

# Family member codebook

Please refer to csv : main\_subject\_xxxx.csv for the latest version

Remarks :

1. 1993年的 who \_help系列跟1996與之後的coding 差蠻多的，目前暫略，因此只有 {1996, 1999, 2003, 2007}

2. Spouses 的年齡在原本的資料內就有1/3的缺失，目前無法解決。

- qser\_no : 問卷樣本編號，建議用此去對 main subject e.g. A\_1231

- survey year : 問卷年份 {1996, 1999, 2003, 2007}

- relationship : 關係，請對照關係表

- gender : 生理性別 : 1男生, 2女生

- age: 年紀 :

- edu : 教育程度 :

0 不識字

1~6 小學 1~6年級

7~9 初中 1~3年級

10~12 高中 1~3年級

13~ 16 大學 1~4年級

17 研究所與以上

88 不適用-> 成員未滿6歲

90 識字

91 空中大學肄業

99 missing

- education : 教育程度（粗略） 僅分成0, 6, 9, 12, 16, 17 為主，可略

- 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）
8. 不適用
9. missing

- 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
88. 不適用 -> 家戶成員未滿15歲
9. missing

- 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
- 8 不適用
- 9 missing

- 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

8 不適用

9 missing

- 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

98, 99 missing

- tele 多久講一次電話

coding 同上

- adl\_who\_help 誰是主要adl的照顧者
- iadl\_who\_help 誰是主要的iadl 照顧者

88 = 不需要幫忙

0 = 需要幫忙但沒人幫忙

1= 自己

其他請參照關係表

— — —

### k-group說明

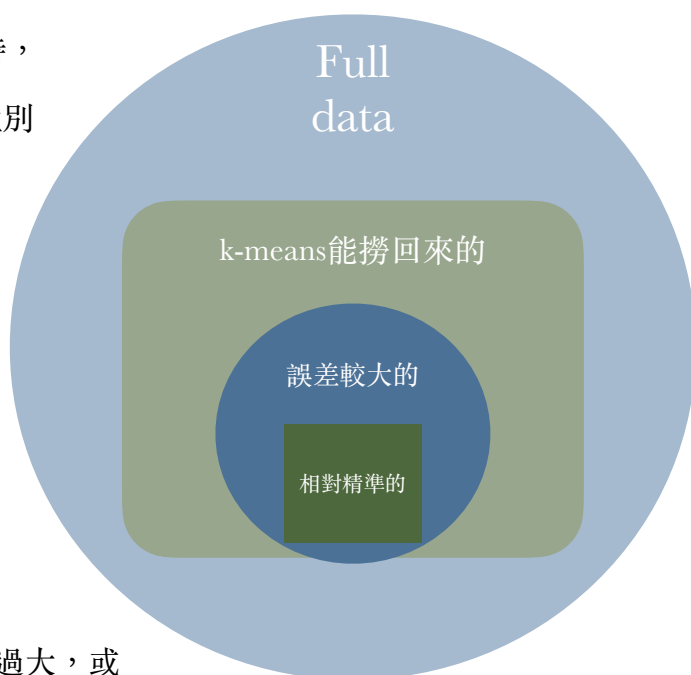
因為受訪者年紀可能偏大，導致有時候有可能記錯家庭成員的資訊，也有可能記錯女兒的順序等等，因此這部分會說明資料的嚴謹度。

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）