MVA Project - Factor Analysis

Chun-Jung Chen & Akshay Arora

Question 1: Deos Family conditions affect students' final grade in Math?

Factor Analysis

Eigenvalue of variables

```
d3q1.pc = principal(d3.q1[-1], nfactors = 3, rotate = "varimax")
round(d3q1.pc$values, 3)
```

```
## [1] 1.189 1.026 0.955 0.831
```

Cross-Loading

- Ratio for Family size
  - $\frac{0.47}{0.10} = 4.6 \rightarrow$ ignorable

- Ratio for Parents Status
  - $\frac{0.67}{0.038} = 17.53 \rightarrow$ ignorable

```
d3q1.pc$loadings
```

```
##
## Loadings:
##         RC1    RC3    RC2
## famsize -0.687 -0.321
## Pstatus  0.817 -0.195
## famrel                  0.999
## famsup          0.942
##
##                  RC1   RC3   RC2
## SS loadings     1.140 1.029 1.000
## Proportion Var  0.285 0.257 0.250
## Cumulative Var  0.285 0.542 0.792
```

Communalities

- All variables with communalities less than 0.50 woulc identify as having sufficient explanation
- According to the result, all variables are considered as having sufficient explanation
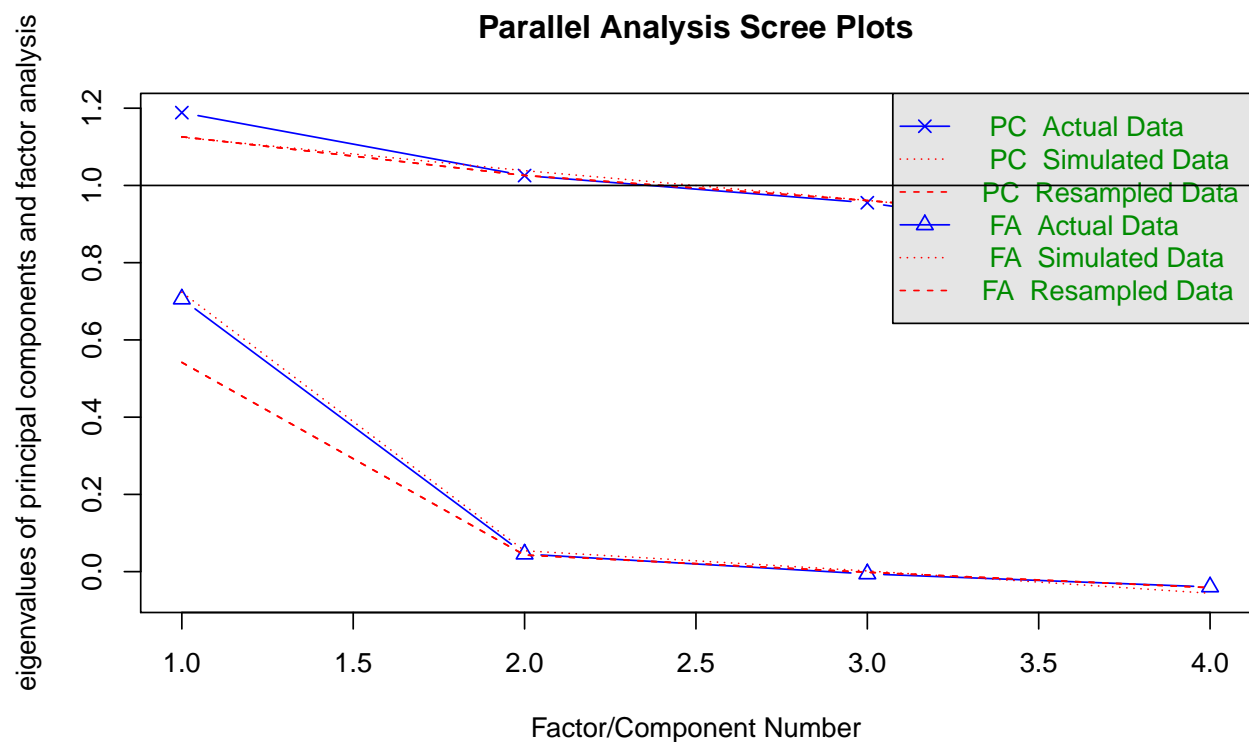
```
# Communalities
d3q1.pc$communality
```

```
##   famsize   Pstatus    famrel    famsup
## 0.5745719 0.7075727 0.9988116 0.8884869
```

Scree Plot

- The reasonable factor number would be 2 or 3.

```
# Factor recommendation
fa.parallel(d3.q1[-1], show.legend = T)
```
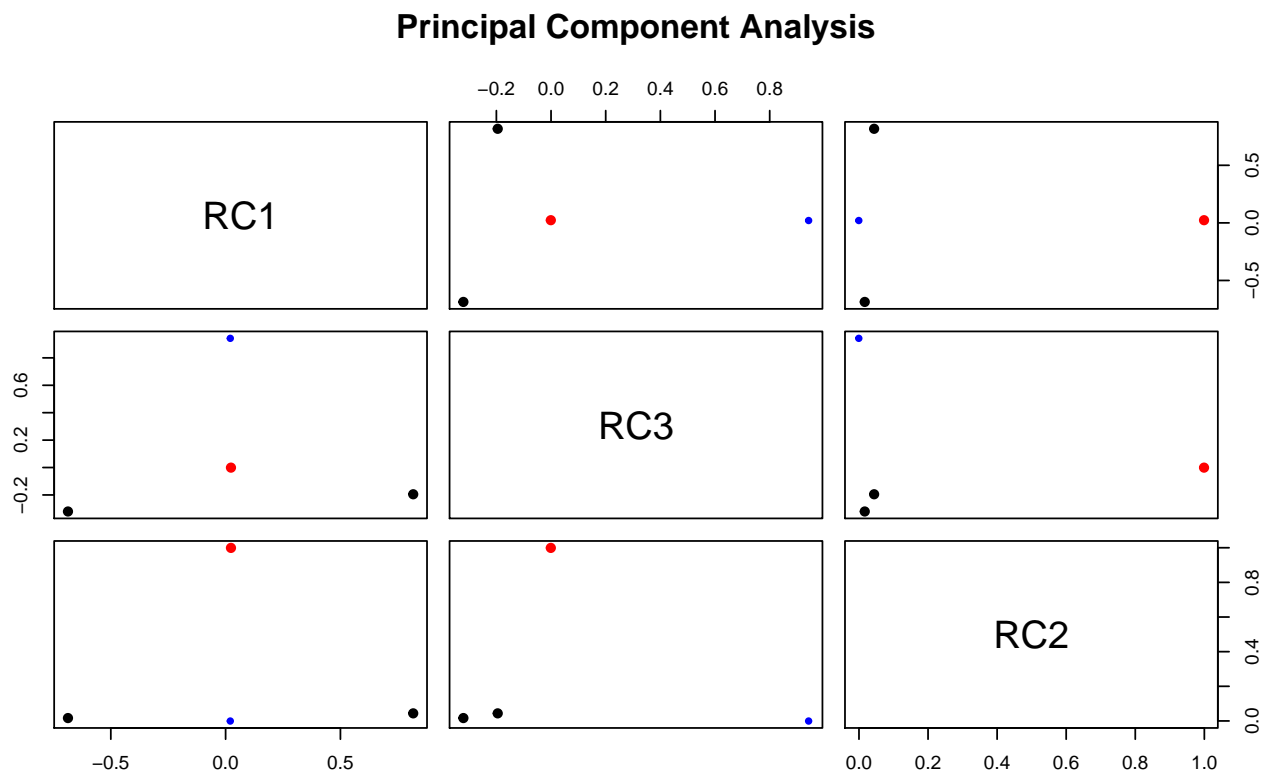

Parallel Analysis Scree Plots

```
## Parallel analysis suggests that the number of factors =  0  and the number of components =  1
```

Factor scatter plot

- It is hard to find the pattern through the scattor plot.

```
fa.plot(d3q1.pc)
```
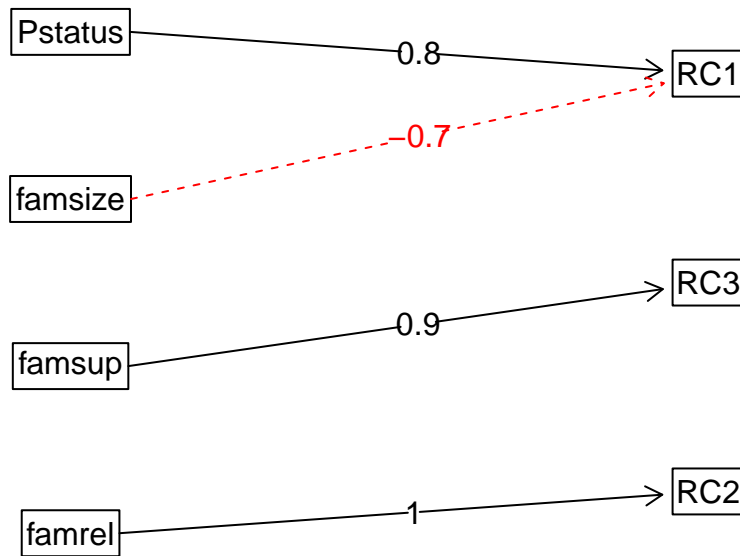
Principal Component Analysis

Component Diagram

- According to the Factor Analysis, there are three latern factors, which could be interpreted as parents status, faily support, and family relationship level.

```
# Visualize the relationship
fa.diagram(d3q1.pc)
```

Components Analysis

Pstatus			

Pstatus ———0.8———→ RC1

famsize ---−0.7---→ RC1

famsup ———0.9———→ RC3

famrel ———1———→ RC2

Question 2: Does parents' jobs and education level influence students' first period of grade in Math?

Factor Analysis

Eigenvalue of variables

```r
d3q2.pc = principal(d3.q2[-1], nfactors = 3, rotate = "varimax")

round(d3q2.pc$values, 3)
```

```
## [1] 2.003 0.926 0.757 0.314
```

Cross-Loading

- Ratio for Mother Education level
  - $\frac{0.677}{0.153} = 4.41 \rightarrow$ ignorable

- Ratio for Father Education level
  - $\frac{0.87}{0.015} = 56.33 \rightarrow$ ignorable

- Ratio for Mother Job Type
  - $\frac{0.93}{0.011} = 81.20 \rightarrow$ ignorable

```r
d3q2.pc$loadings
```

```
##
## Loadings:
##      RC1   RC3   RC2
## Medu 0.823 0.392
## Fedu 0.933       0.124
## Mjob 0.173 0.965 0.107
## Fjob             0.990
##
##                 RC1   RC3   RC2
## SS loadings    1.586 1.094 1.006
## Proportion Var 0.396 0.273 0.252
## Cumulative Var 0.396 0.670 0.921
```

Communalities

- All variables with communalities less than 0.50 woulc identify as having sufficient explanation
- According to the result, all variables are considered as having sufficient explanation
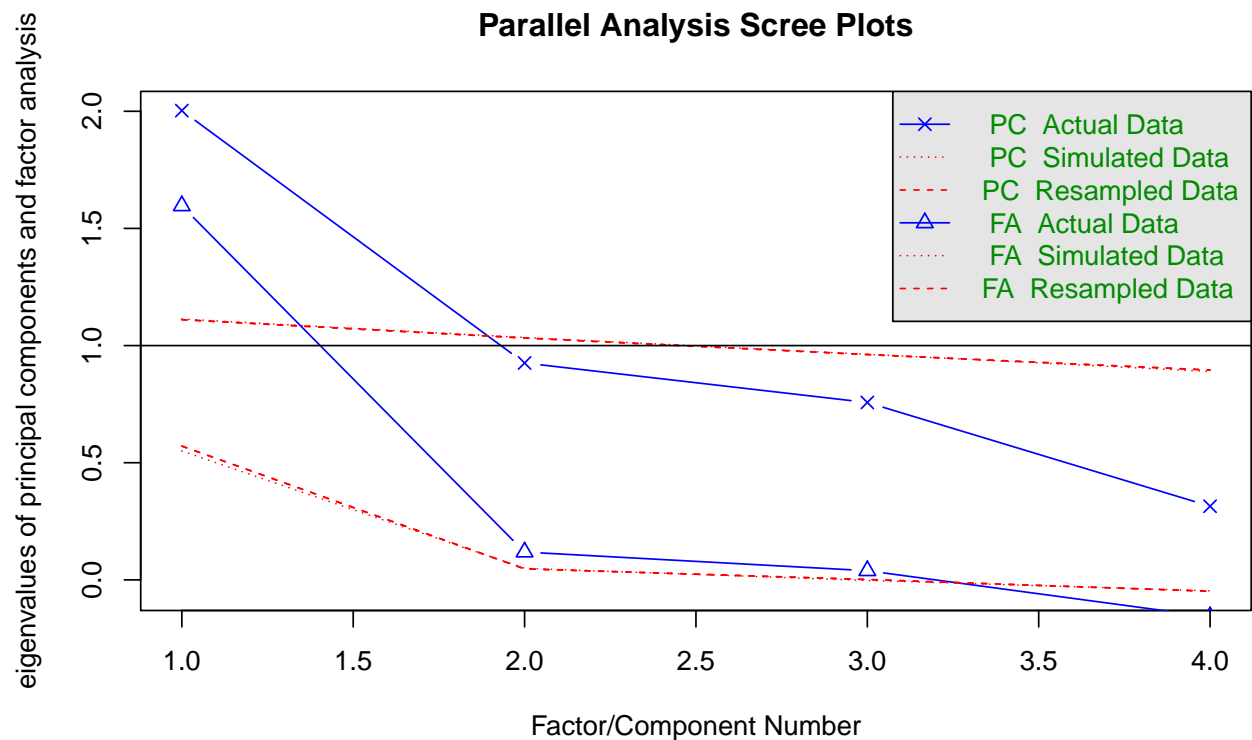
```
# Communalities
d3q2.pc$communality
```

```
##      Medu      Fedu      Mjob      Fjob
## 0.8305494 0.8863545 0.9717988 0.9969812
```

Scree Plot

- The reasonable factor number would be 2 or 3.

```
# Factor recommendation
fa.parallel(d3.q2[-1], show.legend = T)
```
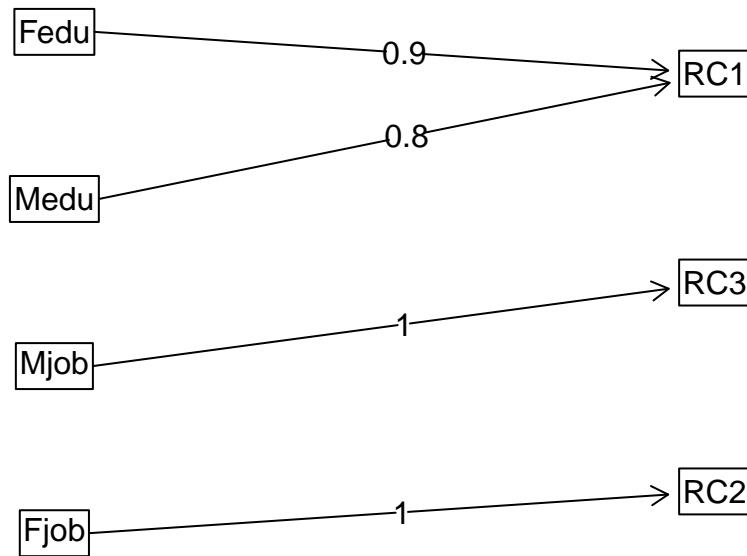

Parallel Analysis Scree Plots

```
## Parallel analysis suggests that the number of factors =  3  and the number of components =  1
```

Component Diagram

- According to the Factor Analysis, there are three latern factors, which could be interpreted as parents education level, Mother's Job, and Father's Job.

```
# Visualize the relationship
fa.diagram(d3q2.pc)
```

Components Analysis

Question 3: Does student's learning conditions really impact students' final grade math score and Portuguese scores in average?

Factor Analysis

Eigenvalue of variables

```
d3q3.pc = principal(d3.q3[, c(2, 4, 5, 6)], nfactors = 3, rotate = "varimax")

round(d3q3.pc$values, 3)

## [1] 1.140 1.012 0.961 0.887
```

Cross-Loading

- Ratio for Internet Assess
    - $\frac{0.291}{0.012} = 23.74 \rightarrow$ ignorable

- Ratio for freetime
    - $\frac{0.773}{0.0164} = 47.12 \rightarrow$ ignorable

```
d3q3.pc$loadings
```

```
##
## Loadings:
##                 RC1    RC3    RC2
## internet       0.560  0.436
## romantic              0.911
## freetime       0.868 -0.143
## normtraveltime               0.997
##
##                 RC1    RC3    RC2
## SS loadings    1.070  1.041  1.002
## Proportion Var 0.267  0.260  0.250
## Cumulative Var 0.267  0.528  0.778
```

Communalities

- All variables with communalities less than 0.50 woulc identify as having sufficient explanation
- According to the result, all variables are considered as having sufficient explanation
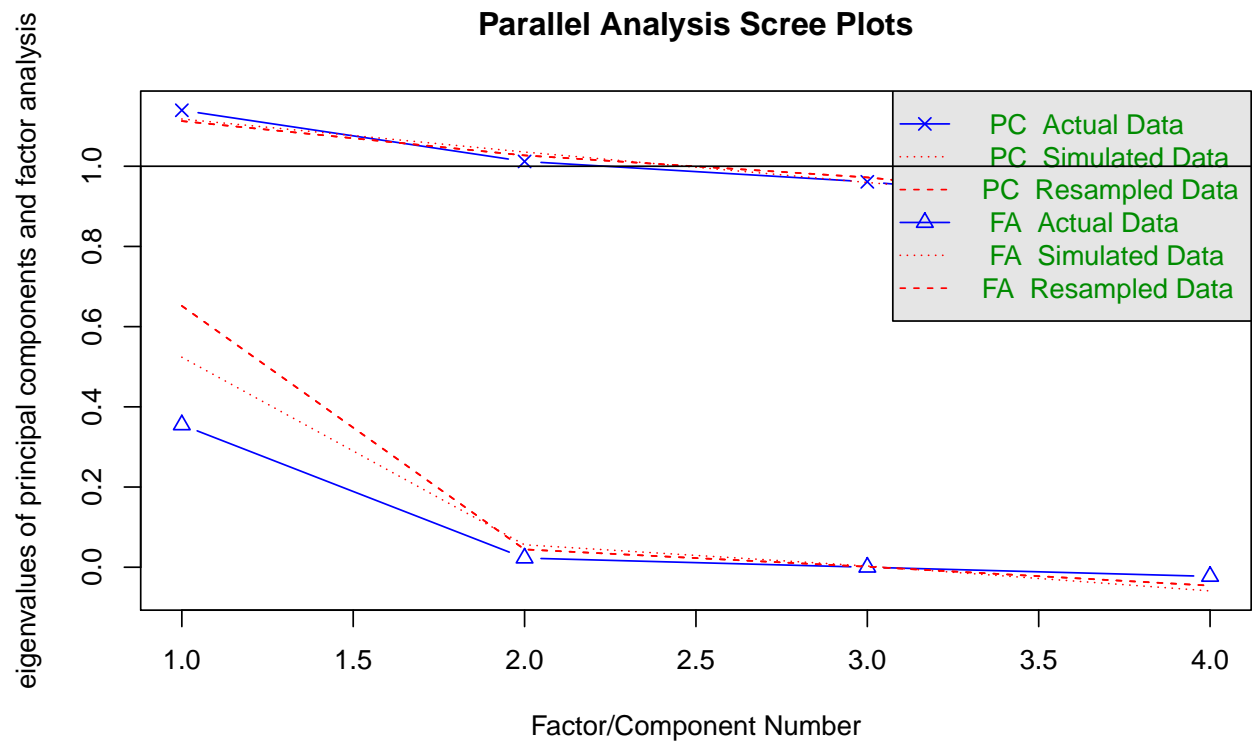
```
# Communalities
d3q3.pc$communality
```

```
##      internet     romantic     freetime normtraveltime
##     0.5100987    0.8327304    0.7765704      0.9932853
```

Scree Plot

- The reasonable factor number would be 2.

```
fa.parallel(d3.q3[, c(2, 4, 5, 6)], show.legend = T)
```


Parallel Analysis Scree Plots

```
## Parallel analysis suggests that the number of factors =  0  and the number of components =  0
```

Component Diagram

- According to the Factor Analysis, there are three latern factors, which could be interpreted as Spare time, Romantic relationship, and Travel time.

```
# Visualize the relationship
fa.diagram(d3q3.pc)
```

Components Analysis