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.

######## **Annotated Output/Results** #########

 # Descriptive statistics indicate low missingness (n = 657). Response distributions for most items are approximately normal; the *discipln*, *negtchrs*, and *conflict* items have excessive skewness, and the *conflict* item also has high kurtosis. Looking forward, we might expect some bias (inaccuracy) in the factor loading estimates for the *conflict* item, particularly, but the maximum likelihood estimation that will be used for FA has some robustness to non-normality.

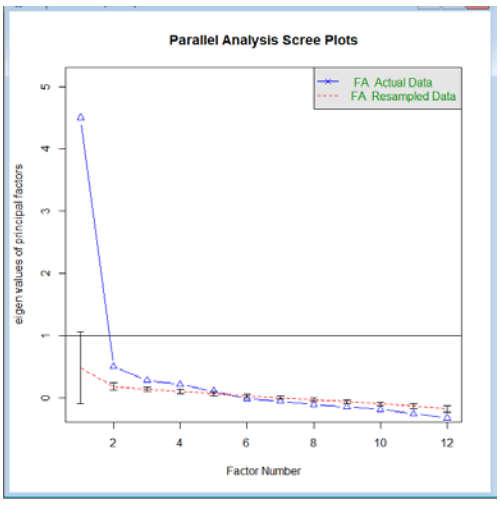
```
> describe(schoolclimatedata)
          vars   n mean   sd median trimmed  mad min max range  skew kurtosis   se
stmorale     1 656 3.94 0.79      4    3.98 0.00   1   5     4 -0.54     0.42 0.03
press        2 655 4.14 0.80      4    4.20 1.48   2   5     3 -0.49    -0.66 0.03
tcmorale     3 648 3.78 0.88      4    3.83 1.48   1   5     4 -0.55     0.18 0.03
```

```
hilearn     4 652 3.45 0.87      3    3.45 1.48   1   5     4  0.00    -0.25 0.03
hwexpect    5 655 4.20 0.88      4    4.30 1.48   1   5     4 -0.87     0.12 0.03
discipln    6 652 4.44 0.73      5    4.57 0.00   2   5     3 -1.21     1.14 0.03
structur    7 652 3.77 0.74      4    3.75 0.00   1   5     4 -0.18    -0.14 0.03
negtchrs    8 655 1.68 0.80      2    1.55 1.48   1   5     4  1.14     1.13 0.03
lowmotiv    9 653 2.34 1.00      2    2.27 1.48   1   5     4  0.44    -0.24 0.04
academic   10 655 4.40 0.72      5    4.51 0.00   1   5     4 -1.00     0.62 0.03
conflict   11 656 1.64 0.74      2    1.53 1.48   1   5     4  1.29     2.30 0.03
indivlrn   12 656 3.97 0.89      4    4.04 1.48   1   5     4 -0.66     0.25 0.03
```

Pearson correlations indicate the items have non-trivial pairwise interrelationships. Particularly, none of the item pairs have near-zero or very small correlations. The weakest correlations appear to be between the 'conflict' item and other items. Some item response pairs have positive, and others have negative, associations.

```
> lowerCor(schoolclimatedata)
         stmrl press tcmrl hilrn hwxpc dscpl strct ngtch lwmtv acdmc cnflc indvl
stmorale  1.00
press     0.51  1.00
tcmorale  0.61  0.54  1.00
hilearn   0.49  0.58  0.45  1.00
hwexpect  0.31  0.54  0.32  0.50  1.00
discipln  0.38  0.38  0.35  0.32  0.43  1.00
structur  0.37  0.50  0.38  0.43  0.42  0.44  1.00
negtchrs -0.34 -0.40 -0.35 -0.37 -0.30 -0.27 -0.29  1.00
lowmotiv -0.36 -0.37 -0.34 -0.48 -0.33 -0.24 -0.25  0.51  1.00
academic  0.31  0.49  0.33  0.36  0.46  0.39  0.39 -0.28 -0.28  1.00
conflict -0.26 -0.25 -0.42 -0.15 -0.09 -0.27 -0.13  0.36  0.22 -0.18  1.00
indivlrn  0.35  0.49  0.41  0.38  0.40  0.27  0.39 -0.36 -0.30  0.40 -0.24  1.00
```

Parallel analysis suggests 4 factors have eigenvalues that clearly exceed those from a randomly-arranged dataset of the same size. The fifth factor is questionable from a parsimony perspective because its eigenvalue just barely exceeds that expected from a random dataset



Parallel analysis suggests that the number of factors = 5

 # Based on the parallel analysis result and our initial hypothesis, we might compare models with 2, 3, 4, and 5 factors.

Preferred models should have the fewest factors possible, but still all or mostly acceptable fit indices (chi-square, RMSR, and TLI) and be substantively interpretable (i.e., the pattern of loadings can be reasonably explained based on theory about the trait and/or knowledge of the item content). Models with more factors will usually fit better because they have more parameters, so it is important that the added factors are interpretable. A model should not have so many factors that many items have no loadings above .4. The order of importance of these criteria in making an overall judgment is *substantive interpretability*, and then *model fit* and *parsimony*.

 # Results for 2-factor model

Chi-square test of model fit (below) tests the null hypothesis that "The model fits the data." Want prob [p-value] > .05 to suggest that model fits data well. With "large" samples (say, greater than 400-500), minor misfit between each person's data and the model may accumulate to produce a significant chi-square value.

The total number of observations was 657 with Likelihood Chi Square = 285.01 with prob < 1.9e-37

Tucker Lewis Index is a model fit statistic. Rule-of-thumb is that models with TLI > .90 fit the data well (Little, 2013, *Longitudinal structural equation modeling*)

Tucker Lewis Index of factoring reliability = 0.867

Root mean square error of approximation (RMSEA) is another model fit statistic. Rule-of-thumb is that models with RMSEA ≤ .05 fit the data well, ≤ .08 fit the data acceptably, and .08 - .10 is marginally acceptable fit (Little, 2013).

RMSEA index = 0.093 and the 90 % confidence intervals are 0.083 0.103
Matrix below gives standardized factor loadings (labeled ML1 – ML2), *communality* (h2; sum of squared loadings), *uniqueness* (u2) for each item, and *complexity* (loosely, an index of the overall extent of cross-loading). Order and numbering of factors is arbitrary.

Standardized loadings (pattern matrix) based upon correlation matrix

	ML1	ML2	h2	u2	com
stmorale	0.15	0.61	0.53	0.47	1.1
press	0.64	0.19	0.62	0.38	1.2
tcmorale	0.01	0.81	0.67	0.33	1.0
hilearn	0.60	0.13	0.49	0.51	1.1
hwexpect	0.93	-0.29	0.57	0.43	1.2
discipln	0.47	0.10	0.30	0.70	1.1
structur	0.61	0.02	0.39	0.61	1.0
negtchrs	-0.28	-0.30	0.29	0.71	2.0
lowmotiv	-0.34	-0.23	0.28	0.72	1.8
academic	0.67	-0.08	0.39	0.61	1.0
conflict	0.16	-0.61	0.26	0.74	1.1
indivlrn	0.48	0.15	0.36	0.64	1.2

Matrix below gives estimated correlation between the two latent factors (0.7)

With factor correlations of

	ML1	ML2
ML1	1.0	
ML2	0.7	1.0

 # Possible conclusion for 2-factor model: None of the model fit indices clearly show acceptable fit of the model to response data from these school climate items. I note, in passing, that the factor loading pattern seems to support our initial hypothesis, with academic items loading highly on Factor 1, and morale items loading highly on Factor 2.

##############################

 # Results for 3-factor model

Chi-square test of model fit (below) tests the null hypothesis that "The model fits the data." Want prob [p-value] > .05 to suggest that model fits data well. With "large" samples (say, greater than 400-500), minor misfit between each person's data and the model may accumulate to produce a significant chi-square value.

The total number of observations was 657 with Likelihood Chi Square = 165.66 with prob < 1.4e-19

Tucker Lewis Index is a model fit statistic. Rule-of-thumb is that models with TLI > .90 fit the data well (Little, 2013, *Longitudinal structural equation modeling*)

Tucker Lewis Index of factoring reliability = 0.905

Root mean square error of approximation (RMSEA) is another model fit statistic. Rule-of-thumb is that models with RMSEA ≤ .05 fit the data well, ≤ .08 fit the data acceptably, and .08 - .10 is marginally acceptable fit (Little, 2013).

RMSEA index = 0.078 and the 90 % confidence intervals are 0.067 0.09

Matrix below gives standardized factor loadings (labeled ML1 – ML3), *communality* (h2), *uniqueness* (u2) for each item, and *complexity* (com).

Standardized loadings (pattern matrix) based upon correlation matrix

	ML2	ML3	ML1	h2	u2	com
stmorale	0.18	0.58	0.00	0.52	0.4836	1.2
press	0.68	0.19	0.04	0.63	0.3675	1.2
tcmorale	0.05	0.86	0.08	0.73	0.2726	1.0
hilearn	0.52	0.06	-0.21	0.50	0.4977	1.3
hwexpect	0.88	-0.26	-0.04	0.56	0.4416	1.2
discipln	0.51	0.11	0.06	0.31	0.6887	1.1
structur	0.66	0.04	0.07	0.41	0.5899	1.0
negtchrs	-0.15	-0.17	0.37	0.35	0.6517	1.7
lowmotiv	0.10	0.07	1.09	0.99	0.0076	1.0
academic	0.68	-0.08	0.00	0.39	0.6071	1.0
conflict	0.13	-0.56	0.02	0.25	0.7533	1.1
indivlrn	0.50	0.14	0.01	0.36	0.6402	1.2

Matrix below gives estimated correlation between the two latent factors (0.7)

With factor correlations of

	ML2	ML3	ML1
ML2	1.00		
ML3	0.67	1.00	
ML1	-0.57	-0.56	1.00

 # Possible conclusion for 3-factor model: Although the chi-square model fit statistic indicates the model does not fit the data (p <.001), perhaps due to large sample size, the RMSEA and TLI indicate an acceptably-fitting model. The loading pattern seems interpretable, with Factor 2 representing academic press, Factor 3 indicating morale, and a single item about difficulty motivating students loading on "Factor" 1. The correlations among the factors indicate that the factors are distinct, and that scoring higher on Factor 1 is associated with lower scores on Factors 2 and 3.

##############################

 # Results for 4-factor model

Chi-square test of model fit (below) tests the null hypothesis that "The model fits the data." Want prob [p-value] > .05 to suggest that model fits data well. With "large" samples (say, greater than 400-500), minor misfit between each person's data and the model may accumulate to produce a significant chi-square value.

The total number of observations was 657 with Likelihood Chi Square = 81.3 with prob < 3.8e-08

Tucker Lewis Index is a model fit statistic. Rule-of-thumb is that models with TLI > .90 fit the data well (Little, 2013, *Longitudinal structural equation modeling*)

Tucker Lewis Index of factoring reliability = 0.943

Root mean square error of approximation (RMSEA) is another model fit statistic. Rule-of-thumb is that models with RMSEA ≤ .05 fit the data well, ≤ .08 fit the data acceptably, and .08 - .10 is marginally acceptable fit (Little, 2013).

RMSEA index = 0.06 and the 90 % confidence intervals are 0.046 0.075

Matrix below gives standardized factor loadings (labeled ML1 – ML3), *communality* (h2), *uniqueness* (u2) for each item, and *complexity* (com).

Standardized loadings (pattern matrix) based upon correlation matrix

	ML1	ML4	ML3	ML2	h2	u2	com
stmorale	0.01	0.68	-0.03	-0.09	0.54	0.46	1.0
press	0.52	0.28	-0.06	0.01	0.62	0.38	1.6
tcmorale	-0.03	0.82	0.12	-0.30	0.74	0.26	1.3
hilearn	0.18	0.39	-0.37	0.27	0.65	0.35	3.2
hwexpect	0.76	-0.08	-0.09	0.18	0.55	0.45	1.2
discipln	0.58	0.03	0.08	-0.12	0.35	0.65	1.1
structur	0.60	0.11	0.05	0.04	0.42	0.58	1.1
negtchrs	-0.07	0.17	0.69	0.33	0.60	0.40	1.6
lowmotiv	0.07	-0.06	0.72	0.04	0.51	0.49	1.0
academic	0.76	-0.12	0.04	-0.04	0.44	0.56	1.1
conflict	0.02	-0.16	0.10	0.54	0.40	0.60	1.2
indivlrn	0.48	0.06	-0.07	-0.12	0.38	0.62	1.2

Matrix below gives estimated correlation between the two latent factors (0.7)

With factor correlations of

	ML1	ML4	ML3	ML2
ML1	1.00			
ML4	0.71	1.00		
ML3	-0.64	-0.61	1.00	
ML2	-0.24	-0.24	0.20	1.00

 # Possible conclusion for 4-factor model: Although the chi-square model fit statistic indicates the model does not fit the data (p <.001), perhaps due to large sample size, the RMSEA and TLI indicate an acceptably-fitting model. The loading pattern seems interpretable, with Factor 1 representing academic press, Factor 4 indicating morale, Factor 3 representing two items related to low pedagogical skill, and a single item about the existence of teacher-administrator conflict loading on "Factor" 2, which has relatively low correlations with the other three factors. The *hilearn* item does not load highly on any of the factors.

##############################

 # Results for 5-factor model

Chi-square test of model fit (below) tests the null hypothesis that "The model fits the data." Want prob [p-value] > .05 to suggest that model fits data well. With "large" samples (say, greater than 400-500), minor misfit between each person's data and the model may accumulate to produce a significant chi-square value.

The total number of observations was 657 with Likelihood Chi Square = 26.34 with prob < 0.049

Tucker Lewis Index is a model fit statistic. Rule-of-thumb is that models with TLI > .90 fit the data well (Little, 2013, *Longitudinal structural equation modeling*)

Tucker Lewis Index of factoring reliability = 0.985

Root mean square error of approximation (RMSEA) is another model fit statistic. Rule-of-thumb is that models with RMSEA ≤ .05 fit the data well, ≤ .08 fit the data acceptably, and .08 - .10 is marginally acceptable fit (Little, 2013).

RMSEA index = 0.031 and the 90 % confidence intervals are 0.002 0.052

Matrix below gives standardized factor loadings (labeled ML1 – ML3), *communality* (h2), *uniqueness* (u2) for each item, and *complexity* (com).

Standardized loadings (pattern matrix) based upon correlation matrix

	ML2	ML5	ML4	ML1	ML3	h2	u2	com
stmorale	-0.05	0.76	-0.05	0.06	-0.01	0.60	0.400	1.0
press	0.70	0.24	0.05	-0.10	-0.01	0.66	0.335	1.3
tcmorale	0.04	0.72	0.09	-0.03	-0.26	0.69	0.312	1.3
hilearn	0.34	0.32	-0.31	-0.04	0.22	0.58	0.415	3.8
hwexpect	0.75	-0.10	-0.08	0.08	0.16	0.54	0.459	1.2
discipln	0.10	0.04	0.00	0.91	-0.13	1.00	0.005	1.1
structur	0.54	0.13	0.06	0.12	0.06	0.41	0.587	1.3
negtchrs	-0.13	0.09	0.52	0.01	0.29	0.49	0.506	1.8
lowmotiv	0.10	-0.04	0.84	-0.01	0.00	0.65	0.353	1.0
academic	0.70	-0.11	0.04	0.06	-0.06	0.41	0.585	1.1
conflict	0.06	-0.10	0.04	-0.07	0.65	0.49	0.506	1.1
indivlrn	0.63	0.01	0.03	-0.12	-0.16	0.42	0.582	1.2

Matrix below gives estimated correlation between the two latent factors (0.7)

With factor correlations of

	ML2	ML5	ML4	ML1	ML3
ML2	1.00				
ML5	0.69	1.00			
ML4	-0.63	-0.56	1.00		
ML1	0.42	0.30	-0.21	1.00	
ML3	-0.31	-0.37	0.33	-0.11	1.00

 # Possible conclusion for 5-factor model: All of the model fit indices indicate good fit of the model to response data from these school climate items. The five factors might be labeled as Factor 2 = Academic Press, Factor 5 = Morale, Factor 4 = Teacher Pedagogical Skill, Factor 1 = Discipline Emphasis, and Factor 3 = Teacher-Administrator Conflict. None of the correlations between factors are so high as to indicate any pairs of factors that are indistinguishable. The *hilearn* item does not load highly on any of the factors.

 # OVERALL CONCLUSION: Perhaps the 3-factor model, which corresponds well to our original hypothesis, is interpretable, and fits the data acceptably should be viewed as the "preferred" model for parsimony reasons. We could consider using the items to generate three subscores – indicating the school's academic press, morale of school community members, and teacher pedagogical skills for motivation. The 4- and 5-factor models are less parsimonious, but also are well-fitting and interpretable. Parallel analysis suggested considering a 4- or 5-factor model.