Exploratory Factor Analysis Using R

Companion Reading: Bandalos, p. 141-143, Furr, Ch. 4 (EFA);
Bandalos, p. 376-379 (Model fit indices)

 # The dataset <ELS04 admin school climate.csv> shows responses of 657 US high school administrators to 12 survey items from the 2004 wave of the Education Longitudinal Study of 2002-2006. These items asked school principals to rate various aspects of school climate (e.g., how hard do students work?, is there much bullying/school violence?) Labels for each item are given below. The rating scale for each item ranged from 1 (not at all accurate) to 5 (very accurate); note that not all items have the same directionality – some are positively, and others negatively, phrased.

Name	Type	Width	Decimals	Label	
stmorale	Numeric	3	0	Student morale is high	
press	Numeric	3	0	Teachers press students to achieve	
tcmorale	Numeric	3	0	Teacher morale is high	
hilearn	Numeric	3	0	Learning is high priority for students	
hwexpect	Numeric	3	0	Students expected to do homework	
discipln	Numeric	3	0	Discipline is emphasized	
structur	Numeric	3	0	Classroom activities are highly structured	
negtchrs	Numeric	3	0	Many teachers are negative about students	
lowmotiv	Numeric	3	0	Many teachers find it difficult to motivate students	
academic	Numeric	3	0	Counselors/teachers encourage students to enroll in academic classes	
conflict	Numeric	3	0	There is often conflict between teachers and administrators	
indivlrn	Numeric	3	0	Teachers usually respond to students' individual needs	

 # Hypothesis: Let's say that based on our knowledge of the literature about school climate, **we hypothesize at least two underlying factors**: one related to the *academic* dimension of school climate, and one related to the *socioemotional* dimension of school climate. Also, we hypothesize that these two dimensions are positively correlated. Unless you have evidence from a previous dataset regarding the underlying number of factors, and which items load highly on each factor, it is generally considered best practice to examine EFA models for different potential numbers of factors. Based on our initial hypothesis, we should plan to compare models up to at least 3 factors.

 # Handling negatively-worded items: We observe that items have mixed directionality or polarity (i.e., some items are negatively worded). Methodologists do not fully agree about how to handle this situation: Traditionally it was suggested to reverse coding of the negative-worded items before FA, but this changes correlations among the items, and so will alter the results. Contemporary statisticians in latent variable modeling (e.g., Muthen, Bentler) recommend leaving the items in their original measurement scales, because reverse coding is not required for meaningful interpretation of the FA results. Some psychometricians suggest avoiding use of negatively-worded items to prevent the occurrence of spurious 'factors' attributable to those items' tone.

 # Import Excel .csv file

```
> schoolclimatedata <- read.table("C:/Users/username/Desktop/ELS04 admin school climate
items.csv", header = TRUE, sep = ",")
```

```
# The equivalent file pathname for Mac would be "/Users/username/Desktop/ELS04 admin
school climate items.csv"


# Use the GPArotation package and Revelle's 'psych' package for exploratory factor
analysis in R [after installing either or both packages, if necessary:
install.packages("psych")]

# Activate both packages

> require(psych)

> require(GPArotation)


# Check basic descriptive statistics for all items

> describe(schoolclimatedata)


# Examine histogram for one item that has a clearly non-normal response distribution

> hist(schoolclimatedata$conflict)


# Compute Pearson correlation matrix for response data from these survey items

> lowerCor(schoolclimatedata)

# Conduct "parallel analysis" comparing eigenvalues for maximum likelihood extraction of
factors from the original dataset, and from 100 random draws from the original dataset
(which, loosely, randomly scramble the original dataset, reducing the observed
correlations among variables)

> fa.parallel(schoolclimatedata, fm="ml", fa="fa", main = "Parallel Analysis Scree
Plots", n.iter=100, error.bars = T, sim=F)


# Considering parallel analysis results, conduct exploratory FA for models with up to 5
underlying factors: (a) indicate that software should compute and analyze the correlation
matrix for scores [ratings] on the ELS items, (b) find solution for [#] latent factors
underlying data [should be greater than or equal to hypothesized maximum number of
underlying factors, and have at least 2 items per factor], (c) use an oblique rotation –
promax – that allows factors to be correlated, and (d) use maximum likelihood as the
estimation method, which will use all available data, even from cases with missing data,
to obtain results

> fa(r = schoolclimatedata, nfactors = 2, rotate = "promax", fm = "ml")
```