Kyle Beitz

STAT 4300

Homework 5

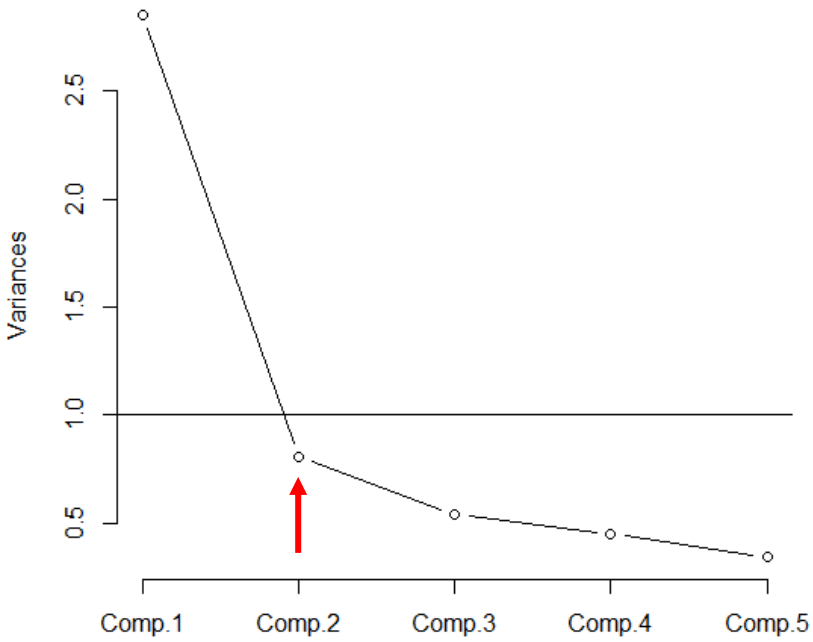
Importance of components:

	PC.1	PC.2	PC.3	PC.4	PC.5
Stddev	**1.6901808**	0.8995353	0.7346259	0.67193757	0.58561928
Prop Var	0.5713422	0.1618327	0.1079350	0.09030002	0.06858999
Cum Var	0.5713422	0.7331749	0.8411100	0.93141001	1.00000000

Loadings:

	PC.1	PC.2
chem	0.784	-0.216
dupont	0.773	-0.458
carbide	0.795	-0.234
exxon	0.712	0.473
texaco	0.712	0.524

Scree Diagram



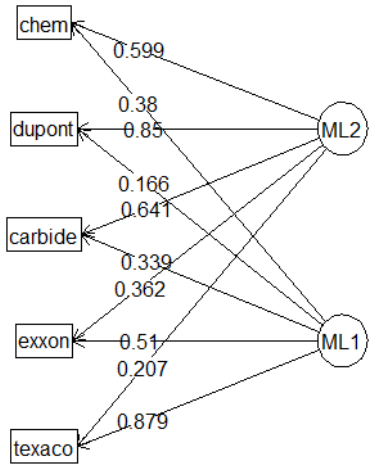
- H.5.1:

- (a):
  - The average eigenvalue approach (m=1) since one pc component have eigen values greater than 1, meaning we should retain 1 component. The percent variance explained (PVE) at 90% (m = 4), meaning we need to retain 4 pc components to exceed 90% variance explained. The skree diagram (m = 2) since the elbow appears to occur at component 2 where the is a sharp degree to the left.
- (b):
  - From the `fa loadings` above we see that PC.1 all have magnitudes that are very high and are going in the same direction. This appears to show the overall return on stocks for the 5 companies. PC.2 appears to show the comparison between chemical companies and oil and gas company returns. We can see that all the chemical companies are going in the same direction and the oil and gas companies are going in the opposite direction. Du Pont, Exxon and Texaco all have the highest magnitudes.

```
fa.ml=fa(R,m,fm="ml",rotate="none",SMC=F,n.obs=n)
> print(fa.ml$loadings,cutoff=0.4)

Loadings:
        ML1    ML2
chem     0.687
dupont   0.704  0.505
carbide  0.686
exxon    0.620
texaco   0.782 -0.453

>(fa.ml$factors,fa.ml$objective,fa.ml$dof,fa.ml$STATISTIC,fa.ml$PVAL)

Nfac   like    df    chisq   pvalue

2.0    0.0059  1.0   0.56    0.45
```

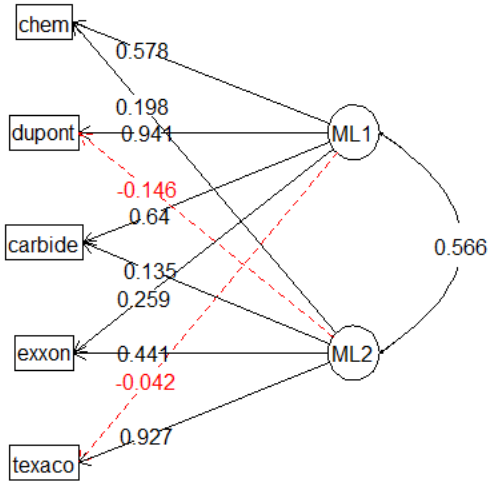
Varimax

```
Loadings:
        ML2   ML1
chem    0.599 0.380              #ML2 Chemical Companies#
dupont  0.850 0.166              #ML1 Oil and Gas Companies#
carbide 0.641 0.339
exxon   0.362 0.510
texaco  0.207 0.879

chem   dupont  carbide   exxon   texaco  mean_complex
1.69   1.08    1.52      1.81    1.11    1.44
```



Oblimin

```
Loadings:
         ML1     ML2
chem      0.578  0.198            #ML1 Chemical Companies#
dupont    0.941 -0.146            #ML2 Oil and Gas Companies#
carbide   0.640  0.135
exxon     0.259  0.441
texaco   -0.042  0.927

chem  dupont   carbide  exxon   texaco  mean_complex
1.23  1.05     1.09     1.62    1.00    1.20
```



No Rotation

```
chem      dupont   carbide  exxon   texaco   mean_complex
1.02      1.56     1.06     1.20    1.82     1.33
```

Residuals w/ no rotations

	chem	dupont	carbide	exxon	texaco
chem	0.00	0.00	-0.00	-0.02	0.00
dupont		0.00	-0.00	-0.00	0.00
carbide			0.00	0.03	-0.00
exxon				0.00	-0.00
texaco					0.00

- H.5.2:
  - (a):
    - $H_0 : \sum = \lambda\lambda' + \Psi$, $H_1 : \sum \neq \lambda\lambda' + \Psi$ at $\alpha = 0.05$
    - Since $X^2_0(1) = 0.56$ with p-value > 0.05, fail to reject H_0 at the 0.05 level and not in favor of H_1. Thus, there is not evidence that the population covariance matrix ($\sum$) does not have factor analytic structure.
  - (b):
    - The varimax rotation with the ML approach, ML1 represents oil and gas companies, ML2 represents chemical companies. With the varimax rotation we can see that ML2 has high magnitude loadings on chem, dupont and carbide grouping them together and all in the same direction. ML1 has high magnitude loadings on exxon and texaco grouping them together and all in the same direction. We do have simple structure with varimax rotation. Considering the Oblimin rotation we see that ML1 now represents the chemical companies and ML2 represents oil and gas companies. ML1 has high magnitudes for chem, dupont and carbide, grouping them together and all moving in the same direction. ML2 has high magnitudes for exxon and texaco, grouping them together and both moving in the same direction. Oblimin also shows simple structure.
  - (c):
    - The rotations above have helped obtain simple structure. We can see that for both varimax and Oblimin rotations all the chemical companies (chem, dupont and carbide) are grouped together in one ML and both the oil and gas companies (exxon and texaco) are grouped together in a different ML. Varimax rotation complexities have all loadings under 2 with chem being the highest at 1.69. Thus, all 5 variables are loading on 1.44 factors on average. No rotations has all complexities under 2 with all 5 factors loading on 1.33 factors on average. However, simple structure with the ML approach is best attained overall by the oblique rotation as average complexity is the smallest at 1.20. The oblique rotation only exxon has a higher individual loading complexity than 1.23. All other individual complexities are close to one and are loading on only one factor.
  - (d):
    - From the correlation matrix above we can see that almost all the residuals are at zero with exxon/chem(-0.2) and exxon/carbide(0.03) being the highest. We are underpredicting exxon/chem and overpredicting exxon/carbide. The two-factor model appears to provide a perfectly adequate fit for this data.