* 11/25 起始進度 HBP dataset已初步整理好

路徑：C:\Users\hsu\Desktop\master\TCHCData\hbp\_devtvs.csv

**待作清單：**

1. 決定預測目標(Y)
2. 挑選適合重複測量的插補資料方法 先看HBP欄位NA比例
3. 挑選變數
4. 輸入不同的input 比較模型結果

* 下次：

1. 決定預測目標(Y)

用hbp\_devtvs.csv裡的**檢測值欄位NA的比例**進行挑選

結果:

總筆數裡NA 比例最低的6個 (和血壓沒有關係的)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 變數 | HR | Total\_C | Creatinine | TG | LDL | glu\_AC |
| NA比(筆數)/有值筆數: | 44.49%  (1349)  1683 | 57.16%  (1733)  1299 | 57.49 %  (1743)  1289 | 58.91%  (1786)  1246 | 60.32 %  (1829)  1203 | 60.75 %  (1842)  1190 |

|  |  |  |  |
| --- | --- | --- | --- |
| 變數 \NA比例(人數)/有值筆數 | V1 Total 1378 | V2 Total 674 | V3 Total 346 |
| HR | 18.43% (254) 1124 | 67.8% (457) 217 | 17.63% (61) 285 |
| Total\_C | 32.15%(443)935 | 66.02%(445)229 | 64.45% (223) 123 |
| Creatinine | 33.02%(455)923 | 65.73%(443)231 | 63.87%(221)125 |
| TG | 34.03%(469)909 | 69.29$(467)207 | 65.90%(228)118 |
| LDL | 37.01%(510)868 | 68.25%(460)214 | 67.63%(234)112 |
| glu\_AC | 38.32%(528)457 | 67.8%(457)217 | 67.34%(233)113 |

和血壓有關係的NA比例最低的 :



1. 挑選適合重複測量的插補資料方法 先看**HBP欄位NA**比例

結果：

|  |  |  |  |
| --- | --- | --- | --- |
| HBP(總人數) | V1(1507) | V2(832) | V3(487) |
| 完整數值人數 | 368 | 223 | 124 |

7天\*4 =28 收縮 舒張共56筆 +平均6筆

56筆中超過3天沒有數值的(NA >12\*2)

|  |  |  |  |
| --- | --- | --- | --- |
| HBP(總人數) | V1(1507) | V2(832) | V3(487) |
| 超過3天沒有數值的 | 155 | 32 | 17 |