<DM-20200516>

做分析的過程(最重要的是前面的前置處理)

1. Preprocessing 前置處理
   1. Read data in R studio
      1. Missing value遺失值比例

那些變數有遺失值? 每個變數遺失值的比例? 畫出熱圖看遺失值分布

* + 1. 確認所有觀察變數(延伸變數不要放在這邊! 要呈現原始data)
  1. Re-codeing (DV) : 類別型? 數值型? 畫個圖
  2. Re-codeing(IV): 類別型? 數值型? 畫個圖

1. EDA
2. 假設檢定

有母數檢定 / 無母數檢定

交互驗證

資料可以隨機的方式分成訓練集與測試集，這個要反覆做很多次

可以將資料分成10個dataset，每個dataset輪流當測試集，其他的當訓練集

然後看一下跑出來的10個結果

<假設檢定 : 無母數檢定>

無母數檢定，就是不假設資料是常態分配

Sign test (31/43)

Wilcoxon signed rank test(paired配對資料)

Wilcoxon signed rank test(unpaired)

Mann-Whitney(兩筆獨立資料)