HW14

107070008

Question 1) Earlier, we examined a dataset from a security survey sent to customers of e-commerce websites. However, we only used the eigenvalue > 1 criteria and the screeplot to find a suitable number of components. Let's perform a parallel analysis as well this week:

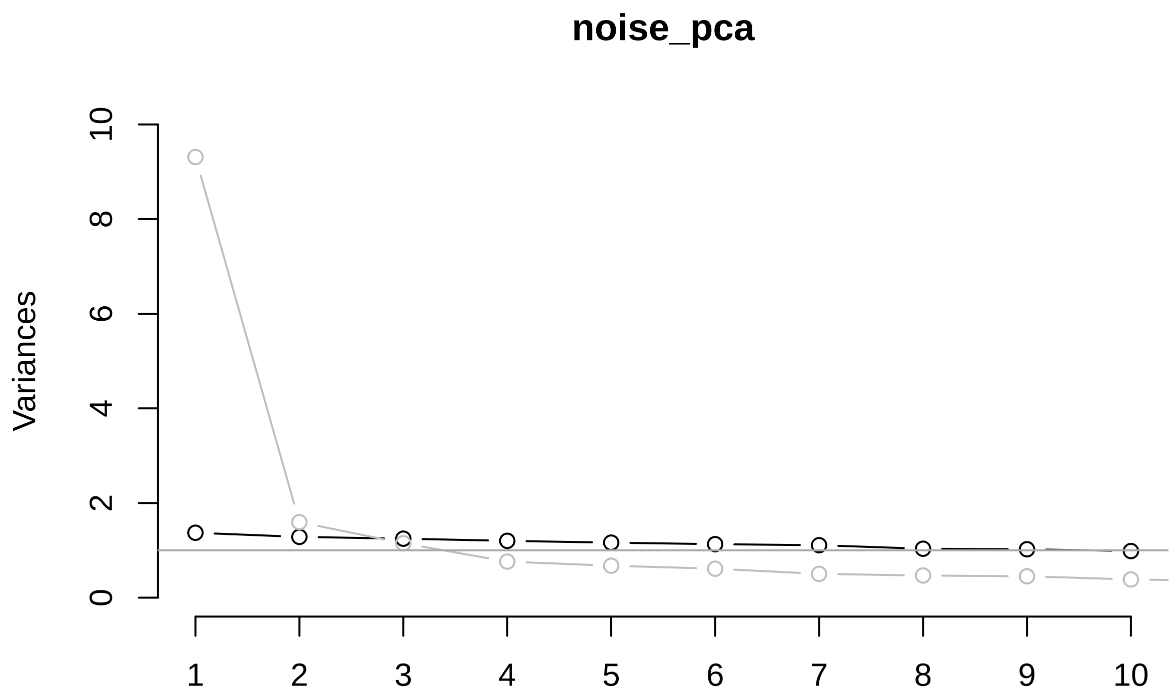
```
library("readxl")
sa <- read_excel("security_questions.xlsx")
```

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.

```
sa_eigen <- eigen(cor(sa))
sa_eigen$values/sum(sa_eigen$values)
```

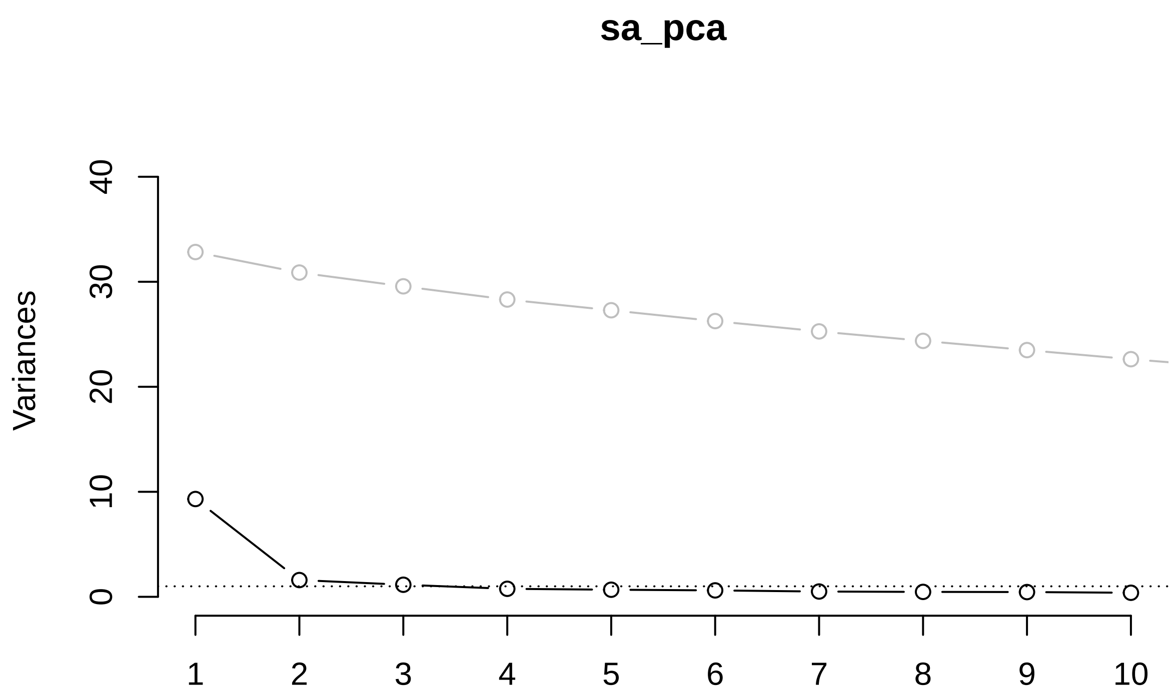
```
##  [1] 0.51727518 0.08868511 0.06386435 0.04233199 0.03750784 0.03398131
##  [7] 0.02794364 0.02601549 0.02510951 0.02139980 0.01971565 0.01673928
## [13] 0.01623763 0.01456354 0.01303216 0.01280357 0.01159706 0.01119690
```

```
sa_pca <- prcomp(sa, scale. = TRUE)

library(magrittr)
noise <- data.frame(replicate(ncol(sa), rnorm(nrow(sa))))
noise_pca <- prcomp(noise, scale. = TRUE)
var_sa <- (sa_pca$sdev)^2

screeplot(noise_pca, type="lines", ylim = c(0,10))
lines(var_sa, type="b", col = "gray")
abline(h=1, col = "darkgray")
```



b. How many dimensions would you retain if we used Parallel Analysis?

```
sa_pca <- prcomp(sa, 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(ncol(sa), nrow(sa)))
evalues_mean <- apply(evalues_noise, 1, mean)
screeplot(sa_pca, type="lines", ylim = c(0,45))
lines(evalues_mean, type="b", col = "gray")
abline(h=1,lty ="dotted")
```




```
Nome of major dimensions seems to exist in these correlates.
```

Question 2) Earlier, we treated the underlying dimensions of the security dataset as composites and examined their eigenvectors (weights). Now, let's treat them as factors and examine factor loadings (use the principal() method from the psych package)

```
library(psych)
sa_principal <- principal(sa, nfactor=3, rotate="none", scores=TRUE)
```

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

```
sa_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
```

PC1 : Q1, Q3, Q8, Q9, Q11, Q13, Q14, Q15, Q16, Q18 /PC2 : none of them /PC3 : none of them

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

```
sa_eigen$values[1:3] /sum(sa_eigen$values)
```

```
## [1] 0.51727518 0.08868511 0.06386435
```

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

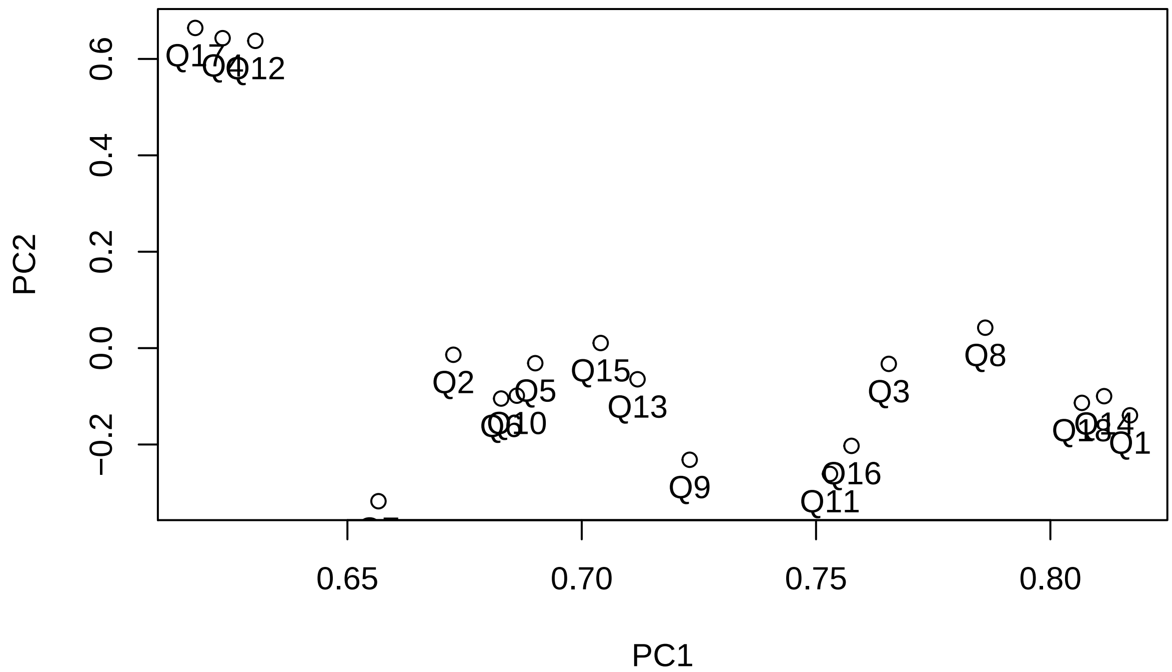
```
sa_principal$uniquenesses
```

```
##        Q1        Q2        Q3        Q4        Q5        Q6        Q7        Q8
## 0.3130959 0.5394567 0.4048641 0.1861853 0.2286580 0.4798896 0.3628631 0.2624488
##        Q9       Q10       Q11       Q12       Q13       Q14       Q15       Q16
## 0.3821333 0.2357097 0.3351446 0.1814443 0.4818957 0.3069979 0.3936244 0.3514148
##       Q17       Q18
## 0.1652968 0.3320337
```

In Q2, its u2 is higer than 50%, so it cannot fully interpreted by these three PCAS.

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

```
plot(sa_principal$loadings)
text(sa_principal$loadings, pos=1, labels = rownames(sa_principal$loadings))
```




```
There are five elements. Q1 & Q14 & Q18, Q4 & Q12 & Q17, Q6 & Q10
```

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

```
The higer component represent that this site respects the confidentiality of the transactions
received from the users, vice versa.
```

Question 3) To improve interpretability of loadings, let's rotate the our principal component axes using the varimax technique to get rotated components (extract and rotate only three principal components)

```
sa_pca_rot <- principal(sa, nfactor=3, rotate="varimax", scores=TRUE)
sa_pca_rot
```

```
## Principal Components Analysis
```

```
## Call: principal(r = sa, 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
```

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

```
sa_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
```

```
sa_pca_rot$loadings
```

```
##
## Loadings:
##     RC1   RC3   RC2
## Q1  0.660 0.450 0.221
## Q2  0.544 0.286 0.288
## Q3  0.621 0.337 0.311
## Q4  0.218 0.193 0.854
## Q5  0.244 0.828 0.162
## Q6  0.652 0.199 0.234
## Q7  0.790 0.103
## Q8  0.382 0.706 0.305
## Q9  0.738 0.234 0.138
## Q10 0.277 0.823 0.102
## Q11 0.757 0.278 0.118
## Q12 0.233 0.186 0.854
## Q13 0.593 0.315 0.259
## Q14 0.719 0.310 0.283
## Q15 0.342 0.656 0.244
## Q16 0.740 0.267 0.174
## Q17 0.205 0.187 0.870
## Q18 0.609 0.495 0.227
##
##                  RC1   RC3   RC2
## SS loadings    5.613 3.490 2.954
## Proportion Var 0.312 0.194 0.164
## Cumulative Var 0.312 0.506 0.670
```

Yes, they are different.

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

The first value is smaller than PCA, and the second and the third values are bigger than PCA.

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?

```
RC2 RC3 are similar to multiple principal components.
```

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)

```
Yes, after reotating our PCA, we can interpret those datas easier.
```

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

```
sa_pca_rot <- principal(sa, nfactor=3, rotate="varimax", scores=TRUE)
sa_pca_rot
```

```
## Principal Components Analysis
## Call: principal(r = sa, 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
```

No, because RC1 and RC2 didn't change a lots.