

作業四

109304019 統計四周佳萱

一、程式碼：

```
setwd("C:/Users/周佳萱/Desktop/商業分析/HW4")
data = read.csv("financialdata.csv")
finance = data[,-1]      ##去掉 id
str(finance)

#將字串轉成數字
op_new <- gsub(",", "", finance$op_profit_growth_rate)
finance$op_profit_growth_rate = as.numeric(op_new)
cr_new <- gsub(",", "", finance$current_ratio)
finance$current_ratio = as.numeric(cr_new)
qr_new <- gsub(",", "", finance$quick_ratio)
finance$quick_ratio = as.numeric(qr_new)

library(stats)
pca<- prcomp(finance, center = TRUE, scale = TRUE)
var = pca$sdev^2
prop = (pca$sdev)^2 / sum((pca$sdev)^2)

#screeplot
screeplot(pca)
abline(h=1, col="blue")

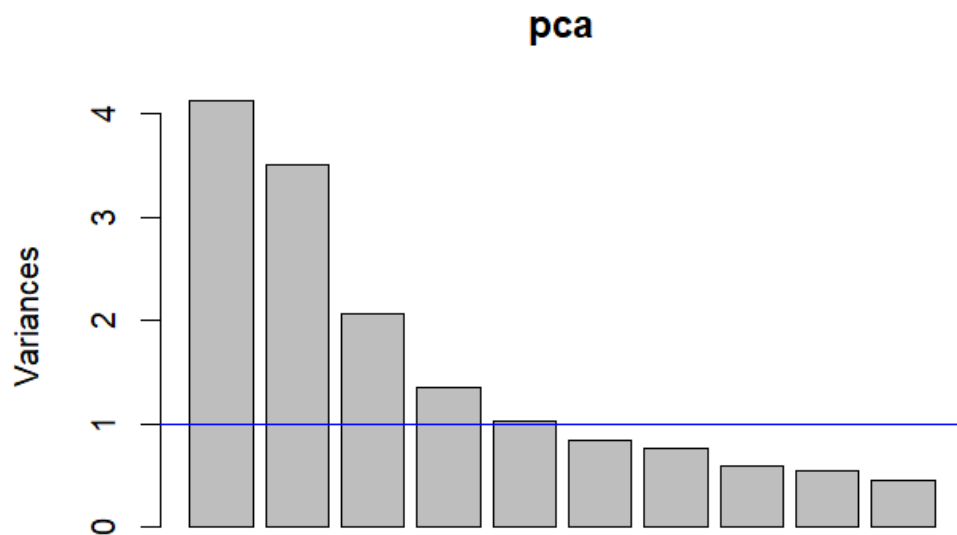
#2.
library(reshape2)
library(tidyverse)
pca$rotation
ggplot(melt(pca$rotation[,1:3]), aes(Var2, Var1)) +
  geom_tile(aes(fill = value), colour = "white") +
  scale_fill_gradient2(low = "firebrick4", high = "steelblue",
                       mid = "white", midpoint = 0) +
  guides(fill=guide_legend(title="Coefficient")) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1),
        axis.title = element_blank())
```

二、結果敘述：

1.

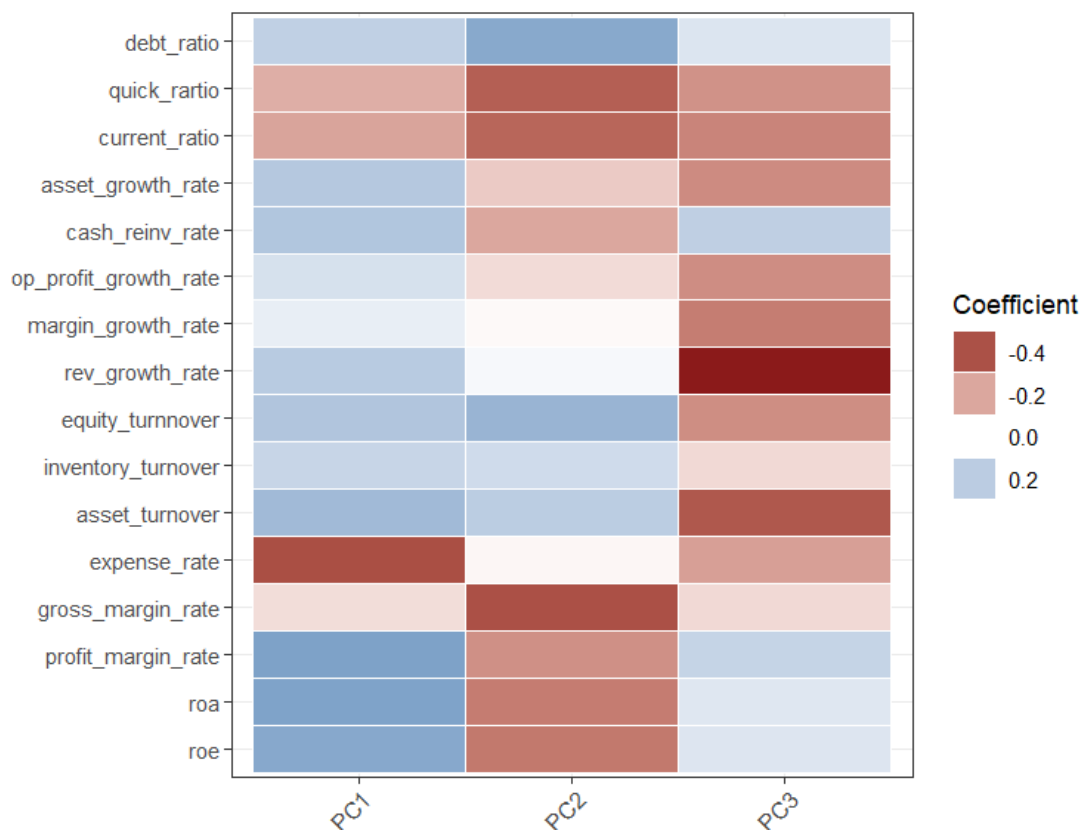
每個主成分能解釋的變異如下：

| PC1 | PC2 | PC3 | PC4 | PC5 |
|--------------|--------------|--------------|--------------|--------------|
| 0.2583148273 | 0.2194629830 | 0.1295813724 | 0.0844127491 | 0.0641653079 |
| PC6 | PC7 | PC8 | PC9 | PC10 |
| 0.0527805214 | 0.0475085399 | 0.0376798772 | 0.0341578302 | 0.0287752595 |
| PC11 | PC12 | PC13 | PC14 | PC15 |
| 0.0196723360 | 0.0158464364 | 0.0048692772 | 0.0014397919 | 0.0010068411 |
| PC16 | | | | |
| 0.0003260494 | | | | |



上圖為 **screeplot**，可以發現前 5 的 PC 的 **variance** 大於 1，所以大約需要 5 個 PC 才能解釋資料。

2.



PC1 的重點變數為 ROE、ROA、profit margin rate、expense rate，多為衡量公司淨利的指標，因此命名 PC1 為公司獲利能力。

PC2 的重點變數為 gross margin rate、current ratio、quick ratio、debt ratio，多為衡量公司可以償還負債的指標，因此命名 PC2 為公司的財務償還能力。

PC3 的重點變數為 revenue growth rate、asset turnover，前者為衡量公司營收增加的指標，後者為衡量公司資產周轉能力，因此命名 PC3 為公司成長能力。

3.

根據第一主成分的結果，公司適合投資股東權益報酬率高、資產報酬率高、營業利益率高、營業費用率低的公司。