BACS_HW_Week15_106071041

106071041

2021/6/6

```
dec <- read_excel("security_questions.xlsx", sheet = "data")
```

#Question 1 | # a. Show a single visualization with **scree plot of data**, **scree plot of simulated noise**, and **a horizontal line showing the eigenvalue = 1 cutoff**

```
dec_pca <- prcomp(dec, scale. = TRUE)
```

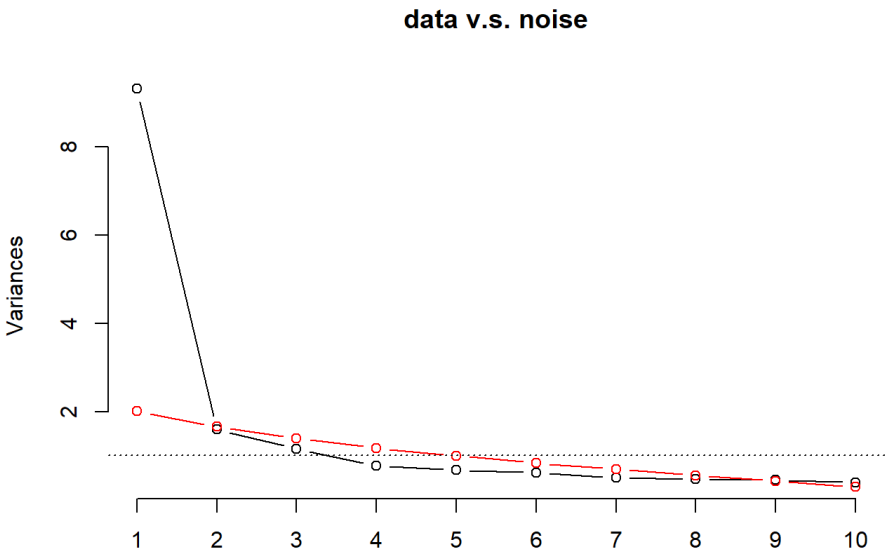
```
noise <- data.frame(replicate(10, rnorm(33)))
```

```
sim_noise_ev <- function(n, p) {
  noise <- data.frame(replicate(p, rnorm(n)))
  return( eigen(cor(noise))$values )
}
```

```
set.seed(24342)
evalues_noise <- replicate(100, sim_noise_ev(33, 10))
```

```
evalus_mean <- apply(evalues_noise, 1, mean)
```

```
# scree plot of data
screeplot(dec_pca, type = "lines", main = "data v.s. noise")
# scree plot of simulated noise
lines(evalus_mean, type = "b", col = "red")
# a horizontal line showing the eigenvalue = 1 cutoff
abline(h=1 , lty="dotted")
```



b. How many dimensions would you retain if we used Parallel Analysis? ONLY ONE: PC1

```
evalus_mean[2]
```

```
## [1] 1.651672
```

```
dec_pca$sdev[2]
```

```
## [1] 1.26346
```

Question 2

a. To which components does each item seem to best belong?

Criteria: >0.7
PC1: “Q1” “Q3” “Q8” “Q9” “Q11” “Q13” “Q14” “Q15” “Q16” “Q18”
PC2: none
PC3: none

```
dec_principal <- principal(dec, nfactors = 3, rotate = "none", scores = TRUE)
```

```
# PC1
names(which(dec_principal$Structure[,1] >= 0.7))
```

```
##  [1] "Q1"  "Q3"  "Q8"  "Q9"  "Q11" "Q13" "Q14" "Q15" "Q16" "Q18"
```

```
names(which(dec_principal$Structure[,1] <= -0.7))
```

```
## character(0)
```

```
# PC2
names(which(dec_principal$Structure[,2] >= 0.7))
```

```
## character(0)
```

```
names(which(dec_principal$Structure[,2] <= -0.7))
```

```
## character(0)
```

```
# PC3
names(which(dec_principal$Structure[,3] >= 0.7))
```

```
## character(0)
```

```
names(which(dec_principal$Structure[,3] <= -0.7))
```

```
## character(0)
```

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

About 67%

```
dec_principal$Vaccounted
```

```
##                             PC1        PC2        PC3
## SS loadings           9.3109533 1.59633195 1.14955822
## Proportion Var        0.5172752 0.08868511 0.06386435
## Cumulative Var        0.5172752 0.60596029 0.66982464
## Proportion Explained  0.7722546 0.13240049 0.09534487
## Cumulative Proportion 0.7722546 0.90465513 1.00000000
```

```
dec_principal$Vaccounted[3,3] %>% round(2)
```

```
## [1] 0.67
```

c. Looking at commonality and uniqueness, which items are less than adequately explained by the first 3 principal components?

“Q1” “Q2” “Q3” “Q6” “Q7” “Q9” “Q11” “Q13” “Q14” “Q15” “Q16” “Q18”

```
names(which(dec_principal$communality < 0.7))
```

```
##  [1] "Q1"  "Q2"  "Q3"  "Q6"  "Q7"  "Q9"  "Q11" "Q13" "Q14" "Q15" "Q16" "Q18"
```

```
names(which(dec_principal$uniquenesses > 0.3))
```

```
##  [1] "Q1"  "Q2"  "Q3"  "Q6"  "Q7"  "Q9"  "Q11" "Q13" "Q14" "Q15" "Q16" "Q18"
```

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

THREE (“Q4” “Q12” “Q17”)

```
dec_principal$loadings
```

```
##
## Loadings:
##     PC1    PC2    PC3
## Q1   0.817 -0.139
## Q2   0.673
## Q3   0.766
## Q4   0.623  0.643  0.108
## Q5   0.690        -0.542
## Q6   0.683 -0.105  0.207
## Q7   0.657 -0.318  0.324
## Q8   0.786        -0.343
## Q9   0.723 -0.232  0.204
## Q10  0.686        -0.533
## Q11  0.753 -0.261  0.173
## Q12  0.630  0.638  0.122
## Q13  0.712
## Q14  0.811         0.157
## Q15  0.704        -0.333
## Q16  0.758 -0.203  0.183
## Q17  0.618  0.664  0.110
## Q18  0.807 -0.114
##
##                  PC1   PC2   PC3
## SS loadings    9.311 1.596 1.150
## Proportion Var 0.517 0.089 0.064
## Cumulative Var 0.517 0.606 0.670
```

e. Can you distinguish a ‘meaning’ behind the first principal component from the items that load best upon it?

PC1: “Q1” “Q3” “Q8” “Q9” “Q11” “Q13” “Q14” “Q15” “Q16” “Q18”
Most of them are related the security of personal information.

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

A caption

Question 3 |

a. Individually, does each rotated component (RC) explain the same, or different, amount of variance than the corresponding principal components (PCs)?

RC1, RC2 and RC3 are different from PC1, PC2 and PC3 respectively.

```
dec_pca_rot <- principal(dec, nfactors = 3, rotate = "varimax", scores = TRUE)
```

```
dec_pca_rot$Vaccounted
```

```
##                            RC1       RC3       RC2
## SS loadings           5.6131484 3.4901395 2.9535556
## Proportion Var        0.3118416 0.1938966 0.1640864
## Cumulative Var        0.3118416 0.5057382 0.6698246
## Proportion Explained  0.4655570 0.2894737 0.2449692
## Cumulative Proportion 0.4655570 0.7550308 1.0000000
```

```
dec_principal$Vaccounted
```

```
##                            PC1        PC2        PC3
## SS loadings           9.3109533 1.59633195 1.14955822
## Proportion Var        0.5172752 0.08868511 0.06386435
## Cumulative Var        0.5172752 0.60596029 0.66982464
## Proportion Explained  0.7722546 0.13240049 0.09534487
## Cumulative Proportion 0.7722546 0.90465513 1.00000000
```

```
dec_pca_rot$Vaccounted[2,] %>% round(2) == dec_principal$Vaccounted[2,] %>% round(2)
```

```
##   RC1   RC3   RC2
## FALSE FALSE FALSE
```

b. Together, do the three rotated components explain the same, more, or less cumulative variance as the three principal components combined?

The same when rounded to 2 decimals.

```
round(dec_pca_rot$Vaccounted[3,3],2) == round(dec_principal$Vaccounted[3,3],2)
```

```
## [1] TRUE
```

c. Looking back at the items that shared similar loadings with multiple principal components (#2d), do those items have more clearly differentiated loadings among rotated components?

(#2d) “Q4” “Q12” “Q17”
Apparently YES, please refer to the comparison dataframe below.

```
# "Q4"
data.frame(dec_pca_rot$loadings[4,],dec_principal$loadings[4,])
```

```
##     dec_pca_rot.loadings.4... dec_principal.loadings.4...
## RC1                 0.2182880                   0.6233733
## RC3                 0.1933627                   0.6430783
## RC2                 0.8536838                   0.1080319
```

```
# "Q12"
data.frame(dec_pca_rot$loadings[12,],dec_principal$loadings[12,])
```

```
##     dec_pca_rot.loadings.12... dec_principal.loadings.12...
## RC1                  0.2327616                    0.6303505
## RC3                  0.1861745                    0.6375312
## RC2                  0.8542346                    0.1215228
```

```
# "Q17"
data.frame(dec_pca_rot$loadings[17,],dec_principal$loadings[17,])
```

```
##     dec_pca_rot.loadings.17... dec_principal.loadings.17...
## RC1                  0.2054021                   0.6175336
## RC3                  0.1869028                   0.6642605
## RC2                  0.8703910                   0.1100612
```

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)

```
options(knitr.duplicate.label = "allow")
```

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

RC1: “Q7” “Q9” “Q11” “Q14” “Q16”

RC2: “Q5” “Q8” “Q10”

RC3: “Q4” “Q12” “Q17”

Interpretation: RC1: The supreme rights of manipulaion of personal information
RC2: The authenticity of the site
RC3: The functionality of the site

```
names(which(dec_pca_rot$loadings[,1] >= 0.7))
```

```
## [1] "Q7"  "Q9"  "Q11" "Q14" "Q16"
```

```
names(which(dec_pca_rot$loadings[,2] >= 0.7))
```

```
## [1] "Q5"  "Q8"  "Q10"
```

```
names(which(dec_pca_rot$loadings[,3] >= 0.7))
```

```
## [1] "Q4"  "Q12" "Q17"
```

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

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

A caption

YES, it changed!!

After components reduction
RC1: “Q1” “Q7” “Q9” “Q11” “Q14” “Q16” “Q18”
RC2: “Q5” “Q8” “Q10” (the same as RC2 with 3 components)

Interpretation
RC1: The functionality of the site including protecting action upon personal information and the transaction.
RC2: RC2: The authenticity of the site

```
dec_pca_rot_2 <- principal(dec, nfactors = 2, rotate = "varimax", scores = TRUE)
```

```
names(which(dec_pca_rot_2$loadings[,1] >= 0.7))
```

```
## [1] "Q1"  "Q7"  "Q9"  "Q11" "Q14" "Q16" "Q18"
```

```
names(which(dec_pca_rot$loadings[,2] >= 0.7))
```

```
## [1] "Q5"  "Q8"  "Q10"
```

(ungraded)

How many components (1-3) do you believe we should extract and analyze to understand the security dataset? Feel free to suggest different answers for different purposes.

I think 3 is better than 2 since the interpretation when it’s only two components is quite hard while when it’s three I can feel some distinct difference between the three components.