

15.2 제약 산업 한 주식 애널리스트는 제약 산업에 대해 연구 중이며, 회사에서 수집한 재무 데이터를 탐색하고 이해하는 데 도움을 받고자 한다. 분석의 주된 목적은 일부 기본적인 재무 변수들을 사용하여 제약 산업의 구조를 이해하는 것이다.

재무데이터는 제약 산업에 속하는 21개의 회사로부터 수집되었고, Pharmaceuticals.csv 파일에 이용가능하다. 각 회사들에 대해서 다음과 같은 변수가 수집되었다.

```
setwd("c:/rdata")
pharma.df <- read.csv("Pharmaceuticals.csv")

#####
row.names(pharma.df) <- pharma.df[,1]
str(pharma.df)

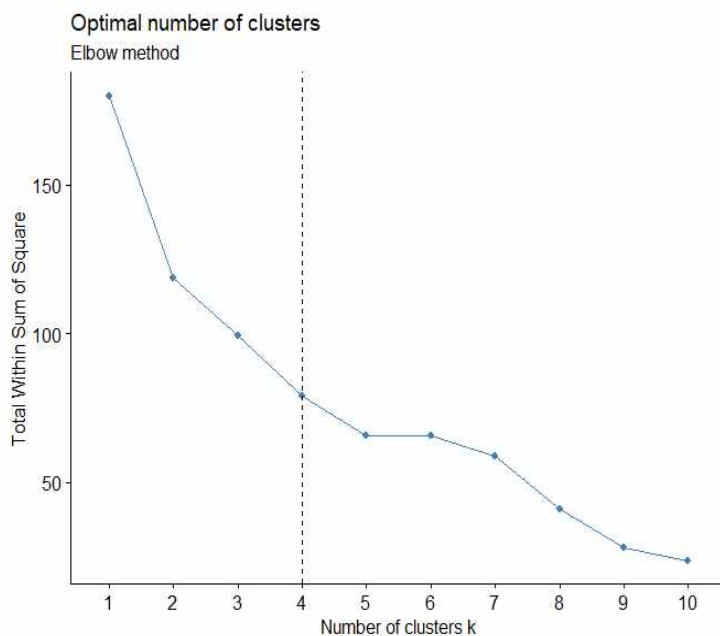
#####
pharma.df.norm <- sapply(pharma.df[,3:11], scale)
pharma.df.norm
row.names(pharma.df.norm) <- row.names(pharma.df)
```

a. 21개의 회사들 군집화하기 위해 양적변수들(1번부터 9번까지의 변수)만을 사용하시오. 군집분석을 수행할 때 고려해야 할 여러 가지 선택사항들, 예를 들어 서로 다른 변수들의 가중치, 특정 군집화 알고리즘, 군집의 수 등에 관하여 정당성을 기술하시오.

다른 변수들의 가중치 : 유클리드 거리는 각 변수의 단위의 영향을 크게 받기 때문에 유클리드 거리를 계산하기 전에 연속형 값들에 대해 레코드들을 정규화한다.

특정 군집화 알고리즘 : 계산 집약적이지 않고, 대용량 데이터에 선호되는 K-평균 군집화 방법을 사용한다.

군집의 수 : 군집의 수는 '엘보우 차트'를 이용해서 군집의 수를 4개로 결정한다.



b. 군집을 형성하는 데 사용된 양적변수들과 관련하여 군집결과를 해석하시오.

```
#####
km <- kmeans(pharma.df.norm, 4)

#####
centers <- aggregate(. ~ km$cluster, data = pharma.df[,3:11], FUN = mean)
centers
```

km\$cluster	Market_Cap	Beta	PE_Ratio	ROE	ROA	Asset_Turnover	Leverage	Rev_Growth	Net_Profit_Margin
1	157.01750	0.4800000	22.22500	44.42500	17.70000	0.9500000	0.2200000	18.532500	19.57500
2	4.37800	0.8880000	21.20000	15.14000	4.60000	0.4800000	1.3920000	16.356000	11.14000
3	62.88286	0.4071429	20.22857	30.71429	13.38571	0.7142857	0.3857143	5.382857	20.51429
4	24.10800	0.3660000	39.64000	14.66000	6.66000	0.7000000	0.3520000	17.440000	10.40000

1번 군집은 Market_cap, ROE, ROA, Asset_Turnover, Rev_Growth이 다른 군집에 비해 높으며 Leverage가 매우 낮다.

2번 군집은 Beta, Leverage가 매우 높으며, Market_Cap, ROA, Asset_Turnover이 매우 낮다.

3번 군집은 Net_Profit_Margin이 매우 높으며, PE_Ratio, Rev_Growth이 매우 낮다.

4번 군집은 PE_Ratio이 매우 높지만, Beta, ROE, Net_Profit_Margin이 매우 낮다.

c. 군집을 형성하는 데 사용되지 않은 수치형 변수들(10번부터 12번까지의 변수)과 관련하여 군집에 어떤 패턴이 존재하는가?

```
table(pharma.df$Median_Recommendation, km$cluster)
```

	1	2	3	4
Hold	2	2	4	1
Moderate Buy	2	2	1	2
Moderate sell	0	1	2	1
Strong Buy	0	0	0	1

```
table(pharma.df$Location, km$cluster)
```

	1	2	3	4
CANADA	0	0	0	1
FRANCE	0	0	0	1
GERMANY	0	1	0	0
IRELAND	0	1	0	0
SWITZERLAND	0	0	1	0
UK	1	0	1	1
US	3	3	5	2

```
table(pharma.df$Exchange, km$cluster)
```

	1	2	3	4
AMEX	0	1	0	0
NASDAQ	0	1	0	0
NYSE	4	3	7	5

1번 군집은 주식을 가지고 있거나 사야 하는 주식이고, 본사의 위치는 US 상장된 주식거래소는 NYSE일 가능성이 높다.

2번 군집은 재무적으로 대부분 안정적이지만, 'moderate sell'을 해야 할 제약 산업이 존재한다.

3번 군집은 재무적으로 대부분 안정적이지만, 'moderate sell'을 해야 할 제약 산업이 2개 존재한다. 본사는 US, 상장거래소는 NYSE일 확률이 높다.

4번 군집은 상장거래소는 NYSE일 확률이 매우 높다.

d. 데이터세트에 있는 일부 혹은 모든 변수를 사용하여 각 군집에 적절한 이름을 부여하시오.

1번 군집 : High Market_cap/ Low reverage (우량주)

2번 군집 : High Beta, reverage / Low Market_cap(고위험주)

3번 군집 : Net_Profit_Margin/ Low Rev_Growth(저정상주)

4번 군집 : High PE_Ratio / Low Beta(안정주)