蔡宜諠 711378912 題目：#4.7

1. 這裡的資料要用標準化後的資料還是原始資料?

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

從這裡可以看的出來每個變數的變異程度較大，因此我們應該使用標準化後的資料，這樣可以讓每個變數在主成分分析中有相同的權重，也比較公平。所以後續的主成分分析採用**標準化後的資料**。#

1. 根據eigen-value-greater-than-one應該保留多少個主成分？

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

根據表 [ 相關矩陣的特徵值 ]，保留特徵值大於1的主成分( 3.9290, 1.6183 )，應該保留下**兩個主成分**。#

1. 如果保留四個主成分，能解釋的總變異量百分比是多少？

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

根據表 [ 相關矩陣的特徵值 ] 中的累積變異百分比，可以看到前四個主成分總共**解釋了 87.37% 的變異**。#

1. 為第一個主成分命名

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

根據表 [ 特徵向量 ]，可以看到L500、L1000和R1000這些變數的係數相對較大，且全部都是正數，其實不是太好命名，但這裡我想出最好的名字大概就是**中低頻率損失變異**。#

1. 為第四個主成分命名。

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

根據表 [ 特徵向量 ]，可以看到L1000和R500這些變數的係數相對較大，但因為L1000為負值，L1000 和 R500 之間在 PRIN4 上存在反向的特性，還可以看到所有的L都是負數、所有的R都是正數，因此在這裡命名為**左右耳中低頻率變異。#**

1. 哪個 ID 的第一個主成分分數最大？

根據Prin1去排序後的結果輸出為下圖。

一張含有 文字, 螢幕擷取畫面, 數字, 正方形 的圖片

自動產生的描述

因此可以看到第一個主成分分數( 5.48987 )是最大的**ID為55**。#

補充：

一張含有 文字, 螢幕擷取畫面, 數字, 字型 的圖片

自動產生的描述

從這個升序的表中可以看出來ID為39的的第一個主成分分數( 5.48987 )是最小的。

一張含有 行, 圖表, 繪圖, 文字 的圖片

自動產生的描述

從此圖可以看出第一個主成分開始快速下降，在前兩個主成分可以解釋大約 70% 的變異，而前四個主成分可以解釋超過 85% 的變異，而後面的變異就趨於穩定，且帶來的解釋變異有限。我們可以保留2到4個主成分獲得較高的解釋力，也不會增加後續建立模型的負擔。