Factor Analysis with R

By: Jovita and Nikitha,

MBA students of JKSHIM, NITTE, NMAMIT, India 574110

Email: jovita.monteiro.3@gmail.com, jackline.niks@gmail.com

Guide: Prof. Mohsin Ahmed (moshahmed@gmail.com)

Agenda

- Introduction to factor Analysis
- Details:
  - FA in R.
  - Scree plots
  - Eigen values
- Conclusion
- Question and Answers
- References

Introduction: Factor Analysis

- Factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors.

- It is used in market research to analyse surveys.

  – From http://en.wikipedia.org/wiki/Factor_analysis

Introduction: SPSS vs R

- Factor Analysis (FA) is typically done in SPSS, a statistical software package from IBM.

- In this talk we show how similar analysis can be done in R, a free and sophisticated statistical programming language.

- One must keep in mind, that R allows complete control on its FA methods, while SPSS may hide the details from the user.

Details: Read the data into R

Read the data into R.

my.data <- read.csv(file.choose())   # choose factor.csv file

head(my.data) # see the data.

Clean the non numeric data.

my.data = my.data[-c(1,3,7,13,37,38,48,50,57,62)]

my.data[is.na(my.data)] <- 3 # change NA to 3 on scale: 1..5.

data.frame(colnames(my.data)) # see the column names

```
> data.frame(colnames(my.data))

                                            colnames.my.data.
1                                   Reading.is.important.for.MBA
2                                  I.sleep.late.and.read.at.night
3                              I.can.change.the.world.with.my.work
```

Details: Read the data into R

head(my.data) # see some of the data, with shortened col names

```
> head(my.data)   # see some of the data.
   X2 X4 X5 X6 X8 X9 X10 X11 X12 X14 X15 X16 X17 X18 X19 X20 X21 X22 X23 X24 X25 X26 X27 X28 X29 X30 X31 X32 X33 X34 X35
1   5  5  5  4  5  3   2   3   5   3   1   5   4   5   3   2   4   4   3   5   3   2   2   2   3   4   2   4   5   4   4
2   4  3  5  5  1  3   1   3   5   3   1   3   5   5   3   5   1   3   3   1   3   3   3   4   5   2   1   5   5   3   5
3   3  4  3  3  1  3   2   3   3   3   2   4   4   5   2   3   3   3   4   4   3   2   3   3   3   3   2   3   3   3   3
4   4  5  3  2  1  2   3   4   3   1   2   3   4   5   4   1   4   4   4   5   3   2   4   4   5   2   1   5   5   1   3
5   3  5  4  5  2  3   4   2   5   4   4   5   5   4   4   4   5   4   4   5   3   4   2   2   4   2   3   5   3   4
6   4  5  5  1  1  5   5   1   1   1   1   5   5   5   5   5   1   1   3   5   5   5   5   4   1   5   5   5   5   5   1
   X36 X39 X40 X41 X42 X43 X44 X45 X46 X47 X49 X51 X52 X53 X54 X55 X56 X58 X59 X60 X61
1    3   3   1   4   3   3   3   2   3   4   1   5   3   4   2   4   4   4   5   5   5
2    5   1   1   1   3   1   4   5   3   2   2   3   1   4   3   3   1   4   1   2   1
3    3   2   1   3   3   3   3   4   1   3   1   4   3   3   3   3   3   4   3   3
4    1   1   1   4   3   2   4   1   4   4   1   1   2   2   2   5   5   4   4   3   5
5    4   4   1   4   3   2   4   3   3   2   3   3   3   4   2   4   2   5   5   4   5
6    1   1   1   1   1   3   5   1   3   3   3   1   1   5   1   5   3   5   5   1   5
>
```

See the correlated columns

> which(cormat > 0.45 & lower.tri(cormat), arr.ind=T,
  useNames=T)

Name   row col

We.must.write.notes.in.all.classes 14 13

I.learn.best.from.fellow.student.presentations. 26 13

There.are.too.many.activities.and.not.enough.interesting.course.material 21 15

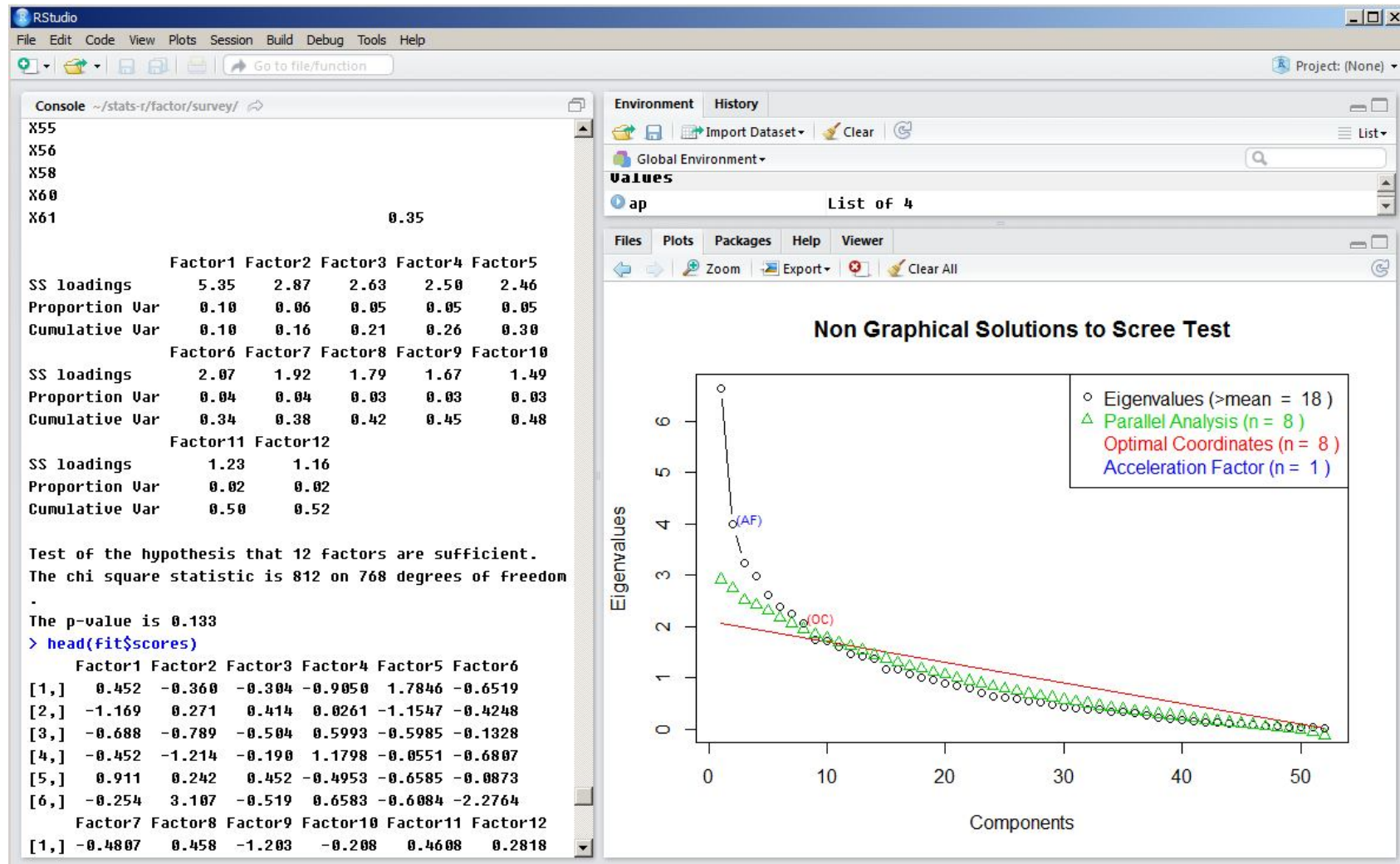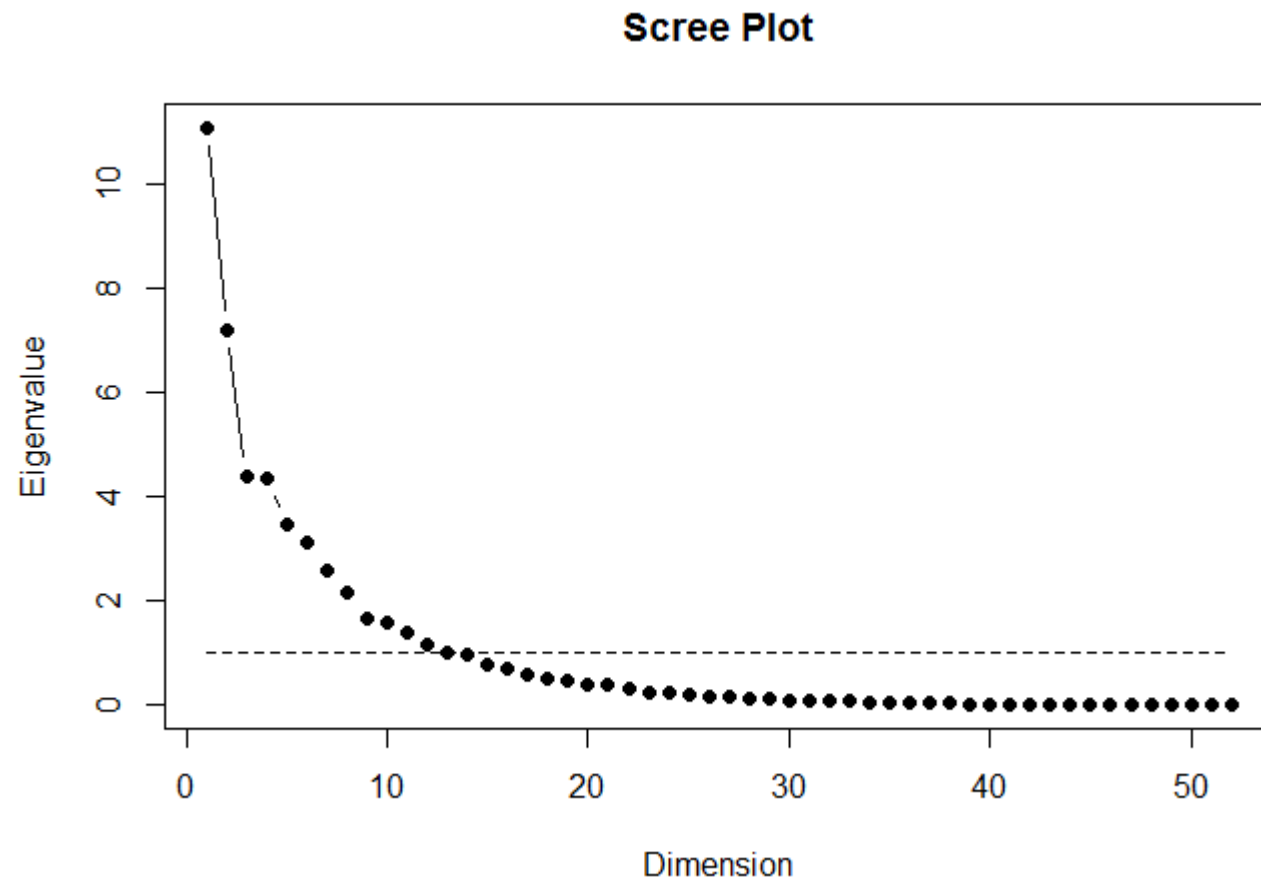I.study.in.a.group.after.college.hours 19 17

I.like.big.classes 33 17

Teachers.empower.me.to.do.great.things 50 17

Teachers.empower.me.to.do.great.things 50 26

R Studio

Scree plot
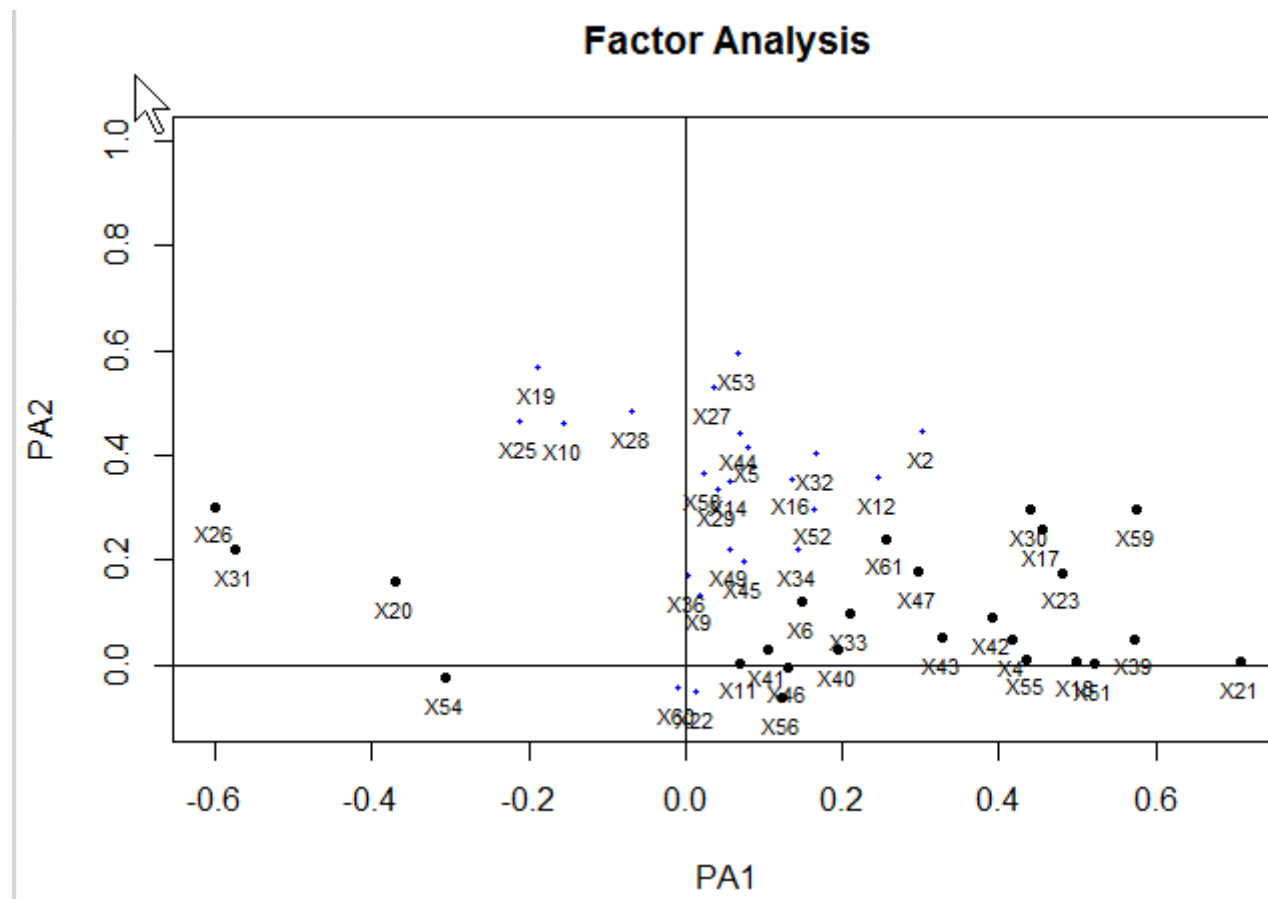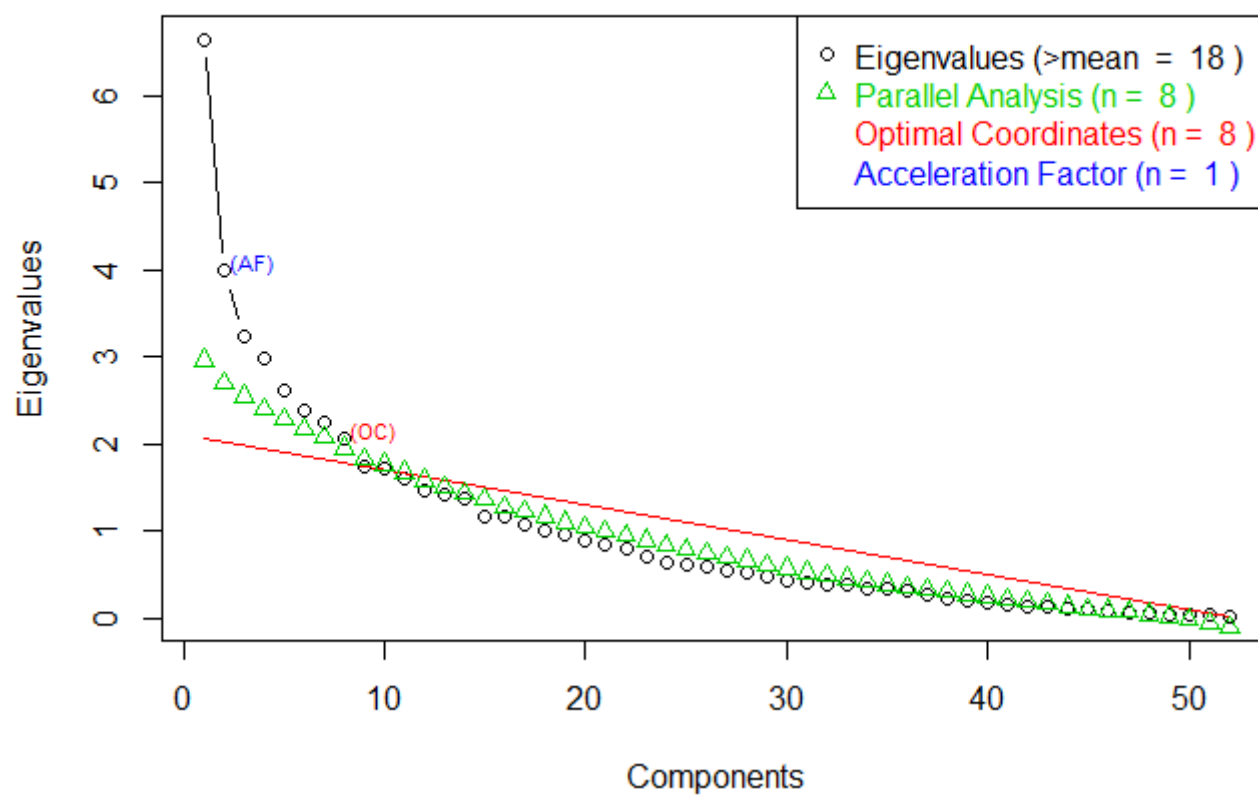
Factor analysis

Non Graphical Solutions to Scree Test

Legend:
- Eigenvalues (>mean = 18)
- Parallel Analysis (n = 8)
- Optimal Coordinates (n = 8)
- Acceleration Factor (n = 1)

(AF)

(OC)

Eigenvalues

Components

Conclusion

- We have shown how to find number of factors using scree plot and eigen vectors in R.

- We also showed which factors are the principal components (factor1 and factor2).

- We recommend MBA students use R for factor analysis and other statistical analysis, as knowing R is a valuable skill asset in management.

References

1. http://en.wikipedia.org/wiki/ Factor_analysis
2. PCA http://www.statsoft.com/ Textbook/Principal-Components-Factor-Analysis
3. SPSS factor analysis http://www.ats.ucla.edu/stat/ spss/output/factor1.htm
4. http://stats.stackexchange.com/questions/1576/what-are-the-differences-between-factor-analysis-and-principal-component-analysi

Question and Answers

- Please email us your unresolved questions or suggestions:
  - Jovita.Monteiro.3 @ gmail.com
  - Jackline.niks @ gmail.com

  - MoshAhmed @ gmail.com