

## 0. 匯入要用到的資料和套件 並確認檔案內容

```
setwd("C:\\Users\\ASUS\\Desktop\\五 234 R\\HW4") #放你的路徑
```

```
financialData <- read.csv("financialdata.csv")
```

```
financialData
```

	comp_id	roe	roa	profit_margin_rate	gross_margin_rate	expense_rate	asset_turnover	inventory_turnover	equity_turnover	rev_growth_rate
1	2303	3.06	2.21	4.40	18.12	14.83	0.38	6.93	0.69	0.96
2	2330	23.56	17.84	39.45	50.62	11.04	0.50	7.88	0.67	3.11
3	2337	25.68	14.29	16.82	36.95	20.12	0.86	2.54	1.59	41.75
4	2342	-3.41	-0.72	3.86	16.86	13.00	0.66	8.87	1.55	12.73
5	2344	10.90	7.69	13.99	34.30	20.31	0.61	3.99	0.89	13.07
6	2408	37.01	28.27	34.22	44.87	10.66	0.38	5.16	0.50	31.91
7	3707	-1.23	-0.57	0.68	10.68	10.00	0.82	5.64	1.43	20.46
8	5346	25.31	13.90	22.26	31.66	10.54	0.77	4.66	1.45	10.68
9	5347	15.93	13.01	20.99	32.01	11.02	0.72	6.79	0.88	-3.56
10	6287	-29.84	-3.89	-4.62	0.15	4.77	0.81	3.73	4.16	11.05

	margin_growth_rate	op_profit_growth_rate	cash_reinv_rate	asset_growth_rate	current_ratio	quick_rartio	debt_ratio
1	-10.93	6.05	4.66	1.93	158.03	118.71	45.69
2	4.21	2.01	11.06	5.59	238.97	215.17	23.55
3	116.49	1,708.73	5.31	24.33	187.85	110.85	44.21
4	13.25	22.41	0.56	17.54	184.99	158.19	51.42
5	35.83	79.26	5.91	29.60	229.31	173.05	30.06
6	91.77	119.72	7.94	10.08	424.44	366.56	12.38
7	228.37	110.18	0.44	22.17	237.72	175.75	41.66
8	24.59	31.03	4.03	21.57	83.4	41.99	46.96
9	-10.65	-14.3	1.06	-1.95	436.4	382.42	18.72
10	102.23	67.78	1.15	14.37	102.94	66.65	84.31

```
# 把第一個變數 id 刪掉
```

```
data = financialData[,-1]
```

```
data$op_profit_growth_rate <- as.numeric(data$op_profit_growth_rate)
```

```
data$current_ratio <- as.numeric(data$current_ratio)
```

```
data$quick_rartio <- as.numeric(data$quick_rartio)
```

```
# 處理 na (用平均值代替)
```

```
mean_col1 <- mean(data[, 11], na.rm = T) # column 1 的平均數(na.rm = T 忽略遺失值)
```

```
na_c1rows <- is.na(data[, 11]) # column 1 中，有遺失值存在的資料(回傳 TRUE/FALSE)
```

```
mean_col2 <- mean(data[, 14], na.rm = T) # column 3
```

```
na_c2rows <- is.na(data[, 14])
```

```
mean_col3 <- mean(data[, 15], na.rm = T) # column 3
```

```
na_c3rows <- is.na(data[, 15])
```

```
# 用 mean( column 1 )，填補第一欄位的遺漏值
```

```
data[na_c1rows, 11] <- mean_col1
```

```
data[na_c2rows, 14] <- mean_col2
```

```
data[na_c3rows, 15] <- mean_col3
```

```
# 變成矩陣
```

```
M = cor(data)
```

```
# 畫熱點圖
```

```
library(reshape2)
```

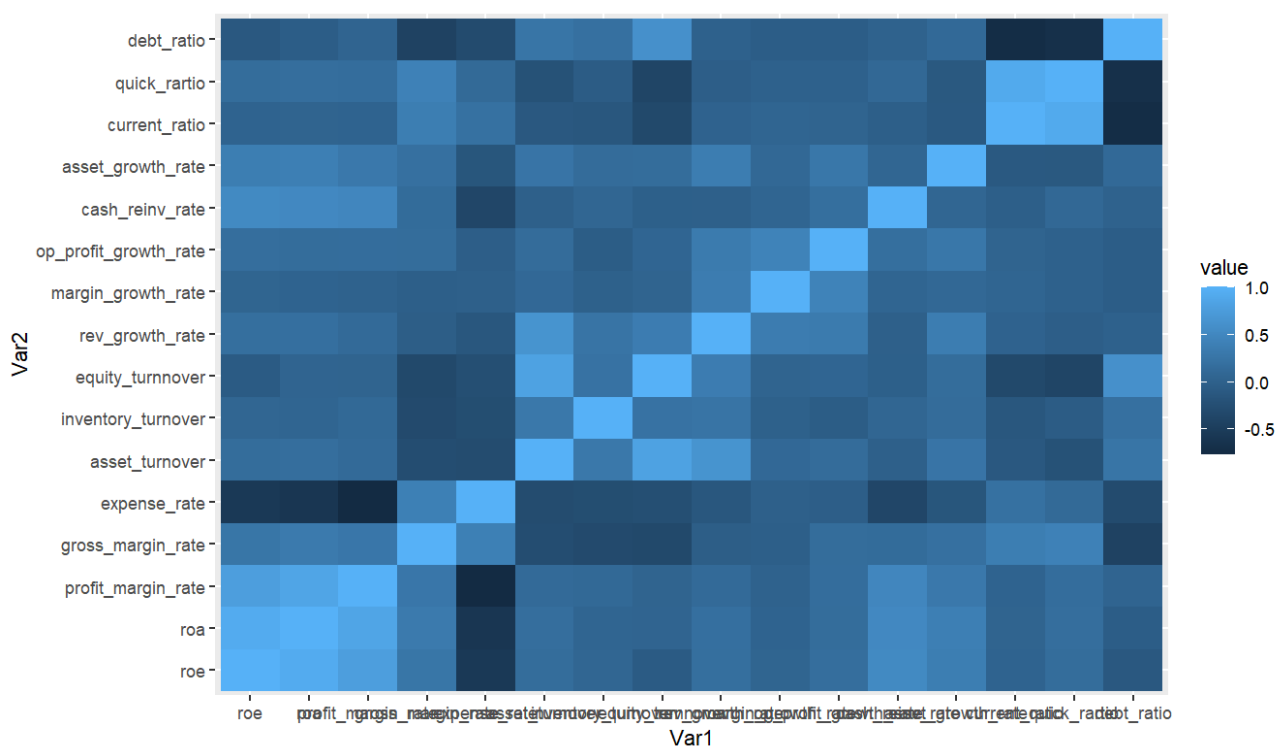
```
melted_cormat <- melt(M)
```

```
head(melted_cormat)
```

	Var1	Var2	value
1	roe	roe	1.0000000
2	roa	roe	0.9276137
3	profit_margin_rate	roe	0.7698366
4	gross_margin_rate	roe	0.2639333
5	expense_rate	roe	-0.5639643
6	asset_turnover	roe	0.1479487

```
library(ggplot2)
```

```
ggplot(data = melted_cormat, aes(x=Var1, y=Var2, fill=value)) +  
  geom_tile()
```



1. 資料 **financialdata.csv** 有 163 間公司的財務指標。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？

```
# install.packages("stats")
```

```
library(stats)
```

```
pca<- prcomp(data, center = TRUE, scale = TRUE)
```

```
names(pca) #sdev=std. deviation, rotation=coefficient 主成份的係數矩陣
```

```
> names(pca) #sdev=std. deviation, rotation=coefficient主成份的係數矩陣
[1] "sdev"      "rotation" "center"    "scale"     "x"
```

```
summary(pca)
```

```
> summary(pca)
```

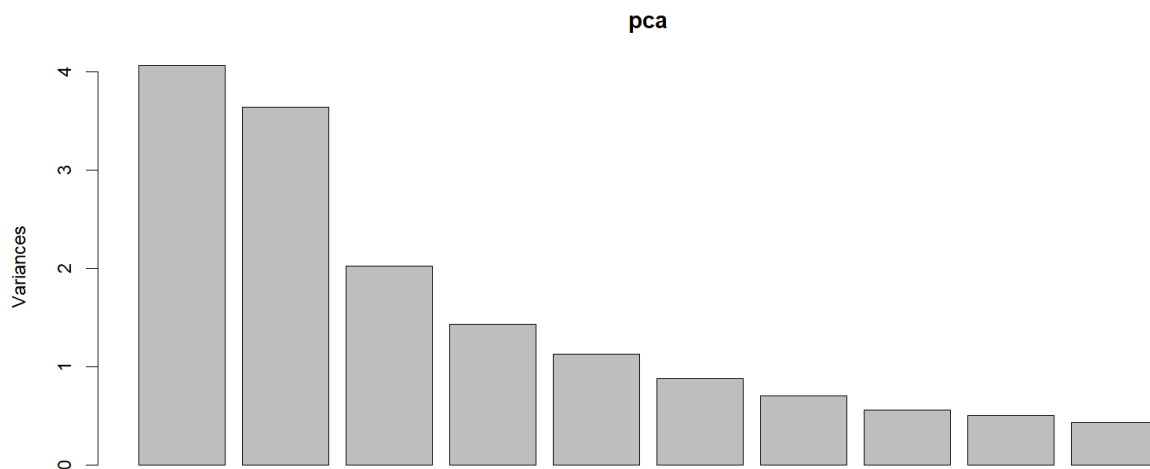
```
Importance of components:
```

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
Standard deviation	2.0165	1.9074	1.4221	1.1940	1.06009	0.93572	0.83472	0.74451	0.7065	0.65378	0.55181	0.44689
Proportion of Variance	0.2541	0.2274	0.1264	0.0891	0.07024	0.05472	0.04355	0.03464	0.0312	0.02671	0.01903	0.01248
Cumulative Proportion	0.2541	0.4815	0.6079	0.6970	0.76727	0.82199	0.86554	0.90018	0.9314	0.95809	0.97712	0.98961

	PC13	PC14	PC15	PC16
Standard deviation	0.28365	0.23800	0.1550	0.07197
Proportion of Variance	0.00503	0.00354	0.0015	0.00032
Cumulative Proportion	0.99464	0.99818	0.9997	1.00000

```
plot(pca) #Variation explained
```



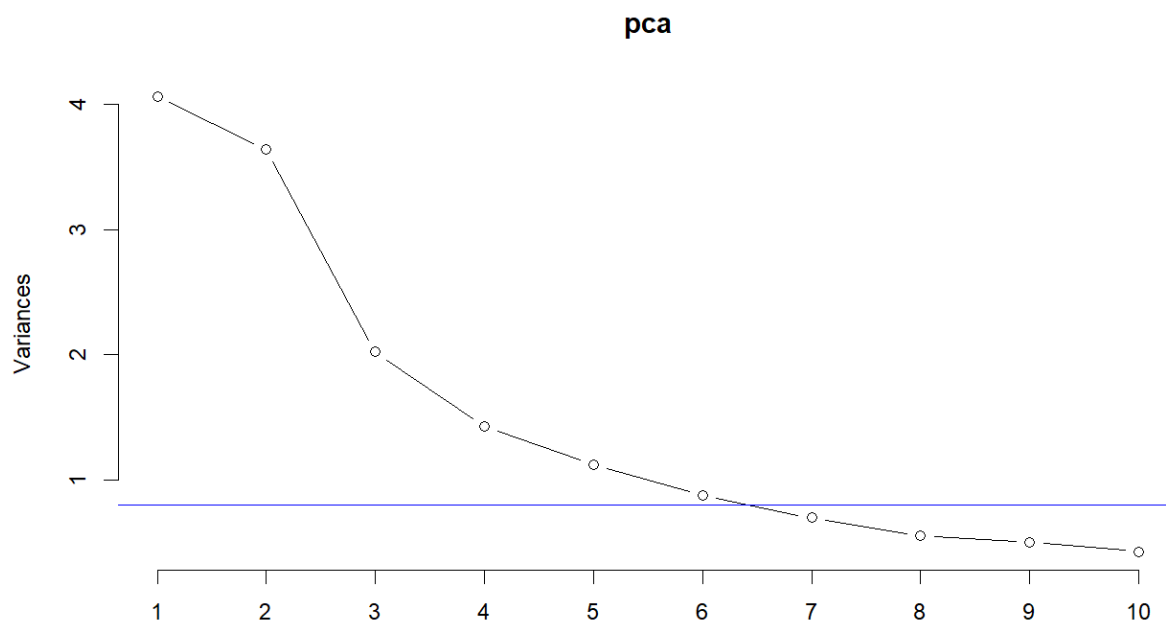
```
## scree plot: variance
```

```
screeplot(pca) #same as plot(pca)
```

```
plot(pca, type="line")
```

```
abline(h=0.8, col="blue")
```

```
#Kaiser eigenvalue-greater-than-one rule, choose pc1~pc5 by Kaiser
```



```
summary(pca)
> summary(pca)
Importance of components:
      PC1      PC2      PC3      PC4      PC5      PC6      PC7      PC8      PC9      PC10     PC11     PC12
Standard deviation  2.0165  1.9074  1.4221  1.1940  1.06009  0.93572  0.83472  0.74451  0.7065  0.65378  0.55181  0.44689
Proportion of Variance 0.2541 0.2274 0.1264 0.0891 0.07024 0.05472 0.04355 0.03464 0.0312 0.02671 0.01903 0.01248
Cumulative Proportion 0.2541 0.4815 0.6079 0.6970 0.76727 0.82199 0.86554 0.90018 0.9314 0.95809 0.97712 0.98961
      PC13      PC14      PC15      PC16
Standard deviation  0.28365 0.23800 0.1550 0.07197
Proportion of Variance 0.00503 0.00354 0.0015 0.00032
Cumulative Proportion 0.99464 0.99818 0.9997 1.00000
```

每個 PC 可以解數 25%到 0.032%的變異，通常會選前 80%具代表性的 PC 來解釋資料，也就是 financialdata 的前六個 PC。

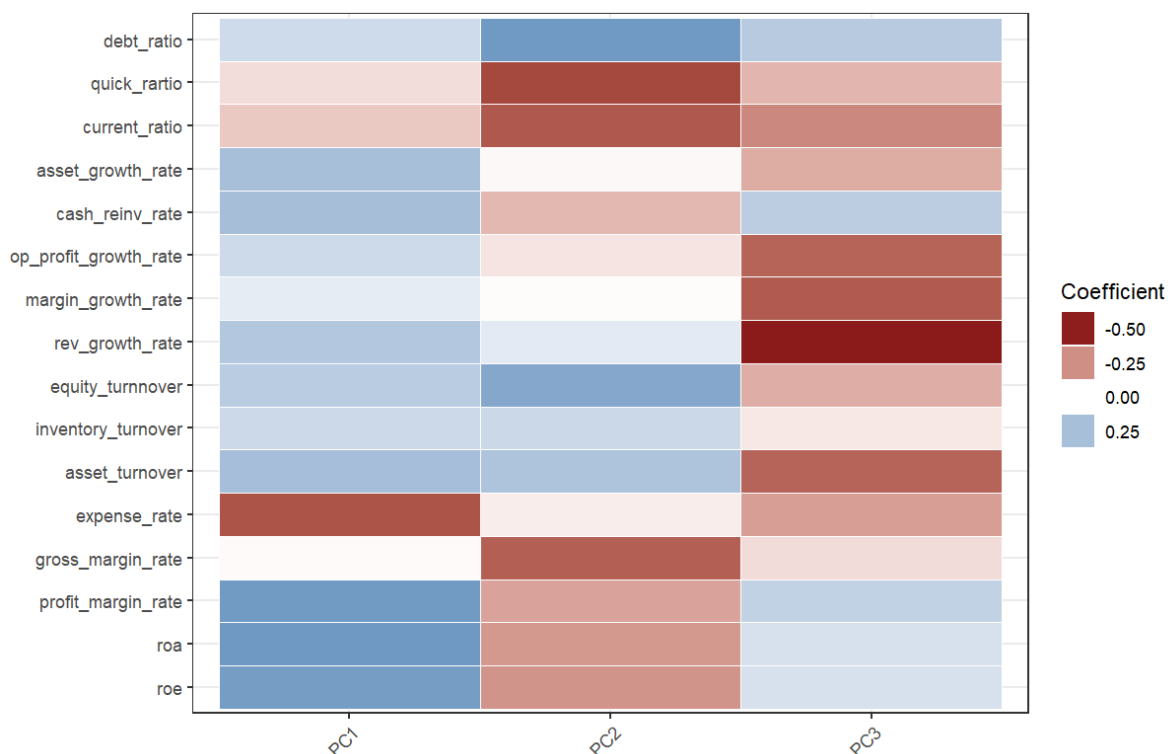
## 2. 找出前三個主成份分別重點變數為何並解釋。

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

```
pca$rotation
```

#visualize

```
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())
```



將三個主成分的特點整理如下表

主成分	正相關	負相關
PC1	expense_rate 營業費用率	roe 股東權益報酬率 roa 資產報酬率 profit_margin_rate 營業利益率
PC2	quick_ratio 速動比例 current_ratio 流動比例 gross_margin_rate 營業毛利率	debt_ratio 負債比例 equity_turnover 淨值週轉率
PC3	rev_growth_rate 營收成長率 margin_growth_rate 毛利成長率 op_profit_growth_rate 營業利益成長率 asset_turnover 資產週轉率	

從 PC1 來看：

1. 營業費用率越高，代表公司運用營業費用的獲利能力越低
2. 可以將 ROE 理解成公司幫股東賺錢的效率，需搭配 EPS 來看。每股盈餘 (EPS) 高、但股東權益報酬率 (ROE) 很低，代表公司賺錢沒效率
3. 如果一家企業資產報酬率低，表示該企業資產的使用效率不佳，因此能創造的利潤也就越少，使得獲利能力表現較差。
4. 營業利益率低，可能因為公司管銷費用控管不佳，但也有可能公司的規模仍不夠大，使得各項基本的人事、總務及管銷費用占有較大的比重。

綜上所述，符合 PC1 的企業，具有營業費用率高、股東權益報酬率低、資產報酬率低、營業利益率低的特性，代表這間公司獲利能力較差。

再從 PC2 來看：

1. 速動比率代表公司每一元的流動負債，有多少流動資產可供償還，係衡量短期償債能力的方法。比率越高，發生短期資金週轉不靈的風險越低，對短期債權人便越有保障。
2. 流動比率一般要求是要在 200% 以上，越高越好，代表短期在還債上越沒有壓力。
3. 毛利率 (Gross Margin) 是指毛利佔營業收入的百分比，毛利率高，代表企業營業成本所佔營業收入的比率低。
4. 企業的負債比率低，表示該企業的資金流相對穩定，財務風險較低，然而過低可能反映出公司未能充分善用財務槓桿，積極創造獲利機會。
5. 淨值週轉率淨值週轉率在表示自有資本在一年期間內從營業收入收回的次數多少。其計算方式為：營業收入 / 平均業主權益。淨值週轉率太低表示自有資本太多或營業額太少。

綜上所述，符合 PC2 的企業，具有速動比例高、流動比例高、營業毛利率高、負債比例略低、淨值週轉率略低的特性，代表該企業還債能力高、財務風險低、營業收入狀態佳。

最後看到 PC3：

1. 營收成長比率逐漸提升，代表市場對公司產品的需求提高。
2. 毛利率高代表公司可以用更低的成本，生產同樣的產品，也就表示公司可以獲取的利潤較

多。

3. 營業利益率越高表示每 1 元的營收中能帶來的獲利越高，代表公司本業的獲利能力越好。
4. 資產周轉率越高，代表公司能有效利用資產來產生收益。

綜上所述，符合 PC3 的企業，具有營收成長率高、毛利成長率高、營業利益成長率高、資產週轉率高的特性，代表該企業具較好的獲利能力，且能妥善運用資產增加收益，市場對產品需求逐漸提升，是已經成長一段時間，且後續表現可期，值得投資的公司類型。

### **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)**

從上述來看具 PC2 和 PC3 特性的公司較值得投資，其中參考最具相關的部分，投資標的應符合以下三點：

1. 速動比例高
2. 流動比例高
3. 營收成長率高

此類公司具有相對安全的財務特性，且獲利能力高，能讓投資人在風險較低的情況下獲得較好的收益，較具投資價值。