Multivariate Analysis Lecture 17: Factor Analysis

Zhaoxia Yu
Professor, Department of Statistics

2023-06-01

Section 1

Motivating Example

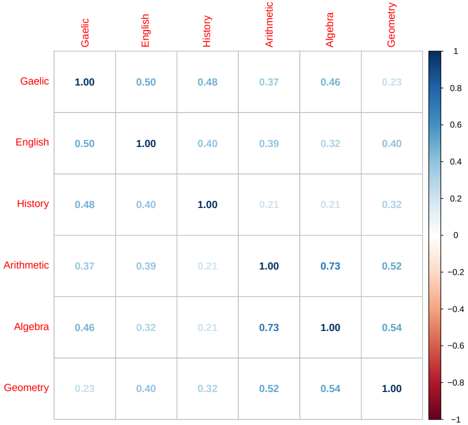
A Motivating Example: Exam Score

```
exam.cor=rbind(
c(1.000,  0.439,  0.410,  0.288,  0.329,  0.248),
c(0.439,  1.000,  0.351,  0.354,  0.320,  0.329),
c(0.410,  0.351,  1.000,  0.164,  0.190,  0.181),
c(0.288,  0.354,  0.164,  1.000,  0.595,  0.470),
c(0.329,  0.320,  0.190,  0.595,  1.000,  0.464),
c(0.248,  0.329,  0.181,  0.470,  0.464,  1.000))
exam.cov=diag(c(2, 3, 2, 3, 2, 2)) %*%exam.cor %*% diag(c(2, 3, 2, 3, 2, 2))

set.seed(2)
exam=round(mvrnorm(n=60, mu=rep(80,6), Sigma=exam.cov))
colnames(exam)=c("Gaelic",
"English",
"History",
"Arithmetic",
"Algebra",
"Geometry")
exam=data.frame(exam)
```

Pairwise Correlation

```r
corrplot(cor(exam), method="number")
```



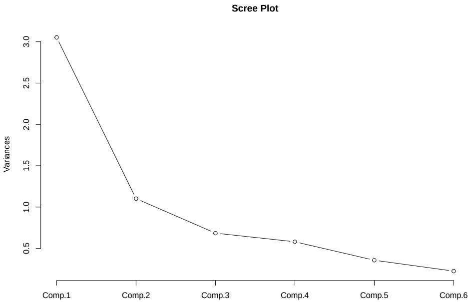
PCA

```
obj=princomp(exam, cor=TRUE)
obj$loadings
```

```
##
## Loadings:
##            Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
## Gaelic      0.403  0.390  0.572         0.488  0.340
## English     0.398  0.322 -0.108 -0.801 -0.168 -0.238
## History     0.328  0.583 -0.324  0.552 -0.358 -0.122
## Arithmetic  0.448 -0.416  0.158        -0.581  0.514
## Algebra     0.453 -0.404  0.274  0.218        -0.708
## Geometry    0.408 -0.262 -0.676         0.511  0.217
##
##                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
## SS loadings     1.000  1.000  1.000  1.000  1.000  1.000
## Proportion Var  0.167  0.167  0.167  0.167  0.167  0.167
## Cumulative Var  0.167  0.333  0.500  0.667  0.833  1.000
```

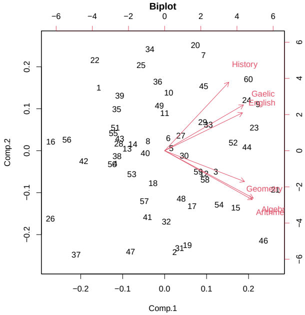
Visualize Eigenvalues

```
plot(obj, type="lines", main="Scree Plot")
```



Visualize First and Second PC (biplot)

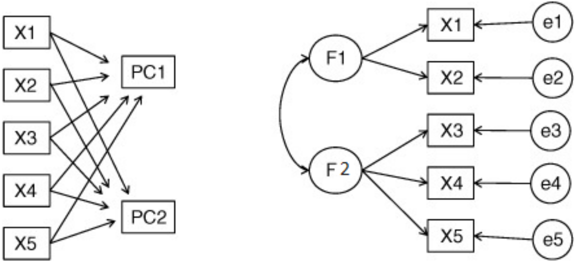
```
biplot(obj, main="Biplot")
```



- How to interpret the results?

Introduction to Factor Analysis

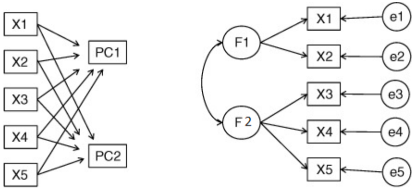
```r
knitr::include_graphics("img/PCA_FA.png")
```



(a) Principal Components Model (b) Factor Analysis Model

https://livebook.manning.com/book/r-in-action-second-edition/chapter-14/6

PCA vs FA



(a) Principal Components Model (b) Factor Analysis Model

- Both reduce dimensionality
- Both use linear combinations
- PCA leads to principal components, which are linear combinations of functions
- FA leads to factors (latent and unobserved)

Section 2

The FA Model

The Factor Model

- Consider a random vector $\mathbf{X} \in \mathbb{R}^p$
- Let $\boldsymbol{\mu} \in \mathbb{R}^p$ denote the population mean
- Let $F \in \mathbb{R}^m$ denote m factors

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}, \mathbf{F} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{pmatrix}$$

The Factor model

- X_j, which is observable, is assumed to be a linear function of the unobservable common factors $f_1, \cdots, f_m$ plus specific errors.

$$
\begin{aligned}
X_1 &= \mu_1 + l_{11} f_1 + l_{12} f_2 + \cdots + l_{1m} f_m + \epsilon_1 \\
X_2 &= \mu_2 + l_{21} f_1 + l_{22} f_2 + \cdots + l_{2m} f_m + \epsilon_2 \\
&\vdots \\
X_p &= \mu_p + l_{p1} f_1 + l_{p2} f_2 + \cdots + l_{pm} f_m + \epsilon_p
\end{aligned}
$$

  where ϵ_j is called the specific factor for feature j.
- The means $\mu_1, \cdots, \mu_p$ are parameters
- The coefficients in the factor loading matrix are also parameters

The Factor model

- Let $\mathbf{L}$ denote the $p \times m$ matrix of factor loadings

$$\mathbf{L} = \begin{pmatrix} l_{11} & l_{12} & \cdots & l_{1m} \\ l_{21} & l_{22} & \cdots & l_{2m} \\ \vdots & \vdots & & \vdots \\ l_{p1} & l_{p2} & \cdots & l_{pm} \end{pmatrix}$$

- A compact expression of the factor model is

$$\mathbf{X}_{p\times 1} = \boldsymbol{\mu}_{p\times 1} + \mathbf{L}_{p\times m}\mathbf{F}_{m\times 1} + \boldsymbol{\epsilon}_{p\times 1}$$

Assumptions of FA

- $\mathbf{F}$ and ϵ are uncorrelated
- The common factors are uncorrelated

$$\mathbb{E}(\mathbf{F}) = 0, Cov(\mathbf{F}) = \mathbf{I}$$

- The specific factors are uncorrelated

$$\mathbb{E}(\epsilon) = 0, Cov(\epsilon) = \Psi$$

where Ψ is a diagonal matrix with non-negative values, i.e.,

$$\boldsymbol{\Psi} = \begin{pmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_p \end{pmatrix}$$

The Covariance

- By the factor model and its assumptions, we have

$$\begin{aligned}\Sigma &= cov(\mathbf{X}) \\ &= cov(\mathbf{LF} + \epsilon) \\ &= \mathbf{L}Cov(\mathbf{F})\mathbf{L}^T + \Psi \\ &= \mathbf{LL}^T + \Psi\end{aligned}$$

The last step is due to our assumption that $cov(\mathbf{F}) = \mathbf{I}$

The Covariance

- For $i \neq j$, the covariance between X_i (feature i) and X_j (feature j) is

$$\sigma_{ij} = cov(X_i, X_j) = \sum_{k=1}^{m} l_{ik} l_{jk}$$

- The variance of X_i is

$$\sigma_{ii} = \sum_{k=1}^{m} l_{ik}^2 + \psi_i$$

Section 3

Communality

Communality and Specific Variance

- From last slide

$$\sigma_{ii} = \sum_{k=1}^{m} l_{ik}^2 + \psi_i$$

- We say that the variance of X_i is partitioned into communality and specific variance where
  - communality is defined as $h_i^2 = \sum_{k=1}^{m} l_{ik}^2$, which is the proportion of variance contributed by common factors
  - specific variance ψ_i, which is the specific variance of X_i

Example

```
obj=factanal(exam, factors=2)
L=obj$loadings[,1:2]
Psi=diag(obj$uniquenesses)
S=diag(sqrt(diag(cov(exam))))
round(cov(exam),2)
```

```
##            Gaelic English History Arithmetic Algebra Geometry
## Gaelic       4.21    3.53    2.17       2.50    2.06     1.10
## English      3.53   11.69    3.01       4.37    2.37     3.21
## History      2.17    3.01    4.84       1.49    1.02     1.66
## Arithmetic   2.50    4.37    1.49      10.73    5.25     4.01
## Algebra      2.06    2.37    1.02       5.25    4.76     2.73
## Geometry     1.10    3.21    1.66       4.01    2.73     5.44
```

Example

```
obj=factanal(exam, factors=2)
L=obj$loadings[,1:2]
Psi=diag(obj$uniquenesses)
S=diag(sqrt(diag(cov(exam))))
round(cov(exam),2)
```

```
##            Gaelic English History Arithmetic Algebra Geometry
## Gaelic       4.21    3.53    2.17       2.50    2.06     1.10
## English      3.53   11.69    3.01       4.37    2.37     3.21
## History      2.17    3.01    4.84       1.49    1.02     1.66
## Arithmetic   2.50    4.37    1.49      10.73    5.25     4.01
## Algebra      2.06    2.37    1.02       5.25    4.76     2.73
## Geometry     1.10    3.21    1.66       4.01    2.73     5.44
```

```
#FA models correlation, L%*%t(L)+ Psi is estimated corr
round(  S%*% (L%*%t(L))%*% S + S%*%Psi%*%S, 2)
```

```
##      [,1]  [,2] [,3]  [,4] [,5] [,6]
## [1,] 4.21  3.38 2.11  2.68 1.85 1.76
## [2,] 3.38 11.69 3.28  3.94 2.72 2.64
## [3,] 2.11  3.28 4.84  1.56 1.06 1.27
## [4,] 2.68  3.94 1.56 10.73 5.27 3.90
## [5,] 1.85  2.72 1.06  5.27 4.76 2.72
## [6,] 1.76  2.64 1.27  3.90 2.72 5.44
```

```
S%*% (L%*%t(L))%*% S
```

Section 4

Non-uniqueness

Non-uniqueness of Factor Loadings

- The factor loading coefficient is NOT unique.
- Suppose $\mathbf{X} = \boldsymbol{\mu} + \mathbf{Lf} + \boldsymbol{\epsilon}$
- Consider any $m \times m$ orthogonal matrix Γ, which satisfies $\Gamma\Gamma^T = \Gamma^T\Gamma = \mathbf{I}$.
- Let $\tilde{\mathbf{L}} = L\gamma$ The model $\mathbf{X} = \boldsymbol{\mu} + \tilde{\mathbf{L}}\mathbf{F} + \boldsymbol{\epsilon}$

give the same Σ because

$$cov(\tilde{\mathbf{L}}\mathbf{F}) = \tilde{\mathbf{L}}cov(\mathbf{F})\tilde{\mathbf{L}}^T = \tilde{\mathbf{L}}\tilde{\mathbf{L}}^T = \mathbf{L}``^T\mathbf{L}^T = \mathbf{L}\mathbf{L}^T = cov(\mathbf{LF})$$

Non-uniqueness of Factor Loadings

```r
#Estimated Sigma
L%*%t(L) + Psi
```

```
##               Gaelic   English   History Arithmetic   Algebra  Geometry
## Gaelic     1.0000011 0.4819908 0.4661743  0.3979476 0.4132344 0.3682496
## English    0.4819908 0.9999986 0.4363080  0.3514418 0.3645679 0.3308447
## History    0.4661743 0.4363080 0.9999990  0.2158592 0.2209179 0.2483943
## Arithmetic 0.3979476 0.3514418 0.2158592  1.0000009 0.7377925 0.5106265
## Algebra    0.4132344 0.3645679 0.2209179  0.7377925 1.0000003 0.5355603
## Geometry   0.3682496 0.3308447 0.2483943  0.5106265 0.5355603 0.9999993
```

Non-uniqueness of Factor Loadings

- Consider a rotation matrix R and define $\tilde{\mathbf{L}} = LR$

```
theta=pi/6
R=matrix(c(cos(theta), sin(theta), -sin(theta), cos(theta)), 2,2)
L.tilde=L%*%R

L.tilde %*% t(L.tilde) + Psi
```

```
##               Gaelic   English   History Arithmetic   Algebra  Geometry
## Gaelic     1.0000011 0.4819908 0.4661743  0.3979476 0.4132344 0.3682496
## English    0.4819908 0.9999986 0.4363080  0.3514418 0.3645679 0.3308447
## History    0.4661743 0.4363080 0.9999990  0.2158592 0.2209179 0.2483943
## Arithmetic 0.3979476 0.3514418 0.2158592  1.0000009 0.7377925 0.5106265
## Algebra    0.4132344 0.3645679 0.2209179  0.7377925 1.0000003 0.5355603
## Geometry   0.3682496 0.3308447 0.2483943  0.5106265 0.5355603 0.9999993
```

Section 5

Computation

Subsection 1

Method 1: Use PCA

Method 1: Use PCA

- By the spectral decomposition of Σ we have

$$\Sigma = \Gamma \Lambda \Gamma^T$$

  where $\Gamma = (\gamma_1, \cdots, \gamma_p)$ is an orthogonal matrix and $\Lambda = diag(\lambda_1, \cdots, \lambda_p)$ be the diagonal matrix of eigenvalues.
- The spectral decomposition can be rewritten to

$$\boldsymbol{\Sigma} = \sum_{i=1}^{p} \lambda_i \gamma_i \gamma_i^T = \sum_{i=1}^{p} (\sqrt{\lambda_i}\gamma_i)(\sqrt{\lambda_i}\gamma_i)^T$$

Subsection 2

Method 1: Use PCA

Method 1: Use PCA

- Suppose that $\lambda_m, \lambda_{m+1}, \cdots, \lambda_p$ are small. Then

$$\mathbf{\Sigma} \approx \sum_{i=1}^{m} (\sqrt{\lambda_i}\gamma_i)(\sqrt{\lambda_i}\gamma_i)^T$$

- Let $\mathbf{L} = (\sqrt{\lambda_1}\gamma_1, \cdots, \sqrt{\lambda_m}\gamma_m)$
- Let $\mathbf{\Psi} = \mathbf{\Sigma} - \mathbf{L}\mathbf{L}^T$

Subsection 3

Method 2: MLE

Method 2: MLE

- We impose multivariate normality on the common and specific factors

$$\mathbf{f} \sim N(\mathbf{0}, \mathbf{I}), \epsilon \sim N(\mathbf{0}, \Psi)$$

- The log-likelihood is

$$l(\mu, \mathbf{L}, \Psi) = -\frac{np}{2}\log 2\pi - \frac{n}{2}\log|\mathbf{L}\mathbf{L}^T + \Psi| - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{X}_i - \mu)^T(\mathbf{L}\mathbf{L}^T + \Psi)^{-1}(\mathbf{X}_i - \mu)$$

where $\mathbf{X}_i \in \mathbb{R}^p$ denotes the i observation (not the ith feature).

The Number of Common Factors

- The $p \times p$ covariance matrix Σ is symmetric. As a result, there are $\frac{p(p+1)}{2}$ parameters.
- A factor mode imposes a structure on Σ
- For a FA model with m common factors
- A FA model a small number common factors, i.e., when m is small, the model uses fewer parameters
  - the model is more parsimonious
  - the model might not be adequate is m is too small
- One can test whether an m is large enough

Choose *m* of Factors Computed using PCA

- Cumulatative variance explained is should be reasonably large, such as $>80\%$
- Look for elbow from the scree plot

A Goodness of Fit Test for the Adequacy of the Number of Common Factors

```r
knitr::include_graphics("img/TestNumberCommonFactors.png")
```

$$H_0: \underset{(p\times p)}{\mathbf{\Sigma}} = \underset{(p\times m)}{\mathbf{L}}\ \underset{(m\times p)}{\mathbf{L}'} + \underset{(p\times p)}{\mathbf{\Psi}}$$

$$-2\ln\Lambda = -2\ln\left[\frac{\text{maximized likelihood under } H_0}{\text{maximized likelihood}}\right]$$

$$= -2\ln\left(\frac{|\hat{\mathbf{\Sigma}}|}{|\mathbf{S}_n|}\right)^{-n/2} + n\left[\operatorname{tr}(\hat{\mathbf{\Sigma}}^{-1}\mathbf{S}_n) - p\right]$$

It can be shown that $\operatorname{tr}(\hat{\mathbf{\Sigma}}^{-1}\mathbf{S}_n) - p = 0$

$$= n\ln\left(\frac{|\hat{\mathbf{\Sigma}}|}{|\mathbf{S}_n|}\right)$$

Test for the Adequacy of the Number of Common Factors

- The number of parameters for covariance in the full model is

$$\frac{p(p+1)}{2}$$

- The number of parameters for covariance in the reduced model is

$$mp + p - \frac{m(m-1)}{2}$$

Note: $-\frac{m(m-1)}{2}$ is due to the nonuniqueness of $\mathbf{L}$.

- The difference is

$$\begin{aligned} df &= \frac{p(p+1)}{2} - [mp + p - \frac{m(m-1)}{2}] \\ &= \frac{1}{2}[(p-m)^2 - p - m] \end{aligned}$$

Test for the Adequacy of the Number of Common Factors

- The result indicates that 1 factor might be inadequate as the p-value is small.

```
factanal(exam, factors=1)
```

```
##
## Call:
## factanal(x = exam, factors = 1)
##
## Uniquenesses:
##     Gaelic     English     History Arithmetic     Algebra    Geometry
##      0.714       0.752       0.876      0.315       0.289       0.593
##
## Loadings:
##            Factor1
## Gaelic     0.535
## English    0.498
## History    0.353
## Arithmetic 0.828
## Algebra    0.843
## Geometry   0.638
##
##                Factor1
## SS loadings      2.462
## Proportion Var   0.410
##
```

Test for the Adequacy of the Number of Common Factors

- The result indicates that 2 factors is adequate becaue the fit is not substantially from the full model.

```r
factanal(exam, factors=2)
```

```
##
## Call:
## factanal(x = exam, factors = 2)
##
## Uniquenesses:
##     Gaelic    English    History Arithmetic    Algebra   Geometry
##      0.478      0.554      0.534      0.298      0.224      0.607
##
## Loadings:
##            Factor1 Factor2
## Gaelic     0.323   0.646
## English    0.275   0.608
## History            0.676
## Arithmetic 0.811   0.211
## Algebra    0.855   0.213
## Geometry   0.554   0.294
##
##                Factor1 Factor2
## SS loadings      1.882   1.421
## Proportion Var   0.314   0.237
## Cumulative Var   0.314   0.551
```

Section 6

Factor Rotation

Subsection 1

Rotation for Better Interpretation

Rotation for Better Interpretation

- Interpretation of final results are easier for some choices of **L** than others.
- We often rotate the factors to gain insights or for better interpretation
- This is one advantage of factor analysis
- In practice,
  - Step 1: fit a factor model by imposing conditions that lead to a unique solution
  - Step 2: the loading matrix L is rotated (multiplied by an orthogonal matrix) in a way that gives a good interpretation of the data. Trial and error
- Well know criteria of rotation exist

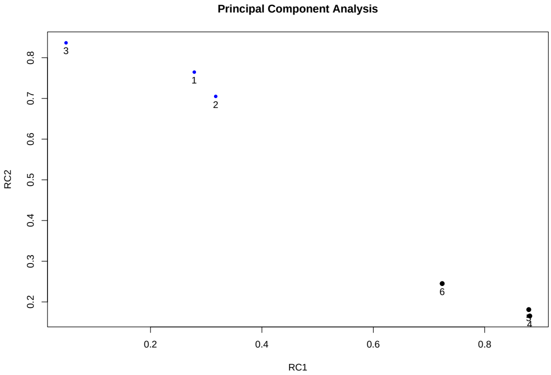
Factor Rotation

```
principal(exam, nfactors=2)
```

```
## Principal Components Analysis
## Call: principal(r = exam, nfactors = 2)
## Standardized loadings (pattern matrix) based upon correlation matrix
##             RC1  RC2   h2   u2 com
## Gaelic     0.28 0.76 0.66 0.34 1.3
## English    0.32 0.70 0.60 0.40 1.4
## History    0.05 0.84 0.70 0.30 1.0
## Arithmetic 0.88 0.17 0.80 0.20 1.1
## Algebra    0.88 0.18 0.81 0.19 1.1
## Geometry   0.72 0.25 0.58 0.42 1.2
##
##                        RC1  RC2
## SS loadings           2.25 1.90
## Proportion Var        0.38 0.32
## Cumulative Var        0.38 0.69
## Proportion Explained  0.54 0.46
## Cumulative Proportion 0.54 1.00
##
## Mean item complexity =  1.2
## Test of the hypothesis that 2 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.11
##  with the empirical chi square  22.04  with prob <  2e-04
##
## Fit based upon off diagonal values = 0.93
```

Factor Rotation

```
plot(principal(exam, nfactors=2))
```



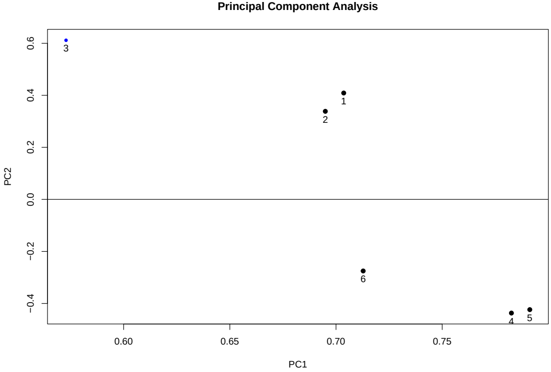
Factor Rotation

```r
principal(exam,nfactors=2, rotate="none")
```

```
## Principal Components Analysis
## Call: principal(r = exam, nfactors = 2, rotate = "none")
## Standardized loadings (pattern matrix) based upon correlation matrix
##             PC1   PC2   h2   u2 com
## Gaelic     0.70  0.41 0.66 0.34 1.6
## English    0.69  0.34 0.60 0.40 1.4
## History    0.57  0.61 0.70 0.30 2.0
## Arithmetic 0.78 -0.44 0.80 0.20 1.6
## Algebra    0.79 -0.42 0.81 0.19 1.5
## Geometry   0.71 -0.28 0.58 0.42 1.3
##
##                        PC1  PC2
## SS loadings           3.05 1.10
## Proportion Var        0.51 0.18
## Cumulative Var        0.51 0.69
## Proportion Explained  0.73 0.27
## Cumulative Proportion 0.73 1.00
##
## Mean item complexity =  1.6
## Test of the hypothesis that 2 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.11
##  with the empirical chi square  22.04  with prob <  2e-04
##
## Fit based upon off diagonal values = 0.93
```

Factor Rotation

```
plot(principal(exam,nfactors=2, rotate="none"))
```



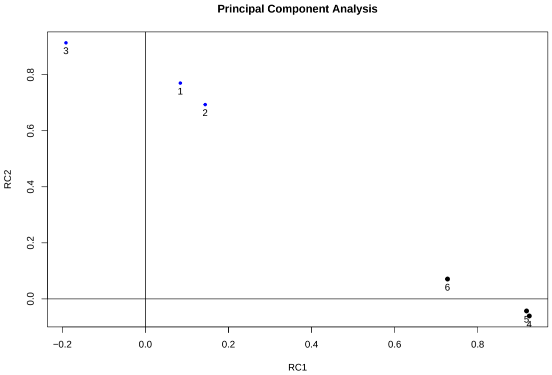
Factor Rotation

```
principal(exam,nfactors=2, rotate="promax")
```

```
## Principal Components Analysis
## Call: principal(r = exam, nfactors = 2, rotate = "promax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##              RC1   RC2   h2   u2 com
## Gaelic      0.08  0.77 0.66 0.34 1.0
## English     0.14  0.69 0.60 0.40 1.1
## History    -0.19  0.91 0.70 0.30 1.1
## Arithmetic  0.92 -0.06 0.80 0.20 1.0
## Algebra     0.92 -0.04 0.81 0.19 1.0
## Geometry    0.73  0.07 0.58 0.42 1.0
##
##                        RC1  RC2
## SS loadings           2.26 1.89
## Proportion Var        0.38 0.32
## Cumulative Var        0.38 0.69
## Proportion Explained  0.54 0.46
## Cumulative Proportion 0.54 1.00
##
##  With component correlations of
##      RC1  RC2
## RC1 1.00 0.48
## RC2 0.48 1.00
##
## Mean item complexity =  1
## Test of the hypothesis that 2 components are sufficient.
##
```

Factor Rotation

```
plot(principal(exam,nfactors=2, rotate="promax"))
```



Section 7

Confirmatory FA

Exploratory or Confirmatory Factor Analysis

- Exploratory Factory Analysis (EFA)
- The FA approach we have discussed is exploratory in nature.
- In fact, we can perform EFA and identify latent factors by using only correlations, not the data
- The purpose of EFA is to explore the possible underlying structure that can explain the observed correlations
- EFA is used when researchers do not have a specific idea about the underlying structure of data
- EFA is hypothesis-generating

Exploratory or Confirmatory Factor Analysis

- Confirmatory Factory Analysis (EFA) is used when a researcher has specific hypotheses or theories about the factor structure of their data.
- It is a “theory-driven” approach.
- In CFA, the researcher specifies the number of factors and which variables load onto which factors.
- CFA is typically used in later stages of research to test or confirm the factor structure suggested by EFA
- CFA is hypothesis testing. A pre-specified model is required

Exploratory or Confirmatory Factor Analysis

- Use EFA when:
  - You are unsure about the underlying structure.
  - You aim to uncover complex patterns.
  - You need to form hypotheses and develop theory.
- Use CFA when:
  - You have a predetermined theory or model.
  - You aim to test the hypothesis about the factor structure.
  - You need to confirm or disconfirm theories.