HW14

109006206

Set Working Directories & Reading Files

```
library(readxl)
library(ggplot2)
library(magrittr)
library(psych)
setwd("/Users/olivia/Documents/Documents/Study/Semester 6/BACS/HW14")
security <- read_excel("security_questions.xlsx", sheet = "data")
```

QUESTION 1

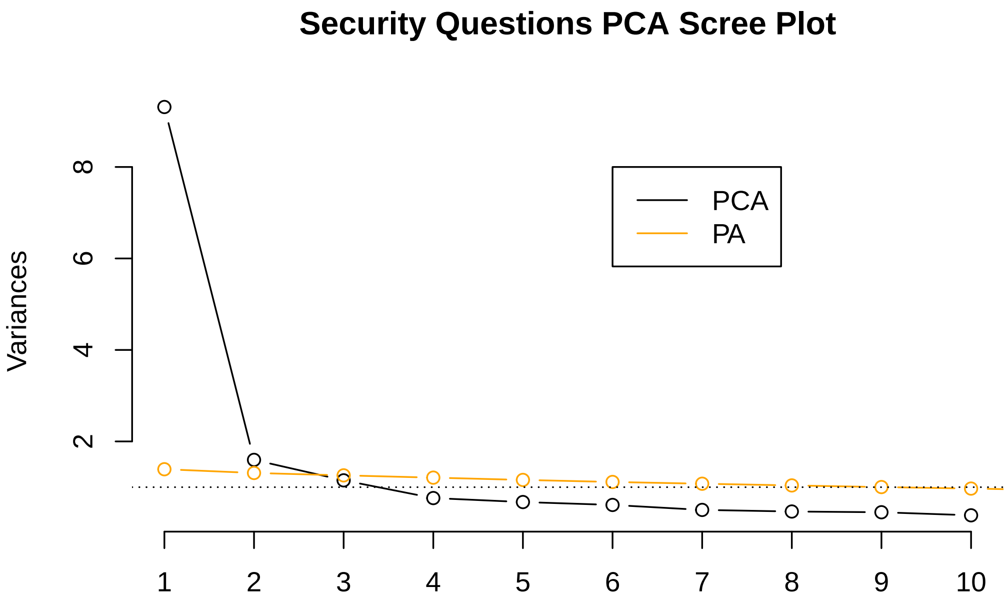
A) Show a single visualization with scree plot of data, scree plot of simulated noise (use average eigenvalues of >= 100 noise samples), and a horizontal line showing the eigenvalue = 1 cutoff.

```
pca1 <- prcomp(security, scale. = TRUE)

sim_noise_ev <- function(n, p) {
  noise <- data.frame(replicate(p, rnorm(n)))
  eigen(cor(noise))$values
}
evalues_noise <- replicate(100, sim_noise_ev(dim(security)[1], dim(security)[2]))
evalues_mean <- apply(evalues_noise, 1, mean)

screeplot(pca1, type="lines",main = "Security Questions PCA Scree Plot")
lines(evalues_mean, type="b",col="orange")
abline(h=1, lty="dotted")

legend(6,8,c("PCA","PA"),lty = c(1,1),col=c("black","orange"))
```



B) How many dimensions would you retain if we used Parallel Analysis?

Answer : I would retain 3 dimensions since it is the point where it starts to cross the data.

QUESTION 2

A) Looking at the loadings of the first 3 principal components, to which components does each item seem to best belong?

```r
pca <- principal(security, nfactors = 3, rotate = "none")
pca
```

```
## Principal Components Analysis
## Call: principal(r = security, nfactors = 3, rotate = "none")
## 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
##
## Mean item complexity =  1.5
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.05
##  with the empirical chi square  258.65  with prob <  1.4e-15
##
```

```
## Fit based upon off diagonal values = 0.99
```

B) How much of the total variance of the security dataset do the first 3 PCs capture?

```
variance_prop <- pca$values[1:3]
variance_prop
```

```
## [1] 9.310953 1.596332 1.149558
```

```
sum_variance <- sum(variance_prop)
sum_variance
```

```
## [1] 12.05684
```

Answer : The total variance from PC1,PC2,PC3 is 12.06

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

```
threshold <- 0.5
less_explained_items <- rownames(pca$loadings)[pca$communality < threshold]
less_explained_items
```

```
## [1] "Q2"
```

```
pca$communality[2]
```

```
##        Q2
## 0.4605433
```

Answer : It would be "Q2" if I set it < 0.5

D) How many measurement items share similar loadings between 2 or more components?

```
shared_loadings_items <- c()
for (i in 1:(ncol(pca$loadings) - 1)) {
  for (j in (i + 1):ncol(pca$loadings)) {
    shared_items <- rownames(pca$loadings)[pca$loadings[, i] >=
                                              0.5 & pca$loadings[, j] >= 0.5]
    shared_loadings_items <- union(shared_loadings_items, shared_items)
  }
}

cat("Measurement items with similar loadings between two or more components:",shared_loadings_items)
```

```
## Measurement items with similar loadings between two or more components: Q4 Q12 Q17
```

E) Can you interpret a 'meaning' behind the first principal component from the items that load best upon it?

Answer: The item that load best upon the first principal component is Q1,Q14,Q18. They suggest that the first principal component reflects customers' perceptions of the site's commitment to safeguarding the confidentiality and accuracy of their transactions.

QUESTION 3

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

```
pca_rc <-principal(security, nfactors=3, rotate="varimax", scores=TRUE)
pca_rc
```

```
## Principal Components Analysis
## Call: principal(r = security, nfactors = 3, rotate = "varimax", scores = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##      RC1  RC3  RC2   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
##
##                        RC1  RC3  RC2
## 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
##
## Mean item complexity =  1.6
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.05
##  with the empirical chi square  258.65  with prob <  1.4e-15
##
```

```
## Fit based upon off diagonal values = 0.99
```

```r
rotated_components <- pca_rc$values[1:3]
rotated_components
```

```
## [1] 9.310953 1.596332 1.149558
```

```r
variance_prop
```

```
## [1] 9.310953 1.596332 1.149558
```

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

```r
total_rc <- sum(pca_rc$values[1:3])
total_not <- sum(pca$values[1:3])

if (total_rc==total_not) {
  print("Yes, the total the three rotated components explain")
  print("the same cumulative variance as the three principal components combined")
} else {
  print("No, the total the three rotated components does not explain")
  print("the same cumulative variance as the three principal components combined")
}
```

```
## [1] "Yes, the total the three rotated components explain"
## [1] "the same cumulative variance as the three principal components combined"
```

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?

```r
items_of_interest <- c("Q4", "Q12", "Q17")
principal_loadings_interest <- pca$loadings[items_of_interest, ]
rotated_loadings_interest <- pca_rc$loadings[items_of_interest, ]
loadings_diff <- abs(principal_loadings_interest - rotated_loadings_interest)
loadings_comparison <- cbind(principal_loadings_interest, rotated_loadings_interest,loadings_diff)
colnames(loadings_comparison)[(ncol(loadings_comparison)-2):
                                ncol(loadings_comparison)] <- c("Loadings Difference1",
print(loadings_comparison)
```

```
##           PC1       PC2       PC3       RC1       RC3       RC2
## Q4  0.6233733 0.6430783 0.1080319 0.2182880 0.1933627 0.8536838
## Q12 0.6303505 0.6375312 0.1215228 0.2327616 0.1861745 0.8542346
## Q17 0.6175336 0.6642605 0.1100612 0.2054021 0.1869028 0.8703910
##     Loadings Difference1 Loadings Difference2 Loadings Difference3
## Q4             0.4050853            0.4497155            0.7456519
```

```
## Q12              0.3975890          0.4513567          0.7327118
## Q17              0.4121315          0.4773577          0.7603299
```

Answer : The items have more clearly differentiated loadings

D) Can you now more easily 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)

Answer : RC1 represents a component related to the security and confidentiality aspects of the website or service.

```
pca_rc$loadings[pca_rc$loadings[,1]>0.5,1]
```

```
##        Q1        Q2        Q3        Q6        Q7        Q9       Q11       Q13
## 0.6602758 0.5437243 0.6206018 0.6524225 0.7895344 0.7378148 0.7573493 0.5931915
##       Q14       Q16       Q18
## 0.7187578 0.7396241 0.6090325
```

Answer : RC2 appears to reflect a component related to user-friendliness, control, and management of privacy settings.

```
pca_rc$loadings[pca_rc$loadings[,2]>0.5,1]
```

```
##        Q5        Q8       Q10       Q15
## 0.2441735 0.3819373 0.2768895 0.3417567
```

Answer : RC3 seems to represent a component related to the accuracy and organization of information provided by the website.

```
pca_rc$loadings[pca_rc$loadings[,3]>0.5,1]
```

```
##        Q4       Q12       Q17
## 0.2182880 0.2327616 0.2054021
```

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

```
reduced <- principal(security, nfactors=2, rotate="varimax", scores=TRUE)
RC1<- reduced$loadings[reduced$loadings[,1]>0.5,1]
RC1
```

```
##        Q1        Q2        Q3        Q5        Q6        Q7        Q8        Q9
## 0.7830951 0.5960420 0.6865878 0.6197912 0.6487494 0.7284256 0.6684679 0.7451939
##       Q10       Q11       Q13       Q14       Q15       Q16       Q18
## 0.6488232 0.7855784 0.6549937 0.7591295 0.6118654 0.7615661 0.7616746
```

```
RC2<- reduced$loadings[reduced$loadings[,2]>0.5,1]
RC2
```

```
##        Q4       Q12       Q17
## 0.2364722 0.2452587 0.2211505
```

Answer: If we reduce the factors by 1 the RC3 will combine into RC1 and RC2 and create a new RC1 and RC2. It wont change the meaning of our rotated components.