Multivariate Analysis

PRINCIPAL COMPONENT ANALYSIS

By Suckwon Hong

1. Open the wiscsem.csv file. Use the following labels. These variables relate to intelligence for children. Use principal component analysis. Use varimax rotation.

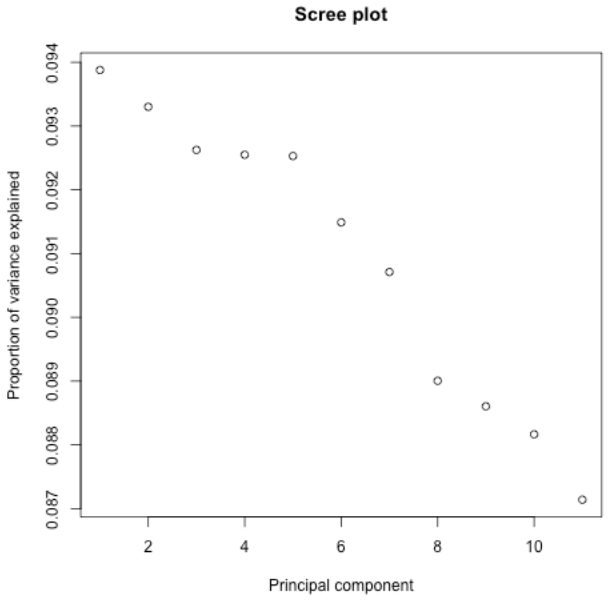
LABEL

info='Information'; comp='Comprehension'; arith='Arithmetic'; simil='Similarities'; vocab='Vocabulary' digit='Digit Span'; pictcomp='Picture Completion'; parang= 'Paragraph Arrangement'; block='Block Design' object='Object Assembly'; coding= 'Coding';

2. How many "real" or "useful" factors emerged from the analysis? In this case you should use the scree plot for this decision, rather than retaining all factors with eigenvalues greater than 1. Attach all factor analysis output here.

Answer:

From the scree plot, I decided to retain 8 factors among 11 factors.



```
Principal Components Analysis
Call: psych::principal(r = data, nfactors = 11, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
           RC7  RC5   RC2  RC4   RC6   RC3  RC8  RC10 RC11  RC9   RC1 h2       u2 com
info      0.23 0.16  0.07 0.07  0.07 -0.02 0.06  0.20 0.27 0.16  0.87  1 -1.8e-15 1.7
comp      0.16 0.09  0.14 0.05  0.18  0.03 0.14  0.20 0.20 0.89  0.16  1  0.0e+00 1.6
arith     0.94 0.11 -0.02 0.09  0.04  0.04 0.11  0.12 0.12 0.14  0.19  1 -4.4e-16 1.3
simil     0.14 0.11  0.11 0.13  0.15 -0.04 0.07  0.89 0.21 0.19  0.19  1 -1.3e-15 1.6
vocab     0.15 0.12  0.05 0.02  0.11  0.05 0.11  0.22 0.88 0.21  0.27  1 -4.4e-16 1.7
digit     0.10 0.97  0.00 0.06  0.02  0.09 0.01  0.09 0.10 0.07  0.12  1  2.2e-16 1.1
pictcomp  0.04 0.02  0.16 0.10  0.94 -0.05 0.16  0.13 0.09 0.15  0.06  1  3.3e-16 1.3
parang    0.09 0.06  0.10 0.97  0.09  0.01 0.14  0.11 0.02 0.04  0.06  1 -4.4e-16 1.1
block     0.11 0.01  0.18 0.16  0.16  0.06 0.93  0.07 0.10 0.13  0.06  1 -6.7e-16 1.3
object   -0.02 0.00  0.95 0.10  0.15  0.03 0.17  0.09 0.05 0.11  0.05  1 -6.7e-16 1.2
coding    0.04 0.08  0.02 0.01 -0.04  0.99 0.05 -0.03 0.04 0.02 -0.01  1 -6.7e-16 1.0

                       RC7  RC5  RC2  RC4  RC6  RC3  RC8 RC10 RC11  RC9  RC1
SS loadings           1.03 1.03 1.02 1.02 1.02 1.01 1.00 0.98 0.97 0.97 0.96
Proportion Var        0.09 0.09 0.09 0.09 0.09 0.09 0.09 0.09 0.09 0.09 0.09
Cumulative Var        0.09 0.19 0.28 0.37 0.46 0.56 0.65 0.74 0.82 0.91 1.00
Proportion Explained  0.09 0.09 0.09 0.09 0.09 0.09 0.09 0.09 0.09 0.09 0.09
Cumulative Proportion 0.09 0.19 0.28 0.37 0.46 0.56 0.65 0.74 0.82 0.91 1.00
```

3. Name the factors. Use your own judgment to name them.

Answer: For each component, I first omit variables that contribute a little (below 0.15). As a result, I get following table.

```
              RC7       RC5       RC2       RC4       RC6       RC3       RC8      RC10
info    0.2309631 0.1617039        NA        NA        NA        NA        NA 0.1970254
comp    0.1632906        NA        NA        NA 0.1768948        NA        NA 0.1955091
arith   0.9368349        NA        NA        NA        NA        NA        NA        NA
simil          NA        NA        NA        NA 0.1529387        NA        NA 0.8887096
vocab          NA        NA        NA        NA        NA        NA        NA 0.2197830
digit          NA 0.9703405        NA        NA        NA        NA        NA        NA
pictcomp       NA        NA 0.1578140        NA 0.9415354        NA 0.1560053        NA
parang         NA        NA        NA 0.9659248        NA        NA        NA        NA
block          NA        NA 0.1819430 0.1613427 0.1608888        NA 0.9311696        NA
object         NA        NA 0.9544194        NA 0.1514552        NA 0.1688858        NA
coding         NA        NA        NA        NA        NA 0.9922682        NA        NA
```

After, I named them as follows:

RC7	Basic knowledge about math and society
RC5	Short term memory
RC2	Space perception
RC4	Logical power
RC6	Ability to find pattern
RC3	Processing speed
RC8	Space perception (similar to RC5)
RC10	Language power

4. What is the highest loading variable? Which factor does it load onto?

Answer: The biggest loading is found in variable *coding*. It is load onto the component *RC3* that I named as processing speed.

5. What is the cumulative percentage of the total variance that is accounted for by your retained factors?

Answer: Eight factors that I retained explain about 74% of the total variance.

6. Use oblique rotation instead of varimax rotation. Do you notice any difference in the results?

Answer: When I use oblique rotation, the result says that it is a Heywood case.

```
Principal Components Analysis
Call: psych::principal(r = data, nfactors = 11, rotate = "oblimin")

 Warning: A Heywood case was detected.
Standardized loadings (pattern matrix) based upon correlation matrix
         TC5 TC2 TC4 TC6 TC7 TC3 TC10 TC1 TC9 TC8 TC11 h2       u2 com
info       0   0   0   0   0   0    0   1   0   0    0  1 -1.8e-15   1
comp       0   0   0   0   0   0    0   0   1   0    0  1  0.0e+00   1
arith      0   0   0   0   1   0    0   0   0   0    0  1 -4.4e-16   1
simil      0   0   0   0   0   0    1   0   0   0    0  1 -1.3e-15   1
vocab      0   0   0   0   0   0    0   0   0   0    1  1 -4.4e-16   1
digit      1   0   0   0   0   0    0   0   0   0    0  1  2.2e-16   1
pictcomp   0   0   0   1   0   0    0   0   0   0    0  1  3.3e-16   1
parang     0   0   1   0   0   0    0   0   0   0    0  1 -4.4e-16   1
block      0   0   0   0   0   0    0   0   0   1    0  1 -6.7e-16   1
object     0   1   0   0   0   0    0   0   0   0    0  1 -6.7e-16   1
coding     0   0   0   0   0   1    0   0   0   0    0  1 -6.7e-16   1
```

7. Calculate the coefficient alphas for each factor.

Answer:

	Cronbach's alpha
RC7	0.701
RC5	0.512
RC2	0.648
RC4	0.510
RC6	0.741
RC3	0.820
RC8	0.648
RC10	0.818

R code

```
library(lattice)
library(psych)
library(GPArotation)
library(psy)
library(dplyr)
##Multivaraite stats hw6 by suckwon hong
rm(list = ls())
getwd()
setwd("/Users/suckwonhong/Desktop/multivariate_stats")

data <- read.csv("wiscsem.csv",header = TRUE)
data$X...client<-NULL
data$agemate<-NULL

#pca varimax
pca.vari<-psych::principal(data, nfactors = 11, rotate = "varimax")
pca.vari.res<-as.data.frame(pca.vari$Vaccounted)
png(filename = "scree plot.png")
plot(c(1:11),pca.vari.res[2,], main = "Scree plot", ylab = "Proportion of variance explained",
    xlab = "Principal component")
dev.off()

#loadings
pca.vari.load<-(unclass(pca.vari$loadings))
pca.vari.load[pca.vari.load<0.15]<-NA

#pca oblique
pca.oblique<-psych::principal(data, nfactors = 11, rotate = "oblimin")
pca.oblique.res<-as.data.frame(pca.oblique$Vaccounted)
```

```
png(filename = "scree plot_obli.png")
plot(c(1:11),pca.oblique.res[2,], main = "Scree plot", ylab = "Proportion of variance explained",
    xlab = "Principal component")
dev.off()

#loadings
pca.oblique.load<-(unclass(pca.oblique$loadings))
pca.oblique.load[pca.oblique.load<0.15]<-NA

#cronbach alpha
pc1<-cronbach(select(data, info,comp,arith))
pc2<-cronbach(select(data, info,digit))
pc3<-cronbach(select(data, pictcomp,block,object))
pc8<-cronbach(select(data, info,parang,block))
pc4<-cronbach(select(data, comp,simil,pictcomp,block,object))
pc5<-cronbach(select(data, coding))
pc6<-cronbach(select(data, pictcomp,block,object))
pc7<-cronbach(select(data, info,comp,simil,vocab))
```