EFA

Ismail Guennouni

2024-03-20

```r
# Convert empty strings to NA and ensure numeric columns are indeed numeric.
# Replace the missing cells with column median
df_clean <- df_efa %>%
  mutate(across(everything(), ~as.numeric(as.character(.)))) %>%
  mutate(across(everything(), ~ifelse(is.na(.), median(., na.rm = TRUE), .)))
```

```r
library(nFactors)
```

```
## Loading required package: lattice

##
## Attaching package: 'nFactors'

## The following object is masked from 'package:lattice':
##
##     parallel
```

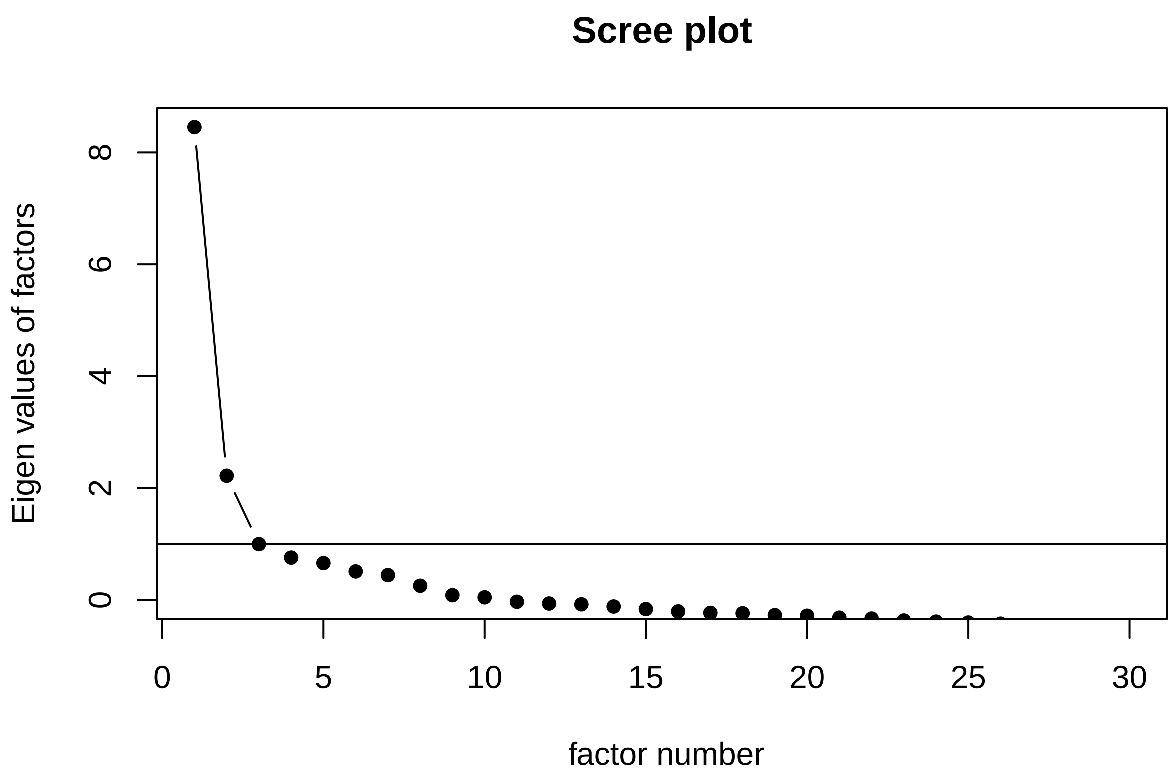
```r
library(psych)
```

```
##
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

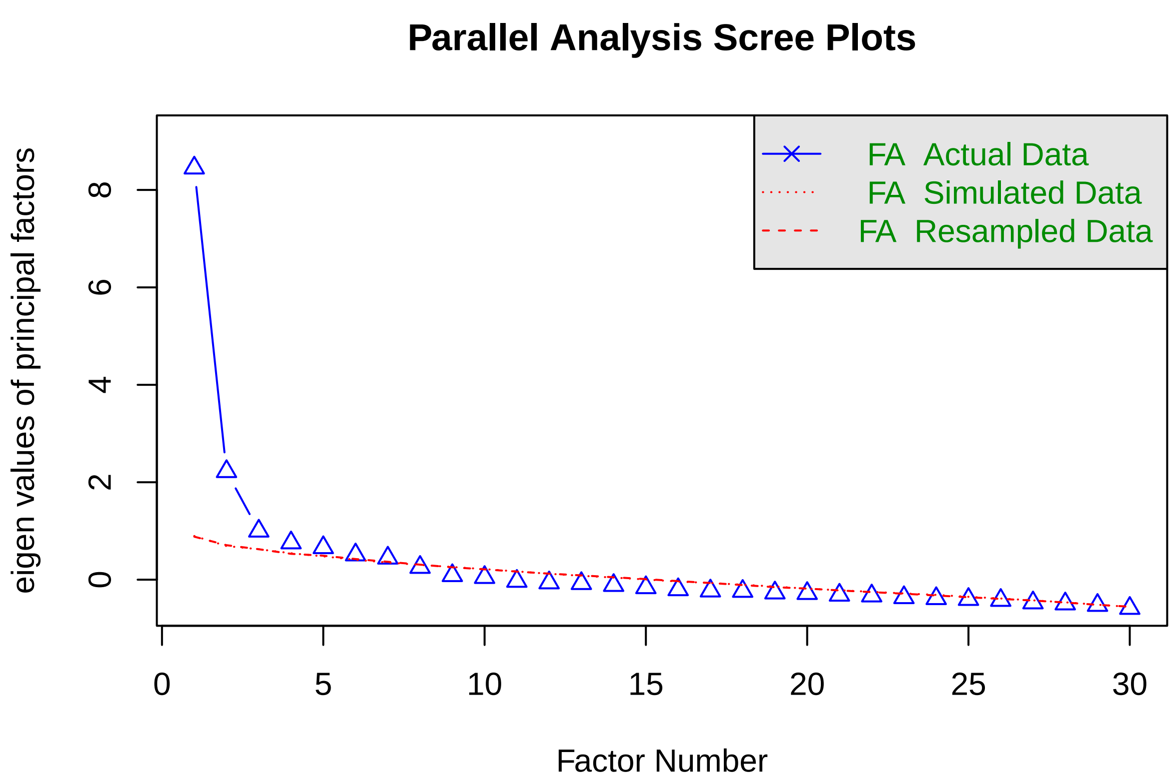
```r
ev <- eigen(cor(df_clean)) # get eigenvalues

ev$values
```

```
##  [1] 9.1422873 2.9211061 1.7303917 1.4591441 1.4296230 1.2973769 1.1637773
##  [8] 1.0258351 0.8429210 0.7910102 0.7207175 0.6943471 0.6235238 0.5851839
## [15] 0.5524589 0.5047850 0.4778792 0.4395217 0.4282085 0.3938658 0.3803313
## [22] 0.3662465 0.3250501 0.3011762 0.2696578 0.2626157 0.2470221 0.2337046
## [29] 0.2082267 0.1820048
```

```r
scree(df_clean, pc=FALSE)  # Scree plot of the eigen values. Use pc=FALSE for factor analysis
```




```r
# another way to plot the eigenvalue
fa.parallel(df_clean, fa="fa")
```




```
## Parallel analysis suggests that the number of factors =  7  and the number of components =  NA
```

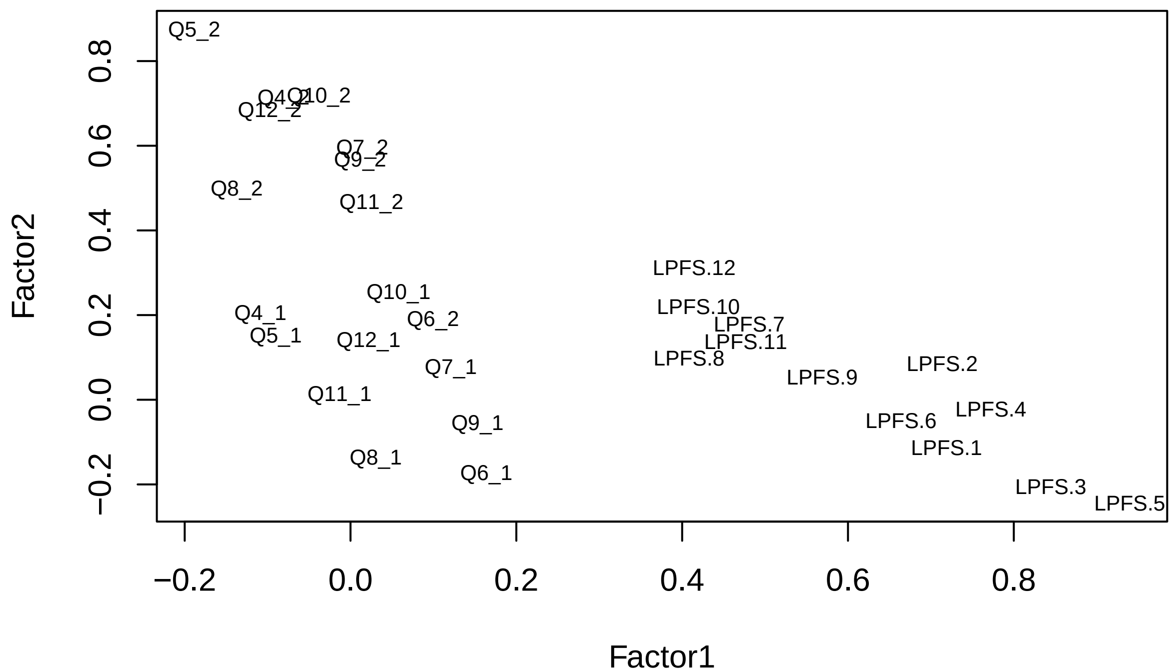
The eigenvalue method ("Kaiser's rule") is telling us that 8 factors may be best. The scree plot is putting us somewhere between three and five factors. Parallel analysis suggests similar. THink 3 is adequate for such a small dataset but we can also use other methods such as CNG.

```
Nfacs <- 3  # Think 3 factors is adequate for this small dataset. You can change this as needed.

# start with assumption of correlation btw factos. so using promax.
fit <- factanal(df_clean, Nfacs, rotation="promax")


print(fit, digits=2, cutoff=0.3, sort=TRUE)
```

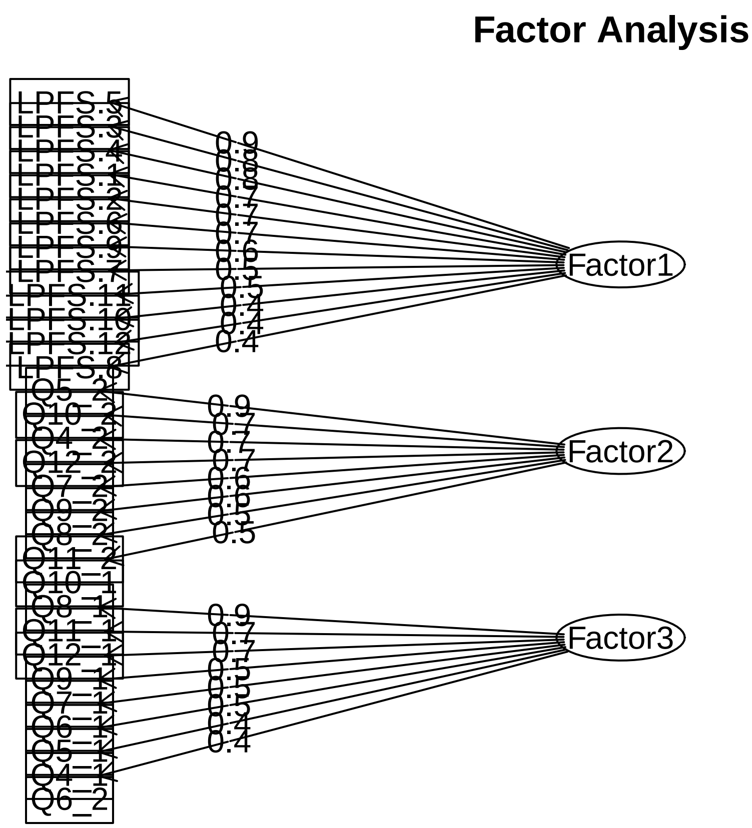
```
##
## Call:
## factanal(x = df_clean, factors = Nfacs, rotation = "promax")
##
## Uniquenesses:
##    Q4_1    Q5_1    Q6_1    Q7_1    Q8_1    Q9_1   Q10_1   Q11_1   Q12_1    Q4_2
##    0.76    0.75    0.79    0.65    0.36    0.68    0.76    0.44    0.41    0.67
##    Q5_2    Q6_2    Q7_2    Q8_2    Q9_2   Q10_2   Q11_2   Q12_2  LPFS.1  LPFS.2
##    0.43    0.83    0.70    0.59    0.65    0.57    0.61    0.44    0.52    0.36
##  LPFS.3  LPFS.4  LPFS.5  LPFS.6  LPFS.7  LPFS.8  LPFS.9 LPFS.10 LPFS.11 LPFS.12
##    0.39    0.45    0.33    0.60    0.71    0.81    0.68    0.71    0.62    0.62
##
## Loadings:
##         Factor1 Factor2 Factor3
## LPFS.1   0.72
## LPFS.2   0.71
## LPFS.3   0.84
## LPFS.4   0.77
## LPFS.5   0.94
## LPFS.6   0.66
## LPFS.9   0.57
## Q4_2             0.71
## Q5_2             0.87
## Q7_2             0.59
## Q9_2             0.57
## Q10_2            0.72
## Q12_2            0.68
## Q8_1                     0.87
## Q9_1                     0.52
## Q11_1                    0.75
## Q12_1                    0.66
## Q4_1                     0.38
## Q5_1                     0.43
## Q6_1                     0.47
## Q7_1                     0.47
## Q10_1
## Q6_2
## Q8_2             0.50
## Q11_2            0.47
## LPFS.7   0.48
## LPFS.8   0.41
## LPFS.10  0.42
```

```
## LPFS.11  0.48
## LPFS.12  0.41    0.31
##
##                  Factor1 Factor2 Factor3
## SS loadings         5.13    3.97    3.16
## Proportion Var      0.17    0.13    0.11
## Cumulative Var      0.17    0.30    0.41
##
## Factor Correlations:
##          Factor1 Factor2 Factor3
## Factor1    1.00    0.45   -0.61
## Factor2    0.45    1.00   -0.64
## Factor3   -0.61   -0.64    1.00
##
## Test of the hypothesis that 3 factors are sufficient.
## The chi square statistic is 801.92 on 348 degrees of freedom.
## The p-value is 7.57e-38
```

```r
load <- fit$loadings[,1:2]
plot(load,type="n") # set up plot
text(load,labels=names(df_clean),cex=.7)
```




```r
loads <- fit$loadings

fa.diagram(loads)
```




```r
dim(fit$loadings)
```

```
## [1] 30  3
```

```r
round(fit$loadings[ 1:30,], 2)
```

```
##         Factor1 Factor2 Factor3
## Q4_1      -0.11    0.20    0.38
## Q5_1      -0.09    0.15    0.43
## Q6_1       0.16   -0.17    0.47
## Q7_1       0.12    0.08    0.47
## Q8_1       0.03   -0.14    0.87
## Q9_1       0.15   -0.06    0.52
## Q10_1      0.06    0.25    0.25
## Q11_1     -0.01    0.01    0.75
## Q12_1      0.02    0.14    0.66
## Q4_2      -0.08    0.71   -0.17
## Q5_2      -0.19    0.87   -0.04
## Q6_2       0.10    0.19    0.19
## Q7_2       0.01    0.59   -0.09
## Q8_2      -0.14    0.50    0.28
## Q9_2       0.01    0.57    0.02
## Q10_2     -0.04    0.72   -0.06
## Q11_2      0.03    0.47    0.19
## Q12_2     -0.10    0.68    0.17
## LPFS.1     0.72   -0.11    0.09
```

```
## LPFS.2      0.71    0.09    0.06
## LPFS.3      0.84   -0.20    0.10
## LPFS.4      0.77   -0.02   -0.04
## LPFS.5      0.94   -0.24    0.02
## LPFS.6      0.66   -0.05   -0.02
## LPFS.7      0.48    0.18   -0.15
## LPFS.8      0.41    0.10   -0.09
## LPFS.9      0.57    0.05   -0.10
## LPFS.10     0.42    0.22   -0.09
## LPFS.11     0.48    0.14    0.08
## LPFS.12     0.41    0.31   -0.07
```