BACS HW (Week 14)

108020024

due on 05/24 (Sun) Helped by 108020033

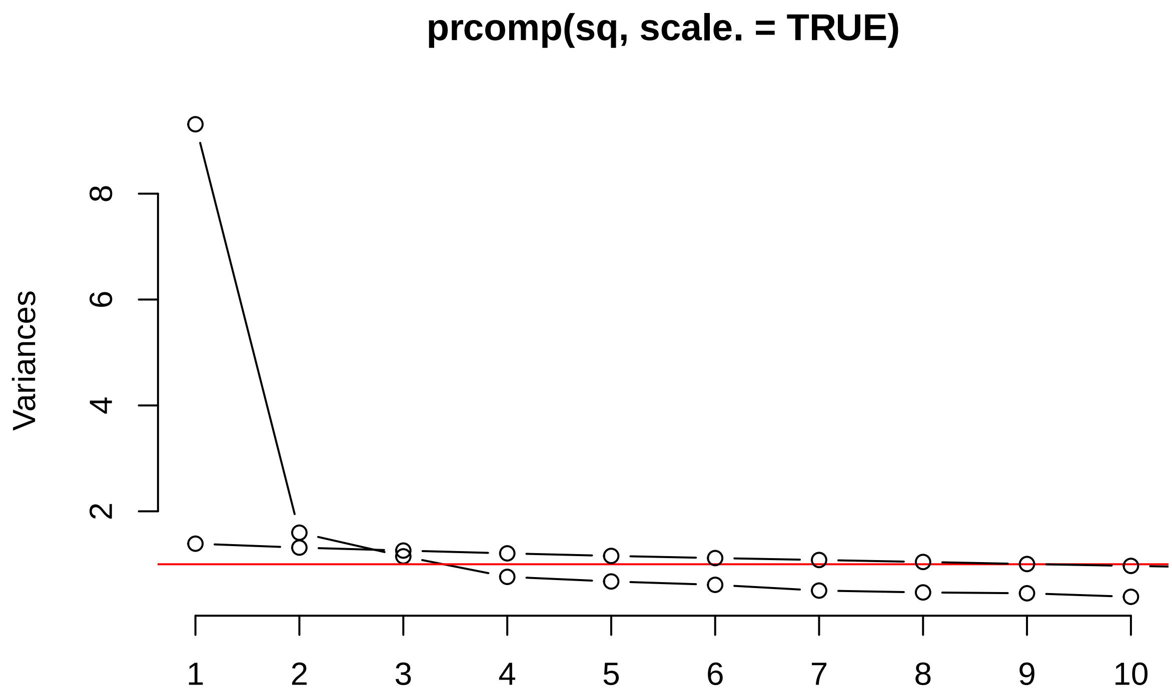
Question 1) Earlier, we examined a dataset from a security survey sent to customers of e-commerce websites. However, we only used the eigenvalue > 1 criteria and the screeplot "elbow" rule to find a suitable number of components. Let's perform a parallel analysis as well this week:

```
sq <- read.csv("security_questions.csv")
```

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.

```
sim_noise_ev <- function(n, p) {
noise <- data.frame(replicate(p, rnorm(n)))
eigen(cor(noise))$values
}
evalues_noise <- replicate(100, sim_noise_ev(405, 18))
evalues_mean <- apply(evalues_noise, 1, mean)

screeplot(prcomp(sq, scale. = TRUE), type="lines")
abline(h = 1, col = "red")
lines(evalues_mean, type="b")
```



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

I would retain 10 dimensions.

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)
sq_principal <- principal(sq, 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?

```
sq_principal
```

```
## 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
##
## 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
```

>0.70 is considered a good loading, so most of variable seem to best belong to PC1, but Q4, Q12, Q17 best belong to PC2.

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

0.67

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

Q2

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

```
sq_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
```

Q4,Q7, Q9,Q12, Q17 ### 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 highest load is Q1, I am convinced that this site respects the confidentiality of the transactions received from me . Since it is a security survey sent to customers of e-commerce websites, if this site respects the confidentiality of the transactions received from me, it is the most important question for the survey.

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

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

```
sq_pca_rot <- principal(sq, nfactor=3, rotate="varimax", scores=TRUE)
sq_pca_rot
```

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

The proportion variance explained by RC1 RC2 RC3 are 0.31 0.16 0.19, proportion variance explained by PC1 PC2 PC3 are 0.52 0.09 0.06, so PC1 explain more variance than RC1 , but RC2 and RC3 explain more variance than PC2 and PC3.

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

The three rotated components explain the same cumulative variance as the three principal components combined, both explained 0.67.

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?

```
sq_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, those items have more clearly differentiated loadings among rotated 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, now we can more easily interpret the meaning of the 3 rotated components from the items.

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

Yes, when you reduce the number of extracted and rotated components from a factor analysis, the meaning of the rotated components can change.