BACS - HW 14 106073401

Question 1) Earlier, we examined the eigenvectors of the security dataset. Now, let's examine *factor loadings* (use the `principal()` method from the `psych` package)

```
sq = read.csv("security_questions.csv", header = TRUE, na.strings = "?")
sq = data.matrix(sq[complete.cases(sq),])
install.packages("psych")
library(psych)
```

a. Looking at the loadings of the first 3 principal components, to which components does each item seem to belong?

All items belong to PC1, Q4,Q12,Q17 belong to PC2, Q5 & Q10 seem to belong to PC3.

```
sec_pca_3<- principal(sq, nfactor=3, rotate="none", scores=TRUE)
Principal Components Analysis
Call: principal(r = sq, nfactors = 3, rotate = "none", scores = TRUE)
Standardized loadings (pattern matrix) based upon correlation matrix
     PC1   PC2   PC3   h2   u2 com
Q1  0.82 -0.14  0.00 0.69 0.31 1.1
Q2  0.67 -0.01  0.09 0.46 0.54 1.0
Q3  0.77 -0.03  0.09 0.60 0.40 1.0
Q4  0.62  0.64  0.11 0.81 0.19 2.1
Q5  0.69 -0.03 -0.54 0.77 0.23 1.9
Q6  0.68 -0.10  0.21 0.52 0.48 1.2
Q7  0.66 -0.32  0.32 0.64 0.36 2.0
Q8  0.79  0.04 -0.34 0.74 0.26 1.4
Q9  0.72 -0.23  0.20 0.62 0.38 1.4
Q10 0.69 -0.10 -0.53 0.76 0.24 1.9
Q11 0.75 -0.26  0.17 0.66 0.34 1.4
Q12 0.63  0.64  0.12 0.82 0.18 2.1
Q13 0.71 -0.06  0.08 0.52 0.48 1.0
Q14 0.81 -0.10  0.16 0.69 0.31 1.1
Q15 0.70  0.01 -0.33 0.61 0.39 1.4
Q16 0.76 -0.20  0.18 0.65 0.35 1.3
Q17 0.62  0.66  0.11 0.83 0.17 2.0
Q18 0.81 -0.11 -0.07 0.67 0.33 1.1

                       PC1  PC2  PC3
SS loadings           9.31 1.60 1.15
Proportion Var        0.52 0.09 0.06
Cumulative Var        0.52 0.61 0.67
Proportion Explained  0.77 0.13 0.10
Cumulative Proportion 0.77 0.90 1.00
```

b. How much of the total variance of the security dataset does the first 3 PCs capture?

0.67

c. Looking at commonality and uniqueness, which item or items' variance is less than adequately explained by the first 3 principal components?

Q2

d. How many measurement items share similar loadings between 2 or more components?

Q4, Q5, Q10, Q12, Q17.

e. Can you distinguish a 'meaning' behind the first principal component from the items that load best upon it? (see the wording of the questions of those items)

Based on the questions, it seems that PC1 is related to "Trust"(the 3 dimensions: benevolence, ability and integrity).

Question 2) To improve interpretability of loadings, let's rotate the our principal component axes to get *rotated components* (extract and rotate only three principal components)

```
principal(sq, nfactors=3, rotate="varimax", scores=TRUE)

Standardized loadings (pattern matrix) based upon correlation matrix
     PC1  PC3  PC2   h2   u2 com
Q1  0.66 0.45 0.22 0.69 0.31 2.0
Q2  0.54 0.29 0.29 0.46 0.54 2.1
Q3  0.62 0.34 0.31 0.60 0.40 2.1
Q4  0.22 0.19 0.85 0.81 0.19 1.2
Q5  0.24 0.83 0.16 0.77 0.23 1.3
Q6  0.65 0.20 0.23 0.52 0.48 1.5
Q7  0.79 0.10 0.06 0.64 0.36 1.0
Q8  0.38 0.71 0.30 0.74 0.26 2.0
Q9  0.74 0.23 0.14 0.62 0.38 1.3
Q10 0.28 0.82 0.10 0.76 0.24 1.3
Q11 0.76 0.28 0.12 0.66 0.34 1.3
Q12 0.23 0.19 0.85 0.82 0.18 1.2
Q13 0.59 0.32 0.26 0.52 0.48 1.9
Q14 0.72 0.31 0.28 0.69 0.31 1.7
Q15 0.34 0.66 0.24 0.61 0.39 1.8
Q16 0.74 0.27 0.17 0.65 0.35 1.4
Q17 0.21 0.19 0.87 0.83 0.17 1.2
Q18 0.61 0.50 0.23 0.67 0.33 2.2

                       PC1  PC3  PC2
SS loadings           5.61 3.49 2.95
Proportion Var        0.31 0.19 0.16
Cumulative Var        0.31 0.51 0.67
Proportion Explained  0.47 0.29 0.24
Cumulative Proportion 0.47 0.76 1.00

#Variance before rotated
Proportion Var        0.52 0.09 0.06
Cumulative Var        0.52 0.61 0.67

#Loading before rotated
Q4  0.62  0.64  0.11 0.81 0.19 2.1
Q5  0.69 -0.03 -0.54 0.77 0.23 1.9
Q10 0.69 -0.10 -0.53 0.76 0.24 1.9
Q12 0.63  0.64  0.12 0.82 0.18 2.1
Q17 0.62  0.66  0.11 0.83 0.17 2.0
```

a. Individually, does each rotated component explain the same, or different, amount of variance than the three principal components?
   PC1:0.31→0.52, PC2:0.19→0.09, PC3:0.16 → 0.06.

b. Together, do the three rotated components explain the same, more, or less cumulative variance as the three principal components combined?
   The cumulative variances of three PC combined are both 0.67.

c. Looking back at the items that shared similar loadings with multiple principal components, do those items have more clearly differentiated loadings among rotated components?
   The rotated components (Q4, Q5, Q10, Q12, Q17) have more clearly differentiated loadings than before.

d. Can you now interpret the “meaning” of the 3 rotated components from the items that load best upon each of them? (see the wording of the questions of those items)

The questions are categorized based on the rotated components loadings. From the questions, we can observe that

PC1 seems to related with trust (benevolence, ability and integrity)

PC2 seems to related with actual security practices.

PC3 seems to related with interaction with real site.

1	Q1	I am convinced that this site respects the confidentiality of the transactions received from me
2	Q2	All communications with this site are restricted to the site and me
3	Q3	This site checks the information communicated with me for accuracy
4	Q4	This site provides me with some evidence to protect against its denial of having received a transaction from me
5	Q5	The transactions I send are transmitted to the real site to which I want to transmit
6	Q6	This site checks all communications between the site and me for protection from wiretapping or eavesdropping
7	Q7	This site never sells my personal information in their computer databases to other companies
8	Q8	This site ascertains my identity before processing the transactions received from me
9	Q9	I can remove my personal information from this site when I want to
10	Q10	The messages I receive are transmitted from the real site from which I want to receive them
11	Q11	This site devotes time and effort to preventing unauthorized access to my personal information
12	Q12	This site takes steps to make sure that the information in transit is not deleted
13	Q13	This site provides me with some evidence to protect against its denial of having sent a message
14	Q14	This site devotes time and effort to verify the accuracy of the information in transit
15	Q15	This site ascertains my identity before sending any messages to me
16	Q16	Databases that contain my personal information are protected from unauthorized access
17	Q17	This site provides me with some evidence to protect against its denial of having participated in a transaction after processing it
18	Q18	This site uses some security controls for the confidentiality of the transactions received from me

e. If we reduced the number of extracted and rotated components to 2, does the meaning of our rotated components change?

It will change, as in the result, the loadings and variance are different, and each principal components captures different questions comparing with the result of 3 components.

```
sec_pca_2 <- principal(sq, nfactor=2, rotate="varimax", scores=TRUE)
Standardized loadings (pattern matrix) based upon correlation matrix
     PC1  PC2   h2   u2 com
Q1  0.78 0.27 0.69 0.31 1.2
Q2  0.60 0.31 0.45 0.55 1.5
Q3  0.69 0.34 0.59 0.41 1.5
Q4  0.24 0.86 0.80 0.20 1.1
Q5  0.62 0.31 0.48 0.52 1.5
Q6  0.65 0.24 0.48 0.52 1.3
Q7  0.73 0.04 0.53 0.47 1.0
Q8  0.67 0.42 0.62 0.38 1.7
Q9  0.75 0.15 0.58 0.42 1.1
Q10 0.65 0.24 0.48 0.52 1.3
Q11 0.79 0.13 0.64 0.36 1.1
Q12 0.25 0.86 0.80 0.20 1.2
Q13 0.65 0.29 0.51 0.49 1.4
Q14 0.76 0.30 0.67 0.33 1.3
Q15 0.61 0.35 0.50 0.50 1.6
Q16 0.76 0.19 0.62 0.38 1.1
Q17 0.22 0.88 0.82 0.18 1.1
Q18 0.76 0.29 0.66 0.34 1.3

                       PC1  PC2
SS loadings           7.52 3.39
Proportion Var        0.42 0.19
Cumulative Var        0.42 0.61
Proportion Explained  0.69 0.31
Cumulative Proportion 0.69 1.00
```