

商業分析：SAS / R HW4

108208004 經濟三 白植允

分析財務報表，了解公司企業特性，以降維的方式，找出有意義的指標，衡量績優公司。資料financialdata.csv 有 163 間公司的財務指標。

1. 以 PCA 或 SPCA 分析，找出每個主成份能解釋多少變異？大概需要多少個 PC 來解釋這筆資料？

```
library(reshape2)
library(ggplot2)
library(stats)
library(nsprcomp)
library(base)
library(ggfortify)
library(tidyverse)
data = read.csv("financialdata.csv",sep=",")
for(i in c(12,15,16)){
  for(j in 1:nrow(data)){
    data[j,i] = gsub('[,]', '', data[j,i])
  }
  data[,i] = as.numeric(data[,i])
}
```

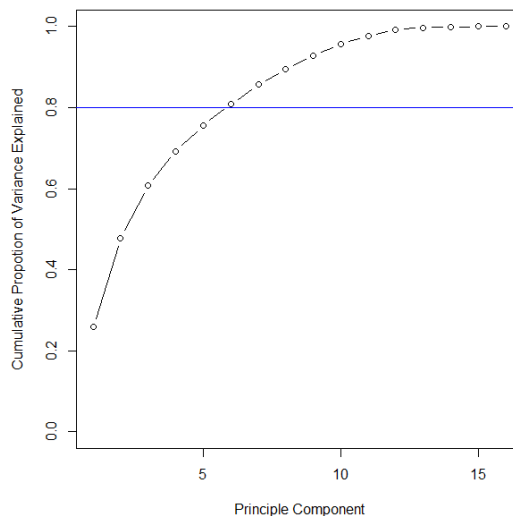
=>將需要的 library 列出以及將資料整理

```
pca <- prcomp(data[,-1],center = T,scale = T)
screeplot(pca)
summary(pca)
```

Importance of components:	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16
Standard deviation	2.0330	1.8739	1.4399	1.16215	1.01323	0.91896	0.87186	0.77645	0.73927	0.67853	0.56103	0.50353	0.27912	0.15178	0.12692	0.07223
Proportion of Variance	0.2583	0.2195	0.1296	0.08441	0.06417	0.05278	0.04751	0.03768	0.03416	0.02878	0.01967	0.01585	0.00487	0.00144	0.00101	0.00033
Cumulative Proportion	0.2583	0.4778	0.6074	0.69177	0.75594	0.80872	0.85623	0.89391	0.92806	0.95684	0.97651	0.99236	0.99723	0.99867	0.99967	1.00000

=>每個 pc 能解釋的變異如上圖「Proportion of Variance」所示

```
pve =(pca$sdev)^2/sum(pca$sdev^2)
plot(cumsum(pve),xlab="Principle Component", ylab="Cumulative Proportion of Variance Explained",ylim=c(0,1),type = 'b')
abline(h=0.8,col = "blue")
```

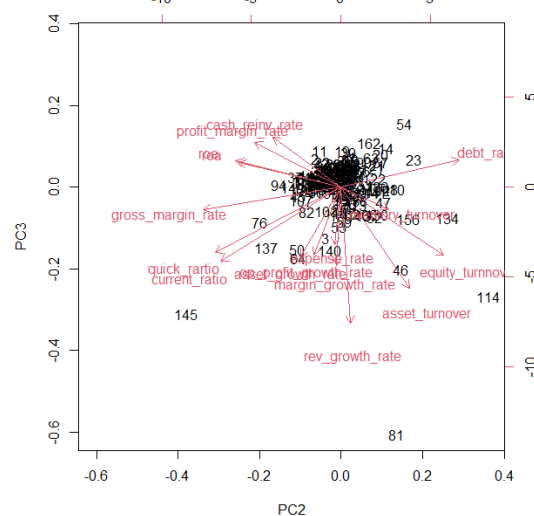
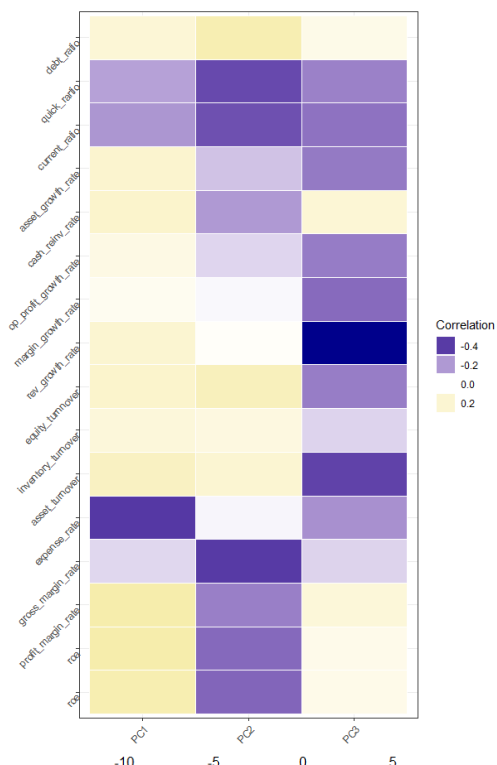


=>大約需 6 個才能解釋 80%的變異

2. 找出前三個主成份分別重點變數為何並解釋。

```
ggplot(melt(pca$rotation[,1:3]), aes(Var2, Var1)) +
  geom_tile(aes(fill = value), colour = "white") +
  scale_fill_gradient2(low = "darkblue", high = "khaki",
    mid = "white", midpoint = 0) +
  guides(fill = guide_legend(title = "Correlation")) +
  theme_bw() +
  theme(axis.text = element_text(angle = 45, hjust = 1, vjust = 1),
    axis.title = element_blank())
```

```
autoplot(prcomp(data, center = T, scale = T), x = 1, y = 2,
  loadings = T, loadings.colour = "red",
  loadings.label = T)
biplot(pca, scale = T, choices = 2:3)
```



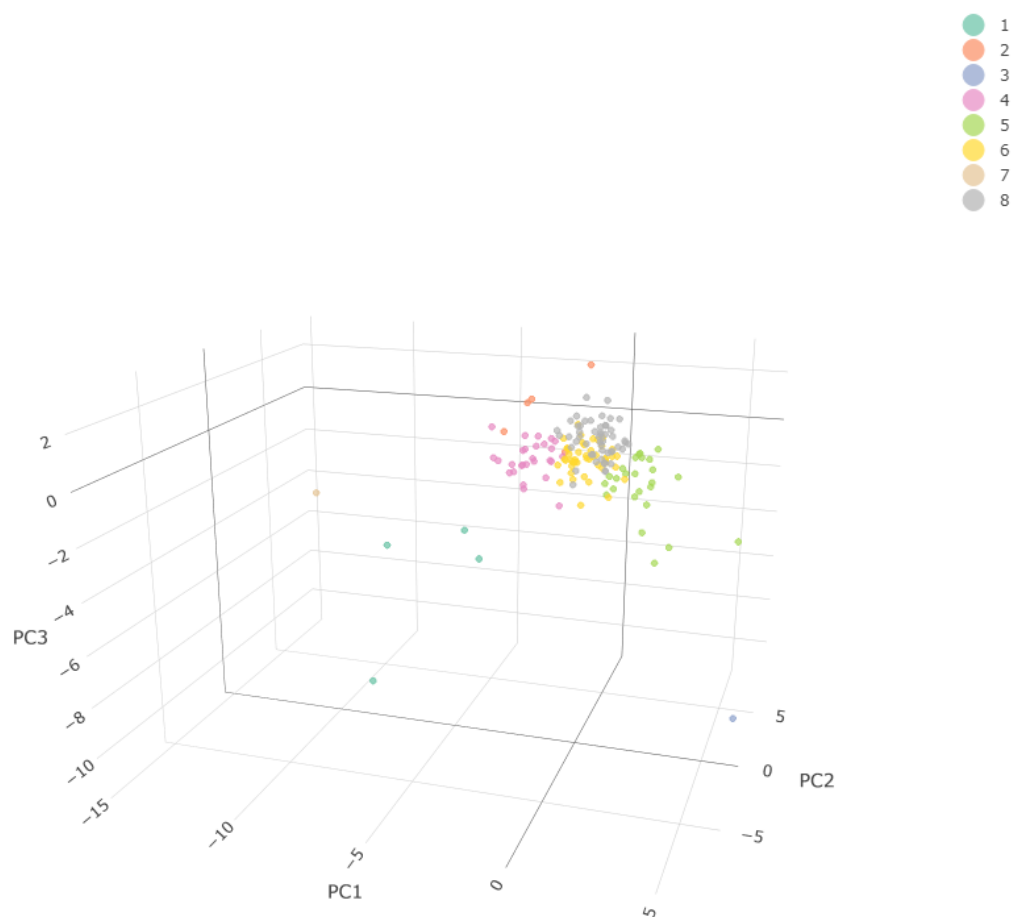
- PC1 的重點變數是「營業費用率，營業利益率，股東權益報酬率，資產報酬率」，可用來關注這家公司對股東權益的影響，用上述指標能看出一家公司獲利的能力。
- PC2 的重點變數是「營業毛利率，速動比率，流動比例，負債比率，淨值週轉率」，可用來觀察一家公司的償債能力，以他的毛利率與其他債務負擔的比率來看這公司的風險高低。
- PC3 的重點變數是「營收成長率，資產週轉率」，這些變數可用來觀測一家公司的資產管理以及營收狀況，看出這家公司運用資產的效率。
- PC1 數字越大，代表公司獲利能力越好，越值得投資。PC2 數字越大，代表公司倒債風險越高，需要注意。PC3 數字越小，資產投資創造的營收效率越好，越值得投資。

3. 找出適合投資的公司。

```
data2 <- cbind(data,pca$x)
```

```
> head(data2[order(data2$PC1,decreasing = T),]$comp_id)
[1] 3219 4967 6462 6684 8084 6238
> head(data2[order(data2$PC2,decreasing = T),]$comp_id)
[1] 5250 6291 3690 8084 8351 4967
> head(data2[order(data2$PC3,decreasing = F),]$comp_id)
[1] 3219 6533 5250 4967 2363 6462
```

=>3219 的獲利能力最高且資產運用效率很好，而 5250 雖然目前營收指數不是前六名，但未來目前資金利用效率也不錯，且倒債風險很低，是值得持續關注的公司，4967 的營運狀況也良好，雖然獲利能力可能低於 3219，但是是屬於風險較低的公司，其他榜上的公司可以依各項指標做投資選擇。



=>可以到這裡看 3D 圖:<https://rpubs.com/Dragalia2001/897858>

附錄:程式碼

```
library(reshape2)
library(ggplot2)
library(stats)
library(nsprcomp)
library(base)
library(ggfortify)
library(tidyverse)

data = read.csv("financialdata.csv", sep=",")
for(i in c(12,15,16)){
  for(j in 1:nrow(data)){
    data[j,i] = gsub('[,]', '', data[j,i])
  }
  data[,i] = as.numeric(data[,i])
}

pca <- prcomp(data[, -1], center = T, scale = T)
screeplot(pca)
summary(pca)

pve = (pca$sdev)^2 / sum(pca$sdev^2)
plot(cumsum(pve), xlab="Principle Component", ylab="Cumulative
Proportion of Variance Explained", ylim=c(0,1), type = 'b')
abline(h=0.8, col = "blue")

ggplot(melt(pca$rotation[, 1:3]), aes(Var2, Var1)) +
  geom_tile(aes(fill = value), colour = "white") +
  scale_fill_gradient2(low = "darkblue", high = "khaki",
                       mid = "white", midpoint = 0) +
  guides(fill=guide_legend(title = "Correlation")) +
  theme_bw() +
  theme(axis.text = element_text(angle=45, hjust = 1, vjust = 1),
        axis.title = element_blank())

autoplot(prcomp(data, center = T, scale = T), x=1, y=2,
         loadings = T, loadings.colour = "red",
```

```
loadings.label = T)

biplot(pca,scale = T,choices = 2:3)

data2 <- cbind(data,pca$x)

head(data2[order(data2$PC1,decreasing = T),]$comp_id)
head(data2[order(data2$PC2,decreasing = T),]$comp_id)
head(data2[order(data2$PC3,decreasing = F),]$comp_id)

k <- kmeans(data2[,18:20],centers = 8)
group <- as.factor(k$cluster)
data3 <- cbind(data2[,c(1,18:20)],group)
library(plotly)
plot_ly(data3, x = ~PC1, y = ~PC2, z = ~PC3,color = group,
size=20,text = ~paste("Comp_id:",comp_id)) %>%
  add_markers() %>%
  layout(scene = list(xaxis = list(title = 'PC1'),
                        yaxis = list(title = 'PC2'),
                        zaxis = list(title = 'PC3')))
```