

Family member codebook

Last updated : 1004

Please refer to csv : main_subject_XXXX.csv for the latest version

Remarks :

1. 1989年的與之後的coding 差蠻多的，目前有 {1993, 1996, 1999, 2003, 2007}

- unif: **追蹤樣本的編碼（將於後面詳述）
- problem : unif可能有的問題，將於後面詳述
- survey year : 問卷年份 {1993, 1996, 1999, 2003, 2007}
- told relationship : subject回報的關係，請對照關係表
- male : 生理性別 : 1男生, 0女生
- age: 年紀 : num
- birthyear 出生年份（民國） : num

- edu : 教育程度 :

0 不識字 或僅識字

1~6 小學 1~6年級

7~9 初中 1~3年級

10~12 高中 1~3年級（空中大學肄業 == 12）

13~ 16 大學 1~4年級

17 研究所與以上

NA 不適用（成員未滿6歲），或是資料缺失

- work 目前工作概況 :

1. 有工作（包含暫時沒有去做）

2. 在學

3. 服役

4. 家務管理（煮飯、洗菜、顧小孩...）

5. 失業或退休
6. 半工半讀
7. 其他（幫助家裡的農事或生意，不算真正工作）

NA：不適用或是資料缺失

- marstat 目前婚姻狀態

1. 有偶
2. 同居
3. 分居
4. 離婚
5. 喪偶
6. 未婚

NA 不適用（家戶成員未滿15歲）或是資料缺失

- have_child：有沒有小孩 dummy

- number_of_children 目前小孩數量：num

NA 不適用（家戶成員未滿15歲，或是未婚）或是資料缺失

- livstat 居住狀況

- 0 住在一起
- 1 正式分開住
- 2 因為求學而外住
- 3 因為工作而外住
- 4 服兵役
- 5 其他

NA 不適用或是資料缺失

- livwhere 目前居住地

0 住在一起

1 隔壁

2 同鄰

3 鄰近地區

4 臺灣他地

5 中國

6 外國

NA 不適用或是資料缺失

- meet 多久見一次面

0 住在一起

1 每天

2 每隔幾天

3 每個禮拜

4 每個月

5 每幾個月

6 每一年

7 每隔幾年

8 不需見面

91 不定期

NA 不適用或是資料缺失

- tele 多久講一次電話

coding 同上

- y_qser_no : 問卷樣本編號，建議用此去對 main subject e.g. 1999_A_1231

- adl_who_help 誰是主要adl的照顧者

- iadl_who_help 誰是主要的iadl 照顧者

NA == 不需要幫忙或是資料缺失

0 = 需要幫忙但沒人幫忙

1= 自己

其他請參照關係表

- subject_residency : subject 現居地（縣市別），請參考附錄

- subject_in_nursing_house :

Family member 的subject是否住在安養機構、榮民之家、仁愛之家..

1: yes, 0:no （住在家裡）

Relationships : “ is.xxx” ,dummy variables

— — —

k-group說明

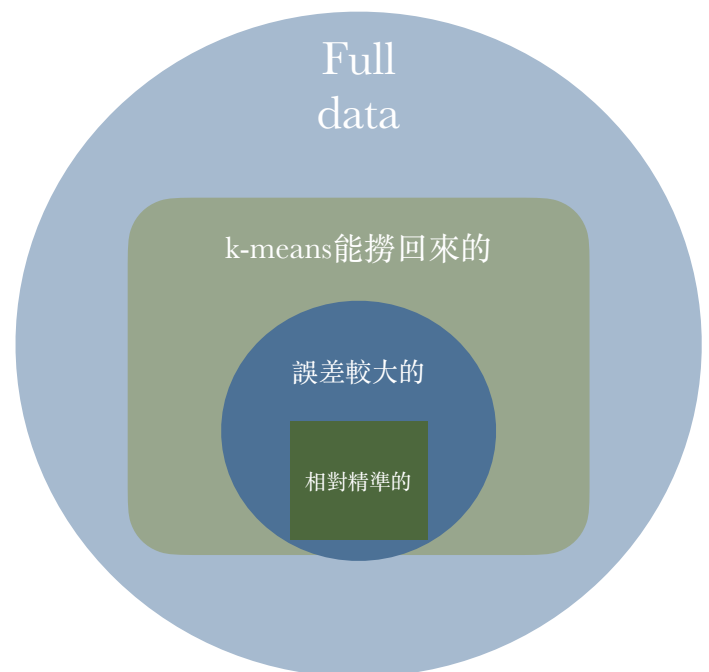
因為受訪者年紀可能偏大，導致有時候有可能記錯家庭成員的資訊，也有可能記錯女兒的順序等等，因此這部分會說明資料的嚴謹度。

首先，全部的資料約莫有10萬多筆，可以分成相對精準的，誤差稍大的，還有最後可以用k-means撈回來的資料，可以從下圖理解

相對精準的是使用subject原本回報的關係時，gender, age, education的誤差皆在一定範圍內（性別完全正對，出生年份誤差在(-3,4)之間，而教育程度落差在(-6,6)以內）。

誤差較大的組別是上面三者的誤差超過了給訂的標準，但把這個subject所有的親屬都跑過一次kmeans分組，分出來的結果跟subject回報的關係是一致的，因此還是可以採信subject回報的關係。

k-means組撈回來的，主要是特定年份誤差過大，或是同一年份有兩個同一項關係個觀察值出現的狀況的資料，例如2007年有三個五女出現，會依循k-means對這個subject分組後的index，來去看看這個index能否指出每一個年份都剛好有一個observation，且gender, age, education都是相似的。



--

在資料裡面：

- problem

0 相對精準組的資料

1 誤差較大的組別的額外資料

2 用kmeans的index撈回來的資料

所以如果要使用到撈回來的資料，當這個observation的 problem 是 {0, 1}，就可以直接使用原本回報的關係 (ID)。如果problem 是 {2}時，就需要使用 k_groups給出的 index作為追蹤此人的index（有點像是我們造出來的，用來追蹤的人造關係）

- youngs:

1: 這位家戶成員age<7 & education == NA，因此在做k_means時，分開來跑。

0: 非上述者

- age_missing :

1：這個家戶成員的年紀曾經是錯的，但用其他年對回來的

0：沒問題

unif 變數詳述：

如果family member 在gender, age, education 上，跨年份的差異不大，或是用 k_group分完後的結果跟原本subject回報的關係一致時， $unif = survey_year + qser_no + relation$.

如果family member 在gender age, education上差異過大，且k_group給出了一個更合理的答案時， $unif = survey + qser_no + k_group' relation$

—

如果family member 是屬於youngs時，在unif後面會再多一個young。

Residency：

Value	Label	Value	Label
1	台北市taipei city	19	屏東縣pingtung county
2	高雄市kaohsiung city	20	宜蘭縣yilan county
3	基隆市keelung city	21	花蓮縣hualien county
4	新竹市hsinchu city	22	台東縣taidung county
5	台中市taichung city	23	澎湖縣penghu county
6	嘉義市chiayi city		
7	台南市tainan city		
8	台北縣taipei county		
9	桃園縣taoyuan county		
10	新竹縣hsinchu county		
11	苗栗縣miaoli county		
12	台中縣taichung county		
13	彰化縣changhai county		
14	南投縣nantou county		
15	雲林縣yunlin county		
16	嘉義縣chiayi county		
17	台南縣tainan county		
18	高雄縣kaohsiung county		