系结构：父母、配偶、多个子女



信息来源二：成人婚姻模块

- 更正或补充上轮调查时婚姻状态及配偶基本信息

更正对象：上轮有婚姻状态且错误

补充对象：上轮未完访或本轮新进样本

- 两轮调查之间的婚姻史：
婚姻状态的变化及采集新婚配偶的基本信息

未婚/离婚/丧偶 → 在婚

在婚 → 在婚

在婚 → 离婚/丧偶



数据特点：家庭关系库

1. 直接采集成员的家庭关系及其婚姻状态
2. 不单独采集非家庭成员（2字头）的家庭关系

3. 对新出现的样本直接采集家庭关系（部分）
4. 后期人工更新原有家庭关系

3. 对新出现的样本直接采集家庭关系（部分）
5. 对进入家庭中的2字头成员直接采集家庭关系（部分）
6. 采集新进和离家人员的婚姻状态
4. 后期人工更新原有家庭关系

同CFPS2016

CFPS2016

CFPS2014

CFPS2012

CFPS2010



- 样本完整，家庭关系人样本编码指代明确，方便与其他数据库进行样本匹配
- 个人基本信息较完整，但是未完访样本存在代答情况
- 后期已经删除离婚和丧偶成员的配偶信息
- 存在基线成员的配偶信息需要修正的情况
 - 初婚 → 在婚且配偶没有进入家庭的
 - 在婚 → 再婚且配偶没有进入家庭的
 - 离婚/丧偶 → 再婚且配偶没有进入家庭的
- 存在新进成员家庭关系不完整的情况
 - 事实上存在的关系人，但是没有进入CFPS样本库，处于缺失状态。
 - 基因成员的子女数量是完整的，但是非基因成员的子女数量存在缺失的可能。



CFPS2014 基因成员和核心成员子女采集数量的对比

年龄组	基因成员子女 数量均值	核心成员子女 数量均值
25岁以下	0.78	0.80
25-29岁	1.12	0.85
30-34岁	1.46	0.99
35-39岁	1.62	1.11
40-44岁	1.74	1.01
45-49岁	1.84	0.85
50-54岁	1.93	0.87
55-59岁	1.99	0.99
60-64岁	2.22	1.11
65-69岁	2.65	1.11
70-74岁	3.27	1.26
75-79岁	3.61	1.06
80岁及以上	3.80	1.02
总成员量	(31766)	(2726)



数据特点：成人婚姻模块

- 婚姻信息是成人自答内容，信息相比代答更准确
- 需要根据CAPI跳转整理当前补充和变更的配偶信息
- 不仅有当前婚姻状态，还有婚姻史的信息
- 没有对应的配偶样本编码，无法与其他数据库样本进行匹配



信息整合

1. 根据成人婚姻模块的信息**整理**出当期调查时新收集的受访者在婚配偶信息，并标记是否是再婚配偶；
2. 利用上述信息**补充**和**更正**家庭关系库中成员的配偶信息。

如何补充？

- 对象：个人样本的配偶编码缺失($pid_s=-8$)，即配偶未进入CFPS样本库
- 方法：直接采用成人婚姻模块的信息

如何更正？

- 对象：往期个人样本的配偶编码不缺失的 ($pid_s \neq -8$)。
- 方法：
 - 对于配偶是新样本且新样本未完访的情况，则优先使用成人婚姻模块的信息；
 - 对于配偶是往期样本且有再婚配偶的记录，首先删除上轮记录的配偶信息，然后直接采用成人婚姻模块的信息



谢 谢