**商業分析作業四**

**統計三 108304010 趙啟宏**

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？

##read in data##

data <- read.csv('financialdata.csv')

data <- data[,-1]

str(data)

summary(data)

##change factor into numeric##

data$op\_profit\_growth\_rate <- as.character(data$op\_profit\_growth\_rate)

data$op\_profit\_growth\_rate <- gsub(',','',data$op\_profit\_growth\_rate)

data$op\_profit\_growth\_rate <- as.numeric(data$op\_profit\_growth\_rate)

data$current\_ratio <- as.character(data$current\_ratio)

data$current\_ratio <- gsub(',','',data$current\_ratio)

data$current\_ratio <- as.numeric(data$current\_ratio)

data$quick\_rartio <- as.character(data$quick\_rartio)

data$quick\_rartio <- gsub(',','',data$quick\_rartio)

data$quick\_rartio <- as.numeric(data$quick\_rartio)

summary(data)

str(data)

##SPCA##

library(nsprcomp)

spca <- nscumcomp(data, k=80, nneg=T, scale=T)

summary(spca)

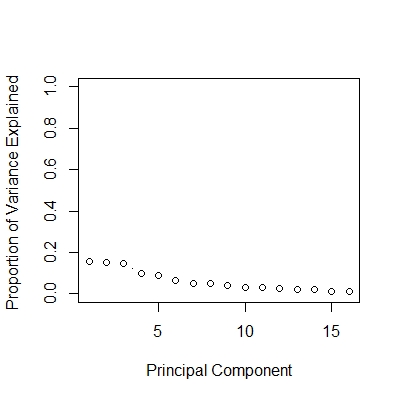
screeplot(spca)

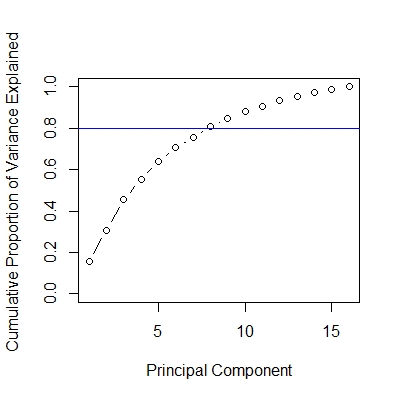
##Q1##pve=(spca$sdev)^2 / (sum(spca$sdev^2))

plot(pve, xlab="Principal Component", ylab="Proportion of Variance Explained ", ylim=c(0,1),type='b')#scree plot

plot(cumsum(pve), xlab="Principal Component", ylab="Cumulative Proportion of Variance Explained ", ylim=c(0,1),type='b')

abline(h=0.8,col='blue')





**詮釋:將資料讀進R後，我先檢查各個變數的資料類別，把其中三個非數值型的變數改為數值型。在處理完資料後，我決定使用SPCA來降維，降維的效果如圖，PC1大約能解釋20%的資料變異，其他的主成分依序下滑，而我希望主成分最少能解釋80%的資料變異，為了達到標準，我大約需要8個PCs。**

1. 找出前三個主成份分別重點變數為何並解釋。

library(ggplot2)

library(reshape2)

ggplot(melt(spca$rotation), aes(Var2, Var1)) +

geom\_tile(aes(fill = value), colour = "white") +

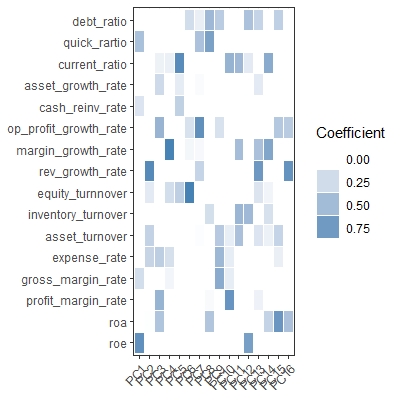
scale\_fill\_gradient2(low = "white", high = "steelblue") +

guides(fill=guide\_legend(title="Coefficient")) +

theme\_bw() +

theme(axis.text.x = element\_text(angle = 45, hjust = 1, vjust = 1),

axis.title = element\_blank())



**詮釋:PC1的重點變數是ROE(return on equity)，股東權益報酬率一直都是投資人在衡量一家公司是否值得投資時的重要指標，代表股東每花一塊錢約能賺進幾塊錢的稅後淨利，一般來說，這個數字越高，表示公司越值得投資。PC2的重點變數是rev\_growth(營收成長率)，表示跟去年相比，這間公司今年賺的錢是變多還是變少的比例，跟ROE類似，一般來說，這個數字越高，表示公司越值得投資。**

**最後，PC3的重點變數是profit\_margin\_rate(營業利益率)和op\_profit\_margin\_rate(營業利益成長率)，這兩個變數高度相關，都是在衡量這間公司的利潤(營收-成本)的表現，由於不同的公司規模和事業種類大不相同，單看營收可能會不夠，因此利潤就是一個很重要的數字，一樣，一般來說，這個數字越高，表示公司越值得投資。**

3.找出適合投資的公司

##Q3##

g1 <- which(data$roe>spca$center[1])

g2 <- which(data$rev\_growth\_rate>spca$center[9])

g3 <- which(data$profit\_margin\_rate>spca$center[3])

g4 <- which(data$op\_profit\_growth\_rate>spca$center[11])

i1 <- intersect(intersect(intersect(g1,g2),g3),g4)

g5 <- which(data$margin\_growth\_rate>spca$center[10])

g6 <- which(data$current\_ratio>spca$center[14])

g7 <- which(data$equity\_turnnover>spca$center[8])

intersect(intersect(intersect(g5,i1),g6),g7)

**詮釋:仔細觀察Q2的各個PCs的重點變數，遵照Q1的邏輯，把前8個PCs的重要變數都選出，依序篩選出每項指標皆高於平均的公司，最後留下2家--編號50和編號103的公司，是適合投資的公司。**