Ch9:Exercise1-stat439

Exercise 9.1 :

Show that covariance matrix $S = \boldsymbol{\rho} = \begin{bmatrix} 1 & 0.63 & 0.45 \\ 0.63 & 1 & 0.35 \\ 0.45 & 0.35 & 1 \end{bmatrix}$

For the $p = 3$ standardized random variables Z_1, Z_2 and Z_3 can be generated by the $m = 1$ factor model

$$Z_1 = 0.9F_1 + \varepsilon_1$$

$$Z_2 = 0.7F_1 + \varepsilon_2$$

$$Z_3 = 0.5F_1 + \varepsilon_3$$

Where $var(F_1) = 1, cov(\varepsilon, F_1) = 0, and\ \psi = cov(\varepsilon) = \begin{bmatrix} 0.19 & 0 & 0 \\ 0 & 0.51 & 0 \\ 0 & 0 & 0.75 \end{bmatrix}$. That is, write $\boldsymbol{\rho}$ in the form $\boldsymbol{\rho} = \boldsymbol{LL^T} + \boldsymbol{\psi}$.

Solution :

Since Z is the standardized r.v, and $\boldsymbol{L} = \begin{bmatrix} 0.9 \\ 0.7 \\ 0.5 \end{bmatrix}$, we have the following:

$$\boldsymbol{LL^T} + \boldsymbol{\psi} = \begin{bmatrix} 0.9 \\ 0.7 \\ 0.5 \end{bmatrix} [0.9 \quad 0.7 \quad 0.5] + \begin{bmatrix} 0.19 & 0 & 0 \\ 0 & 0.51 & 0 \\ 0 & 0 & 0.75 \end{bmatrix} =$$

$$\begin{bmatrix} 0.81 & 0.63 & 0.45 \\ 0.63 & 0.49 & 0.35 \\ 0.45 & 0.35 & 0.25 \end{bmatrix} + \begin{bmatrix} 0.19 & 0 & 0 \\ 0 & 0.51 & 0 \\ 0 & 0 & 0.75 \end{bmatrix} = \begin{bmatrix} 1 & 0.63 & 0.45 \\ 0.63 & 1 & 0.35 \\ 0.45 & 0.35 & 1 \end{bmatrix} = \boldsymbol{\rho}$$

Hence verified, the value of covariance matrix matches with the one provided.

Exercise 9.2 : use the information in Exercise 9.1

a) Calculate communalities $h_i^2, i = 1,2,3$, and interpret these quantities.
b) Calculate $Corr(Z_i, F_1)$ for i=1,2,3. Which variable might carry the greatest weight in "naming" the common factor? Why?
c) The eigenvalues and eigenvectors of the correlation matrix $\boldsymbol{\rho}$ are

$$\lambda_1 = 1.96\,, \quad e_1^T = [0.625 \quad 0.593 \quad 0.507]$$
$$\lambda_2 = 0.68\,, \quad e_2^T = [-0.219 \quad -0.491 \quad 0.843]$$
$$\lambda_3 = 0.36\,, \quad e_3^T = [0.749 \quad -0.638 \quad -0.177]$$

1- Assuming an $m = 1$ factor model, calculate the loading matrix **L** and matrix of specific variances Ψ using the principal component solution method. Compare the results with those in Exercise 9.1.
2- What proportion of the total population variance is explained by the first common factor?

Solution of 9.2:

a) Since m =1, then, by Equ.(9-6), the communalities are calculated as the following:

$$h_i^2 = l_{i\,1}^2 + l_{i\,2}^2 + \cdots + l_{i\,m}^2$$

$$h_1^2 = l_{11}^2 = 0.9^2 = 0.81 \quad ; \quad h_2^2 = l_{21}^2 = 0.7^2 = 0.49 \quad ; \; h_3^2 = l_{31}^2 = 0.5^2 = 0.25$$

b) Since $Corr\left(Z_i, F_j\right) = cov\left(Z_i\,, F_j\right) = l_{i\,j}$. Therefore,

$$cov(Z_1\,, F_1) = l_{11} = 0.9 \quad ; \quad cov(Z_2\,, F_1) = L_{21} = 0.7 \quad ; \quad cov(Z_3\,, F_1) = L_{31} = 0.5$$

Because the first variable Z_1 has the largest correlation with common factor, Z_1 will carry greatest weight in term of F_1.

c) part 1: By principal component solution methods (9-15, 9-16 and 9-17), we have:

$$\tilde{L}_{p\times m} = \left[\sqrt{\hat{\lambda}_1}\;\; \hat{e}_1 \vdots \sqrt{\hat{\lambda}_2}\;\; \hat{e}_2 \vdots \cdots \vdots \sqrt{\hat{\lambda}_m}\;\; \hat{e}_m\right]; \quad \tilde{\Psi} = S - \tilde{L}\tilde{L}^T = \begin{bmatrix} \tilde{\Psi}_1 & 0 \cdots & 0 \\ 0 & \tilde{\Psi}_2 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 \cdots & \tilde{\Psi}_p \end{bmatrix}$$

$$\tilde{L}_{\,3\times 1} = \left[\sqrt{\hat{\lambda}_1}\;\; \hat{e}_1\right] = \sqrt{1.96}\begin{bmatrix} 0.625 \\ 0.593 \\ 0.507 \end{bmatrix} = \begin{bmatrix} 0.875 \\ 0.830 \\ 0.710 \end{bmatrix}$$

$$\tilde{\Psi} = \rho - \tilde{L}\tilde{L}^T = \begin{bmatrix} 1 & 0.63 & 0.45 \\ \square & 1 & 0.35 \\ \square & \square & 1 \end{bmatrix} - \begin{bmatrix} 0.875 \\ 0.830 \\ 0.710 \end{bmatrix} [0.875 \quad 0.830 \quad 0.710] =$$

$$\begin{bmatrix} 1 & 0.63 & 0.45 \\ \square & 1 & 0.35 \\ \square & \square & 1 \end{bmatrix} - \begin{bmatrix} 0.77 & 0.73 & 0.62 \\ 0.73 & 0.89 & 0.59 \\ 0.62 & 0.59 & 0.5 \end{bmatrix} = \begin{bmatrix} 0.23 & 0 & 0 \\ 0 & 0.31 & 0 \\ 0 & 0 & 0.5 \end{bmatrix}$$

Each estimation $\tilde{\Psi}_i$ is different from the results of Exercise 9.1.

c) part 2: By Equ. (9-20), the proportion of the total estimated population variance due to the first common factor is: $\frac{\widehat{\lambda_1}}{p} = \frac{1.96}{3} = 0.653$

Example 9.4 & 9.5 & 9.10: Stock-price data consisting of n = 103 weekly rates of return on p = 5 stocks.

a) Obtain the principal component solution for factor models with m= 2 by using correlation matrix R.

b) List the estimated communalities, specific variances and proportion of total sample variance explained by each factor .

c) What is the ψ matrix based on the PC ?

d) Show that $\Sigma = \hat{L}\widehat{L'} + \psi$.

e) Residual matrix.

f) Obtain the maximum likelihood solution for $m = 2$ common factors by use correlation matrix R.

g) Test adequacy of model ?

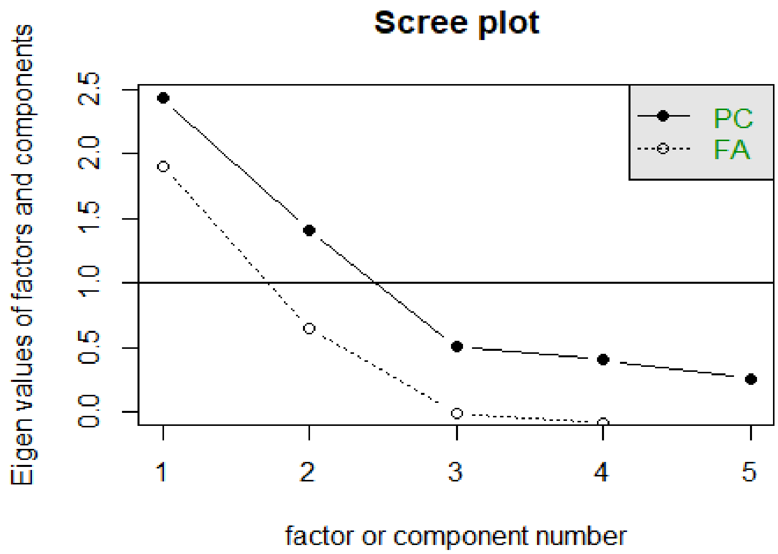
h) Dose your analysis should include Factor Rotation ?

i) Rotate your solutions in Parts (a) and (f). Compare the solutions and comment on them. Interpret each factor.

j) Determine a reasonable number of factors *m,* and compare the principal component and maximum likelihood solutions after rotation. Interpret the factors.

R Code:

```
rm(list=ls())
#install.packages("psych")
#install.packages("GPArotation")
library("psych")
library("GPArotation")
#### Example 9.4 & 9.5 (based on Example 8.5)
data1 <- read.table(choose.files()) #T8-4.DAT
names(data1) <- c("JPMorgan","Citibank","WellsFargo","RoyalDutchShell","Exxon
Mobil")
# Determine Number of Factors
scree(data1,pc=TRUE,main="Scree plot")
```



From ***scree plot*** we probably conclude that there are only two factors in the dataset depend on **PC**.

```
R=cor(data1) # sample correlation matrix (observed).
## part a & b: PC Factor Analysis based upon correlation matrix.
pc <- principal(data1, nfactors=2, rotate="none",covar=FALSE)
print.psych(pc, cut = NULL, sort = FALSE)

Principal Components Analysis
Call: principal(r = data1, nfactors = 2, rotate = "none", covar = FALSE)
Standardized loadings (pattern matrix) based upon correlation matrix
                 PC1   PC2   h2   u2 com
JPMorgan        0.73 -0.44 0.73 0.27 1.6
Citibank        0.83 -0.28 0.77 0.23 1.2
WellsFargo      0.73 -0.37 0.67 0.33 1.5
RoyalDutchShell 0.60  0.69 0.85 0.15 2.0
ExxonMobil      0.56  0.72 0.83 0.17 1.9
```

$$h_1^2 = l_{11}^2 + l_{12}^2 \to 0.73^2 + 0.44^2 = 0.73$$

73% of variation in that **JPMorgan** is explained by the two **factors.**

$$\boldsymbol{u2} = \boldsymbol{\psi}\ ;\ \ \psi_1 = 1 - h^2 = 1 - 0.73 = 0.27$$

```
                      PC1  PC2
SS loadings           2.44 1.41
Proportion Var        0.49 0.28
Cumulative Var        0.49 0.77
Proportion Explained  0.63 0.37
Cumulative Proportion 0.63 1.00
```

sum of squared loadings = **Eigenvalues**

$\frac{2.44}{5} = 0.49$ proportion of total sample variance do to 1-st factor.

0.28 proportion of total sample variance do to 2-ed factor.

With 2 components, we explain 77% of the variance.

```
Mean item complexity =  1.6
Test of the hypothesis that 2 components are sufficient.
The root mean square of the residuals (RMSR) is  0.1
with the empirical chi square  19.17  with prob <  1.2e-05
Fit based upon off diagonal values = 0.95
```

```
eige<-eigen(R) #estimated eigenvalue & eigenvectors for R.

lambda<-pc$values #estimated eigenvalue for R.

Lpc<- unclass(pc$loadings)

h.e<-pc$communality

## part c:
Psi1<-diag(round(pc$uniquenesses,3),5,5)
```

part d: estimated correlation matrix $\widehat{\boldsymbol{\rho}} = \widehat{L}\widehat{L}' + \psi$. We must change all diagonal values to be one.

```
(Rho1 <- Lpc %*% t(Lpc) + Psi1)

                 JPMorgan  Citibank WellsFargo RoyalDutchShell ExxonMobil
JPMorgan        0.99984576 0.7311286  0.6950107       0.1399197 0.09866056
Citibank        0.73112857 0.9995311  0.7084661       0.3079818 0.26645897
WellsFargo      0.69501067 0.7084661  1.0001396       0.1797042 0.14024657
RoyalDutchShell 0.13991966 0.3079818  0.1797042       1.0002571 0.83921362
ExxonMobil      0.09866056 0.2664590  0.1402466       0.8392136 0.99951223
```

part e: estimated residual= $\mathbf{R} - \widehat{\boldsymbol{\rho}} = \mathbf{R} - \mathbf{Rho1}$

```
(residual<-round(pc$residual,3)) # R-LL'

                JPMorgan Citibank WellsFargo RoyalDutchShell ExxonMobil
JPMorgan           0.273   -0.099     -0.185          -0.025      0.056
Citibank          -0.099    0.230     -0.134           0.014     -0.054
WellsFargo        -0.185   -0.134      0.333           0.003      0.006
RoyalDutchShell   -0.025    0.014      0.003           0.153     -0.156
ExxonMobil         0.056   -0.054      0.006          -0.156      0.166
```

(residual2<-round(pc$residual,3)-Psi1) # $R - LL' - \psi = \mathbf{R} - \widehat{\boldsymbol{\rho}}$

```
                JPMorgan Citibank WellsFargo RoyalDutchShell ExxonMobil
JPMorgan           0.000   -0.099     -0.185          -0.025      0.056
Citibank          -0.099    0.000     -0.134           0.014     -0.054
WellsFargo        -0.185   -0.134      0.000           0.003      0.006
RoyalDutchShell   -0.025    0.014      0.003           0.000     -0.156
ExxonMobil         0.056   -0.054      0.006          -0.156      0.000

(residual3<-resid.psych(pc)-Psi1) #pretty printed.

                JPMrg Ctbnk WllsF RylDS ExxnM
JPMorgan         0.00
Citibank        -0.10  0.00
WellsFargo      -0.18 -0.13  0.00
RoyalDutchShell -0.03  0.01  0.00  0.00
ExxonMobil       0.06 -0.05  0.01 -0.16  0.00
```

NOTE:

- In **`Principal()`** since we are using R instead of S, the `covar =FALSE` by default. No rotation is done for now, so the `rotate="none"`.
- Principal components analysis using covariance as follow:
  - ➢ By use data set

```
pc3<- principal(data1, nfactors=2, rotate="none",covar=TRUE, cor="cov")
```

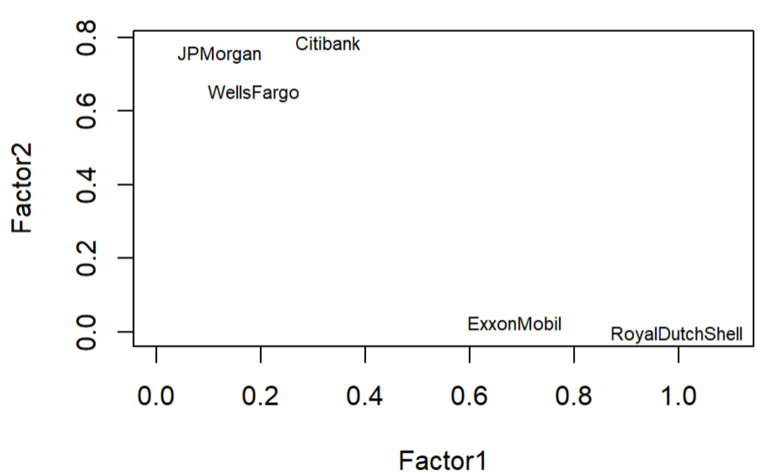
  - ➢ By use covariance matrix

```
v=cov(data1)
pc3<- principal(v, nfactors=2, rotate="none",covar=TRUE)
```

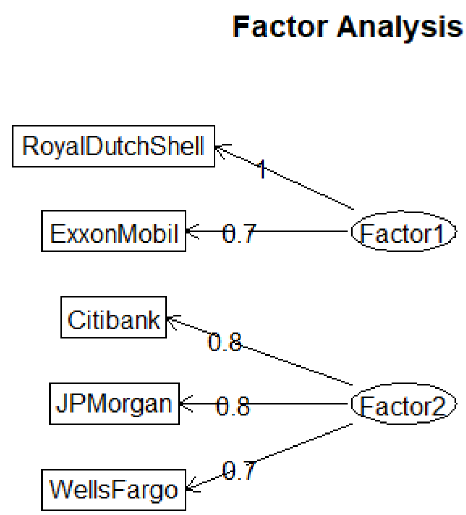
- The word "scree" refers to the loose stone that lies around the base of the mountain.
- In the scree plot , you may face different number of factors depending on principal components and a principal axis factor extraction. you might want to look at both of these factor solutions.

```
## part f:
## Maximum Likelihood Factor Analysis with correlation matrix.
#fa<- factanal(covmat= R ,factors=2,n.obs=103,rotation="none")

fam<- factanal(data1,2,rotation="none")
print(fam,cutoff = NULL)
Call:
factanal(x = data1, factors = 2, rotation = "none")

Uniquenesses:
   JPMorgan        Citibank      WellsFargo RoyalDutchShell      ExxonMobil
     0.417           0.275           0.542           0.005           0.530

Loadings:
                Factor1 Factor2
JPMorgan         0.121   0.754
Citibank         0.328   0.786
WellsFargo       0.188   0.650
RoyalDutchShell  0.997  -0.007
ExxonMobil       0.685   0.026

               Factor1 Factor2
SS loadings      1.622   1.610
Proportion Var   0.324   0.322
Cumulative Var   0.324   0.646

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 1.97 on 1 degree of freedom.
The p-value is 0.16
```

```
#names(fam) #see all the values of "factanal" function
Lml<- fam$loadings
communality<-apply(Lml^2, 1, sum)
Psi2 <- diag(fam$uniquenesses)
(R2 <- fam$correlation) #sample correlation matrix (observed)=R.

                 JPMorgan  Citibank WellsFargo RoyalDutchShell ExxonMobil
JPMorgan        1.0000000 0.6322878  0.5104973       0.1146019  0.1544628
Citibank        0.6322878 1.0000000  0.5741424       0.3222921  0.2126747
WellsFargo      0.5104973 0.5741424  1.0000000       0.1824992  0.1462067
RoyalDutchShell 0.1146019 0.3222921  0.1824992       1.0000000  0.6833777
ExxonMobil      0.1544628 0.2126747  0.1462067       0.6833777  1.0000000

(Rho2 <- Lml %*% t(Lml) + Psi2) #estimated correlation matrix. We must change
all diagonal values to be one.

                 JPMorgan  Citibank WellsFargo RoyalDutchShell ExxonMobil
JPMorgan        0.9999999 0.6322803  0.5130616       0.1149345  0.1024805
Citibank        0.6322803 1.0000000  0.5725336       0.3220805  0.2457536
WellsFargo      0.5130616 0.5725336  0.9999999       0.1825087  0.1456520
RoyalDutchShell 0.1149345 0.3220805  0.1825087       1.0000016  0.6832558
ExxonMobil      0.1024805 0.2457536  0.1456520       0.6832558  0.9999997

# Residual matrix = R-LL'-ψ

(resid<-round(R - Rho2, 4))

                JPMorgan Citibank WellsFargo RoyalDutchShell ExxonMobil
JPMorgan          0.0000   0.0000    -0.0026          -3e-04     0.0520
Citibank          0.0000   0.0000     0.0016           2e-04    -0.0331
WellsFargo       -0.0026   0.0016     0.0000           0e+00     0.0006
RoyalDutchShell  -0.0003   0.0002     0.0000           0e+00     0.0001
ExxonMobil        0.0520  -0.0331     0.0006           1e-04     0.0000

# plot Factor 1 by Factor 2
plot(Lml,type="n",xlim=c(0,1.1)) # set up plot
text(Lml,labels=names(data1),cex=.7)
```



```
#visualize the factor model
psych::fa.diagram(Lml)
```




```
## part g: Test for the Number of Common Factors
n=fam.$n.obs
p=5
m=2
( T.S=(n-1-((2*p+4*m+5)/6))*log(det(Rho2)/det(R)) )
[1] 2.004672

qchisq(0.95,(((p-m)^2)-p-m)/2)

[1] 3.841459
```

Ho: The data fit the model v.s H1: The data does not fit the model

From adequacy test, we can see that the significance level of the χ^2 fit statistic is $0.16 > 0.05$,this indicates that the hypothesis of perfect model fit is **acceptable**.

Bartlett-Corrected Likelihood Ratio Test Statistic:

$$[n-1-(2p+4m+5)/6]ln\frac{|\hat{L}_z\hat{L'}_z+\hat{\Psi}_z|}{|R|} = 1.97\ ;\quad \chi^2_{[(p-m)^2-p-m]/2,\ \alpha} = \chi^2_{1,0.05} = 3.84$$

since **1.97 < 3.84**, so we accept H_0.

NOTE!

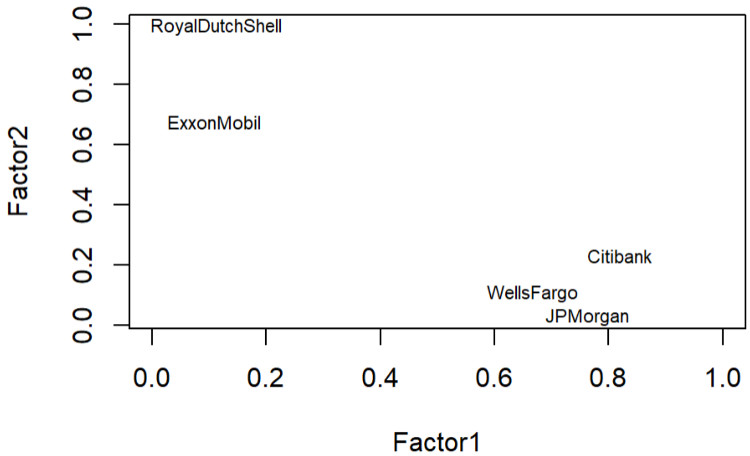
- In **`factanal`** No rotation is done for now, so the rotate argument is set to “none”. Use the **covmat=** option to enter a correlation or covariance matrix directly.
- To see all the values captured by the object use the **names()** function.

Rotated ML & PC Factor Analysis by using varimax.

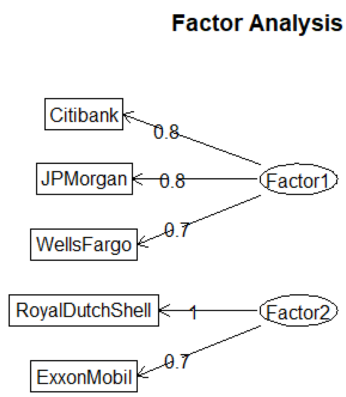
```
## part i & j: Example 9.10
fam.r<- factanal(data1,2,rotation="varimax")
print(fam.r,cutoff = NULL)


Call:
factanal(x = data1, factors = 2, rotation = "varimax")

Uniquenesses:
   JPMorgan        Citibank      WellsFargo RoyalDutchShell      ExxonMobil
    0.417           0.275           0.542           0.005           0.530
Loadings:
                Factor1 Factor2
JPMorgan        0.763   0.029
Citibank        0.819   0.232
WellsFargo      0.668   0.108
RoyalDutchShell 0.113   0.991
ExxonMobil      0.108   0.677

               Factor1 Factor2
SS loadings      1.725   1.507
Proportion Var   0.345   0.301
Cumulative Var   0.345   0.646

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 1.97 on 1 degree of freedom.
The p-value is 0.16

Lml.r<- fam.r$loadings

communality<-apply(Lml.r^2, 1, sum)

Psi2.r <- diag(fam.r$uniquenesses)

R2 <- fam.r$correlation # sample correlation matrix (observed)

(Rho2.r <- Lml.r %*% t(Lml.r) + Psi2.r) # estimated correlation matrix ρ̂.

                 JPMorgan  Citibank WellsFargo RoyalDutchShell ExxonMobil
JPMorgan        0.9999999 0.6322803  0.5130616       0.1149345  0.1024805
Citibank        0.6322803 1.0000000  0.5725336       0.3220805  0.2457536
WellsFargo      0.5130616 0.5725336  0.9999999       0.1825087  0.1456520
RoyalDutchShell 0.1149345 0.3220805  0.1825087       1.0000016  0.6832558
ExxonMobil      0.1024805 0.2457536  0.1456520       0.6832558  0.9999997

# Residual matrix
(resid.r<-round(R2 - Rho2.r, 4))

                JPMorgan Citibank WellsFargo RoyalDutchShell ExxonMobil
JPMorgan          0.0000   0.0000    -0.0026          -3e-04     0.0520
```

```
Citibank          0.0000   0.0000     0.0016          2e-04    -0.0331
WellsFargo       -0.0026   0.0016     0.0000          0e+00     0.0006
RoyalDutchShell  -0.0003   0.0002     0.0000          0e+00     0.0001
ExxonMobil        0.0520  -0.0331     0.0006          1e-04     0.0000
```

```
#additional part: visualize the factor model
plot(Lml.r,type="n",xlim=c(0,1)) # set up plot
text(Lml.r,labels=names(data1),cex=.7)
```




```
psych::fa.diagram(Lml.r)
```




```
# comparing  MLE & PC without rotation
comp1 <- data.frame(unclass(Lml), unclass(Lpc) )
newrow <-c("MLE","MLE", "PC","PC")
df1<-rbind(newrow,round(comp1,4))
row.names(df1)[1]<- "method"
df1

                Factor1 Factor2    PC1     PC2
method              MLE     MLE     PC      PC
JPMorgan         0.1206  0.7543 0.7323 -0.4365
Citibank         0.3285  0.7857 0.8312 -0.2805
WellsFargo       0.1876  0.6502 0.7262 -0.3739
RoyalDutchShell  0.9975 -0.0071 0.6047   0.694
ExxonMobil       0.6852  0.0263 0.5631  0.7186
```

```r
#*pc with "varimax" rotation
pc.r <- principal(data1, nfactors=2, rotate="varimax",covar=FALSE)
Lpc.r<-pc.r$loadings
#*comparing  MLE & PC with "varimax" rotation
comp2 <- data.frame( unclass(Lml.r), unclass(Lpc.r) )
newrow2 <-c("MLE.r","MLE.r", "PC.r","PC.r")
df2<-rbind(newrow2,round(comp2,4))
row.names(df2)[1]<- "method"
df2

                Factor1 Factor2   RC1    RC2
method            MLE.r   MLE.r  PC.r   PC.r
JPMorgan         0.7633  0.0292 0.8518 0.0356
Citibank         0.8195  0.2318 0.8491 0.2203
WellsFargo        0.668  0.1082 0.8124 0.0847
RoyalDutchShell  0.1127  0.9911 0.1262 0.9118
ExxonMobil       0.1084  0.6771 0.0778 0.9096

#*compare the Residuals
residual2 # from PC method

                JPMorgan Citibank WellsFargo RoyalDutchShell ExxonMobil
JPMorgan           0.000   -0.099     -0.185          -0.025      0.056
Citibank          -0.099    0.000     -0.134           0.014     -0.054
WellsFargo        -0.185   -0.134      0.000           0.003      0.006
RoyalDutchShell   -0.025    0.014      0.003           0.000     -0.156
ExxonMobil         0.056   -0.054      0.006          -0.156      0.000
resid #from MLE method

                JPMorgan Citibank WellsFargo RoyalDutchShell ExxonMobil
JPMorgan          0.0000   0.0000    -0.0026          -3e-04     0.0520
Citibank          0.0000   0.0000     0.0016           2e-04    -0.0331
WellsFargo       -0.0026   0.0016     0.0000           0e+00     0.0006
RoyalDutchShell  -0.0003   0.0002     0.0000           0e+00     0.0001
ExxonMobil        0.0520  -0.0331     0.0006           1e-04     0.0000
```

NOTE!

- Rotation primarily affects the factor **loadings**, redistributing the variance among the factors to make the structure more interpretable. The **communality** and specific variance (**uniqueness**) values remain the same after rotation.
- **Interpret the data … .**

Example 9.9 page 529: Rotated Loading for the consumer-preference data.

In a consumer-preference study, a random sample of customers were asked to rate several attributions of a new product. The response, on a 7-point semantic differential scale, were tabulated and the attribute correlation matrix constructed. The correlation matrix is presented next:

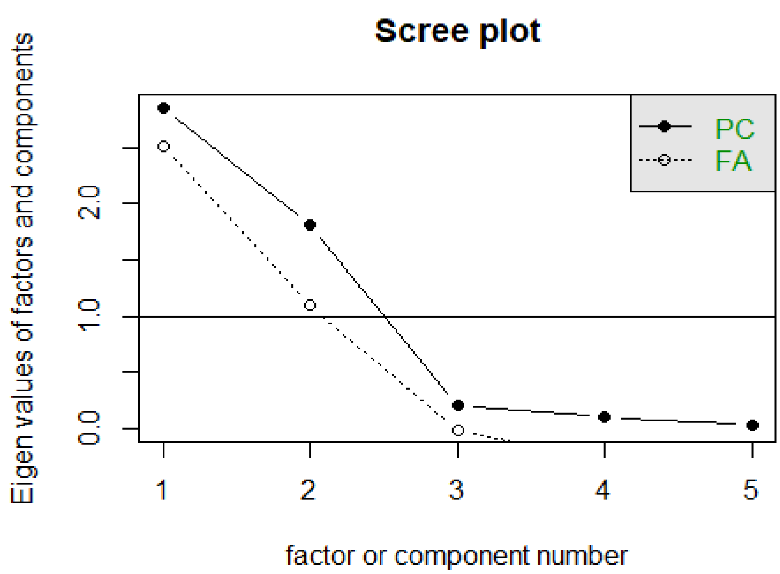
Attribute (Variable)		1	2	3	4	5
Taste	1	1.00	.02	.96	.42	.01
Good buy for money	2	.02	1.00	.13	.71	.85
Flavor	3	.96	.13	1.00	.50	.11
Suitable for snack	4	.42	.71	.50	1.00	.79
Provides lots of energy	5	.01	.85	.11	.79	1.00

a) Do factor analysis by using principal component method.

b) Rotate your solutions in Parts (a). Compare the solutions and comment on them. Interpret each factor. (by Varimax)

c) Determine a reasonable number of factors *m* after rotation. Interpret the factors. (by scree plot)

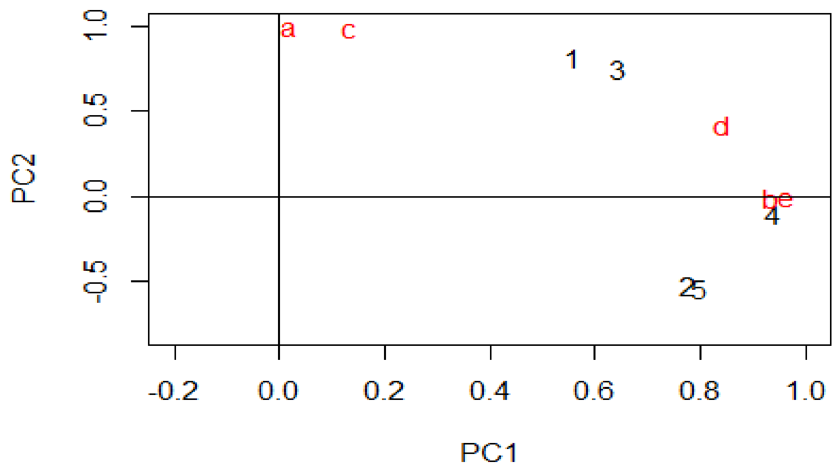
```
rm(list=ls())
library("psych")
R<-matrix(c(1,.02,.96,.42,.01,.02,1,0.13
           ,.71,.85,.96,.13,1,.5,.11,
           .42,.71,.5,1,.79,.01,.85,.11,.79,1)
         ,nrow = 5,byrow=FALSE)
scree(R,pc=TRUE)
```



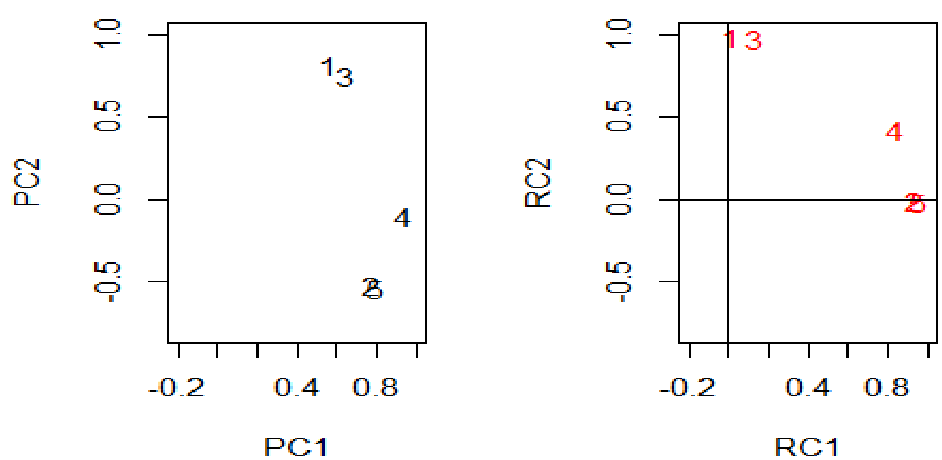
```
## PC Factor Analysis based upon correlation matrix.
pc <- principal(R, nfactors=2, rotate="none",covar=FALSE)
print(pc, cut = NULL, digits=2)

Principal Components Analysis
Call: principal(r = R, nfactors = 2, rotate = "none", covar = FALSE)
Standardized loadings (pattern matrix) based upon correlation matrix
   PC1   PC2   h2    u2 com
1 0.56  0.82 0.98 0.021 1.8
2 0.78 -0.52 0.88 0.121 1.8
3 0.65  0.75 0.98 0.024 2.0
4 0.94 -0.10 0.89 0.107 1.0
5 0.80 -0.54 0.93 0.068 1.8

                       PC1  PC2
SS loadings           2.85 1.81
Proportion Var        0.57 0.36
Cumulative Var        0.57 0.93
Proportion Explained  0.61 0.39
Cumulative Proportion 0.61 1.00

Mean item complexity =  1.7
Test of the hypothesis that 2 components are sufficient.

The root mean square of the residuals (RMSR) is  0.03
Fit based upon off diagonal values = 1


## Rotated loadings by varimax
pc.r <- principal(R, nfactors=2, rotate="varimax",covar=FALSE)
print(pc.r)

Principal Components Analysis
Call: principal(r = R, nfactors = 2, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
   RC1   RC2   h2    u2 com
1 0.02  0.99 0.98 0.021 1.0
2 0.94 -0.01 0.88 0.121 1.0
3 0.13  0.98 0.98 0.024 1.0
4 0.84  0.43 0.89 0.107 1.5
5 0.97 -0.02 0.93 0.068 1.0

                       RC1  RC2
SS loadings           2.54 2.12
Proportion Var        0.51 0.42
Cumulative Var        0.51 0.93
Proportion Explained  0.54 0.46
Cumulative Proportion 0.54 1.00

Mean item complexity =  1.1
```

```
Test of the hypothesis that 2 components are sufficient.
The root mean square of the residuals (RMSR) is  0.03
Fit based upon off diagonal values = 1
```

```
L.r<-pc.r$loadings


# Two plots on the same graph .
plot(L,pch = as.character(c(1:5)),xlim = c(-0.2,1), ylim = c(-0.8,1))
 points(L.r,pch = letters[c(1:9)], col = "red")
 abline(h = 0)
abline(v = 0)
```




```
# split the plot into two columns (2 plots next to each othe)
par(mfrow=c(1,2))
plot(L,pch = as.character(c(1:5)),xlim = c(-0.2,1), ylim = c(-0.8,1))
plot(L.r,pch = as.character(c(1:5)),xlim = c(-0.2,1), ylim = c(-0.8,1),col="r
ed")
 abline(h = 0)
abline(v = 0)
```



Interpret the data based on the rotation (varimax): It is clear that variables "Good buy for money, suitable for snack and provides lots of energy" are define factor 1 (high loadings on factor 1, small loadings on factor 2), while variables "Taste" and "flavor" define factor 2 (high loadings on factor 2, small loadings on factor 1).

We might call factor 1 a nutritional factor and factor 2 a taste factor.

> NOTE! The interpretation of (**varimax**) rotated is much cleaner than that of the original analysis.

Example 9.12: in lecture page 67

Calculate factor scores by the least squares and Regression methods after rotation for first observation of Stock-price data use:

a) Maximum likelihood factor analysis.

b) Principal components analysis.

```
# Example 9.12
data1 <- read.table(choose.files()) #T8-4.DAT
names(data1) <- c("JPMorgan","Citibank","WellsFargo","RoyalDutchShell","Exxon
Mobil")

# part a: Maximum Likelihood Factor with varimax rotation.
## Factor Scores Obtained by "Bartlett" for first observation.
m.b<-factanal(data1,2,rotation="varimax",scores = "Bartlett")
#print(m.b,cutoff = NULL)
Sb<-m.b$scores
Sb[1,]

   Factor1    Factor2
 0.2508994 -1.8536707

#*Explain how to find Bartlett factor Scores for the first observation.
data1.z<- as.data.frame(scale(data1, center = TRUE, scale = TRUE))
z1<-data1.z[1,]  #first "standardized" observation.
psi<-diag(m.b$uniquenesses)
L<-m.b$loadings
(f1.b<- solve(t(L)%*%solve(psi)%*%L) %*% t(L)%*%solve(psi)%*% t(z1))

                 1
Factor1  0.2508994
Factor2 -1.8536707

## Factor Scores Obtained by "Regression" for first observation.
m.r<- factanal(data1,2,rotation="varimax",scores = "regression" )
```

```
Sr<- m.r$scores
Sr[1,]

   Factor1    Factor2 
 0.1653586 -1.8342740 

#*Explain how to find Regression factor Scores for the first observation.
R<- cor(data1)
(f1.r<-t(L)%*% solve(R) %*% t(z1))

                 1
Factor1  0.1653586
Factor2 -1.8342740

# part b: Principal components Factor analysis with varimax rotation
## Bartlett score.
pc.b <- principal(data1, nfactors=2, rotate="varimax",covar=FALSE, scores=TRU
E,method="Bartlett")
#print(pc3,cutoff = NULL)
(sb.p<-pc.b$scores[1,])

        RC1         RC2 
 0.06479191 -1.01648402 

## Regression score
pc.r<- principal(data1, nfactors=2, rotate="varimax",covar=FALSE, scores=TRUE
,method="regression")
(sb.p<-pc.r$scores[1,])

        RC1         RC2 
 0.06479191 -1.01648402 
```