

# SAS/R商業資料分析作業四

107508006 歐西四 陳葳芃

## 財務指標資料分析 Analysis of financial indexes

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？

Ans:

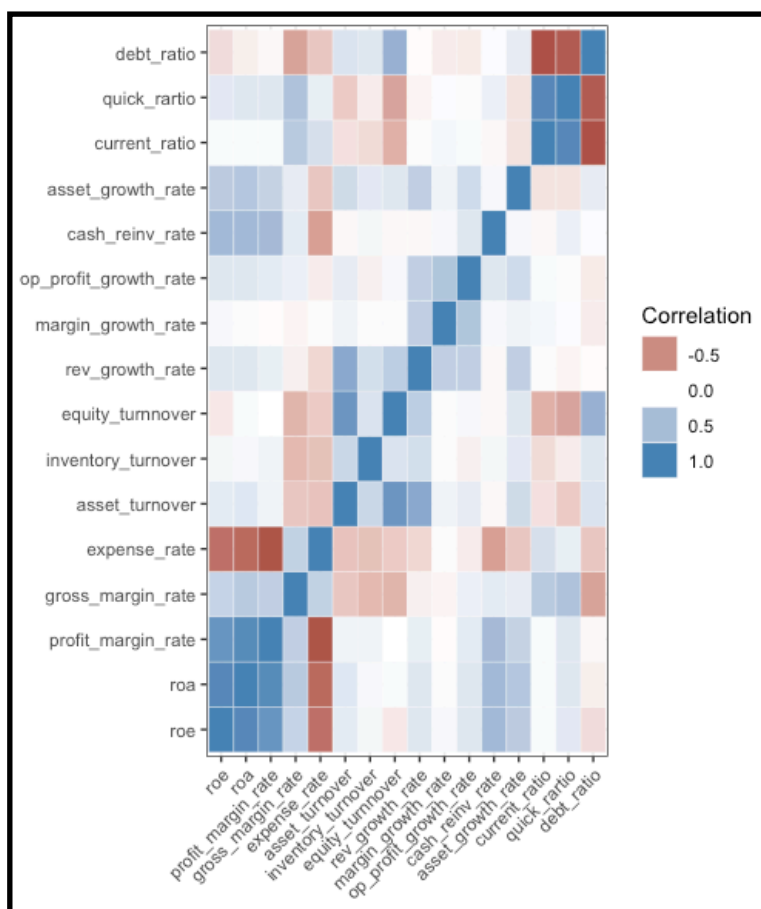
```
> #資料financialdata.csv 有163間公司的財務指標，
> setwd("~/Downloads/1102 R/HW/hw 4")
> library(readr)
> data.fin <- read.csv("financialdata.csv")
> ##先將data變數整理為numeric
> str(data.fin)
'data.frame': 162 obs. of 17 variables:
 $ comp_id      : int  2303 2330 2337 2342 2344 2408 3707 5346 5347 6287 ...
 $ roe          : num  3.06 23.56 25.68 -3.41 10.9 ...
 $ roa          : num  2.21 17.84 14.29 -0.72 7.69 ...
 $ profit_margin_rate : num  4.4 39.45 16.82 3.86 13.99 ...
 $ gross_margin_rate : num  18.1 50.6 37 16.9 34.3 ...
 $ expense_rate   : num  14.8 11 20.1 13 20.3 ...
 $ asset_turnover : num  0.38 0.5 0.86 0.66 0.61 0.38 0.82 0.77 0.72 0.81 ...
 $ inventory_turnover : num  6.93 7.88 2.54 8.87 3.99 5.16 5.64 4.66 6.79 3.73 ...
 $ equity_turnover : num  0.69 0.67 1.59 1.55 0.89 0.5 1.43 1.45 0.88 4.16 ...
 $ rev_growth_rate : num  0.96 3.11 41.75 12.73 13.07 ...
 $ margin_growth_rate : num  -10.93 4.21 116.49 13.25 35.83 ...
 $ op_profit_growth_rate: chr  "6.05" "2.01" "1,708.73" "22.41" ...
 $ cash_reinv_rate : num  4.66 11.06 5.31 0.56 5.91 ...
 $ asset_growth_rate : num  1.93 5.59 24.33 17.54 29.6 ...
 $ current_ratio   : chr  "158.03" "238.97" "187.85" "184.99" ...
 $ quick_ratio     : chr  "118.71" "215.17" "110.85" "158.19" ...
 $ debt_ratio      : num  45.7 23.6 44.2 51.4 30.1 ...
> data.fin <- data.fin[,-1]
> data.fin$op_profit_growth_rate <- as.numeric(data.fin$op_profit_growth_rate)
警告訊息：
強制變更過程中產生了 NA
> data.fin$current_ratio <- as.numeric(data.fin$current_ratio)
警告訊息：
強制變更過程中產生了 NA
> data.fin$quick_ratio <- as.numeric(data.fin$quick_ratio)
警告訊息：
強制變更過程中產生了 NA
> data.fin <- na.omit(data.fin)
```

>>說明：產生na的部分亦使用na omit來使分析順利繼續運行下去。

```

> ##計算correlation
> M = cor(data.fin)
> ##heatmap for correlation
> library(tidyverse)
> library(reshape2)
> melted_cormat <- melt(M)
> head(melted_cormat)
      Var1 Var2      value
1      roe roe  1.0000000
2      roa roe  0.9280363
3 profit_margin_rate roe  0.8313921
4 gross_margin_rate roe  0.3125247
5  expense_rate roe -0.6238273
6  asset_turnover roe  0.1469624
> ggplot(data = melted_cormat,
+       aes(Var1, Var2)) +
+   geom_tile(aes(fill = value), colour = "white") +
+   scale_fill_gradient2(low = "firebrick4", high = "steelblue",
+   mid = "white", midpoint = 0) +
+   guides(fill=guide_legend(title="Correlation")) +
+   theme_bw() +
+   theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1),
+   axis.title = element_blank())

```



>>根據熱度圖的結果來分析:

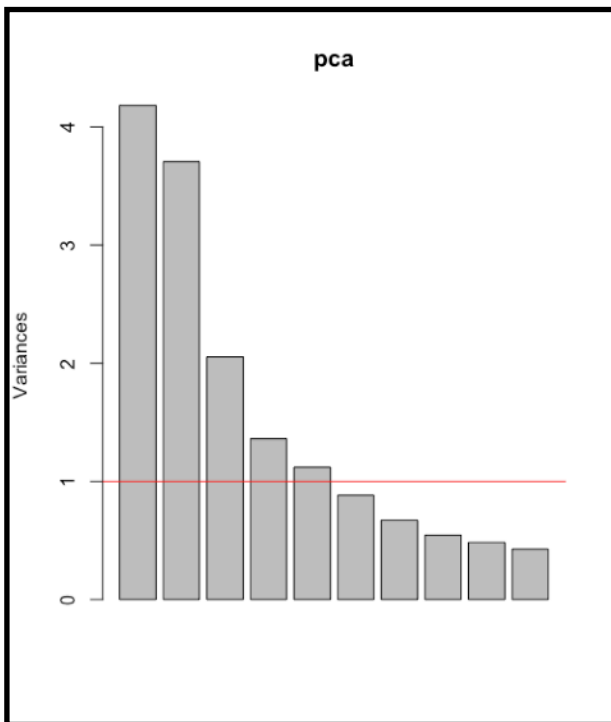
很明顯可以發現前三個變數 “roe”(股東權益報酬率)、“roa”(資產報酬率)、“profit\_margin\_rate”(營業利益率)，兩兩之間各自呈現「高度正相關」;而較明顯呈現「高度負相關」的則是有三組“expense\_rate”(營業費用率)跟“profit\_margin\_rate”(營業利利益率);“debt\_ratio”(負債比例)跟“current\_ratio”(流動比例);“debt\_ratio”(負債比例)跟“quick\_ratio”(速度比例)。

```

> #####2.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(stats)
> pca<- prcomp(data.fin, center = TRUE, scale = TRUE)
> names(pca)
[1] "sdev"      "rotation"  "center"    "scale"     "x"
> summary(pca)
Importance of components:
              PC1      PC2      PC3      PC4      PC5      PC6      PC7      PC8      PC9      PC10     PC11     PC12
Standard deviation  2.0452  1.9253  1.4329  1.1660  1.0574  0.9385  0.8178  0.7379  0.6954  0.6513  0.5427  0.3864
Proportion of Variance 0.2614 0.2317 0.1283 0.0849 0.0698 0.0550 0.0418 0.0340 0.0302 0.0265 0.0184 0.0093
Cumulative Proportion 0.2614 0.4931 0.6214 0.7064 0.7762 0.8313 0.8731 0.9071 0.9374 0.9639 0.9823 0.9916

              PC13      PC14      PC15      PC16
Standard deviation  0.2466 0.2120 0.1458 0.0795
Proportion of Variance 0.0038 0.0028 0.0013 0.0004
Cumulative Proportion 0.9955 0.9982 0.9996 1.0000
> ##大概需要多少5個PC (eigenvalue>1)
> plot(pca) #Variation explained
> abline(h=1, col="red")

```



>>我們可以根據上圖“**Proportion of Variance**”這列的數據來知道每個PC能解釋的變異百分比。我先使用`plot(pca)`把數據視覺化，再透過尋找`eigenvalue>1`的變數，從上圖可以發現，我們大概需要5個PC。

2. 找出前三個主成份分別重點變數為何並解釋。

Ans:

```
> #####3.找出前三個主成份分別重點變數為何並解釋。
> pca.factor <- pca$rotation[,1:3]
> pca.factor[,1:3]
      PC1      PC2      PC3
roe      0.43179880 -0.135780342  0.10910989
roa      0.45023626 -0.124980824  0.10865656
profit_margin_rate 0.43667346 -0.118750815  0.16821679
gross_margin_rate  0.08758386 -0.353245245 -0.04116377
expense_rate      -0.37766540 -0.115803991 -0.19335518
asset_turnover     0.18561418  0.283964640 -0.37030391
inventory_turnover 0.10733895  0.177699324 -0.06297749
equity_turnover    0.10712323  0.380454797 -0.18459864
rev_growth_rate    0.19407798  0.141146622 -0.50295002
margin_growth_rate 0.06509033  0.009712982 -0.36530167
op_profit_growth_rate 0.15519903 -0.014754966 -0.33368481
cash_reinv_rate    0.27745978 -0.087433094  0.19521499
asset_growth_rate   0.25778348  0.073934826 -0.15330787
current_ratio      -0.03806725 -0.404308310 -0.29428980
quick_ratio        0.02013723 -0.427262319 -0.19197578
debt_ratio          0.03348174  0.421101563  0.22565558
> #####
> pc_var=list()
> for (i in 1:3){
+   order=order(abs(pca.factor[,i]),decreasing = TRUE)
+   pc_var[[i]]=pca.factor[,i][order][1:3]
+ }
> pc_var[1]
[[1]]
      roa profit_margin_rate      roe
0.4502363      0.4366735      0.4317988
> pc_var[2]
[[1]]
quick_ratio      debt_ratio current_ratio
-0.4272623      0.4211016      -0.4043083
> pc_var[3]
[[1]]
rev_growth_rate      asset_turnover margin_growth_rate
-0.5029500      -0.3703039      -0.3653017
```

【PC1重點變數】 roe、roa、profit\_margin\_rate、expense\_rate

我認為在PC1是跟公司會不會賺錢以及成本營運的費用有關係，所以推測在影響投資者最重要的因素就是「公司的營運花費及公司會不會賺錢」。

【PC2重點變數】 quick\_ratio、debt\_ratio、current\_ratio

我認為在PC2是著重在財務效能與企業經營的安定性，因當流動比率與速動比率呈現下降趨勢，就要小心企業以短支長狀況經營，所以推測在影響投資者第二重要的因素就是「公司的營運效能及公司是否穩定的發展」。

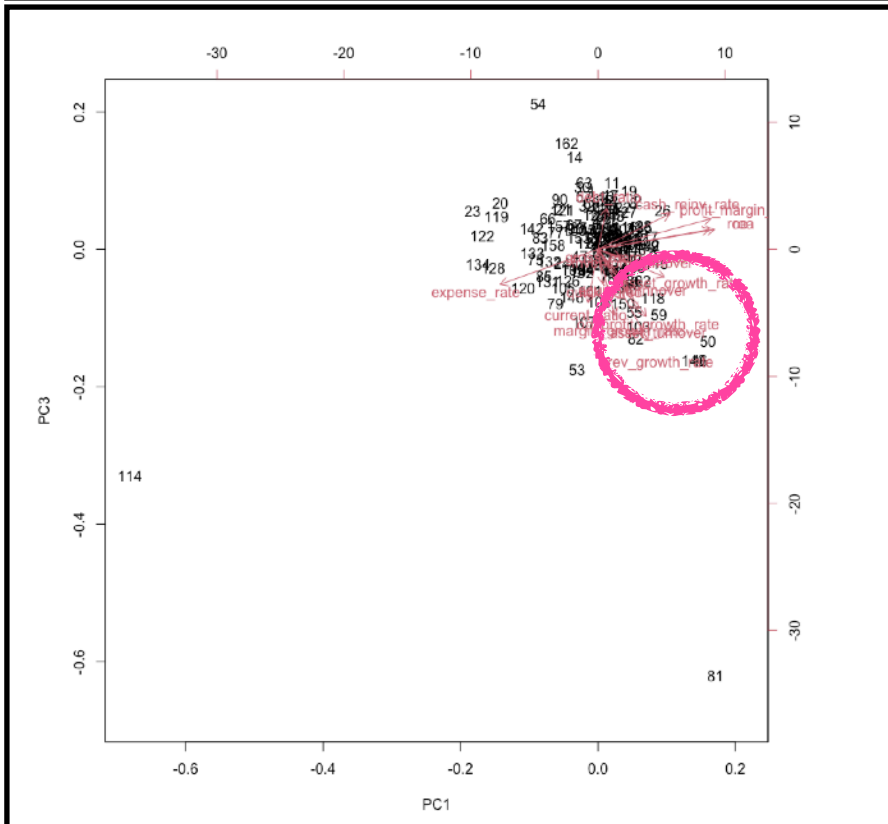
【PC3重點變數】 rev\_growth\_rate、set\_turnover、margin\_growth\_rate

我認為在PC3是跟公司「成長表現」有關係，所以推測在影響投資者第三重要的因素就是「公司營運成長狀況」。

3. 找出適合投資的公司。

Ans:

```
> #####4.找出適合投資的公司。  
> biplot(pca,scale = T,choices = c(1,3))  
> |
```



>>由於PC1(公司會不會賺錢)指標應為越大越好、而PC3(公司營運成長狀況)，本因是越大越好，但當初此主成分3個均為負向組成，故本指標應為越小越好。所以會建議投資越靠近「biplot圖」出來右下方的公司越好(粉色圈圈部分)。

## 附錄: R 程式碼

#HW4

#資料financialdata.csv 有163間公司的財務指標，  
setwd("~/Downloads/1102 R/HW/hw 4")

```
library(readr)
data.fin <- read.csv("financialdata.csv")
#####1.請先將資料探索，計算變數間correlation，並畫熱圖呈現。解釋哪個變數高度正相關或負相關。
```

```
##先將data變數整理為numeric
str(data.fin)
data.fin <- data.fin[,-1]
data.fin$op_profit_growth_rate <-
as.numeric(data.fin$op_profit_growth_rate)
data.fin$current_ratio <- as.numeric(data.fin$current_ratio)
data.fin$quick_rartio <- as.numeric(data.fin$quick_rartio)
```

```
data.fin <- na.omit(data.fin)
```

```
##計算correlation
M = cor(data.fin)
```

```
##heatmap for correlation
library(tidyverse)
library(reshape2)
melted_cormat <- melt(M)
head(melted_cormat)
```

```
ggplot(data = melted_cormat,
       aes(Var1, Var2)) +
  geom_tile(aes(fill = value), colour = "white") +
  scale_fill_gradient2(low = "firebrick4", high = "steelblue",
                      mid = "white", midpoint = 0) +
  guides(fill=guide_legend(title="Correlation")) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust =
1),
        axis.title = element_blank())
```

```
#####2.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(stats)
pca<- prcomp(data.fin, center = TRUE, scale = TRUE)
names(pca)
```

```

summary(pca)

##大概需要多少5個PC (eigenvalue>1)
plot(pca) #Variation      explained
abline(h=1, col="red")

#####3.找出前三個主成份分別重點變數為何並解釋。
pca.factor <- pca$rotation[,1:3]
pca.factor[,1:3]

#####
pc_var=list()
for (i in 1:3){
  order=order(abs(pca.factor[,i]),decreasing = TRUE)
  pc_var[[i]]=pca.factor[,i][order][1:3]
}

pc_var[1]
pc_var[2]
pc_var[3]

ggplot(melt(pca.factor[,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()))

#####4.找出適合投資的公司。
biplot(pca,scale = T, choice= c(1,3)

```