Exercise06

Dennis Perrett

1/15/2022

Exercise 21. A total of 31 athletes took part in the decathlon at the 1996 Olympic Games in Atlanta, USA. The disciplines were (1) 100 metres, (2) long jump, (3) shot put, (4) high jump, (5) 400 metres, (6) 110 metres hurdles, (7) discus throw, (8) pole vault, (9) javelin, (10) 1500 metres. The (10×10)-correlation matrix of the results was analyzed with a principal component analysis. The eigenvalues were found to be

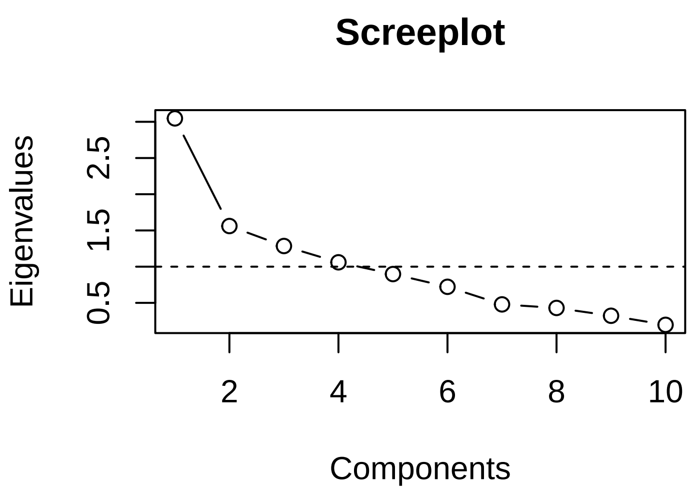
$\lambda = (3.047, 1.561, 1.285, 1.060, 0.898, 0.722, 0.479, 0.430, 0.322, 0.197)'$

and the matrix of eigenvectors:

$$G = \begin{pmatrix} 0.470 & -0.335 & -0.016 & 0.016 & -0.086 & 0.009 & -0.116 & -0.214 & 0.004 & 0.774 \\ -0.377 & -.0236 & -0.201 & -0.357 & 0.036 & 0.244 & -0.535 & 0.477 & 0.164 & 0.182 \\ -0.216 & -0.273 & -0.429 & 0.388 & 0.503 & -0.141 & 0.333 & 0.037 & 0.382 & 0.112 \\ -0.066 & -0.578 & 0.122 & -0.412 & 0.172 & 0.383 & 0.380 & -0.227 & -0.239 & -0.193 \\ 0.316 & -0.514 & 0.046 & 0.142 & -0.349 & -0.178 & -0.209 & 0.037 & 0.434 & -0.476 \\ 0.413 & -0.055 & -0.184 & -0.272 & 0.113 & -0.396 & 0.258 & 0.622 & -0.305 & -0.043 \\ -0.332 & -0.351 & 0.224 & 0.275 & 0.131 & -0.383 & -0.323 & -0.080 & -0.528 & 0.001 \\ -0.288 & 0.033 & -0.255 & -0.529 & -0.283 & -0.547 & 0.128 & -0.367 & 0.183 & 0.086 \\ -0.321 & -0.149 & -0.055 & 0.296 & -0.678 & 0.129 & 0.407 & 0.291 & -0.147 & 0.189 \\ -0.138 & -0.008 & 0.778 & -0.118 & 0.128 & -0.189 & 0.221 & 0.243 & 0.394 & 0.214 \end{pmatrix}$$

was obtained.

a. Draw a scree plot and determine the number of components to be extracted according to the scree test criterion as well as the Kaiser-Guttman criterion.



Basic assumption: Since the factor analysis is a data-reducing procedure , it seems sensible to only keep those factors as the maximum number that explain more variance than the original variables. This is only the case for factors with eigenvalues greater than one.

The Kaiser-Guttman criterion is very easy to use and always leads to a clear solution. It is therefore used as a standard in many statistics packages when performing a factor analysis. It often overestimates the dimensionality and is hardly helpful for gaining meaningful interpretable factors.

b. What is the proportion of the total variance accounted for by the first two principal components?

$$\frac{\sum_{selected} \lambda}{\sum_{all} \lambda} = 0.4607539 = 46\%$$

c. Compute the loading matrix L of a two-component solution.

$$L = G \cdot \begin{pmatrix} \lambda_1^2 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \lambda_p^2 \end{pmatrix}, \text{Take first 2 columns, for first 2 components.}$$

```
L <- G%*%diag(sqrt(lambda))
L <- L[,1:2]
```

d. Compute the loading matrix $\tilde{L}$ after rotation according to the varimax criterion. What is the proportion of the total variance accounted for by each of the two rotated components? Note: Use the rotation matrix

$$M = \begin{pmatrix} 0.783 & 0.623 \\ -0.623 & 0.783 \end{pmatrix}$$

$$\tilde{L} = L \cdot M$$

Portion of explained variance:

$$\frac{\sum \vec{L}_i^2}{10}, \text{ where } \vec{L}_i^2 \text{ is the column sum of the rotated loading vector, and 10 because we have 10 variables.}$$

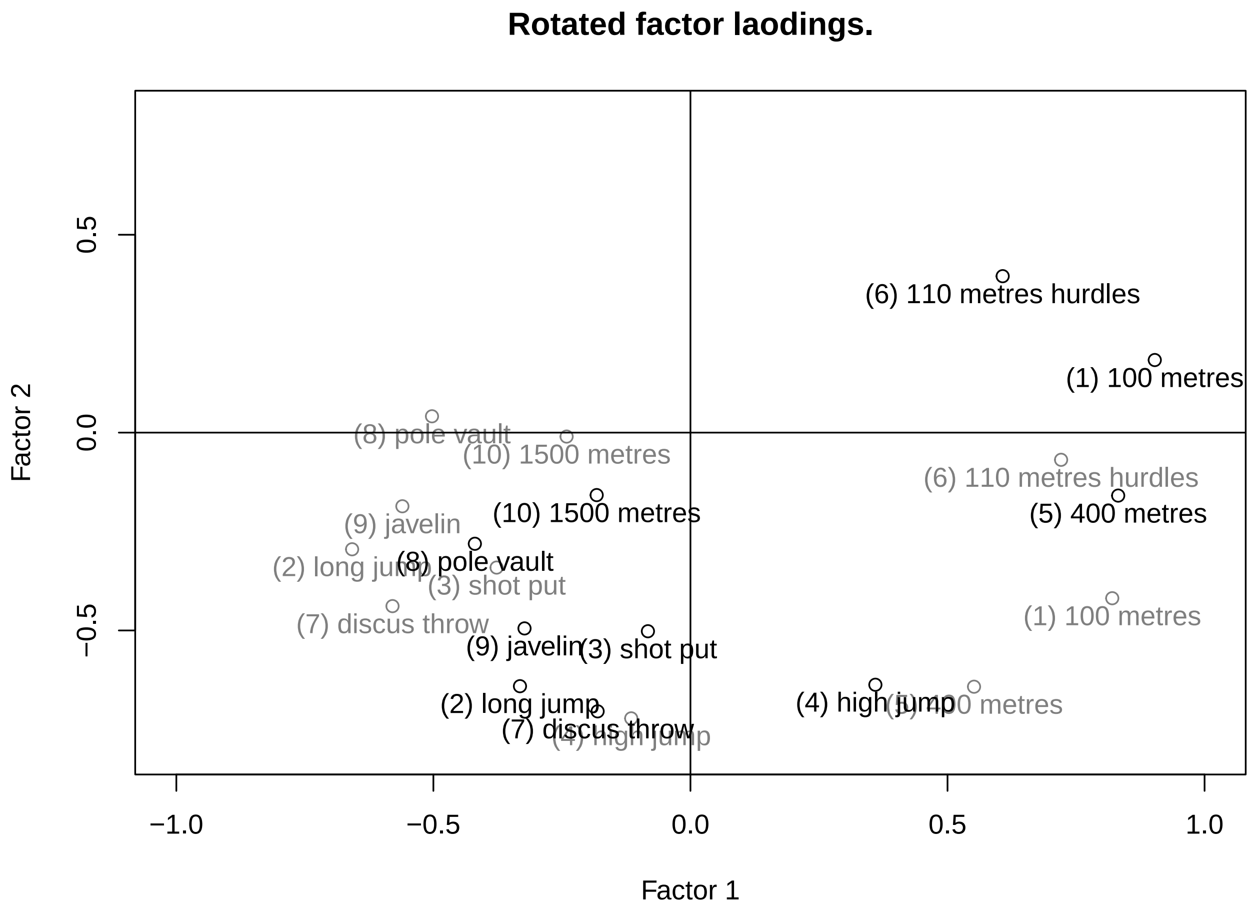
```
M <- matrix(c(0.783,0.623,-0.623,0.783),2,2,T)
Lvari <- L%*%M
print(Lvari)
```

```
##              [,1]       [,2]
##  [1,]  0.90314167  0.1833953
##  [2,] -0.33157856 -0.6408569
##  [3,] -0.08272738 -0.5019678
##  [4,]  0.35969404 -0.6372201
##  [5,]  0.83198718 -0.1591899
##  [6,]  0.60729001  0.3953269
##  [7,] -0.18056029 -0.7044222
##  [8,] -0.41931848 -0.2809132
##  [9,] -0.32275771 -0.4948472
## [10,] -0.18238836 -0.1578995
```

```
# Portion of total variance
print(colSums(Lvari^2)/10) #10 because 10 variables.
```

```
## [1] 0.2468705 0.2128906
```

e. Plot the rotated factor loadings into a diagram. How can the two components be interpreted?



The 2 factors are tough to interpret as most items load equally strong on both factors. Some factors laod positively on both, whereas some load negatively on both. The 2 factors together seem to identify running sports, and strength sports.

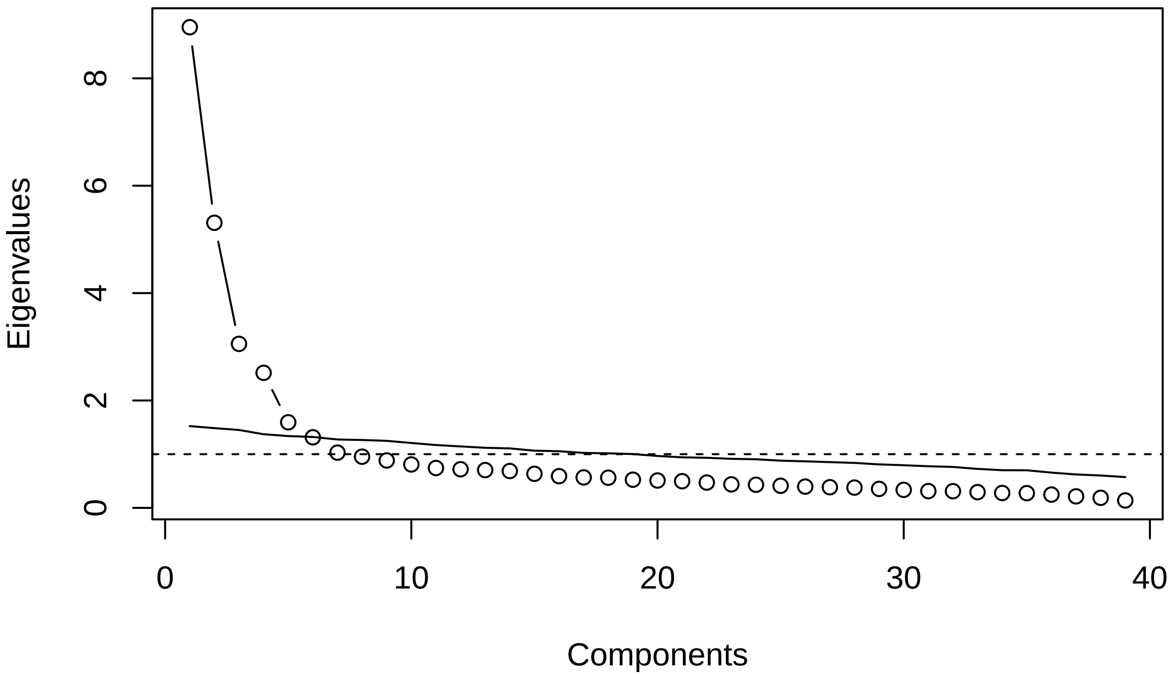
Exercise 22. The file KIMS_codebook.txt lists the items of the Kentucky Inventory of Mindful Skills (KIMS; Baer, Smith, & Allen, 2004) and the file KIMS_data.csv contains the publicly available (https://openpsychometrics.org/_rawdata/) data that were col- lected using an online version of the KIMS (sample size n = 601). The questionnaire collects responses to 39 items (Likert-type responses coded 1 = Never or very rarely true, 2 = Rarely true, 3 = Sometimes true, 4 = Often true, 5 = Very often or always true, and 0 = none selected) in four scales labeled "Observing",Describing","Act With Awareness," and "Accept Without Judgment".

Then dat contains the data matrix for the following analyses.

a. Create a scree plot, and add an illustration of the results of the analysis for random data.

```
R <- cor(dat)
lambda <- eigen(R)$values
G <-eigen(R)$vectors

set.seed(123)
sim_lambda <- eigen(cor(matrix(sample(1:5,nsubjects*nitems,T),nsubjects,nitems)))$values # randoms samp
plot(lambda, xlab = "Components", ylab="Eigenvalues", type="b")
points(sim_lambda, type="l")
abline(h=1,lty=2)
```



b. Decide how many components to extract according to the Kaiser-Guttman criterion and according to parallel analysis. Let the further analyses be based on the components extracted according to parallel analysis.

```
# select 5 values. Bigger than randomly simulated data.
# Parallel analysis
max(which(sim_lambda < lambda))
```

```
## [1] 5
```

```
# KG-Criterion
max(which(1 < lambda))
```

```
## [1] 7
```

Parallel analysis suggests using 5 factors, as 5 eigenvalues have a value than randomly produced data (ie actually hold some information about variance). KG-Criterion suggests using 7, as 7 eigenvalues have a value greater than 1.

c. Compute the factor loadings resulting from a varimax rotation, and assign items to components based on the maximum absolute value.

```
pca <-princomp(dat, cor=T)
eigval <- pca$sdev^2
L <- loadings(pca) %*% diag(pca$sdev)
lt <- varimax(L[,1:5])$loadings

# When assigning to max abs value. Just ignore all non- abs(max(x)) values.

ic2 <- apply(lt, 1, function (x) which.max(abs(x)))

number <- sapply(1:5, function (c) names(ic2[ic2==c]))
content <- sapply(1:5, function (c) paste0(names(ic2[ic2==c]),":",
                           text[which(ic2==c)]))
#Results
number
```

```
## [[1]]
## [1] "Q4"  "Q8"  "Q12" "Q16" "Q20" "Q24" "Q28" "Q32" "Q36"
##
## [[2]]
##  [1] "Q1"  "Q5"  "Q9"  "Q13" "Q17" "Q21" "Q25" "Q29" "Q30" "Q33" "Q37" "Q39"
##
## [[3]]
## [1] "Q2"  "Q6"  "Q10" "Q14" "Q18" "Q22" "Q26" "Q34"
##
## [[4]]
## [1] "Q7"  "Q15" "Q19" "Q31" "Q38"
##
## [[5]]
## [1] "Q3"  "Q11" "Q23" "Q27" "Q35"
```

d. Compare this assignment to the "Observing", "Describing", "Act With Awareness", and "Accept Without Judgment" scales from Baer et al. (2004), and try to interpret any discrepancies.

```r
#head(content)
```

The analysis produced by Baer et al (2004), reported 4 latent variables (as mentioned above). The first 3 factors remain unchanged. The last factor however, has been split into 2 factors: 4 -> 5. This makes sense, because each additional factor, explains additional variance.

e. Compute the factor scores after varimax rotation for each individual. Note: The unrotated factor scores are are returned by princomp() in variable scores, and the rotation matrix M is returned by varimax() in variable rotmat. (For more details and examples see the help pages ?princomp and ?varimax.)

```r
FSrot <- pca$scores[,1:5] %*% varimax(L[,1:5])$rotmat
head(FSrot)
```

```
##        [,1]       [,2]      [,3]        [,4]       [,5]
## 1 -1.480265  1.7219820  2.802827 -3.02819202  2.5970590
## 2 -4.121768  0.9251433  5.150285 -0.96207256 -1.2513448
## 3  3.158037  4.6702212 -3.363733 -1.13907484  2.6984750
## 4 -2.523692 -0.2849671  1.426428  0.35843050  1.0468205
## 5 -2.489814  0.7495833 -1.008468 -0.05926353 -0.2066156
## 6 -1.026629 -1.1398923  4.530939  0.12172614 -3.1619967
```

References: Baer, R.A., Smith, G.T., & Allen, K.B. (2004). Assessment of mindfulness by self-report: The Kentucky Inventory of Mindfulness Skills. Assessment, 11, 191-206. [Link] 2