nFactors Example-1

Gilles Raiche

October 13, 2009

In this example, the recovery of the dimensionality of a simulated factor structure is explored with different indices available in the `nFactors` package. The example is inspired from a simulation study by Zwick and Velicer (1986, table 2, p. 437).

The following code is for the initialisation phase. First, the libraries `nFactors` and `xtable` must be loaded.

```
-------------------------------------------------------------------
 require(xtable)
 library(nFactors)
 nFactors  <- 3
 unique    <- 0.2
 loadings  <- 0.8
 nsubjects <- 180
 repsim    <- 100
 var       <- 36
 pmjc      <- 12
 reppar    <- 100
 index     <- 1:11
 zwick     <- generateStructure(var=var, mjc=nFactors, pmjc=pmjc,
                                  loadings=loadings,
                                  unique=unique)
-------------------------------------------------------------------
```

After that, from 36 variables and 180 subjects a 3 factor solution with 12 variables showing a correlation of 0.5 on their respective factors and 0.2 on the others is simulated. To assure good values of the percentiles, 100 replications of the simulation and 100 replications of the parallel analysis are done.

The following code produces statistics about the number of factors to retain from the 100 replications of a parallel analysis on 100 sampled correlation matrices (factor solution). Note that only the indices corresponding to the `index` parameter are considered. Note also that for the sake of this report the `xtable` function is applied. In a usual R session the print function would be used instead. The same commentary is also of concern for the next table from a principal component analysis.

```
-------------------------------------------------------------------
 mzwick.fa <- structureSim(fload=as.matrix(zwick), reppar=reppar,
                           repsim=repsim, details=TRUE,
                           N=nsubjects, quantile=0.5,
                           model="factors")

 xtable(mzwick.fa[[2]][,index],
        caption="Distribution of the number of factors to retain
        according to different indices (factor analysis)")

 xtable(mzwick.fa[[2]][-4,index] - nFactors,
        caption="Distribution of residuals of the number of factors
        to retain according to different indices (factor analysis)")
--------------------------------------------------------------------
```

	oc	af	par	mean.eig	per	cng	b	t.b	p.b	sescree	R2
mean	5.92	1.00	7.92	7.92	7.92	3.00	4.00	26.20	26.20	19.97	1.04
median	5.50	1.00	8.00	8.00	8.00	3.00	4.00	27.00	27.00	18.00	1.00
quantile	5.50	1.00	8.00	8.00	8.00	3.00	4.00	27.00	27.00	18.00	1.00
sd	2.29	0.00	0.96	0.96	0.96	0.00	0.00	3.41	3.41	5.93	0.20
min	1.00	1.00	5.00	5.00	5.00	3.00	4.00	13.00	13.00	10.00	1.00
max	13.00	1.00	10.00	10.00	10.00	3.00	4.00	30.00	30.00	34.00	2.00

Table 1: Distribution of the number of factors to retain according to different indices (factor analysis)

	oc	af	par	mean.eig	per	cng	b	t.b	p.b	sescree	R2
mean	2.92	-2.00	4.92	4.92	4.92	0.00	1.00	23.20	23.20	16.97	-1.96
median	2.50	-2.00	5.00	5.00	5.00	0.00	1.00	24.00	24.00	15.00	-2.00
quantile	2.50	-2.00	5.00	5.00	5.00	0.00	1.00	24.00	24.00	15.00	-2.00
min	-2.00	-2.00	2.00	2.00	2.00	0.00	1.00	10.00	10.00	7.00	-2.00
max	10.00	-2.00	7.00	7.00	7.00	0.00	1.00	27.00	27.00	31.00	-1.00

Table 2: Distribution of residuals of the number of factors to retain according to different indices (factor analysis)

The following code produces statistics about about the number of factors to retain from a replication of a parallel analysis on 100 sampled correlation matrices (principal components solution).

```
-------------------------------------------------------------------------
 mzwick      <-  structureSim(fload=as.matrix(zwick), reppar=reppar,
                             repsim=repsim, details=TRUE,
                             N=nsubjects, quantile=0.5)

 xtable(mzwick[[2]][,index],
        caption="Distribution of the number of factors to retain
        according to different indices (principal component analysis)")

 xtable(mzwick[[2]][-4,index] - nFactors,
        caption="Distribution of residuals of the number of factors to
        retain according to different indices (principal component
        analysis)")
-------------------------------------------------------------------------
```

	oc	af	par	mean.eig	per	cng	b	t.b	p.b	sescree	R2
mean	2.52	1.00	2.54	7.97	2.57	3.00	4.00	26.33	26.33	7.32	1.04
median	3.00	1.00	3.00	8.00	3.00	3.00	4.00	27.00	27.00	7.00	1.00
quantile	3.00	1.00	3.00	8.00	3.00	3.00	4.00	27.00	27.00	7.00	1.00
sd	0.59	0.00	0.58	0.93	0.56	0.00	0.00	3.37	3.37	1.71	0.20
min	1.00	1.00	1.00	6.00	1.00	3.00	4.00	16.00	16.00	3.00	1.00
max	3.00	1.00	3.00	10.00	3.00	3.00	4.00	30.00	30.00	12.00	2.00

Table 3: Distribution of the number of factors to retain according to different indices (principal component analysis)

	oc	af	par	mean.eig	per	cng	b	t.b	p.b	sescree	R2
mean	-0.48	-2.00	-0.46	4.97	-0.43	0.00	1.00	23.33	23.33	4.32	-1.96
median	0.00	-2.00	0.00	5.00	0.00	0.00	1.00	24.00	24.00	4.00	-2.00
quantile	0.00	-2.00	0.00	5.00	0.00	0.00	1.00	24.00	24.00	4.00	-2.00
min	-2.00	-2.00	-2.00	3.00	-2.00	0.00	1.00	13.00	13.00	0.00	-2.00
max	0.00	-2.00	0.00	7.00	0.00	0.00	1.00	27.00	27.00	9.00	-1.00

Table 4: Distribution of residuals of the number of factors to retain according to different indices (principal component analysis)

An index accuracy plot is produced to compare the number of factors to retain from the principal components analysis solution (upper one) (figure 1). Another plot is also produced for the factor analysis solution (lower one). This plot shows the number of factors retain by each index. It can be seen that the *af* and *R2* indices underestimate the number of factors, while the *p.b* and *t.p* clearly show overestimation. In the context of factor analysis, *sescree* also shows overestimation. The continuous vertical line indicates the initial number of factors of the structure matrix. The dashed line indicates the median number of factors retained by each index. These plots show a general tendency to orverestimate the number of factors to retain.

```
---------------------------------------------------------------------------
par(mfrow=c(2,1))
plot(x=mzwick, nFactors=nFactors, index=index, cex.axis=0.6, col="red",
     main="Principal Component Analysis")
plot(x=mzwick.fa, nFactors=nFactors, index=index, cex.axis=0.6, col="red",
     main="Factor Analysis")
par(mfrow=c(1,1))
---------------------------------------------------------------------------
```

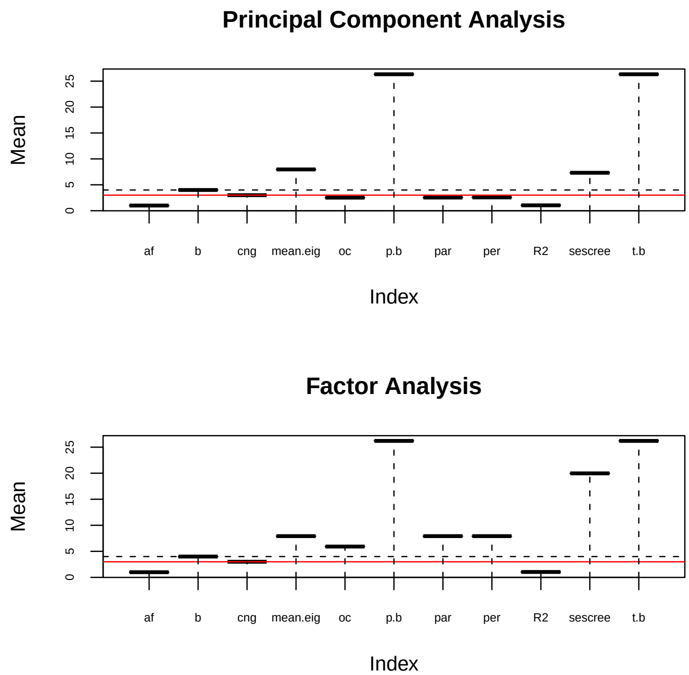

Figure 1: Index accuracy plot of the average number of components retained according to different indices

An eigen boxplot is produced to show the probability distribution of the simulated eigenvalues from the principal components analysis solution (figure 2). Another plot is also produced for the factor analysis solution (lower one). For the sake of a better graphical presentation, only the first 10 eigenvalues are illustrated. It can be seen that the distribution of the first eigenvalue is asymetric for the principal component analysis, while it is somewhat symetric for the factor analysis.

```
---------------------------------------------------------------
par(mfrow=c(2,1))
boxplot(mzwick,    nFactors=3, xlab="Components", cex.axis=0.7,
        eigenSelect=1:10, vLine="blue", col="red")
boxplot(mzwick.fa, nFactors=3, cex.axis=0.7, eigenSelect=1:10,
        vLine="blue", col="red",
        main="Factor Analysis")
par(mfrow=c(1,1))
---------------------------------------------------------------
```

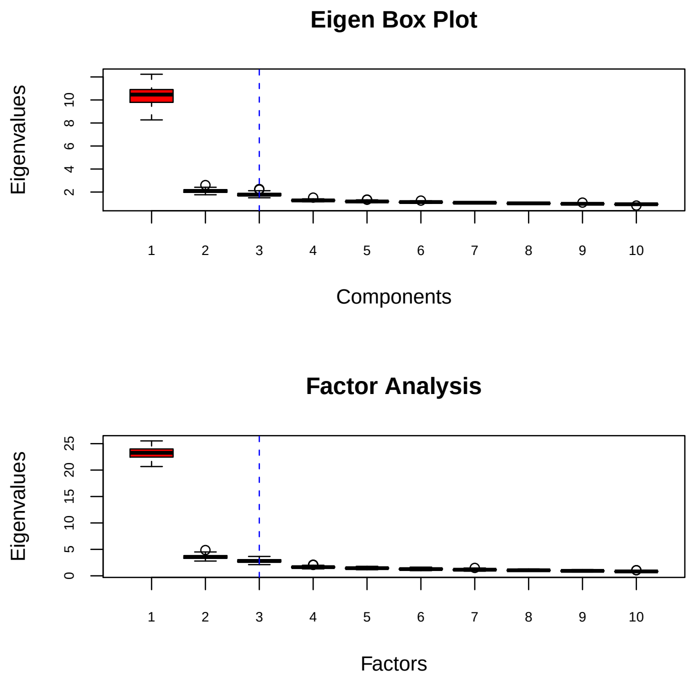

Figure 2: Eigen boxplot (principal component analysis)

References

Raiche, G., Riopel, M. and Blais, J.-G. (2006). *Non graphical solutions for the Cattell's scree test*. Paper presented at the International Annual meeting of the Psychometric Society, Montreal.

Zwick, W. R. and Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432-442.