

110307009 林可婷

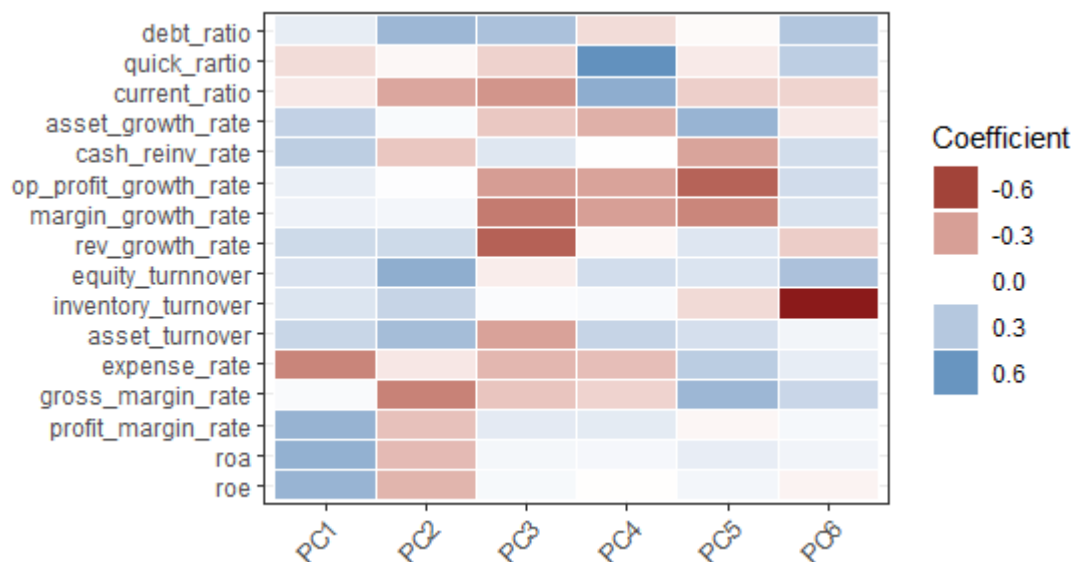
1 以 PCA 或 SPCA 分析，找出每個主成份能解釋多少變異？大概需要多少個 PC 來解釋這筆資料？

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
Standard deviation	2.0087	1.7355	1.3291	1.23626	1.13379	0.98397	0.85364	0.81208	0.75506	0.68057	0.60506
Proportion of Variance	0.2522	0.1883	0.1104	0.09552	0.08034	0.06051	0.04554	0.04122	0.03563	0.02895	0.02288
Cumulative Proportion	0.2522	0.4404	0.5508	0.64635	0.72669	0.78720	0.83274	0.87396	0.90959	0.93854	0.96142
	PC12	PC13	PC14	PC15	PC16						
Standard deviation	0.55798	0.44436	0.28269	0.15290	0.07173						
Proportion of Variance	0.01946	0.01234	0.00499	0.00146	0.00032						
Cumulative Proportion	0.98088	0.99322	0.99822	0.99968	1.00000						

每個主成分能解釋的變異要看 **proportion of variance**，大概需要六個或七個 PC 來解釋這筆資料。

2.找出前三個主成份分別重點變數為何並解釋。



PC1 的三個重點變數是 profit margin rate,roa,roe 越高，expense rate 越低

PC2 第一名的重點變數是 quity turnover，其它前幾名的重要變數有 asset turnover,debt ratio，然後 gross margin rate 越低越好

PC3 的第一名重點變數是 rev growth rate 越低越好，其它前幾名的重要變數有 margin gross rate,op profit growth rate,current ratio 也是越低越好

3.找出適合投資的公司。(不需指出是哪間公司，只需依第一主成份結果說明，例如：適合投資資產報酬率高的公司)

適合投資 profit margin rate,roa,roe 高，expense rate 低的公司

```

library(reshape2)
library(stats)
library(ggplot2)

financial<- read.csv("financialdata.csv")
data<-financial[,-1]

data$op_profit_growth_rate <- as.numeric(as.factor(data$op_profit_growth_rate ))
data$current_ratio <- as.numeric(as.factor(data$current_ratio ))
data$quick_rartio <- as.numeric(as.factor(data$quick_rartio ))

pca<- prcomp(data, center = TRUE, scale = TRUE)
names(pca) #sdev=std. deviation, rotation=coefficient 主成份的係數矩陣
summary(pca)
plot(pca)

## how to get the results of summary(pca)
var = pca$sdev^2 #該主成份解釋變異數的數值
prop = (pca$sdev)^2 / sum((pca$sdev)^2) #該主成份解釋變異數的比率
cum_prop = cumsum((pca$sdev)^2 / sum((pca$sdev)^2)) #該主成份解釋變異數的
累積比率

## scree plot: variance
screeplot(pca) #same as plot(pca)
plot(pca, type="line")
abline(h=1, col="blue") #Kaiser eigenvalue-greater-than-one rule, choose pc1~pc5 by
Kaiser

summary(pca)

## Rotation matrix: Loadings are the percent of variance explained by the variable
pca$rotation

```

```
#visualize
ggplot(melt(pca$rotation[,1:6]), 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())
```